27

MATHEMATICAL MODELLING IN ENGINEERING & HUMAN BEHAVIOUR CONGRESS

JULY 9TH TO 11TH, 2025



BOOK OF ABSTRACTS

Modelling for Engineering & Human Behaviour 2025

València, July 9th-11th, 2025

This book includes the extended abstracts of papers presented at XXVII Edition of the Mathematical Modelling Conference Series at the Institute for Multidisciplinary Mathematics *Mathematical Modelling in Engineering & Human Behaviour 2025*.

ISBN: 978-84-09-74933-1

Edited by:

Clara Burgos Simón, Begoña Cantó Colomina, Alicia Cordero Barberó, Juan Carlos Cortés López, Elena López-Navarro, and María Dolores Roselló Ferragud.
I.U. de Matemàtica Multidisciplinar,
Universitat Politècnica de València.

Cover Page Image by Jaume Hervás Raga (http://www.hagamosalgointeligente.com)

Report any problems with this document to imm@imm.upv.es.
Version of July 24, 2025.







Grant CIAORG/2024/182 of Generalitat Valenciana: Conselleria d'Educació, Cultura, Universitat i Ocupació.

Preface

July 2025

This book includes the works presented at the XXVII Edition of the Mathematical Modelling Conference Series at the Institute for Multidisciplinary Mathematics, *the Mathematical Modelling in Engineering & Human Behaviour 2025 Congress*. We thank the ETS d'Enginyeria Informàtica of the Universitat Politècnica de València for hosting the conference in its comfortable building and generously providing access to its facilities from July 9 to July 11, 2025. This edition of the congress is a tribute to Prof. Antonio Hervás Jorge.

The objective of the Mathematical Modelling in Engineering and Human Behaviour conference is to develop an interdisciplinary forum in areas like Medicine, Sociology, Business and Engineering, where the latest mathematical techniques can be discussed in a common and understandable language with experts in cross-disciplinary areas. The conference is aimed at gathering researchers who need mathematics for the formulation and analysis of models.

This year, the following special sessions were organized:

1. Recent advances in Iterative Processes for Solving Nonlinear Problems.
   Organizers: Neus Garrido and Juan Ramón Torregrosa.
2. Mathematical Methods for Engineering Problems.
   Organizers: Damián Ginestar and Antoni Vidal.
3. Infectious Diseases and Mathematical Modeling in Public Health.
   Organizer: Clara Burgos and Almudena Sánchez.
4. Mathematical Models and Artificial Intelligence for Transport Engineering.
   Organizer: Julia Real.
5. Linear algebra, matrix analysis and applications.
   Organizer: José Mas and Alicia Roca.
6. Math & physics.
   Organizer: Màrius Fullana.
7. Uncertainty Quantification and Modelling.
   Organizer: Dolors Roselló and Vicent Romero.
8. Complex networks graphs and applications.
   Organizer: Begoña Cantó.
9. Financial Mathematics
   Organizers: Julia Martínez and M^{a} Lourdes Gómez.
10. Mathematical models in population dynamics
    Organizers: Luis María Abia and Miguel Ángel López-Marcos.
11. Round table: Mathematical Modelling in Civil Enginering
    Organizers: Julia Real and Juan Carlos Cortés.
12. Student's Projects
    A session for students of the double degrees of mathematics and the students of Data Science.

Organization

Mathematical Modelling in Engineering & Human Behaviour (MME&HB 2025) is organized by the Institute of Multidisciplinary Mathematics (IMM) of Universitat Politècnica de València.

Organizing Committee

The organizing committee of the MME&HB 2025 is composed by:

Begoña Cantó Colomina (Chair)	Universitat Politècnica de València
Clara Burgos Simón (Co-Chair)	Universitat Politècnica de València
Francisco Chicharro López	Universitat Politècnica de València
Juan Carlos Cortés	Universitat Politècnica de València
Màrius Josep Fullana i Alfonso	Universitat Politècnica de València
Elena López Navarro	Universitat Politècnica de València
Julia Real Herráiz	Universitat Politècnica de València
Alicia Roca Martínez	Universitat Politècnica de València
Cristina Santamaría Navarro	Universitat Politècnica de València
Antoni Vidal Ferràndiz	Universitat Politècnica de València

Organizers institutions:

- Instituto Universitario de Matemática Multidisciplinar
- Universitat Politècnica de València

Funded by:

- Grant CIAORG/2024/182 of Generalitat Valenciana: Conselleria d'Educació, Cultura, Universitat i Ocupació.

Sponsors:

- BECSA
- Infrakonsult Europa S.L.
- Vialobra S.L.
- Grupo Rover
- Pavasal
- Fecoval
- Wide Communications Solutions S.L.

Table of Contents

I Plenary Speakers

Use of technological tools and algorithms in Sports Science
Sergio Amat 3

The Co-colonization replicator model: from theory to diversity data
Erida Gjini 4

Modeling the Normal Modes of the Moon and Its Application to Gravitational-wave Detection
Xian Chen 5

What if AI already knows more than we do? Learning Analytics in times of algorithmic dependence
Miguel Á. Conde 6

II Recent advances in Iterative Processes for Solving Nonlinear Problems

Efficient Generation of Dynamical Planes for Stability Analysis in Iterative Methods
Francisco I. Chicharro, Alicia Cordero, Neus Garrido and Juan R. Torregrosa 11

Non-Linear MLS Approximation based on the Partition of Unity method in $\mathbb{R}^n$
Inmaculada Garcés, Juan Ruiz-Álvarez, and Dionisio F. Yáñez 16

Dynamical and numerical analysis of classical methods for multiple roots applied to different multiplicities
Eva G. Villalba, Beatriz Campos, Eulalia Martínez, and Pura Vindel 19

Efficiency and st ability of a new hybrid unconstrained optimization algorithm with quasi-Newton updates and high-order methods
Alicia Cordero, Natanael Ureña, Juan R. Torregrosa, and Javier García M. 27

Iterative methods for multiple roots with memory
Francisco I. Chicharro, Neus Garrido, and Julissa H. Jerezano 33

Increasing the Order of Convergence in Jacobian-Free Iterative Schemes: Applications to Nonlinear Systems and Nonlinear Partial Differential Equations
Alicia Cordero, Miguel A. Leonardo Sepúlveda, María P. Vassileva and Juan R. Torregrosa 39

Extending the applicability of a fifth order iterative method for solving nonlinear Hammerstein-type integral equations
Lovish Dua, Sukhjit Singh, Eulalia Martínez, M.A. Hernández-Verón, Ramandeep Behl 45

On a non-linear approximation method: Shepard WENO-based method
David Levin, José M. Ramón, Juan Ruiz-Álvarez, and Dionisio F. Yáñez .. 49

Two New Maximally Efficient Fifth- and Sixth-Order Methods for Nonlinear Vectorial Problems
Alicia Cordero, Renso V. Rojas-Hiciano, Juan R. Torregrosa, and Maria P. Vassileva 52

QUBO formulation for Multi-Marginal Optimal Transport
S. Torregrosa, E. Gravot, and F. Chinesta 60

A first highly efficient eight-order optimal iterative method for solving non-linear systems
A. Cordero, J. R. Torregrosa, and P. Triguero-Navarro 63

A Hybrid Steffensen-Genetic Algorithm for Finding Simultaneous Roots of Nonlinear Equations
Fiza Zafar, Alicia Cordero, Sadia Mujtaba, and Juan R. Torregrosa 69

A Hybrid Approach Using Tuna-Swarm Optimization Algorithm and Newton's Variant for Solving Non-linear Systems
Sonia Bhalla, Aanchal Chandel, Alicia Cordero, Juan R. Torregrosa, and Ramandeep Behl 75

Analysis and applications of conformable, fractional, fractal and global derivatives
Ali Akgül 81

III Mathematical Methods for Engineering Problems

Interpretable machine learning prediction of the coal gross calorific value relied on ultimate analysis for energy systems by using the SVM and MLP techniques
P.J. García–Nieto, E. García–Gonzalo, J.P. Paredes–Sánchez, L.A. Menéndez–García, L . Álvarez–de –Prado, and A. Bernardo–Sánchez 85

Optimization of Defect Detection and Classification in Textiles Using ARM Architectures
T. Albero-Albero, B. Cantó, N. Garrido, and R. Pérez 93

Prior data driven Physics-Informed Neural Networks for solving source and eigenvalue neutron diffusion problems
E. Navarro-Gamón, A. Vidal-Ferràndiz, D. Ginestar, and G. Verdú 99

Decision-making optimization for redevelopment projects in the marble industry
S. Carpitella, F. Carpitella, and J. Izquierdo 107

An extended LUPOD technique to construct a reduced order model for the steady state neutron diffusion equation
A. Vidal-Ferràndiz, A. Carreño, D. Ginestar, and G. Verdú 115

IV Infectious Diseases and Mathematical Modeling in Public Health

Extending the SIS Epidemic Markov Model Including Coinfection Dynamics
Carlos Andreu-Vilarroig, Sten Madec, and Erida Gjini 125

A Computational Framework for Profiling a Human Bladder Cancer scRNA-seq Sample
C. Lara, G. Rubio, C. Santamaría and B. García-Mora 129

Analysis of RSV Hospitalizations in Infants Under 1 Year: Application of Hybrid Digital Twins
Marcos Llamazares, and Rafael-J. Villanueva 134

Modelling and analysis of the dynamics of infectious diseases described by differential systems with uncertainties
Javier López-de-la-Cruz 142

Modeling Co-Infection Patterns of COVID-19 and Dengue
C.Pinto, J.Calatayud and M. Jornet 145

Using AI to Detect and Predict Infectious Disease Seasonalilty and Trends
J.M. Montañana and A. Hervás 148

Asymptotic behavior for a SIR model with random transmission coefficient and time depends vaccination rate
Javier López-de-la-Cruz, Susana Merchán, Felipe Rivero, and Javier Rodrigo 155

Ex vivo mathematical model for including the reapplication of radiofrequency currents in hepatic ablation
Antonio Romero, Jose Bon, Macarena Trujillo 157

A proposal to model the transmission dynamics of the Respiratory Syncytial Virus using an Agent-Based Model
J. Villanueva-Oller, R. J. Villanueva, I. Hidalgo 161

Mathematical modeling of the innate and adaptive immune systems in a viral infection
B. Chen-Charpentier 162

Mathematical and epidemiological model of dengue virus transmission in Ecuador and Puerto Rico
Juan Ángel Aledo, Mario Alberto Espinoza, Juan Carlos Orengo, Rafael-J. Villanueva 168

V Mathematical Models and Artificial Intelligence for Transport Engineering

Railway Superstructure Inspection System For Detecting Ballast Contamination Level Through Acoustic Signal Analysis
Laura Andrés, Julia I. Real, Anabel Arias, and Lisardo Manuel Fort 175

Disease detection system in black truffle mycorrhized holm oak using unmanned aircraft (drones)
Reynaldo Cabezas, Julia I. Real, Gonzalo Muinelo, Vicente Monfort, and Roberto Herrera 181

Technical-mathematical solution for dynamic railway track inspection
M. Isabel Espinós, Julia I.Real, Juan Ramón Torregrosa, and Polina Batalova 186

Multi-Depth Detection System and Maturity State Identification of Black Truffles in Truffle Cultivation Fields
Salvador Mateo, Julia I. Real, Laia Rubio, and Ferney Gómez 191

Automatic pavement condition assessment system for airports
Julia I. Real, Adrián Miralles, Rodrigo Solano, and Alejandro Grandado ... 197

Structural health monitoring of bridge infrastructures using single-pass SAR data
Guillermo Alandí, Julia I. Real , Ernesto Serrano and Alejandro Rodríguez 202

Monitoring and Structural Health Diagnosis of the Spyckstraβe Bridge in Kleve
Jose Cano, Julia I. Real , Juan Ramón Torregrosa and Jorge Pavón 208

Towards a geometric optimization of Bioreceptive Bricks in Vertical Facade Systems
M.J. López, L. Hilario, P. Verdejo, and N. Montes 216

Innovative and intelligent technology for in situ early diagnosis and prediction of concrete carbonation progress for predictive maintenance of bridges
Javier Maldonado, Julia I. Real, Teresa I. Real and José Ignacio Hervás ... 218

Continuous monitoring system for the safe storage and use of hydrogen in hydrogenerators
Alejandro Marco, Julia I. Real , Sara Vicedo, and Rocío Castelo 226

Probabilistic estimation of the stiffness of a structure from indirect measurements
P. Matos Castañeda, F.J. Navarro-Gonzalez, and Y. Villacampa 234

System for detection and prediction of the phenomenon of hydrogen embrittlement
Juan Ramón Sánchez, Julia I. Real, Juan Alberto Puchol, and Elena Caidas 240

VI Linear Algebra, Matrix Analysis and Applications

Exploring Matrix Exponentials via Charlier Polynomials
José M. Alonso, Emilio Defez, Javier Ibáñez, Jorge Sastre, and Álvaro Ibáñez 251

Characterization of the Combined Matrix of a Tridiagonal Toeplitz Matrix
B. Cantó, R. Cantó and A.M. Urbano 258

Koopman operator approach for data driven models to anticipate trends in cybersecurity incident reports
Miguel V. Carriegos, Noemí DeCastro-García, and Ángel Luis Muñoz Castañeda 264

Leveraging Group Actions for Constructing Convolutional Codes with Optimal Decoding and Distance Properties
Noemí DeCastro-García, Miguel V. Carriegos, and Ángel Luis Muñoz Castañeda 266

On a novel rational-polynomial method to approximate the matrix logarithm
E. Defez, J. Ibáñez, J.M. Alonso, J. Sastre, and A. Ibáñez 268

Computing Functions of Matrices with the Dual Scaling-And-Squaring Algorithm
M. Fasi, B. Iannazzo, and J. Sastre 275

Reducing computational complexity in 3D printing optimization: A quotient-based approach to order matrices
P. Alonso, F. Fueyo, P. García-Medina, J. Jiménez, M. L. Serrano, and A. Tepavčević 279

Decoding convolutional codes over finite rings. A linear dynamical systems approach
Ángel Luis Muñoz Castañeda, Noemí DeCastro-García and Miguel V. Carriegos 280

A new method and simple to get the priority vector for reciprocal matrices
J. Benítez, and C. Serra 283

On the relation between the spectra of linear pencils and linear relations
T. Reis, C. Trunk, and H. Winkler 289

VII Maths & Physics

New Neural Network Approaches to Cosmic Microwave Background Mapping
I. García-Prieto, M. D. Guillen and R. Dale 293

An Analysis of Bell's Cosmic Inequalities: From COBE to Planck Data
R. Dale, and J. M. Gandía 297

Computational Tools in Astrophysics
Màrius Josep Fullana i Alfonso, Pau Amaro Seoane, and Josep Vicent Arnau i Córdoba 305

Extension of analytical regularization methods of the integration step of the elliptic two-body problem to the hyperbolic case
José Antonio López Ortí, María José Marttínez Usó, and Francisco José Marco Castillo 311

xIdeal, a Mathematica package for IDEAL characterizations and determinations
S. Mengual, and A. García-Parrado Gómez-Lobo 312

Quantum Logistics Dynamics
Joan C. Micó 314

Synchronization of adaptative Kuramoto network in presence of contrarians
K. Alfaro-Bittner, G. Contreras-Aso, E. Primo, D. Rodríguez, and M. Romance 322

Relativistic Positioning and Dilution of Precision
R. Serrano Montesinos, and J.A. Morales-Lladosa 325

Relativistic Diatomic Molecules in an Exponential-Type Radial Potential
Michael M. Tung and Guillem Gómez i Blanch 329

Modeling the Gravitational-Wave Forest from Galactic and Extragalactic source Populations
Shao-Dong Zhao, Pau Amaro Seoane 333

VIII Uncertainty Quantification and Modelling

A Study of a Random Fractional Differential Equation with jumps
C. Burgos, A. Navarro-Quiles,A. Sánchez-Sánchez, and L. Villafuerte 343

An Approach to Simulating the Dynamics of Stochastic Urban Spatial Network Models
H. de la Cruz, and M. Munoz 347

Stochastic differential calculus for Gaussian and non-Gaussian white noises
R. Laudani, and G. Falsone 348

Computation and convergence of the approximate 1-PDF for a randomized fractional differential equation with impulsive forcing
C. Burgos, A. Navarro-Quiles, and A. Sánchez-Sánchez 355

A Random Differential Equation Approach for Detecting Variations in Growth Patterns of the Atlantic Herring
Jacob Burbank, Juan Carlos Cortés López, Cristina Luisovna Pérez, and Rafael Jacinto Villanueva 359

The use of d-Choquet integrals for differential privacy: experimental evidence
José Carlos R. Alcantud 365

Modeling Uncertainty in Linear Homogeneous Neutral Delay Differential Equations
J.-C. Cortés, G. González-Parra and S.-M. Sferle 372

Probabilistic study of random oscillators with cross-nonlinearities in position and velocity
Juan-Carlos Cortés, José-Vicente Romero, María-Dolores Roselló and Joaquin F. Valencia Sullca 378

Stability Concepts for Random Semi-dynamical Systems
O. Albescu, L.E. Biris, T. Ceausu and I.-L. Popa 383

IX Complex networks graphs and applications

The Plurality of Consensus: Because One Ranking Cannot Rule Them All
Juan A. Aledo, José A. Gámez and Alejandro Rosete 387

First steps to predict extreme events in complex systems
K. Alfaro-Bittner, J. Tredicce, L. Sarrazin, M. Marconi, and M. Giudici ... 389

$\boldsymbol{\Omega}$-vector spaces
Patricia Ferrero, Jorge Jiménez, María Luisa Serrano, Andreja Tepavčević, and Braminir Šešelja 390

Personalization of Page Rank centrality vector: a fixed-point approach
David Aleja, Julio Flores, Eva Primo, Daniel Rodríguez and Miguel Romance 393

Analysis of the Evolution of Student Enrollment Preferences in Degree University Programs Through Directional Community Analysis (2014–2023)
José-Miguel Montañana, Antonio Hervas, and Pedro-Pablo Soriano-Jiménez 397

PageRank with Time-Varying Personalization for temporal networks
David Aleja, Julio Flores, Eva Primo, and Miguel Romance 408

X Financial Mathematics

Bounded versus unbounded good deal indices in asset pricing
Alejandro Balbás 413

Pricing FX contracts with time-delay stochastic processes
L. Gómez-Valle, J. Martínez-Rodríguez and M. Zani 415

Stochastic Log-Normal Model for price prediction with two cryptocurrencies as underlying assets
C.Burgos-Simón , J.F. Sánchez-Jiménez, and A. Sánchez-Sánchez 418

A coupled stochastic model for cryptocurrency prediction: Mathematical modelling and implementation
C. Burgos, J.F. Sánchez-Jiménez and A. Sánchez-Sánchez 422

XI Mathematical models in population dynamics

Efficient natural grid numerical methods for cell population balance models with asymmetric division
L. M. Abia, O. Angulo, J. C. López-Marcos, and M. A. López-Marcos 429

Multi-serotype nested immuno-epidemiological model for dengue hemorrhagic fever involving backward bifurcation and serotype invasion
Mostafa Adimy, Charlotte Dugourd-Camus, and Ruben Taieb 435

Boundary-value problems of functional differential equations with state-dependent delays
A. Andò, and J. Sieber 437

Numerical integration of a structured population model with two physiological structures
L. M. Abia, O. Angulo, J. C. López-Marcos, and M. A. López-Marcos 442

Harvesting Populations in Random Environments: Shortcomings of the Optimal Harvesting Policy, Proposition of Alternatives and their Assessment
Carlos A. Braumann, and Nuno M. Brites 444

Modelling protective behaviour in response to mosquito-borne epidemics
A. Bizzotto, S. De Reggi, A. Pugliese, M. Sabbatino, and C. Soresina 448

Pseudospectral approximation of Floquet multipliers for state-dependent delay differential equations
A. Andò, D. Breda, D. Liessi, and C. Tanase 450

A meta-population model for non-timber forest product harvesting: connectivity, plant viability, and sustainable yield
Maria C.A. Leite, Ruth I. Oliva-Zúniga, and Pedro J. Torres 452

A Free-boundary Nonlocal Diffusion Logistic Model: Analysis and Computing
V. Egorova, M.-C. Casabán, R. Company, L. Jódar and M. El Fakharany . 458

Vulnerability and risk of forest ecosystems in the face of global change: Design of an integrated observation and modelling system for adaptation
Miguel A. Zavala, Julián Tijerin, Marina Rodes-Blanco, Mariano García, Inmaculada Aguado, and Paloma Ruiz Benito 466

XII Student's Projects

Mathematical modelling of cranial parameters in Ugandan infants: quantitative analysis of normality indices and their implication for infant cranial deformation
Marta Muñoz Domingo, Marc Romeu Ferras, José Luis Lerna and David Vivas Consuelo 471

Introduction to Bifurcations in Two-Dimensional Systems of Differential Equations in Artificial Neuron Models
Damián Oussa Vañó Fernández, Xuan Liu and Antonio Sineiro Abal 479

Machine Learning-Based Prediction of Percutaneous Nephrolithotomy Outcomes in a Large General Hospital
Laura Shalabayeva1, Pilar Bahílo Mateu , Marc Romeu Ferras, Javier Díaz Carnicero, Alberto Budía and David Vivas-Consuelo 487

Part I

Plenary Speakers

Use of technological tools and algorithms in Sports Science

Sergio Amat Plata[1,*]

(1) Universidad Politécnica de Cartagena, Spain
Pza. del Cronista Isidoro Valverde, Edif. La Milagrosa
C.P 30202 Cartagena

This work aims to explore the use of technological tools and algorithms to analyze athletic performance across various sports disciplines. The study focuses on assessing both the physiological and technica behavior of athletes through data collected via electronic devices and processed using advanced, custom-developed techniques [1], [2], [3].

In canoeing, the analysis will center on the evolution of speed throughout the course to identify the point at which cruising speed is reached (linked to power), the stability of that pace (related to technique), and final variations (associated with endurance).

High-Intensity Interval Training (HIIT) will also be examined by analyzing heart rate peaks and troughs, as well as the slopes between them. This will help assess both the correct execution of the training protocol and the athlete's physical condition over time.

In team sports such as futsal, heart rate data will be used to identify different phases of gameplay (rest, activation, exercise, and recovery), aiming to optimize substitution strategies and improve overall performance.

These analyses aim to provide actionable insights that enhance training efficiency, optimize performance, and support evidence-based decision-making in sports science.

References

[1] Manchado, C., Tortosa Martínez, J., Pueo, B., Cortell Tormo, J.M., Vila, H., Ferragut, C., Sánchez Sánchez, F., Busquier, S., Amat, S., & Chirosa Ríos, L.J. High-Performance Handball Player's Time-Motion Analysis by Playing Positions. International Journal of Environmental Research and Public Health. 2020; 17(18):6768. https://doi.org/10.3390/ijerph17186768.
[2] Rico González M., Pino Ortega, J., Clemente F.M., Rojas Valverde D., & Los Arcos A. A systematic review of collective tactical behavior in futsal using positional data. Biology of Sport. 2021; 38(1):23-36. https://doi.org/10.5114/biolsport.2020.96321.
[3] Buchheit, M., & Laursen, P. B. (2013). High-intensity interval training, solutions to the programming puzzle. Sports Medicine, 43(5), 313-338. https://doi.org/10.1007/s40279-013-0029-x
[4] Gomez-Carmona, C. D., & Pino-Ortega, J. (2021). Applications of wearable inertial measurement units in performance analysis of team sports: A systematic review. European Journal of Sport Science, 21(4), 475-487. https://doi.org/10.1080/17461391.2020.1769550

* sergio.amat@upct.es

The Co-colonization replicator model: from theory to diversity data

Erida Gjini[1, *]

(1) CEMAT, Instituto Superior Tecnico,
University of Lisbon, Portugal.

Despite extensive research, modelling and understanding biodiversity remains a challenge across fields such as ecology, epidemiology and microbiology. In the last years, we have approached biodiversity by developing and studying an N-strain SIS system with co-colonization and interactions, leading to a replicator equation [4]. Using mathematical techniques, we have simplified this high-dimensional eco-epidemiological system, finding organizing principles for stability, complexity and coexistence, and parallels across biological scales: from microbial transmission at the population level to multi-species microbiota dynamics within host. In this talk, I will present the model, its key properties, and describe 2 ways to link our replicator framework to data: i) in an epidemiological context [3] and ii) in a microbiological [1] context.

First, we apply the framework of replicator dynamics to 5 datasets of Streptococcus pneumoniae serotype frequencies in different countries. We estimate pairwise strain interactions from population-level frequencies across different endemic settings [3]. In another work, at a different biological scale, we link ecological invasion dynamics to our explicit replicator equation model [2, 4], and quantify selection coefficient between two invaders relative to their host multi-species composition. The method is illustrated on invasion data from E.coli strains in mice with different microbiota [1], but is applicable to other microbiota and co-invasion scenarios. Overall our work opens new theoretical avenues for investigation and fresh perspectives for applying this replicator equation to biological data.

References

[1] Tomas Ferreira Amaro Freire, Sten Madec, and Erida Gjini. Unpacking fitness differences between two invaders in a multispecies context. bioRxiv, pages 2024–11, 2024.
[2] Erida Gjini and Sten Madec. Towards a mathematical understanding of invasion resistance in multispecies communities. R. Soc. Open Sci., 10:231034, 2023.
[3] Thi Minh Thao Le, Sten Madec, and Erida Gjini. Inference of pairwise interactions from strain frequency data across settings and context-dependent mutual invasibilities. (in presss) Bulletin of Mathematical Biology, 2025.
[4] S. Madec and E. Gjini. Predicting N-strain coexistence from co-colonization interactions: Epidemiology meets ecology and the replicator equation. Bulletin of Mathematical Biology, 82:142, 2020.

* erida.gjini@tecnico.ulisboa.pt

Modeling the Normal Modes of the Moon and Its Application to Gravitational-wave Detection

Xian Chen[1,2 *]

(1) Department of Astronomy, School of Physics, Peking University,
Yi He Yuan Lu 5, 100871 Beijing, P. R. China.
(2) Kavli Institute for Astronomy and Astrophysics, Peking University
Yi He Yuan Lu 5, 100871 Beijing, P. R. China

The growing interest in detecting gravitational waves (GWs) via lunar seismic measurements requires a precise understanding of the Moon's response to passing GWs. Previous studies derived two seemingly different response functions—one using a field-theory approach and the other based on tidal forces—raising questions about their equivalence. Here, we analytically and numerically model the normal modes of the Moon excited by GWs. We demonstrate that the aforementioned functions are identical, with differences arising only from coordinate choices. Using the correct response function, we reassess the sensitivity of proposed lunar seismometers (e.g., China's Chang'e and Europe's Lunar GW Antenna), revealing flatter sensitivity curves between 10^{-3}–0.1 Hz than predicted by previous models. Additionally, we reevaluate the feasibility of constraining the stochastic GW background with lunar seismometer networks, deriving updated pattern and overlap reduction functions while relaxing idealized instrument assumptions. To address modeling challenges posed by the Moon's heterogeneous structure, we conduct a 2D finite-element simulation using the spectral element method, validating it against semi-analytical solutions and analyzing its limitations. Our work advances the accurate simulation of lunar GW responses, clarifies scientific objectives for lunar GW observations, and provides critical constraints for future detector design.

References

[1] Yan, H., Chen, X., Zhang, J., Zhang, F., Wang, M., & Shao, L., Toward a consistent calculation of the lunar response to gravitational waves. *Physical Review D*, 109(06):4092, 2024.

[2] Yan, H., Chen, X., Zhang, J., Zhang, F., Shao, L., & Wang, M., Constraining the stochastic gravitational wave background using the future lunar seismometers. *Physical Review D*, 110(04):3009, 2024.

[3] Zhang, L., Yan, H., Chen, X., Zhang, J., 2D numerical simulation of lunar response to gravitational waves using finite element method. *Physical Review D*, 111(06):3014, 2025.

* xian.chen@pku.edu.cn

What if AI already knows more than we do? Learning Analytics in times of algorithmic dependence

Miguel Á. Conde[1 *]

(1) Universidad de Salamanca,
Escuela Politécnica Superior de Zamora,
Av. de Requejo, 33, 49029 Zamora, Spain.

The promise of Learning Analytics (LA) has always been tied to its ability to offer empirical evidence to support decision-making in educational contexts, with the challenging goal of improving student learning [1]. However, the current landscape is changing at an astonishing pace. Learning data is increasingly generated and stored outside institutional environments. External platforms, intelligent assistants, AI-based tutoring systems, and interactive content generators collect information that falls beyond the traditional reach of LA, blurring the boundaries of educational analysis.

This new scenario demands urgent and profound responses from multiple perspectives. First, it will be necessary to define standards that allow the capture and reuse of data generated outside the institutional ecosystem, while preserving its analytical value. At the same time, learning platforms must evolve to offer genuine personalization, delivering to each learner what they "need" at just the right time. Furthermore, the door opens to the integration of intelligent tutors capable of intervening in real time during specific learning tasks. Perhaps most disruptively, Explainable Artificial Intelligence (XAI) emerges strongly [2, 3]: systems that not only predict and recommend, but also justify their decisions in a comprehensible manner—at least, apparently so. This introduces a deep epistemological shift: are humans still the ones interpreting data and making educational decisions, or are we silently handing over that responsibility to opaque yet seductively efficient systems?

This work offers a critical reflection on this evolving context, marked by a growing dependence on algorithmic models that, in many cases, operate beyond our pedagogical oversight. The implications of this transition will be explored: from a mapping of emerging possibilities, through the loss of institutional data sovereignty, to the risks of blindly trusting explanations generated by systems that prioritize optimization over educational understanding.

At the same time, genuine opportunities will be discussed: the potential for radical personalization, the democratization of access to analytical tools, and the rise of a more distributed—yet also more contextualized—Learning Analytics.

Ultimately, one uncomfortable but necessary question remains: if AI already knows (or seems to know) more than we do, what is the new role of Learning Analytics—and what is ours, as educators and researchers

References

[1] Long, P. D., Siemens, G., Penetrating the Fog: Analytics in Learning and Education, *EDUCAUSE Review,* 46(5), 31-40, September 2011. [Online]. Available: http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education.

[2] Došilović, F. K., Brčić, M., Hlupić, N. Explainable artificial intelligence: A survey, *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, 2018: IEEE, 0210-0215.

[3] Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., Atkinson, P. M., Explainable artificial intelligence: an analytical review, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 11(5), 2021.

Part II

Recent advances in Iterative Processes for Solving Nonlinear Problems

Efficient Generation of Dynamical Planes for Stability Analysis in Iterative Methods

Francisco I. Chicharro[1], Alicia Cordero, Neus Garrido and Juan R. Torregrosa

I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

Iterative methods give an approximate solution of equation $f(x) = 0$, $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^n$ when analytical techniques are not able to solve this problem. Root-finding methods design has been a hot topic in the area of numerical analysis for the last decades.

The quality of an iterative method can be interpreted in several ways. On the one hand, the analytical quality involves the order of convergence, the number of function evaluations, the efficiency index, or the optimality, amongst others. Stability quality refers to the number of initial guesses converging to the expected root. Fundamentals of complex dynamics on rational operators can be found at [1, 2]. Several tools facilitate the comprehension of the dynamics of a family of methods. One example is the convergence plane [3, 4], which represents the real dynamical behaviour of a process that includes a parameter. Another example is the isonormal surface [5, 6], which gives information about regions without convergence to the desired roots. However, the most used representation is the dynamical plane.

A dynamical plane is a graphical representation of an iterative method's behaviour when applied to a low-degree complex polynomial. Given a $F \times C$ matrix of initial guesses, each element is iterated until it converges or until a maximum number of iterations M is reached, as stated in [11, 8]. In the worst case, $F \cdot C \cdot M$ iterations are needed. Once the orbit of each initial estimate has been checked, they are classified according to which basin of attraction they have converged to, assigning a different colour to each basin and reserving the colour black for other cases.

In this work, we aim to improve the execution time of the generation of dynamical planes.

2 Materials and Methods

The proposal is based on the ability of specific computer-aided software to work in vector form rather than scalar form. Using vectors instead of scalars improves computational efficiency due to several factors related to hardware optimisation and memory management. This approach is crucial in processing large datasets, such as iterative methods, machine learning, and numerical simulations. The main reason is the reduction of loop overhead, since using a loop to operate on scalars, the CPU has to decode loop instructions, evaluate stopping conditions and manage indices and jumps.

[1] frachilo@mat.upv.es

The initial guess matrix $(z_0)_{F\times C}$ is reshaped into a single-column vector of nodes $(\tilde{z}_0)_{F\cdot C\times 1}$. With this implementation, the rational operator associated with the iterative method can be applied to the entire vector instead of looping through the matrix element by element. Note that the rational operator could be applied over the entire matrix and have an equivalent effect.

The key point is indexing with a boolean flag vector. If every initial estimation converges in the same number of iterates, the flag vector includes memory overhead. However, the typical situation is that convergence is sparse, so the small overhead included with the flag vector is considerably compensated by the faster convergence of the complete algorithm.

In the proposal, when a point $z_{k_{f,c}}$ satisfies the stopping criterion, it stops iterating. Therefore, at each step, a different number of values of $z_{k_{f,c}}$ is evaluated on the rational operator, decreasing or holding as the iterate k progresses. This fact is controlled with a flag vector whose state changes when the stopping criterion is satisfied. This vector modifies a variable-length vector that contains the nodes $\tilde{z}_{k_j}$ where the rational operator is still applied.

3 Results

The high-performance generation of dynamical planes is tested via two iterative methods to evaluate the time spent generating the dynamical planes.

We are comparing the proposed high-performance generation of the dynamical planes with the code whose guidelines were introduced in 2013 [8]. For the comparison, we only consider the time it takes to obtain the endpoints of the orbit of each initial estimate, leaving aside the time to represent the dynamic plane since this part of the code is identical. The values of times are collected using the Profiler App of Matlab, a built-in tool used to analyse the code's performance and identify which parts of the script or function take the most time to run. In particular, it analyses the execution time, measuring the time spent on each function and line of code, and shows which parts of the code take the longest to execute. The metric used to compare the results is the mean execution time because the execution time of the same code can vary between runs due to factors such as background processes, caching and memory use or thermal throttling, amongst others [13]. For this purpose, each program has been executed 100 times, and the values displayed in the summary tables correspond to mean values.

The first comparison is based on Newton's method

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, \ldots,$$

when it is applied on the general quadratic polynomial $p(z) = (z-a)(z-b)$. The resulting rational operator is

$$R(z) = \frac{z^2 - ab}{2z - a - b}.$$

Applying Möbius transformation $M(z) = \frac{z-a}{z-b}$, we can find that

$$O(z) = \left(M \circ R \circ M^{-1}\right)(z) = z^2,$$

is conjugated with $R(z)$, and it satisfies Cayley's Test [7].

The fixed points of operator $O(z)$ satisfy

$$O(z) = z \leftrightarrow z^2 = z \leftrightarrow z(z-1) = 0,$$

so $z_0^* = 0$ and $z_1^* = 1$ are fixed points. The asymptotical behaviour can be found evaluating the fixed points on $O'(z) = 2z$. Since $O'(z_0^*) = 0$, z_0^* is a superattracting fixed point. Since $O'(z_1^*) = 2$, z_1^* is a repelling fixed point.

Following [12], we can classify infinity as a fixed point of an iteration function $O(z)$ as $\lim_{z\to 0} \frac{1}{O\left(\frac{1}{z}\right)} = 0$, so $z^*_\infty = \infty$ is a fixed point. Furthermore, since $\lim_{z\to 0} \frac{1}{O'\left(\frac{1}{z}\right)} = 0$, then z^*_∞ is a superattracting fixed point.

Therefore, operator $O(z)$ has two superattracting fixed points $z^*_0 = 0$ and $z^*_\infty = \infty$, that are the mapping of $z = a$ and $z = b$ of the original rational operator $R(z)$.

The tests are performed using MatlabR2022b on a computer equipped with Intel Core i7-11700 at a frequency of 2.5 GHz with 16 GB of RAM. A mesh of 200×200 initial guesses are chosen to find the final orbit with stopping criterion $|z_k - z^*_0| < 10^{-9}$ or $|z_k - z^*_0| > 10^9$.

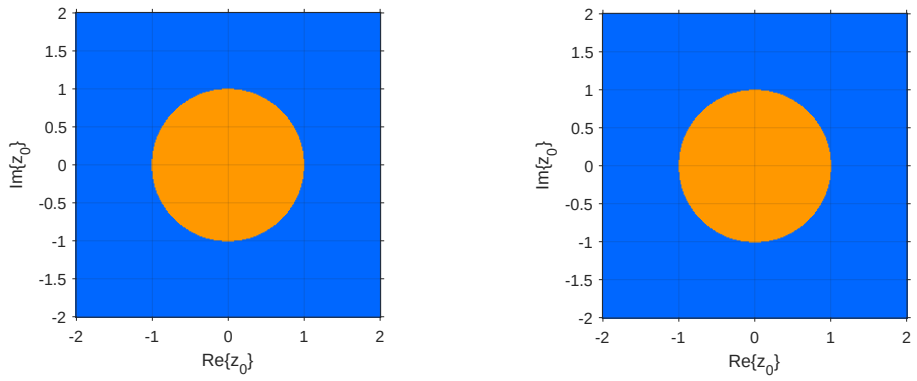
Table 1 collects the information about the times and executions to generate the dynamical planes concerning $O(z)$.

	Reference code		High performance code	
	Average time (ms)	% of total time	Average time (ms)	% of total time
Rational operator	412.46	86.2%	3.33	55.1%
Stopping criterion	22.57	4.7%	5.14	35.7%
Rest of the code	43.38	9.1%	0.86	9.2%
Total	478.41	100%	9.33	100%

Table 1: Time performance for $O(z)$

The performance improvements based on the data in Table 1 are undeniable. The total time with the high-performance code is 1.95% of the time of the reference code. One central insight is the number of calls to the rational operator and the stopping criterion. In the reference code, there are 411444 calls, while in the high-performance code, there are 20 calls. On the other hand, the execution times for the rational operator and the stopping criterion are more balanced in the proposed code. The rest of the code has similar runtime percentages for both versions.

The result of the representation of the dynamic planes is shown in Figure 1. The basin of z^*_0 is represented in orange, while the basin of z^*_∞ is represented in blue. Note that precisely the same representations are obtained by generating the endpoints of the initial estimates with the proposed code (Figure 1a) and with the reference code (Figure 1b).



(a) (b)

Figure 1: Dynamical planes of $O(z)$ generated with (a) the proposed code (b) the reference code

The second comparison is based on Jarratt's parametric family [9] applied on quadratic polynomials of the form $p(z) = (z-a)(z-b)$. The resulting rational operator is conjugated with

$$P(z) = z^4 \frac{23 + 27z + 9z^2}{9 + 27z + 23z^2}$$

after applying Möbuis transformation.

The fixed points of operator $P(z)$ satisfy

$$P(z) = z \leftrightarrow z\left(9 + 27z + 23z^2 + 23z^3 + 27z^4 + 9z^5\right) = 0,$$

so $z_0^* = 0$ and the roots z_p^* of $9 + 27z + 23z^2 + 23z^3 + 27z^4 + 9z^5$, $p = 1, \ldots, 5$, are fixed points. Evaluating the fixed points on

$$P'(z) = \frac{2z^3\left(414z^4 + 1539z^3 + 2230z^2 + 1539z + 414\right)}{\left(23z^2 + 27z + 9\right)^2},$$

we obtain that $P'(z_0^*) = 0$ and $|P'(z_p^*)| > 1$, $p = 1, \ldots, 5$, so z_0^* is a superattracting fixed point and z_p^*, $p = 1, \ldots, 5$ are repelling fixed points. Regarding the infinity, $\frac{1}{P\left(\frac{1}{z}\right)} = P(z)$, so z_∞^* is a fixed point whose behaviour is superattracting. Let us remark that fixed points $z_0^* = 0$ and $z_\infty^* = \infty$ are the mapping of $z = a$ and $z = b$ of the original rational operator, as in the previous case.

Table 2 shows the information about the times and executions after 100 runnings of the codes.

	Reference code		High performance code	
	Average time (ms)	% of total time	Average time (ms)	% of total time
Rational operator	405.12	85.8%	3.31	35.8%
Stopping criterion	28.78	6.1%	5.13	55.4%
Rest of the code	38.50	8.1%	0.81	8.8%
Total	472.40	100%	9.25	100%

Table 2: Time performance for $P(z)$

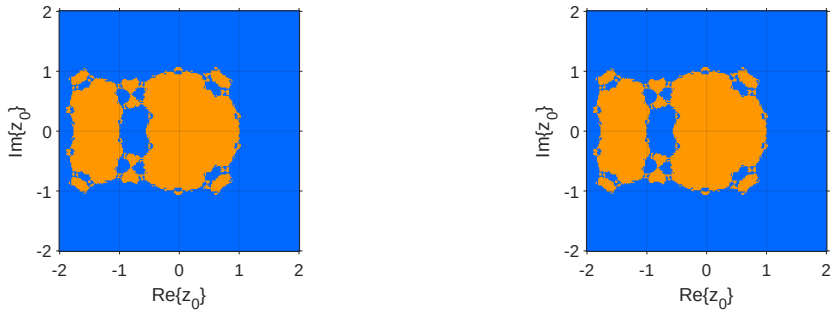
Table 2 shows a similar behaviour concerning the dynamical planes generated for $O(z)$. In this case, the high-performance code is 1.96% of the time of the reference code. Regarding the number of calls to the rational operator and the stopping criterion, the reference code executes 411444 calls, while the high-performance code executes 20 calls. Figure 2 represents the dynamical planes generated with both codes.

4 Conclusion

A new proposal for the generation of dynamical planes has been introduced. The high-performance code is based on a flag vector variable that controls the nodes that have not converged yet. The reference and the proposed code have been compared in terms of execution time with two different rational operators. The results evidence the improvement in the execution time, showing that the new code uses less than 2% of the time regarding the reference code. Some improvements can also be applied, such as reducing half the execution time when the operators have symmetry properties [10].

Acknowledgements

Funded with Ayuda a Primeros Proyectos de Investigación PAID-06-23 and PAID-11-24, Vicerrectorado de Investigación de la Universitat Politècnica de València (UPV).



(a) (b)

Figure 2: Dynamical planes of $P(z)$ generated with (a) the proposed code (b) the reference code

References

[1] Beardon, A. F., Iteration of Rational Functions: Complex Analytic Dynamical Systems. Springer, 1991.

[2] Devaney, R. L., An Introduction to Chaotic Dynamical Systems. Addison-Wesley Publishing Company, 1989.

[3] Magreñán, Á. A., A new tool to study real dynamics: The convergence plane *Appl. Math. Comput.*, 248:215–224, 2014.

[4] Chicharro, F. I., Cordero, A., Garrido, N., Torregrosa, J. R., Impact of stability by the use of memory in Traub-type schemes Math., 8:274, 2020.

[5] Capdevila, R. R., Cordero, A., Torregrosa, J. R., Isonormal surfaces: A new tool for the multidimensional dynamical analysis of iterative methods for solving nonlinear systems *Math. Meth. Appl. Sci.*, 45(6):3360–3375, 2022.

[6] Erfanifar, R., Hajarian, M., High-efficiency parametric iterative schemes for solving nonlinear equations with and without memory, *J. Complexity*, 86: 101896, 2025.

[7] Cayley, A., The Newton-Fourier Imaginary Problem *Amer. J. Math.*, 2(1):97, 1879.

[8] Chicharro, F. I., Cordero, A., Torregrosa, J. R., Drawing Dynamical and Parameters Planes of Iterative Families and Methods *Sci. World J.*, 780153, 2013.

[9] Cordero, A., Segura, E., Torregrosa, J. R., Behind Jarratt's Steps: Is Jarratt's Scheme the Best Version of Itself? *Discrete Dynamics in Nature and Society*, 8840525, 2023.

[10] Campos, B., Canela, J., Vindel, P., Dynamics of Newton-like root finding methods *Numerical Algorithms*, 93:1453–1480, 2023.

[11] Varona, J. L., Graphic and numerical comparison between iterative methods *Math. Intelligencer*, 24:37–46, 2002.

[12] Gutiérrez, J. M., Plaza, S., Romero, N., Dynamics of a fifth-order iterative method *Int. J. Comput. Math.*, 89:822–835, 2012.

[13] Manolache, S., Eles, P., Peng, Z. Real-Time Applications with Stochastic Task Execution Times: Analysis and Optimisation. Springer Dordrecht, 2007.

Non-Linear MLS Approximation based on the Partition of Unity method in $\mathbb{R}^n$

Inmaculada Garcés[♭],[1] Juan Ruiz-Álvarez[♮] [2] and Dionisio F. Yáñez[♭] [3]

(♭) Departamento de Matemáticas, Universidad de Valencia,
Carrer del Dr. Moliner, 50, 46100 Burjassot, Valencia, Spain.

(♮) Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena,
Plaza Cronista Isidro Valverde, 30202 Cartagena, Murcia, Spain.

1 Introduction

The approximation of data is a fundamental task arising in diverse areas such as computer-aided geometric design, numerical solutions of partial differential equations, and the modeling of curves and surfaces. While many approximation techniques perform well under smooth data assumptions, challenges arise when the data contain discontinuities. Among the established techniques, the Moving Least Squares (MLS) method is a widely adopted and robust strategy for data fitting, with applications spanning statistics and applied mathematics. Nonetheless, in the presence of discontinuities, MLS suffers from well-known artifacts, such as the Gibbs phenomenon, that reduce its approximation quality (see [2]).

In this work, we extend previous efforts to integrate MLS with the Weighted Essentially Non-Oscillatory (WENO) method (see, e.g., [3]) and a novel partition of unity framework (see [4]), originally explored in one dimension in [1], to higher-dimensional settings. We propose a nonlinear approximation operator based on the Non-Linear Partition of Unity - Moving Least Squares (NLPU-MLS) method in $\mathbb{R}^n$. This approach enhances accuracy near discontinuities while preserving high-order convergence in smooth regions. We establish theoretical properties of the method and provide numerical evidence demonstrating its performance and reliability.

2 Method

In numerous practical scenarios, data may present abrupt changes or steep gradients. Such situations typically lead to oscillatory artifacts reminiscent of the Gibbs phenomenon and to excessive smoothing near discontinuities. To address these challenges, the Weighted Essentially Non-Oscillatory (WENO) method has been extensively studied and developed in recent years, with a rich body of literature dedicated to its analysis.

WENO constructs a high-order approximation by applying a nonlinear weighting to several Lagrange interpolants of fixed degree. Its core strength lies in dynamically adjusting these weights based on the local regularity of the data, as measured by so-called smoothness indicators. These

[1] inmaculada.garces@uv.es
[2] juan.ruiz@upct.es
[3] dionisio.yanez@uv.es

indicators allow WENO to detect regions of limited regularity and adapt the interpolation accordingly.

In this work, we propose a novel nonlinear approximation operator by combining the WENO strategy with the Moving Least Squares (MLS) framework. The resulting method is capable of suppressing Gibbs-like oscillations near discontinuities while maintaining high accuracy in smooth regions.

2.1 Partition of Unity method

The partition of unity method begins with a covering $\{\Omega_j\}_{j=1}^M$ of the domain Ω, where the sets Ω_j overlap mildly. For this covering, we construct a set of continuous, nonnegative, compactly supported functions $\{w_j\}$ satisfying

$$\sum_{j=1}^{M} w_j = 1 \text{ on } \Omega \quad \text{and} \quad \operatorname{supp}(w_j) \subseteq \Omega_j.$$

Each subdomain Ω_j is associated with an approximation space V_j, and the function f is locally approximated by $s_j \in V_j$. The global approximation is then defined as the weighted combination

$$s_f = \sum_{j=1}^{M} s_j w_j.$$

To be more precise, let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain, and let $\{\Omega_j\}_{j=1}^M$ be a collection of open and bounded subsets such that $\Omega \subseteq \bigcup_{j=1}^M \Omega_j$. Define $\delta_j = \operatorname{diam}(\Omega_j) = \sup_{x,y\in\Omega_j} \|x-y\|_2$. A collection of nonnegative functions $\{w_j\}_{j=1}^M$ with $w_j \in C^k(\mathbb{R}^n)$ is called a *k-stable partition of unity* subordinate to the covering $\{\Omega_j\}$ if the following conditions hold for all $1 \le j \le M$:

1. The support of each function is contained in its corresponding set: $\operatorname{supp}(w_j) \subseteq \Omega_j$,
2. The functions form a partition of unity on Ω: $\sum_{j=1}^M w_j \equiv 1$ on Ω,
3. For every multi-index $\alpha \in \mathbb{N}_0^n$ with $|\alpha| \le k$, there exists a constant $C_\alpha > 0$ such that

$$\|D^\alpha w_j\|_{L^\infty(\Omega_j)} \le \frac{C_\alpha}{\delta_j^{|\alpha|}},$$

for all $1 \le j \le M$.

Now, we want to solve a MLS problem based on this method.

2.2 Integrating MLS with the Partition of Unity method

Finally, we explain briefly the non-linear method based on partition of unity. The idea is quite simple, to divide the domain into several subdomains, to measure the "smoothness" of the data in each subdomain, and to use only those which are free of discontinuities.

We suppose an open domain $\Omega \subset \mathbb{R}^n$, some data points $\{x_i\}_{i=1}^N$ with h the fill distance, and some associated set $\{f_i = f(x_i)\}_{i=1}^N$ with $f : \Omega \to \mathbb{R}^n$ an unknown function. Let $x \in \mathbb{R}^n$ be a point of the domain, and we suppose that the compact support of w is c. We consider that $\{\tilde{x}_k\}_{k=1}^m \subset \mathbb{R}^n$ are some center points and call

$$\Omega_k = \mathcal{B}(\tilde{x}_k, \varepsilon) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \|(x_1, \ldots, x_n) - \tilde{x}_k\|_2 < \varepsilon\}, k = 1, \ldots, m$$

where m is the number of subsets chosen to do the partition of unity, $\varepsilon > 0$ is a constant to determine the radius of the balls. Our goal is to solve a MLS problem for each k

$$p_k = \underset{p \in \Pi_d}{\arg\min} \sum_{x_i \in \Omega_k} \left(p(x_i) - f(x_i)\right)^2 w\left(\|x - x_i\|_2\right).$$

3 Results

In numerical experiments, we have applied this method and achieved the following results. For continuous piecewise functions, we observe that in the results obtained no unwanted effects appear near the discontinuity. In the case of having our data with noise, we obtain noise elimination depending on the weight function we use. The order and continuity of this method will depend on this fact.

4 Conclusions

This work focuses on adapting the well-established partition of unity method, commonly employed with radial basis functions, to the framework of Moving Least Squares (MLS), incorporating a nonlinear strategy to handle cases where the data exhibit discontinuities. To introduce data dependence into the algorithm, we utilize the WENO technique, which selects smooth subsets of data based on the computation of smoothness indicators. We detail the development of this new algorithm, including the design of smoothness indicators tailored to our context. Additionally, we provide theoretical results concerning key properties of the method, such as smoothness, polynomial reproduction, and accuracy order. The numerical experiments included support these theoretical results and confirm the practical performance of the proposed approach.

Bibliography

[1] Garcés, I., Ramón, J.M., Ruiz-Álvarez, J., Yáñez, D.F., Integrating Moving Least Squares with non-linear WENO method: A novel Partition of Unity approach in 1D. *arXiv preprint*, arXiv:2501.14536, 2025.

[2] Amat, S., Levin, D., Ruiz-Álvarez, J., Yáñez, D.F., Non-linear WENO B-spline based approximation method. *Numerical Algorithms*, 100(10): 1–29, 2024.

[3] Baeza, A., Bürger, R., Mulet, P., Zorío, D., Central WENO schemes through a global average weight. *Journal of Scientific Computing*, 78: 499–530, 2019.

[4] Fasshauer, G.E., Meshfree Aproximation Methods with Matlab. Singapore, World Scientific Publishing, 2007.

[5] Wendland, H., Scattered Data Approximation. Cambridge, Cambridge University Press, 2004.

Dynamical and numerical analysis of classical methods for multiple roots applied to different multiplicities

Eva G. Villalba[♭],[1] Beatriz Campos[♮], Eulalia Martínez[♯],[2] and Pura Vindel[♮].

(♭) Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de la Rioja.
Av. de la Paz, 137, Logroño, Spain.

(♮) I. de Matemáticas y Aplicaciones, Universitat Jaume I,
Avinguda de Vicent Sos Baynat, s/n, Castelló de la Plana, Spain.

(♯) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

In this paper we study the classical methods for multiple roots of Newton, Chebyshev and Halley defined in [1], [2] and [3], respectively. The iterative scheme that defines them depends on the parameter m, which represents the multiplicity of the root we want to approximate. One of the disadvantages of using them lies in the fact that we have to adapt the expression of the schemes if we want to approximate solutions with different multiplicities. For this reason, we also work with Schröder's method (see [1]), which approximates the multiple roots of a function without using the multiplicity m in its iterative scheme.

Thus, our main purpose in this work is to check the robustness of the mentioned methods when they are applied to functions whose roots have different multiplicities. To this end, we carry out an exhaustive study of the complex dynamics of the methods on the following polynomial:

$$p(z) = z^m(z-1)^n, \quad z, m \in \mathbb{C}. \tag{1}$$

2 Dynamical study of the methods

In this section we carry out a dynamical study of Newton, Chebyshev, Halley and Schröder methods when they are applied on polynomial (1).

2.1 Newton's method

Newton's method is one of the most commonly used and popular methods for approaching the simple roots of non-linear equations because it is an optimal, efficient and simple method and has quadratic order of convergence. Furthermore, it is one of the first methods which was modified in order to be used to approximate multiple roots of an equation. Its iterative scheme is given by:

$$z_{n+1} = z_n - m\frac{f(z_n)}{f'(z_n)}. \tag{2}$$

[1] egarvil@posgrado.upv.es
[2] eumarti@mat.upv.es

Although the introduction of m in his iterative scheme, in [1] it is proved that it conserves the quadratic order of convergence of the original Newton's method. Nevertheless, this scheme has many limitations since, its very definition for approximating unknown roots implies knowing a priori the multiplicity of these roots. Moreover, it is defined for a single multiplicity m, while the roots of an equation can have different multiplicities.

For this reason, in this section we study the effectiveness of the method for approximating multiple roots with different multiplicities. To do so, we study the dynamics of Newton's method for multiple roots when we apply the scheme on the polynomial (1). The operator we obtain has the form

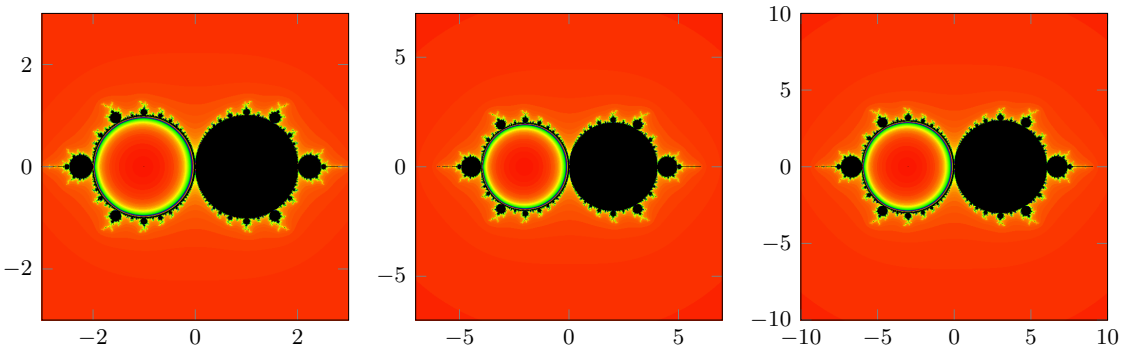
$$O(z,m,n) = \frac{nz^2}{(m+n)z - m}. \tag{3}$$

Firstly, we have to calculate the fixed and the critical points of the operator $O(z,m,n)$ and obtain the parameter and dynamical planes.

Fixed points are defined as the points satisfying $O(z,m,n) = z$. Solving this equation, we obtain the fixed points $z = 0$, $z = \infty$ and $z = 1$. The point $z = \infty$ is called a *strange fixed point*, as it does not coincide with a root of the polynomial p.

On the other hand, we obtain the critical points $z = 0$ and $z = \dfrac{2m}{m+n}$. As $z = 0$ is a fixed point, there is only one free critical point.

Following the iterations of this free critical point we obtain the parameter planes for fixed real values of n and considering m as complex parameter (see Figure 1).



(a) n=1 (b) n=2 (c) n=3

Figure 1: Parameter planes of Newton method for values of $n = 1, 2, 3$.

Now, we obtain the stability of the fixed points analysing the value of the multipliers.

Firstly, the fixed point $z = 0$ is a superattractor for every value of the parameters m and n.

The stability of the other fixed points is studied in the following lemmas.

Lemma 2.1. *The fixed point $z = \infty$ is an attractor inside the circle C_1 defined by*

$$(\alpha + n)^2 + \beta^2 = n^2,$$

with $m = \alpha + i\beta$ and $n \in \mathbb{R}$. In particular, it is a superattractor for $m = -n$. This point is indifferent on the circle C_1 and is a repulsor outside the circle C_1.

If we focus on the real line, then $z = \infty$ is attractor for $-2n < m < 0$, indifferent for $m = 0$ and $m = -2n$ and repulsive in other case. As we consider m and n as multiplicities, the point $z = \infty$ is always a repulsive fixed point.

Lemma 2.2. *The fixed point $z = 1$ is an attractor inside the circle C_2 defined by:*

$$(\alpha - n)^2 + \beta^2 = n^2,$$

with $m = \alpha + i\beta$ and $n \in \mathbb{R}$. In particular, it is a superattractor for $m = n$. This point is indifferent on the circle C_2 and is a repulsor outside the circle C_2.

In particular, if we consider real values for the parameter m, we have that $z = 1$ is attractor for $0 < m < 2n$, indifferent for $m = 0$ and $m = 2n$ and repulsive in other case.

Lemma 2.3. *There is an atractor periodic orbit of period two $\{p_1, p_2\}$ given by:*

$$\left\{ \frac{m(m + 2n - \sqrt{m^2 - 4n^2})}{2n(m + 2n)}, \frac{m(m + 2n + \sqrt{m^2 - 4n^2})}{2n(m + 2n)} \right\} \tag{4}$$

inside de curve C_3 defined by:

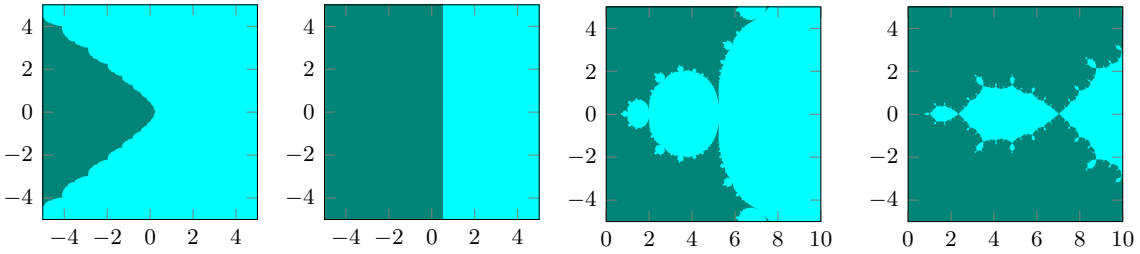
$$(\alpha^2 + \beta^2)^2 + 10n^2(\beta^2 - \alpha^2) + 24n^4 = 0.$$

with $m = \alpha + i\beta$ and $n \in \mathbb{R}$.

For $m, n \in \mathbb{R}$, the periodic orbit is attractor for $-\sqrt{6}n < m < -2n$ and $2n < m < \sqrt{6}n$.

Analysing dynamical planes for natural values of m and n

As the parameter planes for different values of $n > 0$ are topologically equivalent, we consider $n = 3$ and we show some dynamical planes in Figure 2 in order to analyse different behaviours. In these figures, the dark region corresponds to the basin of attraction of $z = 0$. For $0 < m < 6$, the point $z = 1$ is attractor and the free critical point is in its basin of attraction, coloured in cyan. For $m = 6$ the point $z = 1$ bifurcates to the 2-periodic orbit and it is located in the common boundary of the two basins of attraction. For $m = 7$, the point $z = 1$ is repulsive and it is located in the boundary of the two attractor petals of the 2-periodic orbit that becomes attractor; in this case the cyan region corresponds to the basin of attraction of the 2-periodic orbit. For $m \geq 8$, the critical point is in the basin of attraction of $z = 0$ and, therefore, the dynamical plane has only one colour.



(a) m=1 (b) m=3 (c) m=6 (d) m=7

Figure 2: Dynamical planes for $n = 3$ and different values of m.

We can observe that, for $m = n = 3$, we have the case for which the method was designed, when all the roots have the same multiplicity.

2.2 Chebyshev's method

Chebyshev's method is the classical method of order three most commonly used to approximate the simple roots of nonlinear equations. This is due to the fact that, although it is not an optimal method, it has good characteristics such as its dynamics.

Chebyshev's method for multiple roots, studied in [2] is defined by:

$$z_{n+1} = z_n - \frac{m(3-m)}{2}\frac{f(z_n)}{f'(z_n)} - \frac{m^2}{2}\frac{f(z_n)^2 f''(z_n)}{f'(z_n)^3},$$

where m is the multiplicity of the root.

As studied in [2], this scheme preserves the cubic convergence order of the original method. However, as we discuss in this section, the dynamics of Chebyshev's method for simple roots is not preserved when we apply the method for multiple roots on polynomial (1).

Applying the method on this polynomial, we obtain the following operator:

$$O(z,m,n) = z^3 \frac{n}{2}\frac{m(m+3n) - (m+n)(m+2n)z}{(m-(m+n)z)^3}. \tag{5}$$

In this case, the fixed points obtained are $z = 0$, $z = 1$, $z = \infty$ and the points

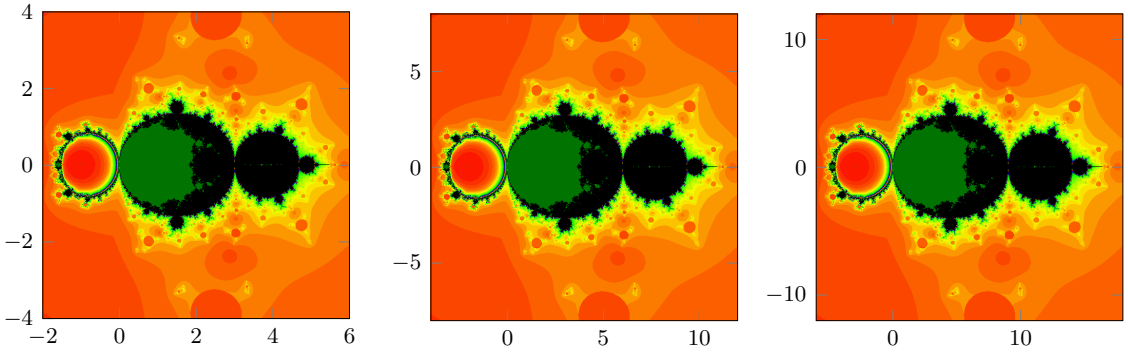
$$z_1 = \frac{m(2m+3n) - m\sqrt{2mn+3n^2}}{(m+n)(2m+3n)} \quad \text{and} \quad z_2 = \frac{m(2m+3n) + m\sqrt{2mn+3n^2}}{(m+n)(2m+3n)}.$$

The points $z = \infty$, z_1 and z_2 are strange fixed points. Note that z_2 coincides with $z = 1$ for $m = 3n$.

In this case, the critical points are $z = 0$ and the free critical points

$$c_1 = \frac{2m(m+2n) - m\sqrt{(m-n)(m+2n)}}{(m+n)(m+2n)} \quad \text{and} \quad c_2 = \frac{2m(m+2n) + m\sqrt{(m-n)(m+2n)}}{(m+n)(m+2n)}.$$

Iterating simultaneously both free critical points, we obtain the parameter planes for m complex and fixing different values of n (see Figure 3). In these figures, red colour means that the critics are in the basins of attraction of $z = 0$ or $z = \infty$ (inside the ring), black color means that one critical point is in the basin of attraction of $z = 0$, while the other critical point is in the basin of attraction of another attractor and green color means that both critical points are in the same basin of attraction of $z = 1$ or another strange attractor.



(a) n=1 (b) n=2 (c) n=3

Figure 3: Parameter planes for Chebyshev methods and for values of $n = 1, 2, 3$.

Now, we study the stability of the fixed points. As $O'(0,m,n) = 0$, $z = 0$ is a superattractor for every value of the parameters m and n.

We study the stability of the other fixed points in the following lemmas.

Lemma 2.4. *The fixed point $z = \infty$ is an attractor inside the curve C_1 defined by*

$$4\beta^4 + (8\alpha^2 + 16\alpha n + 9n^2)\beta^2 + (4\alpha^4 + 16\alpha^3 n + 23\alpha^2 n^2 + 12\alpha n^3) = 0,$$

with $m = \alpha + i\beta$ and $n \in \mathbb{R}$. In particular, it is a superattractor for $m = -2n$. This point is indifferent on the curve C_1 and is a repulsor outside the curve C_1.

If we focus on the real line, we have that $z = \infty$ is attractor for $-\dfrac{3n}{2} < m < 0$ and repulsive in other case. Then, for multiplicities $m, n \in \mathbb{N}$, the point $z = \infty$ is repulsive.

Lemma 2.5. *The fixed point $z = 1$ is an attractor inside the curve C_2 defined by:*

$$\beta^4 + (2\alpha^2 - 6\alpha n + 5n^2)\beta^2 + (\alpha^4 - 6\alpha^3 n + 13\alpha^2 n^2 - 12\alpha n^3) = 0, \tag{6}$$

with $m = \alpha + i\beta$ and $n \in \mathbb{R}$. In particular, it is a superattractor for $m = n$ and $m = 2n$. This point is indifferent on the curve C_2 and is a repulsor outside the curve C_2.

For $z = 1$, if we consider $m \in \mathbb{R}$, we obtain that is attractor for $0 < m < 3n$ and superattarctor for $m = n$ or $m = 2n$. However, $z = 1$ is repulsive for $m < 0$ and indifferent for $m = 0$ or $m = 3n$.

For z_1 and z_2, we obtaing that, for real m and n, the strange fixed point z_1 is always repulsive for $m > 0$. Nevertheless, the strange fixed point z_2 is attractor for $3n < m < kn$, where $k \approx 4.66548$.

Analysing dynamical planes for natural values of m and n

As we have seen, the parameter planes for different n are topologically equivalent. So, we consider $n = 3$ (Figure 3(c)), and we build some dynamical planes in order to analyse different behaviours.

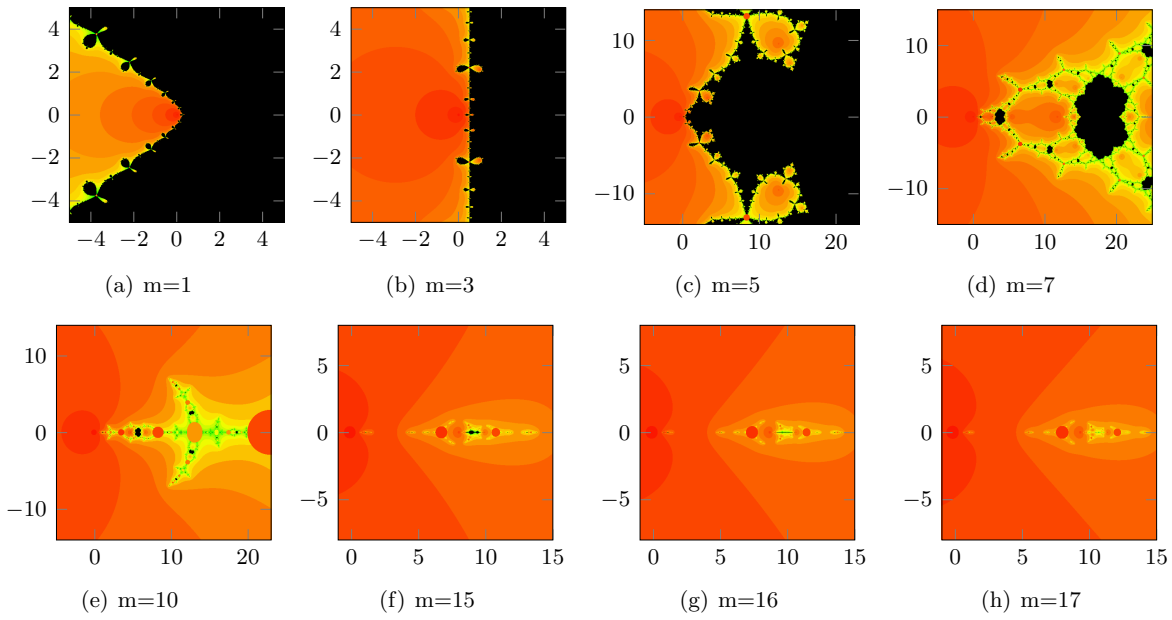


Figure 4: Chebyshev's dynamical planes for n=3.

From the previous study, we know that $z = 1$ is attractor for $0 < m < 9$. From $m = 0$ to $m = 5$, the parameter m is located in the green zone of the parameter plane; this means that both critical points are in the basin of attraction of $z = 1$ and the dynamical planes are divided into two basins of attraction, the corresponding to $z = 0$ in red color and the basins of attraction of $z = 1$ in black color (see (a), (b) and (c) in Figure 4).

For $m = 6, 7, 8$ and 9, the parameter m is located in the black zone of the parameter plane, corresponding to values for which $z = 1$ is attractor and there is only one critical point in the basin of attraction of $z = 1$. Analysing the dynamical planes, we see that the other critical point is in the basin of $z = 0$ (see Figure 4(d)).

For $m = 9$, we have a bifurcation point, the point z_2 coincides with $z = 1$. For $9 < m < 13.996$, z_2 becomes attractor. We show the dynamical planes for $m = 10$ in Figure 4.

The value $m = 13.996$ is a bifurcation point: z_2 becomes a repulsor and an attractor 2-periodic orbit appears. As we can see in the Figure 3(c), for $m = 15$ the parameter is located in the right bulb and, therefore, there is an attractor periodic orbit of period two, like the one shown in Figure 4(f).

For $m = 16$, the parameter m is placed on the antenna and, therefore, a new behaviour occurs (see Figure 4(g)).

For $m = 17$, the parameter is outside the antenna. For this reason, for higher values of m, only the point $z = 0$ is attractor, as can be seen in Figure 4(h).

2.3 Halley's method

Halley's method is a well-known method in the field of iterative methods for approximating simple roots of nonlinear equations. It is a method of order three that is part of the Chebyshev-Halley family of methods.

The Halley method for multiple roots is designed in the article [3] and is defined by the scheme:

$$z_{n+1} = z_n - \frac{f(z_n)}{\frac{(m+1)f'(z_n)}{2m} - \frac{f(z_n)f''(z_n)}{2f'(z_n)}},$$

where m is the multiplicity of the root.

Halley's method for multiple roots, as Chebyshev's method, has cubic convergence.

Applying this method on the polynomial (1), we obtain the operator:

$$O(z, m, n) = z^3 \frac{n(n+m)}{2m^2 - 2m(n+2m)z + (n+m)(n+2m)z^2}. \tag{7}$$

Firstly, we calculate fixed and critical point. Solving $O(z, m, n) = z$, the fixed points obtained are $z = 0$, $z = 1$, $z = \infty$ and $z = \dfrac{m}{n+m}$. The points $z = \infty$ and $z = \dfrac{m}{n+m}$ are strange fixed points.

The critical points are, in this case, $z = 0$ and the free critical points

$$c_1 = \frac{2m(n+2m) - m\sqrt{2(m-n)(2m+n)}}{(m+n)(2m+n)} \quad \text{and} \quad c_2 = \frac{2m(n+2m) + m\sqrt{2(m-n)(2m+n)}}{(m+n)(2m+n)}.$$

Iterating the critical points, we obtain the parameter planes for three fixed values of n, where $m \in \mathbb{C}$ (see Figure 5). The colour pattern of this parameter plane is the same as in Figure 3.

We study the stability of the fixed points. As in the previous methods, $z = 0$ is a superattractor for every value of the parameters m and n.

Lemma 2.6. *The fixed point $z = \infty$ is an attractor inside the circle C_1 defined by*

$$\left(\alpha + \frac{n}{2}\right)^2 + \beta^2 = \left(\frac{n}{2}\right)^2,$$

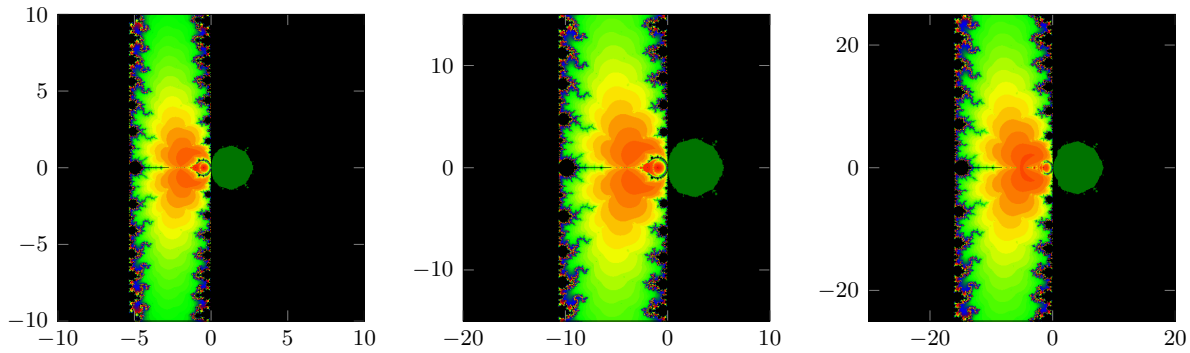
with $m = \alpha + i\beta$ and $n \in \mathbb{R}$. This point is indifferent on the curve C_1 and repulsor outside C_1.

If we focus on real values of m, we have that $z = \infty$ is repulsive for $m > 0$.

Lemma 2.7. *The point $z = 1$ is an attractor for $Re(m) > 0$, it is indifferent on the curve $Re(m) = 0$ and it is a repulsor for $Re(m) < 0$.*

If we focus on real values of m, we have that $z = 1$ is attractor for $m > 0$.

Finally, the strange fixed point $z = \dfrac{m}{n+m}$ is always repulsive.



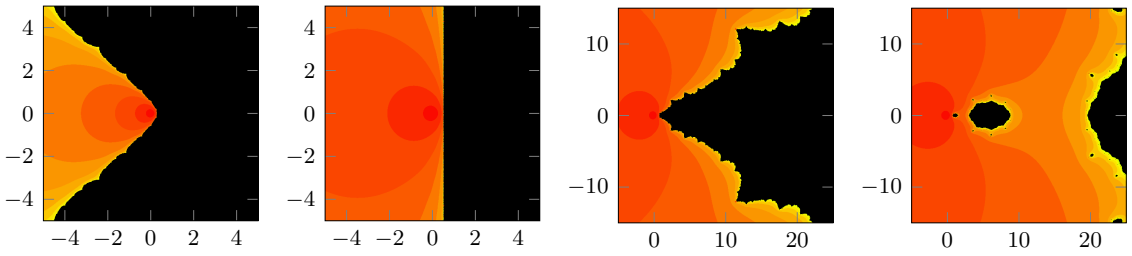
(a) n=1 (b) n=2 (c) n=3

Figure 5: Parameter planes of Halley method for values of $n = 1, 2, 3$.

Analysing dynamical planes for natural values of m and n

As before, we consider the parameter plane for $n = 3$ (see Figure 5(c)) and we build some dynamical planes for $m \in \mathbb{N}$ in order to analyse different behaviours (see Figure 6).

In these cases, $z = 1$ is always attractor, so at least one of the critical points must be in its basins of attraction. We observe that, for m inside the green zone, both critical points are in the basin of attraction of $z = 1$, which is connected for $1 \leq m \leq 8$ (see Figures in 6). For values of m outside this green region, only one critical point belongs to the basin of attraction of $z = 1$ and the other is in the basin of attraction of $z = 0$; as we can observe Figure 6 (d), the basin of attraction of $z = 1$ becomes disconnected.



(a) m=1 (b) m=3 (c) m=7 (d) m=9

Figure 6: Halley's dynamical planes for $n = 3$.

2.4 Schröder's method

Schröder's method is a method of order two, defined in [1] as a multiple root version of Newton's method and which does not include in its definition the multiplicities of the roots to be approximated. Its scheme is defined by the following expression:

$$z_{n+1} = z_n - \frac{f(z_n) f'(z_n)}{(f'(z_n))^2 - f(z_n) f''(z_n)}.$$

This method has a quadratic order of convergence and, unlike the three classical methods above, it is a pure multiple root method that does not depend on the multiplicity of the roots it approximates.

Applying this scheme to polynomial (1), we obtain the operator

$$O(z, m, n) = \frac{nz^2}{m(z-1)^2 + nz^2}. \tag{8}$$

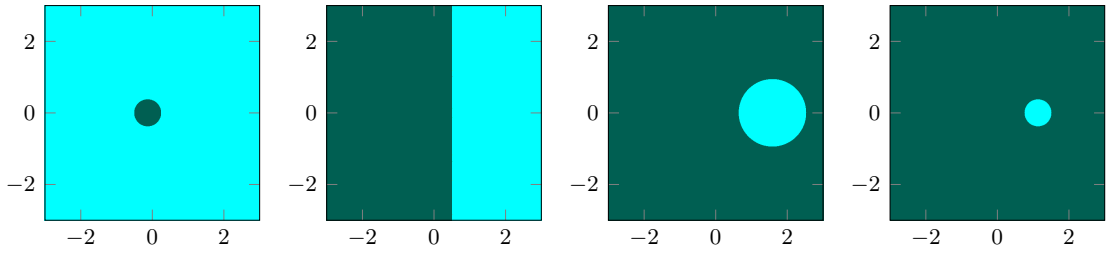
Solving $O(z, m, n) = z$, the fixed points obtained are $z = 0$, $z = 1$ and $z = \dfrac{m}{m+n}$. Let us notice that the point $z = \infty$ is not a fixed point of this method.

From this operator, we obtain that the critical points are $z = 0$ and $z = 1$, which are superattractor fixed points for any value of m and n.

On the other hand, the point $z = \dfrac{m}{m+n}$ is repulsive for any value of m and n.

Analysing dynamical planes for natural values of m and n

As this operator does not have free critical points, the parameter plane, presents only one color and we do not show it. Some examples of dynamical planes for $n = 3$ are shown in Figure 7.



(a) m=1 (b) m=3 (c) m=5 (d) m=9

Figure 7: Dynamical planes of Schöder's method for $n = 3$.

The dynamical conclusions we get from this method is that it seems to be an effective method to find two multiple roots with different multiplicities. This is because both roots, $z = 0$ and $z = 1$, are superattractor roots for any value of m and n. In Figure 7 we can observe the basins of attraction of $z = 0$ (dark green) and $z = 1$ (cyan), for $n = 3$ and different values of m.

If we had finished our analysis of the methods at this point, we would probably declare Schröder's method as one of the best methods for approximating multiple roots since it presents a good dynamical behaviour and, furthermore, it is the only one of the four methods that does not need to know a priori any information of the roots to be approximated. But in Figure 7, we observe that the basin of attraction of the root with the lowest multiplicity decreases rapidly.

References

[1] Schröder, E., Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen. *Mathematische Annalen volume*, 2, 317–365, 1870.

[2] Neta, B., New third order nonlinear solvers for multiple roots. *Applied Mathematics and Computation*, 202, 162–170, 2008.

[3] Hansen E., Patrick, M., A family of root finding methods. *Numerische Mathematik*, 27, 257–269, 1976.

Efficiency and stability of a new hybrid unconstrained optimization algorithm with quasi-Newton updates and high-order methods

Alicia Cordero B. [1♭], Natanael Ureña C.[♮], Juan R. Torregrosa S.[♭] and Javier García M. [2♮]

(♮) Ciencias Básicas y Ambientales (CBA), Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo 10602, Dominican Republic.

(♭) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain.

1 Introduction

Hybrid algorithms [1, 2, 3, 4] have emerged as a promising solution by combining the strengths of different approaches to overcome the inherent limitations of traditional methods. The hybrid Newton-higher-order-Quasi-Newton (NHOQN) method, proposed in this manuscript, offers a robust hybrid strategy for the minimization of convex and nonconvex functions. In its iterative scheme, this algorithm integrates Newton's method, higher order methods such as those of Traub, Chun or Ostrowski, and various quasi-Newton approximations of the Hessian, which allows faster and more efficient convergence compared to classical methods.

1.1 Traub's, Chun's, and Ostrowski's Methods

Traub's method [5] is a multi-step algorithm of convergence order 3, which is typically used to solve nonlinear equations and systems. The iterative expression of this method is obtained by composing Newton's method with itself and substituting, in the second step, the derivative evaluated at the Newton step by the derivative used in the initial step. The iterative formulation of Traub's method for solving nonlinear equations is

$$\begin{aligned} y_k &= x_k - \frac{f(x_k)}{f'(x_k)}, \\ x_{k+1} &= y_k - \frac{f(y_k)}{f'(x_k)} = x_k - \frac{f(x_k)+f(y_k)}{f'(x_k)}. \end{aligned}$$

When applied to optimization problems, where $\nabla f(x)$ denotes the *gradient* of f at x and $\nabla^2 f(x)$ denotes the *Hessian matrix* of f at x, the iterative formulation of Traub's method is expressed as follows:

$$\begin{aligned} y_k &= x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k), \\ x_{k+1} &= x_k - \left[\nabla^2 f(x_k)\right]^{-1} \left(\nabla f(x_k) + \nabla f(y_k)\right). \end{aligned} \tag{1}$$

[1] **acordero@mat.upv.es**
[2] **javier.garcia@intec.edu.do**

On the other hand, two classical fourth-order methods, widely recognized for their optimality, are Chun's method [6] and Ostrowski's method [7]. The iterative expressions corresponding to each of these methods are as follows:

$$\textbf{Chun's Method:} \quad y_k = x_k - \frac{f(x_k)}{f'(x_k)},$$
$$x_{k+1} = y_k - \frac{f(x_k) + 2f(y_k)}{f(x_k)} \frac{f(y_k)}{f'(x_k)}.$$

$$\textbf{Ostrowski's Method:} \quad y_k = x_k - \frac{f(x_k)}{f'(x_k)},$$
$$x_{k+1} = y_k - \frac{f(x_k)}{f(x_k) - 2f(y_k)} \frac{f(y_k)}{f'(x_k)}.$$

For the optimized versions of Chun's and Ostrowski's methods, Cordero et al., in [8], build upon the fundamental idea presented by Singh et al., in [9], which introduces a two-step iterative scheme designed to solve nonlinear systems using a fifth-order convergence method. This scheme includes a **first step**, which employs a standard Newton iteration:

$$w^{(k)} = x^{(k)} - \left[F'\left(x^{(k)}\right)\right]^{-1} F\left(x^{(k)}\right),$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear system of equations, and $F'(x)$ denotes the *Jacobian matrix* of F evaluated at x, that is, the matrix of all first-order partial derivatives of the components of F, and a **second step**, where a weighted corrective term is introduced, based on the ratio of squared norms of the function evaluated at $w^{(k)}$ and $x^{(k)}$, expressed as $\frac{\left\|F\left(w^{(k)}\right)\right\|^2}{\left\|F\left(x^{(k)}\right)\right\|^2}$:

$$x^{(k+1)} = w^{(k)} - \left(1 + \frac{\left\|F\left(w^{(k)}\right)\right\|^2}{\left\|F\left(x^{(k)}\right)\right\|^2}\right) \left[F'\left(w^{(k)}\right)\right]^{-1} F\left(w^{(k)}\right).$$

In [8], Cordero et al. extend this idea by parametrizing the iterative scheme through the **Ermakov Hyperfamily**, deriving methods such as Chun's and Ostrowsky's as particular cases. For Chun's method, by setting $\lambda = 0$ and $\psi = 1$, the iteration is expressed as:

$$x^{(k+1)} = w^{(k)} - \left(1 + \frac{\left\|F\left(w^{(k)}\right)\right\|^2}{\left\|F\left(x^{(k)}\right)\right\|^2}\right) \left[F'\left(w^{(k)}\right)\right]^{-1} F\left(w^{(k)}\right). \tag{2}$$

On the other hand, Ostrowski's method is derived by selecting $\lambda = -4$ and $\psi = 0$, resulting in the following iterative expression:

$$x^{(k+1)} = w^{(k)} - \frac{1}{1 - 4\nu_k} \left[F'\left(w^{(k)}\right)\right]^{-1} \left(F\left(w^{(k)}\right) + 2\nu_k F\left(x^{(k)}\right)\right), \quad \text{where } \nu_k = \frac{\left\|F\left(w^{(k)}\right)\right\|^2}{\left\|F\left(x^{(k)}\right)\right\|^2}$$

The coefficient ν_k provides these versions of Chun's and Ostrowski's methods with greater computational efficiency compared to the classical versions of these methods, which, while maintaining fourth-order convergence, are more computationally expensive. This is particularly important in the context of optimization, where reducing computational cost is key to tackling large-scale and complex problems effectively. In the context of optimization, these methods allow finding stationary points of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$. Given an initial point $x_0 \in \mathbb{R}^n$, Chun's and Ostrowski's methods generate respective sequences of points $\{x_k\}$ using the following iterative formulas:

$$\textbf{Chun's Method:} \quad x_{k+1} = y_k - \left[\nabla^2 f(x_k)\right]^{-1} \left((1+\nu_k)\nabla f(y_k) - 2\nu_k \nabla f(x_k)\right). \tag{3}$$

$$\textbf{Ostrowski's Method:} \quad x_{k+1} = y_k - \frac{1}{1-4\nu_k}\left[\nabla^2 f(x_k)\right]^{-1} \left(\nabla f(y_k) + 2\nu_k \nabla f(x_k)\right), \tag{4}$$

where: $y_k = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$ is the initial approximation obtained using Newton's method y $\nu_k = \frac{\|\nabla f(y_k)\|^2}{\|\nabla f(x_k)\|^2}$ is a coefficient that adjusts the correction based on the function evaluations at the current and previous points.

2 Hybrid-quasi-Newton method (NHOQN)

The hybrid NHOQN method solves unconstrained optimization problems by combining Newton's method, higher-order corrections (such as Chun or Ostrowski) and quasi-Newton updates (such as BFGS or DFP), although it supports other formulas for the Hessian. At each iteration, a Newton direction and an upper correction are computed and combined to define the search direction. The step size is obtained by an inexact linear search (Wolfe or Armijo), and the approximate Hessian is updated with a quasi-Newton scheme. The NHOQN ensures global convergence under standard optimization conditions. To illustrate how these ideas come together in practice, the algorithm 1 below presents the general iterative framework of the hybrid NHOQN method:

Algorithm 1 Iterative Scheme NHOQN

Inputs: $f : \mathbb{R}^n \to \mathbb{R}$, $g(x) = \nabla f(x)$, $x^{(0)}$, $H^{(0)}$, tol, $k_{\max}$
for $k = 0, \ldots, k_{\max} - 1$ **do**
  $g_k \leftarrow g\big(x^{(k)}\big)$
  if $\|g_k\| \leq$ tol **then break**
  end if
  Step 1: Newton
    $d_k^N \leftarrow -H^{(k)} g_k$
    $\alpha_k^N \leftarrow \arg\min_{\alpha \in [0,10]} f\big(x^{(k)} + \alpha\, d_k^N\big)$ (`fminbnd`)
    $y_k \leftarrow x^{(k)} + \alpha_k^N\, d_k^N$
    $g_y \leftarrow g(y_k)$
  Step 2: Alto orden
    Compute d_k^{HO} (Chun/Ostrowski/Traub) using $H^{(k)}, g_k, g_y$
    $\alpha_k^{HO} \leftarrow \arg\min_{\alpha \in [0,10]} f\big(y_k + \alpha\, d_k^{HO}\big)$ (`fminbnd`)
    $x^{(k+1)} \leftarrow y_k + \alpha_k^{HO}\, d_k^{HO}$
    $g_{k+1} \leftarrow g\big(x^{(k+1)}\big)$
  Step 3: Actualización DFP
    $s_k \leftarrow x^{(k+1)} - x^{(k)}$
    $\Delta y_k \leftarrow g_{k+1} - g_k$
    $H^{(k+1)} \leftarrow \text{DFP_update}\big(H^{(k)}, s_k, \Delta y_k\big)$
end for
Return $x^{(k+1)}$, g_{k+1}, $H^{(k+1)}$

2.1 Approximate Computational Order of Convergence (ACOC)

In the context of hybrid optimization algorithms, the calculation of the ACOC, represented by ρ, using the traditional method proposed in [10], does not appear to be a reliable measure, as it

produces significant fluctuations between consecutive iterations. To improve the estimation of ρ, we have employed a technique based on linear regression of the natural logarithms of successive errors $(\ln E_k, \ln E_{k+1})$. This approach takes advantage of the global information from all available iterations, allowing a more consistent estimation and reducing the sensitivity to local variations of the individual errors, in contrast to the traditional formula, which is based only on the last iterations. The slope of the fitted line, ρ, is calculated by the following expression:

$$\rho \approx \frac{\sum_k \left(\ln E_k - \overline{\ln E}\right)\left(\ln E_{k+1} - \overline{\ln E}\right)}{\sum_k \left(\ln E_k - \overline{\ln E}\right)^2}, \tag{5}$$

where E_k is the error at iteration k y $\overline{\ln E} = \frac{1}{N}\sum_k \ln E_k$ is the mean of the natural logarithms of the errors, with N being the total number of iterations.

3 Numerical results

In this section, we numerically compare different variants of the NHOQN method with each other and with the Newton-BFGS method, using the Himmelblau (f_H), and Freudenstein-Roth functions (f_{FR}). For the numerical tests, MATLAB R2024 software was used with a tolerance of $e_k = 10^{-6}$. Tables 1-2 present the optimization results for the Himmelblau and Freudenstein-Roth functions, evaluated under two different initial conditions. Several iterative methods were applied, including the hybrid schemes Newton-BFGS (NB), Newton-Ostrowsky-DFP (NOD), Newton-BFGS-Traub-DFP (NBTD), Newton-BFGS-Chun-DFP (NBCD) and Newton-Ostrowsky-BFGS-DFP (NBOD). Each table summarizes the performance of these methods using metrics such as the number of iterations (Iter), the estimated order of convergence (ρ), the gradient norm $\left\|\nabla f\left(x^{(k+1)}\right)\right\|$, and the computational time in seconds (CT). The numerical results presented in Tables 1-2 show that the order of convergence does not seem to be a conclusive characteristic to compare the evaluated methods, since it varies significantly depending on the function analyzed and the initial conditions selected near the optimal values. It is also observed that the final gradient norm seems to be a measure that depends on the type of optimization problems, since for the Freudenstein-Roth function it was substantially more stable than in the Himmelblau function.

Table 1: Numerical tests for Himmelblau function with two different initial estimates

Himmelblau function (f_H), $\xi = (3, 2)$				
$x_0 = (2.5, 5)$				
Method	Iter	ρ	$\|\nabla f(x^{(k+1)})\|$	CT (s)
NB	6	1.6296	3.56×10^{-11}	0.0112
NOD	4	1.9703	9.73×10^{-9}	0.0393
NTD	5	1.3906	2.45×10^{-9}	0.0293
NBCD	3	2.0926	6.91×10^{-7}	0.0397
NBOD	3	2.7975	1.72×10^{-8}	0.0392
NBTD	4	1.2752	6.12×10^{-8}	0.0566
$x_0 = (2, 8)$				
NB	7	1.4681	4.25×10^{-10}	0.0285
NOD	5	1.6128	3.23×10^{-11}	0.0223
NTD	5	1.2687	4.82×10^{-9}	0.0166
NBCD	3	2.0926	6.91×10^{-7}	0.0397
NBOD	4	1.8886	1.84×10^{-10}	0.0149
NBTD	4	1.3664	8.94×10^{-7}	0.0106

Table 2: Numerical tests for Freudenstein–Roth function with two different initial estimates

Freudenstein–Roth function (f_{FR}), $\xi = (11.41, -0.89)$				
$x_0 = (3, 5)$				
Method	Iter	ρ	$\|\nabla f(x^{(k+1)})\|$	CT (s)
NB	6	1.2781	2.37×10^{-7}	0.0058
NOD	4	1.4388	1.69×10^{-10}	0.0205
NTD	16	0.7334	6.68×10^{-7}	0.0249
NBCD	4	1.3818	1.36×10^{-7}	0.0155
NBOD	4	1.7058	5.01×10^{-8}	0.0118
NBTD	5	1.4566	1.06×10^{-7}	0.0269
$x_0 = (6, 4)$				
NB	7	1.2743	1.79×10^{-8}	0.0089
NOD	19	1.9299	4.32×10^{-7}	0.0292
NTD	11	0.8321	7.51×10^{-7}	0.0202
NBCD	4	1.9313	2.98×10^{-7}	0.0198
NBOD	5	1.0567	2.60×10^{-7}	0.0107
NBTD	6	1.0329	1.86×10^{-7}	0.0123

The numerical results indicate that, for both problems studied, both the number of iterations and the computational time show minimal variations between the hybrid methods analyzed, with no significant numerical differences between them.

4 Conclusions

It is concluded that the hybrid NHOQN scheme is a versatile strategy that coherently integrates Newton's method with high order corrections and several updates of the Hessian, showing excellent numerical and dynamic behavior even in non-convex functions such as those studied.

As a future line of research, it is proposed to design hybrid quasi-Newton methods with parametric families to analyze their dynamic behavior as a function of parameter values. Also, we intend to develop limited memory variants of these hybrid methods, together with the evaluation of their efficiency in large-scale problems, in particular the training of neural networks with complex convolutional architectures, using MNIST and FACE datasets.

It is proposed to design memory-limited hybrid quasi-Newton methods, endowed with a trust region strategy and oriented to deep neural network training. Another interesting line of research is to develop hybrid methods combining quasi-Newton methods with metaheuristic algorithms.

Acknowledgments

This research was financed by Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCT) of Ministerio de Educación Superior Ciencia y Tecnología de la República Dominicana (MESCYT) grant number FONDOCYT 2023-1-1D2-0537.

References

[1] Singh, A., Sharma, A., Rajput, S., Bose, A., Hu, X. An investigation on hybrid particle swarm optimization algorithms for parameter optimization of PV cells. *Electronics*, 11(6):909, 2022.

[2] Farnad, B., Jafarian, A., Baleanu, D. A new hybrid algorithm for continuous optimization problem. *Applied Mathematical Modelling*, 55:652–673, 2018.

[3] Wang, G., Guo, L. A novel hybrid bat algorithm with harmony search for global numerical optimization. *Journal of Applied Mathematics*, 2013(1):696491, 2013.

[4] Garg, H. A hybrid PSO-GA algorithm for constrained optimization problems. *Applied Mathematics and Computation*, 274:292–305, 2016.

[5] Traub, J. F. Iterative Methods for the Solution of Equations. American Mathematical Society: Providence, RI, 1982.

[6] Chun, C. Construction of Newton-like iteration methods for solving nonlinear equations. *Numerische Mathematik*, 104:297–315, 2006.

[7] Ostrowski, A. M. Solution of Equations and Systems of Equations. Elsevier: Amsterdam, 2016.

[8] Cordero, A., Rojas-Hiciano, R. V., Torregrosa, J. R., Vassileva, M. P. A highly efficient class of optimal fourth-order methods for solving nonlinear systems. *Numerical Algorithms*, 95(4):1879–1904, 2024.

[9] Singh, H., Sharma, J. R., Kumar, S. A simple yet efficient two-step fifth-order weighted Newton method for nonlinear models. *Numerical Algorithms*, 93(1):203–225, 2023.

[10] Cordero, A., Torregrosa, J. R. Variants of Newton's method using fifth-order quadrature formulas. *Applied Mathematics and Computation*, 190(1):686–698, 2007.

Iterative methods for multiple roots with memory

Francisco I. Chicharro[♭], Neus Garrido[♭] and Julissa H. Jerezano[♮,♭,1].

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, 46022–València, Spain.

(♮) Departmento de Matemática, Universidad Nacional Autónoma de Honduras,
21102–San Pedro Sula, Honduras.

1 Introduction

Iterative methods for approximating the roots of nonlinear functions play a fundamental role in various applied mathematics, physics, and engineering areas. These methods allow for finding solutions to equations of the form $f(x) = 0$ when obtaining an exact analytical solution is not feasible. However, when a root has a multiplicity greater than one, meaning that $f(x)$ and its derivatives up to a particular order vanish at the root, conventional iterative methods may experience slow convergence or even fail.

The multiplicity of a root x^* is defined as the smallest positive integer m such that $f(x^*) = 0$ and $f'(x^*) = f''(x^*) = \cdots = f^{(m-1)}(x^*) = 0$, but $f^{(m)}(x^*) \neq 0$. In such cases, most iterative methods either do not work properly or do not hold the same convergence properties as when used to approximate a simple root ($m = 1$). Specifically, Newton's method convergence is quadratic when the root is simple since its error equation is $e_{k+1} = \frac{f''(x^*)}{2f'(x^*)} e_k^2 + \mathcal{O}(e_k^3)$, where $e_k = x_k - x^*$, but it changes to linear when $m > 1$, being its error equation $e_{k+1} = \frac{m-1}{m} e_k + \mathcal{O}(e_k^2)$. Reduction in convergence order occurs because the error relationship in each iteration is affected by the root's multiplicity, decreasing the method's efficiency.

Various modifications to Newton's method have been developed to improve convergence in these cases. A common strategy is the modified Newton's method [6], which incorporates information about the multiplicity m, defined by the iteration

$$x_{k+1} = x_k - m\frac{f(x_k)}{f'(x_k)}, \qquad k = 0, 1, 2, \ldots,$$

obtained by applying Newton's method to $\sqrt[m]{f(x)}$, whose error equation is

$$e_{k+1} = \frac{f^{(m+1)}(x^*)}{f^{(m)}(x^*)} \frac{1}{m(m+1)} e_k^2 + \mathcal{O}(e_k^3).$$

This method restores quadratic convergence but requires prior knowledge of m, which is not always feasible. However, Schröder [6] proposed another variant of Newton's method that does not require prior knowledge of multiplicity. Schröder's iterative scheme is obtained applying Newton's method to $\frac{f(x)}{f'(x)}$ and preserves the quadratic order of convergence even for multiple roots, but it is not optimal in the sense of Kung-Traub conjecture.

[1] jhjerull@doctor.upv.es

Alternatively, higher-order methods utilize both the function and its higher-order derivatives to enhance convergence speed. An example is the scheme proposed by Schröder and Ostrowski

$$x_{k+1} = x_k - \frac{1}{2}(3-m)\frac{f(x_k)}{f'(x_k)} - m^2\frac{f''(x_k)}{2f'(x_k)}\left(\frac{f(x_k)}{f'(x_k)}\right), \qquad k = 0, 1, 2, \ldots$$

that converges with order three.

Other approaches include variants of Halley's method, which use the second derivative or even higher-order derivatives to improve accuracy. These methods can offer cubic or higher convergence rates under certain conditions.

Furthermore, some authors design iterative methods based on divided differences that approximate the derivative numerically. In this sense, Cordero et al. [4] generate two iterative methods with and without derivatives based on Kurchatov's method [5].

In practice, selecting an appropriate iterative method depends on several factors, such as the availability of derivatives, computational complexity, and sensitivity to rounding errors. Higher-order methods often require more computations per iteration, making them less efficient in problems where evaluating derivatives is costly.

In this work, the authors present iterative procedures for solving nonlinear equations with multiple roots for problems where the nonlinear function is differentiable and not differentiable. Section 2 is devoted to explaining the procedure of the design of the methods. Section 3 comprises the numerical experiment on three nonlinear functions where the initial guess is obtained after the representation of the dynamical planes. Finally, Section 4 covers the main conclusions.

2 Methods

In the literature of recent decades, one of the most widely used techniques for designing efficient iterative schemes with higher order of convergence has been the use of methods with memory. Such methods use more than one previous iteration to obtain the next approximation to the root. One of the most characteristic schemes is the Secant method, with order of convergence $p \approx 1.6180$ and iterative expression

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}, \qquad k = 1, 2, \ldots$$

In a similar way to [4], in this work we design two iterative methods based on a method with memory. Specifically, the starting iterative scheme is Secant's method. On the one hand, by replacing function $f(x)$ with Newton's quotient, we obtain the SD method. The new method holds the convergence order of the original method and has the following error equation

$$e_{k+1} = -\frac{c_1}{m}e_k e_{k-1} + \mathcal{O}_3(e_k, e_{k-1}),$$

where $\mathcal{O}_3(e_k, e_{k-1})$ denotes all the terms whose sum of the exponents of e_k and e_{k-1} is at least 3.

On the other hand, by replacing function $f(x)$ with a version of Newton's quotient without derivatives and including a parameter β, we obtain the SF family. As in SD method, the new family of iterative methods holds the order of convergence of Secant's scheme. Additionally, it also has the same error equation as the SD method for all values of parameter β.

In the design of a family of iterative schemes with dependence on a real parameter, such as SF's class, for each value of the parameter a scheme belonging to the family would be obtained with the same convergence properties, but probably with different stability characteristics. The appropriate selection of the most stable members of a family of iterative schemes is performed by means of a dynamical analysis in which the convergence of its members is usually analysed when applied to low-degree polynomials. However, we can also verify numerically the performance of

the methods of a family by taking specific parameter values and approximating roots of arbitrary nonlinear functions. We devote the next section to the numerical analysis of the schemes proposed in this paper. For this purpose, we select the scheme obtained by taking the value of the parameter $\beta = 1$ in SF family.

3 Numerical benchmark based on stability representation

We are assessing the proposed iterative schemes solving the nonlinear equations

- $f_1(x) = x^2e^x - \sin(x) + x = 0$, which has one root with multiplicity two, $x^* = 0$.
- $f_2(x) = (x^2 - e^x - 3x + 2)^3 = 0$, which has one root with multiplicity three, $x^* \approx 0.25753029$.
- $f_3(x) = (x-1)^2(x-2)^2(x-3)^2 = 0$, which has three roots with multiplicity two, $x_1^* = 1$, $x_2^* = 2$ and $x_3^* = 3$.

The dynamical planes are generated to establish an appropriate criterion for selecting the initial estimates to apply each method. Figures 1, 2 and 3 represent the dynamical planes corresponding to the nonlinear test functions $f_1(x)$, $f_2(x)$ and $f_3(x)$ generated by the proposed schemes, respectively. The dynamical planes represent the basins of attraction of an attracting point, that is, the set of initial estimations converging to an attracting point, in this case, a root of the nonlinear function under study. In addition, the roots of the nonlinear functions are depicted with white stars.

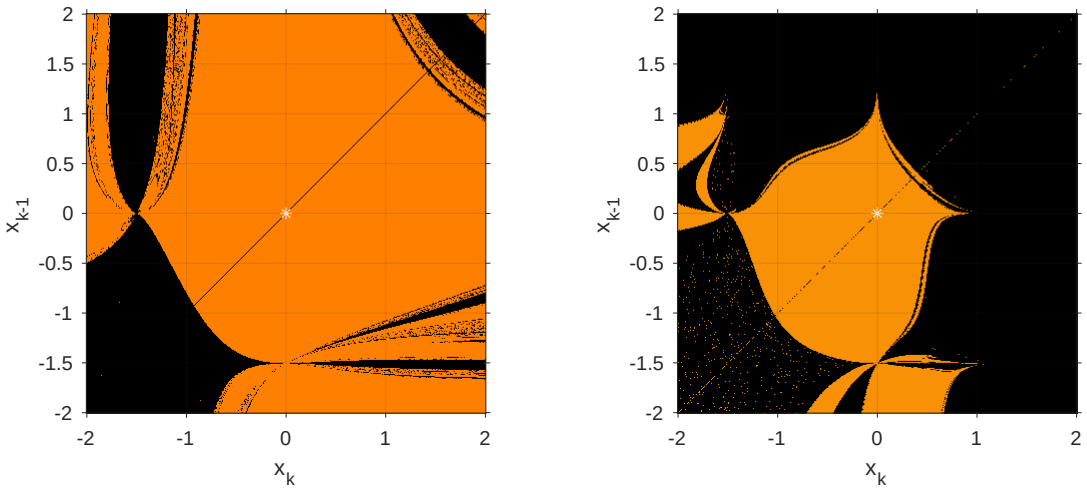
Dynamical planes have been represented following the guidelines on [1]. Iterative schemes SD and SF (for $\beta = 1$) are methods with memory, so they require two initial estimations to apply the method. Therefore, the dynamical planes have been generated using a mesh of 400×400 points in the real plane. The abscissa and ordinate axes represent the current and previous iteration, denoted by x and z, respectively. Each point of the mesh is taken as an initial point to apply the method successively. When any of the points of the sequence generated in the iterative process converges to a root of the nonlinear function, this initial point is represented with a colour different from black. The convergence is set when the difference between the point and a root is lower than 10^{-3} with less than 100 iterations. The point in the plane is represented in black when these conditions are not satisfied, that is, when there is no convergence. Finally, the convergence to the only root of functions $f_1(x)$ and $f_2(x)$ is represented in orange in Figures 1 and 2, while for $f_3(x)$ the initial estimates converging to $x_1^* = 1$, $x_2^* = 2$ and $x_3^* = 3$ are denoted in Figure 3 in orange, blue and green, respectively.

We can see from Figures 1-3 that the methods we propose provide a considerably large set of convergent initial estimates that can be taken to iterate the methods on the three considered nonlinear equations. In the following we will check numerically the information of the dynamical planes.

We compare the proposed methods with the methods coming from [3], which we denote by KM and KMD, with and without derivatives, respectively, and with the technique coming from [2], which we denote by gTM that includes derivatives.

The numerical tests are conducted with Matlab R2023a in a PC equipped with an Intel-Core i7-14700 processor at 2.10 GHz with 16 GB of RAM. The operations use variable precision arithmetics with 500 digits of mantissa to guarantee that the stopping criterion $|x_{k+1} - x_k| + |f(x_k)| < 10^{-25}$ is reached without division by zero problems. We also use a maximum of 100 iterations as a stopping criterion.

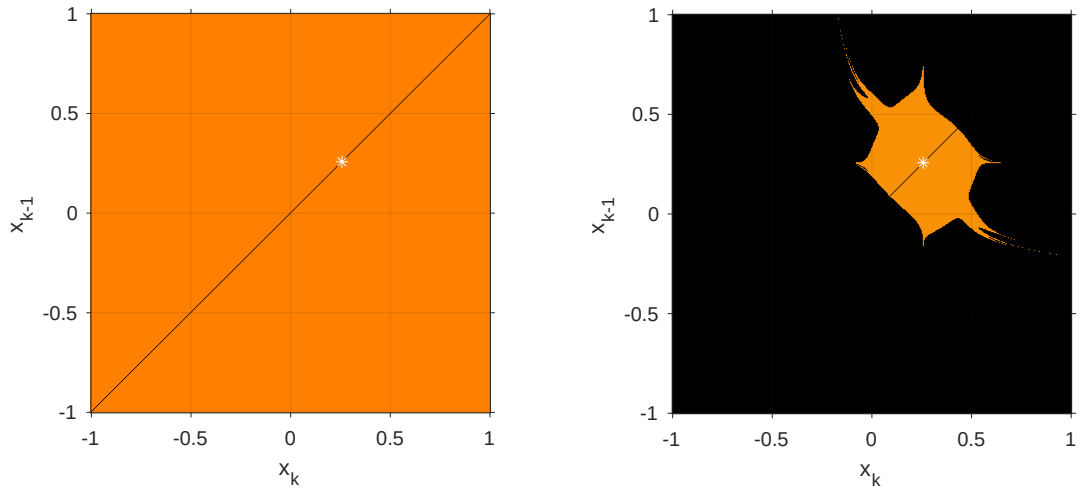
The results are collected in Tables 1, 2, and 3. Tables collect the initial guess x_0, x_{-1}, x_{-2}, the number of iterations k to reach the solution, the approximated computational order of convergence ACOC [4], the residual $|x_{k+1} - x_k|$ at the last iterate, the value of the function $|f(x_{k+1})|$ and the CPU Time (in the seconds) at the last iterate.



(a) SD

(b) SF, $\beta = 1$

Figure 1: Dynamical planes for $f_1(x)$.



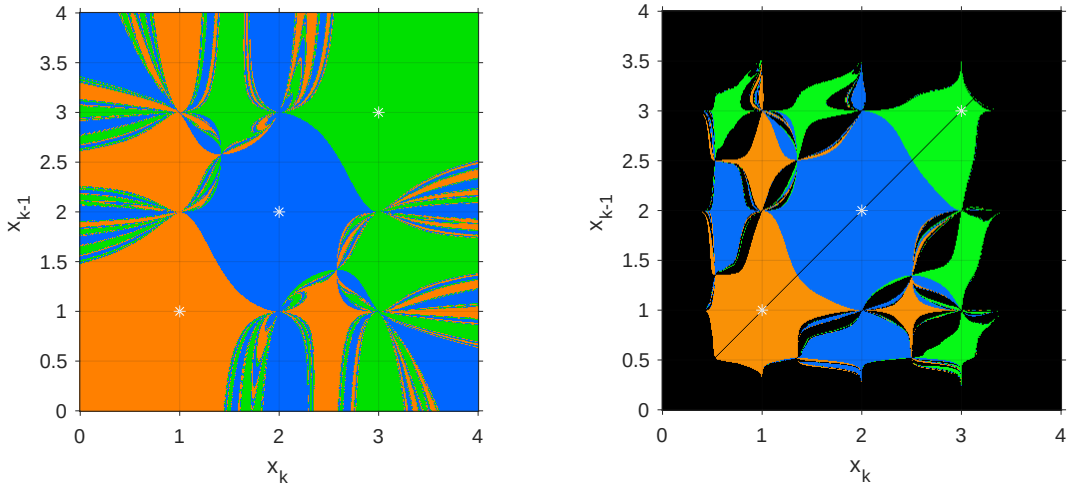
(a) SD

(b) SF, $\beta = 1$

Figure 2: Dynamical planes for $f_2(x)$.

Table 1: Results for equation $f_1(x) = 0$.

Method	x_0	x_{-1}	x_{-2}	k	**ACOC**	$\lvert x_{k+1} - x_k \rvert$	$\lvert f(x_{k+1}) \rvert$	**CPU Time**
SD	-0.5	-0.6		10	1.6184	$3.165767 \cdot 10^{-35}$	$1.173023 \cdot 10^{-112}$	0.00255
SF	-0.5	-0.6		10	1.6184	$6.49214 \cdot 10^{-37}$	$8.69093 \cdot 10^{-118}$	0.00301
KM	-0.5	-0.6		8	2.0002	$3.156288 \cdot 10^{-41}$	$1.137939 \cdot 10^{-162}$	0.00281
KMD	-0.5	-0.6		7	2.0226	$1.938624 \cdot 10^{-32}$	$1.299837 \cdot 10^{-126}$	0.00360
gTM	-0.5	-0.6	-0.7	9	1.8301	$3.003245 \cdot 10^{-26}$	$8.021029 \cdot 10^{-95}$	0.00296



(a) SD

(b) SF, $\beta = 1$

Figure 3: Dynamical planes for $f_3(x)$.

As shown in Table 1, all the methods obtain good results for the chosen initial points. The approximate computational convergence order coincides with the theoretical one. Interestingly, the table shows that, for the initial points chosen, we see that the SD and SF methods require less CPU time than the corresponding methods with and without derivatives, respectively.

Table 2: Results for equation $f_2(x) = 0$.

Method	x_0	x_{-1}	x_{-2}	k	**ACOC**	$\lvert x_{k+1} - x_k \rvert$	$\lvert f(x_{k+1}) \rvert$	**CPU Time**
SD	0.1	0.05		7	1.6028	$7.704388 \cdot 10^{-36}$	$3.008738 \cdot 10^{-285}$	0.00248
SF	0.1	0.05		10	1.5675	$1.381989 \cdot 10^{-26}$	$1.288346 \cdot 10^{-210}$	0.00350
KM	0.1	0.05		6	1.9948	$5.700043 \cdot 10^{-46}$	$2.835103 \cdot 10^{-454}$	0.00280
KMD	0.1	0.05		9	1.9985	$2.495773 \cdot 10^{-46}$	$1.018431 \cdot 10^{-457}$	0.00420
gTM	0.1	0.05	-0.1	6	1.8071	$1.681077 \cdot 10^{-34}$	$4.406565 \cdot 10^{-310}$	0.00282

Table 3: Results for equation $f_3(x) = 0$.

Method	x_0	x_{-1}	x_{-2}	k	**ACOC**	$\lvert x_{k+1} - x_k \rvert$	$\lvert f(x_{k+1}) \rvert$	**CPU Time**
SD	1.8	1.7		7	1.9866	$7.222381 \cdot 10^{-34}$	$4.215194 \cdot 10^{-133}$	0.00252
SF	1.8	1.7		8	1.6212	$1.109661 \cdot 10^{-27}$	$2.570449 \cdot 10^{-88}$	0.00334
KM	1.8	1.7		7	1.9831	$8.220072 \cdot 10^{-41}$	$6.169903 \cdot 10^{-161}$	0.00315
KMD	1.8	1.7		7	1.9919	$8.11058 \cdot 10^{-40}$	$1.217645 \cdot 10^{-156}$	0.03518
gTM	1.8	1.7	1.6	8	1.8318	$2.697373 \cdot 10^{-45}$	$3.503401 \cdot 10^{-164}$	0.00340

As we can see in Table 2 and Table 3, all the methods obtain good results for the chosen initial points. The approximate computational convergence order coincides with the theoretical one, and the number of iterations used to verify the stopping criterion is almost the same for all methods. Although the CPU time is lower in the proposed methods.

4 Conclusions

In this work, we modify a memory-based method to make it applicable to obtaining multiple roots (without needing to know their multiplicity), while maintaining their order of convergence.

We then modify the proposed method to obtain the SF method, which is a derivative-free memory-based method with the same characteristics as the SD method.

By running the KM, KMD gTM, SD, and SF methods on several examples, we can conclude that the proposed SD and SF methods show very good performance and are more efficient than the KM, KMD, and gTM methods in terms of runtime and computational cost.

Acknowledgments

Funded with Ayuda a Primeros Proyectos de Investigación PAID-06-23 and PAID-11-24, Vicerrectorado de Investigación de la Universitat Politècnica de València.

References

[1] Chicharro, F.I., Cordero, A., Torregrosa, J.R., Drawing Dynamical and Parameters Planes of Iterative Families and Methods *Sci. World J.* 2013: 780153, 2013.

[2] Cordero, A., Garrido, N., Torregrosa, J.R., Triguero-Navarro, P., Modifying Kurchatov's method to find multiple roots of nonlinear equations *Applied Numerical Mathematics*, 198: 11–21, 2024.

[3] Cordero, A., Neta, B., Torregrosa, J.R., Memorizing Schröder's Method as an Efficient Strategy for Estimating Roots of Unknown Multiplicity *Mathematics* 9: 2570, 2021.

[4] Cordero, A., Torregrosa, J.R., Variants of Newton's method using fifth-order quadrature formulas *Appl.Math. Comput*, 190: 686-–698, 2007.

[5] Kurchatov, V.A., On a method of linear interpolation for the solution of functional equations *Dolk Akad Nauk SSSR* 198: 524–526, 1971.

[6] Schröder E., Über unendlich viele algorithmen zur auflösung der gleichungen *Mathematische Annalen*, 2: 317–365, 1870.

Increasing the Order of Convergence in Jacobian-Free Iterative Schemes: Applications to Nonlinear Systems and Nonlinear Partial Differential Equations

Alicia Cordero[a], Miguel A. Leonardo Sepúlveda[1c,d], María P. Vassileva[b] and Juan R. Torregrosa[a]

(*a*) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain.

(*b*) Ciencias Básicas y Ambientales (CBA), Instituto Tecnológico de Santo Domingo (INTEC), Av. Los Próceres, Gala, Apartado postal 342-9 and 249-2, Santo Domingo 10602, Dominican Republic.

(*c*) Instituto Superior de Formación Docente Salomé Ureña (ISFODOSU), Recinto Félix Evaristo Mejía, Av. Caonabo con esq. Leonardo Da Vinci, Restauradores, Santo Domingo 10114, Dominican Republic.

(*d*) Instituto Tecnológico de las Américas (ITLA), Autopista Las Américas, Km. 27, PCSD, La Caleta, Boca Chica 11606, Dominican Republic.

1 Introduction

Consider the system of nonlinear equations $F(x) = 0$, where $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$, and f_i, $i = 1, 2, \ldots, n$, are the coordinate functions of F, with

$$F(x) = (f_1(x), f_2(x), \ldots, f_n(x))^T .$$

Solving large-scale nonlinear systems is fundamental to numerous scientific and engineering applications. In this work, we introduce a parametric family of three-step Jacobian-free iterative methods designed specifically for large-scale problems. Our approach bypasses the explicit computation and factorization of the Jacobian matrix often the most expensive component in Newton–Raphson schemes by employing a single divided-difference operator per iteration.

Memory-based techniques, such as those involving the Kurchatov divided difference, have been proposed to enhance the convergence efficiency of classical iterative methods [1, 2]. In our approach, we replace the traditional Jacobian matrix of F with the divided difference $\left[w^{(n)}, x^{(n)}; F\right]$, where $w^{(n)} = x^{(n)} + \mathbf{1}_n \cdot \lambda \left\|F\left(x^{(n)}\right)\right\|_2^2$, and $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector of ones. Here, the squared Euclidean norm is given by $\|F(x)\|_2^2 = F(x)^T F(x)$. This formulation avoids costly evaluations of the Jacobian. To the best of our knowledge, this specific strategy is novel and particularly well suited for nonlinear and non-differentiable problems.

The core contribution of our work is the systematic elevation of convergence order by three: given any Jacobian-free method of order p, our three-step modification achieves order $p+3$ without additional function calls. We build on the Taylor expansion technique for $F(x)^T F(x)$ proposed by Singh et al. [3], ensuring a theoretical gain in asymptotic convergence.

Previous high-order schemes include Wang and Zhang's family of seventh-order methods [4], and the seventh- and eighth-order memory-based approaches by Sharma et al. in [5, 6] and Ahmad et al. [7]. Other authors [8, 10, 11] have similarly employed divided difference operators, while weight-function techniques have been shown to further enhance Jacobian-free schemes [9].

[1] **miguel.leonardo@isfodosu.edu.do, mleonardo@itla.edu.do**

To validate our methods, we solve several benchmark problems, including systems of nonlinear equations with dimensions ranging from $n = 5$ to $n = 60$, one nonlinear integral equations discretized via Simpson's rule and a nonlinear.

We present both the theoretical $p \to p + 3$ convergence and practical performance gains up to 50% CPU time savings when compared to existing seventh- and eighth-order Jacobian-free algorithms.

1.1 Preliminaries and Notation

Let $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ be sufficiently smooth and x^* a simple root, $F(x^*) = 0$. A sequence $\{x^{(k)}\}$ converges with order $p > 1$ if

$$\|x^{(k+1)} - x^*\| \leq M \, \|x^{(k)} - x^*\|^p, \quad k \geq k_0.$$

The *Approximated Computational Order of Convergence* (ACOC) is computed by

$$p \approx ACOC = \frac{\ln\big(\|x^{(k+1)} - x^{(k)}\|/\|x^{(k)} - x^{(k-1)}\|\big)}{\ln\big(\|x^{(k)} - x^{(k-1)}\|/\|x^{(k-1)} - x^{(k-2)}\|\big)}.$$

For comparing the efficiency of the methods, we consider the following indices:

1. **Ostrowski Efficiency Index** [14]
$$I = p^{1/d},$$
where d is the number of function evaluations per iteration.

2. **Traub's Operational Index** [13]
$$C = p^{1/\mathrm{op}},$$
where op denotes the number of elementary operations per iteration (for example, solving r linear systems of size n via LU factorization requires $\frac{1}{3}n^3 + r\,n^2 - \frac{1}{3}n$ multiplications/divisions).

3. **Combined Efficiency Index**
$$\mathrm{CI} = p^{1/(d+\mathrm{op})}. \tag{1}$$

4. Inspired by Traub's theoretical formulation, we also introduce an **Empirical Computational Efficiency index based on real execution time per iteration**. This index is defined as:
$$\mathrm{IEC}_{\mathrm{time}} = p^{1/\tau}$$
where p is the estimated convergence order and τ is the average computational time per iteration, given by:
$$\tau = \frac{t_{\mathrm{total}}}{k},$$
with t_{total} denoting the total execution time of the method and k the total number of iterations required for convergence. This empirical metric enables a practical comparison of methods when theoretical operation counts are either insufficient or unavailable, especially in nonlinear or ill-conditioned problems.

We denote the first-order divided-difference operator by

$$[x, y; F]\,(x - y) \;=\; F(x) - F(y),$$

and then invoke the Genocchi–Hermite representation to obtain higher-order divided differences

$$[x_0, \ldots, x_m; F]$$

directly thus raising the convergence order without requiring any explicit derivatives of F.

2 Methods

2.1 Three-Step Jacobian-Free Scheme

Theorem 1 *Let us consider $F : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^n$ a differentiable enough function defined in a neighbourhood D of x^*, such that $F(x^*) = 0$. Let us also assume that $F'(x^*)$ is non-singular. Therefore, being $x^{(0)}$ is sufficiently close to the required solution x^*, then the iterative scheme defined by (2), whose initial two steps form a method of order p, achieves at least order of convergence $p + 3$ with the addition of the third step.*

Our scheme is

$$\begin{cases} y^{(n)} = x^{(n)} - [\, w^{(n)}, x^{(n)}; F\,]^{-1} F\big(x^{(n)}\big), \\ z^{(n)} = \varphi\big(x^{(n)}, y^{(n)}\big) \quad \textit{(auxiliary method of order p)}, \\ x^{(n+1)} = z^{(n)} - \Big[\alpha I + G\big(x^{(n)}, y^{(n)}\big)\big(\beta I + G\big(x^{(n)}, y^{(n)}\big)\big)\Big]\, [\, w^{(n)}, x^{(n)}; F\,]^{-1} F\big(z^{(n)}\big). \end{cases} \tag{2}$$

with

$$G(x, y) = [w, x; F]^{-1}[z, y; F].$$

For parameters $\alpha, \beta, \gamma \in \mathbb{R}$, given by

$$(\alpha, \beta, \gamma) = (\tfrac{13}{4},\, -\tfrac{7}{2},\, \tfrac{5}{4}).$$

Finally, the error satisfies

$$e^{(n+1)} = \big[C_2\,\big(\tfrac{3}{2}\,\lambda\big(P\,C_2 - C_2\,P + 4\,Q\big) - \big(13\,C_2^2 - C_3\big)\big) + 2\,C_3\big(9\,\lambda\,P + C_2\big)\big]\; M_p\, e^{(n)^{p+3}} + \mathcal{O}\big(e^{(n)^{p+4}}\big), \tag{3}$$

where the parameter λ is determined empirically to optimize the trade-off between convergence accuracy and numerical stability.

One proves that if the first two steps have order p, the full method attains order $p + 3$. No extra evaluations of F are required beyond those in the base method plus the divided differences.

2.2 Analogous High-Order Schemes

In addition to the main three-step family, one can similarly construct three other Jacobian-free Steffensen-type methods that also raise any base order p to $p + 3$ without extra function evaluations:

Case 2.

$$\begin{cases} y^{(n)} = x^{(n)} - \big[w^{(n)}, x^{(n)}; F\big]^{-1} F\big(x^{(n)}\big), \\ z^{(n)} = \varphi\big(x^{(n)}, y^{(n)}\big), \\ x^{(n+1)} = z^{(n)} - \Big\{\tfrac{1}{4} I + G\big(x^{(n)}, y^{(n)}\big)\big(\tfrac{1}{2} I + G\big(x^{(n)}, y^{(n)}\big)\big)\Big\}\, \big[w^{(n)}, x^{(n)}; F\big]\, F\big(z^{(n)}\big), \end{cases}$$

where $G(x^{(n)}, y^{(n)}) = \big[z^{(n)}, y^{(n)}; F\big]^{-1}\big[w^{(n)}, x^{(n)}; F\big]$,, $(\alpha, \beta, \gamma) = \big(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}\big)$ and the error equation is

$$e^{(n+1)} = \big[\,2C_3(2M_{p+1} - 9C_2M_p) + C_2\big(\tfrac{3}{2}\lambda(C_2P - 5PC_2) + (C_2^2 + C_3)\big)M_p\big]\, e^{(n)^{p+3}} + \mathcal{O}(e^{p+4}).$$

Case 3.

$$\begin{cases} y^{(n)} = x^{(n)} - \big[w^{(n)}, x^{(n)}; F\big]^{-1} F\big(x^{(n)}\big), \\ z^{(n)} = \varphi\big(x^{(n)}, y^{(n)}\big), \\ x^{(n+1)} = z^{(n)} - \Big\{\tfrac{5}{4} I - \tfrac{1}{2}\, G\big(x^{(n)}, y^{(n)}\big) + \tfrac{1}{4}\, G\big(x^{(n)}, y^{(n)}\big)^2\Big\}\, \big[z^{(n)}, y^{(n)}; F\big]\, F\big(z^{(n)}\big), \end{cases}$$

where $G(x^{(n)}, y^{(n)}) = \big[w^{(n)}, x^{(n)}; F\big]^{-1}\big[z^{(n)}, y^{(n)}; F\big]$, $(\alpha, \beta, \gamma) = \big(\tfrac{5}{4}, -\tfrac{1}{2}, \tfrac{1}{4}\big)$. and the error equation is

$$e^{(n+1)} = -\tfrac{1}{2}\big[C_2(12\lambda C_2P + 3\lambda PC_2 - 21C_2^2) + 4(C_3C_2 + 3C_2C_3)\big]M_p\, e^{(n)^{p+3}} + \mathcal{O}(e^{p+4}).$$

Case 4.

$$\begin{cases} y^{(n)} = x^{(n)} - \left[w^{(n)}, x^{(n)}; F\right]^{-1} F\left(x^{(n)}\right), \\ z^{(n)} = \varphi\left(x^{(n)}, y^{(n)}\right), \\ x^{(n+1)} = z^{(n)} - \left\{\frac{5}{4} I - \frac{1}{2} G\left(x^{(n)}, y^{(n)}\right) + \frac{1}{4} G\left(x^{(n)}, y^{(n)}\right)^2\right\} \left[z^{(n)}, y^{(n)}; F\right] F\left(z^{(n)}\right), \end{cases}$$

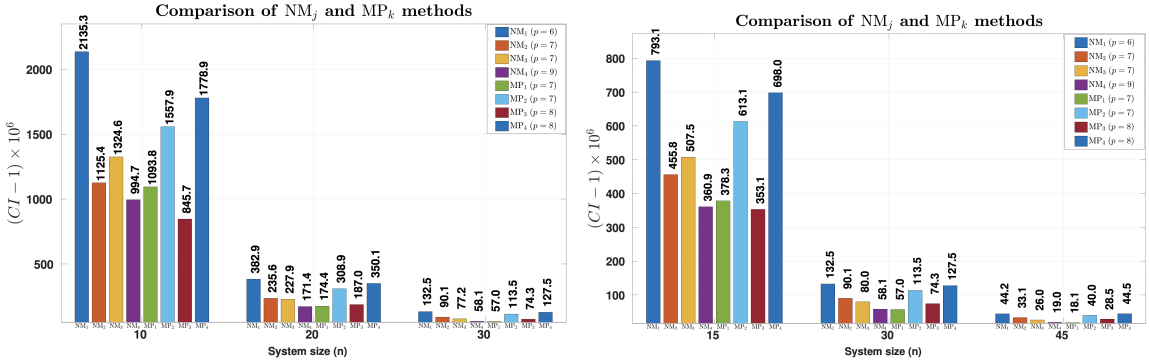
where $G(x^{(n)}, y^{(n)}) = \left[z^{(n)}, y^{(n)}; F\right]^{-1} \left[w^{(n)}, x^{(n)}; F\right]$, $(\alpha, \beta, \gamma) = \left(\frac{5}{4}, -\frac{1}{2}, \frac{1}{4}\right)$ and the error equation is

$$e^{(n+1)} = C_2\left(3\lambda C_2 P + 3C_3 - 2C_2^2\right) M_p \, e^{(n)^{p+3}} + \mathcal{O}(e^{p+4}).$$

In all cases $w^{(n)} = x^{(n)} + I \cdot \lambda \|F(x^{(n)})\|_2^2, \quad \|F(x)\|_2^2 = F(x)^T F(x)$.

3 Numerical Results

The graphs 1.1(a) and 1.1(b) show, for each method tested, bars of height $\Delta\text{CI} = (\text{CI} - 1) \times 10^6$, where CI it's given by 1 with p the convergence order, d the number of function evaluations per iteration, and op the count of elementary operations. Scaling $(\text{CI} - 1)$ by 10^6 renders minute differences (of order 10^{-6}) clearly visible. This facilitates a direct visual comparison of each method's efficiency advantage on large systems.



(a) Efficiency indices for nonlinear systems of size $n = 10, 20, 30$ using the MN_j and MP_k methods.

(b) Efficiency indices for nonlinear systems of size $n = 15, 30, 45$ using the MN_j and MP_k methods.

Figure 1: Computational deficiency index of MN$_j$ and MP$_k$ methods.

Both method families exhibit high CI indices for small and medium systems, but NM$_j$ methods maintain better CI stability as system size grows, confirming their superior robustness and per-iteration efficiency, especially for nonlinear or ill-conditioned problems.

3.1 Solution of the non-linear system

The tables report initial estimates $x^{(0)}$, iterate and residual errors, iteration counts (ITER), ACOC, computational times, and IEC. To solve the different academic problems in all tests used MATLAB with 4000-digit precision and $\epsilon = 10^{-200}$ on a macOS Sequoia 15.1.1 system (M2, 8 GB RAM), applying a uniform stopping criterion.

The following non-linear function is considered:

$$F(x) = x - \cos\left(2\,x - \sum_{i=1}^{4} x_i\right), \quad x \in \mathbb{R}^{(60)} \tag{4}$$

The initial estimate that we will use to solve the system (4) is $x^{(0)} = \left(\frac{11}{20}, \overset{60}{\cdots}, \frac{11}{20}\right)^T$, and its solution is $x^* = \left(0.51493326466112\ldots, \overset{60}{\cdots}, 0.51493326466112\ldots\right)^T$.

$x^{(0)}$	Schemes	$\|x^{(k+1)} - x^{(k)}\|$	$\|F(x^{(k+1)})\|$	Iter	ACOC	e-time	IEC
$x^{(0)} = \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix}$	NM_1	1.41637×10^{-347}	$3.10399 \times 10^{-2086}$	8	6.000000	8.240953	5.693761
	NM_2	1.14139×10^{-531}	$2.01963 \times 10^{-3722}$	8	7.000000	12.510373	3.470678
	NM_3	5.41030×10^{-544}	$5.49362 \times 10^{-3809}$	8	7.000000	11.880413	3.707403
	NM_4	$5.09381 \times 10^{-1123}$	0.0	8	9.000000	12.698978	3.991591
	MP_1	1.00880×10^{-495}	$4.05218 \times 10^{-3470}$	4	7.000000	17.926764	1.543714
	MP_2	2.77838×10^{-525}	$3.88599 \times 10^{-3677}$	4	6.923900	11.965970	1.909475
	MP_3	$3.68258 \times 10^{-1175}$	$4.10627 \times 10^{-5007}$	4	8.000000	51.960590	1.173603
	MP_4	9.09797×10^{-762}	$5.23722 \times 10^{-3446}$	8	8.000000	9.352106	5.922749

Table 1: Numerical results for System Nonlinear Equation (4), including IEC values.

3.2 Nonlinear Integral Equations

We validate our methods on two nonlinear integral equations discretized into nonlinear systems. The first equation, from [12], is given by

$$y(t) = \frac{t}{e} + \int_0^1 2ts\, e^{-y(s)^2}\, \mathrm{d}s, \quad t \in [0,1], \tag{5}$$

For this test, we discretized the integral equation using Simpson's rule with 30 subintervals, set the number of subintervals in the domain to $m = 50$, chose the initial estimate $x^{(0)} = \left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right)^T$, carried out all computations in variable-precision arithmetic with a 1000-digit mantissa, and employed a convergence tolerance of 10^{-100}. The table then reports, for each NM_j and MP_k scheme initialized at $x^{(0)} = \left[\frac{1}{2}, \ldots, \frac{1}{2}\right]^T$, the Euclidean norm of the update step $\left\|x^{(k+1)} - x^{(k)}\right\|$, which quantifies the change between successive iterates; the residual norm $\left\|F\left(x^{(k+1)}\right)\right\|$, measuring how closely the current approximation satisfies the nonlinear equation; the total number of iterations (Iter) required to satisfy the convergence criterion; the Approximate Computational Order of Convergence (ACOC), estimated empirically from the sequence of update norms; the wall-clock execution time in seconds (e-time); and the Computational Efficiency Index (IEC), defined as $\mathrm{IEC}_{\mathrm{time}} = p^{1/\tau}, \tau = \frac{t_{total}}{k}$, which synthesizes the observed convergence order p and the average computational cost per iteration to provide a unified measure of each method's overall efficiency.

$x^{(0)}$	Schemes	$\|x^{(k+1)} - x^{(k)}\|$	$\|F(x^{(k+1)})\|$	Iter	ACOC	e-time	IEC
$x^{(0)} = \begin{bmatrix} \frac{1}{2} \\ \vdots \\ \frac{1}{2} \end{bmatrix}$	NM_1	1.37866×10^{-306}	$4.56978 \times 10^{-1008}$	8	6.000000	191.257114	1.077827
	NM_2	2.92475×10^{-376}	$3.85530 \times 10^{-1008}$	8	7.000000	287.715976	1.055597
	NM_3	3.47957×10^{-377}	$4.64222 \times 10^{-1008}$	8	7.000000	294.095193	1.054359
	NM_4	2.23872×10^{-746}	1.93779×10^{-769}	8	9.000000	305.912979	1.059143
	MP_1	6.56960×10^{-271}	$6.60741 \times 10^{-1008}$	4	6.999997	447.305455	1.017553
	MP_2	9.51629×10^{-292}	0.0	4	6.999998	290.279651	1.027600
	MP_3	5.45603×10^{-123}	2.18896×10^{-992}	3	7.833821	212.089732	1.029545
	MP_4	9.43737×10^{-652}	$4.35290 \times 10^{-2099}$	8	8.000000	282.280214	1.060704

Table 2: Numerical results for the integral Equation (5), using high-precision arithmetic and real-time IEC.

4 Conclusions

In conclusion, in Theorem 1 it was shown that the proposed method raises the convergence order of any Steffensen-type algorithm from p to $p + 3$. Numerical tests performed on the modified schemes corroborate this theoretical improvement and also show that our methods systematically outperform the benchmark methods: not only do they achieve higher accuracy, but they also reduce the computation time by up to 50% in certain cases. Moreover, we introduced and computed the empirical computational efficiency index $\mathrm{IEC}_{\mathrm{time}}$, a runtime-based metric that is essential when theoretical counts omit implementation overhead, and which clearly evidences the advantage of our schemes in terms of number of iterations and real cost. Finally, we applied the technique to several problems nonlinear systems and solving a discretized integral

equation by Simpson's rule and in all cases our methods demonstrated superior performance and remarkable robustness.

Acknowledgments

This work was supported by ISFODOSU, INTEC, and ITLA. The authors also express their gratitude to the Universitat Politècnica de València for its valuable support.

References

[1] Cordero, A., Villalba, E.G., Torregrosa, J.R., Triguero-Navarro, P., Convergence and stability of a parametric class of iterative schemes for solving nonlinear systems, *Mathematics*, 9(1):686–698, 2021.

[2] Chicharro, F.I., Cordero, A., Garrido, N., Torregrosa, J.R., A new efficient parametric family of iterative methods for solving nonlinear systems, *Journal of Difference Equations and Applications*, 25(9–10):1454–1467, 2019.

[3] Singh, H., Sharma, J.R., Kumar, S., A simple yet efficient two-step fifth-order weighted-Newton method for nonlinear models, *Numerical Algorithms*, 93(1):203–225, 2023.

[4] Wang, X., Zhang, T., A family of Steffensen-type methods with seventh-order convergence, *Numerical Algorithms*, 62:429–444, 2013.

[5] Sharma, J.R., Arora, H., A novel derivative-free algorithm with seventh-order convergence for solving systems of nonlinear equations, *Numerical Algorithms*, 67:917–933, 2014.

[6] Sharma, J.R., Arora, H., Efficient higher order derivative-free multipoint methods with and without memory for systems of nonlinear equations, *International Journal of Computer Mathematics*, 95(5):920–938, 2017.

[7] Ahmad, F., Soleymani, F., Khaksar Haghani, F., Serra-Capizzano, S., Higher order derivative-free iterative methods with and without memory for systems of nonlinear equations, *Applied Mathematics and Computation*, 314:199–211, 2017.

[8] Sharma, J.R., Arora, H., Efficient derivative-free numerical methods for solving systems of nonlinear equations, *Computational & Applied Mathematics*, 35(1):269–284, 2016.

[9] Sharma, R.K., Guha, J.R., Sharma, R., A efficient fourth-order weighted-Newton method for systems of nonlinear equations, *Numerical Algorithms*, 62:307–323, 2013.

[10] Singh, A., An efficient fifth-order Steffensen-type method for solving systems of nonlinear equations, *International Journal of Computing Science and Mathematics*, 9(5):501–514, 2018.

[11] Wang, X., Zhang, T., Qian, W., Teng, M., Seventh-order derivative-free iterative method for solving nonlinear systems, *Numerical Algorithms*, 10(3):545–558, 2015.

[12] Hueso, J.L., Martínez, E., Teruel, C., Convergence, efficiency and dynamics of new fourth and sixth order families of iterative methods for nonlinear systems, *Journal of Computational and Applied Mathematics*, 275:412–420, 2015.

[13] Traub, J.F., *Iterative Methods for the Solution of Equations*, Providence, American Mathematical Society, 1982.

[14] Ostrowski, A.M., *Solution of Equations and Systems of Equations: Pure and Applied Mathematics: A Series of Monographs and Textbooks, vol. 9*, New York, Elsevier, 1966.

Extending the applicability of a fifth order iterative method for solving nonlinear Hammerstein-type integral equations

Lovish Dua[a], Sukhjit Singh[a], Eulalia Martínez[b], M.A. Hernández-Verón[c], Ramandeep Behl[d]

(*a*) Department of Mathematics and Computing,
Dr B. R. Ambedkar National Institute of Technology, Jalandhar, India.

(*b*) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València. València, Spain.

(*c*) Department of Mathematics and Computation,
University of La Rioja, Spain

(*d*) Department of Mathematics,
King Abdulaziz University, Jeddah 21589, Saudi Arabia

1 Introduction

In this study, we consider a nonlinear Hammerstein-type integral equations [1],[3] of the form

$$u(x) = f(x) + \mu \int_{\alpha}^{\beta} X(x,t)P(u)(t)\ dt, \quad x \in [\alpha, \beta], \tag{1}$$

where, $\mu \in \mathbb{R}$, $f(x) \in \mathbb{C}[\alpha, \beta]$, $X(x,t) \in [\alpha, \beta] \times [\alpha, \beta]$, is a non-separable kernel, $P : \Omega \subseteq \mathbb{C}[\alpha, \beta] \to \mathbb{C}[\alpha, \beta]$ is Nemytskii operator with $P(u)(x) = \mathcal{P}(u(x))$, where $\mathcal{P} : \mathbb{R} \to \mathbb{R}$ and $\Omega \neq \phi$ is open convex subset, $u(x)$ is a unknown function that needs to be found.

For this, we consider the operator $\mathcal{G} : \Omega \subseteq \mathbb{C}[\alpha, \beta] \to \mathbb{C}[\alpha, \beta]$ with

$$[\mathcal{G}(u)](x) = u(x) - f(x) - \mu \int_{\alpha}^{\beta} X(x,t)P(u)(t)\ dt, \quad x \in [\alpha, \beta], \quad \mu \in \mathbb{R}. \tag{2}$$

Generally, Newton's method is most widely used iterative method for solving nonlinear equations. Various researchers [4]-[5] have considered the following fifth order iterative method

$$\begin{cases} v_n = u_n - [\mathcal{G}'(u_n)]^{-1}\mathcal{G}(u_n), \\ w_n = v_n - 5[\mathcal{G}'(u_n)]^{-1}\mathcal{G}(v_n), \\ u_{n+1} = w_n - \frac{1}{5}[\mathcal{G}'(u_n)]^{-1}(-16\mathcal{G}(v_n) + \mathcal{G}(w_n)), \end{cases} \tag{3}$$

with $u_0 \in \mathbb{C}[\alpha, \beta]$ for solving nonlinear equations under various continuity conditions. However, the existing studies are applicable if the kernel of integral equation (1) is separable, *i.e.,* $X(x,t) = \sum_{j=1}^{m} c_j(x)d_j(x)$, where $c_j, dj \in \mathbb{C}[\alpha, \beta]$ for $j = 1, 2, \cdots, m$. The iterative method (3) presents a significant challenge for non-separable kernels and we are unable to explicitly calculate the operator $[\mathcal{G}'(u_n)]^{-1}$ at each iterative step. Hernández et al. [6] established the semilocal convergence of

method (3) under weaker continuity condition for solving integral equation with non-separable kernel. Recently, Ezquerro and Hernández [7] established the restricted global convergence of Newton type iterative method for solving (1), which is more general as compared to semilocal convergence. Motivated by this work, we will establish the restricted global convergence of a fifth order method (3) which is more efficient than the Newton's method for solving (1).

For this, we onsider the operator $\mathcal{J} : \Omega \subseteq \mathbb{C}[\alpha, \beta] \to \mathbb{C}[\alpha, \beta]$ with

$$[\mathcal{J}(u)](x) = u(x) - f(x) - \mu \int_{\alpha}^{\beta} Y(x,t)P(u)(t)\ dt, \quad x \in [\alpha, \beta], \quad \mu \in \mathbb{R}, \tag{4}$$

where, kernel $X(x,t)$ is approximated by a separable kernel $Y(x,t)$ such that $X(x,t) = Y(x,t) + W(\theta, x, t)$. Here, $Y(x,t) = \sum_{j=1}^{m} \rho_j(x)\zeta_j(x)$, where $\rho_j, \zeta j \in \mathbb{C}[\alpha, \beta]$ for $j = 1, 2, \cdots, m$. Afterthat, we can consider $[\mathcal{J}'(u_n)]^{-1}$ as an approximation of $[\mathcal{G}'(u_n)]^{-1}$ and obtain the following iterative scheme

$$\begin{cases} v_n = u_n - [\mathcal{J}'(u_n)]^{-1}\mathcal{G}(u_n), \\ w_n = v_n - 5[\mathcal{J}'(u_n)]^{-1}\mathcal{G}(v_n), \\ u_{n+1} = w_n - \frac{1}{5}[[\mathcal{J}'(u_n)]^{-1}(-16\mathcal{G}(v_n) + \mathcal{G}(w_n)), \end{cases} \tag{5}$$

There are two primary objectives of this work. The first is to demonstrate that the fifth order iterative method given in (5) is a good approximation of a fifth order iterative method (3) *i.e.,* the estimates provided by the two approaches are identical. To show this, we prove that the operators $[\mathcal{G}'(u_n)]^{-1}$ and $[\mathcal{J}'(u_n)]^{-1}$ are close as we require. The second objective is a qualitative study for a fifth order iterative method (5). Using the technique of recurrence relations, we establish the global convergence domains for the solution of a nonlinear Hammerstein-type integral equations.

2 Restricted global convergence analysis

In this section, we investigate the restricted global convergence of a fifth order iterative method (5) under the following assumptions:

(I) Let $\Omega \neq \phi$ be a open convex subset of $\mathbb{C}[\alpha, \beta]$. Suppose there exists an operator $[\mathcal{J}'(\tilde{u})] \in \mathcal{L}(\mathbb{C}[\alpha, \beta], \mathbb{C}[\alpha, \beta])$ for some $\tilde{u} \in \Omega$, satisfying $\|[\mathcal{J}'(\tilde{u})]^{-1}\| \leq \beta$ and $\|[\mathcal{J}'(\tilde{u})]^{-1}J(\tilde{u})\| \leq \eta$.

(II) $\exists$ a constant $A_0 \geq 0$ such that $\|P(u) - P(\tilde{u})\| \leq A_0\|u - \tilde{u}\|$, for $u \in \Omega$.

(III) $\exists$ a constant $\tilde{A} \geq 0$ such that $\|P'(u) - P'(v)\| \leq \tilde{A}\|u - v\|$, for $u, v \in \Omega$.

Now, we construct the following Lemma's to establish the existence and uniqueness domains for the solution.

Lemma 2.1. *Let $u, v \in \Omega$ and assume that the above conditions (I)-(III) hold. Then*

(a) $\|\mathcal{J}'(u) - \mathcal{J}'(v)\| \leq M\|u - v\|$, *where* $M = |\mu|R\tilde{A}$ *and* $R = \max_{x\in[\alpha,\beta]} \int_{\alpha}^{\beta} |Y(x,t)|\ dt$.

(b) $\|\mathcal{G}'(u) - \mathcal{G}'(v)\| \leq K\|u - v\|$, *where* $K = |\mu|S\tilde{A}$ *and* $S = \max_{x\in[\alpha,\beta]} \int_{\alpha}^{\beta} |X(x,t)|\ dt$.

(c) $\|\mathcal{J}(u) - \mathcal{G}(u)\| \leq \kappa(\|u - \tilde{u}\|)$ *with* $\kappa : \mathbb{R} \to \mathbb{R}$ *and* $\kappa(\xi) = |\mu|Q(\|P(\tilde{u})\| + A_0 t)$ *where* $Q = \max_{x\in[\alpha,\beta]} \int_{\alpha}^{\beta} |W(\theta, x, t)|\ dt$.

Lemma 2.2. *Assume that conditions (I)-(III) hold. If* $\|u-\tilde{u}\| < \frac{1}{M\beta}$ *and* $u \in \Omega$, *then* $\exists\ [\mathcal{J}'(u)]^{-1}$. *Moreover,*

$$\|[\mathcal{J}'(u)]^{-1}\mathcal{J}'(\tilde{u})\| \leq h(M\beta\|u-\tilde{u}\|)$$

and

$$\|[\mathcal{J}'(u)]^{-1}\| \leq \beta h(M\beta\|u-\tilde{u}\|),$$

with $h:(0,1)\to\mathbb{R}$ *and* $h(t)=\dfrac{1}{1-t}$.

Lemma 2.3. *Assume that the conditions (I)-(III) hold, then*

(a) $\|\mathcal{G}(u_n)\| \leq \frac{M}{2}\|u_n-u_{n-1}\|^2+\omega(\|u_{n-1}-\tilde{u}\|+\|u_n-\tilde{u}\|)\|u_n-u_{n-1}\|$, *where* $\omega(t)=|\mu|Q(\|P'(\tilde{u})\|+\frac{\tilde{A}}{2}t)$ *with* $\omega:\mathbb{R}\to\mathbb{R}$, *for* $n\geq 1$.

(b) $\|v_n-u_n\| \leq r(c_n)\|v_{n-1}-u_{n-1}\|$ *with* $r:\mathbb{R}\to\mathbb{R}$ *and* $r(t)=\triangle_n\ (1+p(t))-p(t)$, $p:\mathbb{R}\to\mathbb{R}$ *and* $p(t)=\frac{1}{2}t+\frac{1}{2}t^2+\frac{5}{8}t^3$, *where* $c_n=\dfrac{K\beta\|v_n-u_n\|}{1-M\beta\|u_n-\tilde{u}\|}$ *and*

$$\triangle_n = \beta h(M\beta(\|u_n-\tilde{u}\|)\left[\frac{M}{2}\|u_n-u_{n-1}\|^2+\delta(\|u_{n-1}-\tilde{u}\|+\|u_n-\tilde{u}\|)\right].$$

(c) $\|v_n-\tilde{u}\| \leq \tilde{\triangle}_n$, *where* $\tilde{\triangle}=h(M\beta\|u_n-\tilde{u}\|)\left[\beta\mu(\|u_n-\tilde{u}\|)+\eta+\frac{M}{2}\beta\|u_n-\tilde{u}\|^2\right]$.

(d) $\|w_n-v_n\| \leq \frac{5}{2}c_n\|v_n-u_n\|$.

(e) $\|w_n-\tilde{u}\| \leq \frac{5}{2}c_n\|v_n-u_n\|+\tilde{\triangle}_n$.

(f) $\|u_{n+1}-u_n\| \leq [1+p(c_n)]\|v_n-u_n\|$.

(g) $\|u_{n+1}-\tilde{u}\| \leq [1+p(c_n)]\|v_n-u_n\|+\tilde{\triangle}_n$.

3 Existence and uniqueness domains

In this section, we establish a existence and uniqueness theorem for the solution of a nonlinear Hammerestein-type integral equation (1).

Theorem 1. *Assume that the conditions (I)-(III) hold and let* $\tilde{R}\geq 0$ *such that* $\|u-\tilde{u}\| < \frac{1}{M\beta}, r(t)=\triangle_n\ (1+p(t))-p(t)<1$ *and* $[1+p(t)]\|v_n-u_n\|+\tilde{\triangle}_n \leq \tilde{R}$. *If* $B(\tilde{u},\tilde{R})\subset\Omega$, *then for any starting point* $u_0\in B(\tilde{u},\tilde{R})$, *the iterative method converges to a solution* u^* *of* $G(u)=0$ *in* $B(\tilde{u},\tilde{R})$. *Moreover,* $u_n,v_n,w_n,u^*\in B(\tilde{u},\tilde{R})$ *for* $n\geq 0$.

Theorem 2. *The solution* u^* *of the equation* $\mathcal{G}(u)=0$ *is unique in* $B(\tilde{u},\tilde{r})\cap\Omega$, *under the conditions of Theorem 1, where* $\tilde{r}=\frac{2}{\tilde{A}}\left(\frac{1}{|\mu|S}-\|P'(\tilde{u})\|\right)-\tilde{R}$.

We will also demonstrate the applicability of our approach by verifying the aforementioned theoretical findings for solving the nonlinear Hammerstein type integral equations.

References

[1] Argyros, I. K and Regmi, S., Undergraduate research at Cameron University on iterative procedures in Banach and other spaces, Nova Science Publishers, New york, 2019

[2] Regmi, S., Optimized iterative methods with applications in diverse disciplines, Nova Science Publishers, New york, 2021

[3] Wazwaz, A. M., A first course in integral equations, World Scientific Publishing Company, Singapore, 2015

[4] Arroyo, V., and Cordero, A., and Torregrosa, J. R., Approximation of artificial satellites' preliminary orbits: the efficiency challenge, *Mathematical and Computer Modelling*, 54(7-8):1802–1807, 2011.

[5] Yadav, S., and Singh, S., Global convergence domains for an efficient fifth order iterative scheme, *Journal of Mathematical Chemistry*, 61(10): 2176–2191, 2023.

[6] Hernández, M. A., Yadav, S., Martínez, E., and Singh, S., Solving nonlinear integral equa-tions with non-separable kernel via a high-order iterative process, *Applied Mathematics and Computation*, 409, 2021.

[7] Ezquerro, J. A., and Hernández, M. A., Location, separation and approximation of solutions of nonlinear Hammerstein-type integral equations, *Applied Numerical Mathematics*, 198, 1–10, 2024.

On a non-linear approximation method: Shepard WENO-based method

A. David Levin[♭], B. José M. Ramón[♮],[1] C. Juan Ruiz-Álvarez[♯] and D.Dionisio F. Yáñez[♮]

(♭) University of Mathematical Science,Tel-Aviv University,
229 Schreiber Building, Ramat Aviv,Israel

(♮) Departament de Matemàtiques, Universitat de València,
Carrer Dr Moliner 50, València, Spain.

(♯) Universidad Politécnica de Cartagena,
Plza Cronista Isidro Valverde,Cartagena, Murcia, Spain.

1 Introduction

The Shepard method is a commonly and effective algorithm for interpolating scattered data across multiple dimensions (see, e.g. [1,4,6]). This well-established numerical analysis approach is founded on the idea that data farther from the approximation point should have a smaller influence on the final interpolation. However, accurately approximating piecewise smooth functions in $\mathbb{R}^n$ near discontinuities along a hypersurface in $\mathbb{R}^{n-1}$ remains a challenge for Shepard's method and other linear sparse-data techniques, due to the inherent difficulty in accurately capturing sharp transitions and avoiding oscillations.

This study presents a nonlinear adaptation of the Shepard method [3] by integrating concepts from the weighted essentially non-oscillatory (WENO) interpolation [5]. Rather than exclusively using inverse squared distance weights, we dynamically adjust the weighting function to suppress oscillations near discontinuities [2] and improve interpolation accuracy. Numerical results demonstrate the improved effectiveness of the proposed method, establishing it as a reliable tool for scientific computing and data analysis.

2 Methods

We assume an open and bounded domain $\Omega \subseteq \mathbb{R}^n$, a set of N distinct nodes

$$\chi_N = \{\mathbf{x}_i \in \Omega : i = 1, \ldots, N\}$$

and a corresponding set of function values

$$\mathcal{F}_N = \{f_i = f(\mathbf{x}_i) : i = 1, \ldots, N\},$$

where $f : \Omega \to \mathbb{R}$ is unknown. We define the fill distance as

$$h \equiv \sup_{\mathbf{x}\in\Omega} \min_{\mathbf{x}_i\in\chi_N} \|\mathbf{x} - \mathbf{x}_i\|,$$

[1] Jose.Manuel.Ramon@uv.es

Shepard's method can be viewed as a special case of the moving least squares technique, where the degree of the polynomial is zero, i.e., a constant.

Therefore if we consider a non-negative, compactly supported, radial function $\omega : \mathbb{R}^n \to \mathbb{R}$, and define

$$\omega_i(\mathbf{x}) = \omega\left(\frac{\|\mathbf{x}_i - \mathbf{x}\|}{h}\right), \tag{1}$$

the Shepard's approximant is defined as:

$$\mathcal{I}_{\mathrm{S}}(\mathbf{x}) = \sum_{i=1}^{N} \frac{\omega_i(\mathbf{x})}{\sum_{j=1}^{N} \omega_j(\mathbf{x})} f_i = \sum_{i=1}^{N} W_i(\mathbf{x}) f_i,$$

where $W_i : \Omega \to \mathbb{R}$ are defined as $W_i(\mathbf{x}) = \frac{\omega_i(\mathbf{x})}{\sum_{j=1}^{N} \omega_j(\mathbf{x})}$. It follows that for $i = 1, \ldots, N$, we have

$$0 \leq W_i(\mathbf{x}) \leq 1, \text{ and } \sum_{j=1}^{N} W_j(\mathbf{x}) = 1, \forall \mathbf{x} \in \mathbb{R}^n.$$

We call the W_i "*optimal Shepard weights*".

3 Results

Shepard's method is based on the distances between the evaluation point and the given nodes. However, if a discontinuity contaminates the data, or they present a strong gradient, it is then recommended to use only nodes that are free of singularities. To achieve this, the optimal Shepard weights are replaced with non-linear ones, defined as follows:

$$\mathcal{W}_i(\mathbf{x}) = \frac{\alpha_i(\mathbf{x})}{\sum_{j=1}^{N} \alpha_j(\mathbf{x})}, \quad \alpha_i(\mathbf{x}) = \frac{W_i(\mathbf{x})}{(\epsilon + I_i)^t}.$$

The values $\{I_i\}_{i=1}^{N}$ are smoothness indicators. They determine whether a node is near a discontinuity and should therefore have a minimal contribution to the approximation. The parameters ϵ and t are chosen to achieve maximal approximation order. In our case, we take $\epsilon = 10^{-14}$ and $t = 4$. We will describe the method for constructing the smoothness indicators. With these ingredients the new WENO-Shepard's method is:

$$\mathcal{I}_{\text{WENO-S}}(\mathbf{x}) = \sum_{i=1}^{N} \mathcal{W}_i(\mathbf{x}) f_i. \tag{2}$$

Therefore, in our experiments, we take $\varepsilon = \lfloor \frac{\sqrt{N_l}}{2} \rfloor / \sqrt{2}$. In Tables 1 we present a grid refinement analysis in l^∞ norm, using Wendland 2 and Wendland 4:

(W2)	Shepard		WENO-Shepard		Shepard		WENO-Shepard	
l	MAE_l	r_l^∞	MAE_l	r_l^∞	RMSE_l	r_l^2	RMSE_l	r_l^2
4	6.1891e-02	-	1.4160e-01	-	1.5976e-02	-	4.4803e-02	-
5	2.1657e-02	1.5149	5.4509e-02	1.3773	4.7667e-03	1.7448	1.5903e-02	1.4943
6	1.1315e-02	0.9365	1.7315e-02	1.6545	1.5991e-03	1.5758	4.6052e-03	1.7879
7	5.7795e-03	0.9693	4.6431e-03	1.8989	6.6941e-04	1.2563	9.5190e-04	2.2744
(W4)	Shepard		WENO-Shepard		Shepard		WENO-Shepard	
l	MAE_l	r_l^∞	MAE_l	r_l^∞	RMSE_l	r_l^2	RMSE_l	r_l^2
4	4.7978e-02	-	1.1916e-01	-	1.2350e-02	-	3.6371e-02	-
5	1.6849e-02	1.5097	4.3702e-02	1.4471	3.6356e-03	1.7643	1.2588e-02	1.5307
6	8.8017e-03	0.9368	1.3690e-02	1.6746	1.2311e-03	1.5623	3.6177e-03	1.7989
7	4.4875e-03	0.9719	3.6360e-03	1.9126	5.1835e-04	1.2479	7.3183e-04	2.3055

Table 1: Errors and rates using Shepard and WENO-Shepard's methods for Franke's test function on a regular grid.

4 Conclusions

WENO-Shepard's method presented in this article represents a significant advancement in the interpolation of scattered data, particularly in approximating piecewise smooth functions with discontinuities.

Using non-linear modification of the weight functions, the method demonstrates a remarkable ability to reduce diffusion effects near discontinuities.

References

[1] Gregory E. Fasshauer,Meshfree Approximation Methods with MATLAB,Chicago(USA), World Scientific Publishing, 2007.

[2] D. Gottlieb, C-W. Sh : On the Gibbs phenomenon and its resolution., *SIAM Rev.*, 39 (4), 644-668, 1997.

[3] David Levin, José M. Ramón, Juan Ruiz-Álvarez, Dionisio F. Yáñez: Weighted Essentially Non-Oscillatory Shepard method., *arxiv.org, arXiv:2412.02286*, 2024.

[4] D. Shepard.A two-dimensional interpolation function for irregularly-spaced data, *Proceedings of the 1968 23rd ACM national conference.*, 517-524. 1968.

[5] C.-W. Shu, High Order Weighted Essentially Nonoscillatory Schemes for Convection Dominated Problems", *SIAM Review*, 51(1), 82–126. 1999.

[6] H. Wendland,Scattered Data Approximation,Cambridge(UK),Cambridge University Press, 2004.

Two New Maximally Efficient Fifth- and Sixth-Order Methods for Nonlinear Vectorial Problems

Alicia Cordero[♭], Renso V. Rojas-Hiciano[♮],[1] Juan R. Torregrosa[♭] and Maria P. Vassileva[♭♭]

(♭) Multidisciplinary Mathematics Institute, Universitat Politècnica de València, Camí de Vera s/n, València, Spain.

(♮) Escuela de Ciencias Naturales y Exactas, Pontificia Universidad Católica Madre y Maestra, Sarasota Ave. 10109, Santo Domingo, Dominican Republic.

(♭♭) Área de Ciencias Básicas y Ambientales, Instituto Tecnológico de Santo Domingo, Los Próceres Ave. 10602, Santo Domingo, Dominican Republic.

1 Introduction

Many real-world problems can be modeled as systems of nonlinear equations $F(x) = 0$, where $F : D \subset \mathbb{R}^n$, with $n \in \mathbb{N}$, is a function defined on an open convex domain D.

To solve them, many studies have focused on improving Newton's scheme, which we denote N2S:

$$x^{(k+1)} = \mathcal{N}^{(k)} = x^{(k)} - F'\left(x^{(k)}\right)^{-1} F\left(x^{(k)}\right),\ k = 0, 1, \ldots, \tag{1}$$

where $F'(x) \in \mathscr{L}\left(\mathbb{R}^n, \mathbb{R}^n\right)$ is a linear operator representing the Jacobian matrix of F.

Other schemes are also often used, even with higher order of convergence but designed in the scalar context, whose extension to several variables is not always possible or obvious.

We can analyze and compare different iterative methods using the order of convergence, which indicates how quickly the sequence $\{x^{(k)}\}_{k\geq 0}$ generated by the iterative method approaches the root.

We take $e^{(k)} = x^{(k)} - \xi$ as the measure of the local error at the k^{th} iteration. The order of convergence p of an iterative method can be obtained from the error equation

$$e^{(k+1)} = Le^{(k)^p} + O\left(e^{(k)^{p+1}}\right),\ k = 0, 1, \ldots,$$

where $L \in \mathscr{L}(\overbrace{\mathbb{R}^n \cdots \times \mathbb{R}^n}^{p-\text{ times}}, \mathbb{R}^n)$ is a p linear operator and $\mathscr{L}$ is the set of bounded linear operators. In fact, N2S is a quadratic convergence order scheme.

Traub [1] proposed a two-step scheme that achieves 3rd order, we call it T3S:

$$x^{(k+1)} = \mathcal{N}^{(k)} - F'\left(x^{(k)}\right)^{-1} F\left(\mathcal{N}^{(k)}\right),\ k = 0, 1, \ldots, \tag{2}$$

where $\mathcal{N}^{(k)}$ is N2S.

[1] rv.rojas@ce.pucmm.edu.do

The search for good methods leads to ranking criteria, such as optimality. Kung and Traub established in [4] an upper bound for the order of convergence of scalar schemes. Those that reach it are called optimal. Similarly, the authors did the same for the vectorial case.

Conjecture 1 (Cordero-Torregrosa [5]). *Given any iterative method for solving systems of nonlinear equations, with k_1 evaluations of the Jacobian matrix and k_2 of the nonlinear function at each iteration. Then it is optimal if its order of convergence is $2^{k_1+k_2-1}$ and $k_1 \leq k_2$.*

The efficiency index defined by Ostrowski in [2] is extended to the vectorial case by being calculated using $I = p^{\frac{1}{d}}$, where d is the total number of functional evaluations per iteration, i.e. $d = k_1 n^2 + k_2 n$, k_1 is the number of evaluations of the Jacobian matrix, while k_2 denotes the number of evaluations of the nonlinear function and n the size of the system.

In addition, we consider the computational efficiency index, proposed by the authors in [3], defined as $CI = p^{\frac{1}{d+op}}$, where *op* is the number of products/quotients per iteration. The number of products and quotients needed to solve m linear systems with the same Jacobian matrix, applying LU factorization is $\frac{1}{3}(n^3 - n) + mn^2$.

Not few proposals lead to schemes whose computational cost is unsustainable, compared to others of lower order and lower cost. Therefore, having a high order is not enough to classify one method as better than another.

One indicator we consider is Effort, defined by Carson and Strakošin in [6] as the total number of evaluations and operations in an iterative process. It is calculated by:

$$\mathit{effort} = \mathit{iter} \times (d + op), \tag{3}$$

where d, *iter*, and *op* denote the number of evaluations, iterations, and operations, respectively.

Recently, in [7], the authors presented new criteria for classifying vectorial iterative schemes. This categorization is useful in this article, as it allows us to organize them according to their efficiency.

Definition 1 (Maximally Efficient Methods [7]). *A damped composed Newton-type iterative method of order P is Maximally Efficient at P-order, or MAXP, if no other method with the same order is more efficient.*

Proposition 2 (Maximally Efficient Methods [7]). *A damped composed Newton-type iterative scheme of order P is Maximally Efficient if and only if it has only one evaluation of Jacobian matrix, and as many steps as evaluations of the function F, which must be equal to $\lceil \log_2 P \rceil$.*

Convergence Order Accelerators

This research proposes a strategy to increase the convergence order of iterative schemes in a very efficient way compared to others introduced to date, under certain conditions.

Singh, Sharma, and Kumar in [8] derived a two-step scheme of order 5, two evaluations of the Jacobian matrix, and two linear systems with different inverse matrices, which we denote as SS5S:

$$x^{(k+1)} = \mathcal{N}^{(k)} - (1 + \nu_k)\, F'\left(\mathcal{N}^{(k)}\right)^{-1} F\left(\mathcal{N}^{(k)}\right), \tag{4}$$

where $k = 0, 1, ...$, $\mathcal{N}^{(k)}$ is the *k-th* iteration of N2S, and

$$\nu_k = \frac{\left\|F\left(\mathcal{N}^{(k)}\right)\right\|^2}{\left\|F(x^{(k)})\right\|^2} = \frac{F\left(\mathcal{N}^{(k)}\right)^T F(\mathcal{N}^{(k)})}{F(x^{(k)})^T F(x^{(k)})}. \tag{5}$$

This idea motivated the authors to design an optimal extension to the multivariate case of the Ermakov's Hyperfamily [9]. In this way, we presented in [10] the first optimal family for nonlinear

systems of order 4, according to the Cordero-Torregrosa conjecture. Its cost is lower than that of any other known scheme, due both to the frozen Jacobian matrix and scalar weight function:

$$x^{(k+1)} = \mathcal{N}^{(k)} - F'(x^{(k)})^{-1}\left(p_k F\left(\mathcal{N}^{(k)}\right) + q_k F(x^{(k)})\right), \quad k = 0.1, \ldots \tag{6}$$

where $p_k = \frac{1+\psi\nu_k}{1+\mu\,\nu_k}$, $q_k = \frac{2\,\nu_k}{1+\mu\,\nu_k}$, $(\mu, \psi) \in \mathbb{R}^2$, ν_k as defined in (5) and $\mathcal{N}^{(k)}$ is N2S.

Its best members behave in the scalar case like those of the uniparametric KLAM family, a subclass of the scheme (6) obtained by canceling the parameter ψ, and whose stability has been carefully studied in [9]. From it we have constructed free versions of Jacobian matrix evaluations in [11]. By canceling both parameters in (6), we obtain the simplest member of the scheme, denoted by EH4S. It is an optimal and MAX4 method:

$$x^{(k+1)} = \mathcal{N}^{(k)} - F'\left(x^{(k)}\right)^{-1}\left(F\left(\mathcal{N}^{(k)}\right) + 2\,\nu_k F\left(x^{(k)}\right)\right). \tag{7}$$

On the other hand, most of the known methods use non-scalar weight functions. For comparison we take the following two, high-order and efficient method, which has frozen Jacobian matrix.

Sharma and Arora in [13] have proposed a Jarratt-like scheme of sixth-order three-step, wich we denote by SA6S, with iterative formula

$$\begin{aligned}
y^{(k)} &= x^{(k)} - \frac{2}{3}\ \mathrm{F}'\left(x^{(k)}\right)^{-1}\ \mathrm{F}\left(x^{(k)}\right), \\
z^{(k)} &= x^{(k)} - \left[\frac{23}{8}I - \mathrm{G}^{(k)}\left(3I - \frac{9}{8}\mathrm{G}^{(k)}\right)\right]\ \mathrm{F}'\left(x^{(k)}\right)^{-1}\ \mathrm{F}\left(x^{(k)}\right), \\
x^{(k+1)} &= z^{(k)} - \frac{1}{2}\left(5I - 3\mathrm{G}^{(k)}\right)\ \mathrm{F}'\left(x^{(k)}\right)^{-1}\ \mathrm{F}\left(z^{(k)}\right),
\end{aligned}$$

where $\mathrm{G}^{(k)} = \ \mathrm{F}'\left(x^{(k)}\right)^{-1}\ \mathrm{F}'\left(y^{(k)}\right)$.

New concepts about efficiency

The authors have recently introduced in [12] some concepts and indicators as an alternative to classify iterative schemes according to their efficiency.

Definition 3. *Let Φ be an iterative scheme of order P, evaluative cost d and operational cost op. We call the Functional Performance of a method Φ, or E_Φ, the reduction in evaluative cost per unit of convergence of Φ relative to that of Newton's scheme. It is calculated by*

$$E_\Phi = 1 - \frac{d}{P}\frac{2}{n(n+1)}. \tag{8}$$

Similarly, we define Operating Performance, or T_Φ, as

$$T_\Phi = 1 - \frac{op}{P}\frac{6}{n^3 + 3n^2 - n}. \tag{9}$$

We can interpret functional performance as the evaluative savings over that required by N2S. For example, if $E_\Phi = 80\%$, we mean that the ϕ scheme needs 80% fewer functional evaluations than N2S to increase each order unit per iteration.

Definition 4. *Let Φ be an iterative scheme of order P. The Computational Performance of Φ, or C_Φ, is the average of their functional and operating performances,*

$$C_\Phi = \frac{1}{2}(E_\Phi + T_\Phi). \tag{10}$$

2 Design and analysis of new scalar accelerators and methods

The authors, in [7], have presented two families of maximally efficient iterative methods for solving systems of nonlinear equations. Both classes are fourth- and sixth-order convergent, respectively, being the first one optimal in the vectorial sense. Two members, one from each class, are:

$$x^{(k+1)} = \mathcal{N}^{(k)} - \lambda_k \, F'(x^{(k)})^{-1} F(\mathcal{N}^{(k)}), \quad k = 0, 1, \ldots, \tag{11}$$

wich is optimal two-steps fourth-order scheme, $\mathcal{N}^{(k)}$ is N2S, and

$$x^{(k+1)} = z^{(k)} - \lambda_k \, F'(x^{(k)})^{-1} F(z^{(k)}), \quad k = 0, 1, \ldots, \tag{12}$$

with order 6 and three-steps, $z^{(k)}$ is (11), and the accelerator

$$\lambda_k = 1 + 2\frac{F(x^{(k)})^T F(\mathcal{N}^{(k)})}{F(x^{(k)})^T F(x^{(k)})}. \tag{13}$$

Let us consider the combination of (11) and (12). We conjecture that from any method $\phi(x^{(k)})$ of order $p \geq 2$, with first Newton step, we can design a new method of order $p+2$ by simply adding a damped Traub-type step of the form

$$x^{(k+1)} = \mathcal{T}^{(k)} = w^{(k)} - \lambda_k \, F'(x^{(k)})^{-1} \, F(w^{(k)}), \tag{14}$$

where $w^{(k)} = \phi(x^{(k)})$.

Theorem 5. *Let $F : V \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a sufficiently differentiable function in an open convex neighborhood V of the solution ξ of the nonlinear system $F(x) = 0$, and ϕ represents the fixed-point function of order p with first step N2S. Let us also assume that $F'(x)$ is continuous and nonsingular at ξ. Then the sequence of iterates obtained by scheme (14) converges to ξ, with convergence order $p + 2$. The error equation is*

$$e^{(k+1)} = -\delta[p-2]C_2^3 e^{(k)^4} + \left(C_3 + 2P^{-1}\Omega C_2\right) L_p e^{(k)^{p+2}} + O\left(e^{(k)^{p+3}}\right),$$

where

$$e_k = x^{(k)} - \xi, \quad C_i = \frac{1}{i!}\left[F'(\xi)\right]^{-1} F^{(i)}(\xi), \quad i = 2, 3, \ldots,$$
$$P = \sum_{i=1}^{n} R_i^T R_i, \quad \Omega = \sum_{i=1}^{n} \left(R_i^T H_i\right),$$

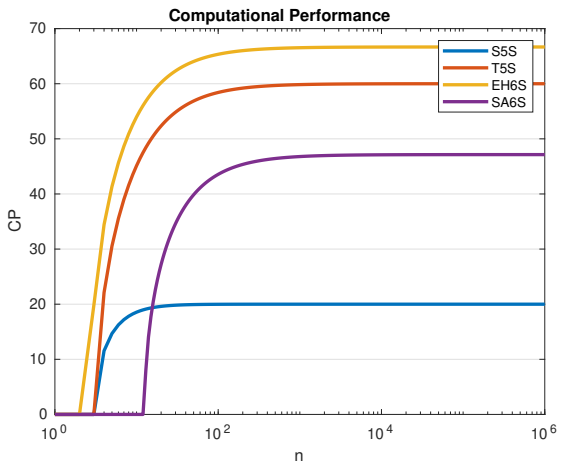
being $R_i = \nabla f_i(\xi) = \left(\frac{\partial f_i}{\partial x_1}(\xi), \frac{\partial f_i}{\partial x_2}(\xi), \ldots, \frac{\partial f_i}{\partial x_n}(\xi)\right)$ and $H_i = Hf_i(\xi)$ the Hessian matrix relative to coordinate function f_i of $F, i = 1, 2, \ldots, n$.

Now, it is immediate that we can generalize the result obtained in Theorem 5 by extending the accelerator scheme. If m is the number of steps added, we increase the order to $p + 2m$. We denote the scheme (13) as MART2M+.

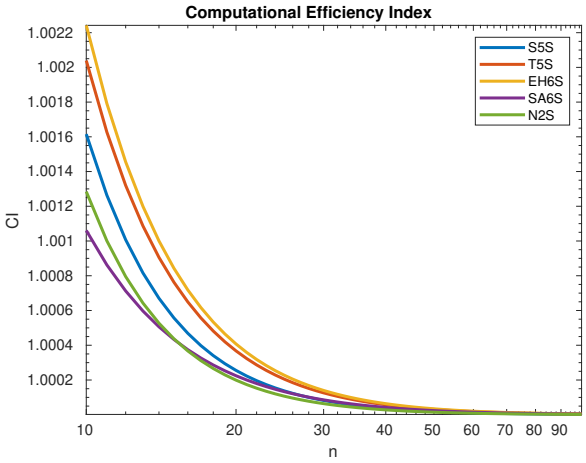
New Maximally Efficient Methods

The Traub and Ermakov's Hyperfamily schemes are maximally efficient with orders 3 and 4, respectively. They are among the best candidates to apply an accelerator, because they are the most efficient schemes in their orders.

We take these two methods as a basis to increase their order by appliying MART2M+. In this way we obtain two methods of order 5 and 6, which turn out to be new schemes that are also



(a) Computational Performance up to $n = 10^6$



(b) Computational Index from $n = 10$ to 100

Figure 1: Computational Performance and Efficiency Comparison

maximally efficient. The new methods use Newton's method as a first step, $\mathcal{N}^{(k)}$, and the scalar accelerators λ_k and ν_k are defined in (13) and (5), respectively.

The new scheme of order 5, adding a step to Traub's method, and which we denote T5S, is

$$\begin{aligned} z^{(k)} &= \mathcal{N}^{(k)} - F'\left(x^{(k)}\right)^{-1} F\left(\mathcal{N}^{(k)}\right), \\ x^{(k+1)} &= z^{(k)} - \lambda_k\, F'(x^{(k)})^{-1} F(z^{(k)}), \quad k = 0, 1, \ldots \end{aligned}$$

The new method of order 6, by adding MART2M+ to the EH4S, is denoted by EH6S, with iterative expression

$$\begin{aligned} z^{(k)} &= \mathcal{N}^{(k)} - F'\left(x^{(k)}\right)^{-1} \left(F\left(\mathcal{N}^{(k)}\right) + 2\nu_k F\left(x^{(k)}\right)\right), \\ x^{(k+1)} &= z^{(k)} - \lambda_k\, F'(x^{(k)})^{-1} F(z^{(k)}), \quad k = 0, 1, \ldots \end{aligned}$$

3 Efficiency Analysis

The computational efficiency of methods of the same order is higher in the one with the lowest cost.

Table 1: Computational Permance and Efficiency Index Comparison

j	***Method***	$\boldsymbol{p_j}$	$\boldsymbol{d_j}$	$\boldsymbol{op_j}$	***CP***	***CI***
1	*N2S*	*2*	n^2+n	$\frac{1}{3}\left(n^3-n\right)+n^2$	$1-\frac{1}{n\,p_1}\left(\frac{d_1}{n+1}+\frac{3\,op_1}{n^2+3n-1}\right)$	$2^{\frac{1}{d_1+op_1}}$
2	*SS5S*	*5*	$2n^2+2n$	$\frac{2}{3}\left(n^3-n\right)+2n^2+3n$	$1-\frac{1}{n\,p_2}\left(\frac{d_2}{n+1}+\frac{3\,op_2}{n^2+3n-1}\right)$	$5^{\frac{1}{d_2+op_2}}$
3	*T5S*	*5*	n^2+3n	$\frac{1}{3}\left(n^3-n\right)+3n^2+3n$	$1-\frac{1}{n\,p_3}\left(\frac{d_3}{n+1}+\frac{3\,op_3}{n^2+3n-1}\right)$	$5^{\frac{1}{d_3+op_3}}$
4	*EH6S*	*6*	n^2+3n	$\frac{1}{3}\left(n^3-n\right)+3n^2+5n$	$1-\frac{1}{n\,p_4}\left(\frac{d_4}{n+1}+\frac{3\,op_4}{n^2+3n-1}\right)$	$6^{\frac{1}{d_4+op_4}}$
5	*SA6S*	*6*	$2n^2+2n$	$\frac{1}{3}\left(n^3-n\right)+11n^2+4n$	$1-\frac{1}{n\,p_5}\left(\frac{d_5}{n+1}+\frac{3\,op_5}{n^2+3n-1}\right)$	$6^{\frac{1}{d_5+op_5}}$

Table 2: System size=2025, $x^{(0)} = (1/10, \ldots, 1/10)^T$

Method	*CPU Time*	*It*	$\left\|x^{(k+1)} - x^{(k)}\right\|$	$\left\|F(x^{(k+1)})\right\|$	$\breve{\rho}$
N2S	376.50	9	2.45×10^{-211}	6.38×10^{-422}	2
S5S	333.51	4	3.03×10^{-106}	1.09×10^{-529}	5
T5S	227.59	4	1.09×10^{-75}	4.59×10^{-374}	5
EH6S	227.79	4	4.00×10^{-140}	1.59×10^{-837}	6
SA6S	458.96	4	2.19×10^{-135}	5.30×10^{-809}	6

It can be seen in Figure 1, based on Table 1 where we show the Method, p_j (order), d_j (number of functional evaluations), op_j (number of products/quotients), CP (Computational Performance) and CI (Computational Efficiency Index), that the computational efficiency index and computational performance of the new schemes compete with others of high order or efficiency.

4 Numerical Performance

The numerical behaviour of the new maximally efficient schemes T5S and EH6S is compared with that of the schemes N2S, SS5S and SA6S, of orders 5 and 6, on the same problem.

All calculations were performed using Matlab R2022b with variable precision arithmetic with a 2000-digit mantissa to minimize rounding errors. The stopping criterion is $\left\|x^{(k+1)} - x^{(k)}\right\| + \left\|F\left(x^{(k+1)}\right)\right\| < 10^{-300}$ or 50 as maximum number of iterations.

In this manuscript, we use $\breve{\rho}$ as an estimator of the order of convergence p, as defined Cordero et al. in [14] :

$$\breve{\rho} = \frac{\ln\left(\left\|x^{(k+1)} - x^{(k)}\right\| / \left\|x^{(k)} - x^{(k-1)}\right\|\right)}{\ln\left(\left\|x^{(k)} - x^{(k-1)}\right\| / \left\|x^{(k-1)} - x^{(k-2)}\right\|\right)}. \tag{15}$$

Problem

Now, we consider a large system of 2025 equations:

$$\begin{aligned} & x_i + \log\left(2 + x_i + x_{i+1}\right) = 0, \quad i = 1, 2, \ldots, 2024, \\ \text{and} \quad & x_{2025} + \log\left(2 + x_{2025} + x_1\right) = 0. \end{aligned}$$

We select the initial estimate $x^{(0)} = \left(\frac{1}{10}, \cdots, \frac{1}{10}\right)^T$.

In Table 2, we show the CPU time in seconds, the amount of iterations It, the approximate and residual error, $\left\|x^{(k+1)} - x^{(k)}\right\|$ and $\left\|F\left(x^{(k+1)}\right)\right\|$, and the approximate order of convergence, $\breve{\rho}$.

We place with red letter the scheme with the best CPU time, with blue letter the second place, and with green letter the third place.

It is observed that our maximally efficient fifth and sixth order T5S and EH6S schemes, respectively, overwhelmingly outperform all the other schemes in CPU time. In third place is SS5. This indicates that our proposals are very competitive, even against high-performance methods such as the comparison ones, and that they make more use of the computational effort.

5 Conclusions

We have presented two new maximally efficient iterative schemes. One of them is an improvement of the classical Traub Method, while the other is an improvement of one of the optimal schemes of the Ermakov Hyperfamily for nonlinear problems. We have achieved this from a new contribution: an accelerator with scalar weight that allows to increase by two units the convergence order of Newton-type methods. This is very efficient with respect to other accelerators, as we show in a brief comparison.

We offered new concepts on efficiency, as a contribution to classify iterative vectorial schemes according to functional, operational or computational performance. We have also shown that the order of the new schemes is 5 and 6, respectively. We have shown that they involve a significant reduction in time even with respect to Newton's Method in numerical tests.

Acknowledgement: This research was financed by Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCYT) of Ministerio de Educación Superior Ciencia y Tecnología de la República Dominicana (MESCYT) grant number FONDOCYT 2023-1-1D2-0537.

References

[1] Traub, J.F., Iterative methods for the solution of equations. New York, USA, Chelsea Publishing Company, 1982.

[2] Ostrowski, A.M., Solutions of equations and systems of equations. New York, USA, Academic Press Inc., 1966.

[3] Cordero, A., Hueso, J.L., Martínez, E., Torregrosa, J.R., A modified Newton-Jarratt's composition. Numerical Algorithms, 55: 87-99, 2010.

[4] Kung, H.T., Traub, J.F., Optimal order of one-point and multipoint iteration. Journal of the ACM (JACM), 21(4): 643-651, 1974.

[5] Arroyo, V., Cordero, A., Torregrosa, J.R., Approximation of artificial satellites' preliminary orbits: the efficiency challenge. Mathematical and Computer Modelling, 54(7-8): 1802-1807, 2011.

[6] Carson, E., Strakoš, Z., On the cost of iterative computations. Philosophical Transactions of the Royal Society A, 378(2166): 20190050, 2020.

[7] Cordero, A., Rojas-Hiciano, R.V., Torregrosa, J.R., Vassileva, M.P., Maximally efficient damped composed Newton-type methods to solve nonlinear systems of equations. Applied Mathematics and Computation, 492: 129231, 2025.

[8] Singh, H., Sharma, J.R., Kumar, S., A simple yet efficient two-step fifth-order weighted-Newton method for nonlinear models. Numerical Algorithms, 93(1): 203-225, 2023.

[9] Cordero, A., Rojas-Hiciano, R.V., Torregrosa, J.R., Vassileva, M.P., Fractal Complexity of a New Biparametric Family of Fourth Optimal Order Based on the Ermakov–Kalitkin Scheme. Fractal and Fractional, 7(6): 459, 2023.

[10] Cordero, A., Rojas-Hiciano, R. V., Torregrosa, J.R., Vassileva, M.P., A highly efficient class of optimal fourth-order methods for solving nonlinear systems. Numerical Algorithms, 95(4): 1879-1904, 2024.

[11] Cordero, A., Rojas-Hiciano, R.V., Torregrosa, J.R., Triguero-Navarro, P., Efficient parametric family of fourth-order Jacobian-free iterative vectorial schemes. Numerical Algorithms, 97(4): 2011-2029, 2024.

[12] Cordero, A., Rojas-Hiciano, R.V., Torregrosa, J.R., Vassileva, M.P., High-Level Convergence Order Accelerators of Iterative Methods for Nonlinear Problems. Applied Numerical Mathematics, submited, 2024.

[13] Sharma, J.R., Arora, H., Efficient Jarratt-like methods for solving systems of nonlinear equations. Calcolo 51(1): 193-210, 2014.

[14] Cordero, A., Torregrosa, J.R., Variants of Newton's method using fifth-order quadrature formulas. Applied Mathematics and Computation, 190(1): 686-698, 2007.

QUBO formulation for Multi-Marginal Optimal Transport

S. Torregrosa[♭], [1] E. Gravot[♮] and F. Chinesta[♭⋆]

(♭) ENSAM Institute of Technology, PIMM, UMR CNRS 8006, ESI/Keysight Chair,
151 Boulevard de l'Hopital, 75013 Paris, France.

(♮) ENSAM Institute of Technology, PIMM, UMR CNRS 8006, RTE Chair,
151 Boulevard de l'Hopital, 75013 Paris, France.

(⋆) CNRS@CREATE Ltd.,,
1 CREATE Way, 08-01, CREATE Tower, 138602 Singapore.

Real-world challenges in areas like logistics and planning often give rise to combinatorial optimization problems, which are notoriously difficult to solve due to their computational complexity. These problems are conventionally modeled as integer programs and tackled using solvers such as CPLEX and Gurobi, which implement branch-and-bound strategies. However, these methods often involve exponential time complexity, making them unsuitable for large-scale problems. As a result, heuristic techniques have become popular for their speed and scalability, despite offering no guarantee of finding optimal solutions.

In recent years, quantum computing has emerged as a promising framework to address such intractable problems. Quantum and quantum-inspired algorithms have attracted considerable research interest, largely because of their potential to leverage specialized hardware for faster problem-solving than is possible with classical machines [1]. Among these, quantum annealing is particularly notable. It harnesses quantum properties such as superposition and tunneling to efficiently avoid local minima. Though the technology is still maturing, devices like D-Wave's quantum annealer have shown promising results on specific problem sets [2]. In addition, quantum-inspired technologies like Fujitsu's Digital Annealer [1] have been developed, replicating many of these capabilities on classical hardware without requiring actual quantum effects.

Quantum annealing approaches combinatorial optimization by reformulating problems as Ising models—or equivalently, as Quadratic Unconstrained Binary Optimization (QUBO) problems [3]—which can then be processed by an annealing device. The QUBO formulation has become a widely adopted standard for expressing diverse combinatorial problems [4], characterized by its unconstrained, quadratic, and binary structure. Although many integer programming problems inherently involve constraints, these can be incorporated into the QUBO formulation through the use of penalty terms [5], as demonstrated in prior work [3]. However, because QUBO operates on binary variables, natural encoding for permutation problems cannot be directly applied.

Permutation problems form a significant subset of combinatorial optimization, with classical examples including the Traveling Salesman Problem and the Quadratic Assignment Problem. Due to their NP-hard nature, extensive research has been conducted on heuristic or metaheuristic

[1] sergio.torregrosa_jordan@ensam.eu

techniques to develop practical solution methods. In this study, we turn our attention to another well-known NP-hard permutation problem: the Multidimensional Assignment Problem (MAP).

In earlier research, the authors developed a parametric metamodel based on Optimal Transport (OT) theory [6], demonstrating enhanced real-time inference of physical fields compared to standard regression approaches. This method is particularly effective in scenarios with limited data and high-dimensional input spaces, and it does not rely on specially structured training datasets. The resulting computational parametric methodology is effective for physical fields characterized by localized regions—a common characteristic in domains such as structural mechanics, fluid flow, and electromagnetism. Traditional regression techniques often struggle with such problems; for instance, Euclidean interpolation of parametric functions with localized features often produces unsatisfactory results. OT, in contrast, offers a more natural approach by progressively translating and scaling features. While this leads to more accurate results, the computational cost of OT remains significant. Even with recent improvements, the use of OT in real-time settings is still limited by its inherent complexity. To address this real-time limitation, the authors developed a regression framework for parametric physical problems [7, 8], along with OT-inspired methods for parameterized spatial domains [9].

During those works, the authors addressed the Multi-Marginal Optimal Transport (MMOT) Monge formulation problem [10, 11], which was rewritten as a MAP and solved using a meta-heuristic approach: a Genetic Algorithm [12]. The current study introduces a new formulation by expressing this NP-hard permutation problem in the QUBO framework. Although the earlier method established a viable solution path, the transition to a QUBO-based model offers improved scalability and optimization efficiency. This reformulation is motivated by the structural compatibility between the original problem and QUBO's encoding paradigm.

References

[1] Aramon, M., Rosenberg, G., Valiante, E., Miyazawa, T., Tamura, H., Katzgrabeer, H., Physics-inspired optimization for quadratic unconstrained problems using a digital annealer. *Frontiers in Physics*, 7(48), 2019.

[2] Venegas-Andraca, S., Cruz-Santos, W., McGeoch, C., Lanzagorta, M., A cross-disciplinary introduction to quantum annealing-based algorithms. *Contemporary Physics*, 59(2):174–197, 2018.

[3] Lucas, A., Ising formulations of many NP problems. *Frontiers in Physics*, 2, 2014.

[4] Verma, A., Lewis, M., Penalty and partitioning techniques to improve performance of qubo solvers. *Discrete Optimization*, 44(2), 2022.

[5] Smith, A., Coit, D., Baeck, T., Fogel, D., Michalewicz, Z., Penalty functions. *Handbook of evolutionary computation*, 97(1), 1997.

[6] Villani, C., Optimal transport, old and new. Verlag Berlin Heidelberg, Springer, 2006.

[7] Torregrosa, S., Champaney, V., Ammar, A., Herbert, V., Chinesta, F., Surrogate Parametric Metamodel based on Optimal Transport. *Mathematics and Computers in Simulation*, 194:36–63, 2021.

[8] Torregrosa, S., Munoz, D., Herbert, V., Chinesta, F., Parametric Metamodelling Based on Optimal Transport applied to Uncertainty Evaluation. *Technologies*, 12(2), 2024.

[9] Munoz, D., Torregrosa, S., Allix, O., Chinesta, F., Empowering PGD-based parametric analysis with Optimal Transport. *Finite Elements in Analysis and Design*, 228(104049), 2024.

[10] Monge, G., Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 666–704, 1781.

[11] Pass, B., Multi-marginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis, Optimal Transport*, 49(6):1771–1790, 2015.

[12] Sastry, K., Goldberg, D., Kendall, G., Genetic Algorithms. *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, 97–125, 2005.

A first highly efficient eight-order optimal iterative method for solving non-linear systems

A. Cordero [♭], J. R. Torregrosa[♭] and P. Triguero-Navarro [♭1]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Abstract

This work introduces a novel eighth-order optimal iterative scheme for nonlinear systems, pushing the boundaries of computational efficiency. This new scheme significantly outperforms existing methods, even those with ninth-order convergence, across both small and large-scale systems. Furthermore, it demonstrates excellent stability, with guaranteed convergence to roots for quadratic polynomials. Extensive numerical testing, including challenging large-scale problems, highlights the method's remarkable performance and low computational demands. Notably, unlike typical higher-order methods that often sacrifice speed, our optimal scheme achieves computational speeds comparable to Newton's method.

2 Introduction

Solving systems of nonlinear equations is a critical challenge in many areas of science and engineering. Since exact solutions are rarely attainable, iterative methods provide essential approximations.

Newton's method, a fundamental tool in numerical analysis, forms a basis for these iterative approaches (see [1]), and is defined for vectorial problems as follows:

$$x^{(k+1)} = x^{(k)} - \left[F'(x^{(k)})\right]^{-1} F(x^{(k)}), \quad k = 0, 1, 2, \ldots, \tag{1}$$

being $F : W \subseteq \mathbb{R}^n \to \mathbb{R}^n$ a non-linear function describing the system of equations $F(x) = 0$, and $F'(x)$ its associated Jacobian matrix. While Newton's method is valued for its efficiency, there's an ongoing quest to develop methods with even higher convergence rates without incurring excessive computational costs. A key conjecture sets a theoretical limit on convergence order, defining "optimal" methods as those that achieve this limit (see [2]). This conjecture is defined as follows:

Conjecture 1 *Any Newton-type iterative method for solving non-linear systems, without memory, has an order of convergence that cannot exceed $2^{k_1+k_2-2}$, where k_1 is the number of functional evaluations in the entries of the Jacobian matrix per iteration and k_2 is the number of functional evaluations of the non-linear function, which where $k_1 \leq k_2$. If the method reaches this bound, it is called optimal in the vectorial sense.*

[1] ptrinav@mat.upv.es

To the best of our knowledge, only one parametric class of optimal schemes for vectorial models, in accordance with the aforementioned conjecture, has been documented in the literature, see [3]. In this work, the authors introduced a family of optimal vectorial fourth-order methods depending on two parameters, characterized by

$$\begin{aligned} y^{(k)} &= x^{(k)} - \left[F'(x^{(k)})\right]^{-1} F(x^{(k)}), \\ x^{(k+1)} &= y^{(k)} - \left[F'(x^{(k)})\right]^{-1} \left(p_k F(y^{(k)}) + q_k F(x^{(k)})\right), \quad k = 0, 1, 2, \ldots \end{aligned} \tag{2}$$

where

$$\nu_k = \frac{\|F(y^{(k)})\|_2}{\|F(x^{(k)})\|_2}, \qquad K_k = \frac{1}{1+\lambda\nu_k}, \qquad p_k = K_k(1+\psi\nu_k), \qquad q_k = 2K_k\nu_k.$$

The use of the variable ν_k was introduced in [4]. Family (2) exhibits fourth-order convergence.

Our primary objective is to develop an optimal method with an eighth-order convergence. Additionally, we aim to perform these modifications without significantly adding computational cost, to obtain a competitive and efficient method comparable to other existing schemes with similar structure. To achieve this objective, we select the member of family (2) corresponding to $\lambda = 0$ and $\psi = 1$, an additional iterative step is implemented, and new accelerating scalar functions, similar to ν_k, are incorporated. The resulting three-steps method, which we denote as CTT (3), is subsequently defined as follows:

$$\begin{aligned} z^{(k)} &= y^{(k)} - \left[F'(x^{(k)})\right]^{-1} \left((1+\nu_k)F(y^{(k)}) + 2\nu_k F(x^{(k)})\right), \\ x^{(k+1)} &= z^{(k)} - \left[F'(x^{(k)})\right]^{-1} \left(F(z^{(k)}) + (2\alpha_k(1-\nu_k) + 4\delta_k)\, F(x^{(k)}) + (2\alpha_k + \beta_k)F(y^{(k)})\right), \end{aligned} \tag{3}$$

where $y^{(k)}$ is a Newton's step and the accelerators are

$$\begin{cases} \nu_k = \dfrac{F(y^{(k)})^T F(y^{(k)})}{F(x^{(k)})^T F(x^{(k)})}, & \alpha_k = \dfrac{F(y^{(k)})^T F(z^{(k)})}{F(x^{(k)})^T F(x^{(k)})}, \\ \beta_k = \dfrac{F(z^{(k)})^T F(z^{(k)})}{F(y^{(k)})^T F(y^{(k)})}, & \delta_k = \dfrac{F(z^{(k)})^T F(z^{(k)})}{F(x^{(k)})^T F(x^{(k)})}, \end{cases}$$

3 Convergence Analysis

This section rigorously establishes the local convergence order of the proposed CTT method, defined by equation (3). The following theorem formally states this result:

Theorem (Local Convergence Order). Let $F : W \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a sufficiently differentiable nonlinear function defined on an open set W in $\mathbb{R}^n$. Suppose that F has a root $\xi \in W$, such that $F(\xi) = 0$, and that the Jacobian matrix of F at the root, $F'(\xi)$, is non-singular. Then, for an initial approximation $x^{(0)}$ sufficiently close to ξ, the sequence $\{x^{(k)}\}$ generated by the CTT method converges to ξ with a local order of convergence of 8.

4 Complex Dynamics

To gain deeper insights into the stability and behavior of the iterative CTT method, we analyze it as a complex dynamical system. This approach involves studying the iterative process when applied to complex functions, allowing us to understand how the method behaves based on different

initial values (seeds) in the complex plane. For the specific case of a second-degree polynomial $p(x) = (x - a)(x - b)$, where a and b are complex numbers, the Scaling Theorem proves to be a valuable tool. It enables us to relate the dynamical behavior of the method applied to one second-degree polynomial to that of any other through a suitable transformation.

Theorem (Scaling Theorem). Let $f(x)$ be an analytical function defined on the Riemann sphere, and let $T(x) = \alpha x + \beta$ be an affine transformation with $\alpha \neq 0$. Consider a function $g(x) = \lambda(f \circ T)(x)$, where $\lambda \neq 0$. Let R_f and R_g be the rational operators (fixed point functions) associated with the CTT scheme (3) when applied to f and g, respectively. Then, R_f and R_g are conjugate via the transformation T, meaning that $(T \circ R_g \circ T^{-1})(x) = R_f(x)$.

By applying the Möbius transformation $h(x) = \frac{x-a}{x-b}$ to the rational operator $R_p(x)$ of the CTT method when applied to the polynomial $p(x)$, we obtain a conjugate rational operator $O(x) = (h \circ R_p \circ h^{-1})(x)$. This conjugated operator simplifies the analysis as it becomes independent of the specific roots a and b. The resulting operator for the CTT method takes the form of a rational function:

$$O(x) = \frac{x^8 n(x)}{d(x)},$$

where $n(x)$ and $d(x)$ are polynomials of degree 22. The degree of these polynomials arises from the complex interactions within the three-step iterative process.

Theorem (Fixed Points and Stability). The fixed points of the rational operator $O(x)$, which are the values of x satisfying $O(x) = x$, are found to be $x = 0$, $x = \infty$, $x = 1$, and the 28 roots of a polynomial $k(t)$ of degree 28. Specifically:

- $x = 0$ and $x = \infty$ are superattracting fixed points. These correspond to the roots of the original polynomial $p(x)$ after the Möbius transformation, and their superattracting nature indicates strong local convergence.

- $x = 1$ is a repelling strange fixed point. Strange fixed points are those that do not correspond to the roots of the polynomial and can hinder convergence if initial seeds are close to them. The repelling nature implies that iterations will generally move away from this point.

- The 28 roots of the polynomial $k(t)$ are also repelling strange fixed points. Their repelling nature is beneficial for the stability of the method, as it reduces the likelihood of convergence to these extraneous points.

The dynamical plane, which visually represents the basins of attraction of the roots and other fixed points in the complex plane, for the CTT method shows no black regions. Black regions typically indicate areas where the iterative method fails to converge to any of the roots or known attractors within a specified number of iterations, suggesting chaotic or unpredictable behavior. The absence of black regions for the CTT method implies a generally well-behaved convergence across a wide range of initial seeds. Furthermore, the fact that all fixed points other than the roots ($x = 0$ and $x = \infty$ in the transformed space) are repulsive and lie on the Julia set (the boundary between basins of attraction) suggests a high degree of stability and predictability for the method.

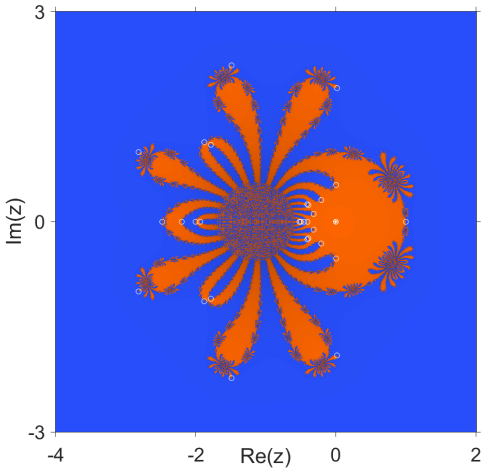


Figure 1: Dynamical plane

5 Efficiency

For a rigorous comparative evaluation of various iterative methods from the perspective of computational cost, the Ostrowski efficiency index is commonly employed. This index is mathematically defined as:

$$I = p^{1/d},$$

where p denotes the order of convergence of the method, and d represents the number of function evaluations required per iteration. A higher Ostrowski efficiency index generally indicates a more efficient method.

Furthermore, the computational efficiency index provides a complementary measure that takes into account not only the function evaluations but also the arithmetic operations involved in each iteration. It is defined as:

$$IC = p^{1/(d+op)},$$

where op signifies the total number of arithmetic operations (such as additions, subtractions, multiplications, and divisions) performed per iteration. This index offers a more comprehensive view of the computational cost.

For the proposed CTT scheme, based on its eighth-order convergence ($p = 8$) and the analysis of its computational requirements, the Ostrowski efficiency index is given by:

$$I_{CTT} = 8^{1/(n^2+3n)},$$

where n is the dimension of the nonlinear system. The term n^2 arises from the Jacobian matrix evaluation (involving n^2 partial derivative evaluations), and $3n$ comes from the three function evaluations ($F(x^{(k)})$, $F(y^{(k)})$, and $F(z^{(k)})$), each requiring n scalar function evaluations.

The computational efficiency index for the CTT method is:

$$IC_{CTT} = 8^{1/(\frac{1}{3}(n^3-n)+4n^2+11n)}.$$

Comparing these efficiency indices with those of other ninth-order and eighth-order methods available in the literature, our analysis indicates that the CTT method exhibits superior or highly competitive performance for solving nonlinear systems, across both small-scale (low n) and large-scale (high n) problems. This suggests that CTT offers a favorable balance between the achieved

convergence order and the computational cost per iteration, making it an efficient choice for a wide range of nonlinear systems.

6 Numerical results

The numerical performance of the proposed CTT method was evaluated against other eighth-order and ninth-order methods using high-precision arithmetic in Matlab. The iterative process was terminated when either the norm of the difference between two successive approximations, $\left\|x^{(k+1)} - x^{(k)}\right\|$, fell below 10^{-100}, or the norm of the function evaluated at the current approximation, $\left\|F(x^{(k+1)})\right\|$, was less than 10^{-100}. A maximum of 50 iterations was allowed for each method. The comparison considered several key metrics, including the norm of the function at the final iteration, the norm of the difference between the last two iterations, the total number of iterations required to meet the stopping criteria, the computational time (CPU time), and the approximate computational order of convergence (ACOC).

Example 1. We first examined a system of $m = 20$ nonlinear equations defined as:

$$x_i - 1.5 \sin\left(\sum_{j=1, j\neq i}^{m} x_j\right) = 0, \quad i = 1, 2, \ldots, m.$$

The approximate solution for this system is $\xi \approx (0.15973, 0.15973, \ldots, 0.15973)^T$. The iterative methods were initiated with the starting guess $x^{(0)} = \left(\frac{1}{10}, \frac{1}{10}, \ldots, \frac{1}{10}\right)^T$. The results indicated that the CTT method achieved convergence in a similar number of iterations and with comparable error norms to the ninth-order methods considered, but with a notably lower computational time.

Example 2. Next, we considered a system of $m = 100$ nonlinear equations given by:

$$x_i - \cos\left(-\left(\sum_{j=1}^{m} x_j\right) + 2x_i\right) = 0, \quad i = 1, 2, \ldots, m.$$

The target solution for this system is $\xi \approx (0.20625, 0.20625, \ldots, 0.20625)^T$, and the initial approximation used for all methods was $x^{(0)} = \left(\frac{1}{10}, \ldots, \frac{1}{10}\right)^T$. The results from this example further demonstrated the efficiency of the CTT method, showcasing the lowest execution time and the smallest residual error among the compared methods.

Example 3. Finally, we investigated a large-scale nonlinear system with $m = 500$ equations, defined as:

$$\begin{aligned} x_i^2 x_{i+1} - 1 &= 0, \quad i = 1, 2, \ldots, m-1, \\ x_m^2 x_1 - 1 &= 0. \end{aligned}$$

The solution for this system is $\xi \approx (1, 1, \ldots, 1)^T$, and the initial guess used was $x^{(0)} = \left(\frac{3}{2}, \cdots, \frac{3}{2}\right)^T$. The outcomes of this experiment reinforced the previous findings, with the CTT scheme consistently exhibiting the lowest computational time required for convergence.

In conclusion, the results obtained from these numerical experiments are in strong agreement with the theoretical efficiency analysis conducted earlier. These findings indicate that the proposed eighth-order optimal vectorial scheme, CTT, is indeed highly competitive and efficient for solving systems of nonlinear equations. It often outperforms other higher-order methods, particularly in terms of computational time, while maintaining comparable or superior accuracy and convergence behavior.

References

[1] Traub, J. Iterative methods for the solution of equations. Prentice-Hall, 1964.

[2] Arroyo, V., Cordero, A. & Torregrosa, J. Approximation of artificial satellites' preliminary orbits: The efficiency challenge. *Mathematical And Computer Modelling*, 54: 1802-1807, 2011.

[3] Cordero, A., Rojas-Hiciano, R., Torregrosa, J. & Penkova, M. A highly efficient class of optimal fourth-order methods for solving nonlinear systems. *Numerical Algorithms*, 95: 1879–1904, 2024.

[4] Singh, H., Sharma, J. & Kumar, S. A Simple yet Efficient Two-Step Fifth-Order Weighted-Newton Method for Nonlinear Models. *Numerical Algorithms*, 93: 203-225, 2022.

A Hybrid Steffensen-Genetic Algorithm for Finding Simultaneous Roots of Nonlinear Equations

Fiza Zafar [♭],[1] Alicia Cordero[♮], Sadia Mujtaba [♭], and Juan R. Torregrosa[♮]
(♭) Centre for Advanced Studies in Pure and Applied Mathematics, Bahauddin Zakariya University, Multan, Pakistan
(♮) I.U. de Matematica Multidisciplinar, Universitat Politècnica de València
Cami de Vera s/n, Valencia, Spain.

1 Introduction

Numerical analysis plays a critical role across various fields, including engineering, applied mathematics, business, physical sciences, and medicine. It encompasses a range of iterative techniques that are essential for solving complex mathematical problems, often when analytical methods are insufficient or infeasible. While some problems can be solved exactly using a finite sequence of analytical steps, many are too intricate for such direct solutions. In these cases, numerical analysis offers the advantage of iterative methods, which enable approximate solutions through systematic approximation.

Many real-world problems can be modeled through non-linear equations, a common form being $f(x) = 0$. Newton's method is a well known approach for its efficiency in finding roots, specially when functions have continuous derivatives. Newton's method iteratively refines an initial guess x_n to approximate the solution, using the formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{1}$$

However, it requires differentiable functions, which can limit its use. To address this, Steffensen,'s method offers an alternative that approximates through forward difference:

$$x_{n+1} = x_n - \frac{f^2(x_n)}{f(x_n + f(x_n)) - f(x_n)}. \tag{2}$$

These iterative methods, including newer higher-order techniques, provides a solution path that does not rely on derivative calculations. In recent years, numerous higher-order iterative methods have emerged, refining these classic approaches to improve convergence speed and computational efficiency.

A Genetic Algorithm (GA) [3] is a metaheuristic search method inspired by Darwin's theory of natural evolution, simulating principles of natural selection and genetic mechanisms like mutation, crossover, and selection. Genetic Algorithm's are known for their powerful global search capabilities, though they can face challenges with convergence efficiency in later stages. They're widely used for finding optimal or near-optimal solutions to complex problems that would otherwise require extensive computation.

[1] fizazafar@bzu.edu.pk

The search for simultaneous multiple roots of non-linear equations [4] represents a significant challenge in both mathematics and engineering, with implications that span various fields such as physics, economics, and engineering design. Quadratically convergent conventional techniques, such as the Newton-Raphson method and Steffesen's method [5], frequently face convergence challenges, especially for complex problems with multi, multiple roots and discontinuities. These constraints have motivated researchers to investigate more resilient computational methodologies, resulting in the creation of hybrid algorithms that integrate the advantages of genetic algorithms (GA) with alternative optimization techniques.

The implementation of evolutionary algorithms for the solution of nonlinear equation also presents several challenges, including premature convergence, parameter tuning, and the equilibrium between exploration and exploitation [2]. Although these algorithms can yield robust solutions, they frequently demand substantial computational resources, posing limitations for real-time applications. It is essential to meticulously evaluate the trade-offs between accuracy, computational efficiency, and robustness when choosing a suitable method for a specific problem.

Several innovative hybrid approaches to root-finding have been developed in the literature. One notable method involves merging Genetic algorithm's with local search techniques, such as gradient descent or Newton's method [7], which has proven considerable gains in convergence speed. Furthermore, some research has conceptualized the root-finding problem as a multi-objective optimization problem [2], facilitating the simultaneous identification of multiple roots through the minimization of the residuals of the nonlinear equations . This comprehensive strategy can be effectively managed by hybrid genetic algorithms, which are adaptable to handling complex optimization landscapes. The efficacy of a hybrid genetic algorithm [6] lies in the integration of a local optimizer's rapidness such as higher order Steffensen's method with the global search capability of a genetic algorithm that can be applied to non-differentiable, discontinuous, and even combinatorial functions.

This paper presents a novel hybrid Steffensen-Genetic Algorithm (SGA) merging the GA evolutionary strategy to explore the solution space effectively and a fourth-order Steffensen's method given by Cordero et al. [1] to refine solutions to high accuracy. To preserve population diversity and convergence efficiency, the method makes use of a standard GA that has been improved by adaptive crossover and mutation operators. To ensure accurate and rapid convergence to roots, a local refinement phase is also included, during which the Steffensen approach is applied to the top-performing individuals from the GA. By combining these methods, the algorithm functions as a robust multi-root solver capable of efficiently finding and refining multiple distinct roots of nonlinear equations.

2 Hybrid Steffensen-Genetic Algorithm

Algorithm 1 A New Hybrid Steffensen-Genetic Algorithm (SGA)

Input: Function $f(x)$, interval $[a, b]$, population size pop_size, maximum generations max_gens, tolerance tol, distinct tolerance $distinct_tol$, crossover probability p_c, mutation probability p_m
Initialize: Random population within $[a, b]$
for $generation = 1$ **to** max_gens **do**
    Calculate fitness for each individual: $fitness(x) = |f(x)|$
    Select individuals using tournament selection
    Define Steffensen type Method Iter(x):
    repeat
        Update x using Steffensen type method
    until convergence or instability **return** x
    Perform tournament selection and crossover
    Perform mutation and update fitness
    Screen top individuals and refine roots using Iter(x)
    Replace population with refined roots
end for
return best solution, best fitness, and all qualified roots

3 Numerical Experiments

We consider the following test problems to compare the performance of Steffensen, Genetic and hybrid Steffensen-Genetic Algorithm (SGA) method. We provide several examples illustrating the efficacy of our method.To systematically assess the efficacy of our Hybrid Steffesen-GA approach (SGA), we define essential parameters that affect the evolutionary process. The parameters encompass the function's domain, crossover and mutation probabilities, the ratio of parents to offspring, and the overall population size. The appropriate selection of these factors guarantees effective convergence and stability in the optimization process.

Table 1: Test Functions

Nonlinear Functions	Domain	Solutions
$f_1(x) = (x^2 - e^x - 3x + 2)\ln(x + 150)$	$(-1, 1)$	0.2575302854380565
$f_2(x) = \ln x + x^{1/2} - 5$	$(2, 10)$	8.309432664200271
$f_3(x) = (xe^{2x^2} - \sin^2 x + 50\cos x + 5)^4$	$(-2, 2)$	-1.20556848926394
$f_4(x) = (x^2 - e^{60x} - 3x + 2)^5$	$(-1, 1)$	0.01188653001928682
$f_5(x) = (2x\lvert\cos x\rvert + x^2 - 1)^2 - 2$	$(-3, 2)$	-2.704784856524704
		-1.821166332738140
		-1.562296722714125
		-0.840953096992933
		-0.244153167113250
		1.350786955284248
$f_6(x) = \frac{x+20}{x-2}\left\lvert\frac{x\lvert x-1\rvert(x-3)\lvert x-4.1234\rvert}{5(x-2)\lvert x+1\rvert}\right\rvert$	$(-25, 25)$	$-20, 0, 1, 3, 4.1234$

Table 2: Parameter Selections

Parameter	Value
Selection	Tournament Selection
Crossover	Linear Crossover
Mutation	Random
Domain	Problem Based
Parents to Offspring Ratio	1:1
Initial Population Size	50
Tolerance	0.1

Example 1: The problem is to find all distinct roots of the nonlinear function given below within domain $(-1, 1)$. The hybrid SGA method is used to efficiently identify and refine all roots.

$$f_1(x) = (x^2 - e^x - 3x + 2)\ln(x + 150) \tag{3}$$

All qualified roots using SGA are $0.25753028534160627, -0.7086371431448026$. All qualified roots using GA are $0.04464052964066578,\ 0.1936547305150758, -0.19424790884764365,\ 0.5014384151770221,\ 0.6860543471574483,\ 0.31853779992783493$.

In Table 3, we now compare our results for $f_1(x)$.

Table 3: Performance of Root-Finding Methods for $f_1(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	4	0.25753	$9.68540e-10$	15
Genetic Algorithm	10	0.29593	0.000626389	42
Hybrid Genetic Algorithm	10	0.25753	$2.18764e-11$	198

Example 2: The problem is to find all distinct roots of the nonlinear function given below within domain $(2, 10)$. A hybrid approach using a genetic algorithm for global search and Steffensen's method for local refinement is used to efficiently identify and refine all roots.

$$f_2(x) = \ln x + x^{1/2} - 5. \tag{4}$$

All qualified roots using SGA are 8.309432593909841. All qualified roots using GA are $3.2603218594806886,\ 7.86489782067479,\ 8.307394637570608, 9.695056671195973$.

In Table 4, we now compare our results for $f_2(x)$.

Table 4: Performance of Root-Finding Methods for $f_2(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	5	8.30941	$5.43433e-06$	4.3
Genetic Algorithm	10	8.20849	$3.23381e-05$	53
Hybrid Genetic Algorithm	10	8.30943	$8.82317e-09$	90

Example 3: The problem is to find all distinct roots of the nonlinear function given below within domain $(-3, 2)$. A hybrid approach using a genetic algorithm for global search and Steffensen's method for local refinement is used to efficiently identify and refine all roots.

$$f_3(x) = (xe^{2x^2} - \sin^2 x + 50\cos x + 5)^4. \tag{5}$$

All qualified roots using SGA are $-0.7346697924239922, -1.2055684892763794, 0.2593151484520361,\ 1.0919490033757175, -0.1859530120249725$. All qualified roots using GA are $-0.3477254519233073,\ 0.19624360533714436,\ -1.3854507429153862,\ 0.5302031660201361,\ -1.2095911373523562$.

In Table 5, we now compare our results for $f_3(x)$.

Table 5: Performance of Root-Finding Methods for $f_3(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	25	-1.11111	3.08728016	47
Genetic Algorithm	10	-1.20959	$1.08775e-09$	43
Hybrid Genetic Algorithm	10	-1.20556	$2.96632e-05$	70

Example 4: The problem is to find all distinct roots of the nonlinear function given below within domain $(-1, 1)$. A hybrid approach using a genetic algorithm for global search and Steffensen's method for local refinement is used to efficiently identify and refine all roots.

$$f_4(x) = (x^2 - e^{60x} - 3x + 2)^5. \tag{6}$$

All qualified roots using SGA are -0.50569231143650, -0.249207775598216, -0.728754465568196, 0.0118865300192786. All qualified roots using GA are -0.5972237062082192, -0.4558348663214321, -0.18505641416905194, 0.11972423777262131, -0.0127540974851747.

In Table 6, we now compare our results for $f_4(x)$.

Table 6: Performance of Multi-Root-Finding Methods for $f_4(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	25	0.02040	6.7077121	288
Genetic Algorithm	10	-0.41725	$5.44048e-05$	46
Hybrid Genetic Algorithm	10	0.01188653	$8.82317e-09$	90

Example 5: The problem is to find all distinct roots of the nonlinear function given below within domain $(-3, 2)$. A hybrid approach using a genetic algorithm for global search and Steffensen's method for local refinement is used to efficiently identify and refine all roots.

$$f_5(x) = (2x|\cos x| + x^2 - 1)^2 - 2 \tag{7}$$

All qualified roots using SGA are -1.8211666262698032, -0.840953097125319, 1.3507866207645483. All qualified roots using GA are -0.7292182525740634, -0.8963994415773797, -1.5330078109106458, 1.3526204886746642.

In Table 7, we now compare our results for $f_5(x)$.

Table 7: Performance of Root-Finding Methods for $f_5(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	4	-0.24415	$2.67570e-09$	18
Genetic Algorithm	10	-0.79759	0.000978809	77
Hybrid Genetic Algorithm	20	-1.821166332	$6.236202e-10$	90

Example 6: The problem is to find all distinct roots of the nonlinear function given below within domain $(-25, 25)$. A hybrid approach using a genetic algorithm for global search and Steffensen's method for local refinement is used to efficiently identify and refine all roots.

$$f_6(x) = \frac{x+20}{x-2}\left|\frac{x\,|x-1|\,(x-3)\,|x-4.1234|}{5\,(x-2)\,|x+1|}\right|. \tag{8}$$

All qualified roots using SGA are 2.9999999971626687, 4.123400011119594, 1.0000000008555652, -20.000084157295, 0. All qualified roots using GA are -2.65020750154822, -6.165051616849415, -19.47550463586774, -9.379646317103019, 3.544717942337133, 3.5639351980430165.

In Table 8, we now compare our results for $f_6(x)$.

Table 8: Performance of Root-Finding Methods for $f_6(x)$

Methods	No. of iterations	Best Solution	Best fitness	Time/(ms)
Steffensen Method	1	-20	0.0	50
Genetic Algorithm	10	1.05537	0.003668218	49
Hybrid Genetic Algorithm	10	-20	$8.82317e-09$	90

4 Conclusion

A comparison of computational times between Steffensen, GA and the hybrid GA demonstrated that our suggested hybrid iterative approach is markedly more efficient in reaching convergence to the root. The examination of optimal iterative schemes in conjunction with GA-based methodologies revealed that the integration of GA with an optimal iterative scheme significantly decreases the number of iterations needed while enhancing accuracy. The increase in efficiency arises from the sensitivity of higher-order iterative methods to their initial points; when initialized distant from the root, these methods either fail to converge or necessitate an excessive number of iterations. The proposed method can determine multiple roots without the knowledge of multiplicity. Such type of hybrid methods can also serve as initial guess generators for higher order root solvers.

References

[1] Cordero, A., Hueso, J. L., Martínez, E., Torregrosa, J. R. Steffensen type methods for solving nonlinear equations. *J. Comput. Appl. Math.* 236(12): 3058-3064, 2012.

[2] Deb, K. Multiobjective optimization using evolutionary algorithms. Wiley, New York, 2001.

[3] Haupt, R.L., Haupt, S.E. Practical genetic algorithms, John Wiley & Sons, Inc., Hoboken, 2004.

[4] Nedzhibov, G. H. Iterative methods for simultaneous computing arbitrary number of multiple zeros of nonlinear equations. *International J. Computer Math.* 90(5): 994-1007, 2013.

[5] Ortega, J.M., Rheinboldt, W.C. Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.

[6] Qiu, T., Hu, H., Chen, R., Zhou, Q., Guan, Q., Li, X. A multi-root solver for discontinuous and non-differentiable equations by integrating genetic algorithm and derivative-free iterative methods. *Appl. Soft Comput.* 109: 107493, 2021.

[7] Wang, Y., Wang, H., Chen, P., Zhang X., Ma, G., Bintao, Y., Al dmour, A. A hybrid computational intelligence method of Newton's method and genetic algorithm for solving compatible nonlinear equations. *Appl. Math. Nonlinear Sci.* 8(1): 1731-1742, 2023.

A Hybrid Approach Using Tuna-Swarm Optimization Algorithm and Newton's Variant for Solving Non-linear Systems

Sonia Bhalla[a], Aanchal Chandel[a], Alicia Cordero[b],
Juan R. Torregrosa[b] and Ramandeep Behl[c,d1]

a Department of Mathematics, Chandigarh University, Gharuan, Mohali 140413 India
b I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.
c Mathematical Modelling and Applied Computation Research Group (MMAC), Department of Mathematics, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia
d Department of Mathematical Science, Saveetha School of Engineering, SIMATS, Chennai 602105, India

1 Introduction

Solving system of nonlinear equations is a common challenge in various engineering fields, including weather forecasting, petroleum geological prospecting, computational mechanics, and control systems. The mathematical expression for system of non-linear equations is represented as

$$\begin{aligned} f_1(u_1, u_2, ...u_n) &= 0, \\ f_2(u_1, u_2, ...u_n) &= 0, \\ &\vdots \\ f_n(u_1, u_2, ...u_n) &= 0. \end{aligned} \tag{1}$$

Generally, finding solutions to such system are difficult due to the lack of an efficient and reliable algorithm. Despite extensive research in this area, existing methods, such as Newton's method [1], Halley's Method [2], Ostrowski's Method [3], Jarratt's Method [4] are widely used but their convergence and performance can be highly sensitive to the initial guess provided. If the initial guess is unsuitable, these algorithms may fail to converge. However, selecting an appropriate initial guess for most nonlinear system is often a challenging task.
With the development of optimization algorithms such as Particle Swarm Optimization (PSO) [5], Differential Evolution (DE) [6], Grey Wolf Optimization (GWO) [7], and Butterfly Optimization Algorithm (BOA) [8], there has been a great interest in utilizing these techniques to solve system of nonlinear equations. The advancement in evolutionary algorithms, in particular, present novel opportunities to address the challenges typically associated with solving nonlinear system. A significant advantage of these methods is their capacity to overcome the common issue of selecting an

[1] ramanbehl87@yahoo.in

appropriate initial guess, a limitation frequently encountered with traditional approaches. Indeed, the system of nonlinear equations (1) can be reinterpreted as an optimization problem in (2).

$$minf(u) = \sqrt{\sum_{i=0}^{n} f_i^2(u)},\ u = (u_1, u_2, ...u_n). \tag{2}$$

The hybridization of iterative methods with optimization algorithms gains significance when addressing complex functions, including various engineering problems. A hybrid algorithm can mitigate the limitations of one approach while leveraging the advantages of the other. Consequently, selecting an appropriate combination of algorithms to create an effective hybrid algorithm is a crucial initial step. The concept of combining evolutionary algorithms with iterative methods for solving complex equations was introduced in 1988 by Karr et al. [9]. They proposed a hybrid approach that utilized the Genetic Algorithm to provide an initial estimate for Newton's method. However, due to its inefficient results, this early hybridization did not gain significant attention at the time. Two decades later, in 2008, Luo et al. [10] addressed complex non-linear equations by integrating Quasi-Newton and Chaos Optimization algorithms. The mathematical convergence of the Chaos Optimization algorithm had not been established at that point, so the convergence analysis of their method was not addressed. More recently, in 2021, Sihwail et al. [11] developed a hybrid algorithm called NHHO, which combined Newton's method with Harris Hawk's optimization technique to solve arbitrary system of non-linear equations. Their work included an examination of non-differentiable functions to demonstrate the method's efficacy. Sihwail et al. in 2022 [12] developed a hybrid method combining Jarratt's method and the Butterfly Optimization algorithm (JBOA) for solving system of non-linear equations. However, in this method, the Butterfly algorithm may get trapped in local optima, resulting in limited exploration of the search space and potentially compromising the accuracy of the solution. Moreover, due to the presence of derivatives at two points in Jarratt's method sometimes, it is difficult or time-consuming to compute the derivative of the function at each iteration. After that, Solaiman et. al in 2023 [13] formulated a modified hybrid algorithm that integrated Newton's method with the Sperm Swarm Optimization Algorithm, achieving enhanced accuracy with fewer iterations. The integration of multiple algorithms can address the limitations of individual methodologies while capitalizing on their respective strengths. Consequently, a crucial initial step is to identify an optimal combination of algorithms to develop an efficacious hybrid approach. Hence, this paper focuses on constructing hybrid approaches that combines derivative-free Newton's iterative scheme [4] and fourth order Newton's variant [16] with the Tuna Swarm optimization algorithm [14]. The objective of this work is to demonstrate the advantages of combining a with derivative and derivative-free iterative approach with an optimization algorithm. This integration aims to reduce the complexity of the iterative method and enhance its accuracy in solving non-linear system and also reduce time-complexity. Moreover, it can address several limitations of Newton's method, including divergence issues, local optima entrapment, and initial point selection and solving large systems with more accuracy and efficiency. The principal contributions of this paper are as follows:

Section 1 presents the introduction of the work. Section 2 includes the derivative-free Newton's Iterative Technique, Newton's fourth order variant the Tuna Swarm optimization algorithm, and its operational mechanism. Section 3 describes the proposed methods and its flowchart. Section 4 presents a comparison of the numerical results of the proposed algorithms with the original optimization algorithms TSO, Newton iterative method. The comparison is based on parameters such as accuracy, stability, fitness value, and convergence speed. Finally, the paper concludes in Section 5.

2 Development of hybrid iterative schemes

Firstly, the Tuna Swarm Optimization (TSO) algorithm and the selected classical iterative methods are discussed in detail. Subsequently, the framework of the proposed hybrid methods namely, derivative-gree Newton-Tuna swarm optimization algorithm (DNTSO) and novel fourth Tuna swarm optimization algorithm (RTSO) is presented and explained.

2.1 Tuna Swarm Optimization Algorithm

In 2021, Lei Xie et al. [14] developed a novel meta-heuristic algorithm called Tuna Swarm Optimization Algorithm. This algorithm is inspired by the foraging behavior of the marine fish species Tuna. The algorithm incorporates two foraging strategies observed in Tuna. The first strategy is spiral foraging, in which Tuna drive their prey into shallow water and attack them following a spiral movement pattern. The second strategy is parabolic foraging, wherein they swim in a parabolic shape to encircle their prey before attacking. Similar to the majority of swarm-based meta-heuristic algorithms, TSO initiates the optimization process by uniformly generating initial populations at random within the defined search space:

$$U_i = rand(ub - lb) + lb, \quad i = 1, 2, 3.....N_p, \tag{3}$$

where N_p is the initial tuna population, $rand$ is a uniformly distributed random vector having value between [0,1] U_i is the i^{th} initial individual, and ub and lb are the upper and lower boundaries of the search space. The spiral foraging behavior in TSO is determined by the following equations:

$$U_i^{t+1} = \begin{cases} \alpha_1(U_{best}^t + \beta|U_{best}^t - U_i^t|) + \alpha_2 U_i^t, & i = 1, \\ \alpha_1(U_{best}^t + \beta|U_{best}^t - U_i^t|) + \alpha_2 U_{i-1}^t, & i = 2, 3....N_p, \end{cases} \tag{4}$$

$$U_i^{t+1} = \begin{cases} \alpha_1(U_{rand}^t + \beta|U_{rand}^t - U_i^t|) + \alpha_2 U_i^t, & i = 1, \\ \alpha_1(U_{rand}^t + \beta|U_{rand}^t - U_i^t|) + \alpha_2 U_{i-1}^t, & i = 2, 3....N_p, \end{cases} \tag{5}$$

$$\alpha_1 = c + (1 - c)\frac{t}{t_{max}},$$

$$\alpha_2 = (1 - c) - (1 - c)\frac{t}{t_{max}},$$

$$\beta = e^{bl} cos(2\pi b),$$

$$l = e^{3cos(((t_{max} + \frac{1}{t}) - 1)\pi)},$$

where, c is a constant used to determine the degree to which the tuna follows the optimal individual and the previous individual in the initial phase α_1 and α_2 are weight coefficients that control the tendency of individuals to move towards the optimal individual and the previous individual; t_{max} is the maximum iterations; t denotes the number of current iterations; and b is a random number uniformly distributed between 0 and 1. U_i^{t+1} is the i^{th} individual of the $t + 1$ iteration, and U_{best}^t is the current optimal individual (food). U_{rand}^t is a reference point in the search space that is generated at random.

The Parabolic behavior in TSO is modelled as:

$$U_i^{t+1} = \begin{cases} U_{best}^t + rand(U_{best}^t - U_i^t) + TFp^2(U_{best}^t - U_i^t), & if\ rand < 0.5, \\ (TF)p^2 U_i^t, & if\ rand \geq 0.5, \end{cases} \tag{6}$$

where,

$$p = \left(1 - \frac{t}{t_{max}}\right)^{\left(\frac{t}{t_{max}}\right)}.$$

where TF is an arbitrary number between $[-1, 1]$. All TSO individuals are iterated and calculated until the termination condition is satisfied throughout the optimization process.

2.2 Derivative-Free Newton-Tuna Swarm Optimization Algorithm (DNTSO):

Newton's method [1] employs an iterative process and exhibits a second-order convergence rate. Consequently, Newton's method demonstrates exceptional efficiency in resolving non-linear equation system, substantially expediting the process of obtaining accurate solutions. The iterative scheme for Newton's method (7) is outlined as follows:

$$u_{n+1} = u_n - [F'(u_n)]^{-1}.F(u_n), \tag{7}$$

where, $F'(u_n)$ denotes the jacobian of $F(u_n)$ and the jacobian here is calculated by using the Finite difference approximation (8) as follows:

$$\frac{df_i}{du_j} \approx \frac{f_i(u+h) - f_i(u)}{h} \tag{8}$$

where $(u + h)$ and (u) involves each element u_j of u and perturbating it by a small step size h. $f_i(u + h)$ and $f_i(u)$ are the function evaluations at $(u + h)$ and (u).

Optimization algorithms have been successfully employed in numerous contexts and possess a wide range of applications. However, as the No Free Lunch (NFL) Theorem [15] states, no algorithm is universally effective for all problems. Moreover, when addressing non-linear system of equations, optimization algorithms may encounter divergence issues or become trapped in local optima. Consequently, the integration of optimization techniques and iterative methods plays a crucial role in solving system of non-linear equations. In this hybrid approach, Newton's method is implemented in each iteration of the Tuna Swarm optimization algorithm. Initially, a best position is considered to be the optimal position as determined by the optimization algorithm. Subsequently, this optimal position is typically refined by using it as input for the Newton's method. Finally, the best positions are evaluated against the output of Newton's method, and the position with the highest fitness is selected. A comparison is conducted between the best position in the optimization algorithm and the best position derived from Newton's method, based on fitness values. The optimal solution is ultimately determined by assessing the best fit at the ideal position.

2.3 Newton's Fourth Tuna Swarm Optimization Algorithm (RTSO):

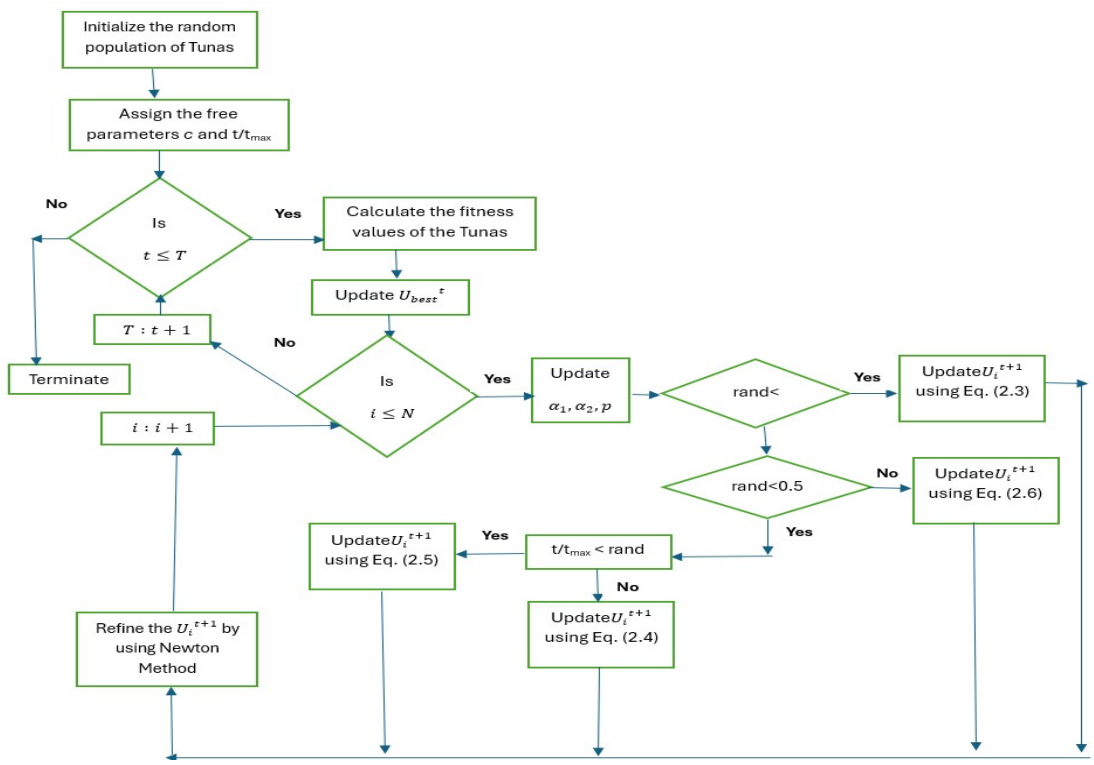
Here, we choose the following fourth-order special case from Behl et al. [16], which is given as follows:

$$\begin{cases} v_n = u_n - \frac{2f(u_n)}{3f'(u_n)} \\ u_{n+1} = v_n - \left(a_1 + a_2\left(\frac{f'(u_n)}{f'(v_n)}\right)^2\right)\frac{f(u_n)}{f'(u_n)}, \quad n = 0, 1, 2... \end{cases} \tag{9}$$

In a similar manner to the expression (7), a fourth-order variant is developed and hybridized with the Tuna Swarm Optimization method (RTSO). This technique is based on a Newton variant that includes the derivative but we follows the same hybridization procedure as DNTSO.

The flowchart presented in Figure 1 illustrates the framework of the proposed DNTSO and RTSO algorithms. In the diagram, the parameter c is a constant that determines the extent to which a tuna follows the optimal individual during the initial phase. The term $\frac{t}{t_{max}}$ represents the ratio of the current iteration t to the maximum number of iterations. Here, t denotes the current iteration, and $t + 1$ denotes the next iteration. Similarly, i refers to the current individual, and $i + 1$ refers to the next individual. The parameters α_1 and α_2 are weight coefficients that help regulate the tendency of an individual to move toward the optimal solution. U^t_{best} represents the best individual found in the current iteration t while U^{t+1}_i denotes the ith individual in the $t + 1$ iteration.

Figure 1: Framework of the proposed DNTSO and RTSO Algorithm



References

[1] Epperson, J.F., An introduction to numerical methods and analysis. Hoboken, New Jersey, John Wiley & Sons, 2013.

[2] Chen, D., Argyros, I.K., Qian, Q.S., A note on the Halley method in Banach spaces, *Applied Mathematics and Computation*, 58(2-3):215–224, 1993.

[3] Ostrowski, A.M., Solution of equations and systems of equations. New York, Academic Press, 1966.

[4] Jarratt, P., Some fourth-order multipoint iterative methods for solving equations, *Mathematics of computation*, 20(95): 434–437, 1966.

[5] Kennedy, J., Eberhart, R., Particle swarm optimization, *Proceedings of ICNN'95-international conference on neural networks*, 4: 1942–1948, 1995.

[6] Storn, R., On the usage of differential evolution for function optimization, *Proceedings of North American fuzzy information processing*, 519–523, 1996.

[7] Mirjalili, S., Mirjalili, S.M., Lewis, A., Grey wolf optimizer, *Advances in engineering software*, 69:46–61, 2014.

[8] Arora, S., Singh, S., Butterfly optimization algorithm: a novel approach for global optimization, *Soft computing*, 23:715–734, 2019.

[9] Karr, C.L., Weck, B., Freeman, L.M. Solutions to systems of nonlinear equations via a genetic algorithm, *Engineering Applications of Artificial Intelligence*, 11:369–375, 1998.

[10] Luo, Y.-., Tang, G.-J., Zhou, L.-N., Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-Newton method, *Applied Soft Computing*, 8(2):1068–1073, 2008.

[11] Sihwail, R., Solaiman, O.S., Omar, K., Ariffin, K.A.Z., Alswaitti, M., Hashim, I., A hybrid approach for solving systems of nonlinear equations using harris hawks optimization and newton's method, *IEEE Access*, 9:95791–95807, 2021.

[12] Sihwail, R., Solaiman, O.S., Ariffin, K.A.Z., New robust hybrid Jarratt-Butterfly optimization algorithm for nonlinear models, *Journal of King Saud University-Computer and Information Sciences*, 34(10):8207–8220, 2022.

[13] Said, S., Obadah, S., Rami, S., Hisham, H., Ishak, A., Kamal, Hybrid Newton–Sperm Swarm Optimization Algorithm for Nonlinear Systems, *Mathematics*, 11(6):1473, 2023.

[14] Xie, L., Han, T., Zhou, H., Zhang, Z.-R., Han, B., Tang, A., Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization, *Computational intelligence and Neuroscience*, 2021(1):9210050, 2021.

[15] Adam, S.P., Alexandropoulos, S.-A. N., Pardalos, P.M., Vrahatis, M.N., No free lunch theorem: A review, Approximation and optimization: Algorithms, complexity, and applications, 57–82, Springer Optimization and Its Applications ((SOIA,volume 145), Cham, Switzerland, Springer, 2019.

[16] Behl, R., Sarría, Í., González, R., Magreñán, Á.A., Highly efficient family of iterative methods for solving nonlinear models, *Journal of Computational and Applied Mathematics*, 346:110–132, 2019.

Analysis and applications of conformable, fractional, fractal and global derivatives

Ali Akgül[1, *]

(1) Siirt University, Art and Science Faculty,
Department of Mathematics, Siirt Turkey

1 Abstract

This work presents a comprehensive analysis of conformable, fractional, fractal, and global derivatives, exploring their mathematical foundations, interrelationships, and diverse applications. These generalized derivative concepts extend classical calculus, enabling more accurate modeling of complex systems characterized by memory effects, non-locality, and self-similar structures. We provide a unified framework that highlights the similarities and distinctions among these derivatives, focusing on their theoretical underpinnings and operational properties. The paper further investigates the practical relevance of each derivative type in various scientific and engineering contexts, including anomalous diffusion, signal processing, control theory, and material science. Emphasis is placed on the advantages and limitations of each approach in modeling real-world phenomena. Through analytical comparisons and illustrative examples, this study aims to advance the understanding and effective utilization of these advanced calculus tools in both academic research and applied problem-solving.

* aliakgul@siirt.edu.tr

Part III

Mathematical Methods for Engineering Problems

Interpretable machine learning prediction of the coal gross calorific value relied on ultimate analysis for energy systems by using the SVM and MLP techniques

P.J. García–Nieto[1,*], E. García–Gonzalo[1], J.P. Paredes–Sánchez[2], L.A. Menéndez–García[3], L. Álvarez–de–Prado[4], and A. Bernardo–Sánchez[4]

(1) Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain,
(2) Department of Energy, College of Mining, Energy and Materials Engineering, University of Oviedo, 33004 Oviedo, Spain,
(3) Department of Industrial Organization and Electronics, Higher School of Engineering and Technology, International University of La Rioja, Avda. de la Paz 137, 26006 Logroño, Spain,
(4) Department of Mining Technology, Topography and Structures, Higher and Technical School of Mining Engineering, University of León, Campus de Vegazana s/n, 24071 León, Spain

1 Introduction

In energy production systems, the higher heating value (HHV), sometimes referred to as the gross calorific value, is a crucial statistic for identifying the main energy source [1,2]. Using support vector machines (SVMs) and the grid search (GS) optimizer [3,4], a new artificial intelligence model was created in this work to describe the coal gross calorific value (dependent variable). The components of coal's ultimate analyses—hydrogen (H), carbon (C), oxygen (O), sulfur (S), and nitrogen (N)—were included in the model. A multilayer perceptron (MLP) and previously established empirical correlations were fitted to the experimental data for comparison. The SVM–based model was used to first illustrate the impact of all independent variables on the coal gross calorific value (dependent variable), and it was later discovered that the SVM–based method was the best forecaster of the coal gross calorific value. Specifically, an SVM regression (SVR) produced a correlation coefficient and a determination coefficient of 0.9991 and 0.9983, respectively, for the coal HHV estimate, based on coal testing samples. The GS/SVM–relied approximation proved to have performed satisfactorily, as evidenced by the agreement between the observed and predicted data. Lastly, a synopsis of the results of these investigations is given.

2 Methods

At the moment, the two primary approaches employed to estimate the HV of fuels are the use of a bomb calorimeter and equations based on those suggested by Dulong [5]. These techniques, which draw from experimental data from proximate or ultimate analyses of solid fuels like coal, are grounded in empirical modeling. Consequently, a plethora of models and mathematical

* pjgarcia@uniovi.es

expressions have been created in order to aid in the characterization of fuel properties [6]. It is possible to use more sophisticated fuel models in this field of study in order to promote energy research in statistical machine learning.

2.1 Experimental dataset

Higher heating values (HHVs) that correspond to the experimental ultimate analyses form the coal data set that was utilized in the present research as crucial input variables. The physico-chemical variables serve as the basis for the dataset used in both the GS/SVM and GS/MLP approaches. Based on past studies on coal as a solid fuel, laboratory analysis provided the data used in this investigation [6], where 318 coal samples values were handled including the following variables: higher heating value (HHV), hydrogen quantity (H), sulfur quantity (S), nitrogen quantity (N), oxygen quantity (O), and finally, carbon quantity (C). In the end, the coal's analysis has choosen the primary input parameters for this investigation with care. The HHV of coals stands as a fundamental characteristic for coal-based energy systems. To assess energy performance and explore viable fuel alternatives, it is imperative to employ more advanced models rooted in machine learning environments. These advanced models are instrumental in aiding researchers in the characterization and management of solid fuels within the energy field.

It is crucial to determine a solid fuel's heating value (HV) [5,6] before using it for thermal applications, especially coal. HV measures the amount of heat of the solid fuel that will be produced by the combustion process of its mass. However, part of this heat is retained as latent heat from the evaporation of the fuel's water during combustion. In this regard, the higher heating value (HHV) represents the maximum energy that can be released during the combustion of the fuel, including the latent heat generated during burning that results from the condensation of water vapor. To estimate HHV in a laboratory setting, an adiabatic bomb calorimeter is typically used; however, this method is not always feasible and can be costly [46]. Furthermore, a solid fuel's ultimate analysis can be used as a substitute to assess its chemical and physical characteristics. Elemental analysis can be employed to determine the weight percentages of hydrogen (H), oxygen (O), nitrogen (N), carbon (C), and sulfur (S).

HHV is relevant because it provides information about the efficiency and energy performance of a fuel. It is used in power generation, heating, industrial processes, and transportation. To characterise in detail the fuel behaviour of coal, these parameters are required for energy management based on standard processes. As a result, the following are the dry-run input variables needed for the model (see Table 1):

- Ultimate analysis:
- Sulphur (S): the sulfur concentration in coal is determined through elemental analysis. It shows the potential contribution to produce SO_x by coal combustion.
- Oxygen (O): elemental analysis uses this to characterize the oxygen content of coal. It is a way of indicating the grade of the evolution of coalification process.
- Nitrogen (N): the part of nitrogen that constitutes coal's atomic structure is this one. It indicates the potential production of NO_x by the combustion of coal.

- Hydrogen (H): the coal's elemental analysis is used to determine the hydrogen content. It demonstrates how volatile matter in coal contributes.
- Carbon (C): it represents the proportion of carbon that constitutes the sample's primary atomic structure. It is an indication of the evolution of the coalification process in coal.

Table 1 lists the five physico-chemical predictors used as input factors for the GS/SVM model studied. In addition, the HHV of the coal, derived from test samples of several varieties of coal, is the objective variable in this research.

Table 1: The physico-chemical input and output factors of the coal, along with their corresponding averages and standard deviations, employed in this investigation.

Input variables	Name of the parameter	Mean	Standard deviation
Carbon content (wt%)	C	78.85	8.11
Hydrogen content (wt%)	H	5.01	0.95
Oxygen content (wt%)	O	13.13	7.95
Nitrogen content (wt%)	N	1.30	0.43
Sulfur content (wt%)	S	1.72	1.88
Output variable			
Higher heating value (MJ/kg)	HHV	30.84	4.03

Note: wt% means *weight percentage*.

2.2 Support vector machines (SVM) method

Next, look at a dataset where the values of the covariates $\mathbf{x}_i \in \Re^p$, $i = 1, 2, ..., m$ and the continuous output factor $y_i \in \Re$, $\forall i = 1, 2, ..., m$ are included in the training set. Consequently, one function $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ is produced via the support vector regression (SVR) technique, where $\mathbf{w}$ is the vector normal to the hyperplane, also referred to as the direction vector. The distance normal between the hyperplane and the origin of coordinates system is denoted by the expression $b/\|\mathbf{w}\|$. Additionally, this approximation must guarantee that the model stays as flat

as possible while permitting a maximum deviation of ε from the true value y_i for all training cases $\mathbf{x}_i$. The Euclidean norm $\|\mathbf{w}\|_2$ is minimized to achieve horizontality, and the sum of differences greater than ε is penalized in order to fit the model. In reality, the support vector regression (SVR) approximation seeks to find a solution of the optimization issue that follows:

$$\min_{\mathbf{w},b,\xi^+,\xi^-} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\left(\xi_i^+ + \xi_i^-\right) \tag{1}$$

liable to

$$\left\{\begin{array}{cc} y_i - \left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq \varepsilon + \xi_i^+ & i = 1,...,m \\ \left(\mathbf{w}^T\mathbf{x}_i + b\right) - y_i \geq \varepsilon + \xi_i^- & i = 1,...,m \\ \xi_i^+, \xi_i^- \geq 0 & i = 1,...,m \end{array}\right\} \tag{2}$$

so that $\boldsymbol{\xi}^+, \boldsymbol{\xi}^- \in \Re^m$ are specified as slack variables and *C* is termed as the regularization constant. In order to limit the penalization applied to data outside of the ε interval and mitigate overfitting, Eq. (1) requires that constant *C* take a numerical value greater than zero. The trade-off between the goal function's horizontal nature and the model's decreased complexity is determined by this value [3,4]. Slack variables, which are linked to each training vector, allow deviations larger than ε while penalizing them in the objective function. Furthermore, an $\varepsilon-$ insensitive tube is the area that $y_i \pm \varepsilon, \ \forall i$ surrounds.

This is a highly nonlinear problem; therefore, we will apply the kernelization method. The original dataset is mapped to a feature space, which is a higher-dimensional space *H*, using this method. A kernel function $K\left(\mathbf{x}_i, \mathbf{x}_j\right)$ is utilized to perform this mapping, permitting a scalar product to be calculated in *H*. We will use its dual form to reformulate the primal optimization problem as it is stated in Eq. (1). The following mathematical expression $f\left(\mathbf{x}\right)$ can be employed to determine the regressive foretelling for a fresh sample **x**:

$$f\left(\mathbf{x}\right) = \sum_{i=1}^{m}\left(\beta_i^+ - \beta_i^-\right)K\left(\mathbf{x}, \mathbf{x}_i\right) + b \tag{3}$$

The technical literature employs a range of commonly used functions as kernels, expressed as [3,4]:

- RBF (radial basis function) kernel:

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = e^{-\sigma\|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{4}$$

- Polynomial type kernel:

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\mathbf{x}_i \cdot \mathbf{x}_j + 1\right)^p \tag{5}$$

Thus, the type of the kernel is determined by the parameters p and σ. In closing, the appropriate kernel type must be selected and its optimal parameters predetermined to transfer nonlinearly separable data into a characteristics space (higher dimensional space termed also *feature space*) and apply the support vector machine (SVM) method to tackle the regression problem. The grid search (GS) optimizer performed effectively in this study.

3 Results

The empirical correlations for the coal gross calorific value (output factor) [3–6], the GS/SVM-based approach with different kernel types, and the determination and correlation coefficients for the MLP-based model are all shown in Table 2 and were assessed using the testing data.

Table 2: Correlation coefficients (r) and coefficients of determination (R^2) for the GS/SVM–relied method fitted with distinct kinds of kernels and the GS/MLP method adjusted together with various correlations found empirically [3–6] adjusted in this study to the coal gross calorific value employed with the test data.

Model	*RBF–SVM*	*Linear–SVM*	*Quadratic–SVM*	*Cubic–SVM*	*MLP*	*E3*	*E5*	*E6*
R^2	**0.9983**	0.783	0.887	0.997	0.935	0.503	0.416	0.421
r	**0.9991**	0.885	0.942	0.998	0.967	0.709	0.645	0.649

The SVM approach with the RBF kernel is the best model for estimating the gross calorific value (HHV) as a dependent variable for various coal types, according to the most recent statistical estimates. This approach achieved a correlation coefficient of 0.9991 and a coefficient of determination of 0.9983 for the coal HHV factor. This selection demonstrates a consistent goodness-of-fit, indicating a suitable agreement between the data of the experimentally obtained measurements of the samples and the SVM approach.

Fig. 1 displays another outcome of these computations, which is the significance ranking of each of the five operation factors foretelling the coal gross calorific value in this comprehensive investigation. According to the SVM model, the carbon quantity (C) is the most significant factor in foretelling the coal gross calorific value (output variable). This is followed by the oxygen (O), hydrogen (H), sulfur (S), and nitrogen (N) concentrations.

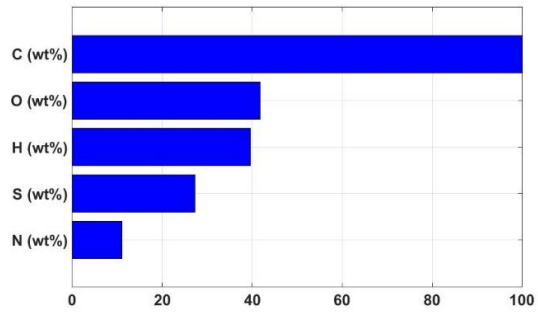
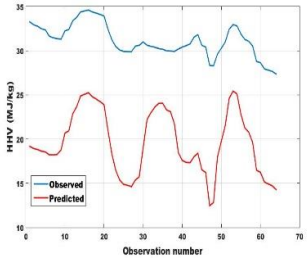
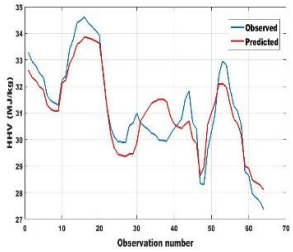


Figure 1: Classification of the operation variables' relative importance used in the best–adjusted GS/SVM–relied technique with RBF kernel for coal HHV forecasting.

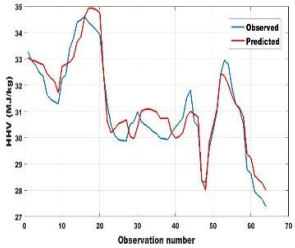
Undoubtedly, the predicted and experimental values of the coal HHV are contrasted in Fig. 2, utilizing the following models: the most accurate empirical correlation E3 (Fig. 2(a)), the linear kernel GS/SVM model (Fig. 2(b)), the GS/SVM model with a quadratic kernel (Fig. 2(c)), the MLP approach (Fig. 2(d)), the GS/SVM model with a cubic kernel (Fig. 2(e)), and the GS/SVM-based approximation with an RBF kernel (Fig. 2(f)). Consequently, the use of an SVM approach proves essential for achieving the most effective solution to the regression problem. These results clearly demonstrate that the GS/SVM-based approach with an RBF kernel provides the best fit while satisfying the critical statistical goodness-of-fit criterion (R^2).



(a)



(b)



(c)

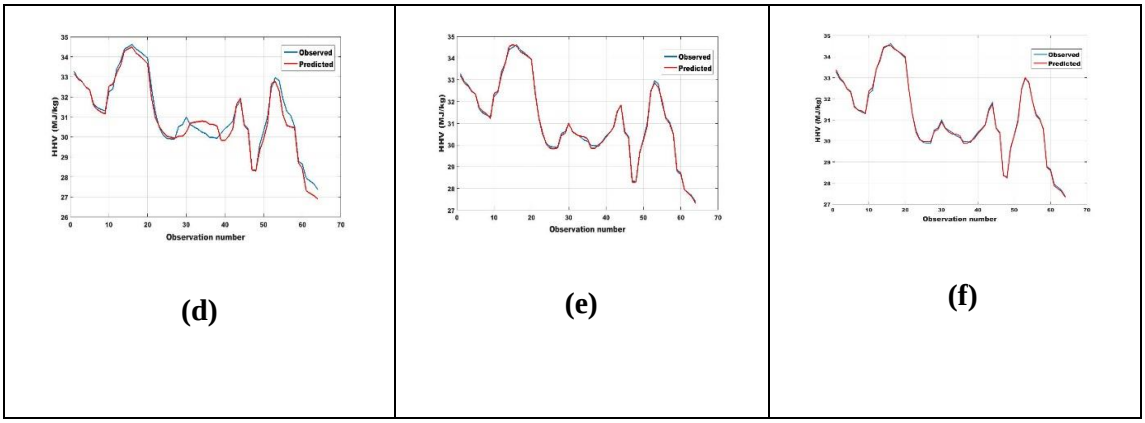


Figure 2: The observed and foretold values of the coal gross calorific value for the test dataset employing: (a) Channiwala and Parikh's empirical correlation (E3) ($R^2 = 0.5025$); (b) GS/SVM model with linear kernel ($R^2 = 0.7832$); (c) GS/SVM model with quadratic kernel ($R^2 = 0.8874$); (d) GS/MLP-type model ($R^2 = 0.9345$); (e) GS/SVM model with cubic kernel ($R^2 = 0.9966$); and (f) GS/SVM model with RBF kernel ($R^2 = 0.9983$).

4 Conclusions

By contrasting the numerical and experimental results, the following succinctly describes the investigation's primary findings:

- Initially, to compute the coal Higher Heating Value factor, one must encounter solution of a challenging heat transfer issue that considers radiation, convection, and conduction—the three distinct forms of heat transmission. The solution to the partial differential equations (PDEs) is derived from the final comprehensive model. Actually, the only way to solve these PDE is numerically (using techniques like the finite differences method, the finite element method, etc.); furthermore, the solutions obtained with additional heuristic approximations differ significantly from one another. Specifically, the most appropriate method to accurately assess the HHV factor in different kinds of coal from distinct deposits and locations is the GS/SVM–relied approximation with RBF kernel employed in this study. It should be noted that the SVM technique with cubic kernel was also a very good option to foretell the coal HHV, but with slightly lower performance than SVM approach with RBF kernel.
- Secondly, it was established that a GS/SVM–relied approximation can be used to calculate the coal HHV precisely in the fuel industry.
- Thirdly, a coefficient of determination of 0.9983 is obtained when the dependent variable, coal HHV, is computed using the GS/SVM–relied approximation (RBF kernel) and the testing data (20% of the observed data not utilized for training).

- Fourthly, for fuel automation applications, a low-cost microcontroller-relied device can be configured with the SVM approximation to accomplish a dependable coal HHV foretelling efficiency.
- Fifthly, the input variables that were used to predict the coal HHV could be arranged according to significance. This is one of the investigation's primary conclusions. Accordingly, the elements O, H, S, and N, in that order, could be regarded as the secondary most important predictors of the coal HHV, after the carbon content (C).
- Sixthly, in subsequent investigations, the concepts showcased herein in order to define hybrid models that incorporate ultimate and/or proximate analyses may be expanded to encompass a greater number of independent variables.
- Finally, the results of this study strongly advise against the widespread application of readily accessible mathematical techniques that heavily depend on the anticipated behavior of the data. A useful GS/SVM–based approximation then provides a useful response to the query of coal HHV estimation in different types of coal relevant to the fuel industry.

In conclusion, other kinds of coal with similar or different origins could be successfully treated with this GS/SVM approach. However, it is always important to take into account the unique characteristics of each basin and deposit.

Acknowledgements

The authors wish to acknowledge the Department of Mathematics at the University of Oviedo for its computational support.

References

[1] Finkelman, R.B., Dai, S., French, D., The importance of minerals in coal as the hosts of chemical elements: a review, *Int. J. Coal Geol.,* 212: 103251, 2019.

[2] Wang, X., Tang, Y., Wang, S., Schobert, H.H., Clean coal geology in China: Research advance and its future, *Int. J. Coal Sci. Technol.,* 7(2): 299–310, 2020.

[3] Hastie, T., Tibshirani, R., Friedman, J.H., The Elements of Statistical Learning, New York, Springer–Verlag, 2003.

[4] Steinwart, I., Christmann, A., Support Vector Machines, New York, Springer, 2008.

[5] Hosokai, S., Matsuoka, K., Kuramoto, K., Suzuki, Y., Modification of Dulong's formula to estimate heating value of gas, liquid and solid fuels, *Fuel Process. Technol.,* 152: 399–405, 2016.

[6] Mason, D.M., Gandhi, K.N., Formulas for calculating the calorific value of coal and coal chars: Development, tests, and uses, *Fuel Process. Technol.,* 7(1): 11–22, 1983.

Optimization of Defect Detection and Classification in Textiles Using ARM Architectures

T. Albero-Albero♮ B. Cantó♭ N. Garrido♭ and R. Pérez♮,[1]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) I.U. Mixto de Tecnología Informática, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Abstract

Defect detection in textile manufacturing is crucial for ensuring product quality and reducing waste. This paper presents an optimized system for defect detection and classification in woven fabrics using artificial vision and classical image processing techniques. The proposed approach adapts image processing software to Advanced Risc Machines (ARM) architecture, enhancing computational efficiency while maintaining high detection accuracy. The study focuses on optimizing traditional image analysis methods to ensure compatibility with embedded hardware, improving real-time processing capabilities for industrial applications.

2 Introduction

The detection and classification of defects in textile fabrics is a critical challenge in the manufacturing industry. Identifying flaws in woven materials ensures product quality, reduces waste, and enhances production efficiency. Traditional defect detection methods rely on manual inspection, which are often time-consuming, inconsistent, and prone to human error [1].

Traditional image processing techniques, such as edge detection, thresholding, and morphological operations, have been widely used for fabric defect detection. These methods are effective in identifying defects like missing threads, stains, and holes, but they often struggle with complex textures and variations in fabric patterns. For instance, edge detection algorithms can highlight discontinuities in the fabric weave, while thresholding can segment regions with different intensities to identify defects. Morphological operations, such as dilation and erosion, are also employed to enhance defect visibility [2].

This study focuses on optimizing a defect detection system based on classical image processing techniques for execution on ARM-based embedded hardware. The optimization process involves transitioning from traditional x86 architectures to Advanced Risc Machines (ARM), which offers advantages in terms of energy efficiency, cost-effectiveness, and scalability for real-world industrial applications. The proposed system employs classical image processing algorithms trained on the AFID dataset, a widely used textile defect database, to identify and classify fabric defects with high reliability [3].

[1] ruperez@disca.upv.es

A key aspect of this research is the development of a simulated linear camera using multiple USB cameras, reducing hardware costs while maintaining high-resolution image capture [4]. Furthermore, motion detection algorithms are implemented to minimize redundant image processing when the fabric is stationary, optimizing computational resources [5]. The final implementation leverages edge computing devices to enable real-time inference directly on the production line.

The primary goal of this research is to optimize a defect detection and classification system for textile fabrics by adapting artificial vision software to ARM architectures. This involves modifying existing image processing algorithms to ensure compatibility and efficiency in low-power embedded systems. Additionally, the study focuses on enhancing the accuracy and robustness of traditional image processing methods to improve defect classification while reducing false positives. Another key objective is the integration of decision-making algorithms into an embedded system, enabling real-time defect detection on industrial production lines to provide a cost-effective, scalable, and high-performance solution for textile quality control.

The remainder of this paper is structured as follows: Section 1 presents the research topic, Section 2 details the methodology used in the process and Section 3 presents experimental results and analysis, and Section 4 concludes the study.

3 Methodology

This section describes the methodology used to develop an optimized textile defect detection system.

3.1 Software Adaptation

The architecture of a processor determines how programs are designed and executed within a system. The implementation of a defect detection system in an industrial setting requires high computational efficiency and real-time processing. Two of the most influential and widely used architectures today are x86 and ARM.

The x86 architecture is known for its Complex Instruction Set Computing (CISC), which allows a single instruction to perform multiple hardware-level operations. This can simplify programming but also increases power consumption and cost, making it less suitable for embedded systems.

Conversely, the ARM architecture follows the Reduced Instruction Set Computing (RISC) paradigm. RISC simplifies instructions, ensuring that each performs a single operation, leading to a more efficient and streamlined processor design. Although this may require more instructions to complete complex tasks, ARM architectures are commonly used in embedded systems and IoT applications, due to their energy efficient and scalability.

While x86 is ideal for applications requiring significant computing power, ARM's energy efficiency and simplified design make it better suited for mobile devices and embedded systems where power consumption and efficiency are key factors.

For this work, both hardware and software have versions available for Linux and Windows, making the choice between x86 and ARM architectures flexible. Thus, migration between these systems is expected to be feasible.

3.2 Image Acquisition

As the goal is to develop a cost-effective vision system, the choice of hardware and software is crucial. USB cameras were selected due to their affordability and ease of use, along with the availability of standard drivers such as Video4Linux (V4L). Most USB cameras exhibit characteristics that make them suitable for vision systems. When implementing ARM-based systems, it is particularly advantageous to use cameras with Video4Linux-compatible drivers. This approach eliminates

dependence on manufacturer-specific drivers, making the application more open and compatible with all cameras supporting V4L.

To reduce costs, image acquisition is triggered based on time intervals rather than events, removing the need for encoders. To prevent redundant image processing, the software will incorporate motion detection control, ensuring that image capture and processing are only performed when the system is in motion.

Video4Linux (V4L) is an application programming interface (API) designed for image capture from various devices, such as TV tuner cards and USB cameras. Its primary goal is to provide a unified interface for a wide range of hardware, enhancing code portability across different applications. Virtually all Linux distributions include kernel support for V4L.

Regarding camera configuration, V4L-compatible cameras offer adjustable parameters that can be optimized for different lighting conditions, improving image clarity and contrast.

3.3 Image Capture Triggering

Vision systems typically trigger image capture based on events. In such systems, the camera or computer receives a signal instructing it to capture an image. In surface inspection applications, this signal is usually linked to an encoder that generates a pulse train based on the configured resolution.

To optimize resources, this system employs time-based image capture instead of event-based triggering. An image is acquired at regular time intervals, denoted as T. The value of T is determined using the same algorithm employed for image processing, ensuring consistency in image acquisition. The same Region Of Interest (ROI) height and number of frames per composite image are maintained for the final application.

A drawback of time-based triggering compared to event-based control is that if the fabric stops moving, the system continues capturing images. This situation frequently occurs in weaving looms and fabric inspection machines due to operator interventions or fabric-related issues such as thread breakage. Although such interruptions are resolved as quickly as possible, they can result in significant downtime and the capture of numerous redundant images, which provide no useful defect detection data.

To mitigate this issue, the system employs a motion detection parameter that evaluates the difference between consecutive images. A threshold is established, below which images are neither captured nor stored, reducing unnecessary processing.

3.4 Image Processing

To decouple image acquisition from processing, these tasks are executed in separate processes. This design choice isolates capture mechanisms from processing algorithms, facilitating their independent modification and exchange. Effective communication between processes is necessary to coordinate information exchange and task synchronization, which is achieved using Inter-Process Communication (IPC) mechanisms.

For synchronization in this project, message queues and semaphores were utilized. A message queue is a data structure managed by the operating system kernel, enabling processes to send and receive messages. Each message has a type and content, allowing selective processing by recipient processes. A semaphore is an abstract data type or variable that regulates access to shared resources among multiple processes, preventing race conditions and ensuring orderly resource usage.

The defect detection program follows a complex structure, as depicted in Figure 1. The system extracts information using Gabor filters, which operate in both spatial and frequency domains. Prior to applying these filters, preprocessing steps and image enhancement techniques are necessary to optimize defect detection accuracy.

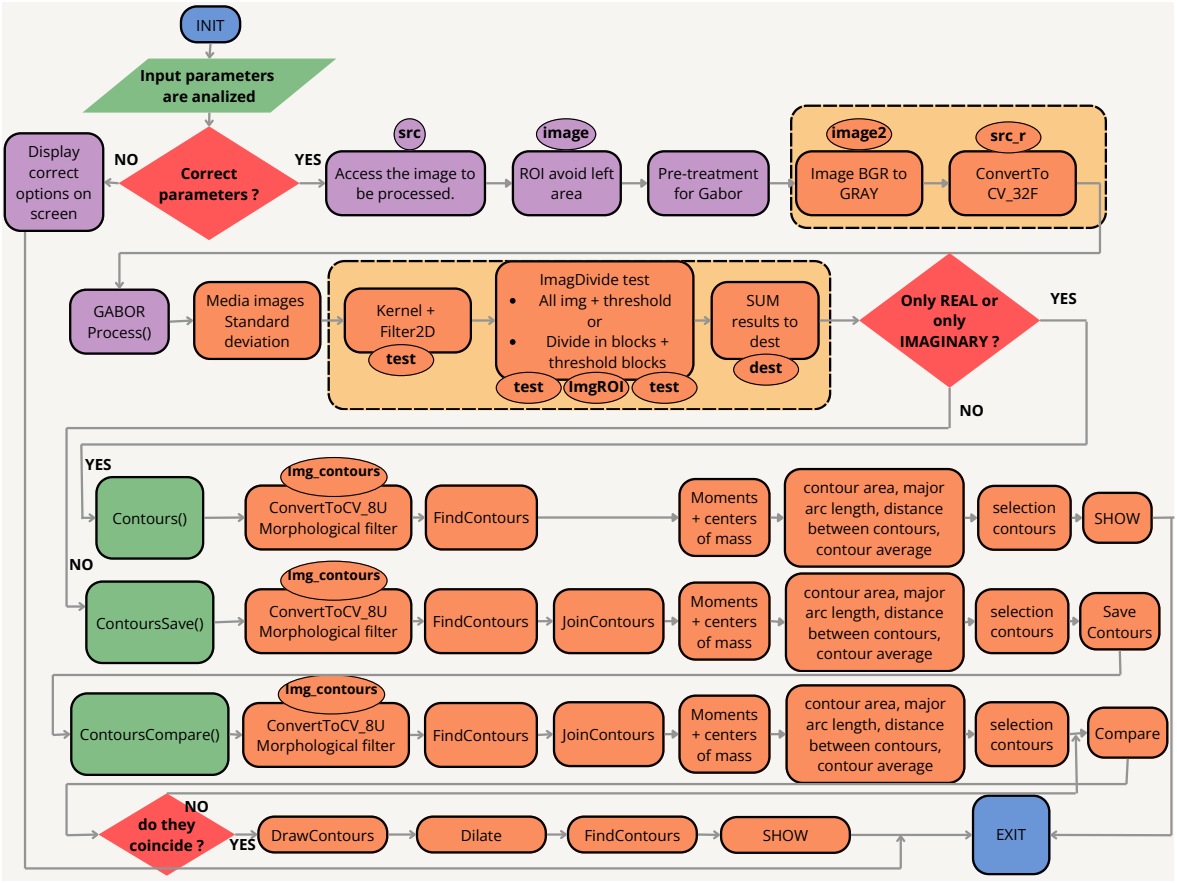


Figure 1: Schematic diagram of the processing algorithm

4 Results and discussion

The defect detection programm was evaluated using the AFID dataset which consists of 245 images from seven different fabric types. Among these, 140 images are defect-free, with 20 images per fabric type. The remaining 105 images contain various types of defects. This database is part of the work published in [3]. Each image has a resolution of 4096×256 pixels.

The evaluation was performed using precision, recall, specificity, and accuracy metrics. One of the most common is measuring the Detection Success Rate (DSR), also known as detection accuracy. This metric is defined as the number of correctly detected samples divided by the total number of samples, including both defective and non-defective ones:

$$DSR = \left(\frac{\text{Number of correctly detected samples}}{\text{Total number of samples}} \right) \times 100$$

Other key metrics include the Detection Rate (DR) and the False Alarm Rate (FAR):

$$DR = \left(\frac{\text{Number of correctly detected defective samples}}{\text{Total number of defective samples}} \right) \times 100$$

$$FAR = \left(\frac{\text{Number of non-defective samples misclassified as defective}}{\text{Total number of non-defective samples}} \right) \times 100.$$

Another common approach is to evaluate specificity and sensitivity. Specificity measures the correct detection of non-defective samples, while sensitivity evaluates the correct detection of defective samples. To compute these values, the following definitions are used:

- TP (True Positive): A defective sample correctly classified as defective.
- FP (False Positive): A non-defective sample incorrectly classified as defective.
- FN (False Negative): A defective sample incorrectly classified as non-defective.
- TN (True Negative): A non-defective sample correctly classified as non-defective.

These values form the confusion matrix given in Table 1

Table 1: Confusion Matrix for Defect Detection

	Classified as Defective	**Classified as Non-Defective**
Actual Defective	TP	FN
Actual Non-Defective	FP	TN

From the confusion matrix, the following performance metrics are calculated:
Sensitivity (Recall):

$$Sensitivity = \left(\frac{TP}{TP + FN}\right) \times 100 \tag{1}$$

Specificity:

$$Specificity = \left(\frac{TN}{TN + FP}\right) \times 100 \tag{2}$$

Using these parameters, the results based on the AFID database provide good results in defect detection, as shown in Table 2.

Table 2: Model Performance on the AFID Dataset

Metric	**Value (%)**
Sensitivity (Detection Rate)	86.6
Specificity (False Positive Avoidance)	88.5
Overall Accuracy	87.7

The results indicate that the system effectively identifies textile defects with high sensitivity, ensuring that most defects are correctly detected. The specificity value shows that the system minimizes false positives, reducing unnecessary alerts in industrial applications.

In addition to AFID, the system was tested on newly captured textile images using the developed real-time vision system. Traditional image processing techniques were applied to ROIs (64×64 pixels) extracted from fabric samples, with performance results summarized in Table 3.

Table 3: Model Performance on Newly Captured Images

Metric	**Value (%)**
Sensitivity	78.5
Specificity	89.2
Overall Accuracy	87.1

5 Conclusions

This study presents an optimized textile defect detection system utilizing classical image processing techniques adapted for ARM-based embedded hardware. The results demonstrate high classification accuracy, with sensitivity and specificity values confirming the system's ability to detect fabric defects reliably. By leveraging edge computing devices, the system achieves efficient defect detection while reducing computational costs.

The model achieved an overall accuracy of 87.7%, ensuring effective defect identification. Additionally, transitioning from x86 to ARM architectures has proven to be cost-effective while maintaining high performance.

Despite these advantages, certain limitations were observed. Sensitivity dropped to 78.5% on newly captured textile images, indicating the need for additional training data and fine-tuning. Variability in lighting conditions affected accuracy, emphasizing the necessity for adaptive preprocessing techniques. Moreover, expanding the dataset with more diverse textile samples would further improve robustness and generalizability.

References

[1] Wei, Q., Zhang, D., A textile architecture-based discrete modeling approach to simulating fabric draping processes *Journal of Industrial Textiles*, 53:1–23, 2023.

[2] Ingale, N., Ansari, A., Yerunkar S., Thakur S., Sayyed S., Fabric defect detection *Int. J. of Creative Research Thoughts*, 11(3):1–5, 2023.

[3] Silvestre-Blanes, J., Albero-Albero T., Miralles I., Pérez-Lloren R., Moreno J., A public fabric database for defect detection methods and results *AUTEX*, 19(4), 2019.

[4] Sathish Kumar, T., Muthuvelammai S., Jayachandran N., AI in Textiles: A Review of Emerging Trends and Applications *Int. J. Research in Applied Science and Engineering Technology*, 12(IX):1–17, 2024.

[5] Ying, W.U., Donghui L.I., Peiyao G.U.O., Yanping L.I.U., Progress in Fabric Defect Detection Based on Machine Learning *J. Shanghai Jiao Tong Univ*, Acepted, 2025.

Prior data driven Physics-Informed Neural Networks for solving source and eigenvalue neutron diffusion problems

E. Navarro-Gamón[♭],[1] A. Vidal-Ferràndiz[♮] D. Ginestar[♮] and G. Verdú[♭]

(♭) I. de Seguridad Industrial, Radiofísica y Medioambiental
Camí de Vera s/n, València, Spain.

(♮) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València,
Camí de Vera s/n, València, Spain.

1 Introduction

Solving the neutron transport equation in a the nuclear reactor core is essential to describe the spatial distribution of power. However, the complexity for the solution of those equations has encouraged the development of several mathematical techniques [1] together with their respective computational implementations [2]. Nevertheless, new deep learning techniques have taken place in the nuclear field, promising more accurate results when solving the neutron diffusion approximation [3, 4].

Recently, the development of Physics-Informed Neural Networks (PINNs) has solved the shortcomings that the initial convolutional neural networks presented. Instead of needing datasets, the training of these neural networks answers and respects the laws of physics that define the problem equations, becoming a tool for solving partial differential equations [5] [6]. In the field of nuclear reactor physics, PINNs have been studied for solving neutron diffusion source and eigenvalue problems [7] and currently, the possibility of using a small dataset to improve the accuracy of the models and reduce the computational time is being evaluated [8].

In this work, we propose several PINN models to solve the one-dimensional and two-dimensional source and eigenvalue neutron diffusion equations, using prior data to strengthen the training. This study shows that using a small dataset of known solutions, provides training PINN models with high accuracy but with less computational time invested.

2 Physically informed Neural Networks

To develop the neural network models, 4 different problems were taken into account. On one hand, a one-dimensional reactor of length l_x of $100cm$ is used to train 2 models, the source problem and the eigenvalue problem for the 1D diffusion equations for 1 group of energy. On the other hand, a two-dimensional reactor with geometry $l_x \times l_y$ and dimensions 100×100 cm is used to also train the source and eigenvalue problem but considering two groups of energy.

[1] emnaga@upvnet.upv.es

2.1 Neutron Diffusion Equation

The neutron diffusion equation for a two-dimensional rectangular homogeneous reactor and 2 groups of energy can be written as equations (1) and (2). These are the equations that the PINN follows to obtain the solution to the fluxes $\phi_1(x, y)$ and $\phi_2(x, y)$, which are considered in the loss function.

$$-\vec{\nabla} D_1 \vec{\nabla} \phi_1(x, y) + (\Sigma_{a1} + \Sigma_{s1\to 2})\, \phi_1(x, y) = \frac{1}{\lambda} \left(\nu\Sigma_{f1}\phi_1(x, y) + \nu\Sigma_{f2}\phi_2(x, y)\right) \tag{1}$$

$$-\Sigma_{s1\to 2}\phi_1(x, y) - \vec{\nabla} D_2 \vec{\nabla} \phi_2(x, y) + \Sigma_{a2}\phi_2(x, y) = 0 \tag{2}$$

Where D_1 and D_2 are the diffusion coefficients for the first and second groups of energy, respectively. The cross-sections for the neutron absorption are Σ_{a1} and Σ_{a2}; the scattering cross-section is $\Sigma_{s1\to 2}$ and the fission cross-sections are $\nu\Sigma_{f1}$ and $\nu\Sigma_{f2}$. Then, λ is the eigenvalue and $\phi_1(x, y)$ and $\phi_2(x, y)$ are the neutron fluxes or eigenvectors of the problem, with analytical solutions expressed in equation (3) and equation (4) respectively.

$$\phi_1(x_i, y_i) = \frac{1}{\lambda} \left(\frac{D_2 \left(\left(\frac{\pi}{l_x}\right)^2 + \left(\frac{\pi}{l_y}\right)^2 \right) + \Sigma_{a2}}{\Sigma_{s1\to 2}} \right) \sin\left(\frac{\pi}{l_x} x\right) \sin\left(\frac{\pi}{l_y} y\right), \tag{3}$$

$$\phi_2(x_i, y_i) = \frac{1}{\lambda} \sin\left(\frac{\pi}{l_x} x\right) \sin\left(\frac{\pi}{l_y} y\right), \tag{4}$$

2.2 Neural Network Design

In order to obtain accurate results, PINN need enough quantity of points to properly define the geometries and understand the problem, and the neural network design must be optimum. Experimentally, the number of collocation points, N_c, to calculate the residual loss was set to $1e^{-4}$, randomly distributed in the 1D and 2D geometries. The number of boundary points, N_b, was set to 250 points for each of the 4 boundaries in the 2D problems and only 2 boundary points in the 1D problems. To reinforce the training, 5 evenly-spaced prior data-points, N_p, were introduced for all the problems, being a few points enough to obtain accurate solutions for the diffusion equation.

The neural networks take the spatial points as input layers and 8 hidden layers with 40 neurons per layer to train the PINN. The number of hidden layers and neurons per layer was determined experimentally and was maintained for every model since the results improved without compromising computational resources.

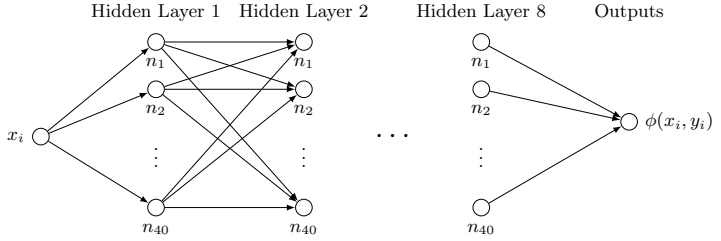


Figure 1: Neural network model for the one-dimensional reactor.

For the one-dimensional reactors, only 1 input is needed, the x_i coordinate, and 1 output is obtained, which refers to the flux $\phi(x_i)$ in each point x_i as is seen in Figure 1. The two-dimensional

reactors need 2 inputs which refer to the (x_i, y_i) coordinates, and as outputs the fluxes $\phi_1(x_i, y_i)$ and $\phi_2(x_i, y_i)$ are obtained for each (x_i, y_i) coordinate, as seen in Figure 2.

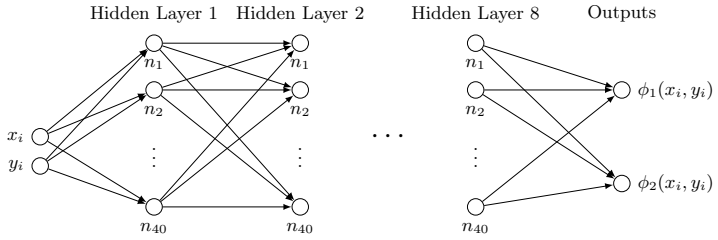


Figure 2: Neural network model for the two-dimensional reactor.

Regarding the python library and activation, the neural network is trained using Keras library [9] from TensorFlow, with the Adam optimizer learning rate set to $1e^{-4}$ for the model and $1e^{-5}$ for the eigenvalue. The activation layer is a hyperbolic tangent and the number of epochs was set to 10^5.

2.3 The Loss Function

The loss function is the definition of the error made by a given neural network state. In other words, is the function minimized by the training process. The loss function [7] is given by

$$Loss = \alpha L_r + \beta L_b + \gamma L_p. \tag{5}$$

The residual loss, L_r, is the term that evaluates the error that the model commits when calculating the neutron diffusion equations in different points of the geometry, called collocation points. The boundary loss, L_b, and the prior data loss, L_p, are the losses due to the mean square error between the known solution and the predicted solution of the neural network in the boundary points and the prior points, respectively. The constants in equation (5) are selected experimentally.

To calculate the residual loss, the neutron diffusion equation for the two-dimensional reactor and 2 groups of energy, is considered. This leads to the residue for the fast neutrons and the residue for the thermal neutrons,

$$R_1(x_i, y_i) = -D_1\vec{\nabla}^2\phi_1(x_i, y_i) + (\Sigma_{a1} + \Sigma_{s1\to 2})\,\phi_1(x_i, y_i) - F_1(x_i, y_i), \tag{6}$$

$$R_2(x_i, y_i) = -D_2\vec{\nabla}^2\phi_2(x_i, y_i) + \Sigma_{a2}\phi_2(x_i, y_i) - F_2(x_i, y_i), \tag{7}$$

where the terms $F_1(x_i, y_i)$ and $F_2(x_i, y_i)$ refer to the source of neutrons, which are taken as in equations (8) and (9),

$$F_{1s}(x_i, y_i) = \frac{1}{\lambda}\left(\frac{D_2\left(\left(\frac{\pi}{l_x}\right)^2 + \left(\frac{\pi}{l_y}\right)^2\right) + \Sigma_{a2}}{\Sigma_{s1\to 2}}\right)\sin\left(\frac{\pi}{l_x}x\right)\sin\left(\frac{\pi}{l_y}y\right), \tag{8}$$

$$F_{2s}(x_i, y_i) = \frac{1}{\lambda}\sin\left(\frac{\pi}{l_x}x\right)\sin\left(\frac{\pi}{l_y}y\right), \tag{9}$$

where λ is the predicted eigenvalue that becomes a trainable variable in the neural network.

$$\lambda = \frac{\nu\Sigma_{f1}\left(D_2\left(\left(\frac{\pi}{l_x}\right)^2+\left(\frac{\pi}{l_y}\right)^2\right)+\Sigma_{a2}\right)+\nu\Sigma_{f2}\Sigma_{s1\to 2}}{\left(D_2\left(\left(\frac{\pi}{l_x}\right)^2+\left(\frac{\pi}{l_y}\right)^2\right)+\Sigma_{a2}\right)\left(\Sigma_{a1}+\Sigma_{s1\to 2}+D_1\left(\left(\frac{\pi}{l_x}\right)^2+\left(\frac{\pi}{l_y}\right)^2\right)\right)}\,. \tag{10}$$

Then, the residual loss that each neural network considers to minimize the total loss, is the mean square error of the residues

$$L_r = \frac{1}{N_c}\sum_{i=1}^{N_c}|R(x_i)|^2\,. \tag{11}$$

Likewise, the models to train the one-dimensional diffusion problems adapt the residual loss to a 1D geometry with one group of energy.

3 Numerical Results

3.1 Results for the one-dimensional reactor

Once the models are trained, what exposes how accurate is the PINN solution, is the comparison between the predicted solution and the analytical solution. First, we test the selected PINN with a one-dimensional reactor, the comparison between the predicted and the expected solution, using the absolute difference, is depicted in Figure 3 for the eigenvalue problem. These results are also reached for the source problem, what confirms that the 1D models can predict with high accuracy the solution, with a maximum absolute difference of 0.030.

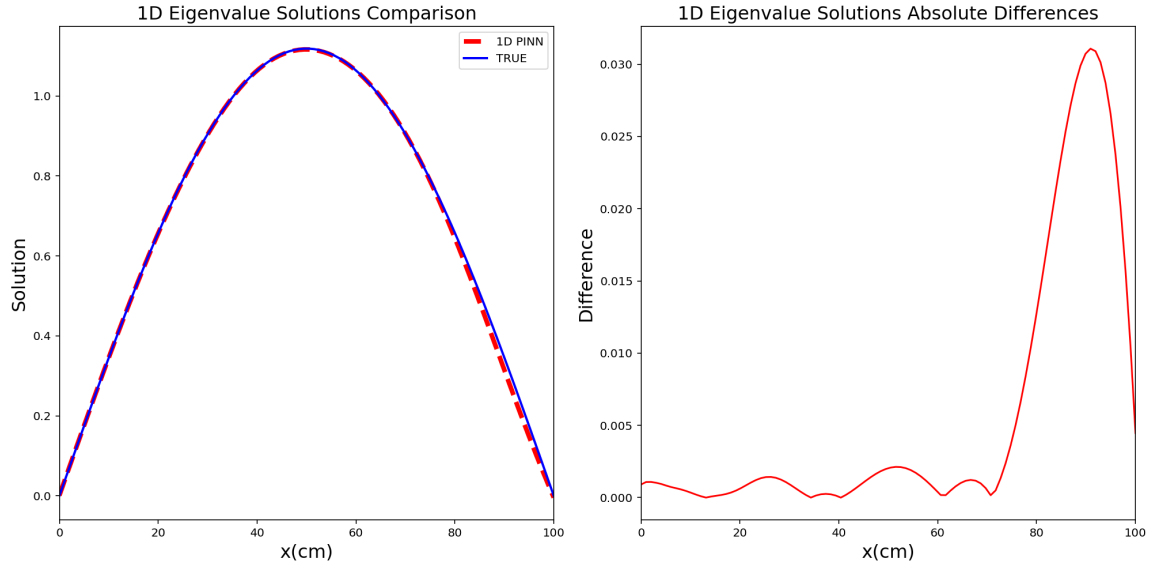


Figure 3: Comparison between the predicted and the analytical solution and absolute differences for the eigenvalue problem in a 1D reactor.

The total loss reached with the 1D eigenvalue model is pictured in Figure 4. The optimizer used is the Adam optimizer, that explains the oscillations of the loss, and it reaches a final loss value under 10^{-5}. Although the loss values oscillate, the tendency is close to reach an asymptote and then, the loss values will not present significant improvements.

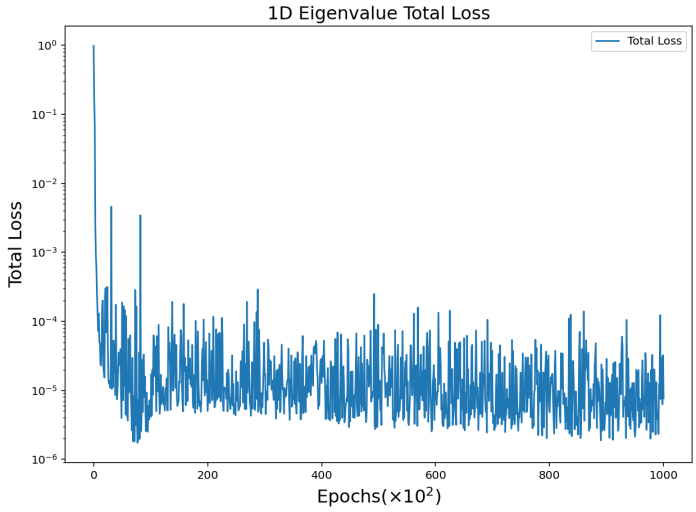


Figure 4: Total loss for the source problem in a 1D reactor.

Regarding the predicted eigenvalue, it tends to the expected value of 0.894106, as can be seen in Figure 5, with an absolute difference between the predicted and real eigenvalue of 14.6 pcm. The computational time needed was about 35 minutes for the one-dimensional source problem and about 61 minutes for the eigenvalue case.

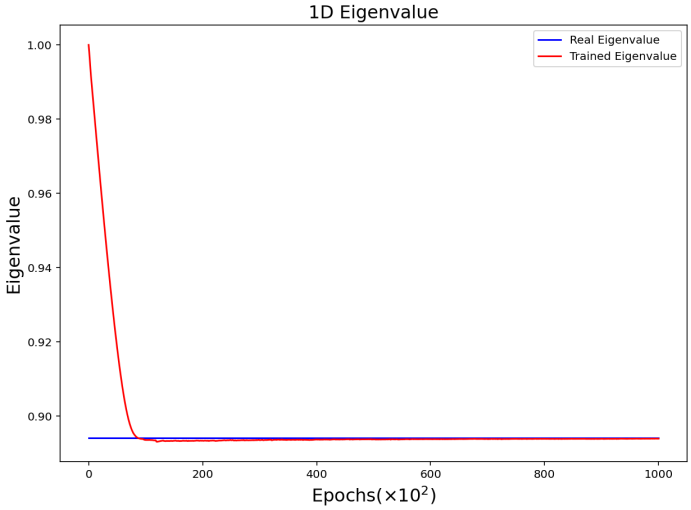


Figure 5: Eigenvalue prediction for a 1D reactor.

3.2 Results for the two-dimensional reactor

The results for the eigenvalue 2D reactor model depict the maximum values in the center of both fluxes as seen in Figure 6 and Figure 7, which is expected for sine distributions, obtaining higher differences close to the boundaries due to the boundary conditions considered. The flux for the

first energy group of energy is four times greater than the flux for the second group of energy, whose maximum value is 1; thus, the absolute differences are approximately 4 times higher for the first group of energy than the absolute differences for the second energy group, with a maximum difference of 0.025 in the boundary areas.

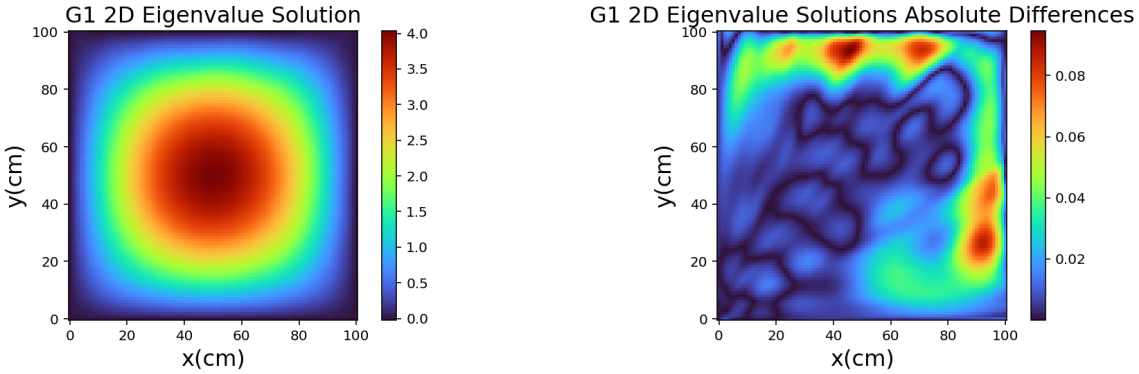


Figure 6: Predicted solution and absolute differences in the fast neutron flux for the eigenvalue problem in a 2D reactor.

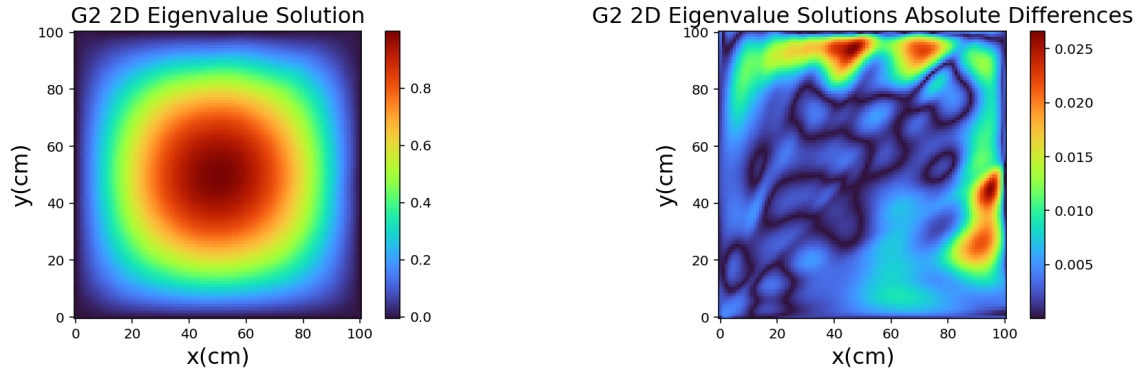


Figure 7: Predicted solution and absolute differences for the thermal neutron flux for the eigenvalue problem in a 2D reactor.

The total loss reached with the 2D eigenvalue model presents the same tendency as the 1D models, since the optimizer used is also Adam. However, the loss values are slightly higher due to the increasing complexity of the model, reaching minimums of $1e^{-4}$. For the predicted eigenvalue for the 2D geometry, pictured in Figure 8, shows a greater difference with respect to the expected value, 273.61 pcm, a result that gives room for improvement.

The results for the two-dimensional source diffusion predicted solutions and the final loss value has the same order as the eigenvalue problem. The computational times were 115 minutes for the source problem and 191 minutes for the eigenvalue problem.

For all the trainings, a GPU NVIDIA GeForce RTX 4060 was used.

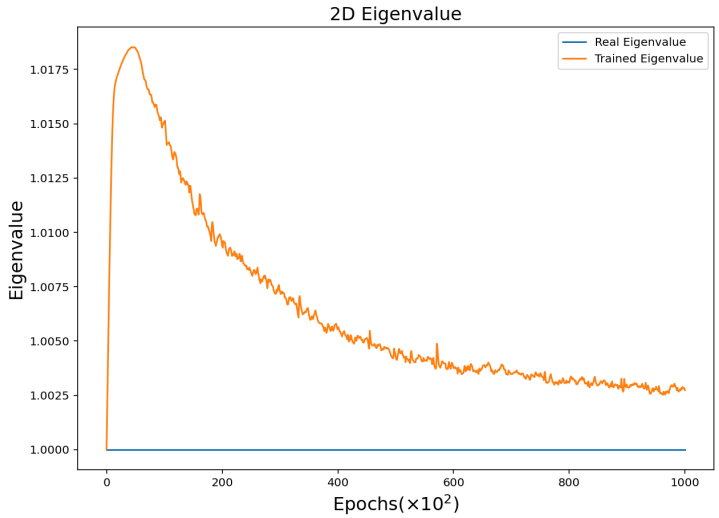


Figure 8: Eigenvalue prediction for a 2D reactor.

4 Conclusions

In this work, we have implemented a Physically Informed Neural Network using prior data points for solving the neutron diffusion equation in homogeneous geometries. After analyzing the results for all the trained models, it can be said that PINNs accurately solve the neutron diffusion equation in a nuclear reactor, although computational times are not competitive with deterministic methods for solving partial differential equations.

The converged solution leads to accurate flux distributions with low differences, which indicates that the loss is close to reaching its minimum and might need a different optimizer. Moreover, the prior datapoints that this model uses help to improve the results and reduce computational time. Only a few prior points were used, which favors the PINN training and avoids the use of previously computed or measured big datasets of known data. However, the minimum number of these datapoints is not studied in this work.

Therefore, due to the results of the 2D models, we will continue to improve the PINNs, focusing on getting lower loss by trying a different combination of optimizers. In future works, the model will be implemented in a two-dimensional reactor with heterogeneous materials.

Acknowledgments

This work was developed under the grant PID2023-152604OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

References

[1] Carreño, A., Vidal-Ferràndiz, A., Ginestar, D., Verdú, G., Development of a diffusive simplified double PN approximation to the neutron transport equation *Annals of Nuclear Energy*, 207: 110675, 2024.

[2] Fontenla, Y., Vidal-Ferràndiz, A., Carreño, A., Ginestar, D., Verdú, G., FEMFFUSION and its verification using the C5G7 benchmark *Annals of Nuclear Energy*, 196: 110239, 2024.

[3] Xie, Y., Wang, Y., Ma, Y., Wu, Z., Neural Network Based Deep Learning Method for Multi-Dimensional Neutron Diffusion Problems with Novel Treatment to Boundary *Journal of Nuclear Engineering*, 2: 533–552, 2021.

[4] Zhang, Q., Zhang, J., Liang, L., Li, Z., Zhang, T.,A Deep Learning Based Surrogate Model for Estimating the Flux and Power Distribution Solved by Diffusion Equations *PHYSOR2020 – International Conference on Physics of Reactors: Transition to a Scalable Nuclear Future*, 247: 3013–3021, 2021.

[5] Raissi, M., Perdikaris, P., Karniadakis, G.E., Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations *Journal of Computational Physics*, 378: 686–707, 2019.

[6] Lagaris, I.E., Likas, A., Fotiadas, D.I., Artificial Neural Networks for Solving Ordinary and Partial Differential Equations *IEEE Transactions on Neural Networks*, 9: 987–1000, 1998.

[7] Elhareef, M.H., Wu, Z., Physics-Informed Neural Network Method and Application to Nuclear Reactor Calculations: A Pilot Study *Nuclear Science and Engineering*, 197: 601–622, 2023.

[8] Yang, Y., Gong, H., Zhang, S., Yang, Q., Chen, Z., He, Q., Li, Q., A data-enabled physics-informed neural network with comprehensive numerical study on solving neutron diffusion eigenvalue problems *Annals of Nuclear Energy*, 183: 109656, 2023.

[9] Chollet, F., et al., Keras. https://keras.io, 2015.

Decision-making optimization for redevelopment projects in the marble industry

S. Carpitella [♭], F. Carpitella [♮], and J. Izquierdo[♢],[1]

(♭) California State University, Northridge, 18111 Nordhoff St, Northridge, California, US.
(♮) Studio di Ingegneria Carpitella, V. Fabrizi no. 3, Trapani, Italy.
(♢) Universitat Politècnica de València, Cno. de Vera s/n, Valencia, Spain.

Abstract

Marble extraction and processing are key contributors to the global economy, especially in such critical sectors as construction and design. With rising demand for sustainable and high-quality materials, the marble industry faces increasing pressure to optimize its operations. This study proposes a decision-making model to support the activities carried out within a redevelopment project for a marble company operating in the territory of Sicily, Italy. In this context, multiple factors are modeled to support strategic planning in a project focused on environmental rehabilitation and socio-cultural enhancement of a quarry, where several potential pathways are assessed. The proposed approach uses expert opinions forming the body of input data, making the model flexible and adaptable to other companies operating in similar sectors and contributing to broader industrial improvement.

1 Introduction and Literature Review

Marble extraction and manufacturing represent significant activities in the global economy, particularly within the construction, design, and luxury goods sectors [1]. In today's economic context, characterized by increasing demand for sustainable building materials and a growing focus on high-end aesthetics, marble represents both a cultural symbol [2] and a strategic commodity [3]. Apart from fueling employment and regional development in several producing countries [4], the marble industry supports a vast international trade network [5]. Furthermore, technological advancements in quarrying and processing have enhanced efficiency and minimized environmental impact [6], reinforcing marble's relevance in modern architecture and global markets [7]. In light of consumer preferences for materials that are natural, long-lasting, and visually appealing, the marble industry continues to represent a cornerstone of both local economies and contemporary industrial innovation [8].

The use of such mathematical models as MCDM (Multi-Criteria Decision-Making) techniques can be highly beneficial for optimizing processes in marble extraction and manufacturing, as well as the overall management of the sites, since these processes often involve complex decisions that must balance multiple factors, e.g., cost, quality, environmental impact, energy use, and productivity. MCDM methods help decision-makers evaluate and compare different options by considering

[1] jizquier@upv.es

all relevant criteria at the same time, rather than focusing on just one aspect [9]. In this way, companies are better equipped to identify efficient solutions, enabling sustainable resource management, while minimizing waste and enhancing the overall performance of their operations. This structured approach to decision-making also supports improved planning and helps reduce the risks associated with daily operations [10]. Lastly, in a competitive and resource-sensitive industry like marble production, using structured decision-making tools ensures that choices are data-driven, transparent, and aligned with both economic and environmental goals [1].

The goal of this research is to develop an MCDM model designed to support a redevelopment project for a marble company based in Sicily, Italy. The model is based on the application of VIKOR (a Serbian acronym that stands for Multi-Criteria Optimization and Compromise Solution), which has already been applied in manufacturing contexts [11] and has proven effective in supporting the achievement of sustainability objectives [12]. Particularly, VIKOR is a decision-making method used to rank and select from among alternatives in the presence of conflicting criteria [13]. In the context of the present research, this method aims to assist in evaluating different strategic and operational choices by considering a range of relevant criteria, to help the company improve efficiency, sustainability, and decision-making across its activities.

The methodological steps for applying VIKOR are detailed in Section 2. A case study involving an Italian marble company is presented in Section 3, while conclusions and directions for future research are provided in Section 5.

2 Methodological Approach

The implementation of the VIKOR method involves the following steps [12].

- Formation of the input decision matrix. The input decision matrix is formulated by gathering the n alternatives, the m criteria, and the evaluations of alternatives x_{ij} under each criterion.

- Standardizing the decision matrix. Normalization of the values in the decision matrix is undertaken to guarantee comparability among the criteria and establish a consistent scale, achieved through the following formula:

$$f_{ij}(x) = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}. \tag{1}$$

- Identifying the optimal and suboptimal performance for each criterion. This involves determining the ideal and anti-ideal solutions, representing the best and worst performance levels, respectively, for each criterion, as elaborated below. If the criterion exhibits a positive preference direction, then:

$$f_j^* = \max f_{ij}; f_j^- = \min f_{ij}; j = 1 \dots m. \tag{2}$$

  If the criterion has a negative preference direction, then

$$f_j^* = \min f_{ij}; f_j^- = \max f_{ij}; j = 1 \dots m. \tag{3}$$

  The positive ideal solution f^* and negative ideal solution f^- can be expressed as follows:

$$f^* = \{f_1^*, f_2^*, \dots, f_m^*\}; \tag{4}$$

$$f^- = \{f_1^-, f_2^-, \dots, f_m^-\}. \tag{5}$$

- Computing the collective utility and individual regret. This step is undertaken to empower decision-makers in discerning the alternative that attains the highest overall performance

(group utility, S_i), simultaneously excelling in individual criteria (minimal individual regret, R_i). These metrics are computed as follows:

$$S_i = \sum_{j=i}^{m} w_j \frac{f_j^* - f_{ij}}{f_j^* - f_j^-}; \tag{6}$$

$$R_i = \max \left[w_j \frac{(f_j^* - f_{ij})}{(f_j^* - f_j^-)} \right]; \tag{7}$$

w_j being the weight assigned to criterion j.

- Achieving the VIKOR index for each alternative. This involves calculating the measure Q_i, which considers both the group utility (proximity to the ideal solution) and individual regret (deviation from the best performer for each criterion). The VIKOR index is computed as:

$$Q_i = \gamma \frac{S_i - S^*}{S^- - S^*} + (1 - \gamma) \frac{R_i - R^*}{R^- - R^*}; \tag{8}$$

where $\gamma = 0.5$ indicates the maximum group utility, $S^* = \min\{S_i\}$, $S^- = \max\{S_i\}$, $R^* = \min\{R_i\}$, and $S^- = \max\{R_i\}$.

- Prioritizing alternatives according to their S, R, and Q values. The ranking is achieved by organizing them in descending order, guaranteeing that the alternative with the smallest VIKOR value receives the highest rank. The result yields three separate ranking lists.

- Recommending a compromise solution. This proposed solution typically entails choosing the alternative that attains the highest overall ranking, considering all three values (S, R, and Q). This alternative represents the most balanced compromise, encompassing considerations of group utility, individual regret, and overall performance.

3 Case Study

This case study focuses on Ducale Marmi A.D. Srl, an Italian company based in Custonaci in the province of Trapani, Sicily, an area known for the extraction and processing of marble. Custonaci is the leading marble production center in Sicily, the second in Italy and in Europe. The area covers a basin of approximately 62 square kilometers, from which 85% of Sicilian marble is extracted. Marble from Custonaci is now exported all over the world.

After World War II, thanks to industrialization, Custonaci gained a prominent role on the international stage due to the marble variety known as *Perlato di Sicilia*. Alongside this variety, the *Perlatino di Sicilia* is also extracted today. is also extracted today. Both are uniform beige-colored marbles, appreciated for their aesthetic qualities and consistency. *Perlato di Sicilia* is the second best-selling marble in the world after the Carrara white marble [14]. It is light ivory with delicate yellow-brown veining caused by concentrations of iron oxides and clay minerals. In high demand both in domestic and international markets, this marble is used in prominent sites such as Milan's Central Station and the cladding of St. Peter's Basilica in the Vatican. In addition to this, the Custonaci marble basin is also rich in so-called "colored" marbles, a term used to describe all non-white varieties. Among them, the colored variety known as *Diaspro* that is sourced from quarries in certain areas of Custonaci is considered the most prestigious ornamental marble produced in Sicily, thanks to its distinctive multicolored aesthetic, featuring shades of red, yellow, and green. In this regard, quarrying activity in the Custonaci marble basin has important historical roots dating back to around 1600, with these materials used to adorn significant buildings such as the Royal Palace of Caserta and St. Peter's Basilica in Rome. Figure 1 shows a group of marble blocks of this type, extracted from the Custonaci area.



Figure 1: *Diaspro* marble blocks and example of its use for floor realization

The production of *Diaspro* is limited, as it almost exclusively originates from a small vein positioned at a 45-degree angle relative to the quarrying plane. This vein is embedded within and contrasted by surrounding materials of different colors. It has an average thickness of about one meter and tends to disappear at a certain depth.

This vein could result from mineral-rich fracture fillings that formed during the uplift of the marble deposits from the sea. These mineral deposits are responsible for the distinctive colors typical of *Diaspro*. As a result, the *Diaspro* vein always has a specific location, orientation, and inclination within the Custonaci marble basin. Due to its historical and artistic value, *Diaspro* is often requested by heritage authorities for the restoration of numerous churches where it was originally used. It is also studied and admired by art students from around the world.

Around the 1950s, colored marbles were extracted from numerous, often small-scale quarries scattered throughout the territory. They met strong demand, especially in the local construction sector and throughout Southern Italy. This is evidenced by their widespread application in flooring, staircases, and wall coverings in residential and public structures. However, over time, shifting market trends began to favor lighter, monochromatic marbles. As a result, the commercial interest of the marble industry in Custonaci gradually shifted toward producing the varieties *Perlato* and *Perlatino*. This change enabled companies to organize for mass production, scaling up operations significantly. Consequently, the Custonaci marble district experienced exponential growth and, until a few years ago, was considered the second most important marble basin in Italy, after Carrara. Due to this shift, many colored marble quarries were progressively abandoned. Today, these can be described as dormant deposits. In recent years, however, there has been a renewed interest in colored marble, particularly within private residential projects led by interior designers. These professionals creatively incorporate marble in innovative ways, such as wall cladding or integrating marble with materials like wood and metal. At the same time, artisans and marble workers are investing in cutting-edge CNC (Computer Numerical Control) machinery that can realize such designs. These emerging trends may breathe new life into the marble market and revitalize Custonaci's marble-based economy.

A current redevelopment project aimed at territorial and environmental recovery plans to make certain vertical and horizontal areas of one of the quarries accessible and functional. Two views of these spaces are shown in Figure 2.

In particular, the innovative concept at the heart of this project involves repurposing these quarry areas to host temporary and/or permanent exhibitions of artworks, transforming the site into a unique cultural and artistic space. The project will begin by involving young Sicilian artists,



Figure 2: Representation of the areas targeted for the redevelopment project

who will have access to the quarry spaces and use locally extracted materials to create their works, while also emphasizing the surrounding environment and the landscape value of the site. The implementation of this project is intended to serve as a platform for connecting the worlds of art, architecture, and design. The focus will be on the natural raw material extracted from this territory, explored through the eyes of artists, architects, and designers, to discover new fields of application beyond its traditional use in construction.

To support the implementation of this project, ten options for potential visitor paths within the quarry have been analyzed under multiple criteria. Each path explores a different route through the site, considering how visitors may engage with viewpoints and natural elements. We apply the VIKOR method to select the path that best balances the following criteria:

- C_1 environmental impact (evaluating how each path minimizes resource use, emissions, and environmental harm during the sourcing, processing, and installation of marble);
- C_2 accessibility (assessing how well the requalified space accommodates users of all abilities, ensuring safe and inclusive access for everyone);
- C_3 functionality (measuring the practicality and durability of marble elements in everyday public use, including maintenance and performance over time);
- C_4 visitor experience (considering the visual, cultural, and emotional impact of the marble design on users, enhancing the identity and appeal of the space).

Table 1: Input matrix with evaluations of paths under criteria

Path	**C1**	**C2**	**C3**	**C4**
P_1	3	4	2	3
P_2	5	4	3	4
P_3	2	3	4	2
P_4	4	3	2	3
P_5	3	2	3	2
P_6	4	4	3	3
P_7	4	3	3	3
P_8	2	3	4	2
P_9	1	4	2	3
P_{10}	5	3	4	3

Table 2: Group utility (S_i), minimal individual regret (R_i), VIKOR index (Q_i) and ranking lists

Path	$\mathbf{S_i}$	**Rank in** S_i	$\mathbf{R_i}$	**Rank in** R_i	$\mathbf{Q_i}$	**Rank in** Q_i
P_1	0.500	5	0.250	3	0.800	5
P_2	0.125	1	0.125	2	0.000	1
P_3	0.562	6	0.250	3	0.850	6
P_4	0.562	6	0.250	3	0.850	6
P_5	0.750	8	0.250	3	1.000	8
P_6	0.312	3	0.125	2	0.150	3
P_7	0.438	4	0.125	2	0.250	4
P_8	0.562	6	0.250	3	0.850	6
P_9	0.625	7	0.250	3	0.900	7
P_{10}	0.250	2	0.125	1	0.100	2

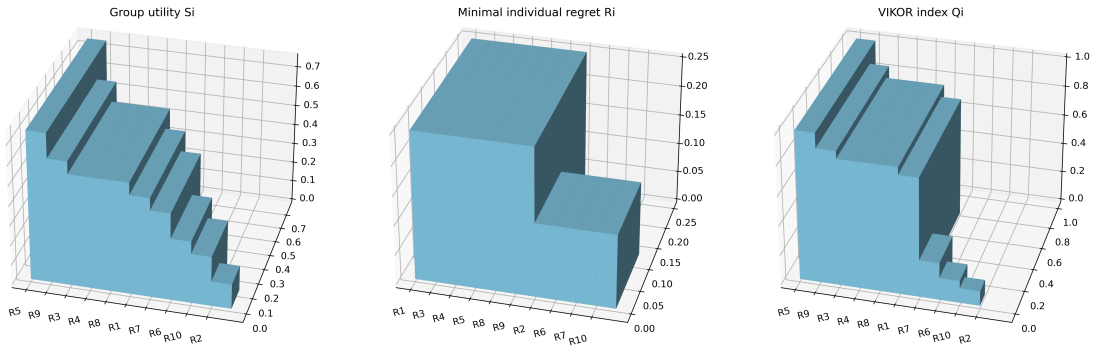


Figure 3: VIKOR results

To ensure the relevance and accuracy of the developed assessment, we involved an expert with extensive experience in the marble industrial sector, who provided valuable insights for the data collection process through collaborative brainstorming sessions and iterative feedback exchanges.

Numerical values in Table 1 correspond to the expert's subjective assessments, informed by specialized sector knowledge. These evaluations indicate the extent to which each of the ten proposed paths satisfies the established criteria within the framework of the redevelopment project. Specifically, a five-point scale was used, where a score of 1 indicates low alignment with the criterion, and 5 indicates strong alignment. All four criteria were regarded as equally significant and assigned equal weight in the VIKOR analysis.

Table 2 reports values of group utility and minimal individual regret, respectively obtained via formulas (1.6) and (1.7) along with the VIKOR index for each alternative, the last one achieved by applying formula (1.8). The ranking list for each of these values is also specified in the same table. Results are graphically synthesized in Figure 3. The alternative with the associated minimum VIKOR index (in our case, P_2) represents the most suitable compromise solution only if the condition of acceptable advantage and the condition of acceptable stability [12] are simultaneously met. As this is not the case, alternatives P_2 and P_{10} (i.e., the alternative ranked second in the Q-based ranking list) are selected as compromise solutions to be prioritized.

In practical terms, this means that among the ten proposed design or development paths, P_2 and P_{10} offer the most balanced solutions when considering all four key aspects of the project: minimizing environmental impact, ensuring accessibility, enhancing functionality, and improving the overall visitor experience. These paths are not necessarily the top performers in every category,

but they provide the most favorable overall compromise across all criteria, making them the most suitable options for implementation in the redevelopment plan and potentially the most appropriate designated routes for visitors within the intervention areas.

4 Conclusions and Future Research

This research investigates the application of mathematical modeling with a specific focus on MCDM methods in the context of sustainable territorial redevelopment within the marble industry. The study demonstrates the potential of the VIKOR method to support complex decision-making processes where multiple and often conflicting criteria including environmental, functional, and socio-cultural factors must be evaluated simultaneously. The proposed methodological approach facilitates the identification of a balanced compromise solution, offering valuable guidance for strategic planning in projects involving environmental rehabilitation and socio-cultural enhancement. A key contribution of this work is the application of the model to a real-world case study set in Italy, within the marble-rich area of Custonaci. In this context, several potential pathways for redeveloping a marble quarry were defined and assessed. The use of the VIKOR method enabled the prioritization of these alternatives through a transparent, criteria-driven process based on expert judgment, ultimately identifying the most suitable paths to guide future development of the site, to recover and enhance key areas of the surrounding territory.

For future research, we aim to formalize actual visitor routes based on the paths identified in the present research through the VIKOR method. This will involve defining dedicated stops at points of particular interest along each path, thereby creating structured itineraries that enhance visitor engagement. Additionally, future research will explore integrating Artificial Intelligence (AI) tools to further refine the evaluation process and support more informed decision-making. AI could also play a key role in the implementation phase of the redevelopment project, aiding in route optimization, real-time visitor flow analysis [15], and adaptive site management to maximize user experience and operational efficiency. This approach would aim to enhance both the experiential quality and functional connectivity of the site, while further promoting its cultural, environmental, and economic value.

Acknowledgments

The authors would like to express their sincere gratitude to the company Ducale Marmi A.D. Srl, located in Custonaci, Italy, for the valuable support and collaboration throughout this research. In particular, special thanks are extended to Francesco Casadei, whose cooperation and insights were instrumental in the development and elaboration of the present work.

References

[1] Azevedo, E., Carpitella, S., Sustainable supply chain control strategies for the natural stone market. *IFAC-PapersOnLine*, 58(4): 562–567, 2024.

[2] Herz, N., Waelkens, M. (Eds.), Classical marble: geochemistry, technology, trade. *Brill Archive*, 153, 1988.

[3] Oweh, S. O., Aigba, P. A., Samuel, O. D., Oyekale, J., Abam, F. I., Veza, I., ..., Sener, R., Improving Productivity at a Marble Processing Plant Through Energy and Exergy Analysis. *Sustainability*, 16(24): 11233, 2024.

[4] Ahmad, Z., Khan, S. M., Ali, M. I., Fatima, N., Ali, S., Pollution indicandum and marble waste polluted ecosystem; role of selected indicator plants in phytoremediation and determination of pollution zones. *Journal of cleaner production*, 236: 117709, 2019.

[5] Al-Haddad, L., Sial, M. S., Ali, I., Alam, R., Khuong, N. V., Khanh, T. H. T., The role of small and medium enterprises (SMEs) in employment generation and economic growth: a study of marble industry in emerging economy. *International Journal of Financial Research*, 10(6): 174-187, 2019.

[6] De Medeiros, W. R. P., Neto, O. D. M. M., Luz, P. M. S. G., de Oliveira, R. K. F., Guedes, L. R., Utilizing marble and granite industry waste in asphalt mixtures for enhanced road performance and sustainability. *Journal of Engineering Research*, 12(4): 670-679, 2023.

[7] Careddu, N., Siotto, G., Marras, G., The crisis of granite and the success of marble: errors and market strategies. The Sardinian case. *Resources Policy*, 52: 273-276, 2017.

[8] Partarakis, N., Zabulis, X., Meghini, C., Dubois, A., Moreno, I., Ringas, C., . . . , Krivokapic, J., A Review, Analysis, and Roadmap to Support the Short-Term and Long-Term Sustainability of the European Crafts Sector. *Heritage*, 2, 70, 2025.

[9] Taherdoost, H., Madanchian, M., Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia*, 3(1): 77-87, 2023.

[10] Bhalaji, R.K.A., Sankaranarayanan, B., Saravanasankar, S., Risk Factors in the Process industry by using MCDM Methods: Literature from 2014 to 2019. *International Journal of Engineering and Advanced Technology*, 9: 840-846, 2019.

[11] Carpitella, S., Abolghasem, S. Kunta, K.R., Enhancing supply chain sustainability in modern manufacturing. *Proceedings of the 5th South American International Conference on Industrial Engineering and Operations Management, Bogota, Colombia*, May 7-9, 1-8, 2024.

[12] Carpitella, S., Overcoming barriers of digital transformation towards greener supply chains in automotive paint shop operations. *Sustainability*, 16(5): 1948, 1-20, 2024.

[13] Çalı, S., Balaman, S.Y., A novel outranking based multi criteria group decision making methodology integrating ELECTRE and VIKOR under intuitionistic fuzzy environment. *Expert Systems with Applications*, 119: 36–50, 2019.

[14] Todaro, S., The potential geosite of the "Libeccio Antico" quarries: A sedimentological and stratigraphic characterisation of ornamental stone from Mt Cocuccio, Custonaci Marble District, Sicily. *Geoheritage*, 11(3): 809-820, 2019.

[15] Centorrino, P., Corbetta, A., Cristiani, E., Onofri, E., Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization. *Journal of Computational Science*, 53: 101357, 2021.

An extended LUPOD technique to construct a reduced order model for the steady state neutron diffusion equation

A. Vidal-Ferràndiz[♭],[1] A. Carreño[♮], D. Ginestar[♭] and G. Verdú[♮]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València,
Camí de Vera s/n, València, Spain.

(♮) I. U.de Seguridad Industrial, Radiofísica y Medioambiental (ISIRYM),
Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

To study the stationary spatial power distribution in a nuclear reactor core, the energy multi-group approximation of the neutron diffusion equation is widely used. This model is a differential generalized eigenvalue problem of the form [4],

$$\mathcal{L}\Phi = \frac{1}{\lambda}\Phi \ , \tag{1}$$

where $\mathcal{L}$ is the neutron loss differential operator and $\mathcal{M}$ is the neutron production operator. Once a spatial discretization is selected, problem (1) is transformed into a large generalized eigenvalue problem that has to be solved to obtain the dominant eigenvalue, which is the k-effective of the reactor configuration and its corresponding eigenvector that gives the neutron flux in steady state for the different energy groups.

In different studies, such as the control rod worth calculations, [8] or uncertainty quantification [10], the partial eigenvalue problem (1) has to be solved many times for different configurations of the core. Since the problem is large, it has a considerable computational cost, and developing a fast and accurate reduced order model (ROM) for this problem is convenient.

Reduced order models approximate complex models, trying to describe the main features of the full order model (FOM) with a small number of state variables. For this, the model is projected onto a given space of snapshots in an offline phase. ROMs based on snapshots generated by means of a proper orthogonal decomposition (POD) for the neutron diffusion equation have been proposed, for example, in [2]. This method uses a set of orthogonal snapshots obtained from the full order solutions of different core configurations. These snapshots are long vectors, and constructing the reduced order model can still be costly. To reduce this cost, the projection of the FOM onto the subspace generated by a set of orthogonal modes is approximated using a reduced scalar product based on a set of collocation points [7]. The selection procedure of the main snapshots and the collocation points is based on a LU decomposition of the snapshots matrix with pivoting of rows and columns. The resulting algorithm is called LUPOD [6]. This method restricts the maximum

[1] anvifer2@upv.es

number of collocation points to the number of snapshots available. In this work, we have extended the LUPOD method to be able to select a larger number of collocation points, and we have tested the performance of this new method to obtain a reduced order model for the neutron diffusion equation.

2 Reduced order model

2.1 The neutron diffusion equation

We consider the two-energy groups without upscattering neutron diffusion equation, given by equation (1) where the operators are given by

$$\mathcal{L} = \begin{pmatrix} -\vec{\nabla}\cdot(D_1\vec{\nabla}) + \Sigma_{a_1} + \Sigma_{12} & 0 \\ -\Sigma_{12} & -\vec{\nabla}\cdot(D_2\vec{\nabla}) + \Sigma_{a_2} \end{pmatrix},$$

$$\mathcal{F} = \begin{pmatrix} \nu\Sigma_{f_1} & \nu\Sigma_{f_2} \\ 0 & 0 \end{pmatrix}, \qquad \Phi = \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix},$$

where Φ_1 and Φ_2 are the fast and thermal scalar neutron fluxes, and the remaining coefficients are the macroscopic cross sections, which are spatially dependent magnitudes. For the spatial discretization of this problem a finite element method is used [9] to get an algebraic problem

$$L\tilde{\Phi} = \frac{1}{\lambda}F\tilde{\Phi} \tag{2}$$

where L and M are matrices $N_d \times N_d$, being N_d the number of degrees of freedom of the problem. These matrices are obtained from the discretization of operators $\mathcal{L}$ and $\mathcal{M}$, respectively. $\tilde{\Phi}$ is the corresponding vector of coefficients of Φ. Problem (2), without further approximations, is known as the Full Order Model (FOM).

2.2 The Proper Orthogonal Decomposition

To construct a ROM using the proper orthogonal decomposition (POD), we will follow two approaches, one, which will be called the *monolithic approach* where it is supposed that $\tilde{\Phi}$ admits the following expansion

$$\tilde{\Phi} = \sum_{m=1}^{M} n_m \psi_m, \tag{3}$$

where ψ_m are orthonormal vectors, known as the POD modes, and n_m are their coefficients. M_0 is the number of modes considered.

Using expansion (3) in problem (2) and multiplying on the left by the orthogonal modes, a reduced eigenvalue problem is obtained with small matrices L_m and M_m,

$$\bar{L}\, n = \frac{1}{\lambda}\bar{F}\, n, \tag{4}$$

where the matrices' elements are given by

$$\bar{L}_{i,j} = \langle \psi_i, L\psi_j \rangle, \quad \bar{F}_{i,j} = \langle \psi_i, F\psi_j \rangle, \tag{5}$$

$n = (n_1, n_2, \ldots, n_M)^T$, and $\langle\, ,\, \rangle$ is the standard Euclidean scalar product. Problem (4) is a generalized eigenvalue problem of size $M \times M$, which is much smaller than problem (2), which is of size $N_d \times N_d$.

Another approach to obtain a reduced order model using the proper orthogonal decomposition consists of assuming the following expansions for the fast and thermal neutron fluxes,

$$\tilde{\Phi}_1 = \sum_{m=1}^{M} n_{1,m}\psi_{1,m}, \quad \tilde{\Phi}_2 = \sum_{m=1}^{M} n_{2,m}\psi_{2,m}. \tag{6}$$

Using these expansions in problem (2), and the block structure of the operators in terms of the energy groups, we obtain the equations

$$\begin{aligned} &\sum_{m=1}^{M} n_{1,m} L_{11}\,\psi_{1,m} = \frac{1}{\lambda}\left(\sum_{m=1}^{M} n_{1,m} F_{11}\,\psi_{1,m} + \sum_{m=1}^{M} n_{2,m} F_{12}\,\psi_{2,m}\right) \\ &\sum_{m=1}^{M} n_{1,m} L_{21}\,\psi_{1,m} + \sum_{m=1}^{M} n_{2,m} L_{22}\,\psi_{2,m} = 0\ . \end{aligned} \tag{7}$$

Multiplying on the left the first equation by the orthogonal modes $\psi_{1,m}$ and the second equation by the modes $\psi_{2,m}$ gives the small eigenvalue problem

$$\begin{pmatrix} \bar{L}_{11} & 0 \\ \bar{L}_{21} & \bar{L}_{22} \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \bar{F}_{11} & \bar{F}_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}\ , \tag{8}$$

where the matrices elements are

$$\begin{aligned} &\left(\bar{L}_{11}\right)_{i,j} = \langle \psi_{1,i}, L_{11}\,\psi_{1,j}\rangle, \quad \left(\bar{F}_{11}\right)_{i,j} = \langle \psi_{1,i}, F_{11}\,\psi_{1,j}\rangle, \\ &\left(\bar{F}_{12}\right)_{i,j} = \langle \psi_{1,i}, F_{12}\,\psi_{2,j}\rangle, \\ &\left(\bar{L}_{21}\right)_{i,j} = \langle \psi_{2,i}, L_{21}\psi_{1,j}\rangle, \quad \left(\bar{L}_{22}\right)_{i,j} = \langle \psi_{2,i}, L_{22}\,\psi_{2,j}\rangle. \end{aligned}$$

This approach will be called *group-wise* proper orthogonal decomposition. Note that in this way, each FOM snapshot is used twice, once for each energy group.

2.3 Generation of Snapshots

Starting from an initial configuration of the reactor core, given by a set of cross sections of the different materials, to obtain the orthogonal modes, in an off-line step, N_s perturbed configurations of the reactor are obtained using an Improved Distributed Hypercube Sampling (IHS)[1]. In the reactor benchmark analysis, there are 7 perturbed cross sections per material and 5 different materials, so the perturbed cross sections are 35 group constants that are modified $\pm 20\%$ using IHS. Solving the FOM (2) for each one of these configurations yields the matrix of snapshots, S, whose columns are the different neutron fluxes obtained. To obtain the orthogonal modes, a singular value decomposition [3] of matrix S is performed,

$$S = U\Sigma V^T\ , \tag{9}$$

identifying the M orthogonal modes ψ_i as the first M columns of matrix U. A similar approach can be followed for the group-wise POD.

2.4 LUPOD reduction

To construct a reduced order model associated with problems with many degrees of freedom, the Euclidean scalar product is used to construct the projected problem (see, for example, equation (5)). If the number of degrees of freedom, N_d, is very large, the use of the scalar product can be

expensive from the computational point of view. To avoid this problem, it is proposed to replace the Euclidean scalar product by a reduced scalar product,

$$\langle \varphi, \psi \rangle_r = \sum_{n=1}^{N} \varphi_{j_n} \psi_{j_n} \ , \tag{10}$$

for a selected set of indices $\{j_1, \ldots, j_N\}$ that defines a set of collocation points corresponding to different degrees of freedom of the problem. To select both a convenient set of collocation points and a subset of dominant snapshots, the LUPOD method is used [7, 6]. This method mainly performs a Gaussian elimination of the snapshots matrix with pivoting of rows and columns, associating the different collocation points with the rows and the dominant snapshots with the columns used in the Gaussian elimination process. This method is outlined in Algorithm 1 and Algorithm 2.

Algorithm 1 LUPOD Algorithm

Require: S (Snapshot matrix of size J, K), ε_N (threshold for collocation points).
1: $\hat{S} \leftarrow S$
2: $snaps \leftarrow \{\}$
3: $points \leftarrow \{\}$
4: $n \leftarrow 1$
5: **while** $\| \hat{S} \|_{\text{Fro}} / \| S \|_{\text{Fro}} > \varepsilon_N$ **do**
6: $\quad \{j_n, \ k_n\} \leftarrow \max(\max(|\hat{S}|))$ ▷ Find the position of the maximum
7: $\quad \hat{S} \leftarrow \text{GaussElimination}(\hat{S}, j_n, k_n)$
8: $\quad points[n] \leftarrow j_n$
9: $\quad snaps[n] \leftarrow k_n$
10: $\quad n \leftarrow n + 1$
11: **end while**
12: **return** points, snaps
13: $\hat{S}[:, k_n] \leftarrow 0$

Algorithm 2 Gauss Elimination Algorithm

Require: $\hat{S}$ (Matrix of size J, K), j_n and k_n (pivot position).
1: $pivot \leftarrow \hat{S}[j_n, k_n]$
2: $\hat{S}[:, k_n] \leftarrow \hat{S}[:, k_n]/pivot$
3: **for** $k = 1$ to K **do**
4: $\quad$ **if** $k \neq k_n$ **then**
5: $\quad\quad \hat{S}[:, k] \leftarrow \hat{S}[:, k] - \hat{S}[j_n, k] \ \hat{S}[:, k_n]$
6: $\quad$ **end if**
7: **end for**
8: **return** $\hat{S}$

The LUPOD method as presented in [7, 6] assumes that the number of collocation points is similar to the number of snapshots, but to obtain a better precision of the reduced order model, it is interesting to use a number of collocation points, N, larger than the number of snapshots available, whose computation is very expensive. We have extended the LUPOD method for the case when the number of collocation points N is quite larger than the number of selected snapshots, K, ($N \gg K$). As a first step, the extended LUPOD starts applying the standard LUPOD to the matrix S to select K collocation points and K snapshots. As a second step, the dominant snapshots are fixed (ordering of columns of S), and the rows corresponding to the selected collocation points

Algorithm 3 LUPOD Extended Algorithm

Require: S (Snapshot matrix of size J, K), N (Requested number of collocation points).

$\hat{S} \leftarrow S$
$snaps \leftarrow \{\}$
$points \leftarrow \{\}$
$n \leftarrow 0$
for $n < N$ **do** ▷ Standard LUPOD algorithm
  $\{j_n,\ k_n\} \leftarrow \max(\max(|\hat{S}|))$ ▷ Find the position of the maximum
  $\hat{S} \leftarrow \text{GaussElimination}(\hat{S}, j_n, k_n)$
  $points[n] \leftarrow j_n$
  $snaps[n] \leftarrow k_n$
  $n \leftarrow n + 1$
end for
$b \leftarrow 1$
for $b < \text{ceil}(N/K)$ **do** ▷ Repeat LUPOD algorithm until N = length(points)
  $\hat{S} \leftarrow S$
  $\hat{S}(points, :) \leftarrow 0$
  for $n < N$ **do**
    $\{j_n,\ k_n\} \leftarrow \max(\max(|\hat{S}|))$
    $\hat{S} \leftarrow \text{GaussElimination}(\hat{S}, j_n, k_n)$
    $points[n] \leftarrow j_n$
    $n \leftarrow n + 1$
  end for
  $b \leftarrow b + 1$
end for
return points, snaps

are eliminated from the matrix S. Next, the LUPOD algorithm is applied to the new matrix S, but only considering the Gauss method with partial pivoting for the rows. This allows the selection of a new set of K collocation points. This process is repeated until the number of desired collocation points is reached. The extended LUPOD algorithm is outlined in Algorithm 3.

3 Numerical Results

The tridimensional Langenbuch reactor [5] is selected to test the methodology proposed. Figure 1 shows the geometry of the reactor model in the steady state, where each number represents a homogenized material. The full description of the reactor can be found in [5]. The calculations for the reactor have been performed with a whole core model, using 1170 cells and finite elements of degree 2. Thus, the FOM has 21546 degrees of freedom split into two energy groups of 10773 degrees each. This FOM has been used as a reference solved with sparse linear algebra [9].

The ROM methodologies are tested using $N_t = 50$ random configurations that are not included in the training set, and computing the following eigenvalue and eigenvector errors for each test

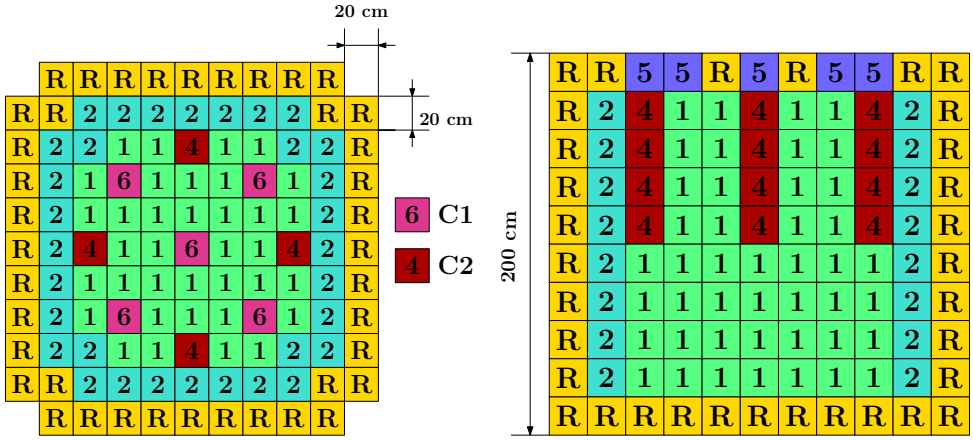
$$\Delta k_{\text{eff}} = |k_{\text{eff,ROM}_i} - k_{\text{eff,FOM}_i}| \ , \tag{11}$$

$$\text{RMS}(\Delta\phi) = \frac{\|\phi_{\text{ROM}_i} - \phi_{\text{FOM}_i}\|}{\|\phi_{\text{FOM}_i}\|} \ . \tag{12}$$

Figure 2 shows the eigenvalue error Δk_{eff}, and the eigenvector error, $\text{RMS}(\Delta\phi)$ depending on the number of snapshots used with the group-wise methodology. This graph shows an exponential

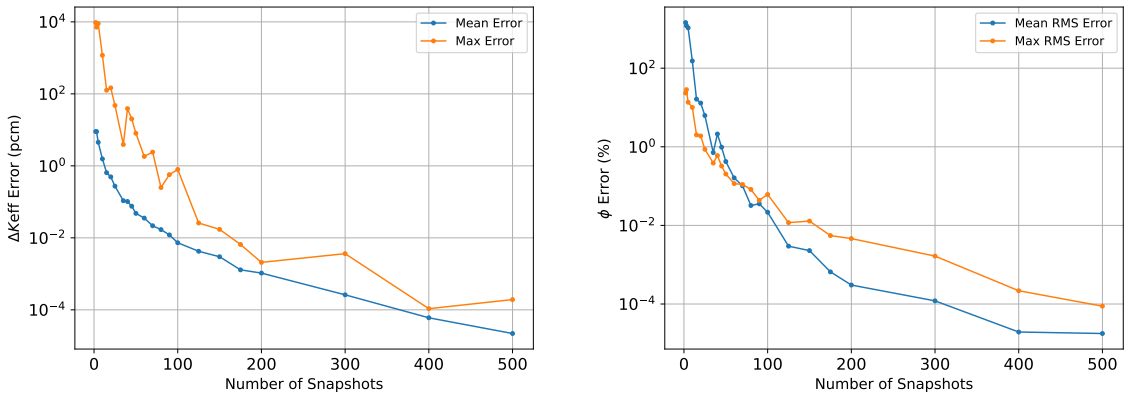
reduction of error with the number of snapshots.

Table 1 shows the average results for the monolithic approach, where it can be seen that inaccurate results are obtained. The appearance of spurious eigenvalues, complex eigenvalues that have no physical meaning due to the approximations, spoils the results. Table 2 presents the average group-wise results that are much more accurate by using each snapshot twice (once per energy group). Simple POD presents, with 20 snapshots, speed-up factors of 3 while maintaining an average error in the eigenvalue of 12.9 pcm and an error in the eigenvector of RMS($\Delta\phi$) = 0.49%, Extended LUPOD results that retain 40% with 20 snapshots of collocation points show a speed-up factor of 8.5 while maintaining an average error in the eigenvalue of 21.3 pcm and an error in the eigenvector of RMS($\Delta\phi$) = 0.47%.



(a) Radial view (b) Axial profile.

Figure 1: Geometry of the Langenbuch reactor.



(a) Eigenvalue Error (pcm) (b) Eigenvector Error (%)

Figure 2: Maximum and average eigenvalue, and eigenvector errors depending on the number of snapshots using POD for the Langenbuch reactor.

Table 1: Average of the results for monolithic ROM.

# of snapshots	Δk_{eff} (pcm)	RMS ($\Delta\phi$) (%)	Speed-Up
POD - Monolithic			
10*	1342.2	8.93	5.6
20*	1800.3	19.29	4.3
40	12.5	0.71	3.0
LUPOD - Monolithic			
10*	18299.0	33.08	100.4
20*	3034.9	23.02	87.6
40*	3721.4	22.93	56.0
LUPOD Extended - 40 % of Collocation Points - Monolithic			
10*	1948.3	10.28	14.2
20*	4305.6	28.02	12.4
40*	134.6	4.35	8.4

* Some test present spurious complex eigenvalues.

Table 2: Average of the results for group-wise ROM.

# of snapshots	(Δk_{eff}) (pcm)	RMS ($\Delta\phi$) (%)	Speed-Up
POD - Group Wise			
10	152.8	1.55	4.3
20	12.9	0.49	3.0
40	2.1	0.10	1.7
LUPOD - Group Wise			
10*	954.0	4.23	87.6
20*	1037.3	7.21	56.0
40*	761.4	6.01	28.9
LUPOD Extended - 40 % of Collocation Points - Group Wise			
10	133.5	1.31	12.5
20	21.3	0.47	8.5
40	9.9	0.12	4.9

* Some test present spurious complex eigenvalues.

4 Conclusions

This work proposes an extended LUPOD algorithm to construct a reduced-order model to solve an eigenvalue problem related to the neutron diffusion equation, which can be used for uncertainty quantification. The extended LUPOD technique uses the dominant orthogonal modes and an arbitrary number of collocation points to define a reduced scalar product. This way, accurate results are found while achieving significant speed-ups. Using a three-dimensional reactor problem, the numerical results show that using a basis dimension of 20 vectors significantly speeds up the solution while maintaining accurate results that differ less than 0.5 % in the eigenvector reference solution. In the extended LUPOD algorithm, at least 40 % of the rows must be used to maintain the accuracy in the reduced model. In these experiments, speed-ups range from 3 in standard POD to 8.4 with extended LUPOD.

Acknowledgments

This work has been partially funded with "Ayuda a Primeros Proyectos de Investigación (PAID-06-24)", Vicerrectorado de Investigación of the Universitat Politècnica de València (UPV) and project PID2023-152604OB-I00 by MICIU/AEI /10.13039/501100011033 and FEDER, EU.

References

[1] Beachkofski, B., Grandhi, R., Improved distributed hypercube sampling. *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, American Institute of Aeronautics and Astronautics, Apr. 2002.

[2] German, P., Ragusa, J. C., Reduced-order modeling of parameterized multi-group diffusion k-eigenvalue problems. *Annals of Nuclear Energy*, 134:144–157, 2019.

[3] Golub, G. H., Van Loan, C. F., Matrix Computations. Baltimore, MD, JHU Press, 2013.

[4] Henry, A., Nuclear-Reactor Analysis. Cambridge, MA, MIT Press, 1975.

[5] Langenbuch, S., Maurer, W., Werner, W., Coarse-mesh flux-expansion method for the analysis of space-time effects in large light water reactor cores. *Nuclear Science and Engineering*, 63(4):437–456, 1977.

[6] Rapún, M.-L., Terragni, F., Vega, J., Adaptive sampling and modal expansions in pattern-forming systems. *Advances in Computational Mathematics*, 47(4):48, 2021.

[7] Rapún, M.-L., Terragni, F., Vega, J. M., LUPOD: Collocation in POD via LU decomposition. *Journal of Computational Physics*, 335:1–20, 2017.

[8] Varvayanni, M., Savva, P., Catsaros, N., Control rod worth calculations using deterministic and stochastic methods. *Annals of Nuclear Energy*, 36(11–12):1718–1725, 2009.

[9] Vidal-Ferràndiz, A., Fayez, R., Ginestar, D., Verdú, G., Solution of the lambda modes problem of a nuclear power reactor using an h–p finite element method. *Annals of Nuclear Energy*, 72:338–349, 2014.

[10] Vinai, P., Macian-Juan, R., Chawla, R., A statistical methodology for quantification of uncertainty in best estimate code physical models. *Annals of Nuclear Energy*, 34(8):628–640, Aug. 2007.

Part IV

Infectious Diseases and Mathematical Modeling in Public Health

Extending the SIS Epidemic Markov Model Including Coinfection Dynamics

A. Carlos Andreu-Vilarroig[♭],[1] B. Sten Madec[♮] and C. Erida Gjini[□]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València,
Camí de Vera s/n, 46020 València, Spain.
(♮) University of Tours, 60 Rue du Plat d'Étain, 37000 Tours, France
(□) Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

1 Introduction

Classical mathematical epidemic models, such as the susceptible-infected-susceptible (SIS) framework, have long been used to understand disease transmission in closed populations. These models capture the essential dynamics of infections that do not confer long-term immunity, allowing individuals to return to the susceptible class after infection. However, empirical evidence in certain infections increasingly highlights the relevance of coinfection, i.e., the possibility that an individual can be infected by multiple strains at the same time. When coinfection occurs, the infection dynamics changes substantially, requiring more extensive models to explain the epidemiological process behavior [1].

In the mathematical modelling field, while deterministic models based on ordinary differential equations (ODEs) have traditionally been used to study epidemic dynamics, stochastic models, such as Markov chains, offer several advantages. In contrast to deterministic models, which analyze the behavior of the system on average, stochastic models capture and analyze the uncertainty inherent in the infection processes. Moreover, Markov chains consider discrete populations, which is more realistic in epidemiology. In this area, an extensive research has been carried out translating classical deterministic models into Markov models [2, 3].

In this paper, we go one step further by extending the Markov SIS model by incorporating coinfection in a continuous-time Markov chain (CTMC) framework. Our goal is to obtain the full mathematical model and analyze its equilibrium by deriving the stationary and quasi-stationary distribution of the model states.

2 Mathematical model

The CTMC mathematical model we have designed considers a closed and finite population with $1, \ldots, N$ individuals, where each individual can be in one of three states: S (susceptible), I (single infected) or D (double infected or coinfected). The population state is fully described by the random vector $X = (I, D) \in \Omega$, where the finite state space is defined as $\Omega = \{x = (i, d) \in \mathbb{N}^2 : \ i + d \leq N\}$. Given I and D, the number of susceptible individuals can be computed as $S = N - I - D$.

[1] caranvi1@upv.es

At each time step Δt, the possible transitions between the states that can happen are the following:

Transition	**States variation**	**Rate**
$S \to I$	$(i,d) \to (i+1,d)$	$\alpha_1(i,d) = \frac{\beta}{N}(i+d)(N-i-d)$
$I \to D$	$(i,d) \to (i-1,d+1)$	$\alpha_2(i,d) = k\frac{\beta}{N}i(i+d)$
$I \to S$	$(i,d) \to (i-1,d)$	mi
$D \to I$	$(i,d) \to (i,d-1)$	md
Nothing	$(i,d) \to (i,d)$	$1 - \left[\frac{\beta}{N}(i+d)(N-d-i+ki) + m(i+d)\right]$

Table 1: Possible transitions at each time step and their associated rate.

Note that, in Markov models, only one individual change is possible at each time step. Note also that the ratios expressing the probabilities of transitions between states are identical to the rates of the differential equations of the equivalent deterministic compartmental model, which is given by

$$\begin{cases} \dot{S} = -\frac{\beta}{N}S(I+D) + m(1-S), \\ \dot{I} = \frac{\beta}{N}S(I+D) - mI - k\frac{\beta}{N}I(I+D), \\ \dot{D} = k\frac{\beta}{N}I(I+D) - mD. \end{cases}$$

As it can be seen, the parameters of this model are: the infection rate $\beta \geq 0$, the clearance rate $m \geq 0$, and the coinfection attenuation factor $k \in (0,1)$.

3 Results

Considering that the model is a stochastic process, it will be completely defined if the probability distribution of all its states at each instant of time is provided. Let $p(i,d)(t) = \mathbb{P}\{(I(t),D(t)) = (i,d)\}$ denote the probability of being in state $(i,d) \in \Omega$ at time t. By the Markov chains theory, we can conclude that the temporal evolution of this probability distribution is given by the so-called forward Kolmogorov equations system:

$$\begin{aligned} \frac{dp_{(i,d)}(t)}{dt} &= p_{(i-1,d)}(t)\alpha_1(i-1,d) + p_{(i+1,d)}(t)m(i+1) + p_{(i+1,d-1)}(t)\alpha_2(i+1,d-1) \\ &\quad + p_{(i,d+1)}(t)m(d+1) - p_{(i,d)}(t)\left(\alpha_1(i,d) + \alpha_2(i,d) + m(i+d)\right), \quad \forall (i,d) \in \Omega, \\ \frac{dp_{(0,0)}(t)}{dt} &= p_{(0,1)}(t)m + p_{(1,0)}(t)m, \end{aligned}$$

At this point, we can express this linear system in its matricial form as

$$\mathbf{p}' = Q\mathbf{p},$$

where Q – the generator matrix – is a block matrix that can be structured as a tridiagonal block matrix

$$Q = \begin{pmatrix} C_0 & U_0 & \mathbf{0} & \dots & \mathbf{0} \\ L_1 & C_1 & U_1 & \dots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \dots & L_{N_p-1} & C_{N_p-1} & U_{N_p-1} \\ \mathbf{0} & \dots & \mathbf{0} & L_{N_p} & C_{N_p} \end{pmatrix}$$

(each non-null block contains a set of transition rates), and the probability distribution vector is defined as

$$\mathbf{p} = \mathbf{p}(t) = (\mathbf{p}_0(t), \ldots, \mathbf{p}_N(t))^T,$$

where

$$\mathbf{p}_a = \mathbf{p}_a(t) = \left(p_{(a,0)}(t), p_{(a-1,1)}(t), \ldots, p_{(1,a-1)}(t), p_{(0,a)}(t)\right)^T \in \mathbb{R}_{a+1}$$

is the vector that contains the probabilities of observe (at time t) certain number of total infected $i + d = a,\ a = 0, \ldots, N$.

Knowing that Q is Metzler, and that the first column have all entries equal to 0, it can be proved that 0 is an eigenvalue and that the rest of the eigenvalues have negative real part. Thus, it can be proved that the stationary distribution to which the model tends is given by $\mathbf{p}^* = \lim_{t\to\infty} \mathbf{p}(t) = (1, 0, \ldots, 0)^T$. This means that, at long term, the infection tends to the disease-free equilibrium point: everyone susceptible and no one infected or coinfected. In other words, the disease is extinguished.

Nevertheless, previous to the extinction time, the random vector $X = (I, D)$ seem to achieve an endemic equilibrium: the so-called *quasistationary distribution* (QSD). Since extinction is inevitable in finite populations, the study of the QSD is fundamental. The QSD is defined as the probability of being in a certain state at time t conditional on non-extinguishment, or formally,

$$q_{(i,d)}(t) = \mathbb{P}\{(I(t + \Delta t), D(t + \Delta t)) = (i, d) | (I(t), D(t)) \neq (0, 0)\} = \frac{p_{(i,d)}(t)}{1 - p_{(0,0)}(t)}.$$

By differentiating $q_{(i,d)}(t)$ expression and applying the chain rule along with Kolmogorov's equations, we derive the equation

$$\mathbf{q}' = \tilde{Q}\mathbf{q} + m\mathbf{1}^T\mathbf{q}_1\mathbf{q},$$

where

$$\mathbf{q} = \mathbf{q}(t) = (\mathbf{q}_1, \ldots, \mathbf{q}_N) = \frac{(\mathbf{p}_1, \ldots, \mathbf{p}_N)^T}{1 - p_{(0,0)(t)}}$$

is the conditional probabilities vector, and $\tilde{Q}$ is the reduced generator matrix (excluding the first row and column). The QSD is the fixed point q^* satisfying

$$\mathbf{0} = \tilde{Q}\mathbf{q} + m\mathbf{1}^T\mathbf{q}_1\mathbf{q}.$$

The QDS can be interpreted as, under certain epidemic parameters (a sufficiently large N and $R_0 = \beta/m > 1$), the distribution quickly converges to $p(t) \simeq (0, \mathbf{q}^*)$ in a large time interval before extinction. Its existence and uniqueness derives from the fact that the Markov model is irreducible, i.e., after eliminating the absorbant state $(0, 0)$, from any state (i_1, d_1), any other (i_2, d_2) can be reached.

Finally, although the QSD is the solution of a nonlinear system, and it can only be computed numerically, we can obtain a good approximation by knowing that (for large N and $R_0 > 1$), the stochastic process

$$\sqrt{N}(\mathbf{x}(t) - \mathbf{x}^*)$$

can be approximated by a bivariate Ornstein-Uhlenbeck process, where

$$\mathbf{x}(t) = (x_1, x_2) = \left(\frac{I}{N}, \frac{D}{N}\right)$$

is the normalized vector of the infected and coinfected variables, and

$$\mathbf{x}^* = \left(\frac{1 - 1/R_0}{1 + k(R_0 - 1)}, \frac{k(R_0 - 1)(1 - 1/R_0)}{1 + k(R_0 - 1)} \right)$$

is the normalized endemic equilibrium point, which can be obtained from the differential equations system of the normalized deterministic compartmental model (considering $N = 1$).

The QSD is approximated by the stationary distribution of the Ornstein-Uhlenbeck process, which is given by a bivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$. The covariance matrix Σ have been computed from the local covariance matrix and the local drift matrix of the Ornstein-Uhlenbeck process. Therefore, we conclude that

$$X(t) = N\mathbf{x}(t) \xrightarrow{t\to\infty} \mathbf{q}^* = \mathcal{N}\left(N\mathbf{x}^*, \frac{1}{\sqrt{N}}\Sigma \right).$$

so that, under the mentioned assumptions, the QSD of the infected and coinfected population is completely characterised, as well as relevant aspects such as its expected value (which coincides with the endemic point of the deterministic model) and its variance (given by matrix Σ).

4 Conclusions

In this work, we have developed a extension of the SIS Markov model incorporating the coinfection phenomenon. Our theoretical analysis has succeeded in defining the structure of the generator matrix of the Markov model, demonstrating the stationary equilibrium from its structure, proving the existence and uniqueness of a previous quasi-stationary equilibrium, and approximating its distribution under certain conditions. As a future line, this framework offers the basis for future studies on n-strain infections in finite populations, by subdividing the infected I and double infected D classes into n subclasses, one for each strain.

Acknowledgments

This work has been funded by the Generalitat Valenciana ACIF program under grant CIACIF/2022/203.

References

[1] Thi Minh Thao Le, Erida Gjini, and Sten Madec. Quasi-neutral dynamics in a coinfection system with n strains and asymmetries along multiple traits. *Journal of Mathematical Biology*, 87(3):48, 2023.

[2] Linda JS Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, 2008.

[3] Ingemar Nåsell. Stochastic models of some endemic infections. *Mathematical Biosciences*, 179(1):1–19, 2002.

A Computational Framework for Profiling a Human Bladder Cancer scRNA-seq Sample

C. Lara[♭],[1] G. Rubio[♭], C. Santamaría[♭] and B. García-Mora[♭]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

Bladder cancer is the most common malignancy of the urinary tract, and it is also one of the most prevalent cancers worldwide [1]. It is characterized by high mutational burden and heterogeneity at the genomic, transcriptional, and cellular levels [2]. This presents a challenge, as it hampers our ability to understand tumor development and progression, thus complicating the identification of effective and personalized treatment strategies [3].

Traditional approaches to understand and treat cancer focus on histological analysis and molecular studies [4], but this limits our ability to detect key molecular drivers of the disease. In recent years, however, single-cell RNA sequencing (scRNA-seq) has emerged as a transformative technology that offers great insight into tumor heterogeneity, cellular trajectories and tumor microenvironment. Making use of this potential, a preliminary analytical pipeline was developed to extract meaningful information from a scRNA-seq dataset of a human bladder cancer sample. This approach integrates a cohort of computational methods, such as unsupervised clustering, CNV inference or trajectory analysis; providing a a framework for profiling tumor cell states and their potential progression paths.

This flexible and reproducible pipeline offers a step forward toward better characterizing tumor behaviour, leading to more personalized - and therefore, more effective - treatment strategies.

2 Methods

2.1 Data Acquisition and Preprocessing

The single-cell RNA sequencing (scRNA-seq) analysis of the human bladder cancer sample was performed by the company *10X Genomics*, who provided the necessary gene expression matrix data in a easy-to-process format.

The data was then exported into R (version 4.4.1) software, and the Seurat (version v5) package was used for the subsequent analysis. The first step was filtering out low quality cells. Looking at the statistical distribution of the sample data, thresholds were defined for the acceptable number of unique feature counts (genes) and mitochondrial content. Cells with < 200 or > 4000 genes or $> 10\%$ UMIs derived from mitochondria were considered low-quality and were therefore filtered.

Then, the count data was normalized using the NormalizeData function. In preparation to run the Principal Component Analysis (PCA), the 2,000 most variable genes - which are more

[1] clarpin@upv.edu.es

likely to capture biologically meaningful information - were selected using the FindVariableFeatures function, and the ScaleData function was also used to standarize the expression data of all genes.

Subsequently, PCA dimensionality reduction was performed with the RunPCA function. Lastly, for the clustering process, the functions FindNeighbors (dims = 1:20) and FindClusters (resolution = 0.4) were used. 10 different clusters were obtained, and the results were visualized with the RunUMAP function.

2.2 Cell annotation

To identify the cell types present in the sample, an initial automated annotation was performed using the SingleR (version 2.8.0) package, using the Human Primary Cell Atlas as reference. Following this, only epithelial cells were kept for further analysis.

Subsequently, a manual refinement of epithelial subtypes was carried out based on literature-informed biological expectations and known marker gene expression profiles. The epithelial cells were then classified into Basal/Intermediate (KRT5, KRT17, KRT13, KRT14, TP63, SHH) or Umbrella (UPK1A, UPK2, UPK3A, KRT20, TJP1, CLDN4, CLDN8, CLDN12, OCLN).

2.3 CNV analysis

Copy number variation (CNV) inference was performed to identify potential chromosomal amplifications or deletions using the R package infercnv (version 1.3.3) with a cutoff of 0.1, ideal for 10X Genomics datasets. This method assumes that tumor cells often show large values of CNVs, whereas normal cells generally maintain a stable genome, and therefore show lower CNV values.

In this particular case, since we didn't have any reference healthy cells, the *null* refernce approach was used, in which expression levels are compared to a computed baseline across all cells. Although no the most optimal option, it provides good results if the sample shows enough heterogeneity; but, ideally, a reference should be used for better comparison.

A CNV score was computed per each cell, quantifying the extent to which it deviated from baseline expression, and then these results were used to calculate the average CNV score per cluster.

Finally, the results were visualized on a histogram plot and a cut-off point of $n = 0.002$ was defined. Three clusters were then defined as malignant and subset to further analysis.

2.4 Trajectory inference

To perform trajectory inference, the R package Monocle3 (version 1.2.7) was used. Monocle3 infers cellular trajectories from scRNA-seq data by arranging the cells in a *pseudotime* order that reflects the underlying biological progression. Pseudotime is not based on actual cronological time, but rather on transcriptional similarity among cells. The underlying idea behind it is that in scRNA-seq data, we can see a snapshot of thousands of individual cells at one point in time, but within that snapshot, different cells will be in different states; and the transition between states is what pseudotime tries to reconstruct.

To begin, a Monocle object was created from the previously obtained Seurat object, only keeping the (suspected-to-be) tumor cells. The data was then preprocessed using the preprocess_cds function (with num_dim = 30). PCA dimensionality reduction was performed with the reduce_dimesion function, and the subsequent clustering, with the cluster_cells function.

Lastly, the trajectory inference was constructed using the functions learn_graph and order_cells.

3 Results

The cells were grouped into 10 different clusters, and Figure 1 shows their UMAP representation:

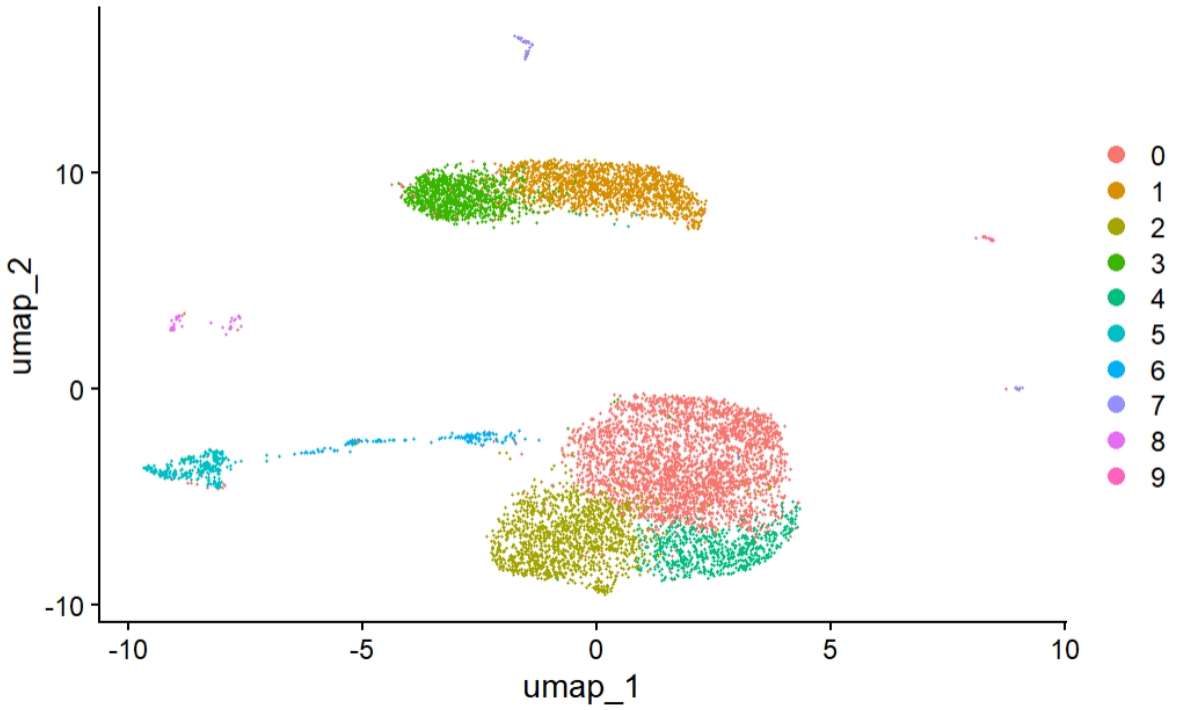


Figure 1: UMAP visualization of the different clusters, res = 0.4

The automatic annotation process showed that most of the cells in the sample were epithelial cells, as expected, but other cell types were also found, such as endothelial cells, fibroblasts or immune cells. These, however, were filtered out, since the focus of the analysis was on the epithelial cells.

Following the automatic annotation process, as was discussed in the previous section, a manual annotation was performed to distinguish between the different epithelial subtypes that were expected to be found in the sample, given its origin.

Table 1 shows the amount of cells of each category present in each one of the nine clusters:

	Basal/Intermediate	Umbrella
0	2907	55
1	1379	114
2	1134	288
3	1011	66
4	597	64
5	93	190
6	140	18
7	42	1
8	34	3
9	13	1

Table 1: Cell types present in each cluster

Therefore, all clusters were considered to be basal/intermediate except for cluster number 5, which had a higher amount of umbrella cells.

Copy number variation (CNV) inference was then performed, and the individual CNV scores calculated per cell were used to calculate the average CNV score per cluster. Higher CNV scores are related to a higher probability of the cells being malignant. Figure 2 shows the histogram plot of the CNV values per cluster:

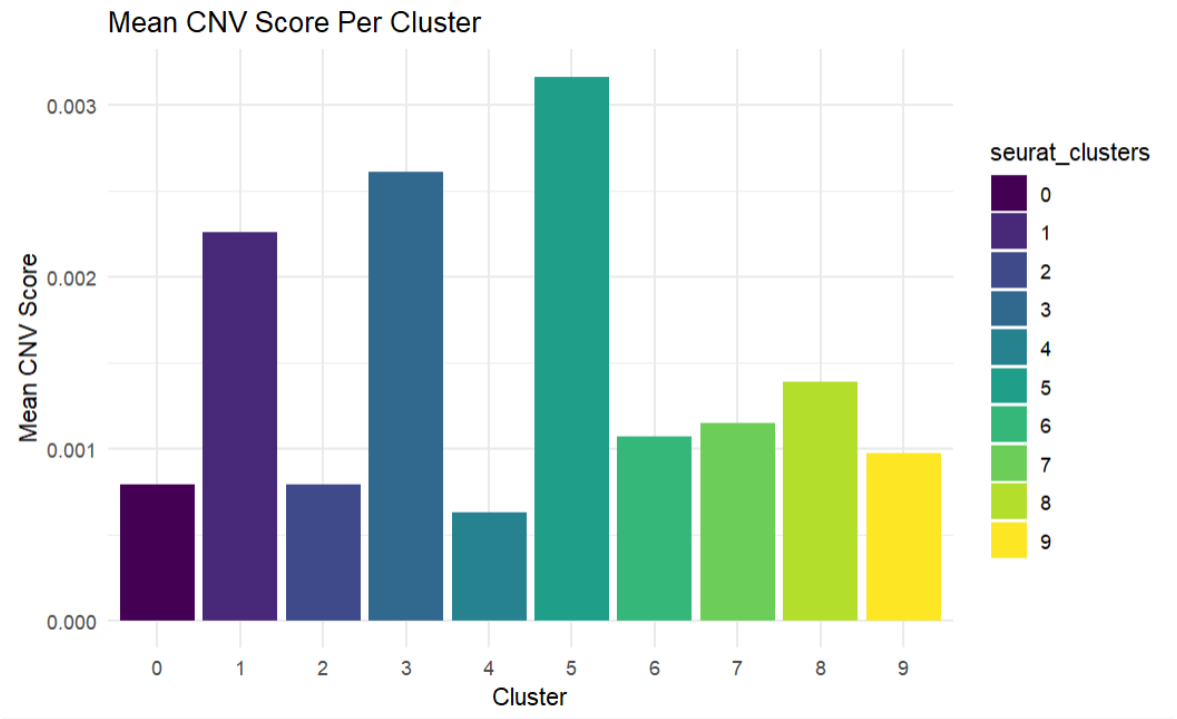


Figure 2: Histogram plot of the CNV scores per cluster

Looking at the histogram plot, a cutoff value of 0.002 was determined, and clusters 1, 3 and 5 were considered malignant. Interestingly, cluster number 5, which was the only "Umbrella" cluster, was also the one with the higher CNV score.

Finally, a subset containing the malignant clusters was created, and then it was converted to a Monocle object to perform trajectory analysis.

A limited factor encountered during the trajectory inference process was that after the necessary preprocessing conducted using Monocle3, the dataset was still fragmented into three distinct clusters. Since the algorithm estimates trajectories by learning a principal graph through the manifold of the cell state space, it requires the cells to form a connected structure. Our cells, however, are distributed in three different clusters, and Monocle3 cannot therefore infer a unified trajectory, limiting the interpretability of pseudotime dynamics across the entire sample.

It remains unclear whether the observed clusters represent distinct tumor subpopulations or reflect different transcriptional states within a common lineage. Further analysis, such as Gene Set Enrichment Analysis (GSEA), are necessary and will be conducted to disambiguate these possibilities and provide a clearer insight of the tumor's evolution.

References

[1] Dobruch, J., & Oszczudłowski, M. (2021). Bladder Cancer: Current Challenges and Future Directions. *Medicina*, 57(8), 749. https://doi.org/10.3390/medicina57080749

[2] Luo, Y., Tao, T., Tao, R., Huang, G., & Wu, S. (2022). Single-Cell Transcriptome Comparison of Bladder Cancer Reveals Its Ecosystem. Frontiers In Oncology, 12. https://doi.org/10.3389/fonc.2022.818147

[3] Meeks, J. J., Al-Ahmadie, H., Faltas, B. M., Taylor, J. A., Flaig, T. W., DeGraff, D. J., Christensen, E., Woolbright, B. L., McConkey, D. J., & Dyrskjøt, L. (2020). Genomic heterogeneity in bladder cancer: challenges and possible solutions to improve outcomes. Nature Reviews Urology, 17(5), 259-270. https://doi.org/10.1038/s41585-020-0304-1

[4] Liu, H., Li, Y., Karsidag, M., Tu, T., & Wang, P. (2024). Technical and Biological Biases in Bulk Transcriptomic Data Mining for Cancer Research. Journal Of Cancer, 16(1), 34-43. https://doi.org/10.7150/jca.100922

Analysis of RSV Hospitalizations in Infants Under 1 Year: Application of Hybrid Digital Twins

Marcos Llamazares[♭],[1] and Rafael-J. Villanueva[♭]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction and Motivation

Respiratory Syncytial Virus (RSV) [1] causes infections in the respiratory tract, primarily affecting infants, the elderly, and immunocompromised individuals. Its high transmissibility and seasonal nature make it a recurrent challenge for healthcare systems, especially during colder months. In particular, RSV is associated with a high incidence of bronchiolitis in children under one year of age, potentially leading to severe complications and hospitalizations.

In this work, we employ a model based on a system of difference equations (SDE) to describe the temporal evolution of RSV-related hospitalizations. This type of model was selected due to its low computational cost and fast execution times, which are especially advantageous when performing repeated simulations or calibrating model parameters. Unlike more complex simulation-based approaches, difference equations offer a balance between interpretability, mathematical rigor, and computational efficiency.

A common strategy in seasonal epidemiological models is to modulate the transmission rate using periodic functions. In our case, we adopt a cosine term to represent the seasonality in the transmission coefficient. However, this formulation fails to accurately reproduce the sharp hospitalization peaks observed in real data. Capturing these peaks, along with accurately predicting the annual average number of hospitalizations, is essential for effective resource planning and public health decision-making.

To address this limitation, we introduce a Hybrid Digital Twin (HDT) framework [3], which combines the mechanistic model based on difference equations with a data-driven correction using artificial intelligence techniques. Specifically, we apply Grammatical Evolution with lexicase selection [6], which learns an explicit corrective function to reduce the discrepancy between the model's output and observed hospitalization data. This hybrid approach significantly improves the ability to capture peak dynamics without increasing the model's complexity and sacrificing interpretability.

One of the key advantages of this methodology is that it maintains the computational efficiency of the base model. Grammatical Evolution produces an explicit functional expression, which can be evaluated instantly once learned. As a result, the HDT-enhanced model retains the fast execution

[1] marllalo@upvnet.upv.es

times of the original difference equations model. This makes the approach not only more accurate in reproducing the peaks and make predictions, but also highly suitable for large-scale applications.

2 Data

This study utilizes hospitalization data for infants under one year of age from the Comunitat Valenciana (CVA), specifically focusing on cases of bronchiolitis, recorded weekly between January 3, 2011 (the first week of 2011) and December 27, 2015 (the final week of 2015) [2]. The dataset spans 260 weeks, covering a period of five years. For analysis purposes, the data is divided into two age categories: infants aged 0 to 5 months, and those between 6 and 12 months. As depicted in Figure 1, a noticeable variation exists between the hospitalization rates of the two age groups. On average, 91.65% of all infants under one year who are hospitalized annually belong to the 0-5 month age group. These findings align with previous studies suggesting that hospitalization rates due to RSV are significantly higher among younger infants, with a decline as they grow older [7].

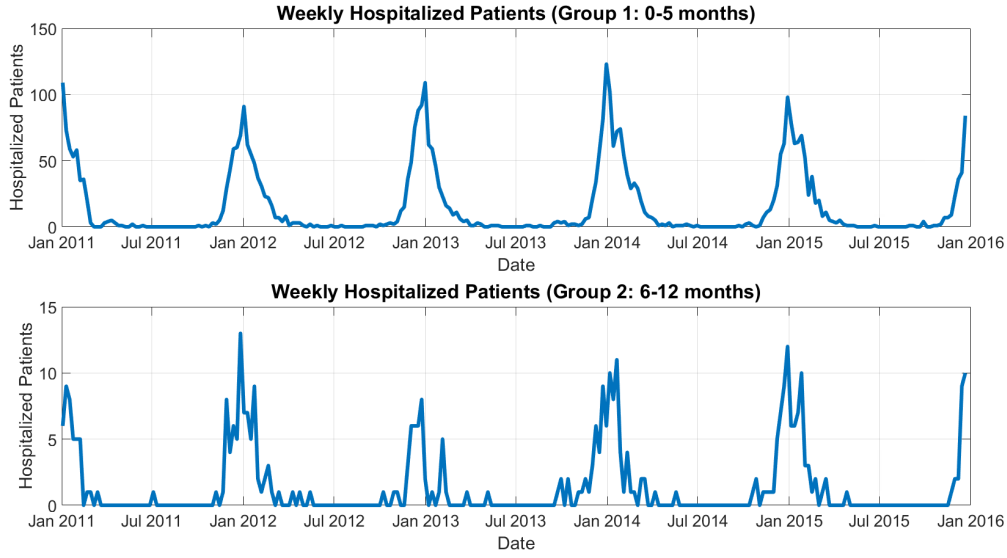


Figure 1: Infants aged 0 to 5 months and 6 to 12 months in the Valencian Community (CVA) who were hospitalized over a 260-week period, from week 1 of 2011 to week 52 of 2015.

3 Methods

3.1 Model

The model used in this study is an SLIRS model and incorporates several additional compartments to better represent the progression of the disease. Based on the information available in [1], the population is divided into five subpopulations: Susceptible (S), Latent (L), Infected (I), Hospitalized (H), and Recovered (R). Susceptible individuals are healthy and at risk of contracting the disease. Latent individuals are infected but not yet contagious. It is important to clarify that, in this model, a latent individual is defined as non-contagious regardless of symptoms, differing from the usual medical definition, where a latent case typically refers to an asymptomatic but

infected person. This distinction is relevant insofar as, for the model, we separate those who infect and those who do not and certain asymptomatic individuals may still transmit the disease, whereas some symptomatic ones might not. Infected individuals (I) are those capable of spreading the disease, while Hospitalized individuals (H) are those whose condition requires hospitalization. Specifically, parameters p_1 and p_2 represent the percentage of Infected individuals from age groups 1 and 2, respectively, who progress to Hospitalized. Recovered individuals (R) have overcome the infection and experience a temporary immunity before becoming susceptible again.

Due to the nature of the hospitalization data, the total population is further stratified into three age groups as described in [1]: Group 1 (G_1) includes infants aged 0 to 5 months, Group 2 (G_2) includes children aged 6 to 12 months, and Group 3 (G_3) comprises individuals older than 1 year. Each age group contains all five epidemiological subpopulations. The model also incorporates demographic processes such as births, transitions between age groups as individuals grow, and natural mortality. Figure 2 shows a flowchart illustrating the transmission dynamics of the disease.

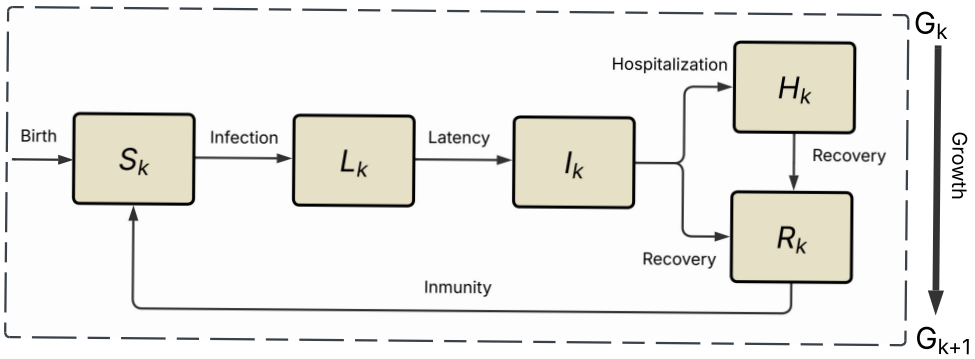


Figure 2: Flowchart illustrating the transmission dynamics and interactions between different compartments of the SLIRS model.

Subpopulation vectors are defined as:

$$\mathbf{S}(t) = \begin{bmatrix} S_1(t) \\ S_2(t) \\ S_3(t) \end{bmatrix}, \quad \mathbf{L}(t) = \begin{bmatrix} L_1(t) \\ L_2(t) \\ L_3(t) \end{bmatrix}, \quad \mathbf{I}(t) = \begin{bmatrix} I_1(t) \\ I_2(t) \\ I_3(t) \end{bmatrix}, \tag{1}$$

$$\mathbf{R}(t) = \begin{bmatrix} R_1(t) \\ R_2(t) \\ R_3(t) \end{bmatrix}, \quad \mathbf{H}(t) = \begin{bmatrix} H_1(t) \\ H_2(t) \\ H_3(t) \end{bmatrix}. \tag{2}$$

Transfer matrix between age groups is defined as:

$$C = \begin{bmatrix} -(c_{12} + m_1) & 0 & 0 \\ c_{12} & -(c_{23} + m_2) & 0 \\ 0 & c_{23} & -m_3 \end{bmatrix} \tag{3}$$

The spread of infection is modeled through the following transmission matrix:

$$B = \beta(t) \cdot \begin{pmatrix} \sum_{i=1}^{3} I_i(t) & 0 & 0 \\ 0 & \sum_{i=1}^{3} I_i(t) & 0 \\ 0 & 0 & \sum_{i=1}^{3} I_i(t) \end{pmatrix} \tag{4}$$

With this notation, the system of equations is expressed as:

$$\mathbf{S}(t+1) = (I+C)\mathbf{S}(t) + \mu\mathbf{e}_1 + r_2 s\mathbf{R}(t) - B\mathbf{S}(t) \tag{5}$$

$$\mathbf{L}(t+1) = \mathbf{L}(t) + C\mathbf{L}(t) - l_2 i\mathbf{L}(t) + B\mathbf{S}(t) \tag{6}$$

$$\mathbf{I}(t+1) = \mathbf{I}(t) + C\mathbf{I}(t) - (i_2 r + i_2 h)\mathbf{I}(t) + l_2 i\mathbf{L}(t) \tag{7}$$

$$\mathbf{R}(t+1) = \mathbf{R}(t) + C\mathbf{R}(t) + h_2 r\mathbf{H}(t) + i_2 r\mathbf{I}(t) - r_2 s\mathbf{R}(t) \tag{8}$$

$$\mathbf{H}(t+1) = \mathbf{H}(t) + C\mathbf{H}(t) + i_2 h\mathbf{I}(t) - h_2 r\mathbf{H}(t) \tag{9}$$

where $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

The model parameters and variables are defined as follows:

- X_k represents the population X and the age group k, with $X = S, L, I, R, H$ and $k = 1, 2, 3$. Therefore, $G_k = \sum X_k$ and the total population is $N = G_1 + G_2 + G_3$.
- μ represents the daily births, based on the monthly birth data from the CVA [4]. Each day, the average daily births for the corresponding month are considered. Births are accounted for in the susceptible subpopulation S_1.
- c_{jj+1} represents the growth rate from age group j to $j+1$, where $j = 1, 2$.
- m_k represents the mortality rate of group k. These rates have been determined by balancing μ, c_{jj+1}, m_i, so that the number of individuals in each age group remains constant, balancing the equations of a demografic model [5].
- The rates $l_2 i, i_2 r_k, i_2 h_k, h_2 r, r_2 s$ represent the average days^{-1} an individual takes to transition between subpopulations: from latent to infected, from infected to recovered (for each age group k), from infected to hospitalized (for each age group k), from hospitalized to recovered, and from recovered to susceptible.

  It is assumed that both $i_2 h_k$ and $i_2 r_k$ are determined by general rates $i_2 h, i_2 r$, multiplied by the probability of hospitalization or recovery in each age group, respectively: $i_2 h_k = i_2 h \cdot p_k$, $i_2 r_k = i_2 r \cdot (1 - p_k)$. For the group older than 1 year, hospitalization probability is considered negligible ($p_3 = 0$), as described in [8]. This implies that infected individuals in group 3 (H_3) are not hospitalized, and the only entry into this subpopulation comes from the growth of hospitalized individuals in group 2 (H_2). Conversely, all infected individuals in group 3 directly transition to the recovered population (R_3).
- β is the transmission rate, describing the transition from susceptible to latent, and follows the expression:

$$\beta(t) = \frac{b_0}{N} \cdot \left(1 + b_1 \cdot \cos\left(\phi + \frac{2\pi t}{365}\right)\right) \tag{10}$$

  The cosine function represents the seasonal variation of the disease, with an annual period. The adjustment parameters are b_0, b_1, and ϕ. Specifically, b_1 determines the magnitude of seasonality: $b_1 = 0$ implies a constant transmission rate, while $b_1 = 1$ indicates extreme seasonality, with maximum transmission during epidemic peaks and dropping to zero during valleys. N is the total population of the CVA. The transmission rate is assumed to be uniform across and within age groups. It is important to properly define the parameter ranges $b_0 > 0$, $b_1 \in [0, 1]$; so that the transmission rate $\beta(t)$ remains strictly positive throughout the year, avoiding biologically implausible negative or zero values.

3.2 Calibration

The calibration process was carried out on 64 processing cores powered by eight Xeon Sandy Bridge E5-4620 CPUs operating at 2.2 GHz. These systems included 16 MB of cache and 512 GB of RAM. Both computers were running Ubuntu 20.04.6 LTS as their operating system.

The set of parameters to be calibrated in the model is $\{l_2i, i_2r_k, i_2h_k, h_2r, r_2s, \phi, p_1, p_2, b_0, b_1\}$. Model calibration was performed using the Particle Swarm Optimization (PSO) [9] technique, with the objective of minimizing the *Calibration Function*. In this work, the *Calibration Function* is defined as the sum of the Root Mean Squared Error (RMSE) between the model predictions and the hospitalization data for age groups 1 and 2.

Although the results of this calibration, as seen in section 4, are generally satisfactory, they fail to capture the sharp peaks of infection. To address this limitation, we will apply the Hybrid Digital Twin (HDT) framework.

3.3 Hybrid Digital Twins

HDT [3] combine traditional mathematical models with artificial intelligence (AI) techniques to enhance predictive accuracy. In this approach, both model outputs and experimental data contribute to refining an error term, E, which represents discrepancies between expected and observed behaviors. AI-based adjustments improve model accuracy without adding complexity to the model.

The HDT formulation follows:

$$H(X, \beta, t) = M(X, \beta, t) + E(X, t), \tag{11}$$

where H denotes the HDT model, M corresponds to the output of the mathematical model (5)-(9), E is the error due to unknown factors that is adjusted using AI techniques, X denotes the experimental data, β are the calibrated model parameters, and t is time.

This study employs Grammatical Evolution with Lexicase Selection [6] to approximate E, which generates an explicit function dependent on the system's input variables to fit the data, specifically the error E. Training is performed iteratively on random subsets of experimental data to prevent overfitting and improve generalization. The final function is selected based on the lowest RMSE across the entire dataset.

4 Results

The set of parameters shown in Table 1 corresponds to the best parameter set obtained for the model calibration using the PSO algorithm. These parameters fall within the appropriate ranges for RSV according to the literature [10].

Table 1: Best set of calibration parameters obtained with the PSO algorithm.

Parameter	l_{2i} days^{-1}	i_{2r_k} days^{-1}	i_{2h_k} days^{-1}	h_{2r} days^{-1}	r_{2s} days^{-1}	ϕ rad	p_1 %	p_2 %	b_0 -	b_1 -
Calibration value	0.216	0.132	0.232	0.178	0.0052	0.650	4.1	0.28	0.20	0.911

Figure 3 presents the results of evaluating the model with the calibrated parameter set. The model successfully reproduces the seasonal pattern of hospitalizations but falls short in capturing

the peak values observed in the data. In contrast, the HDT provide an improved fit in this regard, reaching higher hospitalization peaks that more closely match the observed values. This improvement is quantitatively supported by the RMSE comparison shown in Table 2 for age groups G_1, G_2, and the combined group $G_1 + G_2$, where the HDT consistently yield lower RMSE values than the original model.

Another metric used to assess the goodness of fit is the annual average number of hospitalizations, also reported in Table 2. Based on this metric as well, the HDT outperform the model, providing estimates that are more aligned with the observed data.

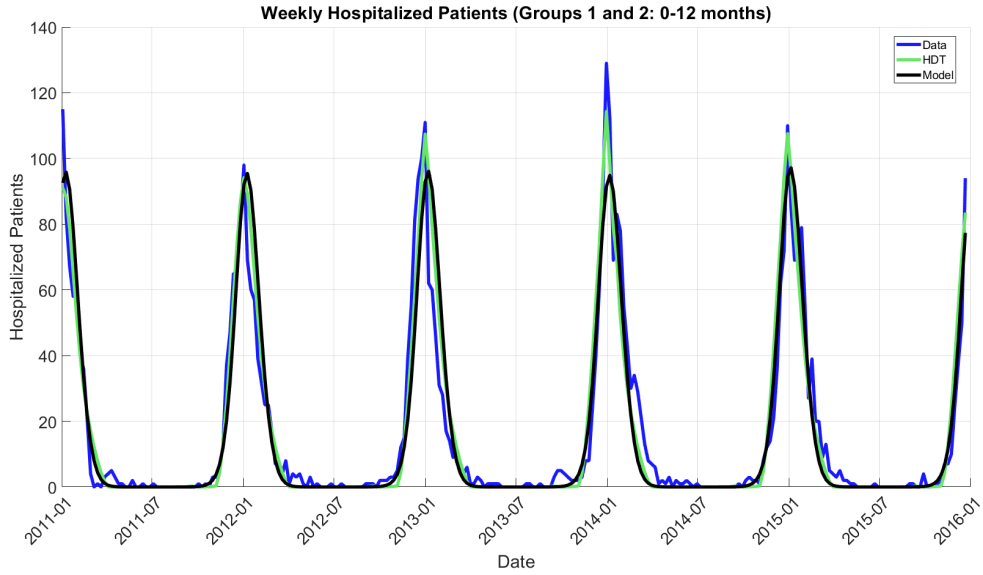


Figure 3: Comparison of the Differences Model and HDT with the Data.

Table 2: Comparison for Annual Hospitalizations and RMSE between Data, Model and HDT.

	(0-5months)	RMSE	(6-12 months)	RMSE	(0-12 months)	RMSE
Data	773.6	-	71.0	-	844.6	-
Model	750.8	8.6	75.8	1.5	826.4	10.1
HDT	765.8	7.5	69.9	1.4	835.7	8.9

Finally, Table 3 compares the execution times of the model and the HDT. It can be observed that the HDT result in only a minimal increase in the model simulation time.

Table 3: Comparison for average Time per Simulation (s) between Agent-Based Model, Differences Model and HDT Model.

	Time per Simulation (s)
Differences Model	0.29
HDT Model	0.30

5 Conclusions

This study introduces a new approach to RSV modeling. The incorporation of HDT has resulted in more accurate predictions, particularly at the critical data values of peaks hospitalizations, while only slightly increasing the computational cost of the model and without adding extra complexity to the equations. This faster and more precise model opens the door for future research, where a detailed analysis of various vaccination strategies and their long-term effects on the disease can be explored.

Acknowledgments

Marcos Llamazares López has been a recipient of a doctoral scholarship from the Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València (UPV).

References

[1] Centers for Disease Control and Prevention (CDC), "Learn about RSV infants and young children" https://www.cdc.gov/rsv/infants-young-children/index.html 2025, accessed: 2025-4-03.

[2] "A multivariate age-structured stochastic model with immunization strategies to describe bronchiolitis dynamics," https://rotapp.shinyapps.io/APP IMPACTO RSV/, accessed: 2024-10-24

[3] Chinesta, Francisco and Cueto, Elias and Abisset-Chavanne, Emmanuelle and Duval, Jean Louis and Khaldi, Fouad El. (2018) "Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data." *Archives of Computational Methods in Engineering, Springer Science and Business Media LLC*. doi: 10.1007/s11831-018-9301-4

[4] Instituto Nacional de Estadística (INE), "Instituto Nacional de Estadística - National Institute of Statistics (INE)," [Online]. Available: `http://www.ine.es`, accessed: 2025-04-29.

[5] H. W. Hethcote, "The Mathematics of Infectious Diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000. [Online]. doi: `https://doi.org/10.1137/S0036144500371907`

[6] Conor, Ryan and Michael, O'Neill and JJ, Collins. (2018) "Handbook of Grammatical Evolution." *Springer International Publishing*. ISBN: 9783319787176. doi: 10.1007/978-3-319-78717-6

[7] E. Goldstein, L. Finelli, A. O'Halloran, P. Liu, Z. Karaca, C. A. Steiner, C. Viboud, and M. Lipsitch, "Hospitalizations associated with respiratory syncytial virus and influenza in children, including children diagnosed with asthma," *Epidemiology*, vol. 30, no. 6, p. 918, 2019. [Online]. doi: https://doi.org/10.1097/EDE.0000000000001092

[8] M. Bardsley, R. A. Morbey, H. E. Hughes, C. R. Beck, C. H. Watson, H. Zhao, J. Ellis, G. E. Smith, and A. J. Elliot, "Epidemiology of respiratory syncytial virus in children younger than 5 years in england during the covid-19 pandemic, measured by laboratory, clinical, and syndromic surveillance: A retrospective observational study," *The Lancet Infectious Diseases*, vol. 23, no. 1, pp. 56–66, 2023. [Online]. doi: https://doi.org/10.1016/S1473-3099(22)00525-4

[9] Marini, Federico and Walczak, Beata. (2015) "Particle swarm optimization (PSO). A tutorial." *Chemometrics and Intelligent Laboratory Systems* 149. doi: 10.1016/j.chemolab.2015.08.020.

[10] T. M. Pollard, N. G. Whitehead, and N. G. Medley, "Modelling the seasonal variations of respiratory syncytial virus," *Epidemiology Infection*, vol. 138, no. 2, pp. 226–233, 2010. [Online]. doi: https://doi.org/10.1017/S0950268809991373

Modelling and analysis of the dynamics of infectious diseases described by differential systems with uncertainties

Javier López-de-la-Cruz*

Dpto. de Matemática Aplicada a las TIC
Escuela Técnica Superior de Ingenieros Informáticos
Campus de Montegancedo, Universidad Politécnica de Madrid
28660 Boadilla del Monte, Madrid, Spain.

Epidemics have influenced the humanity over history, bringing about consequences as catastrophic as the ones caused by wars and causing extinctions of entire populations. Apart from the human losses, epidemics also have important social consequences and have a huge negative impact on the economy.

Many infectious diseases, which have caused epidemics over history, are not eradicated yet and new ones are arising over the years. Hence, it is essential to set up mathematical models to describe the evolution of the diseases and then understand their evolution and make decisions to control the spread of the disease as soon as possible, trying to minimize the consequences.

The first known paper about mathematical modeling of epidemics was published in 1760 (see [2]), where D. Bernoulli studied the spread of smallpox by means of a system of ordinary differential equations. However, the study of mathematical models in Epidemiology was not developed until the very beginning of the twentieth century, when W. H. Hamer formulated in 1906 a discrete mathematical model to describe the transmission of measles (see [4]). Afterward, in 1911, R. Ross provided a mathematical model to predict the spread of an outbreak of malaria (see [7]), where he proved that it was enough to partially reduce the population of mosquitoes to eradicate the epidemics.

Nevertheless, in 1927 Kermack and McKendrick established a mathematical model to describe the spread of an infectious disease (see [5]), which has been the base of mathematical epidemiology ever since. This classical model, called classical SIR model, is a compartmental model, i.e., it divides the population into three groups. The first group contains the susceptible individuals (S), people who are not infected but can contract the disease when contacting an infected. The second group contains the infected individuals (I). The third group contains the recovered individuals (R), people who are recovered with immunity or who pass away because of the disease.

*Corresponding author: javier.lopez.delacruz@upm.es

The aforementioned classical SIR model is given by the following nonlinear system of ordinary differential equations

$$\frac{dS}{dt} = -\gamma \frac{S\,I}{S+I+R}, \tag{1}$$

$$\frac{dI}{dt} = \gamma \frac{S\,I}{S+I+R} - c\,I, \tag{2}$$

$$\frac{dR}{dt} = c\,I, \tag{3}$$

where $S = S(t)$, $I = I(t)$, and $R = R(t)$ denote the number of susceptible, infected, and recovered individuals at time t, respectively, $\gamma > 0$ (days^{-1}) represents the transmission coefficient of the disease and $1/c > 0$ (days) describes the mean time that an individual remains infected when contracting the disease.

Unfortunately, the classical SIR model (1)-(3) assumes strong restrictions. For instance, the total population is supposed to be constant. However, sometimes infectious diseases do not disappear quickly (as the COVID-19), then the size of the population usually varies due to external factors. Another important restriction is to assume that every infected individual either recovers with immunity or passes away, which is not the case with the COVID-19.

To set up a more realistic model and avoid the drawbacks pointed out above, we will introduce vital dynamics and reinfection in the classical SIR model. Then, the resulting model is given by

$$\frac{dS}{dt} = q - \gamma \frac{S\,I}{S+I+R} - a\,S, \tag{4}$$

$$\frac{dI}{dt} = -(a+b+c)\,I + \gamma \frac{S\,I}{S+I+R}, \tag{5}$$

$$\frac{dR}{dt} = c\,I - a\,R, \tag{6}$$

where $q > 0$ denotes the recruitment rate (number of individuals $\times$ days^{-1}), which takes into account the new susceptible individuals joining the population during the epidemic (such as births), $a > 0$ (days^{-1}) takes into account the deaths that are not caused by the disease (such as natural deaths, accidents or others) and $b > 0$ (days^{-1}) describes those infected individuals that recover without immunity.

Nevertheless, we will go further and we will also perturb randomly the transmission coefficient in (4)-(6), motivated by facts observed in real epidemics (see [1]). Hence, we will focus on the following random system

$$\frac{dS}{dt} = q - (\gamma + \Phi(\theta_t\omega)) \frac{S\,I}{S+I+R} - a\,S, \tag{7}$$

$$\frac{dI}{dt} = -(a+b+c)\,I(t) + (\gamma + \Phi(\theta_t\omega)) \frac{S\,I}{S+I+R}, \tag{8}$$

$$\frac{dR}{dt} = c\,I - a\,R, \tag{9}$$

where $\Phi(\theta_t\omega)$ denotes a real (i.e, bounded) noise satisfying certain properties.

In this talk, which will be based on the recently published paper [6], we will first present the previous models. After that, we will analyze the random system (7)-(9) by means of the theory of random dynamical systems and pullback attractors (see [3]). This will allow us to provide useful information about the long-time behavior of the disease and, in addition, to find conditions under which the disease either is eradicated or becomes endemic, the main goal for practitioners. Apart from that, we will interpret the results from the biological point of view. Finally, we will illustrate the theoretical results by depicting several numerical simulations.

References

[1] V. V. L. Albani and J. P. Zubelli. Stochastic transmission in epidemiological models. Journal of Mathematical Biology, vol. 88, no. 25 (2024).

[2] D. Bernoulli. Essai d'une nouvelle analyse de la mortalité cause par la petite vrole, et des aventages de l'inoculation pour la prévenir. In: Mémoires de mathématique et de physiques tires des registres de l'Academie Royale des Sciences (1760), pp. 1–75.

[3] T. Caraballo and X. Han. Applied Nonautonomous and Random Dynamical Systems, Applied Dynamical Systems. Springer International Publishing (2016).

[4] W. H. Hamer. The milroy lectures on epidemic disease in England - The evidence of variability and persistence of type. Lancet (1906).

[5] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, vol. 115 (1927), pp. 700–21.

[6] J. López-de-la-Cruz and Alexandre N. Oliveira Sousa. SIR models with vital dynamics, reinfection, and randomness to investigate the spread of infectious diseases. Communications in Nonlinear Science and Numerical Simulation, vol. 140 (2025), pp. 108359.

[7] R. Ross. The prevention of malaria. 2nd ed. London: John Murray (1911).

Modeling Co-Infection Patterns of COVID-19 and Dengue

C. Pinto[1,*], J.Calatayud[2] and M. Jornet[3]

(1) School of Engineering, Polytechnic of Porto[1],
CIDMA, Aveiro, Portugal.
(2) Departament de Matemàtiques, Universitat Jaume I[2],
12071 Castellón, Spain
(3) Universidad Internacional De La Rioja[3]
Logroño, La Rioja, ES

1 Introduction

We present a novel mathematical model designed to describe the overlapping dynamics of dengue and COVID-19 infections within a susceptible population, formulated as a nonlinear system of ordinary differential equations. Initially, we derive the basic reproduction number and demonstrate its utility in assessing outbreak potential, long-term behavior, and sensitivity to model parameters. Subsequently, we extend the model to a stochastic framework using an Itô formulation and perform numerical simulations to analyze its dynamics, offering a generalization of the deterministic version. The model is calibrated using real-world epidemiological data from Colombia through both global fitting and segmented (sub-stage) fitting strategies. We address key challenges such as parameter unidentifiability and limited data availability. To streamline the least-squares fitting process, some parameters were fixed in advance. As a result, the model's predictions should be interpreted with appropriate caution. Addressing these limitations remains essential for the continued development of robust epidemic models.

2 Methods

The proposed nonlinear model of ordinary differential equations proposed in [1]is the following. The variables of the model are : A_m - & mosquito's aquatic phase, F - uninfected female mosquito, F_1 - uninfected fertilized female mosquito, F_2 - infected mated fertilized female mosquito, M - male mosquito, H_s - susceptible human, H_{iD} - dengue infectious human, H_{rD} - dengue recovered human, H_{iC} - SARS-CoV-2 infectious human, H_{rC} - COVID-19 recovered human, C_{CD} - coinfectious human. The parameters are defined as follows: ϕ_1 - aquatic mosquito recruitment rate, C - carrying capacity of aquatic/adult mosquito, σ_A - mosquitoes' transition rate from aquatic to adult stages, μ_A - mosquitoes' aquatic form mortality rate, p - proportion of female mosquitoes, β - mosquitoes mating rate, μ_m - mosquitoes' natural death rates, β_m - transmission rate from humans to mosquitoes, b - average bite per mosquito per day, μ_h - human natural death rate, μ_c - human coinfection induced death rate, β_h - contact rate from female mosquitoes to humans, β_c - COVID-19 transmission rate,η_D - dengue partial immunity,

* cap@isep.ipp.pt

η_C & COVID-19 partial immunity, α_D - dengue recovering rate, α_C - COVID-19 recovering rate, ξ_C - COVID-19 coinfected recovering rate, ξ_D - dengue coinfected recovering rate.

$$
\begin{aligned}
A'_m &= \phi_1\left(1-\tfrac{A_m}{C}\right)(F_1+F_2)-(\sigma_A+\mu_A)A_m,\\
F' &= p\sigma_A A_m-\tfrac{\beta}{M}FM-\mu_m F,\\
F'_1 &= \tfrac{\beta}{M}FM-\tfrac{b\beta_m}{H}(H_{iD}+C_{CD})F_1-\mu_m F_1,\\
F'_2 &= \tfrac{b\beta_m}{H}(H_{iD}+C_{CD})F_1-\mu_m F_2,\\
M' &= (1-p)\sigma_A A_m-\mu_m M,\\
H'_s &= \mu_h(H-H_s)-\tfrac{b\beta_h}{H}H_sF_2-\beta_C\tfrac{(H_{iC}+C_{CD})}{H}H_s+\eta_D H_{rD}+\eta_C H_{rC},\\
H'_{iD} &= \tfrac{b\beta_h}{H}H_sF_2+\zeta_C C_{CD}-\beta_C\tfrac{(H_{iC}+C_{CD})H_{iD}}{H}-(\alpha_D+\mu_h)H_{iD},\\
H'_{rD} &= \alpha_D H_{iD}-(\eta_D+\mu_h)H_{rD},\\
H'_{iC} &= \beta_C\tfrac{(H_{iC}+C_{CD})}{H}H_s+\zeta_D C_{CD}-\tfrac{b\beta_h}{H}F_2H_{iC}-(\alpha_C+\mu_h)H_{iC},\\
H'_{rC} &= \alpha_C H_{iC}-(\eta_C+\mu_h)H_{rC},\\
C'_{CD} &= \tfrac{b\beta_h}{H}F_2H_{iC}+\beta_C\tfrac{(H_{iC}+C_{CD})}{H}H_{iD}-(\zeta_C+\zeta_D+\mu_c)C_{CD}.
\end{aligned}
\tag{1}
$$

3 Results

Figure 1 shows the stochastic fit. Dengue remains stable with low weekly cases, while COVID-19 rises rapidly, becoming the dominant threat. This aligns with the basic reproduction numbers: R_C=1.12, R_D=0.15.

However, caution is needed when using these values for long-term forecasts—dengue persists in Colombia, and COVID-19 is now largely controlled, limiting the model's validity to short time frames.

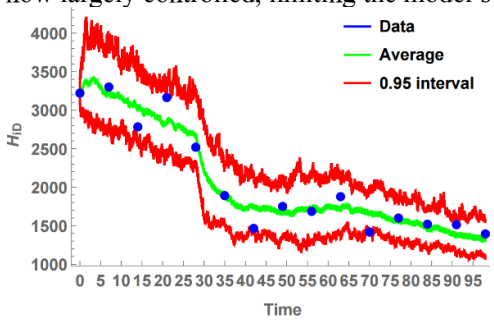


Figure 1: Fitting to the Colombia data with the stochastic model corresponding to model (1) [1].

4 Conclusions

We developed a mathematical model using nonlinear ODEs to describe the co-dynamics of dengue and COVID-19 in a susceptible population, structured into 11 compartments for human and mosquito stages. After analyzing the deterministic system, we extended the model to a stochastic framework with

Brownian motions to capture uncertainty. Numerical simulations confirmed the stochastic model's consistency with deterministic trends. Fitting to Colombian case data from 2020, we addressed challenges such as parameter unidentifiability and limited surveillance by fixing some parameters and optimizing key ones. While the model shows strong potential, its results must be interpreted cautiously due to data and structural limitations. Improving data quality and parameter identifiability will be key for advancing co-infection modeling.

Acknowledgements

C.Pinto's work is supported by CIDMA under the FCT (Portuguese Foundation for Science and Technology) Multi-Annual Financing Program for R&D Units.
The author Marc Jornet was funded by Agencia Estatal de Investigación, Spain, under the project with reference PID2023-146836NB-I00, and by Generalitat Valenciana, Spain, under the project with reference CIAICO/2021/227.

References

[1] Calatayud, J., Jornet M., Pinto, Carla M.A. Patterns of dengue and SARSCoV-2 coinfection in the light of deterministic and stochastic models, *Nonlinear Analysis: Real World Applications* - submitted March (2025).

Using AI to Detect and Predict Infectious Disease Seasonalilty and Trends

J.M. Montañana*[1] and A. Hervás‡

(∗) inpeek GmbH,
lda-Rhodes-Straße 1, 64295 Darmstadt, Germany.
jmmontanana@gmail.com, jose.montanana@inpeek.de

(‡) IUMM, Universitat Politècnica de València, València, Spain,
Camí de Vera s/n, València, Spain.

1 Introduction

Recent advancements in Artificial Intelligence (AI) have opened new possibilities for developing predictive models that can serve as valuable tools for public health institutions. By leveraging machine learning (ML) and deep learning (DL) techniques, AI systems can analyze large volumes of data, even when datasets are inbalanced [4] or incompleted [6], to uncover patterns, forecast disease trends, and identify the most influential factors affecting disease spread.

These modern AI approaches—particularly those incorporating advanced architectures such as convolutional neural networks and recurrent models—can complement traditional statistical methods. Their ability to adapt to complex and often inconsistent epidemiological data makes them particularly suitable for real-world public health applications, where data may not always be recorded with the same frequency or accuracy. This flexibility allows for the creation of robust, scalable, and responsive tools for disease monitoring and prediction.

These tools open the oportunity for timely forecasting for supporting effective disease control, enabling early intervention and the implementation of preventive measures. In this context, the capacity of the AI to manage complex, high-dimensional datasets enables earlier detection and more informed decision-making in the face of potential outbreaks.

Numerous studies have demonstrated the effectiveness of AI models in predicting infectious disease dynamics. For instance, Long Short-Term Memory (LSTM) neural networks[3] have proven capable in forecasting influenza outbreaks [1, 5], and Foodborne Disease risk[9]. Likewise, Random Forest models[2] have shown strong predictive power and interpretability in various epidemiological settings, often outperforming other ML algorithms depending on the nature of the data [7].

In this study, we analyze weekly influenza and monthly campylobacteriosis case data from the canton of Bern, Switzerland, covering the period from 2013 to 2025. We apply both Random Forest and LSTM models to predict disease incidence and to assess their effectiveness to capture underlying seasonal patterns. This work represents a first step toward the broader goal of developing AI-based tools for automatic detection, quantification, and monitoring of seasonality in infectious disease data. By establishing a framework for early identification of seasonal trends, we aim to support the creation of responsive, data-driven public health interventions.

[1] jose.montanana@inpeek.de

2 Methods

We analyzed two time series datasets of notifiable infectious diseases in Switzerland: weekly influenza cases and monthly campylobacteriosis cases from 2013 to 2025, specifically for the canton of Bern. Data were obtained from the Swiss Federal Office of Public Health (FOPH), publicly available at [8], and preprocessed to ensure temporal consistency, proper date formatting, and inclusion of seasonal features (e.g., sine and cosine encodings of week or month).

Two machine learning approaches were implemented for each disease: a Random Forest regression model and a Long Short-Term Memory (LSTM) neural network. Both models were trained using a growing window strategy, predicting one full year (52 or 12 time steps for weekly and monthly data respectively) based on prior observations. The LSTM architecture consisted of a single LSTM layer followed by a dense output layer, trained with mean squared error loss and the Adam optimizer. Feature scaling was applied using MinMax normalization.

Model performance was evaluated yearly, and predictions were plotted against true values to assess temporal accuracy. Statistical analyses, including autocorrelation, ANOVA, and augmented Dickey-Fuller tests, were conducted to assess the presence and strength of seasonal patterns in the data.

2.1 Model Considerations

Random Forest is a tree-based ensemble method suited for tabular data. It does not inherently account for time sequences, treating each data point independently. In contrast, Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) specifically designed to handle sequential patterns in time series data. LSTMs incorporate memory mechanisms that enable them to learn long-term temporal dependencies, making them well-suited for time-series forecasting.

3 Results

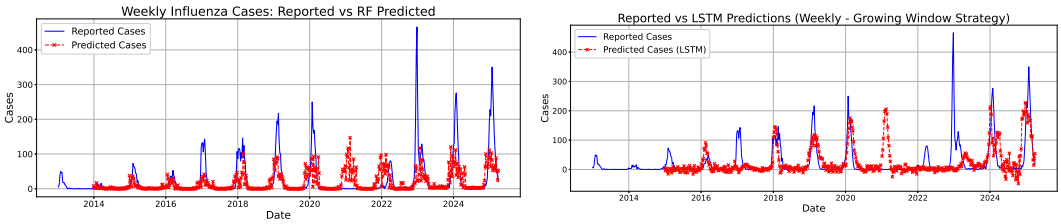
We evaluated the predictive performance of two machine learning models—Random Forest and Long Short-Term Memory (LSTM)—on infectious disease data from the canton of Bern, Switzerland. The datasets included weekly Influenza cases and monthly Campylobacteriosis cases spanning from 2013 to 2025.

3.1 Influenza

The weekly influenza data from 2013 to 2025 reveal distinct temporal patterns shaped by both epidemiological and societal factors. Between 2013 and 2014, reported case counts remained relatively low. However, during the COVID-19 pandemic (2020-2021), case numbers dropped dramatically, likely due to lockdowns, social distancing. From 2022 onwards, case counts began to rise again. These structural shifts introduced non-stationarities into the time series, which contributed to inconsistencies in model performance across years (Figures 1.a and 1.b show the predicted cases from both models).

Despite these disruptions, the data exhibit clear seasonality over the decade. To quantify this, we computed several statistical metrics:

- **Autocorrelation at seasonal lag (52 weeks):** A value of 0.48 indicates moderate annual seasonality, exceeding the typical threshold of 0.3 and suggesting a repeating yearly pattern.
- **Seasonal variance ratio:** A ratio of 0.34 implies that roughly one-third of the total variance is attributable to seasonal components, highlighting a meaningful seasonal structure.



(a) (b)

Figure 1: (a) Influenza cases predicted using Random Forest, and (b) Influenza cases predicted using LSTM. Both models are compared against reported weekly cases to assess forecasting performance.

- **ANOVA F-statistic:** An F-statistic of 6.99 ($p < 0.0001$) confirms statistically significant differences across weeks, supporting the presence of seasonality.
- **ADF stationarity test:** The Augmented Dickey-Fuller test produced a statistic of -8.061 with a p-value of 0.0000, allowing rejection of the null hypothesis of non-stationarity. This indicates that the series is stationary, which is favorable for modeling and consistent with a stable seasonal pattern.

These properties make the time series well-suited for classical machine learning models such as Random Forest, which can leverage both lag-based and seasonal features for effective prediction. While the Random Forest model captured the general trend of influenza incidence reasonably well during non-pandemic years, it consistently underestimated peak values during high-incidence weeks. This limitation suggests that, although the model recognizes seasonal fluctuations, it struggles to extrapolate rare or extreme events—likely due to their underrepresentation in the training data.

The LSTM model, however, struggled more significantly, particularly during and after the pandemic. For this period, seasonality analysis revealed a seasonal variance ratio of only 0.03 and a negative autocorrelation at lag 52 (-0.31), indicating that recent influenza dynamics lacked a clear repetitive structure.

Figure 2 presents the seasonal decomposition of the influenza incidence time series based on weekly data into four panels. This decomposition, which does not rely on predictive modeling, separates the observed series into four components: observed values, trend, seasonality, and residuals. While descriptive in nature, this decomposition can provided a reference for improving the predictive model, such as trends or recurring seasonal patterns that can be explicitly be considered when building the model.

The first panel displays observed influenza cases, with values ranging from 0 to 466. The data demonstrate a clear seasonal pattern, with pronounced annual peaks. The magnitude and timing of these peaks vary slightly across years, reflecting intrinsic epidemiological dynamics and external influences.

The second panel illustrates the extracted trend component, scaled between 0 and 55 on the y-axis. Excluding the anomalous period during 2021 and 2022, when public health interventions were in place, the trend shows a gradual upward trajectory from 2014 to 2024. In future work, we plan to improve the predictive model that incorporates additional covariates such as population growth to better account for this trend.

The third panel represents the seasonal component, characterized by a highly regular pattern that recurs annually. The seasonal fluctuations oscillate between approximately -22 and 78, corresponding to well-established wintertime peaks in temperate regions.

The final panel presents the residuals, accounting for variations not explained by the trend and seasonality. Before 2017, residuals were mainly negative during peak periods, whereas after

2017, they were predominantly positive during peaks and approached zero during non-peak periods. Notably, during the COVID-19-affected years and in 2023, residuals reached unusually high values, suggesting the influence of events or external factors not captured by the decomposition. Future modeling efforts may consider dynamic seasonal components to reduce this variance.

Overall, the decomposition highlights the substantial roles of both seasonality and long-term trends in shaping influenza dynamics.

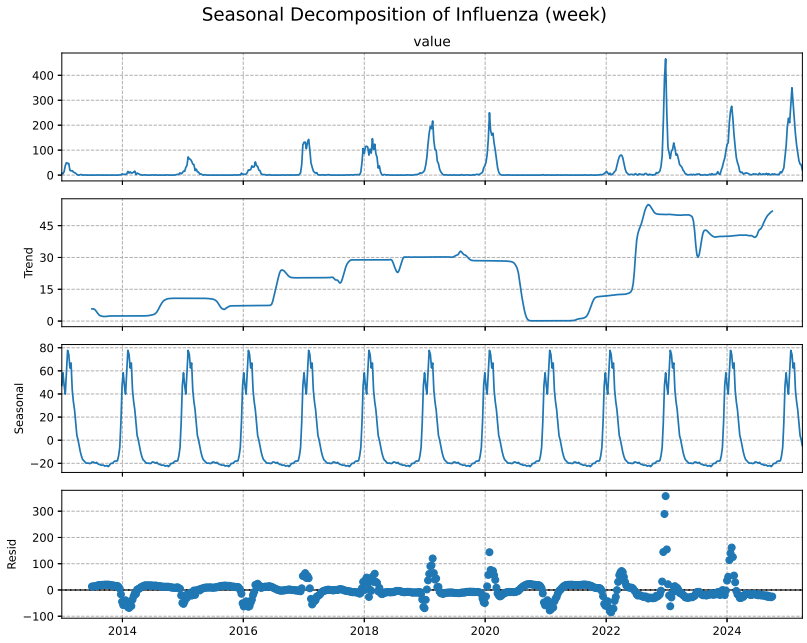
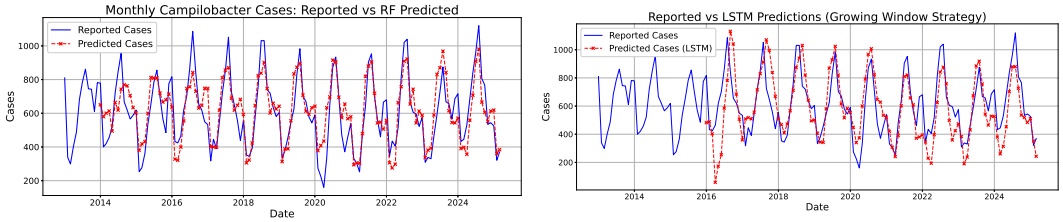


Figure 2: Seasonal decomposition of weekly Influenza cases.

3.2 Campylobacteriosis

Unlike influenza, Campylobacteriosis case patterns were less affected by the COVID-19 pandemic. As a result, both models demonstrated relatively stable performance over the years (Figures 3.a and 3.b).



(a) (b)

Figure 3: (a) Campylobacteriosis cases predicted using Random Forest, and (b) Campylobacteriosis cases predicted using LSTM. Monthly reported cases are compared with model predictions to evaluate forecasting accuracy.

The Random Forest model aligned well with the inherent seasonality in the monthly data. It exhibited strong autocorrelation at lag 12 (0.75) and a high seasonal variance ratio of 0.84,

indicating that over 80% of the variability is attributable to recurring monthly patterns. The ANOVA F-statistic of 65.39 with $p < 0.0001$ confirms a pronounced seasonal structure (Table 1). The ADF statistic of -3.526 ($p = 0.0073$) also supports the stationarity of the series.

The LSTM model performed comparably, benefiting from the strong seasonal regularity. It achieved similar stationarity and seasonality statistics, confirming that Campylobacteriosis data are well-suited for time-series learning models. The consistent performance across both models highlights the robustness of the monthly structure in this dataset.

Figure 4 depicts the seasonal decomposition of the Campylobacteriosis incidence time series based on monthly data.

The first panel shows the raw Campylobacteriosis cases over time, with annual fluctuations between a minimum of 160 and a maximum of 1120 cases.

The second panel displays the trend component, which suggests a possible oscillating pattern with peaks approximately every two years. A sharp decline is evident around July 2020. We wish to develop a model, in a future work, with a larger dataset and including also other regions to investigate if this trend is also associated with pandemic-related factors.

The third panel presents the seasonal component, which remains stable across years. Maximum seasonal values occur in August, corresponding to late summer peaks likely driven by environmental factors, including elevated ambient temperatures favoring bacterial growth and human behavior (e.g., outdoor activities, food handling practices). The seasonal component oscillates between approximately -266.88 and 369.48.

The final panel presents the residuals, which are relatively small in magnitude compared to the trend and seasonal components (ranging from -156.00 to 150.08), indicating a good fit of the model to the data with minimal unexplained variation.

In summary, the Campylobacteriosis dataset displays a stable seasonal pattern with consistent late summer peaks. While the trend oscillates approximately biennially, the sharp drop in 2020 warrants further investigation. The residual component remains small, suggesting a well-captured temporal structure.

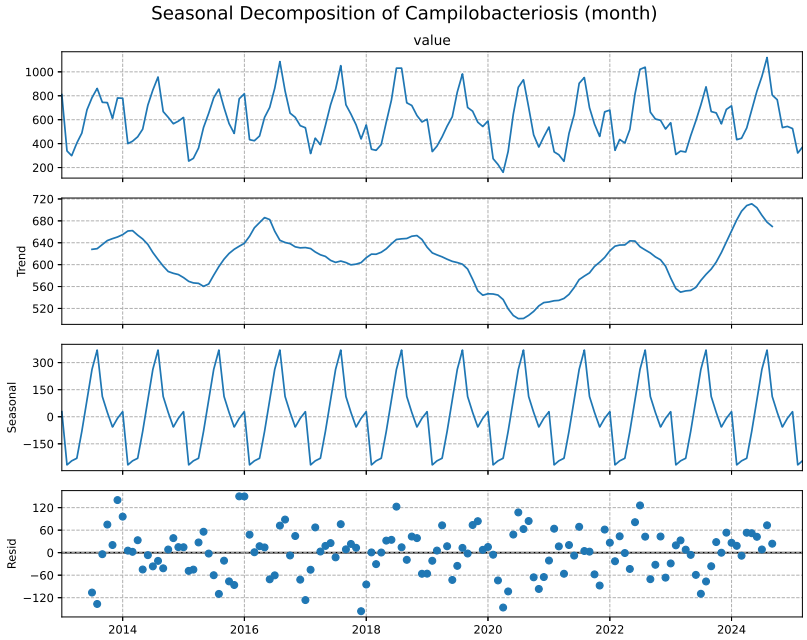


Figure 4: Seasonal decomposition of monthly Campylobacteriosis cases.

To facilitate a direct comparison between the influenza and campylobacteriosis datasets, we summarize the key seasonality and stationarity statistics in Table 1.

Metric	Ideal Range	Influenza	Campylobacteriosis	Interpretation
Autocorrelation (seasonal lag)	> 0.3	0.48	0.75	Moderate (Influenza), Strong (Campy)
Seasonal variance ratio	> 0.2	0.34	0.84	Moderate (Influenza), Very Strong (Campy)
ANOVA F-statistic	> 5	6.99	65.39	Clear weekly/monthly group differences
ANOVA p-value	< 0.05	0.0000	0.0000	Statistically significant seasonality
ADF p-value	< 0.05	0.0000	0.0073	Series are stationary

Table 1: Summary of seasonality and stationarity statistics across datasets.

4 Conclusions

In this study, we explored the predictive performance of Random Forest and Long Short-Term Memory (LSTM) models applied to weekly influenza and monthly campylobacteriosis case data from Bern, Switzerland, spanning from 2013 to 2025. Through our analysis, we observed notable seasonal patterns in both datasets, confirming that disease incidence in both cases is influenced by yearly cycles. However, our results also highlight the challenges posed by disruptions such as the COVID-19 pandemic, which caused notable shifts in influenza patterns and contributed to fluctuating prediction accuracy.

The Random Forest model demonstrated its capability to capture seasonal trends in both datasets, particularly for campylobacteriosis, where the seasonal structure remained more stable across the years. Meanwhile, the LSTM model, despite its strength in handling temporal dependencies, struggled to model the post-pandemic shifts in influenza cases, as evidenced by low seasonal variance ratios and negative autocorrelation values.

Our analysis of the seasonality metrics—autocorrelation, seasonal variance ratio, ANOVA, and stationarity—has provided strong evidence of predictable yearly cycles in both diseases. The statistical tests confirmed that both datasets exhibited significant seasonality, as indicated by high ANOVA F-statistics and p-values below 0.05. Furthermore, the stationarity of the time series, confirmed by the Augmented Dickey-Fuller test, supports the feasibility of future time-series forecasting models to capture long-term trends effectively.

Contribution to AI-Driven Public Health Systems: Our primary goal is to contribute to the development of AI-driven tools that can detect seasonality and provide real-time indicators for public health monitoring. Our results justify this effort by demonstrating that machine learning models, particularly Random Forest and LSTM, can be utilized to identify and track seasonal patterns in disease data. The ability of these models to learn from historical data and adapt to new trends as they emerge provides a solid foundation for building tools that automatically update predictions and seasonal insights as new data becomes available.

By automating the detection of seasonality and continuously updating disease forecasts, AI models like the ones tested here can enhance the responsiveness of public health systems to emerging threats. For instance, these tools could provide early warnings of potential disease outbreaks based on seasonal patterns, allowing for timely interventions and resource allocation. Further work is needed to refine these models and enhance their robustness in the face of unexpected disruptions, such as the ones caused by the COVID-19 pandemic.

In conclusion, this initial exploration of AI-based models for disease prediction underscores the potential for using machine learning to improve the monitoring and forecasting of public health trends. Our ongoing research aims to refine these models and extend their capabilities, ultimately contributing to more dynamic and adaptive public health systems.

References

[1] Amendolara, A.B., Sant, D., Rotstein, H.G. et al. LSTM-based recurrent neural network provides effective short term flu forecasting. *BMC Public Health*, 23, 1788 (2023). https://doi.org/10.1186/s12889-023-16720-6

[2] Breiman, L. Random forests. Machine learning, Springer. 45 (1), pp 5–32. (2001).

[3] Hochreiter, S. and Schmidhuber, J. Long short-term memory. Neural Comput. MIT Press, vol 9 (8) pp 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, (1997).

[4] Khalilia, M., Chakraborty, S. and Popescu, M. Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 11, 51 (2011). https://doi.org/10.1186/1472-6947-11-51

[5] Lu, M., Ilk, N., Tang, X., Liu, Y. Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, 177, 114905 (2021). https://doi.org/10.1016/j.eswa.2021.114905

[6] Mehdipour-Ghazi, M., Nielsen, M., Pai, A. et al, Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling, Medical Image Analysis, Elsevier BV, vol 53, pp 39–46, http://dx.doi.org/10.1016/j.media.2019.01.004 (2019).

[7] Pillai, N., Ramkumar, M., Nanduri, B. Artificial Intelligence Models for Zoonotic Pathogens: A Survey. *Microorganisms*, 10(10), 1911 (2022). Available at: https://www.mdpi.com/2076-2607/10/10/1911

[8] Swiss Federal Office of Public Health (FOPH). *Influenza and Campylobacteriosis Disease Data*. Available at: `https://www.idd.bag.admin.ch/diseases/campylobacteriosis/data` and `https://www.idd.bag.admin.ch/diseases/influenza/data`, accessed on 15 April 2025.

[9] Du Y, Wang H, Cui W, Zhu H, Guo Y, Dharejo FA, Zhou Y Foodborne Disease Risk Prediction Using Multigraph Structural Long Short-term Memory Networks: Algorithm Design and Validation Study JMIR Med Inform;9(8):e29433 doi: 10.2196/29433 PMID: 34338648 PMCID: 8369373 (2021).

Asymptotic behavior for a SIR model with random transmission coefficient and time depends vaccination rate

Javier López-de-la-Cruz[♭],[1] Susana Merchán[♮],[2] Felipe Rivero[♭],[3] and Javier Rodrigo[♯4].

(♭) Dpto. de Matemática Aplicada a las TIC, ETS de Ingenieros Informáticos.
Universidad Politécnica de Madrid, Spain.

(♮) Dpto. de Matemática e Informática Aplicadas a las Ingenierías Civil y Naval
ETS de Ingenieros Caminos, Canales y Puertos. Universidad Politécnica de Madrid, Spain.

(♯) Dpto. de Matemática Aplicada, Escuela Técnica Superior de Ingeniería.
Universidad Pontificia Comillas de Madrid, Spain.

1 Abstract

In this work we studied the asymptotic behavior of a SIR model where the transmission coefficient presents a bounded random variation along the time and the vaccination rate is consider as a time depends function. Specifically, we study the dynamic of the following system

$$\begin{cases} S'(t) = \nu - (\beta + \Phi(z^*(\theta_t\omega)) \dfrac{S(t)I(t)}{S(t)+I(t)+R(t)} - \mu S(t) - p(t)S(t), \\ I'(t) = (\beta + \Phi(z^*(\theta_t\omega)) \dfrac{S(t)I(t)}{S(t)+I(t)+R(t)} - \delta I(t) - \mu I(t), \\ R'(t) = \delta I(t) - \mu R(t) + p(t)S(t), \end{cases}$$

where $S = S(t)$ denotes the number of susceptible individuals, $I = I(t)$ is the number of infected individuals, $R = R(t)$ symbolizes the number of recovered individuals, we have that

- ν : birth rate of the population,
- μ : death rate of the population,
- β : transmission coefficient of the disease,
- δ^{-1} : average incubation time,
- $p(t)$: vaccination rate,

[1]javier.lopez.delacruz@upm.es
[2]javier.lopez.delacruz@upm.es
[3]felipe.rivero@upm.es
[4]jrodrigo@comillas.edu

and $\Phi(z^*(\theta_t\omega))$ models a random bounded perturbation of the transmission coefficient using the the Ornstein-Uhlenbeck process (see [1]).

We were interested in the case that makes biological sense, so we studied the behavior of positive solutions, that is, the dynamic in $\mathbb{R}^3_+ = \{(S, I, R) \in \mathbb{R}^3, S > 0, I > 0, R > 0\}$, giving conditions over the coefficients in order to obtain the extinction or the endemicity of the infection in the population.

References

[1] Caraballo, T., Garrido, M.J., López-de-la-Cruz, J., Modeling and analysis of random and stochastic input flows in the chemostat model. *Discrete and Continuous Dynamical Systems-Series B*, 24 (8), 3591-3614 (2019).

Ex vivo mathematical model for including the reapplication of radiofrequency currents in hepatic ablation

Antonio Romero [1],* Jose Bon[2] Macarena Trujillo[3]

(1) Department of Applied Mathematics, Universitat Politècnica de València
(2) Food Technology Department, Universitat Politècnica de València
(3) BioMIT, Electronic Engineering Department,Universitat Politècnica de València

1 Introduction

Radiofrequency ablation (RFA) is a minimally invasive technique used for the treatment of tumors in various locations [1]. Its main feature is the ability to increase tissue temperature intentionally causing thermal damage by creating a lesion, referred to as the coagulation zone (CZ), that ultimately destroys the tumor. The CZ typically has an ellipsoidal shape due to the geometry of the devices used to apply the RF currents [1]. However, research efforts are currently focused on achieving more spherical CZs, as tumors usually have this shape. Scientific advances have been made in this scope, including the development of various protocols and power settings, as well as design modifications in the devices themselves [2], but achieving a spherical CZ remains as a goal in RFA.

Due to these limitations, in clinical practice, overlapping is a commonly used procedure by surgeons and radiologists to achieve a greater transversal diameter of the CZ that can destroy the entire tumor. Overlapping consists of applying RFA to the tumor in successive steps: apply RF currents once, then remove the device, and reapply again by positioning the device perpendicular to the area where the first application was done (Figure 1). Although this procedure is common in clinical settings, there is no mathematical model that reproduces it. The model is complex since it needs to consider changes that occur in geometry, initial and boundary conditions, as well as in the material properties. Developing a model that can quantify the increase in CZ volume using overlapping procedure is of great interest in clinical practice, as it would help assess the true effectiveness of the method.

Therefore, the main goal of this study is to build an ex vivo mathematical model to mimic the overlapping procedure in hepatic RFA, that can be compare with experimental ex vivo resoults. To achieve this, we consider how the tissue properties change with temperature, we incorporate reversible damage to electrical conductivity, we consider the vaporization that occurs when temperatures exceed 100 ^{o}C during RFA. And, the most important feature, we address the geometric and boundary condition changes that result from the reapplication of radiofrequency currents.

*aromtar@mat.upv.es

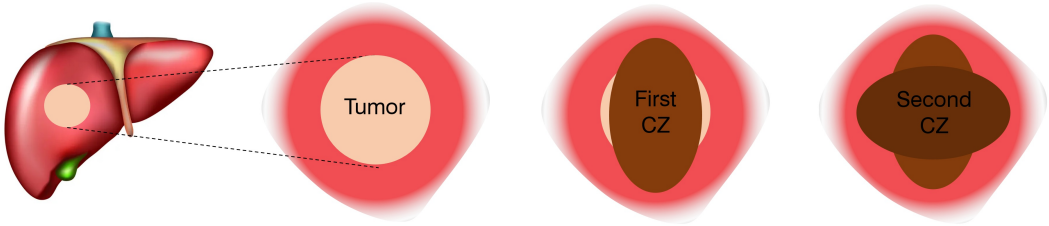


Figure 1: Scheme of the steps of overlapping procedure.

2 Mathematical model

2.1 Model domains

The geometry involves three domains: 1) the liver tissue, 2) the plastic shaft of the device, and 3) the metallic part of the device that corresponds to the electrode through which the RF currents are applied and is in contact with the tissue (Figure 2A). The device we use is a Cool-tip applicator (Covidien, Boulder, CO, USA), which is a conventionally cooled, needle-like device, 1.5 mm in diameter, with a circular section and a 3 cm tip. To model the overlapping we consider a change in geometry between the first and second RF currents applications. As it is shown in Figure 2, in the second RF application (Figure 2B) the device is introduced perpendicular to the first one (Figure 2A) and aligned in the center of the electrode. It is needed to consider a three-dimensional geometry, but due to the symmetry plane only a half of the geometry is represented as shown Figure 2.

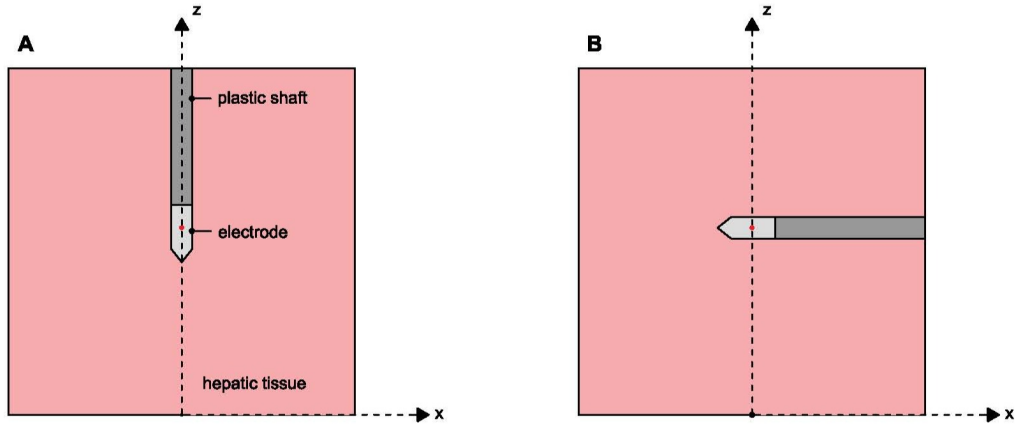


Figure 2: Schemes of the geometries used in the model. A. First RF current application. B. Second RF current application. Red point in the middle of the electrode is the rotation point of the device. Schemes out of scale.

2.2 Governing equations and conditions

The governing equation of the thermal problem is the Bioheat Equation modified by the enthalpy method to include the intratissue water vaporization [3]:

$$\frac{\partial(\rho h)}{\partial t} = \nabla \cdot (k \nabla T) + q + Q_p \tag{1}$$

where $\rho\,(kg/m^3)$ is tissue density, $h\,(J/kg\cdot K)$ the enthalpy, $k\,(W/m\cdot k)$ the thermal conductivity, $T\,(^oC)$ the temperature and $t\,(s)$. Q_p represents the heat loss by blood perfusion, which is considered zero since this model because it is an ex vivo model to compare with ex vivo experimental results. The heat source denoted by q, originates from the electrical problem, serving as the coupling term between the electrical and thermal physics. Specifically, $q = \sigma|E|^2$, where E is the electric field obtained from the electrial problem and σ is the electrial conductivity. The following equation is the governing equation for the electric problem

$$\nabla \cdot \sigma \nabla V = 0 \tag{2}$$

V being the voltage. The electric field is calculated from $E = -\nabla V$.
The entalphy is related with the temperature using the enthalpy method as [3]:

$$\frac{\partial(\rho h)}{\partial t} = \frac{\partial T}{\partial t} \cdot \begin{cases} \rho_l c_l & 0 < \quad T \quad \leq 99^oC \\ h_{fg} C & 99 < \quad T \quad \leq 100^oC \\ \rho_g c_g & \quad T \quad > 100^oC \end{cases} \tag{3}$$

where ρ_i and c_i are the density and specific heat for the tissue when the intra-tissue water is in liquid phase (i=l) and in gas phase (i=g), h_{fg} is the product of the latent heat of vaporization and the water density at 100 °C, and C is the water content in the liver (68 %).
To perform the calculations, we use the material properties given by Table 1.

Table 1: Material properties (Tissue properties assessed at 37°C)

Hepatic tissue	$\rho(kg/m^3)$	$\sigma(S/m)$	$c(J/kg\cdot K)$	$k(W/m\cdot K)$	$\omega(1/s)$
Healthy tissue	1080	0.26	3455	0.502	0.016
Electrode	8000	$7.4\cdot 10^6$	480	18	
Plastic shaft	70	10^{-5}	1045	0.026	

ρ: density, σ: electrical conductivity, c: specific heat, k: thermal conductivity, ω blood perfusion.

To evaluate the tissue damage caused during the procedure, we use the Arrhenius damage model

$$\Omega(t) = \int_0^t A e^{-\frac{\Delta E}{R\cdot T}} d\tau \tag{4}$$

which provides a damage index (Ω), allowing us to estimate the percentage of cells that have been destroyed after the ablations. An Ω value greater than 4.6 indicates that approximately 99% of the cells have been irreversibly destroyed. Therefore, in regions where $\Omega = 4.6$, we can identify the CZ. We use the frequency factor $A = 7.39\cdot 10^{39} s^{-1}$ and activation energy $E = 2.577\cdot 10^5 J/mol$.
To evaluate the model, a zero-flux Neumann boundary condition is applied at the symmetry plane for thermal and electrical problems. For the thermal problem, a Dirichlet condition of constant T_0 temperature is applied at the outers of the geometry. T_0 is also equal to the initial temperature and equal to 21 °C. For the electrical problem, a boundary condition of V=0 V is applied at the lower limit to mimic the dispersive electrode (ground). A constant voltage boundary condition of 80 V is applied at the electrode. The time for the first RF current application is 110 s and for the second is 65 s. Between both applications, a time of 3 min of latency is used to mimic the time needed to extract the device after the first RF current application and introduce again the device for the second.
Given the nonlinearity of the governing equations and the complexity of the geometries involved, the model is solved computationally using COMSOL Multiphysics, a software based on the Finite Element Method.

3 Results

Figure 3A shows the temperature distribution after the first RFA at the symmetry plane. The black line encloses the CZ around the device, which has a visible ellipsoidal shape. Figure 3B shows the temperature distribution after the second ablation at the symmetry plane and the new area of the CZ (enclosed by the black line). Although the CZ is not spherical, it is observed that the area of damaged tissue is larger than the obtained in the first ablation.

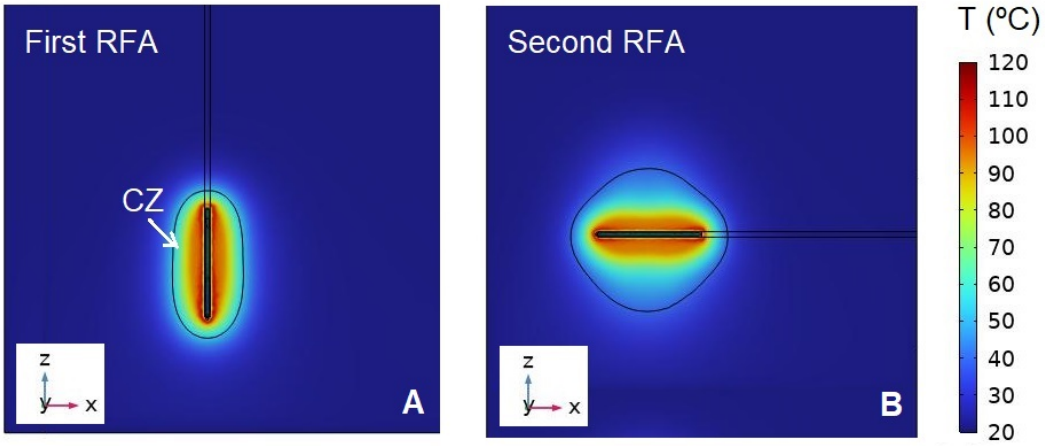


Figure 3: Temperature distribution in the symmetry plane at the end of the first (A) and second (B) ablations. Black line encloses the CZ.

4 Conclusions

The model successfully replicates the overlapping procedure in the RFA of liver tissue and confirms an increase in CZ volume. The second application enlarges de CZ, but its total volume is not the double of that obtained after the first application.

Acknowledgments

This research was funded by Spanish Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, Fondo Europeo de Desarrollo Regional (Grant PID2022-136273OBC31 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE).

References

[1] Berjano, E.J., Theoretical modeling for radiofrequency ablation: state-of-the-art and challenges for the future. *Biomedical engineering online*, 5(1),24, 2006.

[2] Kho, A. S., Foo, J. J., Ooi, E. T., Ooi, E. H., Shape-shifting thermal coagulation zone during saline-infused radiofrequency ablation: A computational study on the effects of different infusion location. *Computer methods and programs in biomedicine*, 184,105289, 2020.

[3] Abraham, J. P., Sparrow, E. M., A thermal-ablation bioheat model including liquid-to-vapor phase change, pressure-and necrosis-dependent perfusion, and moisture-dependent properties. *International Journal of Heat and Mass Transfer*, 50(13-14),2537-2544, 2007.

A proposal to model the transmission dynamics of the Respiratory Syncytial Virus using an Agent-Based Model

J. Villanueva-Oller[1,*], R. J. Villanueva[2], I. Hidalgo[3]

(1) Departamento de Informática y Estadística, Universidad Rey Juan Carlos, Av. del Alcalde de Móstoles, s/n, 28933 Móstoles, Spain.
(2) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, Spain.
(3) Departamento de Arquitectura de Computadores y Autómatas, Universidad Politécnica de Madrid, C\ Profesor José Garcia Santesmases, 9, 28040 Madrid, Spain.

Abstract

Respiratory Syncytial Virus (RSV) is a common and highly contagious virus that affects the respiratory tract. With symptoms similar to the common cold, it is a leading cause of lower respiratory infections in infants and young children. It is the primary cause of bronchiolitis and pneumonia in children under one year of age, accounting for 1.6% to 2.1% of RSV-associated hospitalizations in the overall population cohort.

The Public Health impact of RSV has prompted increased preventives measures against this disease. For example, in the Valencian Community (Spain), Nirsevimab, a long-acting monoclonal antibody, was introduced during the 2023–2024 RSV season as a universal pre-exposure prophylaxis for high-risk infants and all children under 6 months of age, showing a preliminary positive impact.

Under these premises, the use of mathematical models is of great interest for ensuring that vaccination and immunoprophylaxis strategies remain optimal over time and across different scenarios. Such models make it possible to explore alternative approaches, optimize resource allocation, and adapt to changes in virus dynamics or population conditions, thereby maximizing the impact on the reduction of pediatric hospitalizations.

Our proposal, among the various modeling approaches used to study the dynamics of RSV transmission and vaccination, is to use an agent-based modeling (ABM) framework to represent the population of the Valencian Community (CVA). The model incorporates age groups consistent with actual demographic data, as well as annual birth distributions and age-specific mortality rates.

The use of ABM allows randomness to be incorporated in a natural way, reflecting the inherent stochasticity of viral transmission. Once the model is established, it can be used to simulate vaccination strategies in which vaccine-induced protection wanes over time, with the duration of protection varying across individuals.

* Corresponding author Javier.villanueva@urjc.es

Mathematical modeling of the innate and adaptive immune systems in a viral infection

B. Chen-Charpentier[♭],[1]

(♭) Department of Mathematics, University of Texas at Arlington
Arlington, Texas 76019, USA.

1 Introduction

Viruses cause a large number of diseases in humans, animals and plants. In humans the viral diseases include the common cold (Rhinovirus/Enterovirus, Respiratory Adenoviruses), flu (Influenza), COVID-19 (SARS-CoV-2), measles, hepatitis, AIDS (Human Immunodeficiency Virus) and many others. The effects of a viral infection can be mild as for the flu to deadly as for COVID-19, AIDS and Ebola. Viruses are non-cellular entities. Each virus particle or virion consists of a nucleic acid core, an outer protein coating, and sometimes an outer envelope [9]. They replicate by penetrating a healthy cell, and using the cell's RNA or DNA to make copies of its genetic material, and then burst the cell producing many new virions. Vaccines and antivirals can be used to fight the infection. Vaccines, usually administered before the individual gets the disease, train the immune system to fight the virus. Antivirals are medications that reduce the effects and the duration of the disease. Rather than killing a virus directly, antivirals usually suppress the virus's ability to infect and multiply in the organism cells [28]. The human body has physical barriers such as the skin and cilia in the respiratory tract. But viruses can overcome these barriers. Then the innate immune system responds with inflammation, virus phagocytosis, and secretion of cytokines, chemical factors and proteins [9]. The innate immune system is non-specific and responds to the presence of any debris or pathogen. The adaptive immune system takes longer to act than the innate system but is more specific and has memory. Within-host mathematical models can help understand the very complex interactions between the different cells, chemical substances and virions. A basic model consists of three populations, susceptible cells, infected cells and virions. Their interactions and rates of change are given by ordinary differential equations [23, 25, 1]. An improvement is to add effector cells to kill the infected cells [22, 15] or to kill the virus [21, 18]. Both processes are added without giving explicit equations for the immune system in [13]. More elements of the immune system are included in [2, 4, 14]. There are other processes that have been observed for some virus within-host infections. One is the presence of cell to cell infection [11, 26]. Delays are also added in many papers including [27, 33, 34]. The effects of treatment are studied in [31, 3, 6, 32]. In this paper we introduce and analyze a new within-host virus infection model, based on ordinary differential equations. It includes healthy cells, infected cells, virions, and innate and adaptive immune system cells. It also includes cell to cell infection. Two more simple previously published models are included in the study for comparison purposes. Finally numerical simulations for a model of hepatitis B are done, first with no treatment and then with treatment.

[1] bmchen@uta.edu

2 Methods

The first within-host virus infection model we study only includes three populations: susceptible cells, infected cells, and virus particles (virions) [23, 25, 1]. Susceptible cells are recruited and die naturally, and they become infected by invasion of a virus particle. The virus replicates inside the cell and then the cell bursts and releases a certain number of virions. These free virions can infect healthy cells or die. This death process is not specified but it is implied to be caused by the immune system. All parameters are non-negative. The equations that describe the model are:

$$\begin{aligned} \frac{dx}{dt} &= \lambda - \mu_x x - \beta x v \\ \frac{dy}{dt} &= \beta x v - \mu_y y \\ \frac{dv}{dt} &= B\mu_y y - \mu_v v. \end{aligned} \tag{1}$$

Here, x, y, and v, are the concentrations of susceptible cells, infected cells, and virions, respectively. λ is the recruitment rate of new susceptible cells, μ_x is their death rate, β is the infection rate, μ_y is the bursting rate of infected cells by the virus, B is the number of virions produced per infected cell, and μ_v is the death rate of virions due to the immune system. All parameters are non-negative.

A second more realistic model also including the elimination of infected cells and virions by the effector cells of the immune system, as well as cell-to-cell transmission is based on [8, 19, 26, 34] and introduced in [5]. In this model which are the effector cells is unspecified. The equations of the model are

$$\begin{aligned} \frac{dx}{dt} &= \lambda - \mu_x x - \beta x v - \beta_{xy} x y \\ \frac{dy}{dt} &= \beta x v + \beta_{xy} x y - \beta_{ey} y e - \mu_y y \\ \frac{dv}{dt} &= B\mu_y y - \beta x v - \beta_{ev} v e - \mu_v v \\ \frac{de}{dt} &= s + \alpha_1 y - \beta_{ye} y e - \beta_{ve} v e - \mu_e e. \end{aligned} \tag{2}$$

Here, x, y, v, and e are the concentrations of susceptible cells, infected cells, virions, and effector cells, respectively. λ is the recruitment rate of new susceptible cells, μ_x is their death rate, β is the infection rate of susceptible cells by free floating virions, μ_y is the killing rate of infected cells by the virus, B is the number of virions produced per infected cell, and μ_v is the death rate of virions. Susceptible cells can also be infected through direct contact with an infected cell at a rate β_{xy}; effector cells eliminate infected cells at a rate β_{ey}; free virus particles are introduced into susceptible cells at a rate β; and effector cells are recruited at a constant rate s and are also recruited depending on the number of infected cells y at a rate α_1. In addition, effector cells eliminate infected cells and virus particles at rates β_{ye} and β_{ve}, respectively. Finally, effector cells die naturally at a rate μ_e. All parameters are non-negative.

But it is not realistic to consider only one cell type from the immune system since they have different jobs and act at different times. The elimination of the free virions is done by both the innate and adaptive immune systems while the killing of the infected cells is done mostly by the adaptive system. The innate system acts first and then recruits the adaptive system. Herbert in [16] lists neutrophils and natural killer cells as the most important innate cells in fighting virus infections while natural killer T cells are possibly relevant. For the adaptive immune system he has B or plasma cells, cytotoxic T lymphocytes CD4+, and cytotoxic T lymphocytes CD8+. He

mentions several signaling cytokines and factors but only the antibodies as killing agents. He just briefly mentions macrophages but other authors state that they play very important roles [20, 20, 30]. Therefore the variables for the third model are: x susceptible cells, y infected cells, v virons, N neutrophils, m macrophages, K natural killer cells, C_4 CD4+ cells, P plasma cells, C_8 CD8+ cells, a antibodies.

$$\begin{aligned}
\frac{dx}{dt} &= \lambda - \mu_x x - \beta x v - \beta_{xy} x y \\
\frac{dy}{dt} &= \beta x v + \beta_{xy} x y - \beta_{my} m y - \beta_{Ky} K y - \beta_{C_8 y} C_8 y - \mu_y y \\
\frac{dv}{dt} &= B\mu_y y - \beta x v - \beta_{av} a v - \beta_{Nv} N v - \mu_v v \\
\frac{dN}{dt} &= k_N N v - \mu_N N \\
\frac{dm}{dt} &= k_m(m_0 - m) + k_{mv} m v - \beta_{my} m y \\
\frac{dK}{dt} &= k_K m v - \mu_K K \\
\frac{C_4}{dt} &= k_{C_4} K - \mu_{C_4} C_4 \\
\frac{dP}{dt} &= k_P C_4 - \mu_P P \\
\frac{C_8}{dt} &= k_{C_8} - \mu_{C_8} C_8 \\
\frac{da}{dt} &= k_a(a_0 - a) + k_{aP} P - \beta_{va} a v.
\end{aligned} \tag{3}$$

The parameters not include in the previous model are: β_{Ky} the elimination rate of infected cells by neutrophils; $\beta_{C_8 y}$ the elimination rate of infected cells by CD8+ cells; β_{av} the elimination rate of virions by antibodies; β_{Nv} the elimination rate of virions by natural killer cells; k_N the recruitment rate of neutrophils due to the presence of the virus; μ_N the death rate of neutrophils; k_m the natural replenishing rate of macrophages; m_0 the steady state of macrophages in the absence of an infection; k_{mv} the recruitment rate of macrophages due to the virus; β_{my} the elimination of macrophages due to engulfing an infected cell; k_K the activation rate of natural killer cells; μ_K the deactivation rate of natural killer cells; k_{C_4} the activation rate of CD4+ cells; μ_{C_4} the deactivation rate of CD4+ cells; k_P the activation rate of plasma cells; μ_P the deactivation rate of plasma cells; k_{C_8} the activation rate of CD8+ cells; μ_{C_8} the deactivation rate of CD8+ cells; k_a the production rate of antibodies in absence of an infection; a_0 the steady state of the antibodies in absence of an infection; k_{Pa} the production rate of antibodies by activated plasma cells; β_{va} the elimination rate of antibodies after reacting with a virus. All the new parameters are non-negative.

There are some mathematical models that consider different types of immune cells [2, 14, 6] but none has the same immune cells that we do.

There are several different types of antiviral treatments: some block receptors so the virions cannot get into cells, and some lower the viral load by reducing the number of virus copies from an infected cell [9, 10]. Vaccines boost the immune system. They are usually administered before an infection, but they can sometimes be administered after the infection for some slow developing diseases, for example rabies. So we will also consider them as a treatment. Several authors include treatment in their models [24, 31, 7]. The effect of blocking receptors is done by replacing the infection rate β in the equations by $(1-\epsilon)\beta$. The effect of reducing the viral load by multiplying the number of virions produced by a single cell B by $(1-\eta)$. Both ϵ and η are between 0 and

1.To our knowledge the effect of strengthening the immune system has not been introduced. We will model it by multiplying the killing rate of virions and infected cells by effector cells by a factor $(1-\delta)$, with *delta* between 0 and 1. We will also consider that there is a delay of several days tt between the start of the infection and the administration and action of the antivirals. For example, for the first type of antiviral the factor is $(1-H(t-tt)\epsilon)$ with H the Heaviside function.

3 Results

We analyzed the three models. For the first model we repeated the analysis done by others [23, 25, 1] and found the two steady solutions (the disease free DFE and a chronic one), their local stability and also calculated the basic reproductive number R_0, number of susceptible cells infected by a single infected cell at the beginning of the infection [17]. R_0 is given by

$$R_0 \propto \left(\frac{infection}{contact}\right) \times \left(\frac{contact}{time}\right) \times \left(\frac{time}{infection}\right).$$

There are several methods of calculating R_0. The most widely used approach is the next generation matrix method [12, 29]. R_0 is easier to compute than the eigenvalues of the linearized system since its calculation only involves the infected compartments, y and v. The most important information a mathematical model of infection can give is whether the infection will persist which is for $R_0 > 1$. For the second model, the equilibrium solutions can still be calculated as are their local stability. But the formulas are complicated. The basic reproductive number has a simpler formula and is easier to interpret. For the third model only the DFE and R_0 can be calculated. For all models the numerical simulations give very useful information. We did simulations without treatment and then with treatment. Since the data available has large variations due to differences in the individuals, errors and difficulties in the measurements, different estimation methods, it is very important to do a global sensitivity analysis. This way the relative influence of each parameter on the solutions can be established.

4 Conclusions

The immune system is very complex. There are many different types of cells and factors that interact with themselves, the virus, and healthy and infected cells. Including more processes makes the model more realistic, but in practice it is necessary to limit their number, both to simplify the model and better understand the most important processes. The fact that the number of parameters and data available is limited also restricts the complexity of the model. In this paper we only considered one factor, antibodies, since thy kill virions directly. The other factors and cytokines involved in signaling were assumed to have concentrations proportional to the number of cells that produce them. Also for simplicity we did not include some immune cells such as inactivated macrophages. We did comparisons between three models that have different complexities, and studied the effect of introducing antivirals. We also used global sensitivity analysis to establish the parameters that cause the biggest changes in the number of infected cells. This way directions to better measure them can be developed. But most importantly it can encourage research in how to change these parameters to decrease the damage from the viral infection.

References

[1] Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., and Perelson, A. S. Kinetics of influenza a virus infection in humans. *Journal of virology 80*, 15 (2006), 7590–7599.

[2] Bocharov, G. A., and Romanyukha, A. A. Mathematical model of antiviral immune response iii. influenza a virus infection. *Journal of Theoretical Biology 167*, 4 (1994), 323–360.

[3] Canini, L., and Perelson, A. S. Viral Kinetic Modeling: State of the Art. *Journal of pharmacokinetics and pharmacodynamics 41*, 5 (Oct. 2014), 431–443.

[4] Cao, P., Wang, Z., Yan, A. W. C., McVernon, J., Xu, J., Heffernan, J. M., Kedzierska, K., and McCaw, J. M. On the Role of CD8+ T Cells in Determining Recovery Time from Influenza Virus Infection. *Frontiers in Immunology 7* (Dec. 2016), 611.

[5] Chen-Charpentier, B. On population models with delays and dependence on past values. *Axioms 13*, 3 (2024), 206.

[6] Chimal-Eguia, J. C. Mathematical Model of Antiviral Immune Response against the COVID-19 Virus. *Mathematics 9*, 12 (Jan. 2021), 1356.

[7] Ciupe, S. M. Modeling the dynamics of hepatitis b infection, immunity, and drug therapy. *Immunological Reviews 285*, 1 (2018), 38–54.

[8] Ciupe, S. M., Ribeiro, R. M., Nelson, P. W., Dusheiko, G., and Perelson, A. S. The role of cells refractory to productive infection in acute hepatitis b viral dynamics. *Proceedings of the National Academy of Sciences 104*, 12 (2007), 5050–5055.

[9] Clark, M. A., Douglas, M., and Choi, J. *Biology 2e*. Openstax, 2018.

[10] Clinic, C. What Do Antivirals Treat? https://my.clevelandclinic.org/health/treatments/antivirals, 2025.

[11] Culshaw, R. V., Ruan, S., and Webb, G. A mathematical model of cell-to-cell spread of hiv-1 that includes a time delay. *Journal of mathematical biology 46*, 5 (2003), 425–444.

[12] Diekmann, O., Heesterbeek, J., and Roberts, M. G. The construction of next-generation matrices for compartmental epidemic models. *Journal of the royal society interface 7*, 47 (2010), 873–885.

[13] Du, S. Q., and Yuan, W. Mathematical modeling of interaction between innate and adaptive immune responses in covid-19 and implications for viral pathogenesis. *Journal of medical virology 92*, 9 (2020), 1615–1628.

[14] Gómez, M. C., and Yang, H. M. Mathematical model of the immune response to dengue virus. *Journal of Applied Mathematics and Computing 63*, 1-2 (June 2020), 455–478.

[15] Hattaf, K., Yousfi, N., 1 Centre Régional des Métiers de l'Education et de la Formation (CRMEF), 20340 Derb Ghalef, Casablanca, Morocco, and 2 Laboratory of Analysis, Modeling and Simulation (LAMS), Faculty of Sciences Ben M'sik, Hassan II University of Casablanca, P.O Box 7955 Sidi Othman, Casablanca, Morocco. Dynamics of SARS-CoV-2 infection model with two modes of transmission and immune response. *Mathematical Biosciences and Engineering 17*, 5 (2020), 5326–5340.

[16] Herbert, J. A., and Panagiotou, S. Immune Response to Viruses. In *Encyclopedia of Infection and Immunity*. Elsevier, 2022, pp. 429–444.

[17] Jones, J. H. Notes on r0. department of anthropological sciences, 2007.

[18] Kajiwara, T., and and Sasaki, T. Global stability of pathogen-immune dynamics with absorption. *Journal of Biological Dynamics 4*, 3 (May 2010), 258–269.

[19] Kim, H. Y., Kwon, H.-D., Jang, T. S., Lim, J., and Lee, H.-S. Mathematical modeling of triphasic viral dynamics in patients with hbeag-positive chronic hepatitis b showing response to 24-week clevudine therapy. *PLoS One 7*, 11 (2012), e50377.

[20] Meischel, T., Villalon-Letelier, F., Saunders, P. M., Reading, P. C., and Londrigan, S. L. Influenza A virus interactions with macrophages: Lessons from epithelial cells. *Cellular microbiology 22*, 5 (2020), e13170.

[21] Murase, A., Sasaki, T., and Kajiwara, T. Stability analysis of pathogen-immune interaction dynamics. *Journal of Mathematical Biology 51*, 3 (Sept. 2005), 247–267.

[22] Nowak, M. A., and Bangham, C. R. M. Population Dynamics of Immune Responses to Persistent Viruses. *Science 272*, 5258 (Apr. 1996), 74–79.

[23] Nowak, M. A., Bonhoeffer, S., Hill, A. M., Boehme, R., Thomas, H. C., and McDade, H. Viral dynamics in hepatitis B virus infection. *Proceedings of the National Academy of Sciences 93*, 9 (Apr. 1996), 4398–4402.

[24] Perelson, A. S. Modelling viral and immune system dynamics. *Nature Reviews Immunology 2*, 1 (Jan. 2002), 28–36.

[25] Perelson, A. S., and Nelson, P. W. Mathematical Analysis of HIV-1 Dynamics in Vivo. *SIAM Review 41*, 1 (Jan. 1999), 3–44.

[26] Pourbashash, H., Pilyugin, S. S., De Leenheer, P., and McCluskey, C. Global analysis of within host virus models with cell-to-cell viral transmission. *Discrete Contin. Dyn. Syst. Ser. B 19*, 10 (2014), 3341–3357.

[27] Shu, H., Wang, L., and Watmough, J. Global Stability of a Nonlinear Viral Infection Model with Infinitely Distributed Intracellular Delays and CTL Immune Responses. *SIAM Journal on Applied Mathematics 73*, 3 (Jan. 2013), 1280–1302.

[28] Taveira, N. Antivirals and Vaccines. *International Journal of Molecular Sciences 24*, 12 (June 2023), 10315.

[29] Van den Driessche, P. Reproduction numbers of infectious disease models. *Infectious Disease Modelling 2*, 3 (2017), 288–303.

[30] Wang, X.-L., Wang, X., and Ho, W.-Z. Roles of Macrophages in Viral Infections. *Viruses 16*, 10 (Oct. 2024), 1643.

[31] Wang, Z., and Zhao, X.-Q. A Within-Host Virus Model with Periodic Multidrug Therapy. *Bulletin of Mathematical Biology 75*, 3 (Mar. 2013), 543–563.

[32] Xu, Z., Song, J., Zhang, H., Wei, Z., Wei, D., Yang, G., Demongeot, J., and Zeng, Q. A mathematical model simulating the adaptive immune response in various vaccines and vaccination strategies. *Scientific Reports 14*, 1 (Oct. 2024), 23995.

[33] Yang, Y., Zou, L., and Ruan, S. Global dynamics of a delayed within-host viral infection model with both virus-to-cell and cell-to-cell transmissions. *Mathematical Biosciences 270* (Dec. 2015), 183–191.

[34] Zhang, S., Li, F., and Xu, X. Dynamics and control strategy for a delayed viral infection model. *Journal of Biological Dynamics 16*, 1 (2022), 44–63.

Mathematical and epidemiological model of dengue virus transmission in Ecuador and Puerto Rico

Juan Ángel Aledo[1], Mario Alberto Espinoza[2], Juan Carlos Orengo[3 *], Rafael-J. Villanueva[4]

(1) Departamento de Matemáticas, Universidad de Castilla-La Mancha, Albacete, Spain
(2) Departament of Computer Science, Universidad Internacional del Ecuador, Quito, Ecuador
(3) Programa de Salud Pública, Ponce Health Science University, Ponce, Puerto Rico
(4) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain.

1. Introduction

The dengue virus (DENV) belongs to the genus *flavivirus*, family *Flaviviridae* and has four serotypes (DENV 1, DENV 2, DENV 3 and DENV 4), each of these serotypes has different genotypes. DENV causes dengue fever in humans, which is a vector-borne disease. The main vectors are *Aedes aegypti* and *Aedes albopictus*, the female mosquito is the one that transmits the virus[1]. In Ecuador, *Aedes aegypti* and *Aedes albopictus* are endemic[2,3]. In relation to Puerto Rico, the main vector is *Aedes aegypti* and *Aedes mediovittatus* is a competent vector[4,5]. Dengue is classified as a neglected disease and in relation to the 2030 Sustainable Development Goals it is directly included in Goal 3 for Good Health and Well-being[6]. In the Region of Las Americas about 13,000,000 cases of dengue were reported in 2024. Ecuador reported around 61,000 cases (incidence rate 338.18 x 100,000) and Puerto Rico more than 6,000 cases (incidence rate 187.79 x 100,000)[7].

The main objective of our study is to determine the most relevant factors for an adjusted modeling of dengue transmission dynamics to be implemented in Ecuador and Puerto Rico.

2. Results

There are different factors that influencing the transmission and dissemination of dengue. These factors can be classified into four categories: Humans, intrinsic to the pathogen, related to the vector, and finally those belonging to the environment (climatic and others).

We are currently working in the vector category (focusing on the *A. aegypti*), and how the vector and climatic factors interact, defining parameters to include based on the impact on disease transmission and the availability of local data (Ecuador and Puerto Rico) or studies conducted in the field, not just at the experimental level.

* jorengo@psm.edu

The *A. aegypti* is an anthropophilic mosquito, endophilic, endophagic, bites during the day and feeds on multiple occasions. Transmission of the virus can be horizontal or vertical, a mechanical transmission is being assessed.

Aedes aegypti is a competent vector because it can become infected with and transmit DENV. Transmission of the virus can be horizontal or vertical; both routes will be incorporated into the model. The vector capacity of *A. aegypti* (virus transmission efficiency) is based on: 1) mosquito density, which depends on mating, fecundity, fertility, parity, viable eggs, larval survival (1, 2, 3,4), pupa, and adult mosquito (larval survival is dependent on the growing container, density, nutrition, competitors and predators; these factors are a limitation in the model); 2) biting rate (presents individual heterogeneity); 3) external incubation period (time that elapses from when A. aegypti ingests the virus until it reaches the salivary glands, between 5-14 days); 4) gonadotrophic cycle (includes locating the human, biting, ingesting blood (multiples bloodmeals), maturing the ovaries, searching for the place to lay the eggs and laying them), lasts 3-5 days; 5) lifespan[8-18].

Among the climatic variables, temperature, humidity, precipitation, and wind will impact the vectorial capacity of *A. aegypti*, primarily temperature, which will affect lifespan and size of adult, the gonadotrophic cycle, the external incubation period, mosquito activity (from zero to optimal) and others. The main limitation of climatic variables is the spatial variability[19-24].

3. Conclusions

These are the factors related with the vector that any dengue model that aims to explain its dynamics in an adjusted way should include: routes of transmisión (horizontal and vertical), characteristics of vector capacity (density, reproductive factors, life stages, mortality, daily survival, lifespam, behavior, external incubation period, gonadotrophic cycle, biology), and climate variables (temperature, humidty, precipitation)

References

[1] Paz-Bailey G, Adams LE, Deen J, Anderson KB, Katzelnick LC. Dengue. Lancet. 2024 Feb 17;403(10427):667-682. doi: 10.1016/S0140-6736(23)02576-X. Epub 2024 Jan 24. PMID: 38280388.
[2] Lippi CA, Stewart-Ibarra AM, Loor MEFB, Zambrano JED, Lopez NAE, Blackburn JK, Ryan SJ. Geographic shifts in Aedes aegypti habitat suitability in Ecuador using larval surveillance data and ecological niche modeling: Implications of climate change for public health vector control. PLoS Negl Trop Dis. 2019 Apr 17;13(4):e0007322. doi: 10.1371/journal.pntd.0007322. PMID: 30995228; PMCID: PMC6488096.
[3] Garcia-Rejon JE, Navarro JC, Cigarroa-Toledo N, Baak-Baak CM. An Updated Review of the Invasive *Aedes albopictus* in the Americas; Geographical Distribution, Host Feeding Patterns, Arbovirus Infection, and the Potential for Vertical Transmission of Dengue Virus. Insects. 2021 Oct 26;12(11):967. doi: 10.3390/insects12110967. PMID: 34821768; PMCID: PMC8621292.

[4] Moore CG, Cline BL, Ruiz-Tibén E, Lee D, Romney-Joseph H, Rivera-Correa E. Aedes aegypti in Puerto Rico: environmental determinants of larval abundance and relation to dengue virus transmission. Am J Trop Med Hyg. 1978 Nov;27(6):1225-31. doi: 10.4269/ajtmh.1978.27.1225. PMID: 727328.
[5] Otero LM, Medina-Martinez G, Sepúlveda M, Acevedo V, Toro M, Barrera R. Cemeteries as sources of Aedes aegypti and other mosquito species in southeastern Puerto Rico. Trop Med Int Health. 2022 Mar;27(3):300-309. doi: 10.1111/tmi.13723. Epub 2022 Mar 2. PMID: 35118778; PMCID: PMC8954491.
[6] Naciones Unidas (2018), La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe (LC/G.2681-P/Rev.3), Santiago
[7] Panamerican Health Organization (PAHO), Reported Cases of Dengue Fever in The Americas, Retrieved from: https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html, Accessed 04/07/2025
[8] Li HH, Su MP, Wu SC, Tsou HH, Chang MC, Cheng YC, Tsai KN, Wang HW, Chen GH, Tang CK, Chung PJ, Tsai WT, Huang LR, Yueh YA, Chen HW, Pan CY, Akbari OS, Chang HH, Yu GY, Marshall JM, Chen CH. Mechanical transmission of dengue virus by Aedes aegypti may influence disease transmission dynamics during outbreaks. EBioMedicine. 2023 Aug;94:104723. doi: 10.1016/j.ebiom.2023.104723. Epub 2023 Jul 22. PMID: 37487418; PMCID: PMC10382859.
[9] Scott TW, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Zhou H, Edman JD. Longitudinal studies of Aedes aegypti (Diptera: Culicidae) in Thailand and Puerto Rico: population dynamics. J Med Entomol. 2000 Jan;37(1):77-88. doi: 10.1603/0022-2585-37.1.77. PMID: 15218910.
[10] Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Edman JD. Longitudinal studies of Aedes aegypti (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. J Med Entomol. 2000 Jan;37(1):89-101. doi: 10.1603/0022-2585-37.1.89. PMID: 15218911.
[11] Harrington LC, Buonaccorsi JP, Edman JD, Costero A, Kittayapong P, Clark GG, Scott TW. Analysis of survival of young and old Aedes aegypti (Diptera: Culicidac) from Puerto Rico and Thailand. J Med Entomol. 2001 Jul;38(4):537-47. doi: 10.1603/0022-2585-38.4.537. PMID: 11476334.
[12] Styer LM, Minnick SL, Sun AK, Scott TW. Mortality and reproductive dynamics of Aedes aegypti (Diptera: Culicidae) fed human blood. Vector Borne Zoonotic Dis. 2007 Spring;7(1):86-98. doi: 10.1089/vbz.2007.0216. PMID: 17417961.
[13] Chan M, Johansson MA. The incubation periods of Dengue viruses. PLoS One. 2012;7(11):e50972. doi: 10.1371/journal.pone.0050972. Epub 2012 Nov 30. PMID: 23226436; PMCID: PMC3511440.
[14] Ponlawat A, Harrington LC. Factors associated with male mating success of the dengue vector mosquito, Aedes aegypti. Am J Trop Med Hyg. 2009 Mar;80(3):395-400. PMID: 19270288.
[15] Théo Maire, Louis Lambrechts, Felix J.H. Hol. Arbovirus impact on mosquito behavior: the jury is still out. Trends in Parasitology, 2024, 40 (4), pp.292-301. 10.1016/j.pt.2024.02.004. pasteur-04572547
[16] Pruszynski CA, Buckner EA, Burkett- Cadena ND, Hugo LE, Leal AL, Caragata EP (2024) Estimation of population age structure, daily survival rates, and potential to support dengue virus transmission for Florida Keys Aedes aegypti via transcriptional profiling. PLoS Negl Trop Dis 18(8): e0012350. https://doi.org/10.1371/journal. pntd.0012350
[17] Extrinsic Incubation Period Correlate with Varying Dengue Risk in the Arizona–Sonora Desert Region. Viruses 2023, 15, 851. https://doi.org/10.3390/v15040851
[18] Facchinelli L, Badolo A, McCall PJ. Biology and Behaviour of *Aedes aegypti* in the Human Environment: Opportunities for Vector Control of Arbovirus Transmission. Viruses. 2023 Feb 27;15(3):636. doi: 10.3390/v15030636. PMID: 36992346; PMCID: PMC10053764.

[19] Goindin D, Delannay C, Ramdini C, Gustave J, Fouque F. Parity and longevity of Aedes aegypti according to temperatures in controlled conditions and consequences on dengue transmission risks. PLoS One. 2015 Aug 10;10(8):e0135489. doi: 10.1371/journal.pone.0135489. PMID: 26258684; PMCID: PMC4530937.
[20] Gloria-Soria A, Armstrong PM, Powell JR, Turner PE. Infection rate of *Aedes aegypti* mosquitoes with dengue virus depends on the interaction between temperature and mosquito genotype. Proc Biol Sci. 2017 Oct 11;284(1864):20171506. doi: 10.1098/rspb.2017.1506. PMID: 28978730; PMCID: PMC5647301.
[21] Barr JS, Estevez-Lao TY, Khalif M, Saksena S, Yarlagadda S, Farah O, Shivere Y, Hillyer JF. Temperature and age, individually and interactively, shape the size, weight, and body composition of adult female mosquitoes. J Insect Physiol. 2023 Jul;148:104525. doi: 10.1016/j.jinsphys.2023.104525. Epub 2023 May 24. PMID: 37236342.
[22] Christofferson RC, Mores CN. Potential for Extrinsic Incubation Temperature to Alter Interplay Between Transmission Potential and Mortality of Dengue-Infected Aedes aegypti. Environ Health Insights. 2016 Jul 25;10:119-23. doi: 10.4137/EHI.S38345. PMID: 27478382; PMCID: PMC4961054.
[23] Ernst, K.C.;Walker, K.R.; Castro-Luque, A.L.; Schmidt, C.; Joy, T.K.; Brophy, M.; Reyes-Castro, P.; Díaz-Caravantes, R.E.; Encinas, V.O.; Aguilera, A.; et al. Differences in Longevity and Temperature-Driven
[24] Zapletal J, Erraguntla M, Adelman ZN, Myles KM, Lawley MA. Impacts of diurnal temperature and larval density on aquatic development of Aedes aegypti. PLoS One. 2018 Mar 7;13(3):e0194025. doi: 10.1371/journal.pone.0194025. PMID: 29513751; PMCID: PMC5841800.
[25] Sukiato F, Wasserman RJ, Foo SC, Wilson RF, Cuthbert RN. The effects of temperature and shading on mortality and development rates of Aedes aegypti (Diptera: Culicidae). J Vector Ecol. 2019 Dec;44(2):264-270. doi: 10.1111/jvec.12358. PMID: 31729799.

Part V

Mathematical Models and Artificial Intelligence for Transport Engineering

Railway Superstructure Inspection System For Detecting Ballast Contamination Level Through Acoustic Signal Analysis

Laura Andrés[1,*], Julia I. Real[1], Anabel Arias[1] and Lisardo Manuel Fort[2]

(1) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, 46022, València (Spain)
(2) ROVER RAIL Empresa Constructora, S.A.
C/ de Cavanilles, 28, El Pla del Real, 46010, Valencia (Spain)

1 Introduction

Track stiffness is a parameter that controls the relationship between the dynamic loads caused by the passage of rolling stock and the energy it consumes. This stiffness depends on the condition of the railway track components and, fundamentally, on the ballast. Ballast contamination results from a deterioration mechanism caused by normal traffic loads, leading to the breakage and abrasion of the gravel, which directly affects track stiffness. Therefore, it is crucial to maintain the ballast in good condition and properly arranged in order to accommodate the expected increase in rail transport without contributing to greater infrastructure degradation or increasing the risk of derailment.

Among current ballast maintenance practices, it is common to tamp the upper 20 cm of the shoulder, which obscures the extent of contamination by addressing only the surface layer. Existing techniques or systems for determining the degree of ballast contamination include visual inspection, sample extraction, and the use of ground-penetrating radar. Unfortunately, no technique currently exists that enables the assessment of ballast contamination in an automatic, continuous, non-destructive, and cost-effective manner. This highlights the need to develop a system that simultaneously fulfills all four of these criteria.

Consequently, the objective of this project is to develop an innovative system for ballast assessment, capable of: (i) determining the degree of ballast contamination; (ii) identifying the necessary maintenance actions; and (iii) providing a near real-time management and analysis system.

* lauanlo@upv.es

The SONOBAL project (CPP2023-010518) is part of the 2023 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by ROVER RAIL Empresa Constructora, S.A. and the Polytechnic University of Valencia as a public research organisation

2 Methods

Regarding its operation, the system will emit a directional acoustic signal onto the ballast (within one-third octave bands), taking advantage of the ballast shoulder's sound absorption properties at mid and high frequencies [1]. Subsequently, the same system will measure the sound levels at the analyzed frequencies from the underside of the railway vehicle and compare them with values obtained from previously calibrated acoustic models. In this way, it will be possible to infer variations in the ballast's absorption coefficient and, therefore, in its degree of contamination. Finally, the system will transmit the collected data to a cloud-based computational server via a long-range wireless connection (3G/4G), where the calibrated acoustic models and algorithms will be executed in near real time. The diagram of the proposed operating system is illustrated in Figure 1.

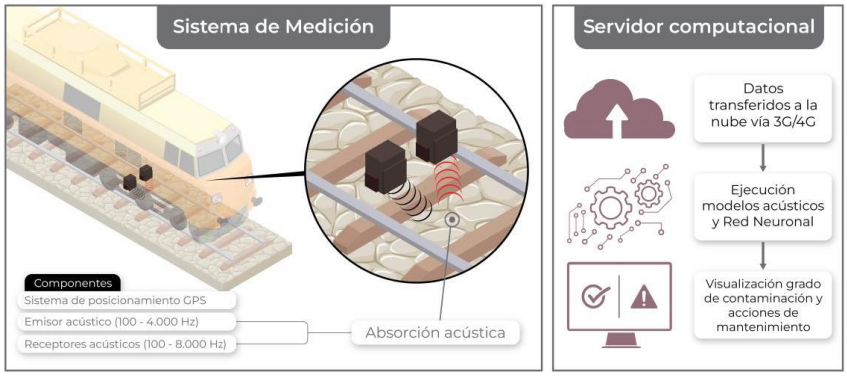


Figure 1. System Operation Diagram.

To achieve this objective, the project will be developed in several stages and will address various challenges:

First, the development of a software system to relate ballast sound absorption to its degree of contamination will be undertaken. This will be based on acoustic models calibrated through experimental testing. The scientific and technological foundation of SONOBAL is related to the acoustic absorption properties of the ballast shoulder, which have been studied since the 1940s, when they were first measured in a reverberation chamber by Kaye et al. [2]. More recently, Attenborough et al. conducted measurements of flow resistivity, density, and porosity

of ballast [3] to assess their influence on railway noise, while Broadbent et al. [1] analyzed the absorption coefficient of a ballast layer by modeling it as a three-layer porous medium.
These studies concluded that the ballast shoulder absorbs sound because it behaves as a porous material due to the discrete arrangement of the gravel particles [4]. When sound propagates through small spaces, such as the interconnected pores of a porous absorber, it loses energy because air is a viscous fluid; consequently, sound energy dissipates through friction with the pore walls [5]. As a result, the sound pressure level for slab track (with maximum stiffness) is approximately 4 dB higher than that of ballasted track [4]. Therefore, as ballast rigidity increases—i.e., as contamination progresses—it is expected that higher sound levels will be measured. This statement is graphically illustrated in Figure 2.

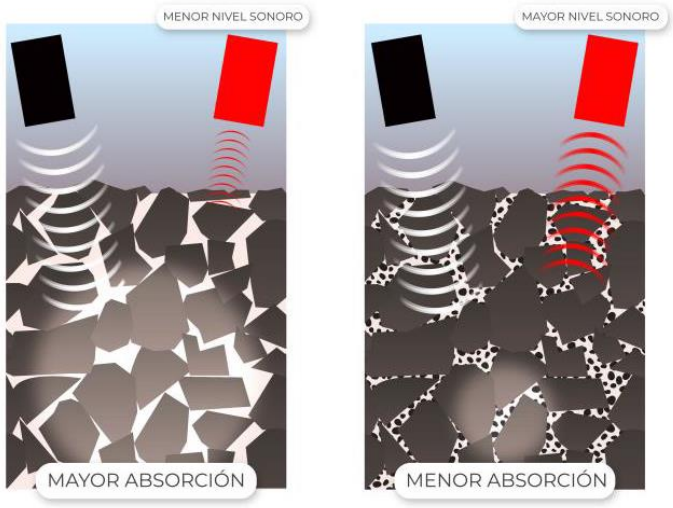


Figure 2. Acoustic absorption with clean ballast (left) and contaminated ballast (right).

In order to ensure that the acoustic model accurately represents the noise within the near sound field (the area where the system's acoustic receivers will be installed), it is proposed to use the three-dimensional wave equation along with the Johnson-Allard model [5], which governs the behavior of acoustic absorption in porous materials. Given that the ballast shoulder is composed of discrete gravel particles, it can be considered a porous material, thus validating the applicability of the aforementioned model.

To address this first challenge, experimental measurements will be conducted using a test bench with varying degrees of contamination, water and/or lubricant saturation, and ballast layer thicknesses. These tests will make it possible to determine the influence of each parameter affecting the acoustic absorption of ballast in relation to its degree of contamination. By fitting the trends observed during testing using a least squares (LSQ) method, relationship curves between each parameter and the ballast contamination level will be obtained. Finally, through the calibration of the acoustic models based on the experimental results, it will be possible to establish a definitive relationship between ballast absorption and its degree of contamination.

Following this, the hardware subsystem responsible for measuring the acoustic absorption of the ballast will be developed. For the acoustic emitter, it will be essential to study factors such as the frequency emission range (100 – 4,000 Hz), the maximum acoustic power (≥ 120 dB), and the directivity pattern, among others. Regarding the microphones, special attention will be given to sensitivity, maximum measurable sound level (≥ 120 dB), and bandwidth (100 – 8,000 Hz), among other characteristics.

"It is crucial to identify the optimal configuration of the hardware subsystem to ensure that the acoustic signal, after being absorbed by the ballast, can be accurately captured from beneath the railway vehicle. The goal of overcoming this challenge is to enable the system to detect variations in sound levels from the receiver, taking into account that it must operate at the working speed of the railway maintenance equipment on which it is installed. To address this second challenge, a test bench will be designed with varying degrees of contamination in a ballast layer, enabling experimental testing using measurement equipment. For this purpose, a scaled model of a railway vehicle will be designed and constructed to house the acoustic emitter and receiver, including rails and sleepers to account for their interaction during the absorption and/or reflection of the acoustic signal. Throughout these tests, the position and orientation of various acoustic emitters and receivers will be adjusted to observe how signal reception varies with different configurations. The optimal configuration will be the one that most clearly and effectively captures the acoustic absorption phenomenon.

Finally, a digital platform will be developed to integrate the entire system and automatically determine the required track maintenance actions based on the type of track, rail traffic, budget, and degree of contamination. To address this final challenge, it will be necessary to collect relevant information on ballast maintenance activities. By analyzing current regulations [6] and conducting surveys among expert groups from ROVER RAIL and the Department of Transport Engineering and Infrastructure at the School of Civil Engineering of the Polytechnic University of Valencia (UPV), a labeled database will be created specifying the various maintenance actions to be performed on degraded ballast. These maintenance actions will not depend solely on the degree of contamination, but also on other factors such as the type of traffic running over a given track segment or the budget allocated to that section.

The information collected through the surveys will serve as a dataset to train a Neural-Backed Decision Tree (NBDT) for decision-making. The NBDT architecture emulates the behavior of a decision tree using neural network structures. As a result, NBDTs provide higher accuracy and greater interpretability compared to classical decision trees [7].

Once the network has been trained, it will be integrated into the system to allow automatic determination of the required ballast maintenance actions upon obtaining its acoustic signature (i.e., the set of acoustic signals that characterizes a specific phenomenon). Based on the acoustic signature associated with a given section of railway track, the degree of ballast contamination in that segment can be determined.

Therefore, by combining the collected acoustic measurements with the operational characteristics of the track segment provided by ROVER RAIL (such as track type, traffic type, and available budget), the NBDT can be used to assess the necessary maintenance actions and their urgency.

Based on the above, the SONOBAL system will consist of the following subsystems:

- Hardware Subsystem. This will include an acoustic signal emitter and receiver operating within a frequency range of 100 to 4,000 Hz; a GPS positioning system to associate the inspection measurements with their corresponding kilometer points along the railway track; a long-range wireless communication system (3G/4G) to transmit data to the cloud for processing; and a storage system to ensure data logging in case of wireless signal loss.
- Software Subsystem. Responsible for processing all the information collected by the hardware. It will comprise advanced signal processing algorithms and calibrated acoustic models. This subsystem will enable the correlation between the received signal and the degree of ballast contamination.
- Digital Platform. It will visually and interactively display the degree of ballast contamination for each railway section, based on the information processed and calculated by the software subsystem.

3 Results

The implementation of the SONOBAL project will result in a system capable of optimally automating the quantification of ballast contamination levels, moving away from inefficient traditional techniques. This will lead to a reduction in both the costs and time required for inspection. Additionally, thanks to the integration of the system with a Neural Network for decision-making, the allocation of financial resources for ballast maintenance can be optimized.

A high scientific and technological impact is anticipated. The project will generate valuable information for key stakeholders, such as infrastructure managers, inspection companies, and researchers in the railway sector. Moreover, significant economic impact and strong market potential are expected, as the system will reduce the cost of inspecting ballast contamination and facilitate better planning and distribution of railway infrastructure maintenance activities

4 Conclusions

The SONOBAL project is an innovative proposal for the automatic, continuous, and non-destructive assessment of railway ballast contamination through acoustic analysis. By developing a system that combines acoustic signal emission and reception technology, physical-mathematical modelling, and advanced processing and artificial intelligence algorithms, the project establishes a new methodology that significantly surpasses the limitations of traditional inspection methods.

Overall, represents a major step forward in the digitalization and optimization of railway infrastructure maintenance, offering tangible benefits in terms of cost reduction, improved safety, and operational sustainability, with strong technological and economic impact potential in the sector.

References

[1] Broadbent, R. A., Thompson, D. J., Jones, C. J., The acoustic properties of railway ballast. *8th European Conference on Noise Control*, 2009.

[2] Kaye, G. W. C., Evans, E. J., The sound-absorbing properties of some common out-door materials. *Proceedings of the Physical Society*, 52(3): 371, 1940.

[3] Attenborough, K., Boulanger, P., Qin, Q., Jones, R., Predicted influence of ballast and porous concrete on rail noise. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 2005(4): 3547--3555, 2005.

[4] Zhuo-ming, L., Qi-liang, L., Zhi-gang, Y., Flow structure and far-field noise of high-speed train under ballast track. *Journal of Wind Engineering and Industrial Aerodynamics*, 220: 104858, 2022.

[5] Cox, T., D'Antonio, P., Acoustic Absorbers and Diffusers: Theory, Design and Application. CRC Press, 2016.

[6] RENFE. P.R.V. 3-4-0.0. Balasto. Pliego de prescripciones técnicas para el suministro y utilización del balasto. 2004.

[7] Wan, A., Dunlap, L., Ho, D., Yin, J., Lee, S., Jin, H., Petryk, S., Bargal, S. A., Gonzalez, J.E., NBDT: Neural-Backed Decision Tree. *9th International Conference on Learning Representations*, 2021.

Disease detection system in black truffle mycorrhized holm oak using unmanned aircraft (drones)

Reynaldo Cabezas[1, *], Julia I. Real[2], Gonzalo Muinelo[2], Vicente Monfort[1] and Roberto Herrera[1]

(1) IDVIA 2020 HORIZONTE 2020 S.L., Plaza de Bandas de Musica 5, Local 11, 46013, Valencia (Spain)
(2) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València,
46022, València (Spain)

1 Introducción:

Black truffle is the result of a fungal fructification (T. melanosporum) which establishes symbiotic relationships (mycorrhized) with several plant species [1], however, due to its climatic and edaphic requirements, in Spain the symbiote mostly used in truffle plantations is found in olm hoak (Quercus ilex) [2]. This relationship entails that free and fungus must be studied as a common entity, as the equilibrium in their association depends on their health status. Truffle cultivation requires a great extent of initial capital for its development (24.480€/ha) [3] and therefore plagues which may threaten the production might endanger the rentability of the whole operation [4].

There are previous instances of several different techniques being employed for the detection of black truffle and holm oak symbiotic system sickness. Methods ranging from simple visual inspection to complex molecular detection [5] or spectroscopy techniques [6]. However, current literature presents a technological gap regarding the detection of white rot such as *Armillaria mellea* and *Rosellinia necatrix* infections individually and their interactions with *Oomiceto Phytophthora Cinnamomi* pathogens on early stages [7], [8], which greatly affect this symbiotic systems.

Therefore, this project aims to develop a high precision, low-cost, real-time holm oak crop health status diagnosis system, which updates the analyzed trees status in terms of infection, stage and location of said infection via a digital platform.

The QUILEXDRON project (CCP2023-010793) is part of the 2023 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific,Technical and Innovation Research 2021-2023. The project consortium is formed by IDVIA 2020 Horizon 2020 S.L. and the Polytechnic University of Valencia as a public research organization.

* reynaldo.cabezas@idvia.es

2 Method

The QILEXDRON system is based on the use of thermal and hyperspectral cameras on board of unmanned aircraft (drone) which will be able to pinpoint and assess the stage of infection of the oak based on its reflectance of specific spectral indicators. To this end, the flight pattern will be optimized to improve the hyperspectral and thermal image quality, based on the flight parameters present in a predefined flight plan. Figure 1 schematically represents the various parameters that can affect image quality. To guarantee the correct performance of the system, several subsystems are developed.

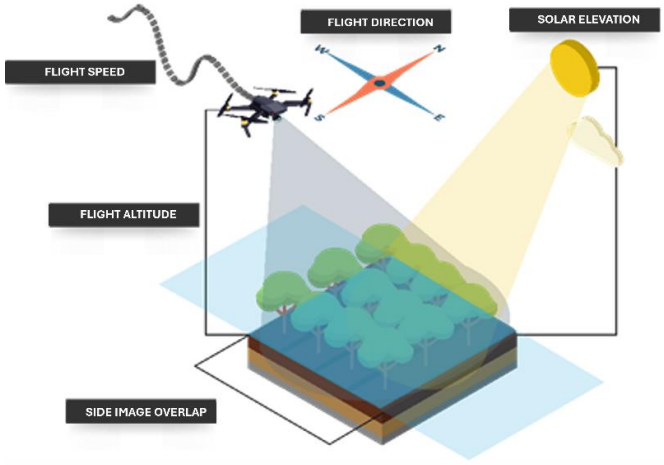


Figure 1. Schematic representation of the different parameters which may affect image quality.

2.1 Hardware and data acquisition subsystem:

The schematic of the data acquisition subsystem is shown in Figure 2. Consisting of the production and assembly of the drone as well as the inclusion of hyperspectral and thermal cameras and the communication subsystem, which will need to be able to filter the background spectral noise contamination in order clear the frequencies of interest and later upload the data base to a cloud from which the processing subsystem will download it.

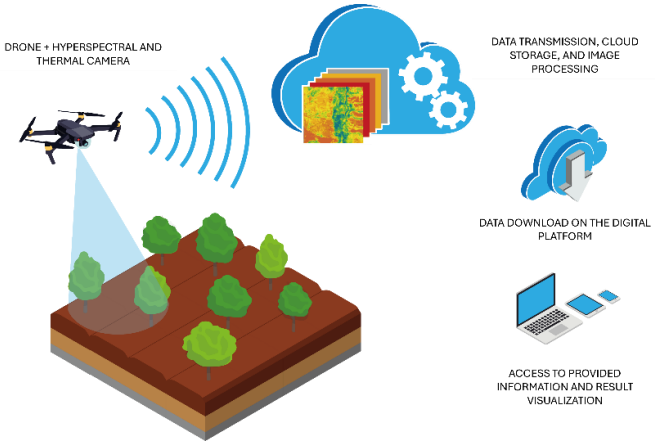


Figure 2. Schematic representation of the data acquisition subsystem.

2.2 Radiative Transfer Model (RTM):

To determine the representative variables to assess the health status of black truffle holm oak trees when infected with white rot infections it is first needed to perform in-situ measurements of currently infected specimens to calibrate the spectral database. Once this step is completed, a Radiative Transfer Model will be developed whose objective will be to evaluate several simulations with different biophysical parameters under different cultivation and climactic conditions. The simulated and measured reflectance parameters relation will be modeled via a Gaussian regression which provides a Bayesian probabilistic approach for the tunning of non-parametric and non-lineal regression nuclei between the simulated reflectance and the biophysical parameters cultivated.

2.3 RTM inversion

After the retrieve model is obtained, it can be inverted as to offer the biophysical parameters based on the reflectance received by the acquisition system in real time. The health status of the crop will be evaluated and assigned via Neural Networking algorithms based on LDA and SVM methods for early and developed stages of infection classification respectively. The workflow of this algorithm is shown in Figure 3.

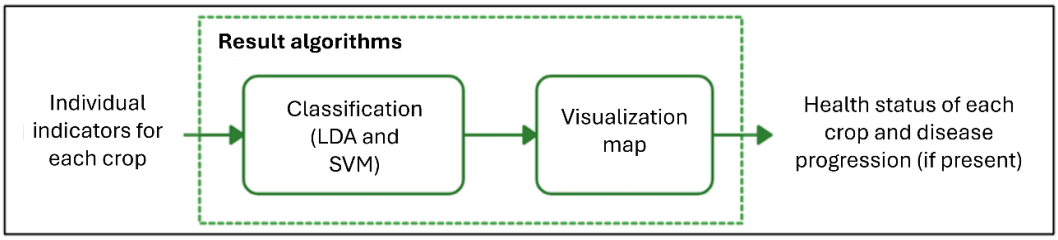


Figure 3. Neural Network workflow for infection diagnosis and location.

2.4 Digital Platform

Which will allow access to the information obtained by the system in real time and in a clear, precise, an easily understandable manner. The results can be monitored from the platform.

3 Results:

Applying the methodology presented in the different subsystems the health status and infection location of the crops can be assessed. In order to do this the QILEXDRON system is based primarily on two levels.

First the acquisition of the reflective spectrum is obtained by the hyperspectral and thermal cameras mounted in the drone, data that is then cleaned of background noise and uploaded to a digital storage cloud. This data is then downloaded and introduced in the inverse RTM model so that the biophysical parameters of the trees can be extracted and their health evaluated by the Neural Network algorithm, allowing it to identify both the

stage and position of the infection. The complete schematic of each of the processes that make up the system can be seen in Figure 4.

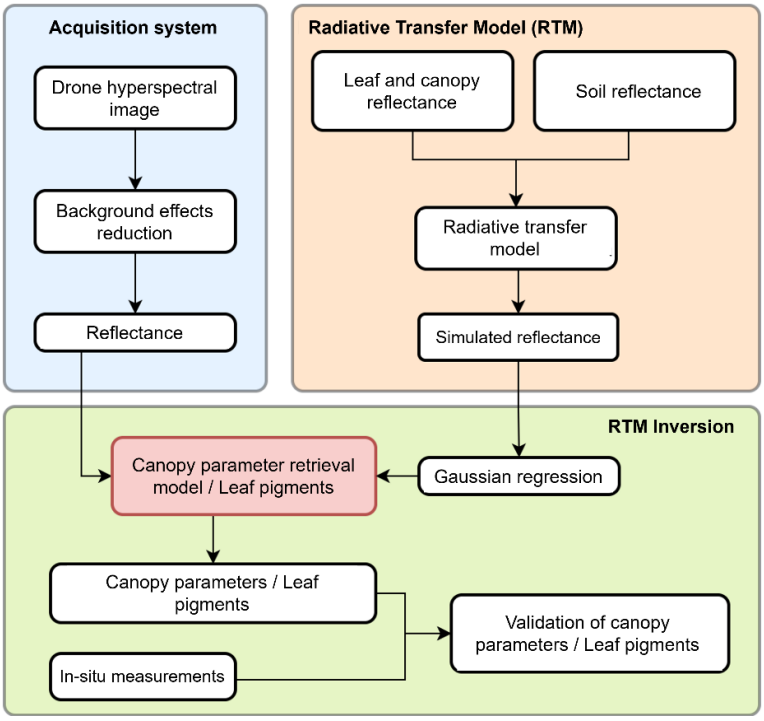


Figure 4. System process flow diagram.

The completed challenges will consist of:

1. Development of a data acquisition and filtering system which continuously uploads real time data on the crop spectral response.
2. Development of RTM model capable of identifying the biophysical parameters of the crop from its spectral reflectance response.
3. Development of Neural Network algorithms capable of assessing the health status of the crop from its biophysical parameters and allowing a real time follow up.

4 Conclusions:

Once this project is finished and some technological maturity has been achieved, some challenges must be overcome to launch a commercial version of the system, which can be enumerated as follows:

1. Guarantee market access through strategical associates for the future commercialization of the system, essentially both the service provider and the consumer.
2. Provide an economically robust version of the data acquisition system.
3. Scale the digital platform to cover the demand to its full expected extent.

4. Stablish new performance demonstrators so that the system can both exhibit its functioning and use to potential buyers and demonstrate its resiliency under different environments and conditions.

References

[1] Ricard, J. M., Bergougnoux, F., La truffe: guide technique de trufficulture. Paris, Centre technique interprofessionnel des fruits et légumes, 2003.

[2] Domenech, S. R., Truficultura. Fundamentos y Técnicas. Ediciones Mundi-Prensa, 2007.

[3] Minguet Pla, J. S., Masses Tarragó, J. M., El cultivo de la trufa negra. *Dossier Tècnic*, (26), 2008.

[4] Santafé, M. M., Plagas y enfermedades asociadas a plantaciones truferas. Centro de Investigación y Tecnología Agroalimentaria de Aragón (Spain), 2021.

[5] Fang, Y., Ramasamy, R. P., Current and prospective methods for plant disease detection. *Biosensors*, 5(3): 537-561, 2015.

[6] Ali, A., Imran, M., Remotely sensed real-time quantification of biophysical and biochemical traits of Citrus (Citrus sinensis L.) fruit orchards - A review. *Scientia horticulturae*, 282: 110024, 2021.

[7] Calamita, F., Imran, H. A., Vescovo, L., Mekhalfi, M. L., La Porta, N., Early identification of root rot disease by using hyperspectral reflectance: The case of pathosystem grapevine/armillaria. *Remote Sensing*, 13(13): 24-36, 2021.

[8] Pérez-Bueno, M. L., Pineda, M., Vida, C., Fernández-Ortuño, D., Torés, J. A., De Vicente, A., Cazorla, F. M., Barón, M., Detection of white root rot in avocado trees by remote sensing. *Plant disease*, 103(6): 1119-1125, 2019.

Technical-mathematical solution for dynamic railway track inspection

M. Isabel Espinós[1,*], Julia I.Real[1], Juan Ramón Torregrosa[1]and Polina Batalova[1]

(1) Instituto Universitario de Matemática Multidisciplinar
Universitat Politècnica de València, 46022, València (Spain)

1 Introduction

Railway auscultation is a key process for the inspection and monitoring of railway infrastructure that requires specialized equipment and skilled labor. It aims to detect potential failures or wear on the superstructure and substructure components, ensuring operational safety, performance, and efficiency. Currently, this process is mainly performed using diagnostic trains equipped with advanced instrumentation and software capable of assessing infrastructure quality and supporting predictive maintenance. However, these systems involve high acquisition and operational costs, significant logistical efforts, and traffic disruptions during inspection works. Additional costs also arise from on-site visual inspections and complementary manual measurements.

Within the framework of the work title “Contribution to the calculation of the five geometric track parameters through the mathematical processing of signals captured by onboard equipment”, whose objective is to contribute to railway engineering by exploring a cost-effective alternative for track quality assessment. Specifically, it investigates whether the five key geometric railway track parameters—track gauge, alignment, longitudinal level, cross level, and twist—can be derived from data acquired by an onboard accelerometer. These parameters are essential indicators of track condition, directly affecting safety, ride comfort, infrastructure reliability, and maintenance planning. The expected outcomes are intended to contribute to the improvement of track inspection systems through a non-invasive methodology that provides quasi-continuous and real-time data.

This paper summarized the first phase of the ongoing work, aimed at determining the most suitable hardware technology and its optimal placement on the locomotive to ensure that the collected data is relevant to the objectives of the study.

* miesgar@upv.es

2 Functional description of the solution

2.1 Hardware elements

First, an exhaustive study was carried out on the structure of railway vehicles according to the type of operation, taking into account the natural vibration frequency of each component, the orientation of surfaces, electromagnetic incompatibility, among other factors. On this basis, technical specifications required for the inertial systems to be installed on the vehicles were defined.

Subsequently, an analysis of the existing hardware technologies with direct applicability or with slight modifications required was conducted to conclude with a proposal of the hardware specifications, including installation conditions considered optimal for application across any type of railway operation: acceleration and frequency ranges, sensitivity, permissible operating temperatures, energy consumption, etc.

Based on all of the above, the system being developed for dynamic auscultation is composed of three main elements:

2.1.1 Accelerometer Sensor

The main characteristics of the selected sensor are:

- Triaxial accelerometer
- ±25g measurement range
- Bandwidth up to 10 kHz
- Galvanically isolated and IP67-ratedResults

2.1.2 Data acquisition module

In order to perform the measurements, an acquisition card is available, which includes a high-performance analog-to-digital converter for processing the accelerometer signals, along with a geolocation module.

The analogue signal from the accelerometer is converted to a digital signal that can be interpreted using a data acquisition module with he following main characteristics:

- Maximum resolution of 24 bits.
- 8 analog signal inputs.
- Sampling frequency up to 15 kHz.

The GPS module from allows reading the vehicle's geolocation data through a communication interface with the computer. It has the following main characteristics:

- Update frequency of 10 Hz.
- Accuracy of 1m.
- Low power consumption.

– Compatibility with GPS, GLONASS, and Galileo systems.

Along with the GPS module, a high-sensitivity, high-gain antenna is required to ensure the proper functioning of the GPS system in all types of situations and environments. This antenna is placed on the outside of the cabin or on one of its windows.

2.1.3 Control computer

This computer is equipped with a high-performance, low-power processor that runs a Linux-based operating system for managing measurement devices and storing data on a removable memory card.
The selected computer has the following main characteristics:

- 1 GHz AM3358 ARM Cortex-A8 processor.
- 512MB DDR3 RAM.
- 4GB 8-bit eMMC internal memory.
- Up to 256 GB of removable memory.
- I2C, SPI, UART, USB, and Ethernet communication.
- Linux operating system.
- Very compact size and low power consumption.

2.1.4 Simplified Diagrams of the Solution

Figure 1 shows a detail of where the components of the measurement system would be located.

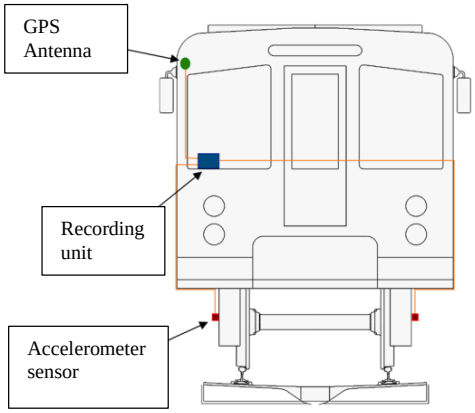


Figure 1. On boarding solution diagram

The information flow diagram of the system is shown below on Figure 2.

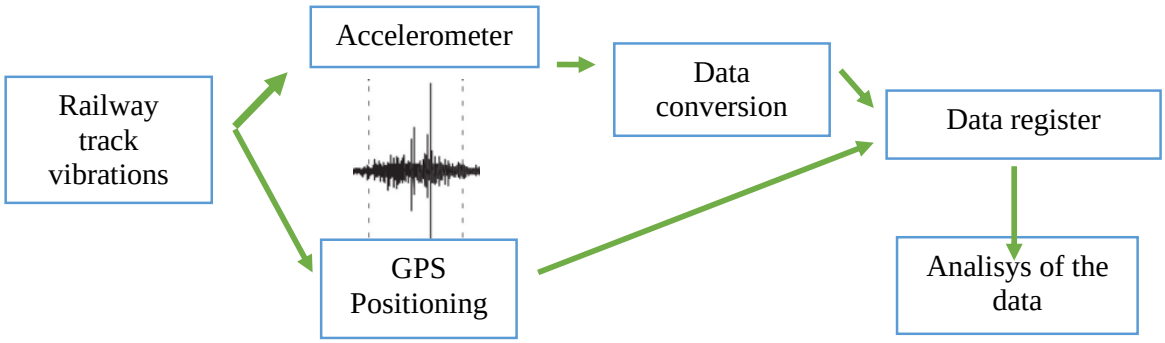


Figure 2. Data collecting diagram

2.2 Mathematical elements for data processing and analysis

Once the data has been collected in the form of files through the described hardware system, its analysis is performed using the signal processing algorithm. The objective of this processing is to characterize the track parameters and its components based on the vibrational signal recorded by the auscultation system installed on the vehicle, and to relate this information by associating the records with the GPS coordinates.
The algorithm will process the acceleration (vibration) data obtained in the time domain. Using acceleration, the displacements that caused it can be derived through integration. By filtering and transforming the data into the frequency domain, the signal can be correlated with the cause that provoked it using an analysis based on a mathematical model of the interaction between the vehicle axle and the track.

When process is completed, it is possible to return to the time domain and, ultimately, obtain in the spatial domain the correlation of the track defects and characteristics with the coordinates of the points where the data was recorded.

The stages or phases involved in this process are as follows:

1. Acceleration integration
2. Discrete Fourier Transform
3. Filtering of recorded data
4. Solution of the model equations
5. Inverse Discrete Fourier Transform
6. Transition to the spatial domain

The data processing will focus on the bogie quarter model as represented in the Figure 3.

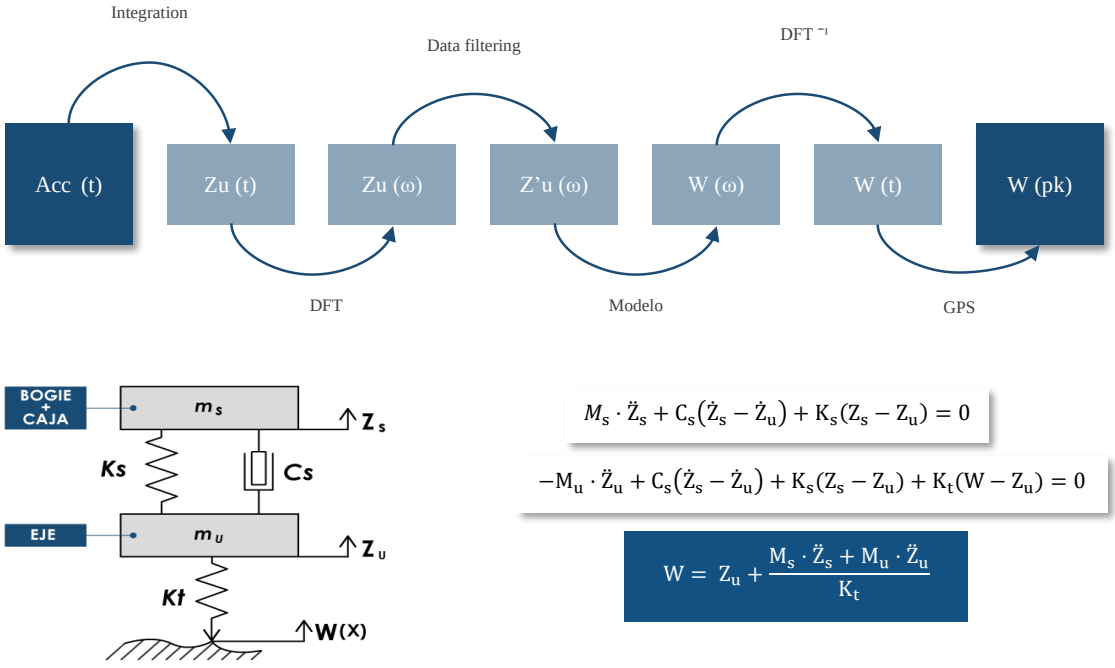


Figure 3. Methodological framework for data processing

Currently, work is being carried out on the design and development of the algorithm that will handle the processing of the obtained inertial signals, so that, from the recorded accelerations, the track geometric parameters can be obtained.

3 Conclusions

A device has been designed that meets the specifications established in relation to the mathematical models that are planned to be used. Based on the established mathematical hypotheses, it will be possible to perform and optimize the calculations of the track parameters.

References

[1] Elkhoury, N., Hitihamillage, L., Moridpour, S., & Robert, D., Degradation prediction of rail tracks: A review of the existing literature. *The Open Transportation Journal,* 12(1), 2018.

[2] Pehlivan, F., Mizrak, C., & Esen, I., Modelling and validation of 2-DOF rail vehicle model based on electro–mechanical analogy theory using theoretical and experimental methods. *Engineering, Technology & Applied Science Research*, 8(6): 3603-3608, 2018.

Multi-Depth Detection System and Maturity State Identification of Black Truffles in Truffle Cultivation Fields

Salvador Mateo[2,*], Julia I. Real[2], Laia Rubio[3] and Ferney Gómez[1]

(1) IDVIA 2020 HORIZONTE 2020 S.L.,
Plaza de Bandas de Musica 5, Local 11, 46013, Valencia (Spain)
(2) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, 46022, València (Spain)
(3) Innovating 4m S.L Calle Cuenca, 64 - p. bj, 08032, Barcelona (Spain)

1 Introduction

The black truffle (*Tuber melanosporum*), considered one of the most prized fungi worldwide, represents a key economic pillar in many rural regions of southern Europe, especially in Spain, which leads international production both in quantity and cultivated area, with an average annual production of 47 tons and over 20,000 hectares of cultivation. Spain dominates global production, reaching up to 104 tons in the last harvest seasons. Despite the sustained growth of this industry and the high market value of black truffles, their harvesting still relies on a traditional method: the use of trained dogs to detect their characteristic underground aroma. Although effective, this system has multiple limitations, such as the high cost of acquiring and maintaining the animals, training practices that may contravene existing animal welfare laws, and variable effectiveness influenced by factors like climate, the dog's age, and physical endurance.

Furthermore, the reliance on truffle dogs not only compromises the efficiency of harvesting but also the sustainability of the crop, due to the risk of damaging the mycelium during excavation. More concerning situations, such as documented cases of dogs being poisoned by competing truffle farmers, further highlight the issues. In this context, there is an urgent need to technify truffle farming with innovative solutions that ensure the long-term economic and environmental viability of this crop.

This project proposes the development of TRUFIND, a pioneering system based on the use of an electronic nose equipped with electrochemical sensors and a biomimetic intake mechanism that simulates canine sniffing. This system will not only allow for the precise detection of volatile compounds emitted by black truffles but will also integrate geolocation capabilities and real-time digital visualization through heat maps. This technology will not only reduce operational costs and improve crop performance but will also provide an essential tool for novice truffle farmers, facilitating intelligent management of their plantations and reinforcing the sector's sustainability.

The TRUFIND project (CPP2023-010735) is part of the 2023 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-

* salmavil@upv.es

2023. The project consortium is formed by INNOVATING 4M S.L, IDVIA HORIZONTE 2020 S.L and the Polytechnic University of Valencia as a public research organisation.

2 Methods

The method of this system is based on the development of an electronic nose, composed of electrochemical sensors and a biomimetic device that simulates canine sniffing, enabling the detection of volatile organic compounds (VOCs) characteristic of black truffles. Additionally, the system will be capable of recording the location of black truffles within the plantation plot, with the aim of optimizing harvesting in future seasons or identifying potential anomalies. This functionality will be continuously visualized through a digital platform using a heat map representation. The schematic of the proposed solution is illustrated in Figure 1.

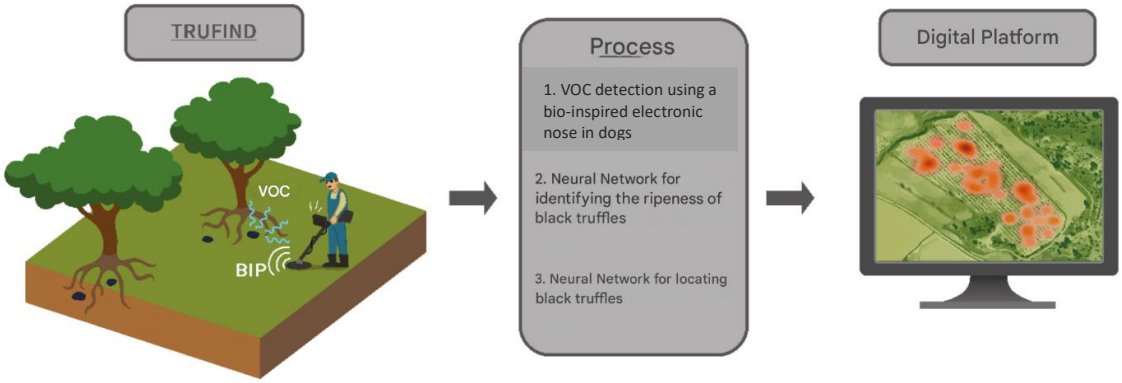


Figure 1. Diagram of the TRUFIND solution.

The bioinspired nose will be equipped with two electrochemical sensors to detect the two main volatile organic compounds (VOCs) perceived by dogs: dimethyl sulfide (DMS) and dimethyl disulfide (DMDS) [1]. The response from these sensors will feed two Neural Networks: the first one will be responsible for determining the depth and maturity stage of the black truffle, and the second will indicate its exact location.

Thus, the TRUFIND system will consist of a hardware subsystem, software, and a digital platform:

HARDWARE AND COMMUNICATION SUBSYSTEM: This unit is responsible for emulating the dog's sniffing cycle to enhance the precision of odor particle detection and identify the main VOCs of black truffles. The core element is an electronic nose composed of specific electrochemical sensors, coupled with a biomimetic sniffing system. The system will include an audible alert to notify the truffle hunter of the presence and maturity level of the truffle. Additionally, it will incorporate a long-range communication device to transmit the recorded data to the digital platform

SOFTWARE SUBSYSTEM: This component will process all the information gathered by the hardware subsystem. It will consist of algorithms based on image processing and time series analysis. Its function will be to detect and identify the maturity stage of buried black truffles at various depths, as well as their direction, following the diagram shown in Figure 2.

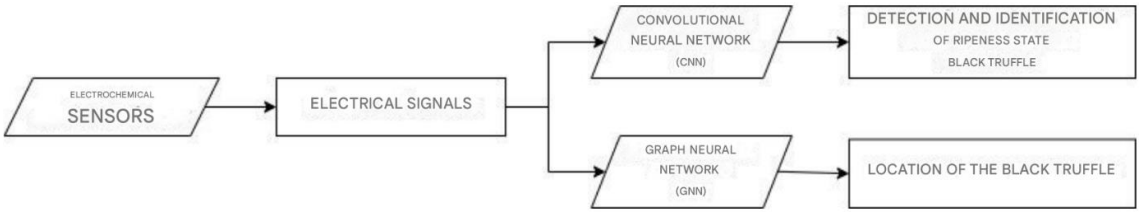


Figure 2. Schematic diagram of the black truffle detection software.

DIGITAL PLATFORM: This platform will display real-time results regarding the truffle's location and maturity level in a clear and user-friendly format. Truffle growers will also be able to access historical records, data analysis, and detect possible anomalies, among other features. Furthermore, it will be implemented to be compatible with various devices, enabling access to results anytime and from any internet-connected device.

The project presents two key challenges for its implementation:

2.1 Identification of the aromatic intensity of the predominant VOCs in black truffles at different depths and maturity stages.

In practice, black truffles are typically found buried at varying depths, ranging from 10 to 40 cm [2]. To ensure the device performs effectively under real operational conditions, tests will be conducted using buried truffles, replicating their natural positioning in the soil.

A classification Neural Network will be developed to help truffle growers interpret and analyze the data generated by the system. This network will be designed to classify buried black truffles based on both their depth and maturity level. The input variables for the Neural Network will consist of a series of images representing each truffle's "digital fingerprint," generated from signals captured by the electrochemical sensors. These signals will be transformed using advanced techniques such as Gramian Angular Field (GAF) and Markov Transition Field (MTF) [3], converting time-series data into visual representations for more accurate classification.

2.2 Development of a Biomimetic Dog Nose with 3D Printing Technology

Dogs detect truffles with great efficiency due to their active sniffing, characterized by a respiratory cycle of 5 Hz and airflow of up to 60 l/min [4]. During exhalation, they direct the air downward and to the sides, preventing the scent from dispersing, while during inhalation, they generate a hemispherical flow that draws nearby air into their nostrils (Figure 3). This mechanism not only preserves the olfactory trail but also creates a vacuum that attracts odor particles toward the nose.

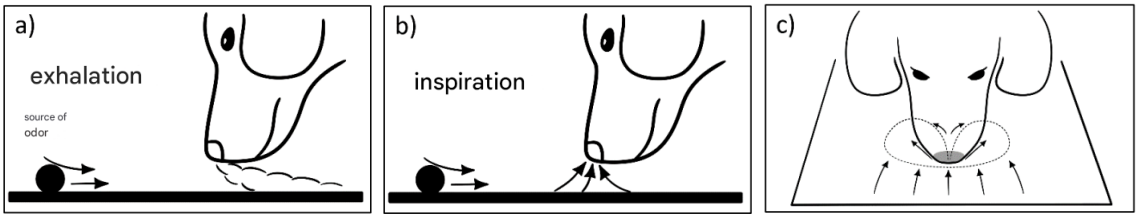


Figure 3. Operation of the exhalation and inhalation cycle. Source: [5].

In this context, the present innovation aims to optimize the capture and localization of the odor sources of the black truffle, improving existing alternatives. This challenge is divided into two parts: the first seeks to replicate the dynamic movement of the dog's nostrils using a flexible material that deforms with electrical impulses, allowing for better direction of airflow during both inhalation and exhalation. The second part involves developing a system that mimics the canine sniffing frequency (5 Hz), using a solenoid valve to control the intake of air into the sensors and regulate its expulsion from a compressed air source [6].

The second challenge focuses on reproducing the omnidirectionality of the sense of smell to improve black truffle detection. To achieve this, the TRUFIND system will feature five bioinspired noses oriented in different directions, each with its own inhalation system and electrochemical sensors, while exhalation will be common. A 3D model of the nasal system will be designed, and computational fluid dynamics (CFD) simulations will be carried out to optimize airflow, adjusting parameters such as nose size, duct length and diameter, and the distance to the ground. The process to be followed for optimizing the design of the bioinspired nose is shown in Figure 4.

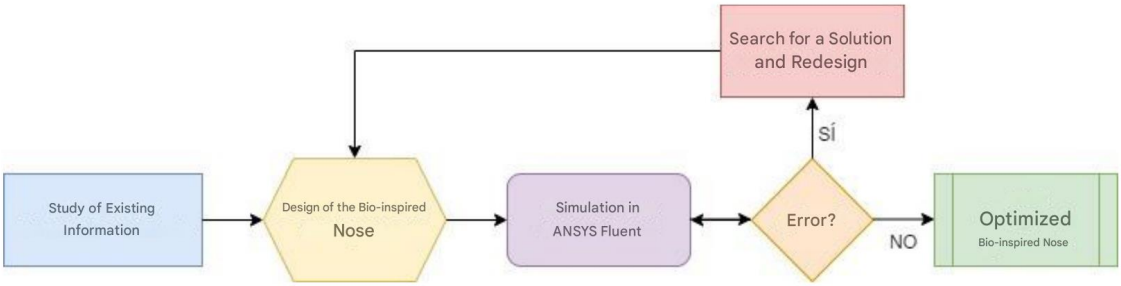


Figure 4. Diagram of the CFD simulation operation for the optimization of the bioinspired nose design.

The third challenge focuses on the design of a Graph Neural Network (GNN) for black truffle localization, inspired by the stereolfaction of dogs. The GNN will classify the signals from the electrochemical sensors, associating them with nodes in a graph that describes their interactions [7]. The system will process this graph to determine which nodes receive the highest aromatic intensity, identifying the direction of the truffle. Training the network will adjust the weights to correlate the sensors with the correct directions, evaluating its effectiveness with test data to achieve accurate localization.

3 Results

The following outcomes are expected from the implementation of the TRUFIND system, which combines biomimetic design, advanced sensor technologies, and intelligent data analysis to modernize black truffle detection:

I. Reliable identification of key VOCs (DMS and DMDS) in diverse soil and maturity conditions using electrochemical sensors.
II. Automated classification of truffles by depth and maturity through neural networks
III. 3D localization of truffles using a Graph Neural Network that mimics canine stereo olfaction to determine odor direction.
IV. Enhanced airflow and odor capture through CFD-optimized biomimetic nose design.
V. Real-time display of truffle location and maturity on a digital platform with alerts and data history for informed decision-making.

Moreover, the advancements achieved will contribute to strengthening knowledge in key areas such as data science, predictive analytics, automation, and precision agriculture, positioning the TRUFIND system as a cutting-edge tool for sustainable agricultural practices.

4 Conclusions

The TRUFIND system introduces an innovative approach to truffle detection by replicating the canine sniffing ability through an electronic nose and advanced sensor systems. The ability to detect and classify black truffles based on VOCs and their maturity is a significant step towards automating the truffle harvesting process. This technology is expected to reduce reliance on trained dogs, which are costly and limited by various factors, providing a more sustainable and efficient alternative.

The outcomes of the project demonstrate the potential of biomimetic systems and machine learning, particularly the use of Graph Neural Networks (GNNs) to improve the precision of truffle localization. By simulating the olfactory capabilities of dogs and employing sophisticated computational models, TRUFIND is expected to enable more accurate and faster detection of truffles. Furthermore, the integration of real-time data visualization is expected to allow truffle farmers to optimize their harvesting techniques and enhance crop management.

References

[1] Mustafa, A. M., Angeloni, S., Nzekoue, F. K., Abouelenein, D., Sagratini, G., Caprioli, G., Torregiani, E., An overview on truffle aroma and main volatile compounds. *Molecules*, 25(24): 5948, 2020.

[2] Sourzat, P., Guide pratique de trufficulture. Station d'expérimentations sur la truffe, 1997.

[3] Wang, Z., Oates, T., Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, 1: 1-7, 2015.

[4] Craven, B. A., Paterson, E. G., Settles, G. S., Lawson, M. J, Development and verification of a high-fidelity computational fluid dynamics model of canine nasal airflow., 2009.

[5] Staymates, M. E., MacCrehan, W. A., Staymates, J. L., Kunz, R. R., Mendum, T., Ong, T., Geurtsen, G., Gillen, G. J., Craven, B. A., Biomimetic sniffing improves the detection performance of a 3D printed nose of a dog and a commercial trace vapor detector. *Scientific reports*, 6(1): 36876, 2016.

[6] Craven, B. A., Paterson, E. G., & Settles, G. S., The fluid dynamics of canine olfaction: unique nasal airflow patterns as an explanation of macrosmia. *Journal of the Royal Society Interface*, 7(47): 933-943, 2010.

[7] Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G. I., King, I., Pan, S., A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Automatic pavement condition assessment system for airports

Julia I. Real[1,*], Adrián Miralles[1], Rodrigo Solano[2] and Alejandro Grandado[3]

(1) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, 46022, València (Spain)
(2) Aena, S.M.E, S.A. C/ Peonías, 12. 28042, Madrid. (Spain)
(3) GRUPO VIARIUM,
C/ Jacinto Benavente, 2B, 2º. Edificio TRIPARK, 28232 Las Rozas, Madrid (Spain)

1 Introduction

Airport pavement condition is a critical factor in ensuring operational safety. Regular and accurate assessments are essential for planning timely maintenance and sustaining optimal performance. However, current inspection methods are predominantly manual and based on visual surveys, which are time-consuming and inefficient. Inspectors must physically traverse the entire area to record surface damage, leading to significant delays and elevated costs—up to €600,000 annually for the 48 airports and heliports managed by AENA (Public Entity Spanish Airports and Air Navigation).

In response to this need, this work was developed with the main objective of implementing a system that streamlines inspection processes and optimizes resource allocation without increasing operational costs. This involves replacing traditional visual inspections with an automated approach leveraging technologies such as computer vision.

The project's objective is to design a low-cost, fully automated evaluation system that can be installed on any vehicle already operated by airport management. This system will enable the automatic detection, classification, and quantification of pavement defects by integrating it into a conventional airport vehicle. It will be capable of assessing both rigid and flexible pavement types using intelligent video analysis techniques supported by Neural Networks. Furthermore, the system will compute the Pavement Condition Index (PCI) automatically through the development and implementation of a fuzzy inference model.

The AEROPAV project (CPP2023-010585) is part of the 2023 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by VIARIUM INGENIERÍA S.L., AENA SME and the Polytechnic University of Valencia as a public research organisation

2 Methods

This new system will enable the automatic detection, categorization, classification, and quantification of pavement distresses in airport runways by being installed on a conventional vehicle. It will combine data collected during the vehicle's routine operations—captured through a set of cameras [1] and a GPS device—with various Neural Networks (NNs) and

* jureaher@upv.es

algorithms responsible for identifying and quantifying the observed distresses. This will allow for the automatic calculation of the Pavement Condition Index (PCI), a numerical rating of pavement condition based on the type and severity of present distresses, following the diagram shown in Figure 1. The PCI value, which is assigned to a specific pavement section, is used to design and update maintenance and rehabilitation programs.

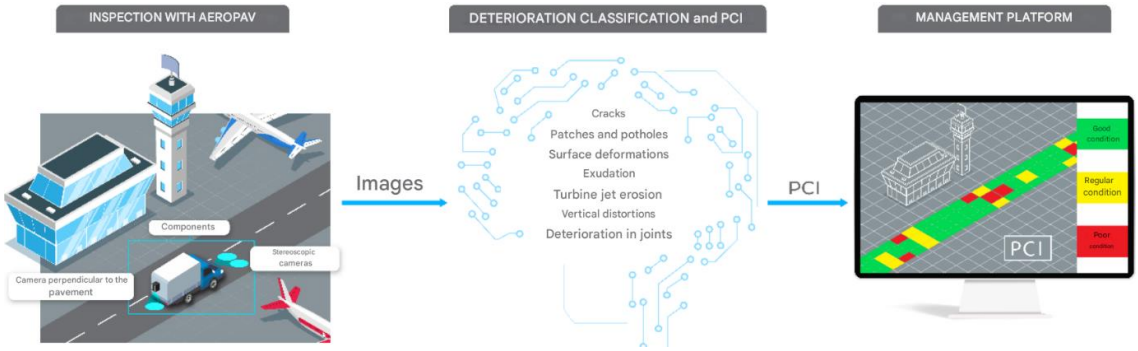


Figure 1. Diagram of the airport pavement condition diagnosis process.

The process will be carried out in four stages:

1. Identification of distressed pavement through entropy analysis
2. Initial categorization of the distress type
3. Specific classification of the distress
4. Evaluation of the severity to determine the pavement condition

Horizontal distortions in pavement represent a discontinuity in its surface texture; therefore, assessing texture homogeneity helps determine whether the pavement is in good condition or if it is likely to exhibit signs of deterioration. To achieve this, the first step is to identify the presence of horizontal distortions using an entropy-based algorithm, aimed at reducing computational load. Entropy can be understood as a measure of uncertainty, that is, the degree of disorder in a system [2]. Shannon (1948) first introduced this concept in the context of image processing, quantifying the amount of uncertainty in a probability distribution where p is the probability of occurrence and N is the number of classes in the analysis:

$$Entropy = E = -\sum_{1}^{N} p \cdot \log_2(p)$$

Once the entropy algorithm determines whether the pavement is in poor condition, the type of damage is categorized using a Neural Network [3]. For flexible pavement, the types include cracks, patches and potholes, surface deformations, vertical distortions, and other forms of damage. In contrast, for rigid pavement, the system will also detect joint deterioration.

In the third stage, each type of damage is further classified using either dedicated neural networks or a hybrid approach that combines the Canny edge detection algorithm, developed by John F. Canny, which refines the final contours of the regions of interest [4].

The final step involves assessing the severity of the deterioration through geometric corrections, color intensity analysis, and shape parameters specific to each damage type. Most forms of deterioration are identified using images from a camera positioned perpendicularly to the pavement surface. However, vertical distortions require depth information, which can only be accurately captured through stereoscopic imaging, making stereoscopic cameras essential for their detection. All stages of the proposed algorithm are outlined in Figure 2.

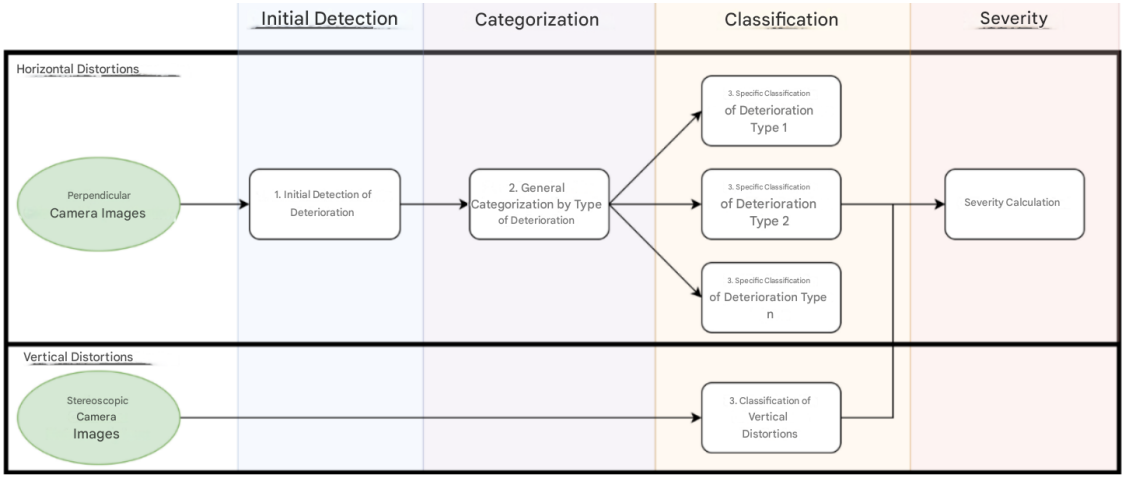


Figure 2. General scheme of the proposed algorithm.

The system will be composed of the following subsystems for data processing:

Hardware Subsystem: Responsible for data acquisition, this subsystem will include a camera mounted perpendicularly to the pavement at the rear of the vehicle, two stereoscopic cameras installed inside the vehicle and oriented in the direction of travel, and a high-precision GPS. It will also feature a storage unit responsible for collecting the various data obtained from both hardware and software sources. A high-capacity hard disk drive (HDD) will be used to store all information, while a solid-state drive (SSD) will serve as a buffer, thus ensuring information traceability through a full encryption system.

Software Subsystem: This component will handle the processing and analysis of all the information collected by the cameras. Its overall operation will follow a decision-tree structure, where an algorithm selects a processing branch based on calculated probabilities, triggering the execution of corresponding algorithms and neural networks until the result is obtained. Specifically, this subsystem is expected to include four advanced image processing algorithms and seven neural networks dedicated to detecting, categorizing, classifying, and quantifying pavement damage, as well as an additional neural network for the automatic calculation of the Pavement Condition Index (PCI) [5].

Digital Platform: This platform will display the georeferenced results in real time, including the current PCI value for each pavement section and its trend, in a format that is simple and easy to interpret by users. Additionally, it will provide access to historical records, data analytics, and other related insights.

3 Results

The expected outcomes of this project include the development of an intelligent system capable of automatically detecting damage in airport pavements in real time, through the combined use of advanced image analysis algorithms and neural networks trained with a broad and representative dataset. These networks are expected to identify, classify, and quantify various types of deterioration in both rigid and flexible pavements, and will also allow for the automatic calculation of the Pavement Condition Index (PCI).

Moreover, the advancements achieved will contribute to strengthening knowledge in key areas such as data science, predictive maintenance, automation, and smart infrastructure.

4 Conclusions

This work presents a solution to the current challenges faced in the maintenance of airport pavements, by proposing the automation of diagnostics through advanced vision technologies and artificial intelligence. This approach represents a significant shift from traditional methods, enabling more frequent, systematic evaluations with reduced reliance on direct human intervention.

The combination of perpendicular and stereoscopic imaging significantly enhances the system's capability to capture both surface texture and depth—key elements for accurate damage detection. Embedding the system into conventional airport operation vehicles enables integration into routine activities without interfering with regular operations. Moreover, automating the Pavement Condition Index (PCI) calculation through entropy-based models and neural networks improves the precision of pavement condition assessments and delivers real-time results.

Moreover, this work highlights the value of interdisciplinary innovation by combining civil engineering with data science and computer science to deliver a practical tool for airport infrastructure managers. The knowledge generated through this experimental development lays the foundation for future improvements in predictive maintenance strategies and smart infrastructure management, aligning with the trend toward more automated, intelligent, and data-driven transportation systems.

References

[1] Hoang, N.D., Image processing based automatic recognition of asphalt pavement patch using a metaheuristic optimized machine learning approach. *Advanced engineering informatics*, 40: 110--120, 2019.

[2] Shannon, C. E., A mathematical theory of communication. *The Bell system technical journal*, 27(3): 379--423, 1948.

[3] Yusof, N. A. M., Ibrahim, A., Noor, M. H. M., Tahir, N. M., Yusof, N. M., Abidin, N. Z., Osman, M. K., Deep convolution neural network for crack detection on asphalt pavement. *Journal of Physics: Conference Series*, 1349(1): 12--20, 2019.

[4] Bruno, L., Parla, G., Celauro, C., Image analysis for detecting aggregate gradation in asphalt mixture from planar images. *Construction and Building Materials*, 28(1): 21--30, 2012.

[5] Issa, A., Samaneh, H., Ghanim, M., Predicting pavement condition index using artificial neural networks approach. *Ain Shams Engineering Journal*, 13(1): 101490, 2022.

Structural health monitoring of bridge infrastructures using single-pass SAR data

Guillermo Alandí[1,*], Julia I. Real[1], Ernesto Serrano[2] and Alejandro Rodrígez[3]

(1) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022,València (Spain)
(2) IMESAPI, Avenida de Ramón y Cajal, 107, 28043 Madrid (Spain)
(3) INECO, Avda. Partenón, 4-6 Planta Baja Puerta 2, 28042, Madrid (Spain)

1 Introduction

Structural Health Monitoring (SHM) stands as an essential discipline to ensure the safety and functionality of critical civil infrastructures, especially bridges, many of which face aging, increased traffic, and adverse environmental conditions. Accurate diagnosis of their structural condition is vital for optimizing maintenance strategies and preventing catastrophic failures. Traditionally, SHM has relied on the installation of sensor networks, such as accelerometers or strain gauges, to measure deformations and dynamic properties [1].

However, this approach has significant limitations: the high costs associated with sensor installation and maintenance, their vulnerability to vandalism or theft, and the inherently limited spatial coverage, which hinders a global and continuous assessment of the structure. In fact, maintenance costs represent a substantial portion of the total infrastructure investment in the EU [2].

In this context, remote sensing using Synthetic Aperture Radar (SAR) has emerged as a powerful and non-invasive technological alternative. SAR systems onboard satellites allow for the quantification of surface deformations with high precision over large areas, regardless of weather or lighting conditions. Nevertheless, conventional SAR techniques, such as Differential Interferometry (DInSAR), often require many historical images with high coherence between them and face challenges in distinguishing structural movements from other factors, like thermal expansion [3].

To overcome these limitations, the SHM-BRISAR project proposed an innovative approach based on the analysis of single-pass SAR data. This methodology allows for the extraction of the complete vibration profile of the infrastructure from a single SAR observation, eliminating the dependence on long time series and significantly reducing the costs and complexity of the analysis.

The present work details the key advances towards the implementation of an operational SHM system based on this technique. It describes the systematic strategy developed for the automated acquisition and processing of SAR data from the Sentinel-1 mission [4], selected for its adequate resolution, high revisit frequency, and open data policy within the European Union's

* guialma3@upv.es

Copernicus program [5]. This automation is crucial for enabling continuous and scalable monitoring of multiple bridges, integrating the results with calibrated numerical models and Digital Twins for a detailed and predictive structural assessment, validated through deformation monitoring techniques [6].

The SHM-BRISAR project (CPP2022-009783) is part of the 2022 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by IDVIA 2020 Horizonte 2020 S.L., IMESAPI, INECO and the Polytechnic University of Valencia as a public research body.

2 Methodology and data source selection

2.1 Single-Pass SAR

The core methodology of this work lies in the application and operationalization of the single-pass SAR interferometry technique. Unlike traditional DInSAR approaches that require multiple coherent acquisitions and have been applied to structures like dams [7, 8] and bridges [9, 10], this technique exploits the information contained within a single SAR image to infer the dynamic behaviour of the structure.

Specifically, the vibrations of the infrastructure during SAR acquisition induce a relative displacement between the sensor and the structure, which manifests as specific artifacts in the resulting SAR image, primarily defocusing and apparent shifts in azimuth and range-azimuth. The magnitude and characteristics of this defocusing are directly related to the velocity and frequency of the vibration of the observed point on the structure.

Through appropriate signal processing of the complex SAR image (SLC), which may include 2D Fourier transforms (2D-DFT), band-pass filtering, and inverse transforms (2D-IDFT), it is possible to estimate this defocusing component and, from it, reconstruct the two-dimensional (2D) vibration profile of the infrastructure, overcoming challenges like separating structural movements from thermal expansion effects [3, 11].

This profile contains valuable modal information, including the natural vibration frequencies of the structure. This modal information extracted from the single-pass SAR analysis becomes the fundamental input for the calibration and validation of a detailed finite element model (FEM). This FEM model is built to faithfully represent the geometry, materials, and boundary conditions of the real structure, constituting its Digital Twin.

The calibration process is iterative: key model parameters, such as material properties (Young's modulus, density) or the stiffness of supports and soil, are adjusted until the natural frequencies and vibration modes calculated by the FEM model match those obtained experimentally through SAR analysis. Once calibrated, the Digital Twin becomes a powerful tool for simulating the bridge's behaviour under various loads, assessing its current structural state, detecting potential damage (which would manifest as changes in dynamic properties), and predicting its future evolution, as shown in Figure 1.

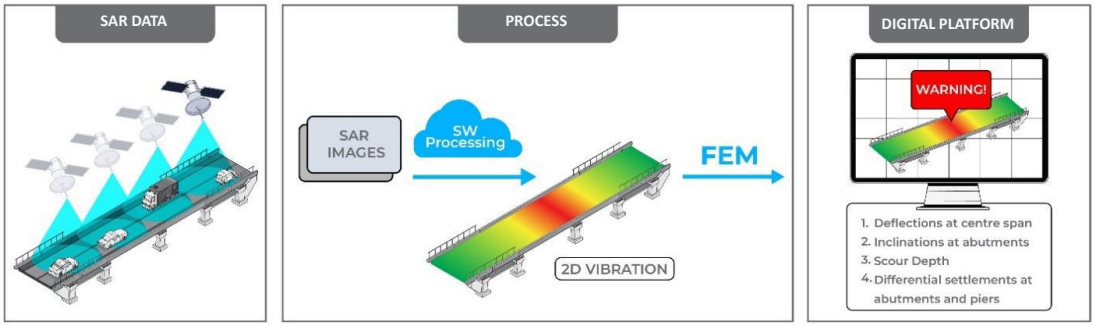


Figure 1. SHM using single-pass SAR data solution schematic. Source: own elaboration.

2.2 Satellite mission selection

A cornerstone for the operational and continuous implementation of this SHM system is the development of a robust and automated data acquisition strategy, specifically tailored to leverage the capabilities of the Copernicus Sentinel-1 mission.

The selection of Sentinel-1 [4] over other high-resolution SAR missions (such as the X-band COSMO-SkyMed [12]) was driven by a pragmatic assessment of operational requirements. Key advantages include Sentinel-1's open and free data access policy, which eliminates significant economic barriers for long-term monitoring projects. Its high revisit frequency, potentially reaching 6 days over Europe with the two-satellite constellation, enables near-continuous observation crucial for tracking dynamic structural behaviour.

The C-band frequency (~5.4 GHz) provides a well-established balance between adequate spatial resolution for infrastructure features and sufficient penetration capabilities through moderate vegetation or atmospheric conditions.

The primary acquisition mode employed, Interferometric Wide Swath (IW), utilizes the advanced TOPSAR (Terrain Observation by Progressive Scans) technique [13]. TOPSAR electronically steers the radar beam from backward to forward in the azimuth direction for bursts of pulses, achieving wide swath coverage (250 km) while maintaining a good geometric resolution (5x20 m) and minimizing scalloping effects, making it highly suitable for systematic infrastructure monitoring over extended areas.

Furthermore, the mission benefits from a mature ecosystem of processing tools and extensive validation, notably through the ESA-provided SNAP (Sentinel Application Platform) toolboxes [5], which facilitates subsequent analysis steps.

3 Results: SAR data acquisition algorithm

To exploit these advantages in an operational context, it is necessary an automated acquisition algorithm, which was scripted in Python. This algorithm automates the entire process of identifying, filtering, and downloading the required Sentinel-1 data by interacting programmatically with the Copernicus Data Space Ecosystem (CDSE) [14]. The core steps of the algorithm (see Figure 2) are as follows:

- Authentication: Obtain a keycloak access token using user credentials to authorize interaction with the CDSE API.
- Query construction: Build a structured query using the OData (Open Data Protocol) API [15]. This query incorporates multiple filtering criteria:
    - Temporal range (start and end dates).
    - Platform (Sentinel-1A/B).
    - Product type (Single Look Complex - SLC).
    - Acquisition mode (IW).
    - Geographic area of interest: Defined by parsing a standard GeoJSON file [16] containing the polygon covering the target bridge. The geometry is included in the OData query using spatial predicates.
- Query execution and product identification: Send the constructed query to the CDSE API endpoint. The API returns a list (JSON file) containing the unique identifiers of all Sentinel-1 products that satisfy the specified filters.
- Automated download: Iterate through the list of retrieved product identifiers. For each identifier, use the authenticated session to request and manage the download of the corresponding SLC data product file to a designated local storage location.

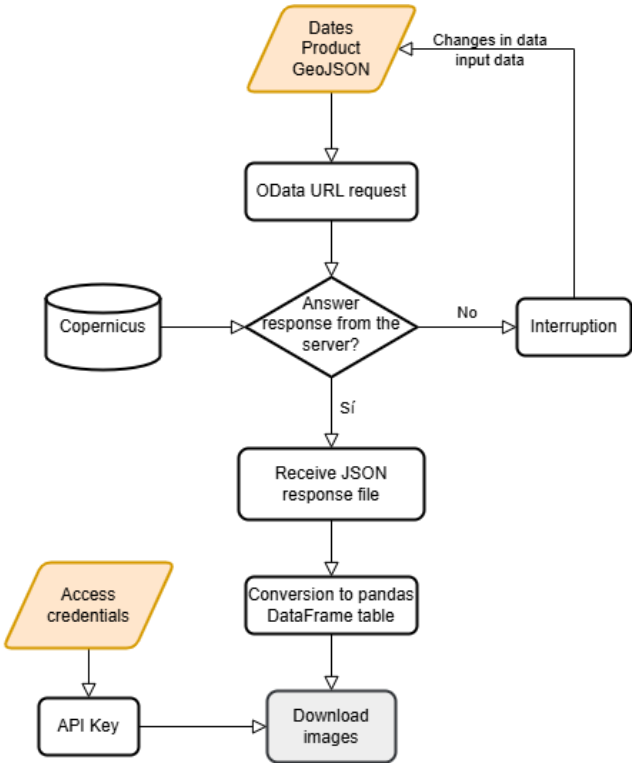


Figure 2. Flowchart of the SAR data acquisition algorithm. Source: own elaboration.

This fully automated pipeline ensures the timely and efficient retrieval of necessary input data, ready for ingestion into the subsequent single-pass SAR processing chain for vibration profile extraction.

The algorithm, developed in Python, has produced promising first results. It can obtain the requested satellite images by filtering by date, platform, product type, acquisition mode and area of interest. Figure 3 shows a satellite image obtained intersected with the region delimited by the El Barquero bridge (A Coruña, Spain).

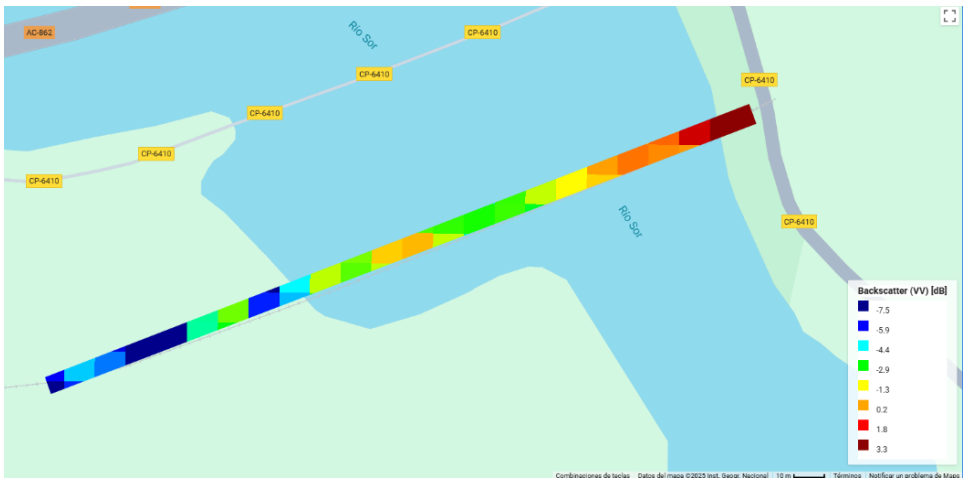


Figure 3. Satellite image obtained by the SAR data acquisition algorithm and intersected to the region of the El Barquero bridge. Source: own elaboration.

4 Conclusions and next steps

This paper describes the progress in implementing a system for the structural health monitoring (SHM) of bridges. This system is based on the single-pass SAR interferometry technique and utilizes data from the Sentinel-1 mission. Preliminary results indicate the feasibility of extracting vibration profiles from individual SAR images and their utility for calibrating numerical models (FEM), facilitating the generation of Digital Twins.

At this stage, the work has focused on the development of an automated algorithm for data acquisition. This algorithm, implemented in Python, interacts with the OData API of the Copernicus Data Space Ecosystem (CDSE) and allows for the systematic retrieval of Sentinel-1 SLC images filtered according to temporal, product, and spatial criteria (defined by a GeoJSON).

The combination of the single-pass SAR method with automated Sentinel-1 data acquisition represents a cost-effective and scalable approach for infrastructure management. Successful development of the system will require overcoming challenges. Below are listed some of the challenges overcome to develop the new tool:

- Extensive validation of the vibration profiles and structural states derived from the system through comparison with measurements from in-situ sensors on instrumented bridges.
- Refinement of the SAR processing algorithms and the criteria for damage detection based on the calibrated Digital Twins.
- Expansion of the system's application to a wider network of infrastructures to evaluate its scalability and operational robustness.

References

[1] Gkoumas, K. et al., Research and innovation in bridge maintenance, inspection and monitoring. Publications Office of the European Union, 2019.

[2] European Union Road Federation, Road Statistics Yearbook. 2021.

[3] Fornaro, G., Reale, D., Verde, S., Bridge thermal dilation monitoring with millimeter sensitivity via multidimensional SAR imaging. I*EEE Geoscience and Remote Sensing Letters*, 10(4): 677--681, 2012.

[4] «Sentinel-1 User Guide». [Online]. Available: https://sentiwiki.copernicus.eu/web/sentiwiki

[5] Jutz, S. & Milagro-Pérez, M. P., Copernicus: The European earth observation programme, *Revista de Teledeteccion*, 2020(56), v-xi, 2020, doi: 10.4995/RAET.2020.14346.

[6] Bigarella, M., Ferretti, M., Rocca, F., & Franceschetti, G., Monitoring structural deformations using InSAR and Sentinel-1 data, *IEEE Geoscience and Remote Sensing Letters*, 12(3), 481-485, 2015.

[7] Ullo, S. L. et al., Application of DInSAR technique to high coherence Sentinel-1 images for dam monitoring and result validation through in situ measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(3): 875-890, 2019.

[8] Sousa, J. J. et al., Potential of C-band SAR interferometry for dam monitoring. *Procedia Computer Science*, 100: 1103-1114, 2016.

[9] Lazecky, M. et al., Potential of satellite InSAR techniques for monitoring of bridge deformations. 2015 Joint Urban Remote Sensing Event (JURSE), 1-4, 2015.

[10] De Corso, T. et al., Application of DInSAR technique to high coherence satellite images for strategic infrastructure monitoring. IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 4235-4238, 2020.

[11] Huang, Q. et al., Displacement monitoring and health evaluation of two bridges using Sentinel-1 SAR images. *Remote Sensing*, 10(11), 2018.

[12] «ASI | Agenzia Spaziale Italiana». [Online]. Available: https://www.asi.it/en/earth-science/cosmo-skymed/

[13] Khan, S. S., Echevarria, E. R., & Hemer, M. A., Ocean Swell Comparisons Between Sentinel-1 and WAVEWATCH III Around Australia, *J Geophys Res Oceans*, 126(2), 2021, doi: 10.1029/2020JC016265.

[14] «Copernicus Browser». [Online]. Available: https://browser.dataspace.copernicus.eu/

[15] «APIs Documentation». [Online]. Available: https://documentation.dataspace.copernicus.eu/APIs.html

[16] «GeoJSON». [Online]. Available: https://geojson.org/

Monitoring and Structural Health Diagnosis of the Spyckstraße Bridge in Kleve

Jose Cano[1,*], Julia I. Real[1], Juan Ramón Torregrosa[1] and Jorge Pavón[1]

(1) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, 46022, València (Spain)

1 Introduction

The Spyckstraße Bridge, built in 1976, spans 92.19 meters across three sections of 30.73 meters each. It features a composite system of slab and prestressed beams, with 9 prestressed concrete beams and 8 longitudinal tendons per beam. The deck slab is transversely prestressed and 1.40 meters high. The bridge has 4 transverse joints, including two at the abutments and two between spans, with pre-curved tendons to prevent horizontal displacement. The infrastructure includes reinforced concrete abutments, two piers with circular shafts, and neoprene bearings supporting the beams. Figure 1 shows an overview of the structure under study.



Figure 1. General Overview of the Spyckstraße Bridge.

The Spyckstraße Bridge, nearly 50 years old, was initially built with the standards and techniques of its time, but now faces stricter load and regulatory requirements, along with aging. A 2022 inspection uncovered issues like concrete deterioration, corrosion of prestressing tendons, damaged bearings, and pavement defects, leading to repair recommendations. A 2024 expert report identified tendon fractures caused by water infiltration. To monitor the bridge, a two-week non-destructive testing system was installed, providing data to calibrate a Digital Twin model, which was refined to match experimental frequencies within a 3% margin. This model helps assess the bridge's current and future health, as well as its performance under various loads.

* jcanmor1@upv.es

* jcanmor1@alumno.upv.es

2 Methods

2.1 Monitoring system

To assess the Spyckstraße Bridge's current health, non-intrusive sensors were installed to measure accelerations, displacements, natural frequencies, and modal shapes across the deck, beams, and piers. The installation scheme included both fixed and mobile equipment. Fixed equipment recorded traffic and other actions over two weeks.

The fixed setup included 26 inertial sensors (■) and 12 displacement sensors (■), following the distribution shown in Figure 2. Beams are numbered 1 to 9 from East to West. Displacement sensors at the abutments monitored longitudinal movements of elastomeric bearings, especially at beams 1 and 9. Sensors along the spans measured deflections and dynamic properties, with accelerometers on beams 3, 4, 6, and 7 detecting vibrations from vehicle passage.

For the piers, inertial sensors at the tops detected movements in three directions, while displacement sensors monitored horizontal movements and checked neoprene bearings and span tendons for regulatory compliance.

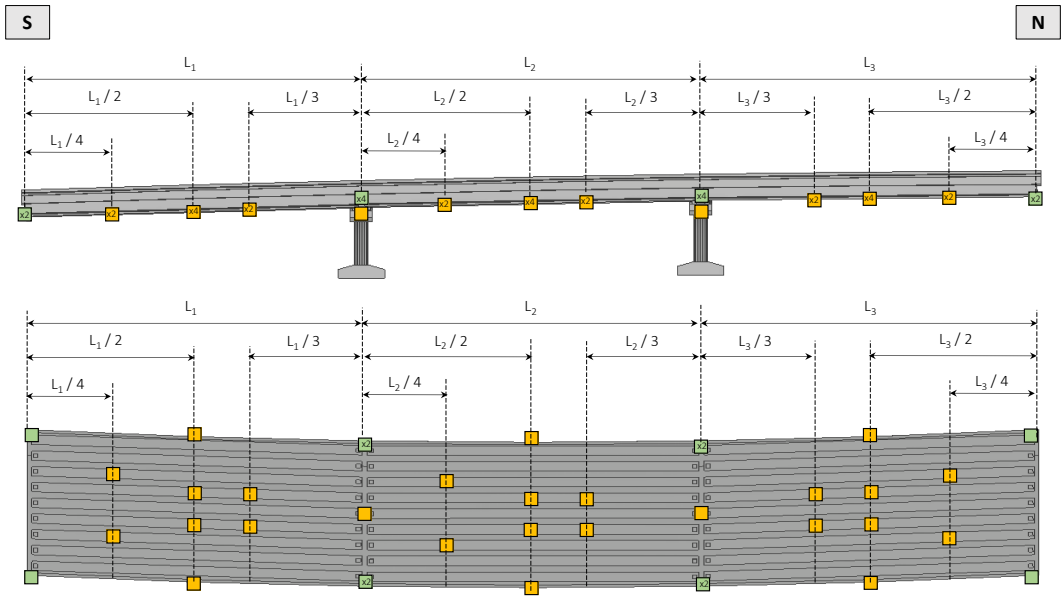


Figure 2. Scheme of fixed equipment installation on the Spyckstraße Bridge. Elevation and underside Views.

A mobile equipment setup, detailed in Figure 3, recorded traffic data for one hour in each section of the bridge. This system used 27 mobile inertial sensors placed across all beams and sections of the deck (L/2, L/3, L/4, and near supports), providing data for the calibration and validation of the bridge's Digital Twin.

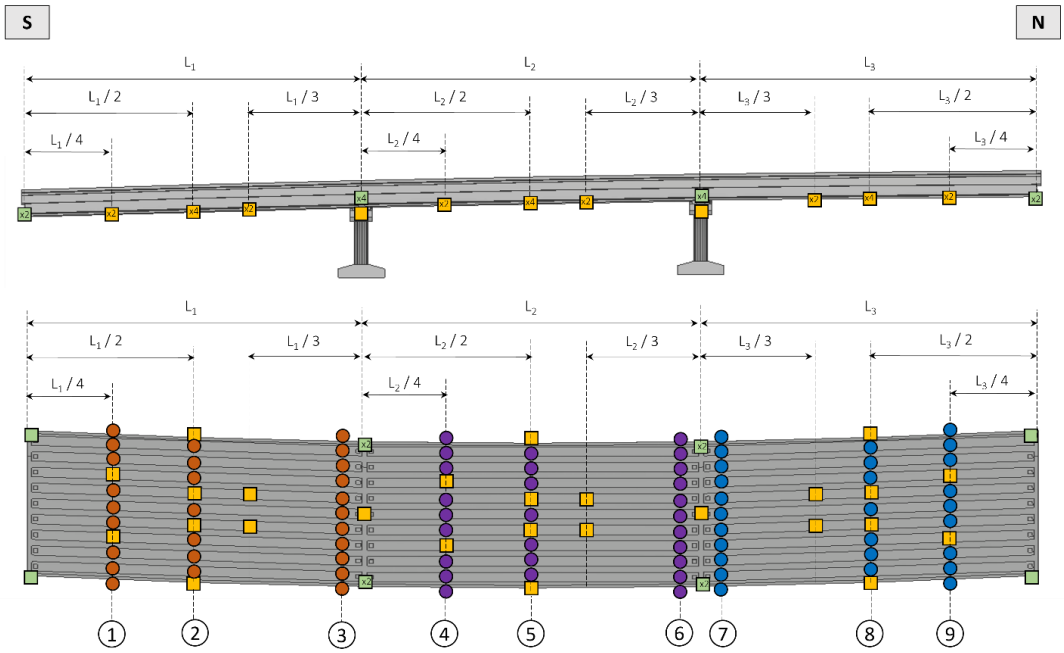


Figure 3. Mobile equipment installation scheme on the Spyckstraße Bridge. Elevation and underside views.

2.2 Digital Twin

The data from the targeted monitoring campaign of the Spyckstraße Bridge facilitated the creation of a numerical model, known as the Digital Twin, which was calibrated and validated using sensor data. This model helps assess the bridge's current and future structural health.

Calibration involved aligning experimental frequencies and vibration modes with those from the model, considering sensor placement, modal mass analysis, and prestressing effects. The model's accuracy was confirmed through site inspections and sensor data, ensuring reliable health assessment.

Sensor data was processed with mathematical algorithms to enhance dynamic identification for model calibration. Power Spectral Density (PSD) was calculated using Welch's method, selecting 80% of the most correlated data (0–40 Hz range). Peak detection identified natural frequencies, and an iterative genetic algorithm was applied until the model's frequencies matched experimental values within 3%. Key calibrated properties included Young's modulus and density of structural elements.

Figure 4 shows the deformed shapes associated with the vertical bending modes of each span in the Digital Twin of the Spyckstraße Bridge, while Table 1 presents the experimental and numerically obtained frequencies for the main vibration modes resulting from the calibration process, along with the corresponding errors between them.

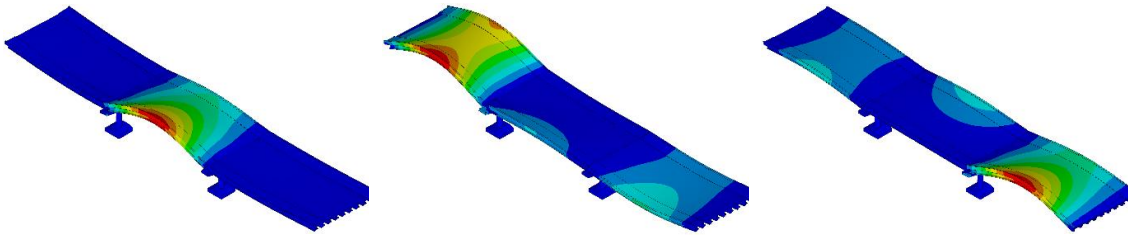


Figure 4. Deformed shapes associated with the vertical bending modes of each span in the Digital Twin of the Spyckstraße Bridge.

	Mode	Mode description	Measured frequency (Hz)	Calibrated model frequency (Hz)	Error
North Span	1	First-order vertical bending	3.75	3.79	0.97 %
	2	First-order torsion around the longitudinal	4.16	4.09	-1.64 %
	3	Second-order torsion around the longitudi-	5.41	5.57	2.96 %
Central Span	1	First-order vertical bending	3.58	3.60	0.45 %
	2	First-order torsion around the longitudinal	3.82	3.85	0.83 %
	3	Transverse bending + first-order torsion	4.88	4.90	0.46%
	4	Second-order torsion around the longitudi-	5.60	5.47	-2.32 %
South Span	1	First-order vertical bending	3.85	3.90	1.21 %
	2	First-order torsion around the longitudinal	4.29	4.23	-1.43 %
	3	Second-order torsion around the longitudi-	5.74	5.69	-0.89 %
North Pier	1	Longitudinal bending of the north pier	1.96	1.98	1.13 %
	2	Transverse bending of the north pier	2.14	2.16	0.99 %
South Pier	1	Longitudinal bending of the south pier	1.98	1.98	0.11 %
	2	Transverse bending of the south pier	2.17	2.16	-0.41 %

Table 1. Vibration modes resulting from the calibration and validation process of the Digital Twin of the Spyckstraße Bridge.

3 Results

3.1 Analysis of sensory data

The analysis of the sensor data provides insights into the structural behaviour of the bridge under traffic load across three spans and piers. The results show that vertical movement, both in acceleration and displacement, is generally larger than horizontal movement. No structural anomalies or risks are identified, but there is evidence of reduced stiffness in the western sections of the spans. This can be observed in Figure 5.

The evaluation of expansion joints reveals temperature-induced displacements within acceptable limits, although some issues related to tendon behaviour at joints suggest further monitoring is needed. Relative movements of 0.4 mm (Figure 6) have been recorded between the central span and the north span in the western area of the slab because of temperature changes—an effect that should be restrained by the joint tendons. This indicates potential anomalies in their performance due to possible deterioration.

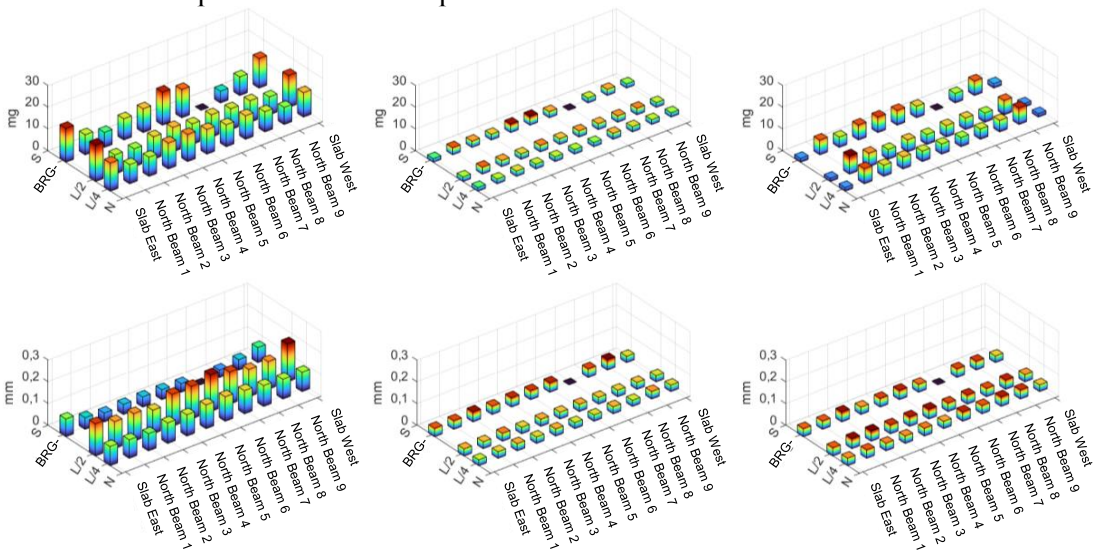


Figure 5. Median of the maximum acceleration (top) and displacement (bottom) values recorded with the mobile sensor scheme located in the north span. Vertical axis (left); longitudinal axis (centre); transversal axis (right).

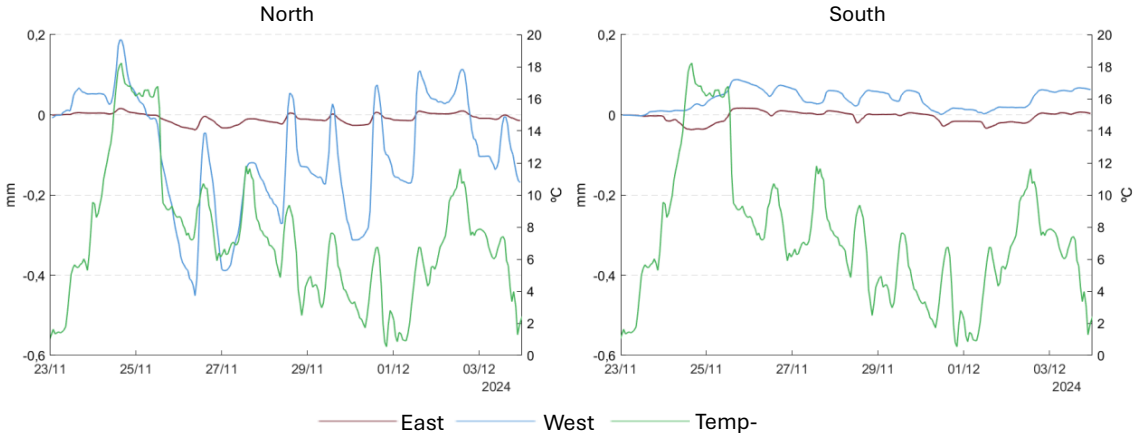


Figure 6. Hourly evolution of relative displacement at the joints between spans.

3.2 Advanced structural diagnosis

A calibrated and validated numerical model is used to assess the safety and functionality of the Spyckstraße Bridge under various load cases, including self-weight, traffic, snow, wind, thermal, and seismic loads, in line with DIN standards and European Eurocodes. The evaluation reflects the bridge's condition at the end of the monitoring campaign, analysing eight demanding load combinations in accordance with DIN EN 1990, providing insights into long-term performance.

The advanced structural diagnosis examines stress levels and utilization percentages to identify critical sections and early signs of deterioration. All structural elements of the bridge perform correctly under the evaluated load combinations, with no depletion of cross-sectional capacity identified in high demand loading scenarios, as shown in Figure 7.

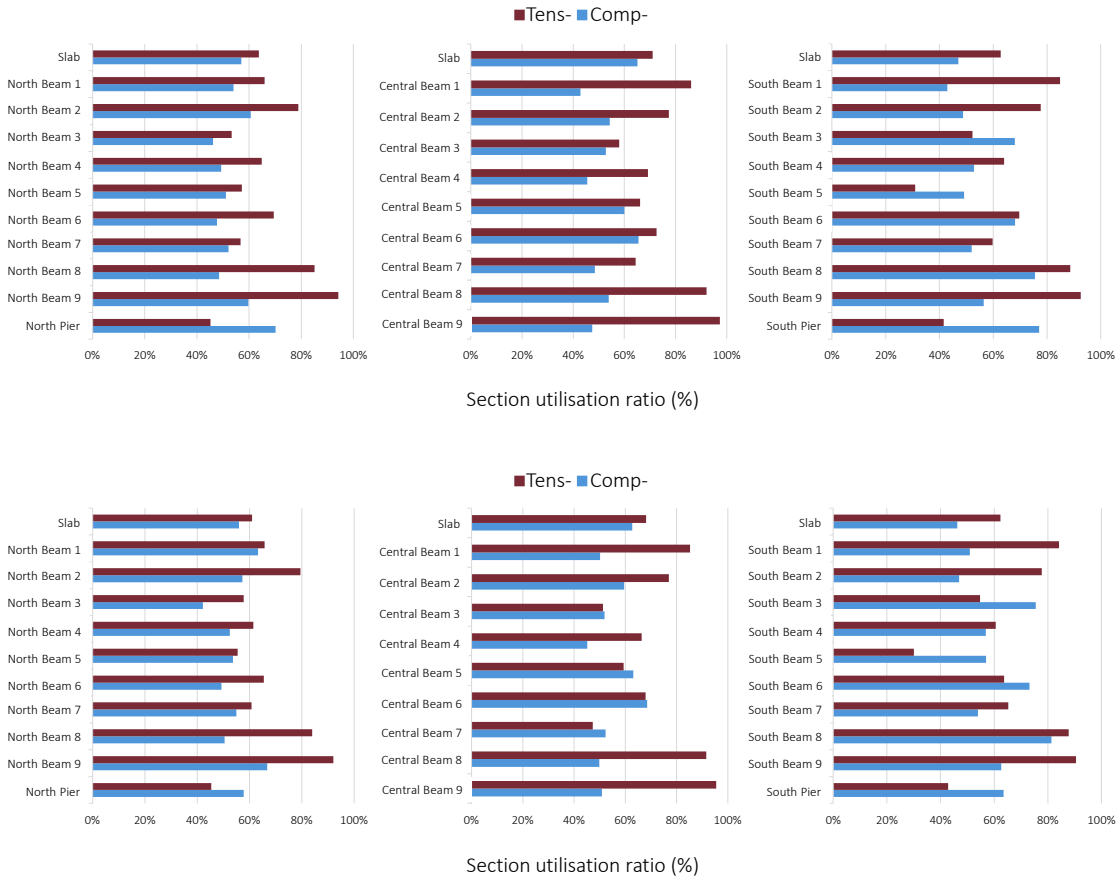


Figure 7. Maximum sectional utilisation factor under tension and compression for each structural element of the northern span (left), central span (centre), and southern span (right) under Load Combination No. 2 (top) and Load Combination No. 6 (bottom).

Compression stresses peak under extreme temperature expansion (load combination No. 6), with the highest utilization factors in beams 9 (north span, 66.8%), 6 (central span, 68.6%), and 8 (south span, 81.3%).

Tensile stresses, which pose a higher risk to concrete and prestressed steel, are most significant under traffic load (combination No. 2) and temperature expansion (combination No. 6). The highest tensile stresses are found in beams closest to the west side, with utilization factors reaching 94.2% for beam 9 (north span), 97.3% for beam 9 (central span), and 92.5% for beam 9 (south span).

The analysis concludes that no structural risk is present under high-demand load scenarios as of the monitoring date. However, structural elements with utilization factors above 80% under extreme conditions should undergo thorough analysis to prevent potential deterioration, especially the beams on the west side (beams 8 and 9), which experience the highest tensile forces.

Due to the particularity found in the expert report of 2024 regarding the corrosion and even fracture of one of the tendons connecting the spans, a tailored simulation for the Spyckstraße Bridge assessed the structural impact of tendon degradation at span joints, focusing on changes in natural frequencies. Results showed that tendon deterioration—especially near the west side—alters the dynamic response, with critical variations in vertical bending and torsional modes (Figure 8). Structural capacity would be exceeded after the failure of 3–4 tendons in key beams. The study highlights the need for continuous monitoring to detect early damage and prevent structural compromise.

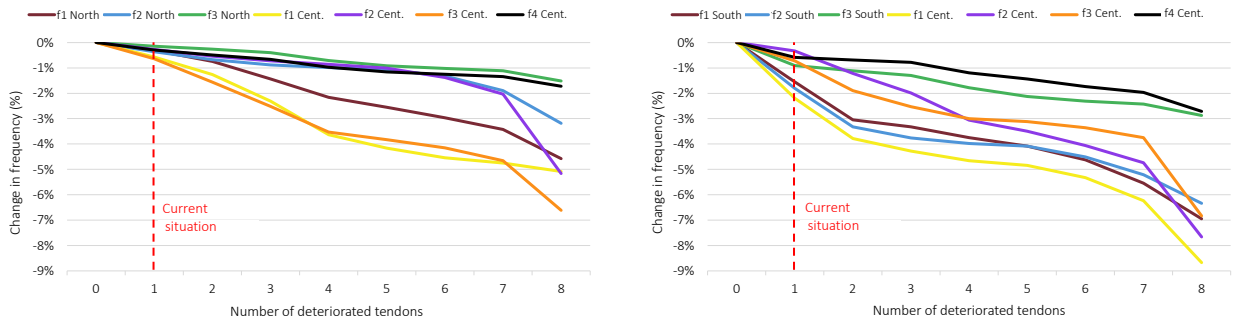


Figure 8. Variation of the natural frequencies of the different vibration modes identified with the number of deteriorated tendons at the joint between the north and central spans (left) and at the joint between the south and central spans (right).

4 Conclusions

A specific diagnostic of the structural health of Spyckstraße Bridge was conducted using a sensor system to measure accelerations and displacements. The data helped calibrate and validate the Digital Twin with an error of less than 3%. The bridge's response to wind,

temperature changes, and traffic was analysed, revealing no structural risks, but minor relative movements (0.4mm) were noted due to temperature variations, indicating potential tendon issues. Future inspections should focus on the northern span joint.

An advanced structural analysis confirmed the bridge is in good health and does not require immediate reinforcement. However, simulations incorporating progressive tendon corrosion in the joints show significant changes in natural frequencies and reduced load-bearing capacity, indicating a potential future risk if deterioration advances. Thus, the implementation of a continuous monitoring is recommended to detect early anomalies, such as tendon or steel corrosion, and improve long-term maintenance, safety, and decision-making. This system would optimize resources and minimize risks by providing real-time data on structural behaviour.

References

[1] Deutsches Institut für Normung (DIN), DIN EN 1990: Eurocodes. Basis of Structural Design. Berlin, Beuth Verlag, 2002.

[2] Deutsches Institut für Normung (DIN), DIN EN 1991-1-1: Eurocode 1: Actions on Structures. Part 1-1: General Actions. Dead Loads, Imposed Loads. Berlin, Beuth Verlag, 2002.

[3] Deutsches Institut für Normung (DIN), DIN EN 1991-1-3: Eurocode 1: Actions on Structures. Part 1-3: General Actions. Snow Loads. Berlin, Beuth Verlag, 2003.

[4] Deutsches Institut für Normung (DIN), DIN EN 1991-1-4: Eurocode 1: Actions on Structures. Part 1-4: General Actions. Wind Actions. Berlin, Beuth Verlag, 2005.

[5] Deutsches Institut für Normung (DIN), DIN EN 1991-1-5: Eurocode 1: Actions on Structures. Part 1-5: General Actions. Thermal Actions. Berlin, Beuth Verlag, 2005.

[6] Deutsches Institut für Normung (DIN), DIN EN 1991-2: Eurocode 1: Actions on Structures. Part 2: Traffic Loads on Bridges. Berlin, Beuth Verlag, 2003.

[7] Deutsches Institut für Normung (DIN), DIN EN 1992-1-1: Eurocode 2: Design of Concrete Structures. Part 1-1: General Rules and Rules for Buildings. Berlin, Beuth Verlag, 2004.

[8] Deutsches Institut für Normung (DIN), DIN EN 1992-2: Eurocode 2: Design of Concrete Structures. Part 2: Concrete Bridges. Design and Construction Rules. Berlin, Beuth Verlag, 2005.

[9] Deutsches Institut für Normung (DIN), DIN EN 1998-2: Eurocode 8: Design of Structures for Earthquake Resistance. Part 2: Bridges. Berlin, Beuth Verlag, 2005.

Towards a geometric optimization of Bioreceptive Bricks in Vertical Facade Systems

M.J. López [1,*], L. Hilario[1], P. Verdejo[2] and N. Montes[1].

(1) University CEU Cardenal Herrera, CEU Universities,
55 San Bartolomé St., 46115 Alfara del Patriarca, Valencia, Spain .
(2) Universitat Politècnica de València,
Camino de Vera s/n, Building 2F, Algirós, 46022 Valencia, Spain.

1 Introduction

Over the past few years, multiple disciplines have converged to explore the relationship between the geometric form of a surface and its capacity to support life. One of the most notable examples of this can be found in so-called vertical gardens (see [1], [2], and [3]) In this context, the design of the geometric pattern on the surface—typically concrete facades or modular panels—plays a crucial role in water retention, biofilm adhesion, and the colonization by plants or moss.

2 Methods

From a mathematical perspective, the problem involves classical questions of surface tessellation, total length and area optimization, and controlled curvature (see [4], [5], [6], and [7]). While regular tessellations (triangular, square, or hexagonal) have traditionally been studied for their significance in geometry and crystallography, recent work has extended these patterns into three-dimensional configurations with indentations (depressions or concavities) (see [8],[9]). Additionally, the application of Catmull-clarck and Laplacian smoothing methods has emerged as a relevant tool in this context (see [10],[12]).

3 Results

The study explores the geometric consequences of applying three-dimensional indentations and smoothing techniques to regular tessellations. These transformations result in a significant increase in the exposed surface area of each unit, which is quantitatively evaluated across different tessellation types.

* migueljose.lopezgil@alumnos.uchceu.es

4 Conclusions

The present study introduces a working framework for the optimal design of a brick capable of maximizing exposed surface area and, consequently, increasing the likelihood of supporting life. This approach combines insights from geometry, mathematical modeling, and computational design to propose a generalizable strategy for enhancing bioreceptivity through form.

References

[1] Manso, M. dan Castro-Gomes, J., Green wall systems: A review of their characteristics, *Renewable and Sustainable Energy Reviews*, 41:863–871, 2015.

[2] Mustafa, K.F., Prieto, A. dan Ottele, M., The role of geometry on a self-sustaining bio-receptive concrete panel for facade application, *Sustainability (Switzerland)*, 13(13), 2021.

[3] Lundholm, J., Green Roofs and Facades: A Habitat Template Approach, *Urban Habitats*, 4, 2006.

[4] Grünbaum, B. dan Shephard, G.C., Tilings and Patterns - Branko Grünbaum, G. C. Shephard - Google Libros, 1987. [Daring]. Tersedia di: https://books.google.es/books?hl=es&lr=&id=0x0vDAAAQBAJ&oi=fnd&pg=PR10&ots=kfYsBs4xMD&sig=LAEKifW9yRFivEVI7VYhFpOGJoM&redir_esc=y#v=onepage&q&f=false. [Diakses: 15-Feb-2025].

[5] Kaplan, C.S. dan Salesin, D.H., Dihedral Escherization, *SIGGRAPH2000*, hal. 499–510, 2000.

[6] Senechal, M., Quasicrystals and geometry, 1995.

[7] Stewart, I., Does God play dice?: The new mathematics of chaos, 1997.

[8] Pottmann, H., Eigensatz, M., Vaxman, A. dan Wallner, J., Architectural geometry, *Computers & Graphics*, 47, 145–164, 2015.

[9] Lewandowski, D., Robain, H., Clergeau, P. dan Le Roy, R., Bioreceptivity of living walls: Interactions between building materials and substrates, and effect on plant growth, *Urban Forestry and Urban Greening*, 83, 2023.

[10] Catmull, E. dan Clark, J., Recursively generated B-spline surfaces on arbitrary topological meshes, *Seminal graphics*, 183–188, 1998.

[11] Herrmann, L.R., Laplacian-Isoparametric Grid Generation Scheme, *Journal of the Engineering Mechanics Division*, 102(5), 749–756, 1976.

[12] Salazar, S.S., Hortal, J.C., Menárguez, F.J.M., Salazar, S.S., Hortal, J.C. dan Menárguez, F.J.M., Modelos de brotes arbustivos o algas en arquitectura: O cómo replicar un vegetal mediante la Agregación Limitada por Difusión (DLA), *[i2] : Investigación e Innovación en Arquitectura y Territorio*, 5(1), 2017.

Innovative and intelligent technology for in situ early diagnosis and prediction of concrete carbonation progress for predictive maintenance of bridges

Javier Maldonado[1], Julia I. Real[1], Teresa I. Real[1] and José Ignacio Hervás[2]

(1) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022, València (Spain)

(2) IMESAPI, Avenida de Ramón y Cajal, 107, 28043 Madrid (Spain)

1 Introduction

The durability of concrete structures, particularly bridges, is affected by carbonation, a process that degrades the material and compromises long-term structural integrity. Monitoring and predicting the progression of carbonation is necessary to implement predictive maintenance strategies and ensure the longevity of these infrastructures. However, traditional assessment methods often lack the capability for early and continuous in-situ diagnosis. This work presents the development of an innovative and intelligent technology based on a low-cost multi-sensor unit for early in-situ diagnosis and prediction of the progress of carbonation in concrete.

The system integrates electrical resistivity measurements using an in-house designed Wenner probe, together with sensors for humidity, ambient temperature, internal concrete temperature, and the presence of surface moisture. The goal is to use the collected data, in combination with probabilistic algorithms and artificial neural networks (ANN), to predict the evolution of the carbonation depth, considering multiple material and environmental factors, thus facilitating predictive bridge maintenance.

This paper details the methodology employed, the development of the hardware, the selection of pilot test sites, the installation of the system and the preliminary results obtained.

The CAREC project (CPP2021-008857) is part of the 2021 public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by ImesAPI S.L., API Movilidad, IDVIA 2020 Horizonte 2020 S.L. and the Polytechnic University of Valencia as a public research body.

2 How to measure concrete carbonation

2.1 Measuring the carbonation penetration

The methodology of this study involves a detailed analysis of the factors relevant to the experimental design, as well as the identification of the sensor modules and other components necessary for the assembly of a required multi-sensor unit.

* javier.maldonado@idvia.es

The electrical resistivity of concrete, measured using the Wenner Probe [1], is a crucial factor in the study of carbonation. This method involves the excitation of the concrete sample by the ends of the four-pronged electrode, measuring the voltages at the intermediate points. In addition, a sensor is incorporated which measures the circulating current, as illustrated in Figure 1. Dividing the voltage by the measured current gives the electrical resistance of the sample, which is then used to calculate the resistivity of the concrete, assuming a homogeneous distribution of voltages over a half-sphere. This provides valuable information on the degree of carbonation and the state of deterioration of the concrete.

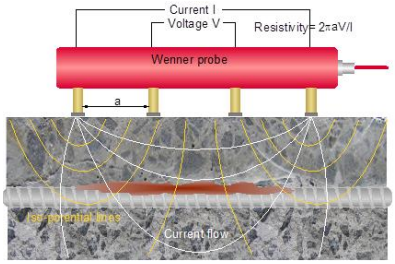


Figure 1. Wenner Probe configuration. Source: [2].

$$\rho = \frac{2\pi aV}{I}$$

Where: ρ is the electrical resistivity (kΩ-cm), a = distance between electrodes (cm), V is the voltage (V) and I is the current (A).

It is proposed to develop a new multi-sensor unit that also incorporates humidity and temperature recordings on the surface of the concrete element, to improve the accuracy of electrical resistivity measurements.

Based on the carbonation depth measurements obtained with the newly developed equipment, probabilistic algorithms will be created to predict how carbonation depth evolves over time. These algorithms will be formulated using an advanced mathematical framework that inputs results from various laboratory tests on concrete specimens, each produced with different mix designs and subjected to accelerated CO_2 exposure levels [3].

To extract actionable insights from the collected data, advanced processing techniques will be employed, leveraging artificial neural networks (ANN). The neural network model will incorporate a range of influential parameters, including cement type, water-to-cement ratio, aggregate composition, and admixtures, as well as environmental conditions such as atmospheric CO_2 concentration, temperature, humidity, and specimen saturation. Accounting for all these factors is essential for accurately and efficiently predicting concrete carbonation depth.

$$x(t_f) = x_0 + \int_{t_0}^{t_f} k(t) \cdot \sqrt{t}\, dt = x_0 + \sum_{t_0}^{t_f} K_{tx} \cdot \sqrt{t_x - t_{x-1}}$$

Where x represents the carbonation penetration in mm, K is a factor depending on environmental conditions and t is the elapsed time in years.

2.2 Hardware development

To enable reliable correlation between electrical resistivity changes and concrete carbonation, a compact multisensor unit was fabricated, integrating measurements of environmental conditions, internal concrete temperature, moisture presence, and surface resistivity. Sensor selection prioritized accuracy, stability, ease of integration, and cost-effectiveness.

For ambient temperature and relative humidity, the AM2315C thermohygrometer was chosen due to its integrated I²C interface, compact dimensions (98 mm × 16 mm), low power consumption (< 10 mA at 3.5–5.5 V DC), and measurement accuracy (±0.1 °C and ±2 % RH), supporting full-range humidity sensing (0–100 %) at up to 0.5 Hz sampling.

Internal concrete temperature, which directly influences ionic mobility and thus resistivity, is monitored by the coaxial DS18B20 (DFR0198). This stainless-steel–encapsulated sensor (35 mm × 6 mm) operates from –10 to 85 °C with ±0.1 °C accuracy at up to 0.5 Hz, draws 3.5–5.5 V DC, and communicates via a single shared 1-Wire bus—allowing multiple devices on one microcontroller pin. Its resilience to standard structural environments, adjustable 9–12-bit resolution, and built-in alarm thresholds for real-time monitoring make it an optimal choice.

To account for pore water conduction effects on resistivity readings, a simple optical raindrop detector (FC-37) was incorporated for real-time detection of surface moisture. Operating from 3.3–5 V DC, it provides an analog output proportional to water coverage (0–100 %, ±10 %) and is housed in an outdoor-rated enclosure (DRD11A). Its low cost and continuous-response capability enables precise temporal correlation between precipitation events and resistivity fluctuations.

These sensors, developed as part of the hardware system of the solution and previously discussed, are illustrated in Figure 2.



Figure 2. Temperature and relative humidity sensor, internal concrete temperature sensor and surface moisture sensor. Source: Digikey.

Surface resistivity is measured non-destructively by a custom Wenner-style four-electrode probe. Commercial alternatives (e.g., Proceq Resipod) were considered to have high acquisition costs, prompting an in-house design comprising four stainless-steel electrodes (25 mm long, 6 mm diameter) spaced uniformly at 38 mm. Embedded within a printed circuit board (PCB) like the one shown in Figure 3, the unit integrates a regulated power supply, an analog-to-digital converter for electrode voltage readings, and a microcontroller for data processing and external communication. Environmental compensation is achieved via the on-board temperature and humidity sensor, while a status LED and reed-switch–activated magnetic on/off control facilitate remote operation.

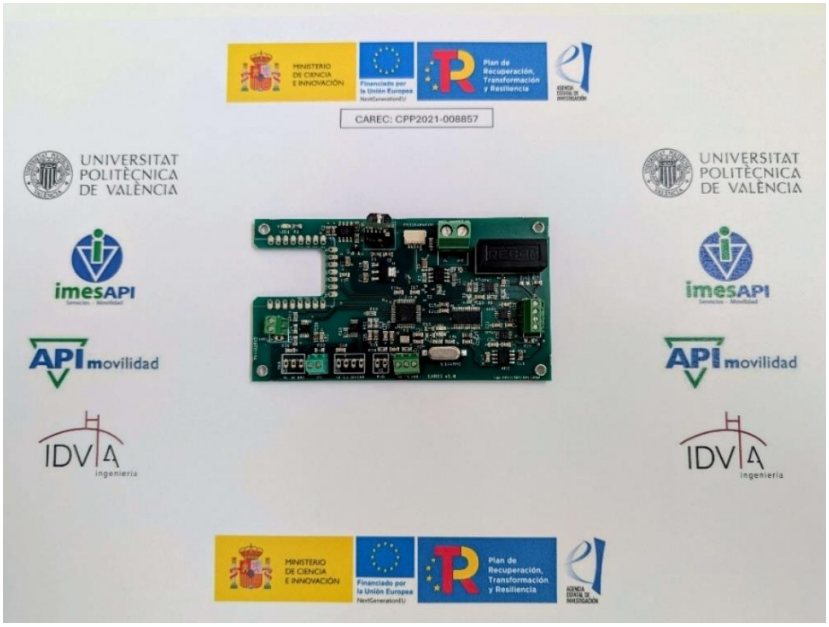


Figure 3. Manufactured PCB. Source: Own elaboration.

Probe geometry follows the principle that electrode spacing should exceed penetration depth but remain below concrete cover thickness to minimize rebar interference; in practice, fixing electrode length and varying interelectrode distance optimizes accuracy in reinforced sections [4]. The probe housing of the four electrodes has been designed and manufactured using ABS 3D printing technology, allowing for low cost and rapid prototyping capabilities for testing. This probe housing is shown in Figure 4.



Figure 4. 3D designed probe housing. Source: own elaboration.

This assembled multi-sensor unit offers a low-cost, modular solution for continuous monitoring of the key parameters governing concrete resistivity and carbonation, thereby supporting probabilistic modeling and long-term durability assessment.

3 Progress and results

3.1 Test site selection

Once the hardware system has been designed, laboratory tests have been conducted in a controlled environment on all the elements to verify that they function properly and that the measurements are within the admissible tolerances. The next step was to select two locations

to start testing the system. Two bridges (the P.I. above CV-309 and Barranco Ratils) were selected based on a prior susceptibility assessment, with documented carbonation damage levels ranging from superficial (N1) to severe (N3).

The Barranco Ratils bridge, near of Burriana (Castellón) town has a first carbonation damage in the abutment element and in the gable/parament sub-element is significant, as significant degradation of the concrete is observed (Figure 5), with spalling and loss of material affecting key areas of the structure, equivalent to a damage level N1.



Figure 5. Damage 1 by carbonation on the Barranco Ratils bridge. Source: Imesapi.

The second damage is in the span element and in the deck/substructure support slab sub-element, severity level N2, it is new damage, with moderate damage to the concrete, as shown in Figure 6.



Figure 6. Damage 2 by carbonation on the Barranco Ratils bridge. Source: Imesapi.

The third damage is in the pile element and in the shaft sub-element, severity level N2, without evolution, with moderate damage to the concrete that has not shown progress over time (Figure 7).



Figure 7. Damage 3 by carbonation on the Barranco Ratils bridge. Source: Imesapi.

3.2 System instalation

The installation of the system on the Ratils Ravine bridge was carried out according to the installation scheme shown in Figure 8.

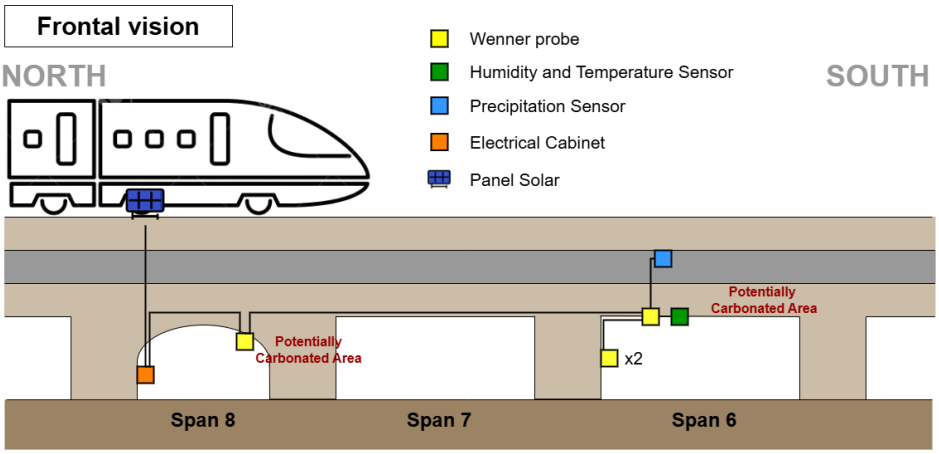


Figure 8. Installation diagram of the CAREC system on the Barranco Ratils Bridge. Source: own elaboration.

In the central cabinet, the data acquisition module, the 4G wireless communication module and the system power supply, consisting of a 12V battery and a solar charge regulator connected to a 10 W solar panel, were installed. Both the central cabinet and the solar panel are shown in Figure 9.



Figure 9. Central cabinet and solar panel installed on Barranco Rentils Bridge. Source: own elaboration.

For the installation of the self-designed Wenner probes, it is necessary to drill four holes in the concrete for the sensor nodes, 38 mm apart. These holes should be 25 mm deep and 6 mm in diameter, ensuring a precise fit for the insertion of the four nodes. In addition, two additional holes should be drilled to fix the probe to the concrete using plugs and screws, located 23 mm from the ends of the nodes. Figure 10 shows some images of the installation of these sensors.



Figure 10. Multi sensor node installation on Barranco Rentils bridge. Source: own elaboration.

3.3 First results

After the installation, the first measurements of the system were carried out on the Barranco Rentils bridge, which can be seen in the graphs in Figure 11 and Figure 12.

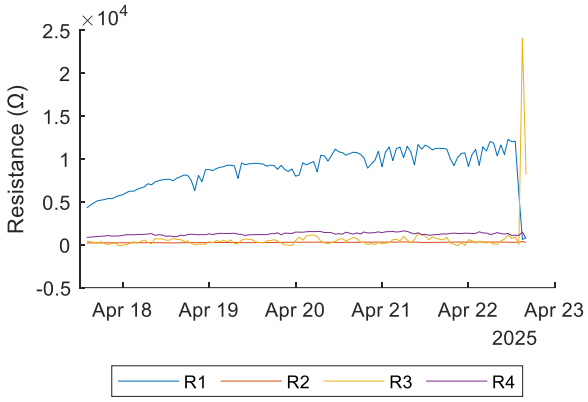


Figure 11. Measurement of concrete resistivity in Barranco Rentils bridge. Source: own elaboration.

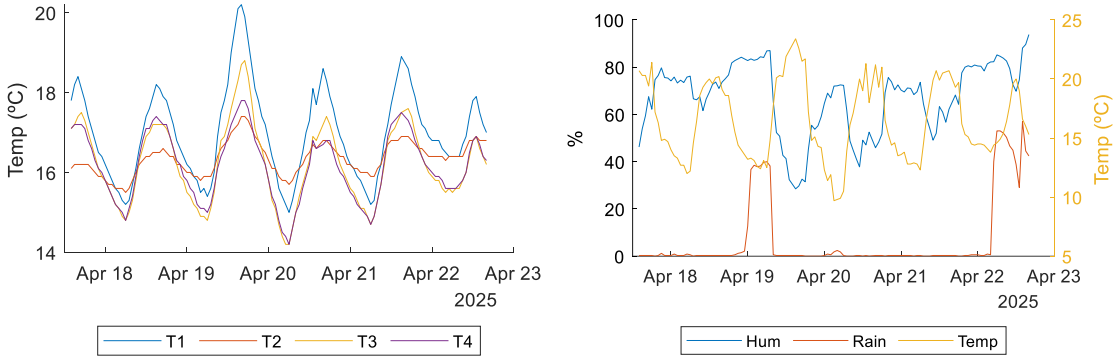


Figure 12. Measurements of internat concrete temperature (left) and environmental conditions (right) in Barranco Rentils bridge. Source: own elaboration.

4. Conclusions and next steps

This study has successfully designed, manufactured, and validated a low-cost multi-sensor unit for in-situ monitoring of concrete carbonation in bridges. The system integrates in a compact form the non-destructive measurement of concrete surface resistivity by means of a four-electrode Wenner-type probe of its own design, together with the monitoring of environmental conditions (temperature and relative humidity), the internal temperature of the concrete and the detection of surface moisture.

The data collected in the field installation at the Ratils Ravine bridge will serve as input for the development of probabilistic algorithms and predictive models based on Artificial Neural Networks (ANN). These models will seek to predict the temporal evolution of the carbonation depth considering a wide range of material parameters and environmental conditions.

Preliminary results from the field installation demonstrate the functionality of the system and its ability to record relevant data on a continuous basis. This technology represents a new tool for early diagnosis and prediction of deterioration, laying the foundation for optimised management and predictive maintenance of concrete bridges.

The next steps will focus on long-term data collection, laboratory testing through accelerated carbonation and refinement of predictive models.

References

[1] Morris, W., Moreno, E. I., Sagüés, A. A., Practical evaluation of resistivity of concrete in test cylinders using a Wenner array probe. *Cement and concrete research*, 26(12): 1779-1787, 1996.

[2] Federal Highway Administration, Electrical Resistivity Measurement Techniques for Concrete. Infotechnology [online], available in: https://infotechnology.fhwa.dot.gov/electrical-resistivity

[3] Campos Silva, A. R., Fajardo, G., Mendoza-Rangel, J. M., Estudio del comportamiento del avance de la carbonatación del concreto reforzado en ambiente natural y acelerado. *Concreto y cemento. Investigación y desarrollo*, 8(1): 14-34, 2016.

[4] Chinè-Polito, B., Cuevas-Kauffmann, R., Jiménez-Salas, R., Ortiz-Quesada, G., Estudio experimental de la carbonatación del concreto. *Revista Tecnología En Marcha*, 32(2): 68-81, 2019.

Continuous monitoring system for the safe storage and use of hydrogen in hydrogenerators

Alejandro Marco[1,*], Julia I. Real[1] , Sara Vicedo [2] and Rocío Castelo[3]

(1) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022, València (Spain)
(2) Vainsa Infraestructuras S.L., Edificio Albufera Center, 46910 Alfafar, Valencia (Spain)
(3) Instituto Tecnológico y de Energías Renovables, S.A., Polígono Industrial de Granadilla, s/n 38600 Granadilla de Abona, Santa Cruz de Tenerife (Spain)

1 Introduction

Transport is one of the main contributors to climate change. Emitting 32% of CO2 emissions in the European Union (EU) [1]. In this context, a key technological issue revolves around the hydropower infrastructure (HRS) needed for the energy transition.

It is precisely during the storage stage that hydrogen presents the greatest challenges, the most notable being its tendency to explode when mixed with oxygen in a sudden manner: as in a storage tank or valve rupture that causes a sudden release of hydrogen at high pressure.

Current techniques or systems for leak identification include electromechanical impedance, hydrogen sensitive sensors, fiber optics, acoustic cameras or acoustic sensors. Unfortunately, these systems are not able to effectively and inexpensively identify leaks in real time and on a continuous basis. In this context, the need is identified to develop a system capable of detecting leaks quickly and effectively, with the ability to intervene before serious accidents occur.

Therefore, the objective is to develop a novel integral system for detection and management of gas leaks in HRS "LEAKWIT". Capable of I) performing real-time and continuous leak detection; II) obtaining the location and size of the leak in the main components of the HRS; and III) providing a real-time risk management and analysis system.

The LEAKWIT project (CPP2021-008524) is part of the 2021 public-private collaboration projects, framed in the State Program to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by ITER (*Insituto Tecnológico y de Energías Renovables*), VAINSA, and the Polytechnic University of Valencia as a public research organisation.

*almargu1@upv.es

2 Methods

To achieve success in the development of this innovative hydrogenator management system, the system will consist of the following subsystems.

2.1 Data acquisition subsystem

It will consist of ultrasonic microphones distributed throughout the HRS, which will be connected to a single data acquisition system (DAQ) [2]. The system also requires knowledge of the temperature and pressure of the tanks, so these variables will be monitored through the use of thermocouples and pressure sensors. (Figure 1)

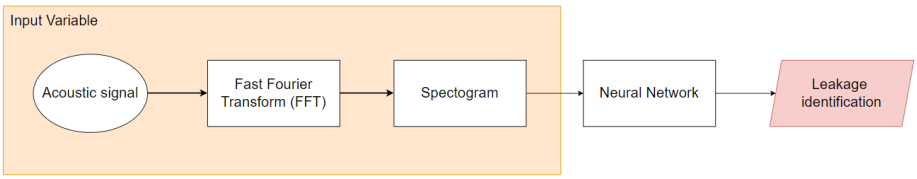


Figure 1. Data acquisition flowchart and Neural Network operation. Source: own elaboration.

The neural network will be based on the use of spectrograms, which are a three-dimensional representation of the sound intensity (amplitude) of a time-varying signal at different frequencies [3].

2.2 Data processing subsystem

Composed of the two-dimensional triangulation algorithm to accurately locate the position of the leak; and a Neural Network in charge of first level leak detection, which will be trained from the Digital Twin. (Figure 2)

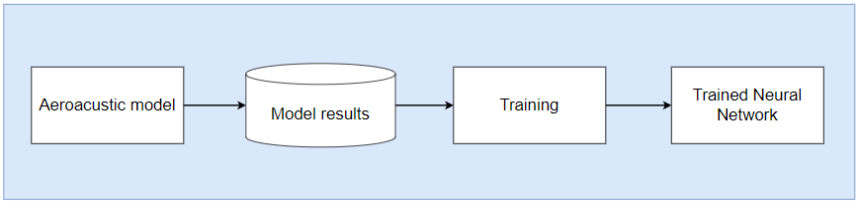


Figure 2. Neural Network training flowchart. Source: own elaboration.

2.3 Digital Twin

It is defined as a comprehensive physical and functional description of the system. In the context of the present proposal, it can be defined as the replication of tanks and piping arranged in the HRS (geometry, state of conservation and other relevant characteristics) [4].

The Digital Twin is fed by the data coming from the instruments and sensors installed in the components such as flow meters, temperature and pressure registers, etc. allowing it to be constantly adjusted and calibrated.
In this way the Digital Twin will allow to replicate the leaks detected at the first level by the acoustic sensors and the Neural Network, generating scenarios similar to reality with which to verify the leak. (Figure 3)

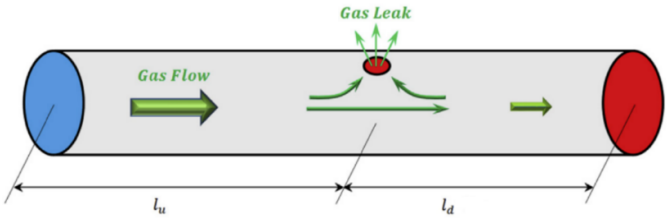


Figure 3. Schematic of the physical model implemented for the Digital Twin of hydrogen distribution pipelines. Source: own elaboration.

In this way, the Digital Twin will make it possible to replicate the leaks detected at the first level by the acoustic sensors and the Neural Network, generating scenarios similar to reality with which to verify the leak, thus reducing the number of false positives.

2.4 Digital Platform

Which will allow access to the information obtained by the system in real time and in a clear, precise an easily understandable manner. The results can be monitored from the platform. The platform will also be responsible for risk management.

3 Results

The solution is based on two independent levels of leak detection in order to reduce the number of false positives. The first level is based on the use of acoustic sensors for continuous monitoring of storage tanks and pipelines present in the hydrogenerators.

In the second level, the detected leak is automatically entered into a Digital Twin of the components that make up the hydrogenerator. The simulation is performed in order to verify the temperature and pressure, as well as the flow in the pipeline corresponds to the one recorded in-situ.

The solution provides an automatic management system and risk analysis of the hydrogen plant based on the obtained results. (Figure 4)

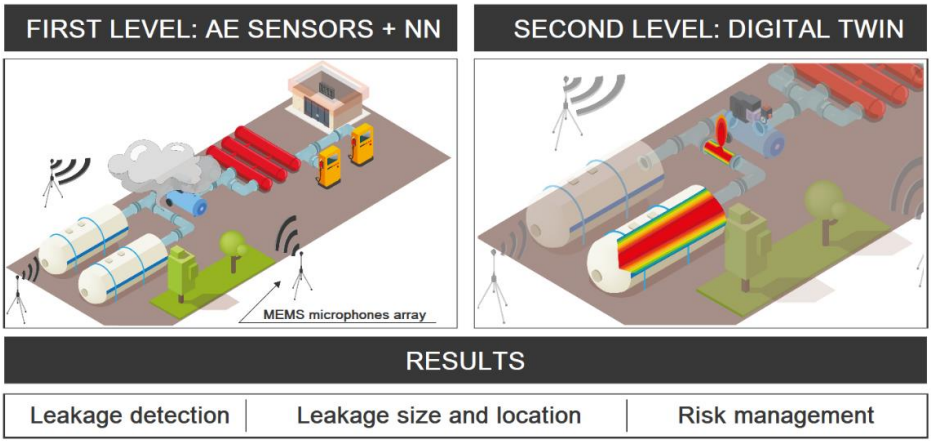


Figure 4. Solution schematic. Source: own elaboration.

For this purpose, following the proposed methodology and in accordance with the development carried out so far, the complete system is made up of the following developments.

I) Development of a low cost and high sensitivity gas detection sensor from the use of an ultrasonic microphone array (MEMS)

The design of the leak detection system for hydrogen tanks and pipelines consists therefore of an array of ultrasonic microphone sensors in charge of collecting the acoustic signals of leaks at different flow rates.

The objective of this will be to determine the frequency response in high sensitivity that characterize this type of leaks, allowing to know the occurrence of hydrogen leaks.

Furthermore, after its manufacture (Figure 5), it has been tested with high-frequency sound emissions, where it has been possible to verify the correct acquisition with precision. (Figure 6)



Figure 5. MEMS microphone array. Source: own elaboration.

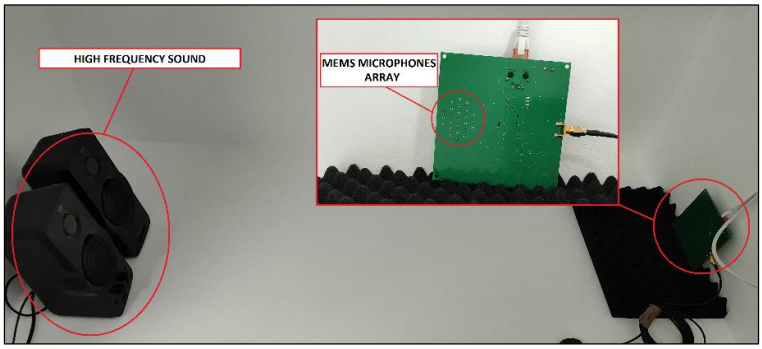


Figure 6. MEMS microphone array acquiring high frequency sound. Source: own elaboration.

II) Development of a signal triangulation algorithm and detection neural network.

The two-dimensional triangulation method is chosen as the leakage location algorithm using the different acoustic microphone arrays.
The proposed triangulation method relies on the calculation of the Time Difference Of Arrival (TDOA), with which the sound source is located using geometric relationships between the sound source and the microphones. (Figure 7)

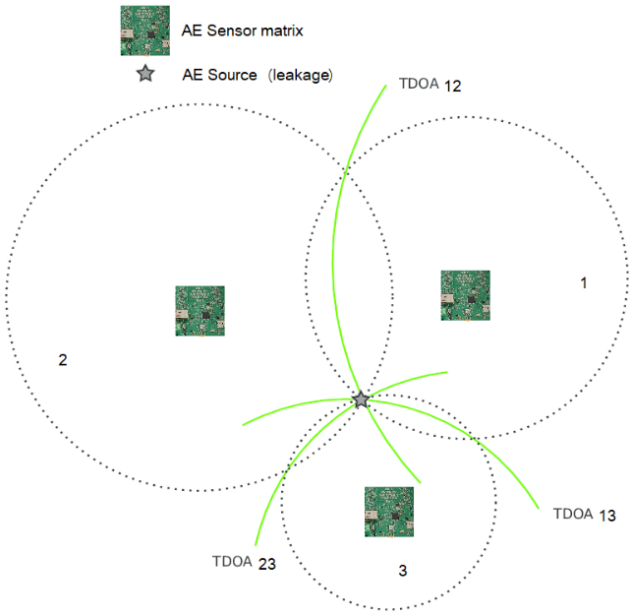


Figure 7. TDOA leakage location algorithm. Source: own elaboration.

On the other side, a model based on convolutional neural networks (CNN) with *Fbank* spectrograms is used for the detection of hydrogen leakage [3]. (Figure 8)

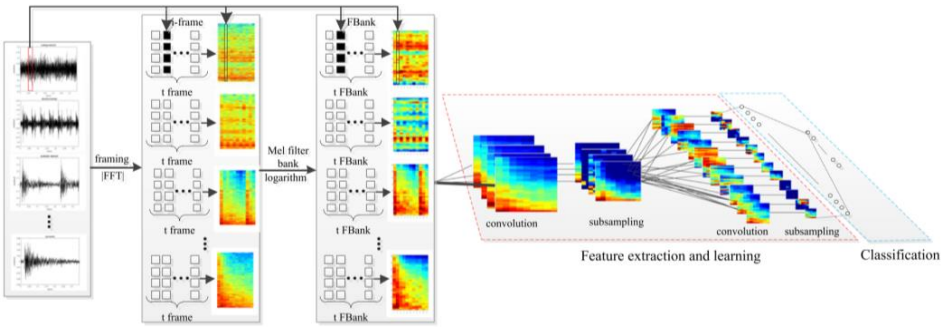


Figure 8. CNN and Fbank spectrum algorithm based on noise recognition. Source: [3].

III) Development of a Digital Twin (with capability to simulate leaks) integral to the hydrogen storage tanks and distribution piping commonly present at HRSs.

In the digital twin of the HRS, in charge of the second level of leak detection, the leaks are entered into the developed model of the hydrogen plant elements, i.e. storage tanks (Figure 9) and distribution pipelines (Figure 10).

Verification simulations of temperatures, pressures, flow rates... and their extrapolation to reality will be carried out for comparison.

In addition, the digital model will allow calibration of the software models developed for the identification of hydrogen leaks.

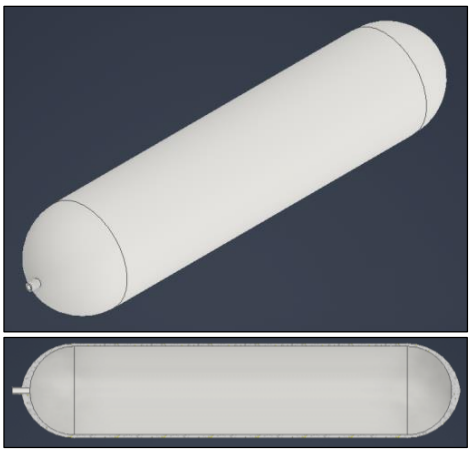


Figure 9. Digital model of the hydrogen storage tank. Source: own elaboration.

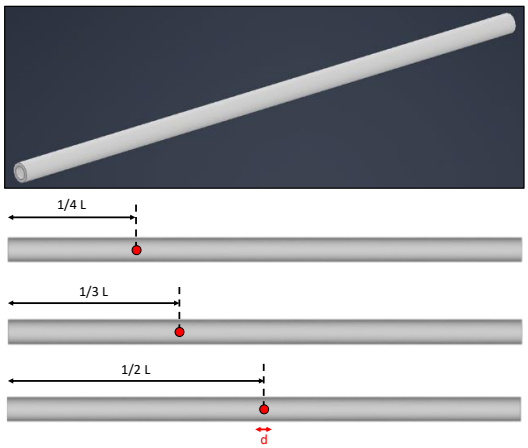


Figure 10. Digital model of hydrogen distribution pipeline. Source: own elaboration.

IV) Digital management and monitoring platform.

Through the digital management platform, it will be possible to visualize real-time data such as the sound acquired at all times and the mathematical processing performed. In addition, in case the system detects a leak event, the platform will be available to carry out the relevant risk management. (Figure 11)



Figure 11. Digital platform for managing hydrogen leaks. Source: own elaboration.

4 Conclusions

Once the project has been completed and a degree of technological maturity has been reached, a series of challenges should be overcome in order to reach the future version of the system. These would be the following:

I) Ensure access by seeking strategic partners for the future commercialization of the system, essentially the service provider and/or distributor of the system.

II) Optimization of the system installation process with the objective of reducing the time required for it since the installation requires the interruption of the normal operation process of the HRS.

III) Establish demonstrators with a double objective: 1) to demonstrate the operation and use of the system to potential customers; and 2) to demonstrate the applicability of the system in other relevant operational environments with different casuistry.

References

[1] Fuel Cells and Hydrogen 2 Joint Undertaking, “Hydrogen Roadmap Europe: A Sustainable Pathway for the European Energy Transition,” Brussels: Fuel Cells and Hydrogen 2 Joint Undertaking, 2019.

[2] Li, J., Li, Y., Huang, X., Ren, J., Feng, H., Zhang, Y., Yang, X., High-sensitivity gas leak detection sensor based on a compact microphone array *Measurement*, 174, 2021.

[3] Cao, J., Cao, M., Wang, J., Yin, C., Wang, D., Paul Vidal, P., Urban noise recognition with convolutional neural network, *Multimedia Tools and Applications*, 78, 2019.

[4] Boschert, S., Rosen, R., Digital Twin—The Simulation Aspect, in *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*, 59-74, 2016.

Probabilistic estimation of the stiffness of a structure from indirect measurements

Matos Castañeda, P. [♭], Navarro-Gonzalez, F.J[♭] [1] and Villacampa, Y.[♭]

(♭) Department of Applied Mathematics, University of Alicante
Carretera San Vicente del Raspeig s/n, Alicante, Spain.

1 Introduction

1.1 Importance of mechanical stiffness in structural analysis

Most civil works today are designed using reinforced concrete due to its accessibility and properties that have been extensively studied for decades. These properties allow structures built with this material to be able to withstand large loads, mainly because of its stiffness, which is determined by the intrinsic characteristics of the material, such as the modulus of elasticity, as well as its geometry.

Stiffness is an essential parameter in structural design, as it directly influences the stability and performance of concrete constructions. However, being a stone material, concrete is subject to various external conditions, such as environmental factors, age of the material and time under load, among others. These conditions can lead to creep and shrinkage phenomena, which significantly reduce the original designed stiffness, including increases in deflections of elements such as beams and plates.

Both destructive and non-destructive techniques are used to determine the stiffness of reinforced concrete structures. Destructive techniques usually involve direct testing of specimens taken from the structure, which can lead to permanent damage. On the other hand, non-destructive techniques allow stiffness-related parameters such as displacements or deformations and natural frequencies to be obtained indirectly without compromising the integrity of the structure. From the analysis of these parameters, it is possible to estimate the stiffness value using mathematical models or calibrated experimental procedures.

In particular, the relationship between a known applied force and the corresponding displacement generated in the structure, which can be measured by specific instruments, can be studied from the equations of dynamics, and theoretical values for stiffness can be obtained from these data. Among the most common methods are the finite element method, the simple bending method and the direct stiffness method [1, 2, 3, 4]. Another approach is through techniques to determine the stiffness of concrete from its response to oscillations, such as mechanical impedance testing. These involve applying an impact force to the structure, which generates vibrations. The response is measured through the propagation velocity of the waves. At low frequencies, this velocity allows to determine the stiffness of the structure.

These proposed methods are not free from inaccuracies and errors that deviate the value of the stiffness that the structure may have in real time from the theoretical assumptions obtained

[1] francisco.navarro@ua.es

through these methods. Recognising these errors and uncertainties allows control over the degree of deviation committed, with the aim of reducing it as far as possible. The Structural Code compiles mathematical expressions that allow this phenomenon to be studied through the analysis of the variables involved in them.

For very slender structures, modal analysis using ambient vibrations can be used [5], which is based on recording the structure's own frequencies by cleaning the signals around the building. This procedure makes it possible to identify the dominant signals specific to the building, from which parameters such as the global stiffness are calculated.

Structural stiffness is influenced by various physical factors, such as ambient humidity, the dimensions of the structure, the initial strengths of the material, the time considered in the analysis and the duration of the sustained loads. These factors cause a decrease in the actual stiffness value compared to the initial value predicted in the design, thus introducing a degree of uncertainty about its actual behaviour.

The Structural Code addresses these effects through the creep coefficient, which measures the temporal increase in the ratio between the deformation of concrete under constant loads maintained over time [6]. However, this coefficient may also be subject to uncertainties, which can be analysed by probabilistic studies [7]. The behaviour of the creep coefficient is not always linear, especially in structures subjected to particular conditions, and it is necessary to adjust its expression by means of mathematical functions, in order to represent more accurately the evolution of the creep phenomenon over time [7].

The expression for the total deformation under constant load is given by:

$$\varepsilon_\sigma(t, t_0) = \varphi(t, t_0) \cdot \frac{\sigma_c(t_0)}{E_{ci,28}}, \tag{1}$$

where $\varepsilon_\sigma(t, t_0)$ represents the total deformation of the concrete at age t, with load applied at time t_0, $\sigma_c(t_0)$ is the constant stress applied to the concrete at time t_0, $E_{ci,28}$ is the modulus of elasticity of the concrete at 28 days and $\varphi(t, t_0)$ is the creep coefficient of the concrete.

This creep coefficient can be defined from the initial values of the system by means of the variable φ_0 (basic creep coefficient), which represents the total expected creep under standard conditions and deriving the study of the evolution from this value to a function $\beta_c(t, t_0)$ that describes the development of the creep with the passage of time since the loading, so that $\varphi(t, t_0) = \varphi_0 \cdot \beta_c(t, t_0)$, modifying the equation (1) as:

$$\varepsilon_\sigma(t, t_0) = \beta_c(t, t_0) \cdot \frac{\varphi_0 \cdot \sigma_c(t_0)}{E_{ci,28}} \tag{2}$$

Beyond the point estimate of the value of the stiffness of the structure, it is important to have information that serves to determine the degree of uncertainty associated with this result, since a high uncertainty in the result can pose a real danger to the dynamic behaviour of the system.

Our study addresses this problem by means of a probabilistic analysis based on the random variable transformation, in order to analyse the creep behaviour of the material and estimate the moment at which significant changes occur in the stiffness of the concrete. This estimation will allow timely interventions, avoiding potential damage and optimising the preventive maintenance of the structures.

1.2 Random variable transformation

The use of the random variable transformation (RVT) technique is typically used in the study of the probabilistic evolution of dependent variables defined from random differential equations [8, 9, 10, 11].

However, it is a technique applicable in any relationship between random variables, as can be seen in [12, 13, 14], or under the name of Probabilistic transformation method (PTM) in [15, 16, 17, 18, 19].

Let us consider two sets of random variables of dimension n, denoted by $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_n)$ related by a one-to-one transformation by a function $g : \mathbb{R}^n \to \mathbb{R}^n$ of class C^1, so that $Y = g(X)$. Then, the relation between the joint probability densities of the respective variables, $f_X(x)$ and $f_Y(y)$ is given by:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |J|, \tag{3}$$

being J the jacobian of the transformation $X = h(Y) = g^{-1}(Y)$:

$$J = det\left(\frac{\partial X^T}{\partial Y}\right) = det\begin{pmatrix} \frac{\partial X_1}{\partial Y_1} & \cdots & \frac{\partial X_n}{\partial Y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial X_1}{\partial Y_n} & \cdots & \frac{\partial X_n}{\partial Y_n} \end{pmatrix} \tag{4}$$

This approach allows to use the uncertainty around variables that can be measured directly to determine the probability distribution of other variables related by an equation and that cannot be estimated by direct means.

1.3 Proposed model

To exemplify the proposed methodology, a structure in the form of a thin rectangular plate is presented, as shown in Figure 1.

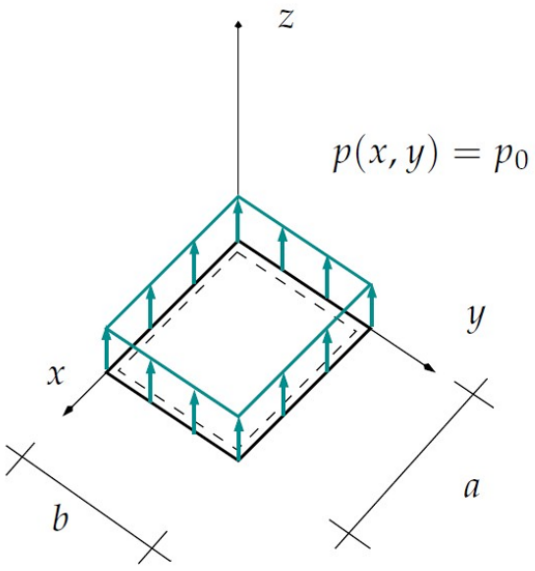


Figure 1: Thin plate model with load distribution.

The dynamics of the plate is modelled according to Kirchhoff's thin plate bending equation, which relates the deflection $w(x, y)$ and the applied load distribution $P(x, y)$ as:

$$\frac{\partial^4}{\partial x^4}w(x,y) + 2\frac{\partial^4}{\partial x^2 \partial y^2}w(x,y) + \frac{\partial^4}{\partial y^4}w(x,y) = \frac{P(x,y)}{D}, \tag{5}$$

with $D = \frac{Eh^3}{12(1-\nu^2)}$ the stiffness, where E is the modulus of elasticity of the material, ν the Poisson's coefficient and h the thickness of the plate [20].

The solution to the equation (5) will be obtained by means of the Navier method [21], in which the functions $w(x, y)$ and $P(x, y)$ are expanded in double Fourier series:

$$w(x,y) = \sum_{n=1,2,3}^{\infty} \sum_{m=1,2,3}^{\infty} w_{mn} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right) \tag{6}$$

$$P(x,y) = \sum_{n=1,2,3}^{\infty} \sum_{m=1,2,3}^{\infty} p_{nm} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right), \tag{7}$$

where the coefficients w_{mn}, p_{nm} are the factors corresponding to each of the modes and their superposition. n and m are integers representing the modes in the x and y directions, respectively.

From this function the deformations can be obtained:

$$e_x = -z \cdot \frac{\partial^2 w(x,y)}{\partial x^2} \tag{8}$$

$$e_y = -z \cdot \frac{\partial^2 w(x,y)}{\partial y^2} \tag{9}$$

The measurement of these deformations will provide us with the relationship between the deformations and the stiffness of the system, so that a study of the creep can be made from the definition given by (1) and (2).

As an example, in the case of a test point load of magnitude P applied at (ξ, η), the coefficients of the development of the function $P(x, y)$ defined in the equation (7) are given by:

$$p_{nm} = \frac{4P}{ab} \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right) \tag{10}$$

Applying the development of (6) to the differential equation (5), it is obtained that the solution for the function $w(x, y)$ is given by:

$$w(x,y) = \frac{4P}{abD\pi^4} \sum_{n,m}^{\infty} \frac{\sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right)}{\left(\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2\right)^2} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right). \tag{11}$$

This results in a deformation with respect to the direction X:

$$e_x = \frac{4zP}{abD\pi^2} \sum_{n,m}^{\infty} \frac{\sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right)}{\left(\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2\right)^2} \left(\frac{n}{a}\right)^2 \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right) \tag{12}$$

And a deformation with respect to the direction Y:

$$e_y = \frac{4zP}{abD\pi^2} \sum_{n,m3}^{\infty} \frac{\sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right)}{\left(\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2\right)^2} \left(\frac{m}{b}\right)^2 \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right) \tag{13}$$

In this particular case, we can consider a set of variables related to the dynamic behaviour of the structure as a function of its stiffness, such as the deformation under the effects of creep. In our example, we can consider the creep coefficient as a random variable, in order to study it and analyse how it behaves under conditions of uncertainty. The uncertainty in the measurements of these variables will be used to infer the probabilistic behaviour of the stiffness of the structure,

studying different configurations and comparing the accuracy of the proposed method against other alternatives.

A first case would be the total deformation $e = e_x + e_y$, considering the errors associated with the measurements of e_x and e_y. From the equations (12) and (13) there is the following relationship between the deformation e and the stiffness D:

$$e = \frac{4zP}{abD\pi^2} \sum_{n,m}^{\infty} \frac{\sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right)}{\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2} \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right) \tag{14}$$

Thus, the relationship between the variables e and D considered as random variables, according to the equation (3) will be:

$$f_D(D) = f_E(e) \cdot \left|\frac{de}{dD}\right| \tag{15}$$

And using the solution (14), that is:

$$f_D(D) = f_E(e) \cdot \frac{e}{D} \tag{16}$$

References

[1] Cruz Noguez, C. A., Murià Vila, D. y Camargo Pérez, J., Método para calcular las rigideces y los amortiguamientos dinámicos de grupo de pilotes de fricción *Revista de Ingeniería Sísmica*, 77:23–51, 2007.

[2] Ávila, J., Puertas, J., Martínez Castro, J. y Gallego, A., Aplicación del método de rigidez sistematizado en una estructura. *Dialnet*, 2021. Avaliable at: `https://dialnet.unirioja.es/descarga/articulo/9718928.pdf`

[3] Departamento de Estructuras (Universidad Politécnica de Cartagena (UPCT)), Análisis de Estructuras. Avaliable at: `https://upct.es/~deyc/publicaciones/AE_TGP.pdf`

[4] Centro de Estudios y Experimentación de Obras Públicas (CEDEX), Ensayos no destructivos para el control de la integridad de pilotes y pantallas in situ. Ministerio de Fomento, Gobierno de España, 2001.

[5] Michel, C., Guéguen, P. & Bard, P., Dynamic parameters of structures extracted from ambient vibration measurements: an aid for the seismic vulnerability assessment of existing buildings in moderate seismic hazard regions. *Soil Dynamics and Earthquake Engineering*, Preprint arXiv:0710.1210v1, 2006.

[6] Urtubeya, E., Lorefice, R., Fluencia básica del hormigón, factores que intervienen, estado del arte y modelos de predicción en el rango no lineal. *Mecánica Computacional*, XXV: 328, 2006.

[7] Ministerio de Transportes, Movilidad y Agenda Urbana (España), Código Estructural. Anejo 19: Proyecto de estructuras de hormigón. Reglas generales y reglas para la edificación. Gobierno de España, 2021. Avaliable at: `https://www.mitma.gob.es/recursos mfom/paginabasica/recursos/codigo estructural 2021.pdf`

[8] Soong, TT. Random differential equations in science and engineering. New York: Academic Press; 1973.

[9] Casabán, M. C., Cortés, J. C., Navarro-Quiles, A., Romero, J. V., Roselló, M. D., & Villanueva, R. J. A comprehensive probabilistic solution of random SIS-type epidemiological models using the random variable transformation technique. *Communications in Nonlinear Science and Numerical Simulation*, 32: 199-210, 2016.

[10] Casabán, M. C., Cortés, J. C., Romero, J. V., & Roselló, M. D. Probabilistic solution of random SI-type epidemiological models using the Random Variable Transformation technique. *Communications in Nonlinear Science and Numerical Simulation*, 24(1-3): 86-97, 2015.

[11] Casabán, M. C., Cortés, J. C., Romero, J. V., & Roselló, M. D. Determining the first probability density function of linear random initial value problems by the random variable transformation (RVT) technique: a comprehensive study. In Abstract and Applied Analysis (Vol. 2014, No. 1, p. 248512). Hindawi Publishing Corporation, 2014.

[12] Glen, A. G., Leemis, L. M., & Drew, J. H. A generalized univariate change-of-variable transformation technique. *INFORMS Journal on Computing*, 9(3): 288-295, 1997.

[13] Taher, M. A., Singh, M. J., Ismail, M. B., Samad, S. A., & Islam, M. T. Reducing the PAPR of OFDM systems by random variable transformation. *ETRI Journal*, 35(4): 714-717, 2013.

[14] El-Wakil, S. A., Sallah, M., & El-Hanbaly, A. M. Random variable transformation for generalized stochastic radiative transfer in finite participating slab media. *Physica A: Statistical Mechanics and its Applications*, 435: 66-79, 2015.

[15] Kadry, S. On the generalization of probabilistic transformation method. *Applied mathematics and computation,* 190(2): 1284-1289, 2007.

[16] Falsone, G., & Laudani, R. A probability transformation method (PTM) for the dynamic stochastic response of structures with non-Gaussian excitations. *Engineering Computations*, 35(5): 1978-1997, 2018.

[17] Laudani, R., & Falsone, G. An evolutive probability transformation method for the dynamic stochastic analysis of structures. *Probabilistic Engineering Mechanics,* 69: 103313, 2022.

[18] Ouhimmou, S., El Hami, A., Ellaia, R., & Tkiouat, M. Approximate solution of the structural problems using probabilistic transformation. *Key Engineering Materials*, 446: 91-99, 2010.

[19] Kassem, H., & Kadry, S. Stochastic fatigue of a mechanical system based on the technique of probabilistic transformation. E*uropean Journal of Scientific Research*, 29(2): 147-156, 2009.

[20] Gallego Sevilla, R., Martínez Castro, A., Puertas García, M. & others, Análisis de placas, 2022.

[21] Pérez Aparicio, J. L., Teoría de placas y láminas en Ingeniería Aeroespacial. Colección Manual de referencia. Editorial Universitat Politècnica de València, 2019.

System for detection and prediction of the phenomenon of hydrogen embrittlement

Juan Ramón Sánchez[1,*], Julia I. Real[1], Juan Alberto Puchol[1] and Elena Caidas[2]

(1) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, 46022, València (Spain).
(2) GRUPO VIARIUM, C/ Jacinto Benavente, 2B, 2º. Edificio TRIPARK, 28232 Las Rozas, Madrid (Spain)

1 Introduction

The European Union aims to use the Natural Gas infrastructure (gas pipelines) for the transport of Hydrogen (a mixture of natural gas and hydrogen). The problem arises in that the steel with which these pipelines are made is affected by the phenomenon of "hydrogen embrittlement" (HE). Hydrogen embrittlement (HE) is the loss of ductility of a metal and reduction of load-bearing capacity due to the absorption of hydrogen atoms or molecules by the metal. At the molecular level, hydrogen ions are introduced between the granular boundaries of the steel where they recombine into molecular hydrogen, occupying more space and weakening the bonds between grains. The result of HE is that the components crack and fracture at stresses below the yield strength of the metal. In the case of the transport of the NG and Hydrogen mixture, Hydrogen penetrates by diffusion into the steel due to the service conditions (pressure and flow) in which it is found. A detailed description of the hydrogen embrittlement process can be found in Figure 1. From the analysis of 372 hydrogen-related accidents in [2] from the ARIA database (France), 27% involved hydrogen-induced leaks or stresses in materials.

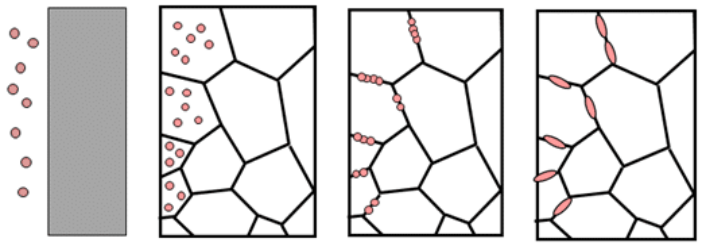


Figure 1. Scheme of the effect of Hydrogen brittleness. Source: [1]

In this work we propose as the optimal solution the technique of using the infrastructure without modifications by performing a preventive monitoring of HE. Specifically, the general objetive of this project is to develop a novel system for the management of the combined transport of Natural Gas and Hydrogen, capable of:

- Detect the degree of Hydrogen embrittlement of the steel pipelines used for Natural Gas transport in real time and continuously.
- Manage the Hydrogen transport in order to limit the Hydrogen concentration in the steel.

* juasanv6@upv.es

The **Hytari** project (**CPP2021-008683**) is part of the **2021** public-private collaboration projects, framed in the State Programme to Promote Scientific-Technical Research and its Transfer, included in the State Plan for Scientific, Technical and Innovation Research 2021-2023. The project consortium is formed by **Viarium** and the **Polytechnic University of Valencia** as a public research body.

2 Methods

The system is based on the electromechanical impedance (EMI) technique, which employs PZT sensors to monitor variations in the dynamic characteristics of the structure (mass, damping and/or stiffness). The EMI technique is used in all types of structural analysis such as corrosion detection in [3] or aircraft structural health in [4].

The system presented is based on the relationship between the elastic modulus (which decreases due to the effects of HE, this relationship is detailed in [5]) and the natural frequencies of vibration in the pipe, since these decrease as the elastic modulus decreases, as demonstrated in [6]. PZT sensors act simultaneously as an actuator (deforming when a strain is applied) and as a strain sensor (generating an electric potential when deforming). Thus, the EMI technique procedure is summarized in the measurement of the impedance in a resonance frequency.

$$f_{\mathrm{N}} = K \cdot \sqrt{E}$$

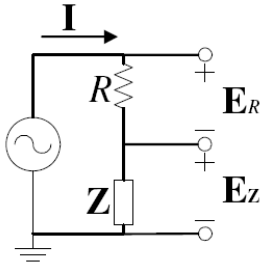
Where K depends of geometry

2.1 Electromechanical impedance

Electromechanical impedance is a structural monitoring method that uses piezoelectric transducers, relating the mechanical impedance of the evaluated structure with the electrical impedance measured in the transducers, so that the mechanical resonance spectrum of the structure is reflected in an almost identical spectrum of the real part of the electromechanical impedance measured in the piezoelectric transducers. Thus, the electromechanical coupling is exploited, associating the electrical response of the piezoelectric transducer with the mechanical impedance of the evaluated structure, allowing an indirect measurement of the changes in the mechanical impedance of the structure, circumventing the difficulties to perform a direct measurement of this variable, which provides information of the structural properties of the mechanical system. A more extensive description can be found in [7].

2.2 Frequency analysis of the EMI technique for the detection of fragility

The EMI technique consists of measuring the impedance of the piezoelectric material on the structure. The developed technique uses the measurement of two voltages (E_R and E_Z) to obtain the electromagnetic impedance (|Z| and Ø), according to the following expressions in Figure 2:



$$a_R = \frac{2}{N}\sum_{n=0}^{N-1} e_R(nT_s)\cdot cos\left(\frac{2\pi}{N}n\right)$$
$$b_R = \frac{2}{N}\sum_{n=0}^{N-1} e_R(nT_s)\cdot sin\left(\frac{2\pi}{N}n\right)$$
$$a_Z = \frac{2}{N}\sum_{n=0}^{N-1} e_Z(nT_s)\cdot cos\left(\frac{2\pi}{N}n\right)$$
$$b_Z = \frac{2}{N}\sum_{n=0}^{N-1} e_Z(nT_s)\cdot sin\left(\frac{2\pi}{N}n\right)$$

$$\rightarrow \quad |Z| = \frac{\sqrt{a_z^2 + b_z^2}}{\sqrt{a_R^2 + b_R^2}}\cdot R$$
$$\emptyset = \ tan^{-1}\left(\frac{a_z b_R - a_R b_Z}{a_z a_R + b_z b_R}\right)$$

Figure 2. Impedance measurement circuit. Source: [8]

2.3 Development of measurement system

Excitation driver

The piezoelectric sensor is excited by means of a sinusoidal signal of the frequency corresponding to the frequency at which impedance measurement is to be required. In the developed circuit in Figure 3, the sinusoidal form is achieved by means of a DAC output of the microcontroller and this is centered to a voltage $V_{DD}/2$ that acts as virtual ground, where V_{DD} = 5V.

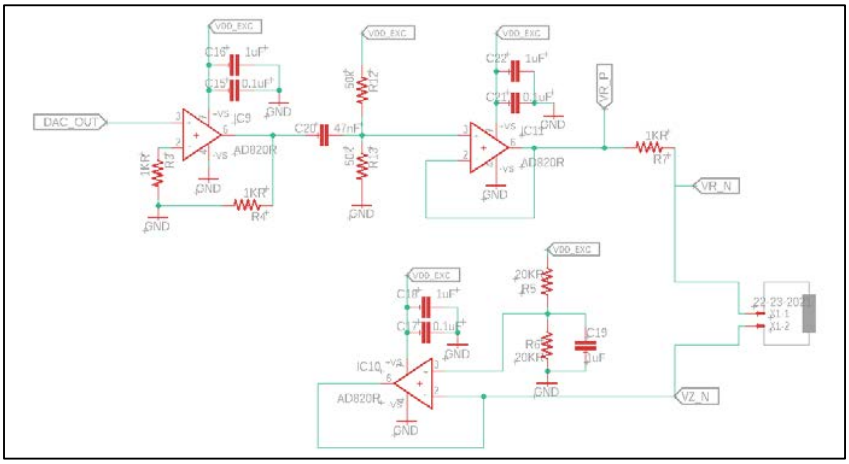


Figure 3. Piezoelectric sensor excitation circuit. Source: self-made

Measurement circuit

The circuit requires the measurement of the voltage drop across the series resistor E_R and the drop across the piezoelectric E_Z. For this purpose, a circuit in Figure 4 based on the ADC AD4001 has been implemented that allows an acquisition up to 2 MHz.

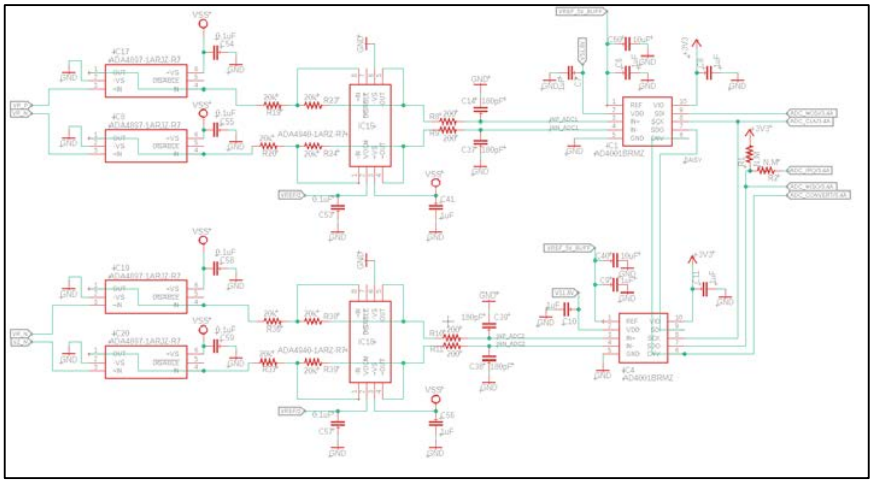


Figure 4. Acquisition circuit for ER and EZ. Source: self-made

System developed



Figure 5. PCB developed. Source: self made.

2.4 Algorithms

The previous sections have described how the phenomenon of brittleness directly affects the electromechanical properties of the material. However, there are other deteriorations and/or environmental conditions that also affect the physical, chemical and structural characteristics. These alterations are also reflected in the electromechanical measurements, which complicates the identification of hydrogen embrittlement. To confront this problem, the influence of other deteriorations and/or environmental conditions on the electromechanical properties has been analyzed.

Corrosion adversely affects metals, weakening their structure and altering their properties. In the context of the EMI technique, corrosion causes changes in the electrical properties of materials, which is reflected in impedance measurements. In the study [9], which investigates the feasibility of EMI technique to monitor corrosion in steel using a steel beam with three PZT sensors, which was subjected to accelerated corrosion using NaCl solution. Over 117 days, the electromechanical response was measured at various corrosion stages.

Cracks in metallic structures have a significant impact on the structure of metallic piping, as they alter the ability to transmit vibrations efficiently, which is reflected in variations in

structural impedance. In the study [10], where cracks in a metallic prosthesis are simulated by introducing contact elements. Cracks of different lengths (2 mm to 6 mm) are created representing various severities of damage are shown in Figure 6**.**

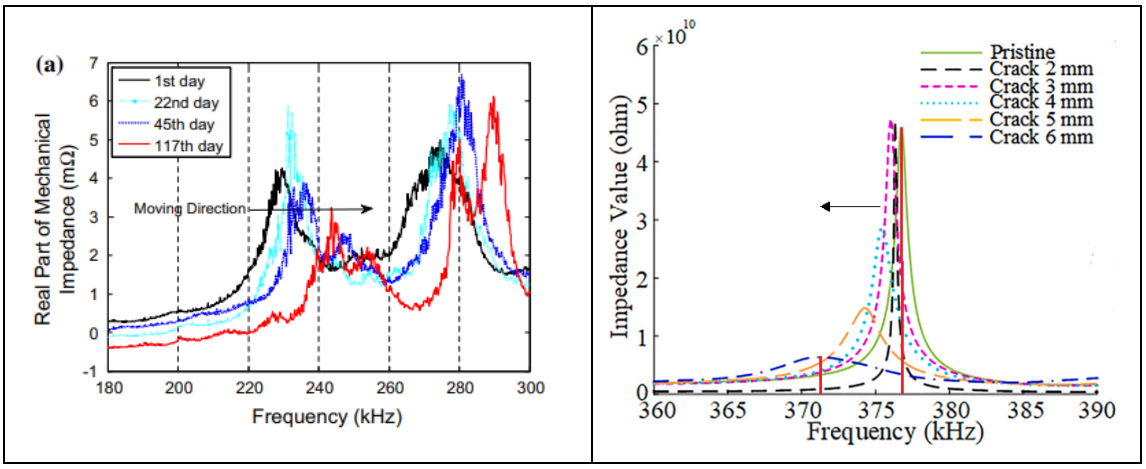


Figure 6. Impedance spectra for different corrosion evolution (1st-117st day) (left). Impedance spectra for cracks of different lengths (2-6 mm) (right) . Source [9] and [10].

Support condition is a critical factor that also significantly influences EMI response. Both support variability and progressive mismatch of threaded connections affect the dynamic properties of the pipe metal, varying the vibration patterns and load distribution. In the study [11], the behavior of the resonance frequency versus the mismatch of the support conditions is investigated by using four sensors based on PZT transducers placed on the connection bolts between two metal plates. These plates simulate a system with bolts that progressively loosen due to cyclic loading. The loosening of bolts P1 and P2 is shown in Table 1, where for each bolt the degree of torque changes from 90 N-m to 0 N-m at 30 N-m intervals, where 0 N-m represents the total loosening state.

Loading cases	Bolt No. 1	Bolt No. 2
L1	90	90
L2	60	90
L3	30	90
L4	0	90

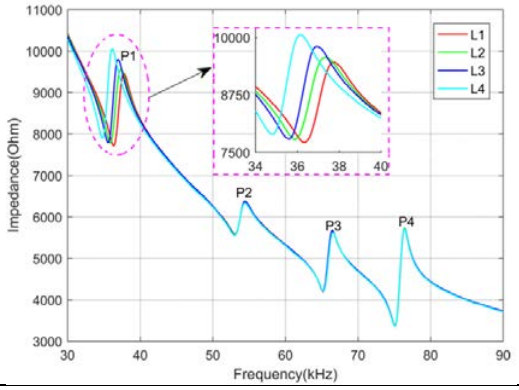


Figure 7. Impedance spectra for different support conditions . Source [11].

Ambient temperature variation is a critical problem for EMI-based applications, as small changes in temperature can generate significant changes in the electromechanical impedance response, both in the pipe structure and in the PZT sensor parameters. From the study [12] where damage is simulated by applying additional masses on the pipe surface and focuses on

the analysis of how thermal variation influences the impedance properties individually as is shown in Figure 8.

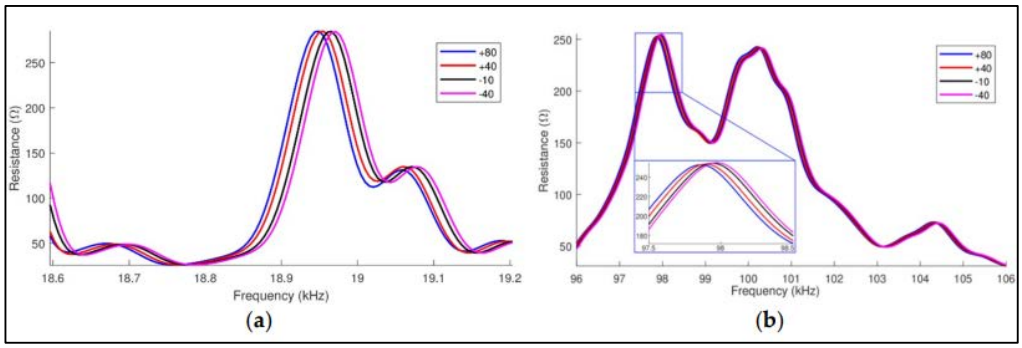


Figure 8. Impedance spectra for different ambient temperature conditions . Source [12].

The metric proposed to evaluate each type of impairment is mainly focused on the use of the correlation coefficient (CC), being this the variable on which the system is based to train the Neural Network (NR). According to studies [13], [14] and [15] using these parameters, their effectiveness has been demonstrated to characterize and detect different types of damage through the analysis of electromechanical impedance (EMI).

3 Results

A test plan has been designed in different phases to evaluate the capability of the system to measure hydrogen embrittlement.

- **Phase 0: evaluation of the system to measure electrical impedance.** For this purpose, several RLC circuit assemblies will be developed, and their impedance will be measured using the developed system. The results obtained will be compared with the theoretical curves and with the graphs obtained with a commercial impedance analyzer.
- **Phase 1: evaluation of the system for electromechanical impedance measurement.** For this, a piezoelectric sensor is installed on a metallic beam of known section (known E and I) and the boundary conditions (D) are modified to vary its resonance frequencies and the impedance spectra obtained will be compared with the theoretical resonance frequencies, this scheme is shown in Figure 9.

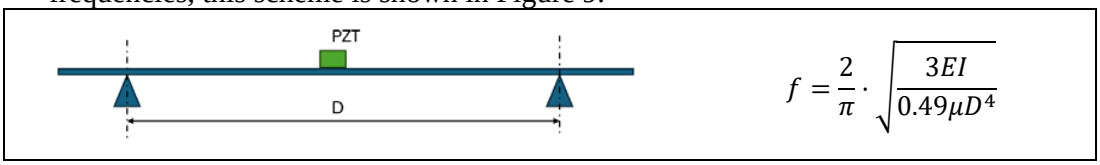


Figure 9. Phase 1. Electromechanical impedance measurement. Source: self-made.

- **Phase 2: evaluation of the system for measuring hydrogen embrittlement.** The aim of this test phase is to check the viability of the system for detecting hydrogen embrittlement and to obtain the frequency range required to detect HE, as well as to obtain the necessary inputs for the calibration of the identification algorithms and correlation with elastic modulus and limit of rupture.

Firstly, 60 samples will be obtained, corresponding to metallic pipes with a diameter of 25 mm and a length of 60 cm. The samples will be grouped in groups of 6 units, each group will be subjected to a hydrogen loading process to force their absorption from 0 ppm to 3 ppm. Five samples from each of the groups shall be subjected to a compression test (according to ISO 8492) to obtain the relationship between the elastic modulus (E) and the limit of rupture (R) with the hydrogen concentration (C_o^H), ($E = f(C_o^H)$ y $R = f(C_o^H)$). The remaining 10 samples will have their electromechanical impedance measured using the system developed to obtain the relationship between the electromechanical impedance (Z(f)) and the hydrogen concentration (C_o^H) ($Z(f) = f(C_o^H)$).

Hydrogen loading will be carried out at the AIDIMME Technological Institute by means of a cathodic galvanostatic loading in H_2SO_4 solution. To check the hydrogen concentration, a sample (obtained by a freezing and brittle breaking process) of one of the specimens of each group will be analysed using the Thermoscientific element analyser. The hydrogen loading and mechanical tests are performed in the same laboratory to reduce the possibility of loss of absorbed hydrogen.

- **Phase 3: training of the neural network for identification of hydrogen embrittlement, corrosion, support condition and cracks.** The case of a pipe without damage is initially analysed as a reference. Subsequently, the influence of each of the four types of damage on the pipe is assessed individually. Next, the combination of each type of damage with hydrogen embrittlement is studied. Finally, the simultaneous interaction of all the types of damage as a whole is considered, with HE embrittlement being the case of study in 5 of the 9 scenarios analysed in the pipes.
- **Phase 4: evaluation of the system in a real operating environment.** For this, the system developed will be installed on the *Instalación Plataforma de I+D para la inyección de hidrógeno en la red de GN* belonging to the *Fundación de hidrógeno de Aragón* during static operation. This installation recreates the injection of different flows of green hydrogen in a natural gas (carrier) of variable composition. This allows extensive experimental validation to be carried out in real environment conditions.

4 Conclusions

- The degree of hydrogen embrittlement affects the elastic modulus of the material. Even in advanced stages, it can lead to brittle fracture.
- Variations of the elastic modulus directly influence the resonance frequencies of the construction element.
- The most standardized method for measuring resonance frequencies is based on the measurement of electromagnetic impedance by means of a piezoelectric element.
- A system capable of measuring the electrical impedance of a piezoelectric sensor attached to the surface of the building element has been developed.
- The developed system has the necessary components, a controller, a long-distance communication system and a power supply system based on a solar system, to be installed along a natural gas transport network.

- Deteriorations and/or environmental conditions that also affect the physical, chemical and structural characteristics. These alterations are also reflected in the electromechanical measurements, which complicates the identification of hydrogen embrittlement.
- A neural network-based algorithm can identify hydrogen embrittlement from impedance spectrum analysis.
- The system allows a real time monitoring of a natural gas transport network.

References

[1] Industrial Metallurgists LLC, "Hydrogen Embrittlement of steel". https://www.imetllc.com/hydrogen-embrittlement-steel/

[2] BARPI, Hydrogen and transport: the risks should not be underestimated, 2020

[3] Zhu, H., Luo, H., Ai, D., and Wang, C. Mechanical impedance-based technique for steel structural corrosion damage detection, *Measurement*, 88: 353-359, 2016

[4] Vargas, L. Steffen, V. and Finzi, R. Monitoreo de salud estructural de aeronaves basado en mediciones de impedancia electromecánica, *Avances Investigación en Ingeniería*, 8: 18-27, 2011

[5] Ortiz, M. and Ovejero-Garcia, J. Effect of hydrogen on Young's modulus of AISI 1005 and 1070 steels. *Journal of Materals Science*, 27: 6777-6781, 1992

[6] Ahmadpar, M., Hoseinzadeh, S., Nakhaei, F. and Memon, S. Experimental Modal Analysis of Distinguishing Microstructural Variations in Carbon Steel SA516 by Applied Heat Treatments, Natural Frequencies, and Damping Coefficients, *Journal of Materials Engineering and Performance*, 30:9256-9261, 2021

[7] Marulanda, D. and Robledo, L. Identificación Y Localización De Daños En Estructuras Activas Con La Técnica De La Impedancia Electromecánica (EMI) Enfocada Al Monitoreo De La Integridad Estructural (Shm). *Trabajo de grado Universidad Tecnológica de Pereira*, 2022

[8] Wang, X. Impedance Measurement Method Based on DFT. *Life System Modeling and Intelligent Computing*. 499-504, 2010.

[9] Hong-ping, Z., Luo, H., Demi A. and Wang, C. Mechanical impedance-based technique for steel structura corrosion damage detection. *Measurement*, 88: 353-359, 2016.

[10] Lu R., Shen, Y., Zhang, B. and Xu, W. Nonlinear Electro-Mechanical Impedance Spectroscopy for fatigue crack monitoring. *Mechanical Systems and Signal Processing*, 184, 2023

[11] Luo, Z., Deng, H., Li L., and Luo, M. A simple PZT transducer design for electromechanical impedance (EMI)-based multi-sensing interrogation. *Journal of Civil Structural Health Monitoring*, 11: 234-249, 2020.

[12] Antunes, R., Cortez, N., Gianesini, B., Vieira, J. Modeling, Simulation, Experimentation, and Compensation of Temperature Effect in Impedance-Based SHM Systems Applied to Steel Pipes. *Sensors*, 19 2019

[13] Wang, D., Song, H. and Zhu, H. Numerical and experimental studies on damage detection of a concrete beam based on PZT admittances and correlation coefficient. *Construction and Building Materials*, 49: 564-574, 2013.

[14] Guofeng, D., Linsheng, H., Qingzhao, K. and Gangbing, S. Damage detection of pipeline multiple cracks using piezoceramic transducers. *Journal of Vibroengineering*, 18: 2828-2838, 2016.

[15] Rabelo, D., Finzi, R., and Steffen, V. Impedance-Based Structural Health Monitoring incorporating compensation of temperature variation effects, *23rd ABCM International Congress of Mechanical Engineering,* 2015

Part VI

Linear Algebra, Matrix Analysis and Applications

Exploring Matrix Exponentials via Charlier Polynomials

José M. Alonso[♮1], Emilio Defez[*], Javier Ibáñez[*], Jorge Sastre[†] and Álvaro Ibáñez[†]

(♮) Instituto de Instrumentación para Imagen Molecular,
(*) Instituto de Matemática Multidisciplinar,
(†) Instituto de Telecomunicaciones y Aplicaciones Multimedia
Universitat Politècnica de València. Camino de Vera s/n, 46022, Valencia, Spain.

Abstract

This paper addresses the application of Charlier polynomials to the matrix exponential function computation. The new numerical method developed, in its different varieties due to a parameter on which it depends, has been compared with another code based on Padé approximants. The results of the numerical test performed have shown that the new method is much more accurate than the one based on Padé, with quite similar response times.

1 Introduction

The matrix exponential is a fundamental function with wide-ranging applications across science and technology. Its significance also stems from the challenges involved in its computation [1, 2]. Numerous recent studies depend on the evaluation of the matrix exponential, such as [3, 4, 5, 6]. Consequently, various methods have been developed to compute it. For example, Padé approximants are explored in [7, 8], while Taylor polynomial approximations are discussed in [9, 10]. More recently, alternative approaches based on other polynomial families (such as Bernoulli and Euler) have been proposed in [11, 12, 13]. Computationally speaking, based polynomial methods have resulted to be more efficient than rational ones for approximating the matrix exponential.

Orthogonal matrix polynomials are a developing field of research that has already achieved significant results, both theoretically and practically. Regarding applications, in [11], Hermite matrix polynomials have been successfully used to compute the matrix exponential.

The monic Charlier polynomials $C_n^{(a)}(x)$ are defined in [14, pp.170–172] by the explicit expression

$$C_n^{(a)}(x) = \sum_{k=0}^{n} \binom{n}{k} \binom{x}{k} k!(-a)^{n-k} \ , \ a \neq 0, \tag{1}$$

or through the three-terms relation:

[1] jmalonso@dsic.upv.es

$$\left.\begin{aligned} C_{n+1}^{(a)}(x) &= (x-n-a)\,C_n^{(a)}(x) - a\,n\,C_{n-1}^{(a)}(x)\ ,\ n\geq 1 \\ C_0^{(a)}(x) &= 1\ ,\ C_1^{(a)}(x) = x-a \end{aligned}\right\}.$$

This sequence of polynomials $\left\{C_n^{(a)}(x)\right\}_{n\geq 0}$ satisfies the orthogonality relation

$$\int_0^{+\infty} C_m^{(a)}(x)C_n^{(a)}(x)d\psi^{(a)}(x) = a^n n!\delta_{nm},$$

where $d\psi^{(a)}(x)$ is the step function whose jumps are

$$d\psi^{(a)}(x) = \frac{\mathrm{e}^{-a}\,a^x}{x!}\ ,\ \text{at } x = 0,1,2,\ldots$$

or the *dual orthogonality* relation

$$\sum_{x=0}^{\infty} C_m^{(a)}(x)C_n^{(a)}(x)\frac{\mathrm{e}^{-a}\,a^x}{x!} = a^{-n}n!\delta_{nm},$$

(which gives an example of discrete orthogonal polynomials) and have the generating function

$$G(x,t) = \mathrm{e}^{-at}(1+t)^x = \sum_{n\geq 0}\frac{C_n^{(a)}(x)}{n!}t^n. \tag{2}$$

Others properties of Charlier polynomials can be found in the reference by T. Chihara [14, Chapter 6]. From (1), we can define the Charlier matrix polynomials taking the matrix parameter $A\in\mathbb{C}^{r\times r}$ as follows:

$$C_n^{(A)}(x) = \sum_{k=0}^{n}\binom{n}{k}\binom{x}{k}k!A^{n-k}, \tag{3}$$

and it is fulfilled that

$$\mathrm{e}^A = \frac{1}{2^x}\sum_{n\geq 0}\frac{C_n^{(A)}(x)}{n!}. \tag{4}$$

Using the partial sums of the series given in (4), the expression we are trying to test to approximate the exponential of a matrix A becomes:

$$\mathrm{e}^A \approx P_m(A) = \frac{1}{2^x}\sum_{n=0}^{m}\frac{C_n^{(A)}(x)}{n!}, \tag{5}$$

where x is a parameter to be accordingly determined based on the experimental tests we will conduct. Thus, $P_m(A)$ is a m-order polynomial matrix polynomial of the form:

$$P_m(A) = \sum_{k=0}^{m} c_k A^k. \tag{6}$$

From expression (5) and together with the scaling and squaring technique, a new numerical method to compute the matrix exponential has been designed and evaluated in this paper.

2 Numerical experiments

In this section, we will compare the performance of our implementations based on Charlier polynomials with respect to another code related to Padé rational approximations. The codes to be evaluated, all developed in MATLAB, are as follows:

- `expm_charlier_x`: these are 3 codes developed from expression (5), where x indicates the numerical value of this parameter. In our case, we will work with x values equal to 1, 3 and 5. The degree of the polynomial and the value of the scaling parameter were obtained from a theoretical formulation of an absolute forward type error, always with the aim of reducing the error committed in the calculation of the exponential. Matrix polynomials in the form of (6) were efficiently evaluated by means of the Paterson–Stockmeyer method. The allowed values of the degree of the polynomial to be used were between 20 and 56.

- `expm_new`: function that computes the matrix exponential by using a scaling and squaring algorithm with a Padé approximation [8, 15].

A heterogeneous test battery consisting of 3 types of square matrices was used to evaluate the numerical performance of the newly implemented method. All of them have a size that is a power of 2, between 4 and 1024, with 180 diagonalizable complex matrices (**Type 1**), 180 non-diagonalizable complex matrices (**Type 2**), and 359 matrices (277 from the Matrix Computation Toolbox [16] and 82 from the Eigtool MATLAB Package [17]) (**Type 3**).

The metric used to show the goodness of each method in computing the exponential of each matrix A with respect to the exact solution was the normwise relative error $Er(A)$, obtained as:

$$Er(A) = \frac{\| \exp(A) - \widetilde{\exp}(A) \|_2}{\| \exp(A) \|_2},$$

where $\exp(A)$ represents the approximate solution obtained by each of the codes used in the comparison and $\widetilde{\exp}(A)$ stands for the exact solution. All the numerical experiments were executed on a a Microsoft Windows 11 x64 PC equipped with a 12th Generation Intel Core i7-12700H @2.70 GHz processor and 32 GB of RAM, using MATLAB R2023b.

Table 1 presents the percentage of matrices, compared to the total of each group, for which the relative error of `expm_charlier_1` was better, worse or equal to that of the other methods. As it can be seen, there were a remarkable number of matrices in which the errors committed by `expm_charlier_1` and `expm_charlier_3` were identical, specially for matrices in Type 3. Notwithstanding, the percentage of cases in which `expm_charlier_1` incurred in a smaller error than `expm_charlier_3` was always higher than vice versa. These differences in favor of `expm_charlier_1` were more pronounced for `expm_charlier_5`, and especially for `expm_new`, where the improvement percentage exceeded 93%.

On the other hand, Figures 1, 2 and 3 show graphically the results for each kind of matrices, respectively. First, Figures 1a, 2a, and 3a depict the normwise relative error introduced by the four codes for each of the matrices used. Theoretically, the continuous black line drawn indicates the relative error expected to be committed in the computation of each matrix. This means that values below this line are desirable, thus reflecting the numerical stability of the method. In general terms, the relative errors corresponding to `expm_charlier_1` always occupied positions below the line mentioned above, being the most stable method. On the other hand, for certain matrices, the other methods take locations above the highlighted line, resulting in higher-than-expected errors.

For all the methods, the result of ordering from highest to lowest their previous mentioned relative error is drawn in Figures 1b, 2b, and 3b. It should be noted that these graphs do not allow appreciating the error generated by all the methods in the computation of a same matrix. It is

Table 1: Percentage of cases in which the error of `exp_charlier_1` was less than, greater than or equal to that of the other codes, for the 3 types of matrices.

	Type 1	Type 2	Type 3
Er(`expm_charlier_1`)<Er(`expm_charlier_3`)	33.33%	42.22%	15.32%
Er(`expm_charlier_1`)>Er(`expm_charlier_3`)	19.44%	31.67%	11.42%
Er(`expm_charlier_1`)=Er(`expm_charlier_3`)	47.22%	26.11%	73.26%
Er(`expm_charlier_1`)<Er(`expm_charlier_5`)	58.89%	65.56%	32.59%
Er(`expm_charlier_1`)>Er(`expm_charlier_5`)	38.33%	33.89%	30.64%
Er(`expm_charlier_1`)=Er(`expm_charlier_5`)	2.78%	0.56%	36.77%
Er(`expm_charlier_1`)<Er(`expm_new`)	95.56%	93.89%	93.04%
Er(`expm_charlier_1`)>Er(`expm_new`)	3.89%	6.11%	6.96%
Er(`expm_charlier_1`)=Er(`expm_new`)	0.56%	0%	0%

Table 2: Execution times (in seconds) for `expm_charlier_x`, with x=1, 2, and 3, and `expm_new`, for the 3 distinct matrix sets.

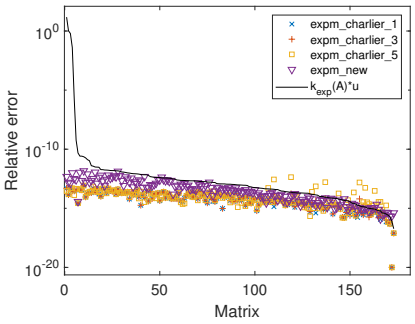
	Type 1	Type 2	Type 3
T(`expm_charlier_x`)	21.24	19.16	1.73
T(`expm_new`)	20.45	18.84	2.49

clear that `expm_charlier_1` always appeared in the bottom positions, while `expm_new` is usually positioned at the top.

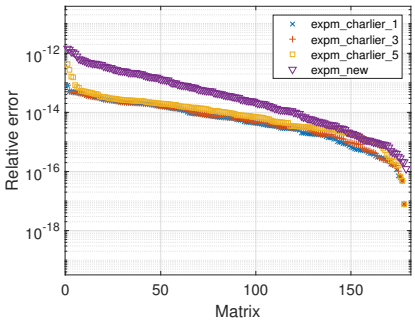
Then, Figures 1c, 2c, and 3c provide the performance profile. This is certainly one of the best indicators, for quantitative and qualitative comparison, of the numerical performance of each method. Each point on the plots represents the percentage of matrices, expressed as a percentage of one, for which the error assumed by each method is less than or equal to the smallest one incurred by all of them multiplied by α. The desired values to be reached are therefore those located in the highest parts of the graphics, which are clearly occupied from the beginning to the end by `expm_charlier_1`. The results corresponding to `expm_charlier_3` and `expm_charlier_5` can be found below `expm_charlier_1`, where `expm_new` once again occupied the most modest positions.

Next, Figures 1d, 2d, and 3d depict the percentage of matrices for which each method achieved the best or worst relative error when computing their exponential, in the form of a pie chart. As we can see, for the 3 types of matrices that constitute our testbed, `expm_charlier_1` was the most accurate method in 36%, 36%, and 33% of cases, respectively. In contrast, `expm_new` was consistently the least reliable code, since it provided the results that were furthest away from the exact ones in 80%, 74%, and 86% of matrices. Alternatively, `expm_charlier_1` only resulted in the worst scores in 1%, < 1%, and 4% of cases. Intermediate percentages between `expm_charlier_1` and `expm_new` were always provided by `expm_charlier_3` and `expm_charlier_5`.

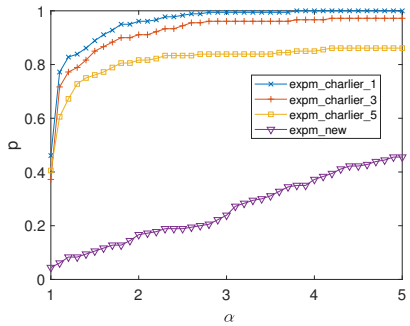
Finally, Table 2 outlines the response times for computing the exponential of matrices belonging to each of the three types used in the experiments. The three possible alternatives corresponding to the Charlier polynomials had identical response times, which is why they have been grouped in the same row. As it can be appreciated, `expm_new` offered slightly lower times for matrices from Types 1 and 2, but lightly higher ones for matrices belonging to Type 3.



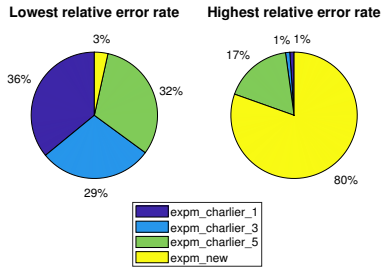
(a) Normwise relative error.



(b) Ordered normwise relative error.

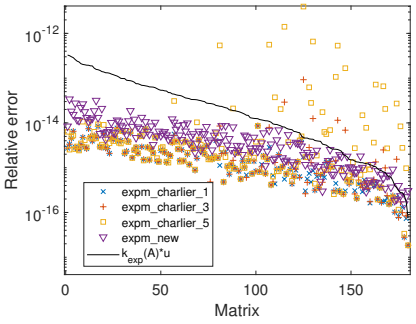


(c) Performance profile.

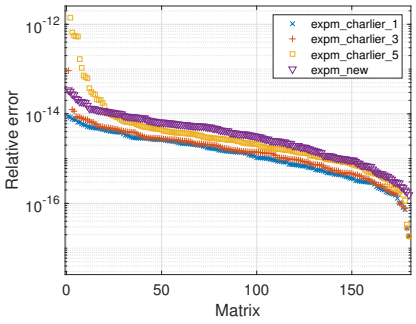


(d) Lowest and highest relative error rate.

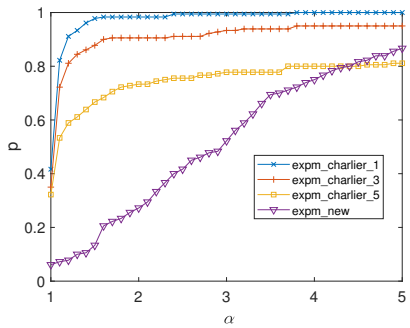
Figure 1: Results for Type 1.



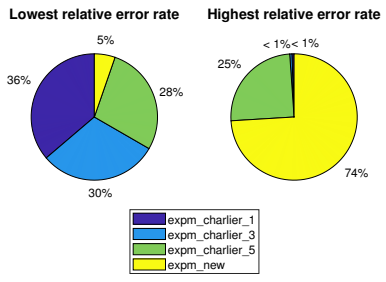
(a) Normwise relative error.



(b) Ordered normwise relative error.

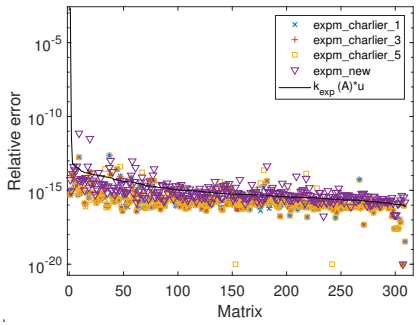


(c) Performance profile.

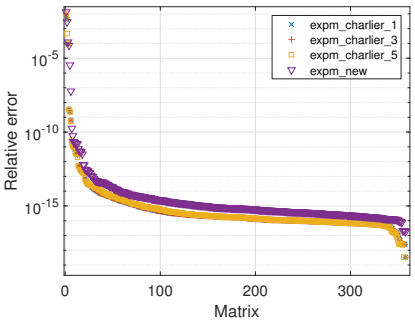


(d) Lowest and highest relative error rate.

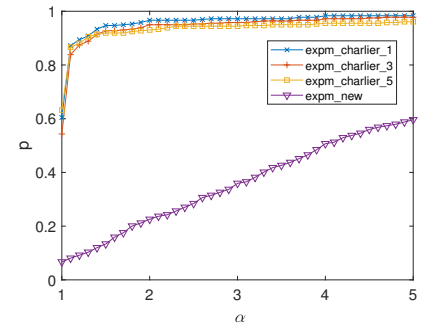
Figure 2: Results for Type 2.



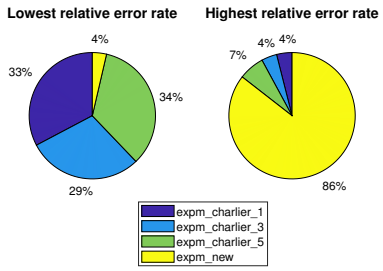
(a) Normwise relative error.



(b) Ordered normwise relative error.



(c) Performance profile.



(d) Lowest and highest relative error rate.

Figure 3: Results for Type 3.

3 Conclusions

The exponential is undoubtedly the matrix function whose computation is most frequently required in many areas of science and technology. This article describes a new method for computing matrix exponential using Charlier polynomials in conjuntion with the scaling and squaring technique. The resulting matrix polynomial, which approximates the exponential of each matrix, was efficiently evaluated by means of the Paterson–Stockmeyer method.

Three codes based on Charlier polynomials have been developed, with the variable x in their definition taking on values of 1, 3 and 5. These codes were compared with a state-of-the-art implementation, which consists of Padé approximants and constitutes the core of the MATLAB built-in function devoted to compute the matrix exponential.

Numerical experiments have revealed that the most accurate results have been obtained for Charlier polynomials and when x equals 1. Slightly worse results were obtained for x=3 and 5, although either of these alternatives was always better than the Padé-based implementation. Execution times were very similar among the 4 codes, with only minor differences between the Charlier and Padé developments.

Acknowledgments

This work has been supported by the Generalitat Valenciana Grant CIAICO/2023/275 and the Vicerrectorado de Investigación de la Universitat Politècnica de València (PAID-11-23).

References

[1] Moler, C., and Van Loan, C., Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.

[2] Higham, N.J., Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[3] Jarlebring, E., Fasi, M., and Ringh, E., Computational graphs for matrix functions. *ACM Transac-tions on Mathematical Software*, 48(4):1–35, 2023.

[4] Caglio, L., Stang, H., Brincker, R., and Katsanos, E., FE-aided Kalman filter for nonlinear state estimation with unknown input. *Mechanical Systems and Signal Processing*, 200:110513, 2023.

[5] Caliari, M., and Cassini, F., Direction splitting of φ-functions in exponential integrators for d-dimensional problems in Kronecker form. arXiv preprint, arXiv:2304.02327, 2023.

[6] Albi, G., Caliari, M., Calzola, E., and Cassini, F., Exponential integrators for mean-field selective optimal control problems. arXiv preprint, arXiv:2302.00127, 2023.

[7] Arioli, M., Codenotti, B., and Fassino, C., The Padé method for computing the matrix exponential. *Linear algebra and its applications*, 240:111–130, 1996.

[8] Higham, N.J., The scaling and squaring method for the matrix exponential revisited. *SIAM review*, 51(4):747–764, 2009.

[9] Sastre, J. Ibáñez, J., and Defez, E., Boosting the computation of the matrix exponential, *Applied Mathematics and Computation*, 340:206–220, 2019.

[10] Bader, P., Blanes, S., and Casas, F., Computing the matrix exponential with an optimized Taylor polynomial approximation. *Mathematics*, 7(12):1174, 2019.

[11] Sastre, J. Ibáñez, J., Defez, E., and Ruiz, P., Efficient orthogonal matrix polynomial based method for computing matrix exponential. *Applied Mathematics and Computation*, 217(14):6451–6463, 2011.

[12] Defez, E., Ibáñez, J., Alonso-Jordá, P., Alonso, J. M., and Peinado, J., On Bernoulli matrix polynomials and matrix exponential approximation. *Journal of Computational and Applied Mathematics*, 404:113207, 2022.

[13] Alonso, J.M., Ibáñez, J., Defez, E., and Alonso-Jordá, P., Euler polynomials for the matrix exponential approximation. *Journal of Computational and Applied Mathematics*, 425: 115074, 2023.

[14] Chihara, T.S., An introduction to orthogonal polynomials. Courier Corporation, 2011.

[15] Al-Mohy, A.H., and Higham, N.J., A new scaling and squaring algorithm for the matrix exponential. *SIAM Journal on Matrix Analysis and Applications*, 31(3):970–989, 2009.

[16] Higham, N.J., The Matrix Computation Toolbox. URL: http://www.ma.man.ac.uk/ higham/mctoolbox, 2002.

[17] Wright, T.G., Eigtool, version 2.1. URL: http://www.comlab.ox.ac.uk/pseudospectra/eigtool, 2009.

Characterization of the Combined Matrix of a Tridiagonal Toeplitz Matrix

B. Cantó,[1] R. Cantó and A.M. Urbano

I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Abstract

Combined matrices associated with tridiagonal Toeplitz matrices are studied. Concretely, we examine the structural properties of their combined matrix, with an emphasis on its symmetry, stochastic nature, and tridiagonal Jacobi form.

2 Introduction

In this work, we focus on real tridiagonal Toeplitz matrices, which are a specific type of tridiagonal Jacobi matrix. A tridiagonal Jacobi matrix is a structured square matrix in which nonzero elements are restricted to the main diagonal and the two adjacent diagonals directly above and below it, while all other entries are zero.

Tridiagonal Toeplitz matrices are a further refine the structure of Jacobi matrices, by ensuring that the values along each of the these three diagonals remain constant. Specifically, every entry on the main diagonal takes the same fixed value, while the entries on the first superdiagonal and subdiagonal also remain unchanged throughout the matrix. This structured repetition makes tridiagonal Toeplitz matrices particularly useful in various mathematical and engineering applications, as their properties allow for efficient computations and analysis [1], simplifying many matrix operations, such as matrix multiplication and inversion, making them highly efficient for computational purposes [2]. Their spectral properties and inverse computations play an important role in numerical treatments of boundary value problems for second-order linear difference equations [3]. Given their computational advantages, they are essential in solving linear systems with fast and stable numerical methods.

These matrices are particularly advantageous in applications involving large-scale systems where computational efficiency is paramount. For example, in signal processing and control theory, tridiagonal Toeplitz matrices are used to model and solve problems related to filtering and system stability. Their predictable structure allows for the development of specialized algorithms that can exploit their properties to achieve faster and more accurate results [4].

Our main objective is to characterize the combined matrix of a tridiagonal Toeplitz matrix. The combined matrix of a nonsingular matrix $A = (a_{ij})$ is defined as $U = \mathcal{C}(A) = A \circ A^{-T}$ where $\circ$ denotes the Hadamard (entrywise) product, and A^{-T} represents the inverse transpose, $(A^{-1})^T$, of A. In control theory, the combined matrix, also known as the Relative Gain Array (RGA) [5], is a

[1] bcanto@mat.upv.es

key tool for analyzing multivariable process control systems. It provides insight into optimal input-output pairings by quantifying the interaction between different control loops. This analysis is crucial for designing control systems that are both effective and efficient, as it helps identify which input-output pairs should be coupled or decoupled to achieve the best performance. The RGA evaluates fundamental closed-loop system properties such as stability, robustness, and sensitivity, which are essential for ensuring that the system can maintain desired performance levels despite disturbances or uncertainties.

From a mathematical perspective, a fundamental property of combined matrices is that the sum of the entries in each row and column is always equal to one. The combined matrix is classified as a doubly stochastic matrix when all entries are nonnegative. More generally, for real-valued matrices, combined matrices are classified as doubly quasi-stochastic [6]. This classification allows for the inclusion of negative values, which can represent inhibitory interactions in control systems. Doubly quasi-stochastic matrices retain the row and column sum properties but extend the applicability of the RGA to a broader range of systems, including those with both positive and negative interactions.

This paper studies the combined matrix associated with a tridiagonal Toeplitz matrix. We establish its characterization as a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix and we demonstrate that constructing the combined matrix requires computing only a subset of its entries.

3 Results

In this study we focus on the combined matrices of tridiagonal Toeplitz matrices. We recall the following concepts.

Definition 1 *A tridiagonal Toeplitz matrix of order n, denoted as $A_n = T(n; c, a, b)$, is given by*

$$A_n = T(n; c, a, b) = \begin{pmatrix} a & b & 0 & \cdots & 0 & 0 \\ c & a & b & \cdots & 0 & 0 \\ 0 & c & a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a & b \\ 0 & 0 & 0 & \cdots & c & a \end{pmatrix}.$$

Lemma 1 *[7] The combined matrix $\mathcal{C}(A) = (c_{ij}) = A \circ A^{-T}$ of a nonsingular matrix $A = (a_{ij})$ satisfies*

(a) If D_1 and D_2 are two diagonal nonsingular matrices then $\mathcal{C}(A) = \mathcal{C}(D_1 A D_2)$.

(b) If P and Q are two permutation matrices, then $\mathcal{C}(PAQ) = P\mathcal{C}(A)Q$.

(c) $\mathcal{C}(A^{-T}) = \mathcal{C}(A)$.

(d) If $c_{ij} \neq 0$ then $a_{ij} \neq 0$, $\forall\ i, j$.

Definition 2 *[8] A real matrix $U = (u_{ij})$ of order n is called a doubly quasi-stochastic matrix if $\sum_{i=1}^{n} u_{ij} = 1$, $\forall j = 1, 2, \ldots, n$ and $\sum_{j=1}^{n} u_{ij} = 1$, $\forall i = 1, 2, \ldots, n$. If $u_{ij} \geq 0$, it is called a doubly stochastic matrix.*

Definition 3 *A real matrix $U = (u_{ij})$ of order n is said to be a bisymmetric matrix if it is symmetric with respect to both its main diagonal and its secondary diagonal.*

In this work, we want to obtain the structure of the combined matrix $U = \mathcal{C}(A_n)$, with $A_n = T(n; c, a, b)$. For that, we assume that b or c are non-zero, since otherwise $\mathcal{C}(A) = I$, and we factorize A_n using nonsingular diagonal matrices to obtain a diagonally equivalent matrix $\tilde{A}_n$.

Proposition 1 *Let $A_n = T(n; c, a, b)$. Then A_n is diagonally equivalent to $\tilde{A}_n$ as follows:*

a) If $a \neq 0$, defining $D_1 = diag\left(\frac{c^{n-1}}{a^n}, \frac{c^{n-2}}{a^{n-1}}, \cdots, \frac{c}{a^2}, \frac{1}{a}\right)$ and $D_2 = diag\left(\frac{a^{n-1}}{c^{n-1}}, \frac{a^{n-2}}{c^{n-2}}, \cdots, \frac{a}{c}, 1\right)$, we obtain $\tilde{A}_n = D_1 A_n D_2 = T(n; 1, 1, d)$ with $d = bc/a^2$.

b) If $a = 0$, defining $D = \frac{1}{c} I_n$, $c \neq 0$, we have $\tilde{A}_n = DA_n = T(n; 1, 0, g)$ with $g = b/c$.

Using the relationship between A_n and $\tilde{A}_n$, we give the following result about its determinant.

Proposition 2 *Let $A_n = T(n; c, a, b)$.*

a) If $a \neq 0$, obviously $\det(A_1) = a$, and

$$\det(A_n) = a^n \sum_{j=0}^{q} (-1)^j \begin{pmatrix} n-j \\ j \end{pmatrix} d^j,$$

where $q = \lfloor n/2 \rfloor$, for $n = 2, 3, \ldots$

b) If $a = 0$, $\det(A_1) = 0$ and $\det(A_n) = c^n(-g)^p$, for $n = 2p$, and $\det(A_n) = 0$, for $n = 2p + 1$, $p = 1, 2, 3 \ldots$

Proof *a)* We have $\det(\tilde{A}_n) = \det(D_1)\det(A_n)\det(D_2) = \frac{1}{a^n}\det(A_n)$. Then, $\det(A_n) = a^n \det(\tilde{A}_n)$. To find $\det(\tilde{A}_n)$, we use the recursive relation derived from Laplace's expansion

$$\det(\tilde{A}_n) = \det(\tilde{A}_{n-1}) - d\,\det(\tilde{A}_{n-2}),$$

with initial conditions $\det(\tilde{A}_0) = 1$ and $\det(\tilde{A}_1) = 1$. We can solve this recurrence relation by noting that $\det(\tilde{A}_n)$ is a polynomial in d of degree $\lfloor n/2 \rfloor$. The coefficients of this polynomial are combinatorial numbers derived from the Tartaglia's triangle. Thus, we have

$$\det(\tilde{A}_n) = \sum_{j=0}^{q} (-1)^j \begin{pmatrix} n-j \\ j \end{pmatrix} d^j,$$

where $q = \lfloor n/2 \rfloor$, for $n = 2, 3 \ldots$

b) The matrix A_n becomes a tridiagonal matrix with zero diagonal elements and, in this case, $\det(\tilde{A}_n) = \det(D)\det(A_n) = \frac{1}{c^n}\det(A_n)$. Then, $\det(A_n) = c^n \det(\tilde{A}_n)$. To find $\det(\tilde{A}_n)$, we use the same recurrence relation

$$\det(\tilde{A}_n) = -g\,\det(\tilde{A}_{n-2})$$

with initial conditions $\det(\tilde{A}_0) = 1$ and $\det(\tilde{A}_1) = 0$.

$$\begin{array}{lll} \text{For } n = 2p & \rightarrow & \det(\tilde{A}_{2p}) = (-g)^{p-1}\det(\tilde{A}_2) = (-g)^p. \\ \text{For } n = 2p+1 & \rightarrow & \det(\tilde{A}_{2p+1}) = (-g)^p \det(\tilde{A}_1) = 0. \end{array}$$

Thus, we have $\det(A_n) = c^n(-g)^p$, for $n = 2p$, and $\det(A_n) = 0$, for $n = 2p+1$, $p = 1, 2, 3, \ldots$. $\square$

Since A_n is diagonally equivalent to $\tilde{A}_n$, Lemma 1 implies $U = \mathcal{C}(A_n) = \mathcal{C}(\tilde{A}_n)$. For simplicity, we use $\tilde{A}_n$ to obtain results concerning the entries of U.

Theorem 1 *Let $\tilde{A}_n = T(n; 1, 1, d)$. Then,*

a) To construct U it is only necessary to obtain the entries $u_{i,i}$ and $u_{i,i+1}$, $i = 1, 2, \ldots, \lfloor n/2 \rfloor$.

b) $U = \mathcal{C}(\tilde{A}_n)$ is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix.

c) $\mathcal{C}(T(n; 1, 1, d)) = \mathcal{C}(T(n; d, 1, 1))$.

Proof *a)* Since $\tilde{A}_n = T(n; 1, 1, d)$, by Lemma 1 $U = (u_{ij}) = \mathcal{C}(\tilde{A}_n)$ is a doubly quasi-stochastic tridiagonal matrix, being

$$U = \mathcal{C}(\tilde{A}_n) = \tilde{A}_n \circ \tilde{A}_n^{-T}.$$

Let $\tilde{A}_n^{-1} = S = (s_{ij})$. The entries are given by

$$s_{i,i} = \frac{1}{\det(\tilde{A}_n)} C_{i,i}(\tilde{A}_n), \quad i = 1, 2, \ldots, n,$$

where $C_{i,i}(\tilde{A}_n) = \det(\tilde{A}_{i-1}) \det(\tilde{A}_{n-i})$ is the cofactor of the entry (i, i) of $\tilde{A}_n$.

Since $u_{ii} = s_{ii}$, for $i = 1, 2, \ldots, n$ and U is a doubly quasi-stochastic tridiagonal Jacobi matrix, we only need to calculate the entries $u_{i,i}$ and $u_{i,i+1}$, for $i = 1, 2, \ldots, \lfloor n/2 \rfloor$.

Finally, for $n = 2q+1$, $q = 1, 2, 3, \ldots$, $u_{q+1,q+1}$ can be directly obtained, considering that U is a doubly quasi-stochastic tridiagonal matrix.

b) From the structure of matrix $\tilde{A}_n$ we have that $C_{i,i}(\tilde{A}_n) = C_{n+1-i,n+1-i}(\tilde{A}_n)$ for $i = 1, 2, \ldots, \lfloor n/2 \rfloor$. By using a) we conclude that $U = \mathcal{C}(\tilde{A}_n)$ is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix.

c) Since $U = \mathcal{C}(\tilde{A}_n)$ is a symmetric matrix, we have

$$\begin{aligned}
\mathcal{C}(T(n; 1, 1, d)) &= T(n; 1, 1, d) \circ (T(n; 1, 1, d))^{-t} \\
(\mathcal{C}(T(n; 1, 1, d)))^t &= (T(n; 1, 1, d))^t \circ \left((T(n; 1, 1, d))^{-t}\right)^t \\
\mathcal{C}(T(n; 1, 1, d)) &= T(n; d, 1, 1) \circ (T(n; d, 1, 1))^{-t} \\
\mathcal{C}(T(n; 1, 1, d)) &= \mathcal{C}(T(n; d, 1, 1)).
\end{aligned}$$

$\square$

We clarify Theorem 1 with the following example.

Example 1 *Consider the matrix $A_5 = T(5; c, a, b)$ with $d = bc/a^2 = 2$. Then, $\tilde{A}_5 = T(5; 1, 1, 2)$. To obtain $U = \mathcal{C}(\tilde{A}_5)$ we calculate*

$$\begin{aligned}
u_{1,1} &= u_{5,5} = \frac{\det(\tilde{A}_4)}{\det(\tilde{A}_5)} = -\frac{1}{5} \\
u_{2,2} &= u_{4,4} = \frac{\det(\tilde{A}_1) \det(\tilde{A}_3)}{\det(\tilde{A}_5)} = -\frac{3}{5}.
\end{aligned}$$

Now we find the rest of the entries of U,

$$u_{1,2} = u_{2,1} = 1 - u_{1,1} = \frac{6}{5} \quad \textit{and} \quad u_{2,3} = u_{3,2} = 1 - u_{2,1} - u_{2,2} = \frac{2}{5}.$$

$u_{3,3}$ is calculated as $u_{3,3} = 1 - u_{3,2} - u_{3,4} = \dfrac{1}{5}$.

Thus, the matrix U is given by

$$U = \frac{1}{5}\begin{pmatrix} -1 & 6 & 0 & 0 & 0 \\ 6 & -3 & 2 & 0 & 0 \\ 0 & 2 & 1 & 2 & 0 \\ 0 & 0 & 2 & -3 & 6 \\ 0 & 0 & 0 & 6 & -1 \end{pmatrix}.$$

Note that U is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix.

We consider the case when $a = 0$, and we can achieve similar results as a previous general case when $a \neq 0$.

Theorem 2 *Let $\tilde{A}_n = T(n; 1, 0, g)$ with $g = b/c$. Then,*

a) For $n = 2q + 1$, $q = 0, 1, 2, \ldots$, $\tilde{A}_n$ is singular and $\mathcal{C}(\tilde{A}_n)$ does not exist.

b) For $n = 2q$, $q = 1, 2, 3, \ldots$, $U = \mathcal{C}(\tilde{A}_n)$ is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix, whose nonzero entries are $u_{i,i+1} = 1$, $u_{i+1,i} = 1$, $i = 1, 3, 5, \ldots, n-1$.

c) For any g_1 and g_2, $\mathcal{C}(T(2q; 1, 0, g_1)) = \mathcal{C}(T(2q; g_1, 0, 1)) = \mathcal{C}(T(2q; 1, 0, g_2)) = \mathcal{C}(T(2q; g_2, 0, 1))$.

Proof *a*) From Proposition 2 for $n = 2q + 1$, $q = 0, 1, 2, \ldots$, A_n is singular and $\mathcal{C}(A_n)$ does not exist.

b) Given the structure of $\tilde{A}_n$, we obtain $u_{i,i} = 0$ for $i = 1, 2, \ldots, n$. Since U is a doubly quasi-stochastic matrix, we have $u_{1,2} = u_{2,1} = 1$ and $u_{2,3} = u_{3,2} = 0$. Furthermore, for $i = 3, 5, \ldots, n-3$, we have $u_{i,i+1} = u_{i+1,i} = 1$ and $u_{i+1,i+2} = u_{i+2,i+1} = 0$. Finally, since $u_{n-1,n-1} = u_{n,n} = 0$, we obtain $u_{n-1,n} = u_{n,n-1} = 1$. Therefore, the nonzero entries of U are

$$u_{i,i+1} = u_{i+1,i} = 1, \qquad i = 1, 3, 5, \ldots, n-1.$$

c) Since $U = \mathcal{C}(\tilde{A}_n)$ does not depend on g, the result follows immediately.

□

The following example shows the results given in Theorem 2.

Example 2 *Consider the matrix $A_6 = T(6; c, 0, b)$ with $g = b/c = 2$. Then, $\tilde{A}_6 = T(6; 1, 0, 2)$. To obtain $U = \mathcal{C}(\tilde{A}_6)$ we calculate*

$$\begin{aligned} &u_{i,i} = 0, \;\; i = 1, 2, \ldots, 6 \\ &u_{1,2} = u_{2,1} = 1, \;\; u_{2,3} = u_{3,2} = 0 \\ &u_{3,4} = u_{4,3} = 1, \;\; u_{4,6} = u_{6,4} = 0 \\ &u_{5,6} = u_{6,5} = 1. \end{aligned}$$

Thus, the matrix U is given by

$$U = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Note that U is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix.

4 Conclusions

In this work, we study the structural properties of the combined matrix $\mathcal{C}(A_n)$ associated with a tridiagonal Toeplitz matrix $A_n = T(n; c, a, b)$. Our analysis covers two distinct cases based on the value of a. When $a \neq 0$, we establish that A_n is diagonally equivalent to $\tilde{A}_n = T(n; 1, 1, d)$ with $d = bc/a^2$. Conversely, when $a = 0$, $A_n = T(n; c, 0, b)$ is equivalent to $\tilde{A}_n = T(n; 1, 0, g)$ with $g = b/c$. We calculate the determinants of these matrices, showing that for $a \neq 0$, $\det(A_n)$ is expressed as a polynomial in d. For $a = 0$, $\det(A_n)$ depends on whether n is even or odd. If n is even, $\det(A_n)$ is expressed as a polynomial in g, but if n is odd, $\det(A_n)$ is zero. Additionally, we prove that the combined matrix $U = \mathcal{C}(A)$ is a bisymmetric, doubly quasi-stochastic tridiagonal Jacobi matrix in both cases. Moreover, for $\tilde{A}_n = T(n; 1, 1, d)$ the construction of U requires only specific entries, and for $\tilde{A}_n = T(n; 1, 0, g)$ we establish that U exists only for even n, has a particular pattern of nonzero entries and its construction is independent of the value of g.

References

[1] Liu, Z., Li, S., Yin, Y., Zhang, Y., Fast solvers for tridiagonal Toeplitz linear systems. *Comput. App. Mathematics*, 39(315):1–10, 2020.

[2] Encinas, A.M., Jiménez, M.J., Explicit inverse of a tridiagonal (p,r)-Toeplitz matrix. *Linear Algebra App.*, 542:402–421, 2018.

[3] Encinas, A.M., Jiménez, M.J., Boundary value problems for second order linear difference equations: application to the computation of the inverse of generalized Jacobi matrices. *RACSAM*, 113:3795–3828, 2014.

[4] Noschese, S., Pasquini, L., Reichel, L., Tridiagonal Toeplitz Matrices: Properties and Novel Applications. *Numerical Linear Algebra with Applications*, 13(8):1–20, 2006.

[5] Bristol, E., On a new measure of interaction for multivariable process control. *IEEE Trans. Autom. Control*, 11(1):133–134, 1966.

[6] Cantó, B., Cantó, R., Gassó, M.T., Urbano, A.M., Doubly stochastic and combined matrices. *Linear Mult. Algebra*, online:1–16, 2024.

[7] Horn, R.A., Johnson, C.R., Topics in Matrix Analysis. New York, Cambridge University Press, 1991.

[8] Minc, H., Nonnegative Matrices. New York, John Wiley and Sons Inc., 1988.

Koopman operator approach for data driven models to anticipate trends in cybersecurity incident reports

Miguel V. Carriegos[♭♮1] Noemí DeCastro-García[♭♮] and Ángel Luis Muñoz Castañeda[♭♮]

(♭) Departamento de Matemáticas, Universidad de León.
Campus de Vegazana, León, Spain.

(♮) Research Institute of Applied Sciences in Cybersecurity.
Campus de Vegazana, León, Spain.

1 Introduction

Cybersecurity reports may be aggregated into time-series in order to be studied properly [1]. Then time-series of aggregates would be forecasted in order to give an assessment of cybersecurity risk level, cybersecurity trends, and in general of any interesting measure we are able to format when the data is aggregated [2].

The talk is devoted to introduce a concrete example of a data-driven approximation to the problem of describing the cybersecurity incident report traffic.

2 Methods

Data-driven discovery of governing equations to describe dynamic processes is based in the Koopman operator approach [4] to non-linear dynamical systems. This is performed by taking into account that those systems are in fact presented as linear systems over a infinite dimensional vector space of complex-valued functions $\mathcal{F}$ by means of the Koopman operator $\mathcal{K} : \mathcal{F} \to \mathcal{F}$ based on a finite set of so-called observables. Then analyzing a nonlinear dynamical system over a finite dimensional space is traded by working on a linear system over a infinite dimensional $\mathbb{C}$-vector space.

When we deal with temporal data given by (say) a vector time-series obtained from concrete experiments we may perform the Dynamic Mode Decomposition (DMD) [5] to obtain a finite dimensional approximation to operator $\mathcal{K}$, that is to say, a matrix $K : \langle \mathcal{D} \rangle \to \langle \mathcal{D} \rangle$. Of course the choice of finite set of functions $\mathcal{D} \subset \mathcal{F}$ is crucial in the procedure.

[1] miguel.carriegos@unileon.es

3 Results

The talk focuses the following tasks: (a) Obtaining matrix K by means of Dynamic Mode Decomposition which involves Penrose pseudoinverse of data matrices. (b) Refreshing the matrix when new data is considered [6]. (c) Evaluate the information we are able to effectively forecast.

In order to give context to above task (c), note that the data we deal with is not obtained from a natural experiment where an underlying dynamical law is ruling the process. Cybersecurity incidents deal with human behavior and hence social activity. Thus the problem of finding an underlying law may be deceptive. Here we refer [3] as a work where a social activity which appears to be inherently non-predictable is shown to have trends.

Acknowledgments

This talk is partially supported by project *Técnicas básicas en ciberseguridad: evaluación y certificación automática y perspectiva post-cuántica*. A project of Instituto Nacional de Ciberseguridad (INCIBE) under NextGenerationEU funding of (PRTR) Plan de Recuperación, Transformación y Resiliencia. C081/23.

References

[1] Carriegos, M.V., Muñoz Castañeda, A.L., Trobajo, M.T., De Zaballa, D.A., On aggregation and prediction of cybersecurity incident reports. *IEEE Access*, 9:102636–102648, 2021.

[2] Cartier, P., Perrin, Y. Integration over finite sets, [in: Diener, F. & al. (eds.): Nonstandard Analysis in Practice]. Springer Verlag, 1995.

[3] Fliess, M., Join, C., A mathematical proof of the existence of trends in financial time series. *arXiv:0901.1945 [q-fin.ST]*, 2009.

[4] Koopman, B.O., von Neumann, J., Dynamical systems of continuous spectra. *Proc. Nat. Acad. Sci.*, 17(5):255–263, 1932.

[5] Schmidt, P.J., Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5–28, 2010.

[6] Thompson, R.C., Invariant factors under rank one perturbations. *Can. J. Math.*, XXXII(1):240–245, 1980.

Leveraging Group Actions for Constructing Convolutional Codes with Optimal Decoding and Distance Properties

Noemí DeCastro-García [♭♮1] Miguel V. Carriegos [♭♮] and Ángel Luis Muñoz Castañeda [♭♮]

(♭) Departamento de Matemáticas, Universidad de León

(♮) Research Institute of Applied Sciences in Cybersecurity (RIASC), Universidad de León
Campus de Vegazana s/n, León, Spain.

Abstract

The talk addresses the construction of observable convolutional codes with good decoding and distance properties for erasure channels using Input/State/Output (I/S/O) representation. There results are developed in [1].

Convolutional codes are a class of error-correcting codes used in communication systems to detect and correct errors that occur during data transmission. Formally, a convolutional code C is defined as a finite submodule of a free module $F[z]^n$, where F is a finite field, n is the number of output symbols, and k is the number of input symbols. A convolutional encoder $G(z)$ represents the transformation of an input sequence $u(z)$ to an output sequence $v(z)$, given by $v(z) = G(z) \cdot u(z)$ where $G(z)$ is a matrix over the ring $F[z]$ (polynomials over the finite field F).

One way to approach the construction of convolutional codes are the I/S/O representations. An I/S/O representation is a mathematical framework used to model convolutional codes as linear dynamical systems. The idea is to represent the encoding and decoding process using a system of linear equations that describe the state evolution over time. The I/S/O representation is described by the following system of equations:

$$x_{t+1} = A \cdot x_t + B \cdot u_t$$
$$y_t = C \cdot x_t + D \cdot u_t$$

where $x_t \in F^\delta$ is the state vector with δ is the dimension of the state space, the input vector $u_t \in F^k$, representing the information symbols at time t, and the output vector $y_t \in F^{n-k}$, representing the parity-check symbols at time t. The matrix triple (A, B, C, D) forms the I/S/O representation of the convolutional where $A \in F^{\delta\times\delta}$ is the state transition matrix, $B \in F^{\delta\times k}$ is the input matrix, $C \in F^{(n-k)\times\delta}$ is the output matrix, $D \in F^{(n-k)\times k}$ is the direct transmission matrix.

The primary goal in constructing convolutional codes is to ensure that they have good distance properties such as Maximum Distance Profile, which help improve their error correction performance. These properties guarantee that the code is capable of correcting a wide range of errors and erasures in the transmitted data. Another desired property is to create observable convolutional codes that are non-catastrophic, meaning that errors do not propagate uncontrollably. Also, it is

[1] ncasg@unileon.es

needed to know how to construct Good Decodable Property codes, which can efficiently recover from erasures in the transmission.

This talk focuses on algebraic methods to construct these codes, ensuring that they have desirable non-catastrophic, decoding and distance properties. A major contribution of this paper is the use of group actions to generate new convolutional codes with desired properties, starting from a given code. These group actions transform the I/S/O representation in ways that maintain the essential properties, such as reachability, observability and output observability, but lead to different convolutional codes. The types of transformations used include:

- Group actions in the state vector given by

$$(A', B', C', D') = (S^{-1}AS, S^{-1}B, CS, D)$$

- Group actions in the parity vector given by

$$(A', B', C', D') = (A, BQ, CQ, DQ)$$

  with an invertible matrix Q.

- Group actions in the information vector given

$$(A', B', C', D') = (A, B, H^{-1}C, H^{-1}D)$$

  with an invertible matrix H.

These transformations preserves the reachability, observability and output-observability, ensuring that the associated convolutional code remains good properties.

Acknowledgments

This work was conducted within the framework of the project *Algebraic Methods for the Recovery, Correction, and Security of Digital Information (MARCSID)*, funded by the Spanish Ministry of Science and Innovation under the State Plan for Scientific and Technical Research and Innovation 2021-2023, with grant number TED2021-131158A-I00.

References

[1] DeCastro-García,N., Carriegos, M. V. and Muñoz Castañeda, A.L. Construction of observable and MDP convolutional codes with good decodable properties for erasure channels by I/S/O representations, *arXiv:2412.00274*, `https://arxiv.org/abs/2412.00274`,2024.

On a novel rational-polynomial method to approximate the matrix logarithm

E. Defez[♭],[1] J. Ibáñez[♭], J.M. Alonso[♮], J. Sastre[◇] and A. Ibáñez[◇]

(♭) Instituto de Matemática Multidisciplinar,
(♮) Instituto de Instrumentación para Imagen Molecular,
(◇) Instituto de Telecomunicaciones y Aplicaciones Multimedia,
Universitat Politècnica de València. Camino de Vera s/n, 46022, Valencia, Spain.

Abstract

In this paper, a new algorithm for approximating the matrix logarithm based on a novel rational-polynomial approximation is presented. This new algorithm is consistently more accurate than state-of-the-art algorithms based on Padé approximations.

1 Introduction

As it is well known, for a non-singular square matrix $A \in \mathbb{C}^{n\times n}$ whose eigenvalues are found in $\mathbb{C} - (-\infty, 0)$, the matrix logarithm of A is defined as a matrix $X \in \mathbb{C}^{n\times n}$ such that it satisfies the matrix equation

$$A = \mathrm{e}^X, \tag{1}$$

where, of course, $\mathrm{e}^X = \sum_{i\geq 0} \frac{X^i}{i!}$ is the usual matrix exponential of X. Although any non-singular matrix A has infinite logarithms, we will only consider the principal logarithm represented by $\log(A)$. This is the only one whose eigenvalues all belong to the set $\{z \in \mathbb{C}; -\pi < Im(z) < \pi\}$, see [1].

The principal matrix logarithm is widely used in various disciplines of science and engineering. The applicability of the matrix logarithm in numerous distinct fields has driven the development of various approaches for its evaluation. The best known is the one based on Padé approximants on which the Matlab function `logm(A)` is based, see [2, 3].

The inverse scaling and squaring procedure, initially proposed in [4], is an extension to the matrix domain of the technique used by Briggs in the XVII century to compute his table of logarithms [5]. This method takes advantage of the matrix identity

$$\log(A) = 2^s \log(A^{1/2^s}), \tag{2}$$

and evaluates $\log(A)$ by combining argument reduction and approximation. By taking a certain number of square roots of matrix A, the problem is reduced to the computation of the logarithm

[1] edefez@imm.upv.es

of a matrix with eigenvalues close to 1. In fact, the matrix logarithm approximation is performed in three distinct steps, see the description given in [4]:

1. Find an integer s, so that matrix $A^{2^{-s}}$ is close to the identity matrix I. For that purpose, an algorithm that computes matrix square roots must be employed.
2. Approximate $\log\left(A^{2^{-s}}\right)$ using a known method.
3. Compute approximation $\log(A)$ using relation (2).

Taking into account the previous three-stage procedure, our proposal in this work is to compute the matrix logarithm in three phases somewhat similar to those described above, by using Daehee approximation under rational and polynomial approximations. Later, the most appropriate type of approach from the point of view of accuracy will be determined.

Throughout this paper, we will denote the identity matrix of order n as I_n, or I. The set of all the eigenvalues of a matrix $A \in \mathbb{C}^{r\times r}$ and any multiplicative norm are designated by $\sigma(A)$ and $\|A\|$, respectively. If $f(z)$ and $g(z)$ are holomorphic functions of the complex variable z, which are defined in an open set Ω of the complex plane, and A is a matrix in $\mathbb{C}^{r\times r}$ such that $\sigma(A) \subset \Omega$, then from the properties of the matrix functional calculus, [6, p. 558], it follows that $f(A)g(A) = g(A)f(A)$.

Daehee polynomials $D_n(x)$ are defined using its generating function (formula (1.1) of [7]) as

$$\left(\frac{\log(1+t)}{t}\right)(1+t)^x = \sum_{n\geq 0} D_n(x)\frac{t^n}{n!}\ ,\ |t|<1. \tag{3}$$

The explicit expression of $D_n(x)$ is given in formula (2.11) of [7], in the form:

$$D_n(x) = \sum_{l=0}^{n} S_1(n,l)B_l(x) \tag{4}$$

where $B_l(x)$ is the Bernoulli polynomial of degree l and $S_1(n,l)$ are the Stirling numbers of first type. Note that we can rewrite the formula (3) in the form

$$\log(1+t) = (1+t)^{-x}\left(\sum_{n\geq 0} D_n(x)\frac{t^{n+1}}{n!}\right)\ ,\ |t|<1. \tag{5}$$

By application of the holomorphic functional calculus to function $\log(1+t)$, acting on the matrix A, it follows from (5) that

$$\log(I+A) \quad = \quad [(I+A)^x]^{-1}\sum_{n\geq 0} D_n(x)\frac{A^{n+1}}{n!}\ ,\ \|A\|<1. \tag{6}$$

Note that by taking x as a positive integer, rational approximations are obtained. In contrast, a polynomial approximation is achieved if $x = 0$.

Using the partial sums of the series given in (6), the expression we are trying to test becomes:

$$\log(I+A) \approx [(I+A)^x]^{-1}\sum_{n\geq 0}^{m} D_n(x)\frac{A^{n+1}}{n!}. \tag{7}$$

The most adequate value of the parameter x has to be determined according to the numerical experiments to be carried out. Nevertheless, the Expression (7) can be reformulated as

$$\log(I+A) \approx [(I+A)^x]^{-1} P_m(A), \tag{8}$$

where $P_m(A)$ is a matrix polynomial of order m. In the next section, an algorithm built upon this approximation will be described.

2 Algorithm based on Daehee approximation

Algorithm 1 computes the matrix logarithm using approximation based on Expression (8). Broadly speaking, it is quite similar to Algorithm 1 developed by authors in [8]. In the first line, the pre-processing is performed. It consists of applying a similarity transformation to the matrix A to obtain a more balanced matrix $A = T^{-1}AT$. In Lines 2–12, the polynomial degree $m_{\bar{k}}$ and the scaling factor s are calculated to compute efficiently the logarithm approximation expressed in (8). Additionally, the matrix A is scaled according to the value s. For this purpose, a comparable study to the one developed by the authors in [8] has been carried out. It includes relative forward and backward error analyses, determining in each case the values of Θ_{m_k} which appears in Line 6. Then, the approximation defined in (8) is computed in Lines 13–14. Next, in Line 15, the matrix L is readjusted, according to the scaling implemented in Lines 2-12. Finally, the post-processing stage is applied in Line 16 having in mind the pre-processing phase performed in Line 1.

Algorithm 1 Given a matrix $A \in \mathbb{C}^{n\times n}$, a maximum approximation order m_k, $k \in \{1, 2, 3, \cdots, 18\}$ and $m \in \{2, 4, 6, 9, 12, 16, 20, 25, 30, 36, 42, 49, 56, 64, 72, 81, 90, 100\}$, and a positive integer x, this algorithm computes $L = \log(A)$ by a Daehee approximation of order lower than o equal to m_k.

1: Pre-processing of matrix A: $A \leftarrow T^{-1}AT$
2: $s = 0$
3: $finish = ||A - I|| \leq \Theta_{m_k}$
4: **while not** $finish$ and $s < maxiter$ **do**
5: $\quad \alpha = \max\{||(A-I)^{m_k}||^{1/m_k}, ||(A-I)^{m_k+1}||^{1/(m_k+1)}\}$
6: $\quad finish = \alpha \leq \Theta_{m_k}$
7: $\quad$ **if not** $finish$ **then**
8: $\quad\quad A \leftarrow A^{1/2}$, using scaled DB iteration (see [1, pp. 147–148])
9: $\quad\quad s = s + 1$
10: $\quad$ **end if**
11: **end while**
12: Find the smallest integer $\bar{k} \leq k$ such that $\alpha \leq \Theta_{m_{\bar{k}}}$
13: Compute $L = P_{m_{\bar{k}}}(A - I)$ by means of the Paterson-Stockmeyer method (see [1, pp. 73–74])
14: $L = [(I + A)^x]^{-1} L$
15: $L \leftarrow 2^s L$
16: Post-processing of matrix L: $L \leftarrow TLT^{-1}$

3 Numerical Experiments

In this section, we show the results of a comparative of the MATLAB implementation of Algorithm 1, named `logm_daehee`, and other MATLAB state-of-the-art codes. All tests were carried out on an IMac computer with an M4 processor, with a 10-cores CPU, a 10-cores GPU, and a RAM memory of 24GB. The implementations were executed in MATLAB version 2024b.

The values of the input parameters used in the tests for MATLAB implementation of Algorithm 1 were $m_k = 100$ ($k = 18$), and x=0, 1, 2. Additionally, the selected Θ_{m_k} of Algorithm 1 come from

relative backward errors. The MATLAB state-of-the-art codes used in the tests are the following ones:

- `logm_new`: this function transforms the input matrix A to the Schur triangular form $A = QTQ^{\star}$. Then, the logarithm of the matrix T is computed by using the inverse scaling and squaring technique and the Padé approximation. The square roots of matrix T are worked out by the Björck and Hammarling algorithm [1, eq. (6.3)] (see Algorithm 4.1 from [2]).
- `logm_iss_full`: the matrix logarithm is computed by means of the transformation-free form of the inverse scaling and squaring method with the Padé approximation. In this case, the matrix square roots are computed by the Denman–Beavers iteration (see Algorithm 5.2 from [2]).

The following matrix sets have been used in the numerical experiments. They are the same than those ones appearing in [9]:

- **Set 1**: one hundred diagonalizable 128×128 complex matrices computed as $A = V \cdot D \cdot V^T$, where D is a diagonal matrix with complex eigenvalues and V is an orthogonal matrix such as $V = H/\sqrt{n}$, where H is the Hadamard matrix of order n. The 2-norm of these matrices varies from 0.1 to 300. The "exact" matrix logarithm was calculated as $\log(A) = V \cdot \log(D) \cdot V^T$.
- **Set 2**: one hundred non-diagonalizable 128×128 complex matrices computed as $A = V \cdot J \cdot V^T$, where J is a Jordan matrix with complex eigenvalues whose algebraic multiplicity is randomly generated between 1 and 3, in progressively wider intervals from $[-2.5, 2.5]$, for the first matrix, to $[-250, 250]$, for the last one. The 2-norm of these matrices range from 3.39 to 337.72. The "exact" matrix logarithm was computed as $\log(A) = V \cdot \log(J) \cdot V^T$.
- **Set 3**: fifty-two matrices A from the Matrix Computation Toolbox (MCT) [10] and twenty from the Eigtool MATLAB Package (EMP) [11], all of them with a 128×128 order. To determine the "exact" matrix logarithm, the matrix logarithm is calculated firstly through the diagonalization of the matrix and secondly by means of the built-in MATLAB function `logm`. When the relative error in computing the matrix logarithm using these two alternatives is smaller than the unit roundoff error in floating-point arithmetic, then either of the two resulting matrices can be used as the "exact" matrix logarithm.

  For Set 3, a total of forty-six matrices were finally considered for the numerical experiments: forty from the MCT and six from the EMP. The others were excluded for the following reasons:

  - Matrices 17, 18, and 40 from the MCT collection, as well as matrices 7, 9, and 14 from the EMP set, failed to meet the requirements of the previously described method for computing the exact matrix logarithm.
  - For matrices 4, 6, 12, 26, 35, and 38 from the MCT, and matrices 1, 4, 10, 19, and 20 from the EMP, some of the tested algorithms produced relative errors equal to or greater than one. This was due to the poor conditioning of these matrices with respect to the matrix logarithm operation.
  - Matrices 2, 9, and 16 from the MCT, and matrices 15 and 18 from the EMP, led to a failure because the function `sqrtm_dbp`, which computes the principal matrix square root using the product form of the Denman-Beavers iteration did not converge within the maximum limit of 25 iterations.
  - Matrices 8, 11, 13, and 16 from the EMP collection were already included in the MCT dataset.

In all the calculations performed to compute the "exact" logarithmic of previous matrices, the built-in MATLAB function `vpa` from the Symbolic Math Toolbox with 256 decimal digits precision was employed.

The results of the numerical experiments carried out for the three sets of matrices are graphically collected, respectively, in Figure 1. They revealed that `log_daehee`, represented in the legends of the graphs as `log_daehee_x`, provided greater accuracy than the other functions in comparison, for any values of x=1, 2, and 3. The normwise relative errors plots (see graphs on the left side of Figure 1) showed that `log_daehee` and `logm_new` are more stable numerically than `logm_iss_full`.

These results are corroborated by the performance profiles (see graphs on the right side of Figure 1), where `log_daehee` always occupied the highest positions in the plots. Moreover, `log_daehee` provided similar results for the values $x = 0$, $x = 1$, and $x = 2$ for matrices in Sets 1 and 2, however offering worse results when $x = 2$ for Set 3.

On the other hand, Table 1 shows the response times corresponding to `log_daehee` ($x = 0$), `logm_new` and `logm_iss_full`. The longest execution times were required by `log_daehee`, and its shortest response times were obtained when $x = 0$. This is because, for x=0, it is not necessary to solve a linear equation system (the term $[(I + A)^x]^{-1}$ does not appear in Expression (7)).

Table 1: Execution times (seconds) in numerical tests.

Set	`log_daehee` ($x = 0$)	`logm_new`	`logm_iss_full`
1	8.25	4.31	6.93
2	7.96	4.32	6.13
3	2.26	2.27	2.25

From the results obtained, we can highlight the following observations:

- The best results in terms of numerical accuracy were offered by `log_daehee`, regardless of the value of x used.
- The normwise relative error figures shown that `log_daehee` and `logm_new` are numerically more stable than `logm_iss_full`.
- The performance profile graphs indicated that `log_daehee` could provide very similar results in accuracy terms for x=0, 1 or 2, but achieving the best results for x=0 and 1.
- The longest execution times corresponded to `log_daehee`.
- The shortest execution times of `log_daehee` were obtained when $x = 0$.

4 Conclusions

In this work, a new polynomial–rational algorithm based on Daehee polynomials has been introduced. When the approximation under consideration is of a polynomial type (i.e, $x = 0$), the new proposed algorithm presents better results in terms of numerical accuracy than the state-of-the-art ones based on Padé approximations, although the response times can be higher.

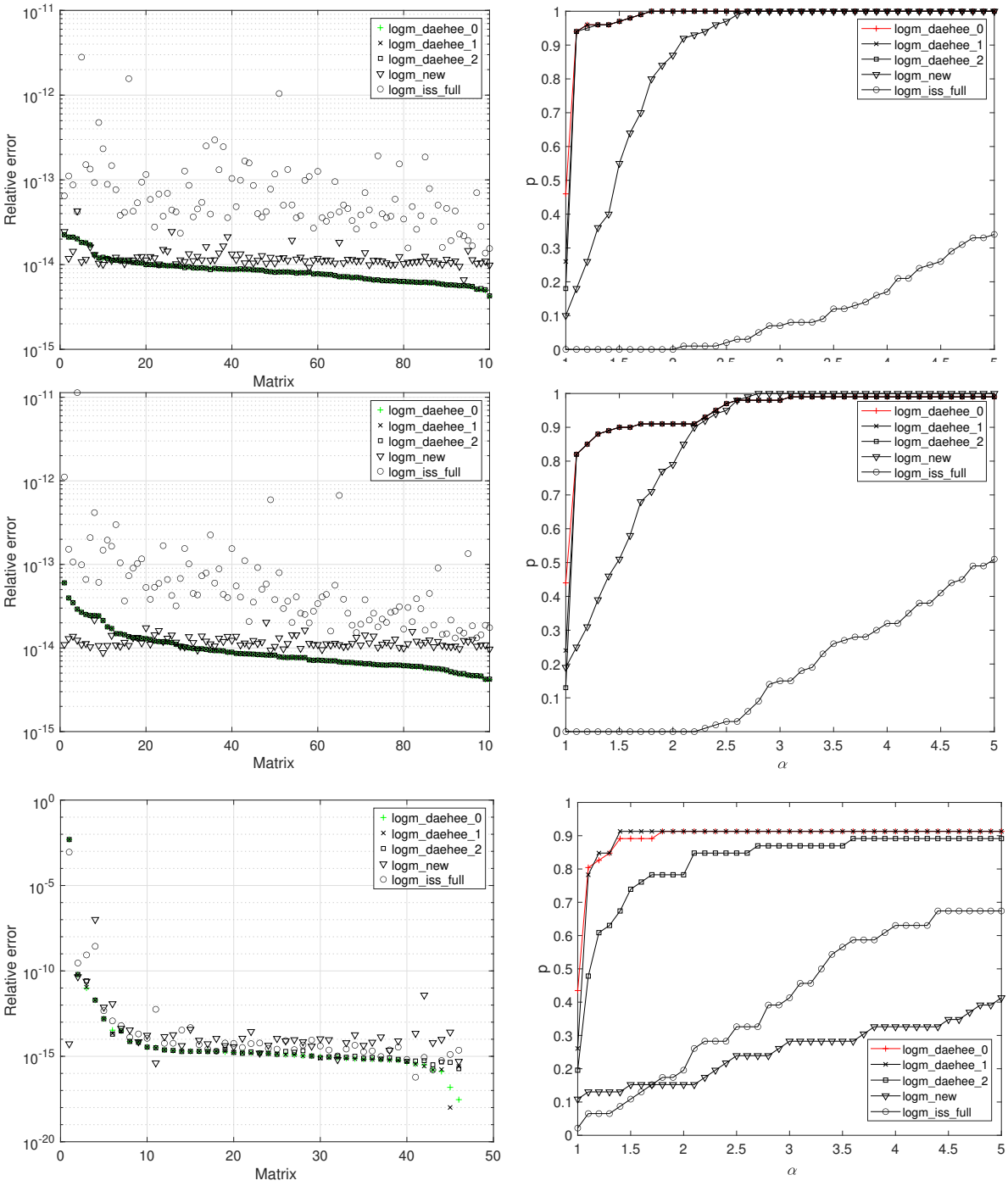


Figure 1: For the 3 sets of matrices in the testbed (from top to bottom), normwise relative errors (left) and performance profiles (right), for all the codes in comparison.

Acknowledgments

This work has been supported by the Generalitat Valenciana Grant CIAICO/ 2023/275 and the Vicerrectorado de Investigación de la Universitat Politècnica de València (PAID-11-23).

References

[1] Higham, N. J., Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[2] Al-Mohy, A. H., and Higham, N. J., Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM Journal on Scientific Computing*, 34(4):C153–C169, 2012.

[3] Al-Mohy, A.H., Higham, N. J., and Relton, S. D.. Computing the Fréchet derivative of the ma-trix logarithm and estimating the condition number. *SIAM Journal on Scientific Computing*, 35(4):C394–C410, 2013.

[4] Kenney, C., and Laub, A. J., Condition estimates for matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 10(2):191–209, 1989.

[5] Briggs, H., Logarithmorum chilias prima, London, 1617.

[6] Dunford, N. and Schwartz. J. T., Linear Operators, Part I: General Theory. John Wiley & Songs, Inc., 1988.

[7] San Kim, D, and Kim, T., Daehee numbers and polynomials. *Applied Mathematical Sciences*, 7(120):5969–5976, 2013.

[8] Ibáñez, J, Alonso, J. M., Defez, E., Alonso-Jordá, P., and Sastre, J., Computing the matrix logarithm with the Romberg integration method. *Algorithms*, 16(9):434, 2023.

[9] Ibáñez, J. Sastre, J., Ruiz, P., Alonso, J. M., and Defez, E., An improved Taylor algorithm for computing the matrix logarithm. *Mathematics*, 9(17):2018, 2021.

[10] Nicholas J. Higham. The Matrix Computation Toolbox. URL: http://www.ma.man.ac.uk/ higham/mctoolbox, 2002.

[11] Thomas G. Wright. Eigtool, version 2.1. URL: http://www.comlab.ox.ac.uk/pseudospectra/eigtool, 2009.

Computing Functions of Matrices with the Dual Scaling-And-Squaring Algorithm

M. Fasi,[♭,1] B. Iannazzo,[♮,2] and J. Sastre[♯,3]

(♭) School of Computer Science
University of Leeds, Woodhouse Lane, Leeds LS2 9JL, UK.

(♮) Dipartimento di Matematica e Informatica,
Università degli Studi di Perugia, Via Luigi Vanvitelli 1, Perugia, Italy.

(♯) Instituto de Telecomunicaciones y Aplicaciones Multimedia,
Universitat Politècnica de València, Camino de Vera s/n, 46022 València, Spain.

Abstract

A logarithm of the matrix $A \in \mathbb{C}^{n\times n}$ is any solution X to the equation

$$e^X = A. \tag{1}$$

If A is nonsingular, then (1) has infinitely many solutions [7, Thm. 1.27]. Among these, applications are typically interested in the *principal logarithm* of A, denoted by $\log A$, which is a solution to (1) whose spectrum lies in the strip $\{z \in \mathbb{C} : |\mathrm{Im}(z)| < \pi\}$. If A has no eigenvalues on the negative real semi-axis, then A has a principal logarithm, which is unique and real if A is real [7, Thm. 1.31]. The principal logarithm is the one usually returned by numerical implementations, including the MATLAB function `logm`.

The principal logarithm can also be viewed as a primary matrix function [7, Chap. 11] by extending the principal logarithm of a complex number to matrices. If A is diagonalizable, with eigendecomposition

$$A = M\mathrm{diag}(\lambda_1, \dots, \lambda_n)M^{-1},$$

then

$$\log A = M\mathrm{diag}\big(\log(\lambda_1), \dots, \log(\lambda_n)\big)M^{-1}, \tag{2}$$

and the extension to nondiagonalizable matrices can be made by continuity [7, Chap. 1].

If A is normal, then M is unitary and the principal logarithm can be computed with a straightforward evaluation of (2). For non-normal matrices, on the other hand, this approach may encounter numerical issues, as M can be ill conditioned. A good all-round method that can tackle these difficulties is the inverse scaling and squaring algorithm [10] which, in its basic implementation, performs four steps.

S1. Triangularise the matrix by computing its Schur decomposition $A =: UTU^*$, where U is unitary and T is upper triangular.

[1] m.fasi@leeds.ac.uk
[2] bruno.iannazzo@unipg.it
[3] jsastrem@upv.es

S2. Compute s square roots of T, where s is chosen so that the spectrum of $Y := T^{2^{-s}}$ lies in a neighbourhood of 1.

S3. Approximate $\log Y$ by computing $Z := p(Y)$, where p is a polynomial or rational approximant to $\log z$ at $z = 1$.

S4. Return the approximation $U 2^s Z U^*$.

The first and last step exploit the matrix identity $\log X^\alpha = \alpha \log X$, which implies that

$$\log A = U 2^s \log(T^{2^{-s}}) U.$$

The inverse scaling and squaring approach can be applied to other matrix functions, such as inverse trigonometric functions, which satisfy similar, albeit more involved, identities. This class of functions can benefit from an improved backward error analysis that, unlike standard approaches, does not use inverse compositional functions [15]. This is the case, for example, for the matrix cosine, whose backward error approximations are based on the arccos function and cannot be used for either Padé [2, Sec. 2.2] or Taylor approximants [12, Sec. 2.2]. The new backward error analysis technique [15] can be used in these contexts and, compared with other state-of-the-art methods, provides larger regions of convergence [11].

In the last few decades, the inverse scaling and squaring algorithm has been improved in various ways, including more refined heuristics to choose the number s of square roots to compute and better choices of the approximant p. Eventually, the scaling and squaring algorithm has become the customary method for computing the logarithm of a nonnormal matrix [1, 2, 8, 9].

Often, the function one wants to evaluated is defined implicitly as the solution to a matrix equation, as is the case for the matrix logarithm, which can be defined in terms of the matrix exponential, as shown in (1). Our idea is to follow a dual approach and modify step S3 in the standard algorithm above to read:

S3b. Approximate $\log Y$ by solving the equation $q(X) = Y$, where $q(z)$ is an approximant to $\exp z$ at $z = 0$.

This may seem to be an unnecessary complication of the direct approach, because, usually, solving an equation is more difficult than evaluating the function that defines it. But there are some notable exceptions: for example, computing a matrix–vector product with a triangular matrix requires exactly as many operations as solving a linear system with a triangular matrix coefficient – even though the latter problem is more difficult in general.

In prior work [3, 4], we have shown that if T is an upper triangular matrix and r is a rational function, then an evaluation scheme for r provides an algorithm for computing an upper triangular solution X to $r(X) = T$ that has the same asymptotic computational cost as evaluating $r(T)$. The focus on triangular matrices is not a restriction, because in the scaling and squaring algorithm for the matrix logarithm the approximant is evaluated at a triangular matrix.

This equivalence is not an exciting feature per se, and alone is not sufficient to justify our dual approach [5]. Solving the dual equation, however, has two benefits over a direct evaluation of the polynomial or rational approximant. First, note that while p, the diagonal Padé approximants to $\log z$ at $z = 1$, has no apparent structure, the $[2\ell + 1, 2\ell + 1]$ Padé approximant to $\exp z$ at $z = 0$ has the form

$$q(z) = \frac{g(z^2) + z h(z^2)}{g(z^2) - z h(z^2)},$$

where g and h are polynomials. Second, the approximants to the exponential [2] achieve binary64 accuracy in a region that is typically much larger than the region where an approximant to the matrix logarithm of same degree achieves that level of accuracy [1]. In existing literature, these

approximants are usually evaluated using the Paterson–Stockmeyer method [14], which satisfies some optimality properties [6] but is not the most efficient technique for evaluating a polynomial at a matrix. Sastre [13] has proposed new methods for polynomial evaluation that can have lower computational cost [9, Sec. 5].

Since $Y = T^{2^{-s}}$ is upper triangular, computing X such that $q(X) = Y$ will require fewer arithmetic operations compared with evaluating $Z = p(Y)$, and if X and Z have comparable approximation accuracy, then using our dual approach will deliver a faster algorithm without compromising the quality of the result. Numerical experiments comparing the new algorithms with the state of the art confirm that this is indeed the case, and that the potential performance improvement does not affect the accuracy of the computed result.

The dual approach is worth considering whenever the function $f(A)$ satisfies an equation $g(f(A)) = A$, and a polynomial or rational approximant to f is more expensive to compute than an approximant to g expected to deliver comparable accuracy. In such cases, instead of evaluating $p(Y)$, where p is an approximant to f, we can obtain a more efficient algorithm by solving the polynomial or rational equation $q(X) = Y$, where q is an approximant to g.

Acknowledgments

This work has been supported by the Generalitat Valenciana Grant CIAICO/2023/275 and the Vicerrectorado de Investigación de la Universitat Politècnica de València (PAID-11-23).

References

[1] Al-Mohy, A. H., Higham, N. J., Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM Journal on Scientific Computing*, 34(4):C153–C169), 2012.

[2] Al-Mohy, A. H., Higham, N. J., Relton, S. D., Computing the Fréchet derivative of the matrix logarithm and estimating the condition number. *SIAM Journal on Scientific Computing*, 35:C394–410, 2013.

[3] Fasi, M., Iannazzo, B., Computing primary solutions of equations involving primary matrix functions. *Linear Algebra and Its Applications*, 560:17–42, 2019.

[4] Fasi, M., Iannazzo, B., Substitution algorithms for rational matrix equations. *Electronic Transactions on Numerical Analysis*, 53:500–521, 2020.

[5] Fasi, M., Iannazzo, B., The dual inverse scaling and squaring algorithm for the matrix logarithm. *IMA Journal of Numerical Analysis*, 42:2829–2851, 2022.

[6] Fasi, M., Optimality of the Paterson-–Stockmeyer method for evaluating matrix polynomials and rational matrix functions, *Linear Algebra and Its Applications*, 574:182–200, 2019.

[7] Higham, N. J., Functions of Matrices. Society for Industrial and Applied Mathematics, Philadelphia, 2008.

[8] Ibáñez, J., Sastre J., Ruiz, P., Alonso, J. M., Defez, E., An improved Taylor algorithm for computing the matrix logarithm. *Mathematics*, 9:2018, 2021.

[9] Jarlebring, E., Sastre, J., Ibáñez, J., Polynomial approximations for the matrix logarithm with computation graphs. *Linear Algebra and its Applications*, in press, available online 2024.

[10] Kenney, C. S., Laub, A. J., Condition estimates for matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 10:191–209, 1989.

[11] Nadukandi, P., Higham, N.J., Computing the wave-kernel matrix functions. *SIAM Journal on Scientific Computing*, 40(6):A4060–A4082, 2018.

[12] Sastre, J. , Ibáñez, J., Alonso, P., Peinado, J., Defez, E., Two algorithms for computing the matrix cosine function, *Applied Mathematics and Computation*, 312:66–77, 2017.

[13] Sastre, J., Efficient evaluation of matrix polynomials, *Linear Algebra and its Applications*, 539:229–250, 2018.

[14] Paterson, M. S., Stockmeyer, L. J., On the number of nonscalar multiplications necessary to evaluate polynomials, *SIAM Journal on Computing*, 2(1):60–66, 1973.

[15] Sastre, J., Ibáñez, J., On the backward and forward error of approximations of analytic functions and applications to the computation of matrix functions , *Journal of Computational and Applied Mathematics*, 419, 114706, 2023.

Reducing computational complexity in 3D printing optimization: A quotient-based approach to order matrices

P. Alonso ♭, F. Fueyo ♭, P. García-Medina♭,[1] J. Jiménez ♭, M. L. Serrano ♭ and A. Tepavčević ♮

(♭) Departamento de Matemáticas, Universidad de Oviedo
Facultad de Ciencias Avda. de Calvo Sotelo s/n 33007 Oviedo, Asturias, Spain.

(♮) Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad
Trg Dositeja Obradovića, 21000, Novi Sad, Serbia.

Abstract

A key challenge in 3D printing optimization is minimizing printing time, where partially ordered sets (POSets) play a crucial role. This has recently been explored in [1], where the concept of compatible totally ordered sets (TOSets) is applied to additive manufacturing. In particular, constructing the full-order matrix from directly related pairs (i.e., those without intermediate elements) becomes computationally expensive for large matrices, such as those of size 14000×14000.

This work presents a novel approach to computing the full-order matrix from a reduced one. The core idea relies on defining a congruence in the POSet (i.e., an equivalence relation compatible with the order, whose classes are convex subsets [2]): Reducible elements form singleton classes, while irreducible elements are grouped according to irreducible chains. We demonstrate that each congruence constitutes a TOSet with well-defined minimum and maximum elements.

Using this structure, we construct an induced-order relation on the quotient set, significantly simplifying the computation. Finally, we prove that the original order relation can be reconstructed from the reduced-order relation using a box-based approach. This method offers an efficient strategy for handling large-order matrices, enhancing the feasibility of optimizing 3D printing processes.

References

[1] Abascal, P., Fueyo, F., Jiménez, J., Palacio, A., Serrano, M. L., Compatible TOSets with POSets: an Application to Additive Manufacturing, *Journal of Optimization Theory and Applications*, 201(3): 177–198, 2024.

[2] Williams, N. J., A survey of congruences and quotients of partially ordered sets, *EMS Surveys in Mathematical Sciences*, 11(1): 153–203, 2024.

[1] garciamedpaloma@uniovi.es

Decoding convolutional codes over finite rings. A linear dynamical systems approach

Ángel Luis Muñoz Castañeda[♭♮1], Noemí DeCastro-García[♭♮] and Miguel V. Carriegos[♭♮]

(♭) Departamento de Matemáticas, Universitad de León
Campus de Vegazana s/n, León, Spain.

(♮) Research Institute of Applied Sciences in Cybersecurity (RIASC), Universidad de León
Campus de Vegazana s/n, León, Spain.

1 Introduction

Convolutional codes are a central component in coding theory for reliable communication over noisy channels. Originally developed over finite fields, these codes have also been extended to more general algebraic structures such as finite rings. This paper studies convolutional codes over the finite chain ring $\mathbb{Z}_{p^r}$ and addresses the decoding problem using techniques from linear dynamical systems and p-adic analysis.

This work is basesd on the fact that observable convolutional codes with the Predictable Degree Property (PDP) over $\mathbb{Z}_{p^r}$ admit minimal input/state/output (I/S/O) representations [3]. These representations offer a useful algebraic framework that aligns convolutional codes with state-space systems and enables a structured approach to decoding.

2 Methods

Let $\mathcal{C} \subset \mathbb{Z}_{p^r}[z]^n$ be an (n, k) convolutional code over $\mathbb{Z}_{p^r}$, defined as the image of a polynomial encoder matrix $G(z) \in \mathrm{Mat}_{n\times k}(\mathbb{Z}_{p^r}[z])$. The convolutional code is said to have the PDP if for all inputs $u(z)$, the degree of $v(z) = G(z)u(z)$ satisfies:

$$\deg v(z) = \max_{1\leq i\leq k} \{\deg u_i(z) + \deg g_i(z)\},$$

where $g_i(z)$ is the i-th column of $G(z)$.

The I/S/O representation of the PDP convolutional code corresponds to a system of the form:

$$\Sigma = \begin{cases} x_{t+1} = Ax_t + Bu_t, \\ y_t = Cx_t + Du_t, \end{cases}$$

where $x_t \in \mathbb{Z}_{p^r}^{\delta}$ is the state, $u_t \in \mathbb{Z}_{p^r}^{k}$ the input, and $y_t \in \mathbb{Z}_{p^r}^{n-k}$ the output. The decoding strategy follows Rosenthal's algorithm, which uses this system representation. It uses two derived linear block codes:

[1] amunc@unileon.es

- The *reachability code*, defined by the kernel of the reachability matrix $\Phi_T = [A^{T-1}B \cdots B]$.
- The *observability code*, defined by the image of the observability matrix $\Psi_\Theta = \begin{bmatrix} C^\top & (CA)^\top & \cdots & (CA^{\Theta-1})^\top \end{bmatrix}^\top$.

The key idea is to decode time segments using these block codes. To carry this out over $\mathbb{Z}_{p^r}$, the authors combine Rosenthal's method [4] with:

- The Greferath-Vellbinger (GV) algorithm [2], which iteratively corrects p-adic digits using decoding over $\mathbb{Z}_p$.
- An adapted Torrecillas-Lobillo-Navarro (TLN) algorithm [1] for syndrome decoding using parity-check matrices with p-adic expansion.

3 Results

We prove that under standard assumptions—such as the invertibility of A and full-rank conditions for Φ_T and Ψ_Θ—the decoding algorithm can correct any error vector $e(z)$ whose weight in a time window of size T satisfies:

$$\mathrm{wt}(e(z)) \leq \lambda := \min\left\{ \left\lfloor \frac{d_{\mathrm{reach}} - 1}{2} \right\rfloor, \left\lfloor \frac{T}{2\Theta} \right\rfloor \right\},$$

where d_{reach} is the minimum distance of the reachability block code over $\mathbb{Z}_{p^r}$ (or its reduction mod p).

We also demonstrate that decoding can be performed blockwise by combining the system equations and block code decoders, enabling parallelism and error localization.

The theoretical performance aligns with:

$$t_{\mathrm{correctable}} = \left\lfloor \frac{d_{\mathrm{free}}(\mathcal{C} \bmod p) - 1}{2} \right\rfloor,$$

as the free distance over the field $\mathbb{Z}_p$ governs the worst-case behavior over the ring $\mathbb{Z}_{p^r}$.

4 Conclusions

This work generalizes a powerful decoding strategy for convolutional codes to the setting of finite rings. The use of I/S/O representations offers both conceptual clarity and computational advantages. Combining p-adic decoding techniques with system-theoretic representations leads to a practical algorithm with provable guarantees under mild conditions.

Future directions include extending this method to erasure channels and exploring its integration with cryptographic schemes or network coding scenarios where ring structures are advantageous.

Acknowledgments

This work is part of the project TED2021-131158A-I00, funded by MCIN/ AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR.

References

[1] Gómez-Torrecillas, J., Lobillo, F. J., Navarro, G., Decoding Linear Codes over Chain Rings Given by Parity Check Matrices, *Mathematics*, 9(16):1878, 2021.

[2] Greferath, M., Vellbinger, U., Efficient decoding of $\mathbb{Z}_{p^k}$-linear codes, *IEEE Transactions on Information Theory*, 44(3):1288–1297, 1998.

[3] Muñoz Castañeda, A. L., DeCastro-García, N., Carriegos, M. V., On the State Approach Representations of Convolutional Codes over Rings of Modular Integers, *Mathematics*, 9(22):2962, 2021.

[4] Rosenthal, J., An Algebraic Decoding Algorithm for Convolutional Codes, *Dynamical Systems, Control, Coding, Computer Vision*, Eds.: G. Picci and D. S. Gilliam, Basel: Birkhäuser, 343–360, 1999.

A new method and simple to get the priority vector for reciprocal matrices

J. Benítez[♭],[1] C. Serra[♮ 2]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) Universitat Politècnica de València,
Camí de Vera s/n, València, Spain.

1 Introduction

Analytic hierarchy process (AHP) is one of the most important methods in leading multi-attribute decision-aiding model. This process is designed to help make better choices when an expert (a decision maker) deals with complex decisions involving several choices, and it was developed by T. Saaty [11]. AHP has been applied in several areas, such as business, manufacturing, industry, government, education, ... [7].

The expert evaluates the possible choices by using pairwise comparisons between the alternatives forming a pairwise comparison matrix $A = (a_{ij})_{i,j=1}^{n}$. When the expert makes the comparisons, he can use concrete data about the choices and his judgments about the elements employing his knowledge about the matter. Therefore, human judgments is used in performing the evaluations. Each entry $a_{ij} > 0$ of a pairwise comparison matrix measures the preference of alternative i over alternative j, and if we use a multiplicative criterion, $a_{ij}a_{ji} = 1$ holds. Evidently, from this we deduce $a_{ii} = 1$, which says that an alternative i is equal to itself.

2 Methods

2.1 A mathematical review of AHP and the eigenvector method

The set of $n \times m$ real matrices and its subset composed of positive matrices will be denoted, respectively, by $M_{n,m}$ and $M_{n,m}^{+}$. We simplify the notation for square matrices: $M_n = M_{n,n}$ and $M_n^{+} = M_{n,n}^{+}$. Any vector in $\mathbb{R}^n$ will be regarded as a column vector, i.e., an element of $M_{n\times 1}$. If A is a matrix, then A^T will denote the transpose of A. The vector $[1,\ldots,1]^T \in \mathbb{R}^n$ will be denoted by $\mathbf{e}$.

A **reciprocal matrix** $A = (a_{ij})$ is a square in M_n^{+} that satisfies $a_i a_j = 1$ for all $1 \leq i, j \leq n$. The element a_{ij} expresses the multiplicative importance of the ith alternative over the jth

[1] jbenitez@mat.upv.es
[2] cserjim@upv.edu.es

(many times Saaty's scale is used, see, e.g., [11]). Evidently, by setting $i = j$ in $a_{ij}a_{ji} = 1$, one has that $a_{ii} = 1$. A **consistent matrix** $A = (a_{i,j}) \in M_n^+$ satisfies $a_{ij}a_{jk} = a_{ik}$ for all $1 \leq i, j, k \leq n$. It is trivial to prove that any consistent matrix is reciprocal (set $k = i$ in the previous equality).

As it was said, usually the decision maker builds a comparison matrix $A \in M_n^+$ which is reciprocal, but is not consistent. In such case, it is important to determinate the priority vector Let $\boldsymbol{v} \in \mathbb{R}^n$ be a vector that is not explicitly known to the decision maker.

By using the Perron theorem (see, e.g., [9]), we get that n is the Perron root of any consistent $n \times n$ matrix and the Perron vector is the priority vector. However, if A is a reciprocal matrix which is not consistent, Saaty [11] recommends to use the Perron vector of A as the priority vector under the assumption that A is close to be consistent. If the inconsistency of a reciprocal matrix A is larger than desired, then we cannot use the Perron vector of A as the priority vector, and some extra computation are needed. Most of the methods to find an appropriate priority vector $\mathbf{v} = [v_1, \ldots, v_n]^T$ of A use $a_{ij} \approx v_i/v_j$ (see [5]). In this paper, we propose a method to find the priority vector which agrees with the idea of the eigenvector method proposed by Saaty and has a very simple solution unlike other ways to find the priority vector because only a linear system is required.

3 Results

3.1 A new optimization method to find the priority vector

Let $A \in M_n^+$ be a reciprocal matrix which will be assumed throughout this section non consistent. Also, we will assume that $A - nI_n$ is a non singular matrix (which is the most common of the cases). Hence the linear system $A\mathbf{x} = n\mathbf{x}$ has the unique trivial solution $\mathbf{x} = \mathbf{0}$.

Therefore, the following system

$$(A - nI_n)\mathbf{x} = \mathbf{0}, \qquad x_1 + \cdots + x_n = 1$$

has no solution. Observe that the normalisation condition $x_1 + \cdots + x_n = 1$ can be rewritten as $\mathbf{e}^T\mathbf{x} = 1$.

The new method that we propose to find the priority vector $\mathbf{x}$ of A is given by the solution of the following problem.

$$\text{Minimize } \|(A - nI_n)\mathbf{x}\| \text{ restricted to } \mathbf{e}^T\mathbf{x} = 1. \tag{1}$$

Here, $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$, i.e., $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}$.

Observe that when A is consistent, the solution of the problem formulated in (1) is the priority vector of A. If $A - nI_n$ is singular, the solution of (1) is trivial: solve $(A - nI_n)\mathbf{x} = 0$ constrained to $\mathbf{e}^T\mathbf{x} = 1$

We will use the following theorem whose proof will be given in an appendix. We give a purely algebraic proof that only uses standard linear algebra (inner product spaces) avoiding the use of the method of the multipliers Lagrange. Notice that this last method only establishes the critical points of the objective function, and not the points minimising the objective function.

Theorem 1. *Let M and C be matrices of sizes $n \times m$ and $m \times p$, respectively; let $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{d} \in \mathbb{R}^p$. Suppose there exist vectors $\boldsymbol{z} \in \mathbb{R}^p$ and $\boldsymbol{x}_0 \in \mathbb{R}^m$ such that*

$$\begin{bmatrix} M^T M & C \\ C^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{z} \end{bmatrix} = \begin{bmatrix} M^T \boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}. \tag{2}$$

Then:

(i) $C^T \boldsymbol{x}_0 = \boldsymbol{d}$.

(ii) $\|M\boldsymbol{x}_0 - \boldsymbol{b}\| \leq \|M\boldsymbol{x} - \boldsymbol{b}\|$ *for any* $\boldsymbol{x} \in \mathbb{R}^m$ *which satisfies* $C^T \boldsymbol{x} = \boldsymbol{d}$.

(iii) If the columns of $\begin{bmatrix} M \\ C^T \end{bmatrix}$ *are linearly independent and the rows of* C^T *are linearly independent, then the solution to* (2) *is unique.*

In order to employ Theorem 1 to solve the problem given in (1), we set $M = A - nI_n$, $C = \mathbf{e}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{d} = 1$. Next, we will prove that condition (iii) of Theorem 2 holds. Since $C^T = \mathbf{e}^T$ is a nonzero row, then the rows of C^T are independent. Furthermore, $M = A - nI_n$ is nonsingular. So, the n columns of M are independent, hence the columns of $\begin{bmatrix} M \\ C^T \end{bmatrix}$ are independent. We have proved the following result.

Theorem 2. *The problem stated in (1) has a unique solution.*

3.2 Using the QR decomposition

According to Theorem 1, the unique solution $\boldsymbol{x} \in \mathbb{R}^n$ of the problem given in (1) must satisfy

$$\begin{bmatrix} (A - nI_n)^T (A - nI_n) & \mathbf{e} \\ \mathbf{e}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}.$$

Hence

$$(A - nI_n)^T (A - nI_n)\mathbf{x} = -z\mathbf{e}, \qquad \mathbf{e}^T \mathbf{x} = 1.$$

Obviously, $z \neq 0$. Define $B = (A - nI_n)^T (A - nI_n)$, a nonsingular matrix. If $\mathbf{x}_0$ is the unique solution of $B\mathbf{x}_0 = \mathbf{e}$, then $\mathbf{x}_0 = B^{-1}\mathbf{e} = -z^{-1}\mathbf{x}$, so $\mathbf{e}^T \mathbf{x}_0 = -z^{-1}\mathbf{e}^T \mathbf{x} = -z^{-1} \neq 0$ and

$$\frac{1}{\mathbf{e}^T \mathbf{x}_0}\mathbf{x}_0 = \mathbf{x} \text{ is the unique solution of problem (1).}$$

Solving $B\mathbf{x}_0 = \mathbf{e}$ directly is not adequate when n is large because the cross matrix $B = (A - nI_n)^T (A - nI_n)$ is ill-conditioned. The QR factorization is suitable to avoid numerical errors. Let $A - nI_n = QR$ be the QR factorization of $A - nI_n$. We have $(A - nI_n)^T (A - nI_n) = (QR)^T (QR) = R^T Q^T QR = R^T R$. Therefore, the system $M\mathbf{x}_0 = \mathbf{e}$ $B\mathbf{x}_0 = \mathbf{e}$ is equivalent to $R^T R\mathbf{x}_0 = \mathbf{e}$. Therefore, the solution $\mathbf{x}$ of problem (1) can be got by solving

$$R^T \mathbf{y} = \mathbf{e}, \quad R\mathbf{x}_0 = \mathbf{y}, \quad \mathbf{x} = \frac{1}{\mathbf{e}^T \mathbf{x}_0}\mathbf{x}_0. \tag{3}$$

Theorem 3. *Let $A \in M_{n,n}$ be a reciprocal, non consistent matrix such that $A - nI_n$ is nonsingular. Let QR be the QR factorization of $A - nI_n$,*

Then the solution $\mathbf{x}$ *of Problem (1) is given by (3).*

3.3 Using the singular value decomposition

One of the most useful decompositions is the singular value decomposition (SVD), see e.g. [6, Chapter 2]. If $A - nI_n = U\Sigma V^T$ is the SVD of $A - nI_n$, then since the Euclidean norm is preserved under orthogonal premultiplication, problem (2) is equivalent to

$$\text{Minimize } \|\Sigma V^T \mathbf{x}\| \text{ restricted to } \mathbf{e}^T\mathbf{x} = 1. \tag{4}$$

Setting $\mathbf{y} = V^T\mathbf{x}$, then (4) is equivalent to

$$\text{Minimize } \|\Sigma \mathbf{y}\| \text{ restricted to } \mathbf{e}^T V\mathbf{y} = 1 \tag{5}$$

because V is orthogonal. Let $\mathbf{c} = [c_1, \ldots, c_n]^T = V^T\mathbf{e}$ and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the singular values of $A - nI_n$. Observe that $\mathbf{c} \neq \mathbf{0}$ (because V is non singular) and $\sigma_n > 0$ (because $A - nI_n$ is non singular). Problem (5) is equivalent to

$$\text{Minimize } \sigma_1^2 y_1^2 + \cdots + \sigma_n^2 y_n^2 \text{ restricted to } c_1 y_1 + \cdots + c_n y_n = 1.$$

By using the method of Lagrange multipliers, there exists $\alpha \in \mathbb{R}$ such that...

$2\sigma_i^2 y_i = \alpha c_i$ for $i = 1, \ldots, n$. Therefore,

$$\mathbf{x} = V\mathbf{y} = \frac{\alpha}{2} V \left[\frac{c_1}{\sigma_1^2}, \ldots, \frac{c_n}{\sigma_n^2}\right]^T.$$

If we define $\mathbf{x}_0 = V\left[c_1/\sigma_1^2, \ldots, c_n/\sigma_n^2\right]^T = 2\alpha^{-1}\mathbf{x}$ (observe that $\alpha \neq 0$ since $\mathbf{x} \neq \mathbf{0}$), then

$$\frac{1}{\mathbf{e}^T\mathbf{x}_0}\mathbf{x}_0 = \frac{1}{2\alpha^{-1}\mathbf{e}^T\mathbf{x}} 2\alpha^{-1}\mathbf{x} = \mathbf{x} \tag{6}$$

because $\mathbf{e}^T\mathbf{x} = 1$. Therefore, the solution of the problem stated in (1) is the following:

$$\mathbf{x} = \frac{1}{\mathbf{e}^T\mathbf{x}_0}\mathbf{x}_0, \qquad \mathbf{x}_0 = V \begin{bmatrix} c_1/\sigma_1^2 \\ \vdots \\ c_n/\sigma_n^2 \end{bmatrix} = V\Sigma^{-2}\mathbf{c} = V\Sigma^{-2}V^T\mathbf{e}. \tag{7}$$

3.4 Comparison with other methods.

There are many methods in the literature to get a priority vector ([8, Chapter 3]), we can cite the Eigenvector Method (EM), the Geometric Mean Method (GMM) and a family of optimizaton methods such as the Least Square Method (LSM), Weighted Least Square Method (WLSM), Least Worst Squares Method (LWSM), and many others. All of these (including the proposed in this article) are based on certain heuristic or on an optimization problem. Hence the obtained rankings depend on the optimization function. Therefore we cannot claim that our method is better than the others because our method is better if we consider the minimization problem (1), but if we change the optimization function, we do not get the optimal result.

However, we consider that the minimization problem (1) considered here is a natural consequence of the fact that the priority vector of a consistent matrix $n \times n$ is an eigenvector associated to the eigenvalue n.

It is important to note that many of the methods to get the priority vector require to solve non linear equations. For example, the EM requires to solve the equation $\det(A - \lambda I_n) = 0$; the LSM problem is difficult to solve as the objective function is non linear, usually non convex, and no unique solution exists [4]. The LWSM has the same handicaps as the LSM. Other methods such as Weighted Least Absolute Error Method and Weighted Least Worst Method require to solve a linear program. The unique two methods with closed solutions is the Weighted Least Square Method (WLSM) and the Logarithm Least Square Method (LLSM). The method proposed in this paper only requires to solve a square linear system, and we give two robust approaches: the QR and the SVD decomposition.

4 Conclusions

We give a new and simple method to find the priority vector of a $n \times n$ non consistent matrix. This method leads to a constrained least square problem, which is solved in an efficient manner by two standard matrix decompositions: the QR and the SVD of a non singular $n \times n$ matrix. We also compare our method with other methods to get the priority vector showing that, although the minimisation problem changes, we get close results. As a future work, it is needed extensive numerical simulations to compare the proposed method with others concerning the POIP condition.

References

[1] C.A Bana e Costa, J. Vansnick, A critical analysis of the eigenvalue method used to derive priority vectors in AHP. *European Journal of Operational Research*, 187(3): 1422-1428. 2008.

[2] J. Barzilai, Deriving weights from pairwise comparison matrices. J. Oper. Res. Soc., 48: 1226-1232, 1997

[3] J. Benítez, X. Delgado-Galván, J. Izquierdo, R. Pérez-García, Improving consistency in AHP decision-making processes, *Applied Mathematics and Computation,* 219: 2432-2441, 2012.

[4] S. Bozóki. Solution of the least squares method problem of pairwise comparison matrices. *Central European Journal of Operational Research*, 16(4): 345-358, 2008.

[5] E.U. Choo, W.C. Wedley, A common framework for deriving preference values from pairwise comparison matrices, *Computers & Operations Research*, 31(6): 893-908, 2004.

[6] G. Golub, C.F. Van Loan. Matrix computations (4th Edition). The John Hopkins University Press, 2013.

[7] W. Ho. Integrated analytic hierarchy process and its applications - A literature review. *European J. Oper. Res.* 186: 211-228, 2008.

[8] K. Kułakowski. Understanding the Analytic Hierarchy Process. CRC Press. 2021.

[9] C.D. Meyer. Matrix Analysis and Applied Linear Algebra. SIAM

[10] L. Mikhailov A fuzzy programming method for deriving priorities in the analytic hierarchy process *J. Oper. Res. Soc.*, 51 (3): 341-349, 2000.

[11] T. Saaty, Relative measurement and its generalization in decision making: Why pairwise comparisons are central in mathematics for the measurement of intangible factors— The Analytic Hierarchy/Network Process, *Revista de la Real Academia de Ciencias Exactas, F ísicas y Naturales, Serie A, Matemáticas* 102: 251–318, 2008.

[12] T.L. Saaty Decision-making with the AHP: why is the principal eigenvector necessary *European J. Oper. Res.,* 145: 85-91, 2003.

[13] T.L. Saaty, L.G. Vargas Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios Math. Model., 5(5): 309-324, 1984

[14] B. Srdjevic Combining different prioritization methods in the analytic hierarchy process synthesis Comput. Oper. Res., 32 (7): 1897-1919, 2005.

On the relation between the spectra of linear pencils and linear relations

T. Reis[♭], C. Trunk[♭], [1] and H. Winkler[♭]

(♭) TU Ilmenau Institute for Mathematics, Technische Universität Ilmenau, Postfach 10 05 65, D-98684 Ilmenau, Germany

The relation between the spectra of operator pencils with unbounded coefficients and of associated linear relations is investigated. It turns out that various types of spectrum coincide and the same is true for the Weyr characteristics. This characteristic describes how many independent Jordan chains up to a certain length exist.

Linear relations are just subspaces of the cartesian product $X \times X$ of a normed Banach space X. They allow a rich spectral theory. Here we associate with an operator pencil

$$\lambda E - A$$

the two subspaces

$$\ker[A, -E] \quad \text{and} \quad \operatorname{ran} \left[\begin{smallmatrix} E \\ A \end{smallmatrix}\right],$$

which are called, for obvious reasons, the *kernel* and the *range representation*. The spectrum and the Weyr characteristic of the operator pencil coincides with the spectrum and the Weyr characteristic for the range and the kernel representation. Furthermore, we consider Hermitian pencils and their correspondig range and kernel representations.

[1] carsten.trunk@tu-ilmenau.de

Part VII

Maths & Physics

New Neural Network Approaches to Cosmic Microwave Background Mapping

I. García-Prieto[♭], M. D. Guillen[♭] and R. Dale[♭],[1]

(♭) Departament d'Estadística, Matemàtiques i Informàtica, Universitat Miguel Hernández d'Elx
Avinguda de la Universitat d'Elx, s/n, 03202 Elx, Alacant, Spain.

1 Introduction

In contemporary cosmological research, maps of the so-called Cosmic Microwave Background (CMB) radiation—offering a "snapshot" of the early Universe when it was merely 375,000 years old, a tiny fraction of its current age of 13.77 billion years—constitute a crucial source of observational evidence.

Since its discovery in 1965, the CMB has been studied extensively through various observational programmes. Among these, the three major satellite missions—COBE (NASA, 1989), WMAP (NASA, 2001), and Planck (ESA, 2009)—are particularly noteworthy [1, 2, 3]. Each mission has successively improved in both angular resolution and measurement precision, enabling increasingly refined mapping of the early Universe.

These observations suggest that the microwave background exhibits statistical homogeneity, isotropy, and a Gaussian distribution. This allows for the calculation of an average temperature (denoted T_B), from which we can define a temperature contrast function in a given direction $\vec{n}$ on the sky as follows:

$$\frac{\Delta T}{T}(\vec{n}) = \frac{T(\vec{n}) - T_B}{T_B} \tag{1}$$

The statistical study of this Gaussian distribution $\Delta T/T(\vec{n})$ has been the subject of numerous investigations. One of the most common approaches involves computing the two-point correlation function, defined as:

$$C(\theta) = \left\langle \frac{\Delta T}{T}(\vec{n}_1), \frac{\Delta T}{T}(\vec{n}_2) \right\rangle \tag{2}$$

Here, $\vec{n}_1$ and $\vec{n}_2$ are unit vectors corresponding to two directions of observation that subtend an angle θ, such that $\vec{n}_1 \cdot \vec{n}_2 = \cos\theta$. This expression can be generalised to other physical quantities and to correlations across N points. Contemporary analyses have explored two- and three-point correlations for the Stokes parameters I, Q, and U (see [4, 5]).

With the considerable increase in angular resolution and measurement precision, the latest temperature (Stokes I) and polarisation (Stokes Q and U) maps released by the Planck team contain $12 \times (N_{\text{side}})^2 = 50{,}331{,}648$ pixels. The parameter N_{side} is commonly used to characterise the resolution of a CMB map following the HEALPix pixelisation scheme.

Owing to the high computational cost associated with statistical analyses on such high-resolution data, these studies typically employ degraded maps rather than the full-resolution versions. The

[1] rdale@umh.es

standard degradation techniques adopted in the WMAP and Planck missions are twofold: the pixel-averaging method and spherical harmonic decomposition.

The first method averages the values of multiple adjacent pixels in the original map, combining them into a single pixel in a lower-resolution version. The magnitude assigned to the new pixel corresponds to the mean of the values from the contributing high-resolution pixels. The second method is based on the mathematical property that any scalar field defined on the sphere can be expressed as a sum of spherical harmonics. The quantity of interest is expanded into spherical harmonics, with coefficients computed up to a multipole moment of $3N_{\text{side}} - 1$. This truncation is chosen as, for a given resolution N_{side}, the harmonic coefficients can be reliably recovered up to this limit—and, conversely, used to reconstruct the field at the same resolution.

In recent years, machine learning has emerged as a promising alternative approach. These methods are particularly advantageous when dealing with sparse or noisy data, as they allow integration with physical models. For instance, physics-informed neural networks can avoid the need for costly mesh generation. Deep learning has already been successfully applied in contexts such as visual physics learning, fluid flow reconstruction governed by the Navier–Stokes equations, and beyond. In astronomy and astrophysics, neural network architectures—coupled with both supervised and unsupervised learning techniques—have enabled comprehensive, multiscale analysis of complex datasets with unprecedented speed and accuracy. Applications include star cluster analysis, exoplanetary systems, galaxy mergers, sunspot evolution, and more.

Specifically in the context of CMB studies, neural networks have recently been utilised for signal reconstruction [6, 7], power spectrum estimation [8], reionisation optical depth inference, foreground cleaning [9], dipole modulation detection [10], and all-sky large-scale power spectrum analyses [11], among others.

In this paper, we present our work employing convolutional neural networks as an alternative tool for CMB map degradation. Our primary focus is on the latest full-resolution maps released by the ESA Planck mission, which pose significant challenges in terms of computational efficiency. This has prompted us to initially refine our methods using the lower-resolution WMAP maps, which contain approximately an order of magnitude fewer pixels (specifically, 3,145,728).

2 Methods

2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly well-suited to the analysis of data with a grid-like structure, such as images, time series, or spatially ordered signals. Originally developed for image recognition tasks, CNNs have since become a foundational architecture in machine learning due to their capacity to automatically learn hierarchical and spatially coherent features from raw input data in an efficient and scalable manner [12].

What distinguishes CNNs from traditional feedforward networks is their use of convolutional layers composed of trainable filters (or kernels) that scan across local regions of the input. This local connectivity exploits the spatial correlation inherent in many datasets and significantly reduces the number of trainable parameters compared to fully connected layers. Additionally, the sharing of weights across spatial locations introduces an inductive bias that enhances the ability of CNNs to detect translation-invariant features. These convolutional operations are typically followed by nonlinear activation functions and, optionally, pooling layers that perform downsampling to reduce dimensionality while preserving essential spatial information [13].

CNNs are trained via backpropagation and gradient descent-based optimisation algorithms. During training, the model's parameters (weights and biases) are updated to minimise a loss function that quantifies the discrepancy between predicted and true values. The weight update rule

takes the general form:

$$w := w - \eta \cdot \frac{\partial \mathcal{L}}{\partial w} \tag{3}$$

Here, w denotes the weight being updated, η is the learning rate (a small positive scalar controlling the update step size), $\mathcal{L}$ is the loss function, and $\frac{\partial \mathcal{L}}{\partial w}$ is the gradient of the loss with respect to the weight.

This update rule is iteratively applied to all trainable parameters of the network, allowing the model to progressively reduce its error and improve performance.

2.2 Data

The input data comprise full-sky theoretical CMB temperature maps generated at resolution level $r = 11$, which corresponds to a HEALPix grid with $N_{\text{side}} = 2^r = 2048$ and a total of 50,331,648 pixels.

The corresponding target outputs are degraded versions of these maps, computed at a coarser resolution $r = 6$ ($N_{\text{side}} = 64$), yielding 49,152 pixels. These degraded maps are produced using standard HEALPix degradation techniques and serve as ground-truth targets during the training phase.

The CNN is trained using a large number of paired $(\text{input}_{r=11}, \text{target}_{r=6})$ maps, where each pair represents a high-to-low resolution mapping. Once trained, the model is applied to a real Planck map at resolution $r = 11$ to predict its $r = 6$ counterpart. This final application step assesses the model's ability to generalise beyond synthetic training data.

2.3 Model Architecture and Training

The CNN architecture employed in this study consists of multiple one-dimensional convolutional layers, each designed to extract localised spatial features from the input signal. These convolutional layers apply trainable filters to overlapping windows across the flattened input map, enabling the detection of both fine and large-scale patterns.

Following the convolutional blocks, the intermediate feature maps are flattened into a one-dimensional vector and processed by fully connected (dense) layers. These layers are responsible for modelling global dependencies and enable the network to learn complex nonlinear mappings between the high-resolution input and the corresponding low-resolution output.

The final dense layer produces a vector of length equal to the number of pixels in the degraded map. A linear activation function is applied at this output stage.

The model is compiled using the Adam optimiser and trained to minimise the mean squared error (MSE) loss, defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4}$$

where y_i and $\hat{y}_i$ represent the true and predicted pixel values, respectively. Training is carried out using mini-batch stochastic gradient descent, with batch size, learning rate, and number of epochs tuned to ensure convergence while avoiding overfitting.

Acknowledgments

This work was partially supported by the Conselleria d'Educació, Universitats i Ocupació of the Generalitat Valenciana through the projects CIAICO/2022/252 and ASFAE/2022/005. M. D.

Guillen thanks the grant PID2022-136383NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. The authors gratefully acknowledge the computer resources at Artemisa and the technical support provided by the Instituto de Fisica Corpuscular, IFIC(CSIC-UV). Artemisa is cofunded by the European Union through the 2014-2020 ERDF Operative Programme of Comunitat Valenciana, project IDIFEDER/2018/048.

References

[1] NASA Astrophysics, Cosmic Background Explorer/FIRAS, `https://lambda.gsfc.nasa.gov/product/cobe/firas_products.html#cmbmap`, 2013.

[2] NASA/WMAP Science Team, The Wilkinson Microwave Anisotropy Probe (WMAP), `https://wmap.gsfc.nasa.gov/`, 2024.

[3] Planck collaboration, Planck mission (ESA), `https://www.cosmos.esa.int/web/planck`, 2024.

[4] Gjerløw, E., Eriksen, H. K., Banday, A. J., Górski, K. M., Lilje, P. B., The two- and three-point correlation functions of the polarised five-year WMAP sky maps. *Astrophys. J.*, 710(1):689–697, 2010.

[5] Planck Collaboration, Planck 2018 results. VII. Isotropy and statistics of the CMB. *Astron. Astrophys.*, 641(A7):1–61, 2020.

[6] Wang, G.-J., et al., Recovering the CMB Signal with Machine Learning. *Astrophys. J. Suppl. Ser.*, 260(13):1–20, 2022.

[7] Yan, Y.-P., Wang, G.-J., Li, S.-Y., Xia, J.-Q., Recovering Cosmic Microwave Background Polarization Signals with Machine Learning. *Astrophys. J.*, 947(29):1–7, 2023.

[8] Pal, S., Chanda, P., Saha, R., Estimation of the Full-sky Power Spectrum between Intermediate and Large Angular Scales from Partial-sky CMB Anisotropies Using an Artificial Neural Network. *Astrophys. J.*, 945(77):1–14, 2023.

[9] Petroff, M. A., Addison, G. E., Bennett, C. L., Weiland, J. L., Full-sky Cosmic Microwave Background Foreground Cleaning Using Machine Learning. *Astrophys. J.*, 903(104):1–8, 2020.

[10] Khan, M. I., Saha, R., Detection of Dipole Modulation in CMB Temperature Anisotropy Maps from WMAP and Planck Using Artificial Intelligence. *Astrophys. J.*, 947(47):1–16, 2015.

[11] Chanda, P., Saha, R., An Unbiased Estimator of the Full-sky CMB Angular Power Spectrum at Large Scales Using Neural Networks. *Mon. Not. R. Astron. Soc.*, 508(3):4600–4609, 2021.

[12] Schmidhuber, J., Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[13] LeCun, Y., Bengio, Y. and Hinton, G., Deep learning. *Nature*, 521(7553): 436–444, 2015.

An Analysis of Bell's Cosmic Inequalities: From COBE to Planck Data

R. Dale[♭],[1] and J. M. Gandía[♭]

(♭) Departament d'Estadística, Matemàtiques i Informàtica, Universitat Miguel Hernández d'Elx
Avinguda de la Universitat d'Elx, s/n, 03202 Elx, Alacant, Spain.

1 Introduction

The cosmic microwave background (CMB) radiation has undeniably played a pivotal role in shaping our understanding of the origin, evolution, and potential fate of the universe since its discovery by Arno Penzias and Robert Wilson in 1965. Among the numerous observational and measurement systems for the CMB, the satellite missions COBE (1989), WMAP (2001), and Planck (2009) are particularly noteworthy. Each successive mission has refined the angular resolution and precision of CMB measurements, providing an extensive dataset that forms the foundation of the present study.

A fundamental question arises as to whether the data obtained from these missions can help discern the classical or quantum nature of primordial density fluctuations within the cosmic inflationary model. To explore this, we begin by assuming a highly entangled universe at the end of the inflationary period—an assumption contingent upon the existence of inflation as an explanation for specific observed cosmic phenomena. If this assumption holds, it may be possible to construct an analogue of Bell's inequalities [1] by substituting spin measurements with measurements of cosmic density fluctuations. In the subsequent section, we introduce a more general class of inequalities known as the Clauser-Horne-Shimony-Holt (CHSH) inequalities [2] and derive new constraints, which we term 'cosmic CHSH inequalities'. These are obtained by replacing spin operators with their corresponding cosmic density operators, demonstrating that violations occur only when certain commutators are nonzero.

The primordial density fluctuations generated during the inflationary period have left an imprint on the CMB. While the CMB has been subject to various perturbations, both at the surface of last scattering (e.g. Sachs-Wolfe and Doppler effects) and along its journey to present-day observers (e.g. Rees-Sciama and Sunyaev-Zel'dovich effects), the impact of such perturbations is minimal on sufficiently large scales [3]. Within the framework of standard cosmology, where the universe post-inflation evolves according to classical physics, any violation of local realism by primordial density fluctuations at the end of inflation should be detectable through an analysis of the CMB fluctuation spectrum. To ensure that the local nature of primordial density fluctuations is preserved throughout cosmic evolution, we consider observation directions that are causally separated, as their separation increases over cosmic time.

[1] rdale@umh.es

2 The CHSH Inequalities

Consider a quantum system comprising two entangled spin-$\frac{1}{2}$ particles, prepared such that they propagate towards two spin measurement devices positioned at locations A and B, with corresponding measurement directions $\vec{x}$ and $\vec{y}$. We assume that these measurements occur in a space-like configuration, thereby preserving causality. The experiment is repeated multiple times: in each trial, $\vec{x}$ is randomly assigned one of the two possible values $\vec{a}$ or $\vec{a'}$, while $\vec{y}$ is independently assigned one of $\vec{b}$ or $\vec{b'}$.

Under the assumption of local realism, the following CHSH inequality can be derived:

$$|\langle C\rangle| = \big|\big\langle A(B+B')\big\rangle + \big\langle A'(B-B')\big\rangle\big| \leq 2, \tag{1}$$

where the operator C is defined as

$$C \equiv A(B+B') + A'(B-B'). \tag{2}$$

For clarity, the operators A and A' represent the measurement operators at location A along directions $\vec{a}$ and $\vec{a'}$, respectively, with a similar convention for B and B'. Using this definition, we obtain

$$C^2 = 4 - [A, A'][B, B'], \tag{3}$$

where the involution conditions $A^2 = A'^2 = I$ and $B^2 = B'^2 = I$ hold, and $[A, A']$ and $[B, B']$ denote the respective commutators.

Clearly, the CHSH inequality (1) can only be violated if both commutators $[A, A']$ and $[B, B']$ are nonzero.

2.1 A Cosmic CHSH Inequality

We now consider the quasi-isotropic CMB temperature distribution. For any direction $\hat{\boldsymbol{x}}$ on the sky, we can express the temperature as:

$$T(\hat{\boldsymbol{x}}) = T_0 + \delta T(\hat{\boldsymbol{x}}), \tag{4}$$

where T_0 is the mean uniform temperature, approximately 2.725 K, and $\delta T(\hat{\boldsymbol{x}})$ is the perturbation dependent on $\hat{\boldsymbol{x}}$. Following Ref. [4], we define the dichotomised values of $\delta T(\hat{\boldsymbol{x}})$ using the function F, such that $F(\hat{\boldsymbol{x}}) = +1$ (or -1) if $\delta T(\hat{\boldsymbol{x}})$ is positive (or negative). A similar definition applies for the sky direction $\hat{\boldsymbol{y}}$. To derive a cosmic CHSH inequality, we invoke the hypothesis of local realism. This assumption implies the existence of hidden variables λ, which predate the measurements, such that the dichotomised CMB anisotropy values (± 1) are well-defined functions of λ and the corresponding measurement directions. Specifically,

$$\delta T_A \equiv \delta T_A(\hat{\boldsymbol{x}}, \lambda) = \pm 1, \quad \delta T_B \equiv \delta T_B(\hat{\boldsymbol{y}}, \lambda) = \pm 1.$$

Local realism is satisfied when these quantities are independent of $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{x}}$, respectively.

Upon performing a pair of measurements, we repeat the experiment, without requiring the same initial state preparation, and continue this process. Each measurement yields either $+1$ or -1. In each trial, $\hat{\boldsymbol{x}}$ randomly assumes one of two possible values, $\hat{\boldsymbol{a}}$ or $\hat{\boldsymbol{a'}}$, with equal and independent probability. Similarly, $\hat{\boldsymbol{y}}$ takes the value $\hat{\boldsymbol{b}}$ or $\hat{\boldsymbol{b'}}$. We consider the expectation value $\langle \delta T_A \delta T_B \rangle_{(\hat{\boldsymbol{x}}\hat{\boldsymbol{y}})}$ and define the quantity:

$$C \equiv \delta T_A(\delta T_B + \delta T_{B'}) + \delta T_{A'}(\delta T_B - \delta T_{B'}).$$

As detailed in Ref. [4], the aforementioned concepts lead to the following inequalities, which we designate as cosmic CHSH inequalities:

$$|\langle C\rangle| \equiv \left|\left\langle F(\hat{\boldsymbol{a}})F(\hat{\boldsymbol{b}}),\alpha_1\right\rangle + \left\langle F(\hat{\boldsymbol{a}})F(\hat{\boldsymbol{b}'}),\alpha_2\right\rangle + \left\langle F(\hat{\boldsymbol{a}'})F(\hat{\boldsymbol{b}}),\alpha_3\right\rangle - \left\langle F(\hat{\boldsymbol{a}'})F(\hat{\boldsymbol{b}'}),\alpha_4\right\rangle\right| \leq 2. \tag{5}$$

To compute the expectation value above, the following considerations are necessary:

(a) Let $\alpha \in [0,\pi]$ be the angle between $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$, such that $\cos\alpha \equiv \hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}$. We denote by $\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle$ the mean value of the product $F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}})$ over all direction pairs forming a constant angle α. For notational brevity, we use $\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle \equiv \langle\alpha\rangle$.

(b) The expectation value $\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle$ represents the mean value of the product $F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}})$ over all direction pairs forming a constant angle α.

(c) Given these four expectation values, there exist (see Ref. [4]) four unit vectors, $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{a}'}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{b}'}$, such that $\hat{\boldsymbol{a}}\cdot\hat{\boldsymbol{b}} = \cos\alpha_1$, $\hat{\boldsymbol{a}}\cdot\hat{\boldsymbol{b}'} = \cos\alpha_2$, $\hat{\boldsymbol{a}'}\cdot\hat{\boldsymbol{b}} = \cos\alpha_3$, and $\hat{\boldsymbol{a}'}\cdot\hat{\boldsymbol{b}'} = \cos\alpha_4$.

(d) As demonstrated in Ref. [4], the set of these four angles $\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\}$ is such that there exists an angle β satisfying the following inequalities:

$$|\cos\alpha_3 - \cos\beta\cos\alpha_1| \leq \sin\beta\sin\alpha_1, \tag{6a}$$

and

$$|\cos\alpha_4 - \cos\beta\cos\alpha_2| \leq \sin\beta\sin\alpha_2. \tag{6b}$$

(e) In fact, given a quadruple of angles α_i, there exists, modulo a generic rotation R, a two-parameter quadruple of unit vectors $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{a}'}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{b}'}$ satisfying the required angular conditions.

Each expectation value $\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle \equiv \langle\alpha\rangle$ is obtained via the expression:

$$\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle = P(+,+,\alpha) + P(-,-,\alpha) - P(+,-,\alpha) - P(-,+,\alpha), \tag{7}$$

where $P(X,Y,\alpha)$ denotes the probability of obtaining the result $X(\pm1)$ for $F(\hat{\boldsymbol{x}})$ and $Y(\pm1)$ for $F(\hat{\boldsymbol{y}})$. To calculate these probabilities, considering $\varepsilon_a(\delta T)$ as the absolute error in the quantities δT, we introduce the probabilities $p_\pm(\hat{\boldsymbol{x}})$ defined below:

$$p_+(\hat{\boldsymbol{x}}) \equiv \frac{\varepsilon_+}{2\varepsilon_a(\delta T)}, \tag{8}$$

and

$$p_-(\hat{\boldsymbol{x}}) \equiv \frac{2\varepsilon_a(\delta T) - \varepsilon_+}{2\varepsilon_a(\delta T)}, \tag{9}$$

where, for simplicity, the $\hat{\boldsymbol{x}}$ dependence in δT is omitted, and ε_+ represents the positive section of the interval $[\delta T - \varepsilon_a(\delta T), \delta T + \varepsilon_a(\delta T)]$. We then obtain the probabilities $p_{XY}(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}) = p_X(\hat{\boldsymbol{x}})\cdot p_Y(\hat{\boldsymbol{y}})$, representing the probability of obtaining a XY value (i.e., $++$, $+-$, $-+$, and $--$). From these quantities $p_{XY}(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$, we compute the probabilities $P(X,Y,\alpha)$ using the expression:

$$P(X,Y,\alpha) = \sum_{i=1}^{N_\alpha} \frac{p_{XY}(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})_i}{N_\alpha}. \tag{10}$$

Where N_α represents the total number of pairs of vectors $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ subtending a given angle α. Finally, we calculate $\langle F(\hat{\boldsymbol{x}})F(\hat{\boldsymbol{y}}),\alpha\rangle$ using expression (5).

3 Computing and Main Results

Before presenting the numerical results, it is helpful to summarise some of the preparatory tasks that enabled us to process the data appropriately.

While the treatment of COBE [5] data proved relatively straightforward—requiring minimal preprocessing—the same cannot be said for the WMAP and Planck datasets [6, 7], which demanded a more involved computational approach.

Recall that our primary aim is to present the numerical results for the quantities $\langle F(R\hat{\boldsymbol{x}})F(R\hat{\boldsymbol{y}}), \alpha\rangle \equiv \langle\alpha\rangle$, as defined in the preceding section. These quantities constitute the core ingredients required for evaluating the inequalities expressed in Eq. (5).

3.1 Data Preprocessing

For the COBE mission, the FITS (Flexible Image Transport System) file contains a total of 6067 records, each corresponding to a distinct direction in the sky. These records include, among other parameters, the measured temperature and its associated uncertainty. From this information, a table is constructed that associates each sky direction—specified in galactic coordinates (longitude θ and latitude φ)—with its respective temperature and uncertainty.

By combining these directions in pairs, we generate a total of 18,401,211 unique angular separations.

Let (θ_i, φ_i) and (θ_j, φ_j) denote the galactic coordinates of two unit vectors, $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{y}}_j$, respectively. The angle α_{ij} subtended between them, where $0 < \alpha_{ij} \leq \pi$, is computed via the standard spherical coordinate relation:

$$\alpha_{ij} = \arccos\left[\cos\varphi_i \cos\varphi_j \cos\left(\theta_j - \theta_i\right) + \sin\varphi_i \sin\varphi_j\right]. \tag{11}$$

To each such realisation α_{ij}, two temperatures T_i and T_j are assigned, corresponding to the respective directions $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{y}}_j$.

Considering the above framework—along with various additional requirements such as ensuring statistical robustness of the sample, confirming the causal disconnection between the two sky directions involved, and other secondary technical considerations—we are now in a position to proceed with the calculation, using COBE data, of the fundamental quantities entering the inequalities (5).

In the cases of WMAP and Planck, whose sky maps contain on the order of 10^6 and 10^7 directions, respectively, the direct computation of the quantities $\langle\alpha\rangle$ becomes prohibitive in terms of both computational resources and processing time. To overcome this, the original maps are pre-processed by degrading their resolution, thereby reducing the number of points on the sphere and making the calculations tractable. For WMAP, the original maps—provided at a resolution level $r = 9$ ($N_{\text{side}} = 2^9 = 512$, corresponding to $12 \times 512^2 = 3,145,728$ pixels per map)—were downgraded to a coarser resolution of $r = 6$, yielding 49,152 pixels per map. Similarly, for Planck, the original maps at resolution $r = 11$ (50,331,648 pixels per map) were likewise degraded to the same lower resolution level, $r = 6$. Both the WMAP and Planck datasets include an additional component known as a *mask*, which identifies whether a given pixel should be considered in the analysis. Typically, this exclusion is due to contamination or systematic effects that render the corresponding measurements unreliable or biased.

The standard approach to managing this issue involves degrading the original high-resolution maps to produce equivalent lower-resolution versions, thereby reducing the computational load while preserving the essential structure. These degraded maps retain the HEALPix format. Two conventional techniques exist for this purpose: one based on averaging neighbouring pixels and another employing a spherical harmonic decomposition. Although the former method is more commonly used for downgrading Stokes I (intensity) maps, we have generated degraded maps using both techniques for completeness.

Lastly, to enable a consistent comparison across the three satellite missions, the monopole and dipole components—previously subtracted in the publicly available WMAP and Planck maps—must be reinstated. These components are essential for ensuring that the statistical analysis, including the computation of the quantities $\langle\alpha\rangle$, is performed under equivalent conditions for all datasets.

3.2 Results with COBE Data

The expected values $\langle F(\hat{x})\, F(\hat{y}), \alpha\rangle$ are shown in Fig. 1. The resulting curve $\langle\alpha\rangle$ closely follows a cosine-like behaviour and is well approximated by the function $0.6\cos(\alpha)$. For comparison, both this scaled cosine and the standard cosine function are plotted in the figure.

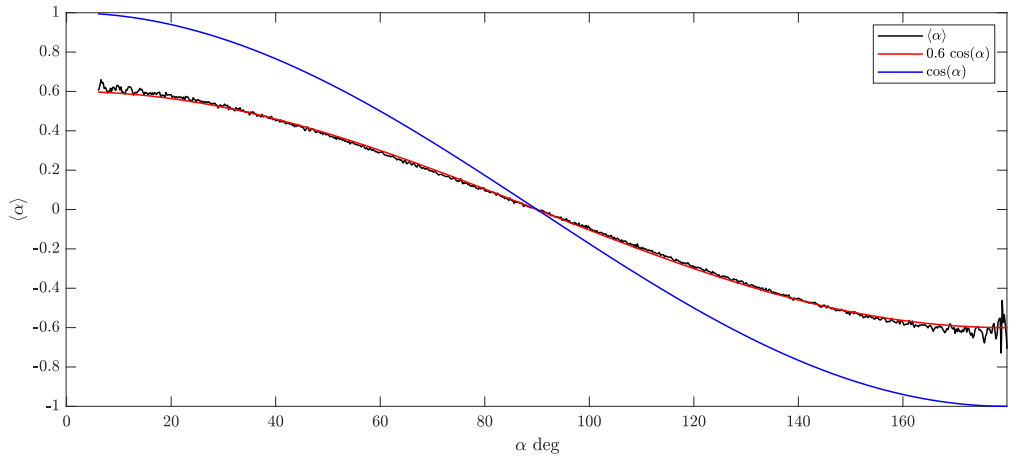


Figure 1: Expected values $\langle\alpha\rangle \equiv \langle F(\hat{x})\, F(\hat{y}), \alpha\rangle$ as a function of angle α. The empirical curve is well fitted by the function $0.6\cos(\alpha)$ (solid red line), while the solid blue line corresponds to $\cos(\alpha)$.

To identify potential violations of the CHSH inequality, a scanning algorithm was employed to search for four angular values that maximise $|\langle C\rangle|$, under the constraints defined in Eqs. (6a) and (6b). The maximum value obtained was $|\langle C\rangle| = 1.71 \pm 0.03$, corresponding to the angles $\{\alpha_1 = 125^\circ, \alpha_2 = 135^\circ, \alpha_3 = 139^\circ, \alpha_4 = 39.2^\circ\}$. This confirms that no violation of the CHSH inequality (5) is present in the COBE data.

3.3 Recent Results: Summary of WMAP and Planck Outcomes

Following our analysis of the COBE FIRAS dataset, we extend the investigation by incorporating more advanced data from the WMAP and Planck missions. It is worth highlighting that, while COBE FIRAS provided an angular resolution of approximately 7 degrees, WMAP and Planck achieved significant improvements—reaching resolutions of about 0.3 degrees and 5 to 10 arcminutes, respectively. In addition to this enhanced resolution, both missions offered superior measurement precision. Notably, whereas COBE FIRAS measured absolute temperature values, WMAP and Planck recorded temperature differences in the CMB between pairs of sky points.

The calculation of the expected value curve—previously presented for COBE FIRAS (Fig. 1)—has been replicated using data from both the WMAP and Planck missions. In each case, the analysis was carried out using degraded maps at a resolution of $r = 6$, employing both the neighbour-averaging method and spherical harmonic expansion. For WMAP, the Q, V, and W band maps were processed independently for each of these degradation techniques. In contrast, for Planck,

maps from all four standard component separation pipelines—COMMANDER, NILC, SEVEM, and SMICA—were utilised.

Among all the curves obtained, we present below those that yield the lowest maximum values of $|\langle C \rangle|$. These are illustrated in the two-panel figure that follows. The left panel displays the resulting curve derived from WMAP Q-band intensity data, based on maps degraded via spherical harmonic expansion. The right panel, in turn, shows the corresponding curve obtained from Planck data processed with the COMMANDER pipeline, using maps downgraded through the neighbour-averaging method.

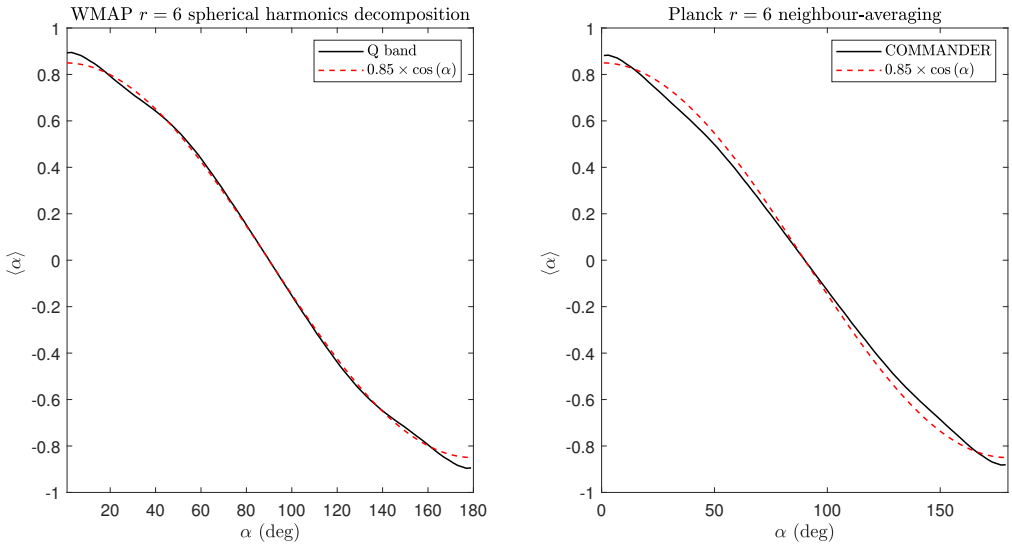


Figure 2: This two-panel figure displays the expected value curves $\langle \alpha \rangle \equiv \langle F(\hat{x}) F(\hat{y}), \alpha \rangle$ as a function of the angle α, derived from the WMAP and Planck datasets (black solid lines in the left and right panels, respectively). As in the case of COBE, we have also included the best-fitting curve of the form $A\cos(\alpha)$ (dashed red line) for each case, providing an approximate representation of the observed $\langle \alpha \rangle$ behaviour.

From the obtained expected value curves, we employed a scanning algorithm to identify four angular values, α, that maximise the quantity $|\langle C \rangle|$. Table 1 presents the three sets of four angles which, in each respective case, yield the highest values of $|\langle C \rangle|$.

Table 1: This table presents three representative four-angle sets (in degrees) for each satellite—WMAP and Planck. For each set, the corresponding maximum values of $|\langle C \rangle|$ obtained via the scanning procedure are listed, along with their associated absolute errors.

Satellite	Four-angle set				$\lvert\langle C \rangle\rvert$	Absolute Error
	α_1	α_2	α_3	α_4		
WMAP Q Band	133	135	137	45	2.4248	0.0009
	133	135	139	47	2.4234	0.0009
	133	137	139	49	2.4208	0.0009
Planck COMMANDER	133	135	137	45	2.1960	0.0009
	133	135	139	47	2.1950	0.0009
	43	45	47	135	2.1949	0.0009

In light of the table above, it is evident that while the COBE FIRAS data do not yield any acceptable four-angle combinations that violate the cosmic Bell inequalities (5), the same cannot be said for the WMAP and Planck datasets. In both of these latter cases, violations of the inequalities are observed. The most conservative scenario—i.e., the smallest deviation from the inequality bounds—occurs with the Planck data processed via the COMMANDER pipeline, where a violation close to 9% is found. In contrast, the analysis of the WMAP Q-band data reveals a more significant departure, with the inequality exceeded by nearly 20%.

4 Conclusions

Through an exhaustive statistical and numerical analysis of multiple datasets—of which we have presented a representative summary—it has become evident that, although initial results did not indicate any violation of the inequalities (5), subsequent investigations yielded a markedly different outcome. Without doubt, the utilisation of data acquired by instruments offering significantly higher angular resolution and measurement precision was instrumental in revealing this result.

Using WMAP and Planck data, we have presented results demonstrating that—even under the most conservative scenarios—it is possible to identify four causally disconnected directions in the sky from which (based on the Stokes I data provided by these satellites) values of $|\langle C\rangle|$ can be obtained that lie between 2 (the quantum threshold for violation) and $2\sqrt{2}$ (the Cirel'son bound).

Naturally, one might question the robustness of these results, noting that the analysis was performed on degraded maps rather than on raw, unaltered measurements. However, we have also conducted consistency checks using the original resolution data, and we indicate at the end of this section where further details on this validation may be found.

Finally, since both WMAP and Planck missions also provide polarisation data (Stokes Q and U), our ongoing efforts are directed towards extending this analysis to include these additional observables. This involves not only the derivation of new inequalities tailored to such quantities, but also the development of appropriate data processing techniques. In parallel, we are re-evaluating the implications of our findings within different theoretical frameworks.

A comprehensive and detailed account of the methods, computations, and results is available in [8].

Acknowledgments

This work was partially supported by the Conselleria d'Educació, Universitats i Ocupació of the Generalitat Valenciana through the projects CIAICO/2022/252 and ASFAE/2022/005. We are also grateful to our colleagues J.A. Morales-Lladosa and R. Lapiedra for their invaluable support and insightful contributions.

References

[1] Bell, J. S., On the Einstein Podolsky Rosen paradox *Physica*, 1(3):195–200, 1964.

[2] Clauser, J. E., Horne, M. A., Shimony, A., and Holt, R. A., Proposed Experiment to Test Local Hidden-Variable Theories *Phys. Rev. Lett.*, 23(15):880–884, 1969.

[3] Peacock, J. A., Cosmological Physics. Cambridge, Cambridge University Press, 1998.

[4] Dale, R., Lapiedra, R., Morales-Lladosa, J. A., Cosmic primordial density fluctuations and Bell inequalities *Phys. Rev. D*, 107(2):023506-1–023506-12, 2023.

[5] NASA Astrophysics, Cosmic Background Explorer/FIRAS, https://lambda.gsfc.nasa.gov/product/cobe/firas_products.html#cmbmap, 2013.

[6] NASA/WMAP Science Team, The Wilkinson Microwave Anisotropy Probe (WMAP), https://wmap.gsfc.nasa.gov/, 2024.

[7] Planck collaboration, Planck mission (ESA), https://www.cosmos.esa.int/web/planck, 2024.

[8] Dale, R., Gandía J. M., Morales-Lladosa, J. A., Lapiedra, R., Violation of Bell inequalities from Cosmic Microwave Background data arXiv:2502.13846.

Computational Tools in Astrophysics

Màrius Josep Fullana i Alfonso[♭],[1] Pau Amaro Seoane[♭]
and Josep Vicent Arnau i Córdoba[♭]

(♭) I.U. de Matemàtica Multidisciplinària, Universitat Politècnica de València
Camí de Vera s/n, 46022, València, Spain.

1 Introduction

This paper summarizes numerical approaches for solving ordinary differential equation (ODE) systems in cosmological and astrophysical contexts. We focus on adaptive step-size methods, particularly Runge-Kutta-Fehlberg algorithms, for initial value problems. The discussion emphasizes numerical stability, error control, and variable rescaling approaches applicable to diverse problems including gravitational perturbations, Cosmic Microwave Background (CMB) anisotropies, and N-body computations. Key aspects include avoiding catastrophic cancellation, objective function optimization, and local error resizing. The methods were demonstrated through applications to Lemaitre-Tolman-Bondi (LTB) spacetime solutions and other cosmological models [1] and [2].

2 Theoretical Framework

Numerical solutions of differential equations are essential in modern Cosmology and Astrophysics where analytical solutions are often unavailable. In our presentation we emphatize numerical techniques for solving systems of ODEs that arise in various physical contexts, focusing on adaptive step-size methods and their optimization.

We discuss general approaches applicable to:

- Gravitational perturbations in cosmological models
- CMB anisotropy calculations
- N-body systems in Cosmology
- Relativistic Positioning Systems (RPS)
- N-body systems in Astrophysics

The methods are particularly relevant for problems requiring high precision over wide dynamic ranges, such as calculating temperature contrasts of order 10^{-5} in CMB studies [3] and [4].

[1] mfullana@mat.upv.es

3 Numerical Implementation

Some numerical technics are presented (for a more detailed explanation see [1] and [2]):

3.1 Adaptive Step-size Control

For initial value problems, we employ Runge-Kutta-Fehlberg (RKF45) methods with adaptive step-size control [5]. The Cash-Karp variant [6] provides flexibility in order selection, crucial for problems with rapidly varying gradients. Key features include:

$$h_{new} = h \left(\frac{\text{tol}}{\text{error}} \right)^{1/(p+1)} \tag{1}$$

where h is step size, tol the tolerance, error the estimated local error, and p the method order.

3.2 Numerical Stability

Maintaining numerical stability requires:

- Avoiding catastrophic cancellation in subtractions
- Careful handling of extremal points
- Variable rescaling to maintain precision

For derivatives near zero, we use logarithmic differentiation [2]:

$$f'(x) \approx \frac{f(x)}{h} \log \left| \frac{f(x+h)}{f(x)} \right| \tag{2}$$

4 Results and Discussion

We now briefly describe different numerical problems in Cosmology and Astrophysics which we have dealt with; and some others we are treating at present, applying our previous experience.

4.1 LTB Spacetime Solutions

The LTB metric describes spherically symmetric pressureless perturbations in an expanding universe. Numerical solution requires:

$$ds^2 = \frac{1}{\Phi^2} \left(\frac{\partial R}{\partial M} \right)^2 dM^2 + R^2 d\Omega^2 - dt^2 \tag{3}$$

where $\Phi^2 = 1 + f(M)$ and $f(M) > -1$ [7]. The Christoffel symbols must be computed with high precision to maintain stability in geodesic integration.

4.2 CMB Anisotropies

For CMB anisotropy calculations, we solve null geodesic equations:

$$\frac{dK^\alpha}{d\lambda} = -\Gamma^\alpha_{\beta\gamma} K^\beta K^\gamma \tag{4}$$

with initial conditions at the observer and integrated backward in time. The temperature contrast is:

$$\delta_B = \frac{a(t_0)K_0^4}{a(t_D)K_D^4} - 1 \tag{5}$$

requiring relative precision better than 10^{-10} for meaningful results [7] and [8].

4.3 N-body systems in Cosmology

For more than twenty years, we have documented the progress in our CMB anisotropy calculations using N-body codes. The most precise and high-resolution codes we employed were those from the Hydra Consortium, all designed by its members. Our work involved: 1) Baryon-free codes. 1.a) Sequential versions. 1.b) Parallel versions. With these, we computed both the weak lensing and Rees-Sciama effects on the CMB angular power spectrum.

Our numerical approach yielded a stronger lensing signal compared to prior methods. By computing CMB anisotropies at each simulation step, we minimized interpolations and approximations, potentially explaining the excess power in our lensing results. The enhanced resolution played an important role. For a more detalied description, see for instance [9] and [10]

4.4 RPS

For over two decades, our research group has also focused on developing Relativistic Positioning Systems (RPS). A novel algorithm to calculate positioning errors was designed, which has been refined over time to enhance its accuracy. While our simulations have primarily used the ESA Galileo satellite constellation as a reference, the methodology is adaptable to other satellite systems at varying altitudes. We used numerical techniques to improve the description of the satellite orbits and to reduce the positioning errors associated with such computations. An accurate description of such algorithms can be found in [11], [12] and [10].

4.5 N-body systems in Astrophysics

Gravitational Wave (GW) sources-such as Extreme-Mass-Ratio Inspirals (EMRIs), where a compact object is captured by a supermassive black hole have traditionally been studied in Relativistic Astrophysics using post-Newtonian (PN) methods (e.g., [14] and [15]). However, this approximation breaks down beyond the two-body problem. In reality, gravitational wave emitting binaries are frequently subject to external perturbations. First-order (PN1) corrections accounting for these effects can be derived using cross terms in the post-Newtonian expansion [16] and [17].

Nowadays, we are working in this line of research. In our presentation we will discuss some interesting questions of the numerical algorithms we are using and the improvements we are making. Some interesting implications could be obtained for the EMRI detection and also for the description of GW. Our approach involves numerically computing PN1 corrections to Newtonian orbits, including cross terms, using an N-body code (see [18] for implementation details). We simulate astrophysical objects orbiting a central Super Massive Black Hole (SMH), demonstrating that the code –though originally designed for Keplerian motion– remains stable with PN1 adjustments. However, computational costs escalate significantly with additional stellar bodies, requiring alternative methods. One promising candidate is the AR-CHAIN code [19], which inherently includes post-Newtonian terms. This code should be studied in a near future.

5 Conclusions

The numerical techniques presented provide robust solutions for diverse cosmological and astrophysical problems. Key aspects include:

- Adaptive step-size control with error monitoring
- Variable rescaling for stability
- Careful handling of numerical derivatives
- Comprehensive validation tests

These methods have proven effective for LTB space-time solutions and CMB anisotropy calculations. They show promise for other applications including N-body simulations. Nowadays, we are extending these algorithms to additional Astrohysical scenarios and further optimize computational efficiency.

At present, we are applying our numerical knowledge to the computation of PN cross-terms in EMRIs evolution. Some of such numerical approaches are presented in the 27 Mathematical Modelling for Engineering & Human Behaviour 2025 Meeting. Our results seem to demonstrate that PN cross-terms can significantly impact the gravitational wave phase evolution of N-body systems. Our first computations highlight the need for more complete relativistic treatments in N-body simulations, particularly for LISA source modeling. Future work should extend these calculations to include higher-order PN terms and dissipative effects.

Acknowledgments

This work was partially supported by Generalitat Valenciana Project CIAICO/2022/252.

References

[1] Arnau i Córdoba, J.V., Fullana i Alfonso, M.J., Resolution of Initial Value Problems of Ordinary Differential Equations Systems *Mathematics*, 10(4), 593 (1-27), 2022. *https://doi.org/10.3390/math10040593*

[2] Amat, S., Aràndiga, F., Arnau, J.V.; Donat Beneito, R.M., Mulet Mestre, P., Peris Sancho, R. Aproximació numèrica, Servei de Publicacions de la Universitat de València, València, Spain. 2002.

[3] Fullana i Alfonso, Màrius Josep. Anisotropies de la radiació de fons de microones produïdes per inhomogeneïtats cosmològiques no lineals, Servei de Publicacions de la Universitat de València, València, Spain. 1996.

[4] Fullana, M.J., Arnau, J.V., Sáez, D. Looking for the imprints of nonlinear structures on the cosmic microwave background. *Vistas Astron.* 41, 467–492, 1997.

[5] Fehlberg, E., *Low-Order Classical Runge-Kutta Formulas with Stepsize Control and Their Application to Some Heat Transfer Problems*; National Aeronautics and Space Administration: Washington, USA, 1969; Volume 315.

[6] Cash, J.R., Karp, A.H., A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides. *ACM Trans. Math. Softw. TOMS* 16: 201–222, 1990.

[7] Arnau, J.V., Fullana, M.J., Monreal, L., Sáez, D., On the microwave background anisotropies produced by nonlinear voids. *Astrophys. J.* 402: 359–368, 1993.

[8] Fullana, M.J.; Sáez, D.; Arnau, J.V., On the microwave background anisotropy produced by great attractor-like structures. *Astrophys. J. Suppl. Ser.* 94: 1–16, 1994.

[9] Fullana i Alfonso, M.J., Arnau, J.V., Thacker, R.J., Couchman, H. M. P., Sáez, D., Estimating Small Angular Scale Cosmic Microwave Background Anisotropy with High-Resolution N-Body Simulations: Weak Lensing *The Astrophysical Journal*, Volume 712, Number 1, 367-379, March 2010. *https://doi.org/10.1088/0004-637X/712/1/367*

[10] Fullana i Alfonso, M.J., Arnau i Córdoba, J.V., *Computational Tools in Cosmology. In 24 Mathematical Modelling for Engineering & Human Behaviour*. València, July 14th-16th, 2022. Edited by: I.U. de Matemàtica Multidisciplinària, Universitat Politècnica de València. J.R. Torregrosa, J.C. Cortés, J. A. Hervàs, A. Vidal-Ferràndiz and E. López-Navarro, 2022. *ISBN: 978-84-09-47037-2*

[11] Puchades Colmenero, N., Arnau i Córdoba, J.V., Fullana i Alfonso, M.J.. Relativistic positioning: Including the influence of the gravitational action of the sun and the moon and the earth's oblateness on galileo satellites. *Astrophys. Space Sci.* 366: 1–19, 2021.

[12] Fullana i Alfonso, M.J., Arnau i Córdoba, J.V., Puchades Colmenero, N., *Some advances in Relativistic Positioning Systems. In 23 Mathematical Modelling for Engineering & Human Behaviour*. València, July 14th-16th, 2021. Edited by: I.U. de Matemàtica Multidisciplinària, Universitat Politècnica de València. J.R. Torregrosa, J.C. Cortés, J. A. Hervàs, A. Vidal-Ferràndiz and E. López-Navarro, 2021. *ISBN: 978-84-09-36287-5*

[13] Fullana i Alfonso, M.J., Puchades Colmenero, N., Arnau i Córdoba, J.V., *Some Computational Tools in RPS. In 25 Mathematical Modelling for Engineering & Human Behaviour*. València, July 11th-14th, 2023. Edited by: I.U. de Matemàtica Multidisciplinària, Universitat Politècnica de València. J.R. Torregrosa, J.C. Cortés, A. Vidal-Ferràndiz, 2023. *ISBN: 978-84-09-57681-4*

[14] Kupi, G. and Amaro-Seoane, P., Spurzem, R., Dynamics of compact object clusters: a post-Newtonian study *Monthly Notices of the Royal Astronomical Society: Letters*, (371):L45–L49, Issue 1, September 2006, https://doi.org/10.1111/j.1745-3933.2006.00205.x

[15] Amaro-Seoane, P., Relativistic dynamics and extreme mass ratio inspirals, Living Rev Relativ (2018) 21:4, https://doi.org/10.1007/s41114-018-0013-8

[16] Will, C. M., Incorporating post-Newtonian effects in N-body dynamics *Physical Review D*, (89): 044043-1–044043-14, 26 February 2014, https://doi.org/10.1103/PhysRevD.89.044043

[17] Fullana i Alfonso, M.J., Amaro Seoane, P., Arnau i Córdoba, J.V., *On the study of PN1 cross terms in the evolution of EMRIs. In 26 Mathematical Modelling for Engineering & Human Behaviour*. València, July 10th-12th, 2024. Edited by: I.U. de Matemàtica Multidisciplinària, Universitat Politècnica de València. A. Vidal-Ferràndiz and J.R. Torregrosa, 2024. *ISBN: 978-84-09-63274-9*

[18] Aarseth, S. J., From NBODY1 to NBODY6: The Growth of an Industry, *The Publications of the Astronomical Society of the Pacific*, (111): 1333-1346, November 1999, https://doi.org/10.1086/316455

[19] Mikkola, Seppo and Merritt, David, Implementing Few-Body Regularitation with Post-Newtonian Terms, The Astronomical Journal 135: 2398-2405, 2008. https://iopscience.iop.org/article/10.1088/0004-6256/135/6/2398

Extension of analytical regularization methods of the integration step of the elliptic two-body problem to the hyperbolic case

José Antonio López Ortí[♭],[1] María José Martínez Usó[♮] and Francisco José Marco Castillo[♭]

(♭) Departamento de Matemáticas, Universitat Jaume I de Castellón
Av. Sos Baynat s/n, Castellón, Spain.

(♮) Departamento de Matemática Aplicada, Universitat Politécnica de Valencia,
Camino Vera s/n, Valencia, Spain.

Abstract

One of the primary subjects in celestial mechanics is the study of the two-body problem, which has been extensively examined by numerous authors using both analytical and numerical methods.

In the case of elliptical orbits, numerical methods generally perform reliably. However, when dealing with very high eccentricities, these methods can encounter specific challenges, often due to a concentration of integration nodes near the apoapsis, while the periapsis region has very few. This uneven distribution of points does not align well with the dynamics of the system. To address these issues, techniques such as variable step methods and analytical step regularization can be employed.

This study concentrates on the latter scenario by implementing reparameterizations of the time variable. This approach aims to create a distribution of integration base points that better reflects the dynamics of the two-body problem, particularly by increasing the concentration of points in the periapsis region, where motion is most rapid. These techniques have been systematically explored within the context of elliptical motion using families of parameter-dependent time transformations.

Acknowledgments

This work has been partially supported by grant GACUJIMC/2024/20 of University Jaume I of Castellón..

[1] lopez@mat.uji.es

xIdeal, a *Mathematica* package for IDEAL characterizations and determinations

S. Mengual[1,*], A. García-Parrado Gómez-Lobo[2]

(1) Departament d'Astronomia i Astrofísica, Universitat de València,
E-46100 Burjassot, València, Spain.
(2) Departamento de Matemáticas, Universidad de Córdoba,
Campus de Rabanales, 14071, Córdoba, Spain.

Abstract

A longstanding problem in General Relativity is determining whether two explicit solutions to Einstein's equations represent the same spacetime, differing only by a coordinate transformation [1, Ch.9]. One approach to this question that has recently attracted attention is the so-called *IDEAL characterization* method. An IDEAL (Intrinsic, Deductive, Explicit and ALgorithmic) characterization consists of a set of tensorial equations covariantly constructed out of spacetime metric concomitants that are satisfied if, and only if, the given spacetime locally belongs to the desired class. This philosophy can also be employed to algorithmically determine expressions for intrinsic quantities and properties of the spacetime under consideration using only metric concomitants. We will refer to such algorithms as *IDEAL determinations*.

Among many other advantages, the algorithmic structure of the IDEAL approach makes it especially well-suited for implementation in a formal computational framework. An example of such a formal calculation program is *xAct*, a collection of *Mathematica* packages designed for tensor computer algebra [2]. In this talk, I will explain how some of these algorithms have been implemented on *xIdeal*, an *xAct* package that can be found in its current version at [3]. In particular, I will show the IDEAL characterization of the Schwarzschild spacetime, the IDEAL determination of the Petrov-Bel type of a given metric and the IDEAL determination of the multiple Debever null directions, among others.

References

[1] Stephani, H., Kramer, D., MacCallum, M., Hoenselares, C., Herlt, E., Exact solutions of Einstein's field equations. Cambridge, Cambridge University Press, 2009.

* Corresponding author salvador.mengual@uv.es

[2] Martín-García, J. M., xAct: Efficient tensor computer algebra for the Wolfram Language, http://www.xact.es/

[3] García-Parrado Gómez-Lobo, A., Mengual, S., xIdeal: IDEAL characterizations for xAct, https://github.com/wtbgagoa/xIdeal

Quantum Logistics Dynamics

Joan C. Micó[1,*]

(1) Insitut Universitari de Matemàtica Multidisciplinar. Universitat Politècnica de Valéncia
Camí de Vera s/n, ciutat de Valéncia, Valéncia (Spain).

1 Introduction

The translation of the quantum formalism from the Physical Systems to the Abstract Dynamical Systems (ADS) has been able to be done in [1]. The work [1] presents a first application to the ADS given by the quantum logistics dynamics. However, this application fails in some parts of its development, although the general formalism is well presented and developed. Taking these considerations into account, the objective of the present work is to provide correctly the application of the quantum formalism of ADS to the logistics dynamics application.

Section 2 presents the main findings of the Quantum Mechanics formalism for ADS, Section 3 provides its adaptation to the one-dimensional autonomous case, and Section 4 provides the correct development for the quantum logistics dynamics application.

2 Main results of the quantum formalism for ADS

The main results of the quantum formalism for ADS, stated in [1], are here presented. First, an ADS is defined as a coupled system of differential equations ($1 \leq k \leq n$):

$$\dot{q}_k(t) = f_k(t, \boldsymbol{q}) \tag{1}$$

In Eq. 1 t is the time variable and $\boldsymbol{q} = (q_1, q_2, \dots, q_n)$ represent the abstract dynamical variables. In many contexts they represent population dynamics, but they can also represent chemical species, economic variables evolution and so on. This is the reason to be called as "abstract".

As it is demonstrated in [1], the quantum version of Eq. 1 is provided by the following coupled system of differential equations ($1 \leq k \leq n$):

* Joan C. Micó jmico@mat.upv.es

$$\left.\begin{aligned} \dot{q}_k(t) &= f_k(t,\boldsymbol{q}) - \frac{x_k}{2m_k(t,\boldsymbol{q})} \\ \dot{x}_k(t) &= \frac{\partial V_q(t,\boldsymbol{q})}{\partial q_k} \end{aligned}\right\} \tag{2}$$

In Eq. 2, x_k $(1 \leq k \leq n)$ are auxiliary variables that provide the quantum correction of Eq. 1, such that the initial conditions are $q_k(0) = q_0^{(k)}$ (with known values) and $x_k(0) = 0$. The function $V_q(t,\boldsymbol{q})$ is the quantum potential:

$$V_q(t,\boldsymbol{q}) = -\frac{\sigma^2}{4}\sum_{j=1}^{n}\frac{\partial^2}{\partial q_j^2}\left(\frac{1}{m_j(t,\boldsymbol{q})}\right) - \frac{\sigma^2}{2A(t,\boldsymbol{q})}\sum_{j=1}^{n}\frac{\partial}{\partial q_j}\left(\frac{1}{m_j(t,\boldsymbol{q})}\frac{\partial A(t,\boldsymbol{q})}{\partial q_j}\right) \tag{3}$$

In Eq. 3, σ is the system Planck constant, particular (by hypothesis) of each dynamical system. The $m_j(t,\boldsymbol{q})$ $(1 \leq j \leq n)$ functions are the generalized masses, which hold the condition:

$$\frac{\partial\left(m_j(t,\boldsymbol{q})\cdot f_j(t,\boldsymbol{q})\right)}{\partial q_k} = \frac{\partial\left(m_k(t,\boldsymbol{q})\cdot f_k(t,\boldsymbol{q})\right)}{\partial q_j} \quad 1 \leq j,k \leq n \tag{4}$$

In addition, in Eq. 3, $A(t,\boldsymbol{q}) = |\Psi(t,\boldsymbol{q})|$, i.e., the modulus of the wave function $\Psi(t,\boldsymbol{q})$, which holds the Micó-Schrödinger equation:

$$i\sigma\frac{\partial\Psi(t,\boldsymbol{q})}{\partial t} =$$

$$= -\frac{1}{2}\frac{\sigma^2}{2}\left(\sum_{j=1}^{n}\frac{1}{m_j(t,\boldsymbol{q})}\frac{\partial^2\Psi(t,\boldsymbol{q})}{\partial q_j^2} + \sum_{j=1}^{n}\frac{\partial^2}{\partial q_j^2}\left(\frac{1}{m_j(t,\boldsymbol{q})}\Psi(t,\boldsymbol{q})\right)\right) -$$

$$-\frac{1}{2}\sum_{j=1}^{n} m_j(t,\boldsymbol{q})\cdot f_j^2(t,\boldsymbol{q})\cdot\Psi(t,\boldsymbol{q}) + \frac{1}{2}h(t,\boldsymbol{q})\cdot\Psi(t,\boldsymbol{q}) \tag{5}$$

In addition to the restrictions of Eq. 4, the $h(t,\boldsymbol{q})$ function of Eq. 5 must hold:

$$\frac{\partial h(t,\boldsymbol{q})}{\partial q_k} = -\frac{\partial\left(2m_k(t,\boldsymbol{q})\cdot f_k(t,\boldsymbol{q})\right)}{\partial t} \qquad 1 \leq k \leq n \tag{6}$$

The corresponding quantum Hamiltonian to Eq. 5 is given by:

$$H_q(t,\boldsymbol{q},\boldsymbol{p}) = H(t,\boldsymbol{q},\boldsymbol{p}) + V_q(t,\boldsymbol{q}) =$$

$$= \frac{1}{2}\sum_{j=1}^{n}\frac{p_j^2}{m_j(t,\boldsymbol{q})} - \frac{1}{2}\sum_{j=1}^{n} m_j(t,\boldsymbol{q})\cdot f_j^2(t,\boldsymbol{q}) + \frac{1}{2}h(t,\boldsymbol{q}) + V_q(t,\boldsymbol{q}) \tag{7}$$

Note in Eq. 7 that $H(t,\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\sum_{j=1}^{n}\frac{p_j^2}{m_j(t,\boldsymbol{q})} - \frac{1}{2}\sum_{j=1}^{n} m_j(t,\boldsymbol{q})\cdot f_j^2(t,\boldsymbol{q}) + \frac{1}{2}h(t,\boldsymbol{q})$ is the classical Hamiltonian, and that $V_q(t,\boldsymbol{q})$ is the quantum potential computed by Eq. 3. If the classical Hamiltonian is time-independent, then $H(t,\boldsymbol{q},\boldsymbol{p}) = H(\boldsymbol{q},\boldsymbol{p}) = E$, being E the conserved system energy.

Therefore, to solve the quantum version of the ADS (Eq. 2), the following steps must be followed, with the order indicated: (1) find out the $m_k(t,\boldsymbol{q})$ $(1 \le k \le n)$ generalized masses from Eq. 4; (2) compute the $h(t,\boldsymbol{q})$ function from Eq. 6; (3) solve the Micó-Schrödinger equation (Eq. 5); compute the quantum potential (Eq. 3); find out the quantum dynamics (or quantum trajectories) by Eq. 2. It must be emphasized that all this formalism has been developed from the generalized dynamics by Dirac [2], the hydrodynamic interpretation of quantum mechanics by Madelung [3], the non-local interpretation of the theory by Bohm and Hiley [4] and several previous works of the present author's work, cited in the work [1].

Note that the quantum dynamics given by Eq. 2 should be stochastic due to the stochastic nature of the quantum potential, which depends on the wave function modulus, which is also stochastic. In addition, it is expectable that, due to the boundary conditions, the wave function is a linear superposition of stationary states depending on several integers $\boldsymbol{m} = \{m_1, m_2, \ldots, m_r\}$, such that $\Psi_{\boldsymbol{m}}(t,\boldsymbol{q}) \in L^2(\mathbb{R}^n)$, associated to the different quantized system's amounts, as for instance, energy.

In addition, if the system of Eq. 1 is autonomous, i.e, $f_k(t,\boldsymbol{q}) = f_k(\boldsymbol{q})$ $(1 \le k \le n)$, the generalized masses can be also chosen as $m_k(t,\boldsymbol{q}) = m_k(\boldsymbol{q})$ $(1 \le k \le n)$, and $h(t,\boldsymbol{q}) = 0$ by Eq. 6. Therefore, the Micó-Schrödinger equation (Eq. 5) can be solved by the common change in the Schrödinger equation, $\Psi(t,\boldsymbol{q}) = \mathrm{e}^{-i\frac{E}{\sigma}t}\psi(\boldsymbol{q})$, which provides the time-independent Micó-Schrödinger equation:

$$E\cdot\psi(\mathbf{q}) = \\ = -\frac{1}{2}\frac{\sigma^2}{2}\left(\sum_{j=1}^{n}\frac{1}{m_j(\boldsymbol{q})}\frac{\partial^2\psi(\mathbf{q})}{\partial q_j^2} + \sum_{j=1}^{n}\frac{\partial^2}{\partial q_j^2}\left(\frac{1}{m_j(\boldsymbol{q})}\psi(\mathbf{q})\right)\right) - \frac{1}{2}\sum_{j=1}^{n} m_j(\boldsymbol{q})\cdot f_j^2(t,\boldsymbol{q})\cdot\psi(\mathbf{q}) \quad (8)$$

In Eq. 8, E is the conserved energy corresponding the second order classical Hamiltonian provided above. As a consequence, the wave function modulus $A(t,\boldsymbol{q}) = |\Psi(t,\boldsymbol{q})| = \left(\Psi^*(t,\boldsymbol{q})\cdot\Psi(t,\boldsymbol{q})\right)^{1/2} = |\psi(\mathbf{q})|$, i.e., $A(t,\boldsymbol{q}) = A(\boldsymbol{q}) = |\psi(\mathbf{q})|$, being $\psi(\mathbf{q})$ real and $|\psi(\mathbf{q})|$ its absolute value.

Finally, an important correction to the formalism presented in [1] must be here considered. If $t = 0$ is the initial time, the initial conditions $q_k(0) = q_k^{(0)}$ are known in Eq. 2. However, in

[1], the considered initial conditions $x_k(0) = 0$. However, they are inconsistent with the formalism. Note that the Hamilton equations for both the classical and quantum Hamiltonians of Eq. 7 provide (see Eqs. 34 and 50 of [1]), for $1 \leq k \leq n$: $m_k(t, \boldsymbol{q}) \cdot \dot{q}_k = \big(p_k + m_k(t, \boldsymbol{q}) \cdot f_k(t, \boldsymbol{q})\big)/2$, and Eq. 2 provides $m_k(t, \boldsymbol{q}) \cdot \dot{q}_k = m_k(t, \boldsymbol{q}) \cdot f_k(t, \boldsymbol{q}) - x_k/2$, therefore, $x_k = m_k(t, \boldsymbol{q}) \cdot f_k(t, \boldsymbol{q}) - p_k$. Then, for the initial time, $x_k(0) = m_k\big(0, \boldsymbol{q}(0)\big) \cdot f_k\big(0, \boldsymbol{q}(0)\big) - p_k(0)$. In addition, in the autonomous case, $p_k(0)$ can be related with the quantized energies.

3 The one-dimensional autonomous ADS

In this case, the only generalized mass can be taken as $m(t, q) = 1$ and $h(t, q) = 0$. Due to, the time-independent Micó-Schrödinger equation (Eq. 8) becomes:

$$\psi''(q) + \left(\frac{f^2(\mathrm{q})}{\sigma^2} + \frac{2E}{\sigma^2}\right)\psi(q) = 0 \tag{9}$$

In addition, the quantum potential (Eq. 3) is computed as:

$$V_q(q) = -\frac{\sigma^2}{2A(q)}\frac{d^2A(q)}{dq^2} \tag{10}$$

Then, the quantum dynamics (Eq. 2), considering the correction for the initial values of $x(t)$ of the last section (and Eq. 8 that relates the momentum variable and the energy as $p_0 = \sqrt{2E + f(q_0)}$), is written for this case as:

$$\left.\begin{aligned} \dot{q}(t) = f(q) - \frac{x}{2}\ ;\ q(0) = q_0 \\ \dot{x}(t) = V_q'(q)\ ;\ x(0) = f(q_0) - \sqrt{2E + f(q_0)} \end{aligned}\right\} \tag{11}$$

Consider now that the boundary conditions provide the wave function and the energy quantization depending on a set of *n* integers (do not confuse with the dynamical dimension of Eq. 1 or Eq. 2):

$$\Psi(t, q) = \sum_n \mathrm{e}^{-i\frac{E_n}{\sigma}t}\psi_n(q) \tag{12}$$

In Eq. 12, every stationary state represented by the wave function $\psi_n(q)$ must hold that is $L^2(\mathbb{R})$ and, in addition, it must obey Eq. 9. Therefore, for any stationary state the modulus is $A_n(t, q) = A_n(q) = |\psi_n(\mathrm{q})|$, and then, the quantum potential becomes $V_q^{(n)}(q) = -\frac{\sigma^2}{2|\psi_n(\mathbf{q})|}\frac{d^2|\psi_n(q)|}{dq^2}$. However, note that the derivative of the absolute value of a real function, such that $|\psi_n(\mathbf{q})|$, is computed as $\frac{d|\psi_n(\mathrm{q})|}{dq} = \frac{\psi_n(\mathrm{q})}{|\psi_n(\mathrm{q})|}\psi_n'(\mathrm{q})$. From this last result, from Eq. 9 and from that $|\psi_n(\mathrm{q})|^2 = \psi_n^2(\mathrm{q})$ it is easy to derive that:

$$V_q^{(n)}(q) = \frac{f^2(q)}{2} + E_n \quad ; \quad \frac{dV_q^{(n)}(q)}{dq} = f(q) \cdot f'(q) \tag{13}$$

Eq. 13 implies that, for any stationary state, $\Psi_n(t,q) = \mathrm{e}^{-i\frac{E_n}{\sigma}t}\psi_n(q)$, the wave function, through the quantum potential, has no influence on the dynamics given by Eq. 11. This amazing result for the one-dimensional case of the time-independent Schrödinger equation in physical systems has also been emphasized by other authors, such that Wyatt in [5] or Poirier in [6]. In fact, in addition of [6], the work [5] provides different approaches of different authors to overcome this problem, which is faced provisionally in the following section as a provisional deterministic approach (see Section 5).

4 The logistics quantum dynamics

The application case considered is the quantum dynamics for the logistic function, $f(q) = a \cdot q - b \cdot q^2$, with positive parameters *a*=0.01 and *b*=0.001. Note that the function has two critical points, a repulsor $q = 0$ and an attractor $q = c = a/b = 10.0$. The classical dynamics provided by Eq. 1 ($\dot{q}(t) = f(q) = a \cdot q - b \cdot q^2$) is showed in Figure 2 (below) for the initial conditions: $q(0) = q_0 = 2 < c = 10.0$ and $q(0) = q_0 = 18 > c = 10.0$; jointly the corresponding quantum dynamics.

To solve Eq. 9 for this case, the following mathematical transformations have been considered: (1) $\psi(q) = \mathrm{e}^{\theta(q)}\Omega(q)$; (2) $z = \mathrm{z}(q)$, then $w(z) = \Omega\big(q(z)\big)$. They provide by being considered in Eq. 9:

$$(\mathrm{z}')^2 w''(z) + (\mathrm{z}'' + 2\theta' \cdot \mathrm{z}')w'(z) + \left(\theta'' + (\theta')^2 + \frac{f^2(\mathrm{q})}{\sigma^2} + \frac{2E}{\sigma^2}\right)w(z) = 0 \qquad (14)$$

In [1], Eq. 14 derivation fails because the term multiplying to $w'(z)$ is presented as $\theta'(\mathrm{z}'' + 2 \cdot \mathrm{z}')$. Here, once carefully revised, Eq. 14 is the correct derivation. Then, the term multiplying to $w'(z)$ can be dropped as $\mathrm{z}'' + 2\theta' \cdot \mathrm{z}' = 0$, i.e., $\mathrm{z}'(q) = \mathrm{e}^{-2\theta(q)}$, or $z(q) = \int_0^q \mathrm{e}^{-2\theta(y)}\, dy$, due to Eq. 14 becomes $\mathrm{e}^{-4\theta}w''(z) + \left(\theta'' + (\theta')^2 + \frac{f^2(\mathrm{q})}{\sigma^2} + \frac{2E}{\sigma^2}\right)w(z) = 0$. In the beginning, it is considered that $\theta'' + (\theta')^2 + \frac{f^2(\mathrm{q})}{\sigma^2} + \frac{2E}{\sigma^2} = k^2\mathrm{e}^{-4\theta}$, where $k^2 \neq 0$. This last consideration provides $w''(z) + k^2 \cdot w(z) = 0$, i.e., $w(z) = C \cdot cos(k \cdot z) + D \cdot sin(k \cdot z)$, where C and D are arbitrary but real constants. Undoing both changes, the final result is $\psi(q) = \mathrm{e}^{\theta(q)}\left(C \cdot cos\left(k\int_0^q \mathrm{e}^{-2\theta(y)}\, dy\right) + D \cdot sin\left(k\int_0^q \mathrm{e}^{-2\theta(y)}\, dy\right)\right)$. However, about the critical points $q \cong 0, c$, $f(q) \cong 0$, and Eq. 9 becomes $\psi''(q) + \frac{2E}{\sigma^2}\psi(q) \cong 0$, with solution $\psi(q) = \bar{C} \cdot cos\left(\frac{\sqrt{2\mathrm{E}}}{\sigma}q\right) + \bar{D} \cdot sin\left(\frac{\sqrt{2\mathrm{E}}}{\sigma}q\right)$. On the other hand, also about the critical points $q \cong 0, c$, $\theta'' + (\theta')^2 + \frac{2E}{\sigma^2} \cong k^2\mathrm{e}^{-4\theta}$, which has the solution $\theta(q) \cong 0$ if $k^2 = \frac{2E}{\sigma^2}$, which also produces $\psi(q) \cong \mathrm{e}^0\left(C \cdot cos\left(k\int_0^q dy\right) + D \cdot sin\left(\int_0^q dy\right)\right) = C \cdot cos(k \cdot q) + D \cdot sin(k \cdot q)$. Then, the conclusion is that $k = \frac{\sqrt{2E}}{\sigma}$, $C = \bar{C}$ and $D = \bar{D}$. The result can be written as:

$$\psi(q) = e^{\theta(q)}\left(C \cdot cos\left(\frac{\sqrt{2E}}{\sigma}\int_0^q e^{-2\theta(y)}\,dy\right) + D \cdot sin\left(\frac{\sqrt{2E}}{\sigma}\int_0^q e^{-2\theta(y)}\,dy\right)\right) \quad (15)$$

With the associated boundary problem:

$$\left.\begin{array}{c} \theta''(q) + \left(\theta'(q)\right)^2 + \frac{f^2(q)}{\sigma^2} = \frac{2E}{\sigma^2}\left(e^{-4\theta(q)} - 1\right) \\ \theta(0) = \theta(c) = 0 \end{array}\right\} \quad (16)$$

The numerical solutions of Eq. 16, computed with the software MATHEMATICA 14.0, show that for any energy value $E > 0$, $\theta(q) \to -\infty$ as $q \to \pm\infty$, then $\psi(q) \to 0$ as $q \to \pm\infty$ (see below Figure 1, although the trend $\theta(q) \to -\infty$ is very slow), therefore $\psi(q)$ is a bounded function in all $\mathbb{R}$. In addition, the boundary conditions $\theta(0) = \theta(c) = 0$ guarantee that the solutions of the differential equation match those of the linear approximation about these boundary conditions and, consequently, they also guarantee that $k = \frac{\sqrt{2E}}{\sigma}$.

However, in [1] the differential equation of Eq. 16 fails because the term $e^{-4\theta(q)}$ of Eq. 16 appears as $\theta^2(q)$, which is a consequence of the first failure described above. However, in that paper, unlike the failure, the wave function also works as a bounded function in all $\mathbb{R}$.

The proposed boundary conditions on the wave function (Eq. 15) are: (1) $\psi(q) = 0$ for $q \leq 0$; (2) $\psi(c) = 0$. By its continuity in $q = 0$, $\psi(0) = 0$, which implies $C = 0$. Then, the second one implies (for a non-null solution, i.e., $D > 0$) that $sin\left(\frac{\sqrt{2E}}{\sigma}\int_0^c e^{-2\theta(y)}\,dy\right) = 0$, i.e., $\frac{\sqrt{2E}}{\sigma}\int_0^c e^{-2\theta(y)}\,dy = \pi \cdot n$, with $n \in \mathbb{Z}_{>0}$ (due to the left hand side is positive). The solutions of this equation are non-trivial for the energy values $E_n > 0$ (with $n \in \mathbb{Z}_{>0}$), due to E_n are also involved in the differential equation of Eq. 16. Then, instead the integral, the auxiliary variable $\varphi(q) = \frac{\sqrt{2E}}{\sigma}\int_0^q e^{-2\theta(y)}\,dy$ is used, and the solutions of the quantized energies are set up as the following boundary problem:

$$\left.\begin{array}{c} \theta_n''(q) + \left(\theta_n'(q)\right)^2 + \frac{f^2(q)}{\sigma^2} = \frac{2E_n}{\sigma^2}\left(e^{-4\theta_n(q)} - 1\right);\ \theta_n(0) = \theta_n(c) = 0 \\ \varphi_n'(q) = \frac{\sqrt{2E_n}}{\sigma}e^{-2\theta_n(q)};\ \varphi_n(0) = 0;\ \varphi_n(c) = \pi \cdot n;\ n \in \mathbb{Z}_{>0} \end{array}\right\} \quad (17)$$

To get the solutions $E_n > 0$ from Eq. 17, a C++ program has been set up. The energies are found out by generating random numbers for E_n (for any $n \in \mathbb{Z}_{>0}$), with the four-order Runge-Kutta numerical solutions of the coupled differential equations in the interval $q \in [0, c]$, selecting those $E_n > 0$ that make the objective variable $\varphi_n(c) - \pi \cdot n$ as low as possible. In addition, several attempts have been done varying the value of the system Planck constant σ, trying that the quantum dynamics (Eq. 2) be as similar as possible to the classical dynamics (Eq. 1) for the fundamental value of the energy E_1 by Eq. 18, written below. Inserting the quantized energy values in Eq. 11, it is rewritten as:

$$\left.\begin{aligned} \dot{q}_n(t) &= f(q_n) - \frac{x_n}{2}\ ;\ q_n(0) = q_0 \\ \dot{x}_n(t) &= V_q'(q_n)\ ;\ x_n(0) = f(q_0) - \sqrt{2E_n + f(q_0)} \end{aligned}\right\} \tag{18}$$

In addition, the wave functions become:

$$\Psi_n(t,q) = e^{-i\frac{E_n}{\sigma}t}\psi_n(q) = D_n \cdot e^{-i\frac{E_n}{\sigma}t} e^{\theta_n(q)} \cdot sin\left(\frac{\sqrt{2E_n}}{\sigma}\int_0^q e^{-2\theta_n(y)}\,dy\right)\ ;\ n \in \mathbb{Z}_{>0} \tag{19}$$

Note in Eq. (19) that the wave function is defined only in the interval $[0, +\infty[$, being null for the values $q \leq 0$. In addition, the value obtained for the system Planck constant is $\sigma = 0.08$, different as $\sigma = 10.0$ in [1], and the first value obtained for the quantized energy is $E_1 = 0.000070$, with an error lesser than 0.000163. Figure 1 presents the result of $\theta_1(q)$ in the interval $[0, 30c] = [0, 300]$, the time-independent wave function $\psi_1(q)$ in the same interval and the shorter interval $[0, 2c] = [0, 20]$. The corresponding normalizing constant has been computed as usual as $D_1 = \left(\int_0^{+\infty} \psi_1^2(q)\,dq\right)^{-1/2} \cong \left(\int_0^{30c} \psi_1^2(q)\,dq\right)^{-1/2} = 0.674269$ (computed approximately by the Simpson's method). Therefore, the wave function is presented already normalized to the unit by multiplying by the D_1 value.

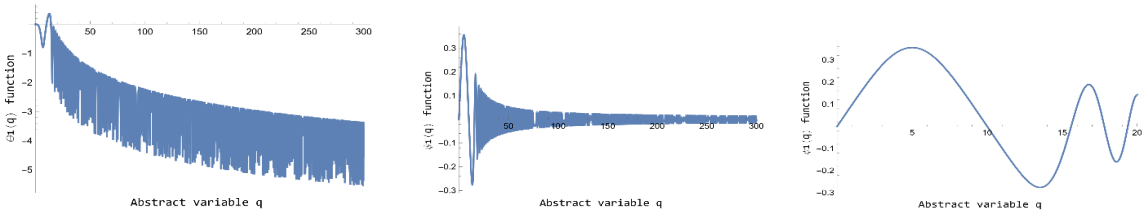


Figure 1. (Left) $\theta_1(q)$ function for $q \in [0, 30c]$; (Middle) normalized quantum wave $\psi_1(q)$ for $q \in [0, 30c]$; (Right) normalized quantum wave $\psi_1(q)$ for $q \in [0, 2c]$.

Figure 2 shows the corresponding quantum dynamics of Eq. 18 for $n = 1$, $q_1(t)$, in the time interval $t \in [0,1000]$ for two initial conditions: $q(0) = q_0 = 2 < c = 10.0$ and $q(0) = q_0 = 18 > c = 10.0$. Note that although, for $q_0 < c$, the quantum trajectory is so like the classical one for the fundamental quantized energy E_1, such as it happens with the quantum trajectory for $q_0 > c$.

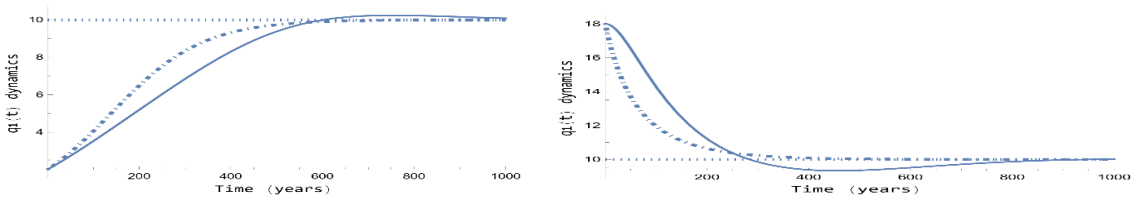


Figure 2. Classical and quantum trajectories for the logistics dynamics in the time interval $t \in [0,1000]$: (Left) initial condition $q_0 = 2 < c = 10.0$; (Right) initial condition $q_0 = 18 > c = 10.0$. The quantum dynamics corresponds to the continuous curve, while the dotted straight line represents the attractor value $c = a/b = 10.0$, and the dot dashed curve represents the classical logistic dynamics.

The allowed length of this document does not permit to present more figures for different quantized values. However, other values can be presented. The second energy level is $E_2 = 0.001090$ with an error lesser than 0.000465, and the third energy level is $E_3 = 0.002670$ with an error lesser than 0.008821. In addition, the corresponding normalizing constants are: $D_2 = \left(\int_0^{+\infty} \psi_2^2(q)\, dq\right)^{-1/2} \cong \left(\int_0^{30c} \psi_2^2(q)\, dq\right)^{-1/2} = 0.327388$, $D_3 = \left(\int_0^{+\infty} \psi_3^2(q)\, dq\right)^{-1/2} \cong \left(\int_0^{30c} \psi_3^2(q)\, dq\right)^{-1/2} = 0.306326$. The fact that the set $\{\Psi_n(t,q)\}_{n=1}^{+\infty}$, with $\Psi_n(t,q)$ given by Eq. 19, can be approximately proved because: $\int_0^{+\infty} \psi_1(q) \cdot \psi_2(q)\, dq \cong \int_0^{30c} \psi_1(q) \cdot \psi_2(q)\, dq = -0.118755$; $\int_0^{+\infty} \psi_1(q) \cdot \psi_3(q)\, dq \cong \int_0^{30c} \psi_1(q) \cdot \psi_3(q)\, dq = -0.0390836$; $\int_0^{+\infty} \psi_2(q) \cdot \psi_3(q)\, dq \cong \int_0^{30c} \psi_2(q) \cdot \psi_3(q)\, dq = 0.0308762$.

5 Conclusions

The work must be reviewed considering the [5] cited work. It offers different solutions to get the quantum trajectories for the one-dimensional time-independent case of ADS. As showed in Section 4, the one-dimensional autonomous case is not solved yet, although the solution for the Micó-Schrödinger equation is provided. Then, a provisional hypothesis provides the way to get deterministic quantum trajectories, although these trajectories should be stochastic.

References

[1] Micó, J., C., A Quantum Formalism for Abstract Dynamical Systems. *Mathematics*, 12: 1076, 2024.

[2] Dirac, P. A. M. Lectures on Quantum Mechanics. New York, Yeshiva University, 1964.

[3] Madelung, E. Quantum Theory in Hydrodynamical Form. *Z. Für Phys. (J. Phys.)*, 40: 1–4, 1927.

[4] Bohm, D., Hiley, B.J. The Undivided Universe: An Ontological Interpretation of Quantum Theory. London, New York, 1995.

[5] Wyatt, R. E. Quantum Dynamics with Trajectories. New York, Springer, 2005.

[6] Poirier, B. Reconciling Semiclassical and Bohmian Mechanics: I Stationary states. *arXiv:0802.3472v1* [quant-ph], 2008.

Synchronization of adaptative Kuramoto network in presence of contrarians

K. Alfaro-Bittner[1], G. Contreras-Aso[1,2], E. Primo[1,2], D. Rodríguez[1 1], and M. Romance[1,2,3]

([1]) Departamento de Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica, Universidad Rey Juan Carlos
28933 Móstoles (Madrid), Spain

([2]) Laboratory of Mathematical Computation on Complex Networks and their Applications, Universidad Rey Juan Carlos,
28933 Móstoles (Madrid), Spain

([3]) Data, Complex networks and Cybersecurity Research Institute, Universidad Rey Juan Carlos, 28028 Madrid, Spain

1 Introduction

Synchronization is a collective phenomenon arising in systems of interacting units, and it is pervasive across nature, society, and technology. Recent research has remarked the crucial role of interaction topology in the emergence of synchronized states, even when their interaction patterns change over time.

For instance, in Ref. [1], the authors consider classical Kuramoto model of N phase oscillators, where the coupling is directed and its value varies on time. The influence of node dynamics on the interaction is governed by the competition of two attributes: the homophily and homeostasis. Moreover, this competition improves not only the syncrhonization, but also make possible emergence of a mesoscale of communities and a scale-free distribution of connection weights.

On the other hand, in Ref. [2], the authors consider a generalization of the Kuramoto model where oscillators are coupled to the mean field with random signs. In this setup, oscillators with positive coupling, called "conformists", are attracted to the mean field and tend to synchronize with it. Conversely, oscillators with negative coupling, known as "contrarians", are repelled by the mean field and prefer to be in a phase that is diametrically opposed to it. This model, exhibits some surprising behaviors such as a desynchronized state and a partially synchronized state where conformists and contrarians are locked in antiphase.

To date, most studies have focused on either the Kuramoto model with adaptive networks of coupled oscillators (coevolution) or on the dynamics of oscillators with positive and negative coupling (conformists and contrarians), but few have analyzed these two aspects together.

[1] danieljose.rodriguez@urjc.es

2 Methods

The model we examine is a set of N coupled oscillators

$$\dot{\theta}_i = \omega_i + m_i \frac{\lambda}{N} \sum_{j=1}^{N} W_{ij} \sin(\theta_j - \theta_i), \quad i = 1, \ldots, N,$$

where θ_i is the phase of the i^{th} oscillator at time t and ω_i is its natural frequency, $m_i = \pm 1$ and λ corresponds to the coupling strength of the i^{th} oscillator. The adaptative evolution of the weights W_{ij} is governed by

$$\dot{W}_{ij}(t) = p_{ij}(t) - \left(\sum_{k \in N_i} p_{ik}(t) \right) W_{ij}(t) \quad \text{and} \quad p_{ij}(t) = \left| \frac{1}{T} \int_{-\infty}^{t} e^{-\frac{(t-u)}{T}} e^{i[\theta_j(u) - \theta_i(u)]} \, du \right|,$$

where N_i is a fixed number of randomly selected neighboring units of the i^{th} oscillator and p_{ij} is a average phase correlation between oscillators i and j over a characteristic memory time $T > 0$ [**?**]. In addition, in the spirit of [**?**], in this work the parameter p is the percentage of the total oscillators randomly selected that are conformist ($m_i = 1$) and, therefore, the $1 - p$ remaining oscillators are contrarians ($m_i = -1$).

3 Results

Our preliminary results show that:

1. The global synchronization of the system decreases by the presence of contrarians, but there is still synchronization among conformist, as it can be observed in Figure 1 for different percents of contrarians (10%, 25%, 50% and 75%) and differnte values of the coupling parameter λ.

2. The greater the number of contrarians, the worse their level of synchronization.

3. It seems that there is an optimal number of contrarians that maximizes the modularity measure.

4. There exists a power-law distribution for the weights W_{ij} in all cases, independently the number of contrarians.

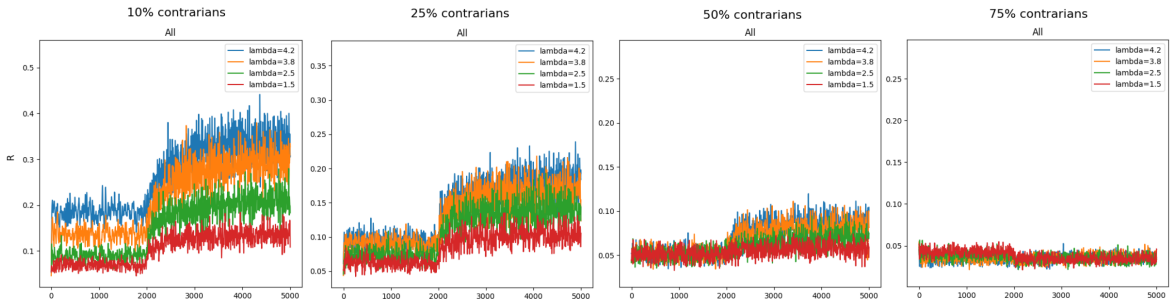


Figure 1: Order parameter for different values of p.

4 Conclusions

In this work, we explored the gap in understanding how the interplay between coevolutionary dynamics and the contrasting behaviors of conformists and contrarians can influence synchronization and collective behavior in complex systems.

In addition, we compute modularity measures for community detection that results from the synchronization phenomena and the possible scale-free distribution of connection strengths among the units.

Acknowledgments

This work has been supported by INCIBE/URJC Agreement M3386/2024/0031/001 within the framework of the Recovery, Transformation and Resilience Plan funds of the European Union (Next Generation EU).

References

[1] Gutiérrez, R., Amann, A., Assenza, S., Gómez-Gardeñes, J., Latora, V. and Boccaletti, S., Emerging Meso- and Macroscales from Synchronization of Adaptive Networks, *Phys. Rev. Lett.*, 107 (23): 234103, 2011.

[2] Hong, H. and Strogatz, S. H., Kuramoto model of coupled oscillators with positive and negative coupling parameters: An example of conformist and contrarian oscillators, *Phys. Rev. Lett.*, 106 (5): 054102, 2011.

Relativistic Positioning and Dilution of Precision

R. Serrano Montesinos[♭],[1] and J.A. Morales-Lladosa[♮]

(♭) Departament d'Astronomia i Astrofísica, Universitat de València,
46100 Burjassot, València, Spain.

(♮) Departament d'Astronomia i Astrofísica, Universitat de València,
46100 Burjassot, València, Spain.
Observatori Astronòmic, Universitat de València,
46980 Paterna, València, Spain.

1 Introduction

To gain insight into the theory of Relativistic Positioning Systems (RPS) [1–9], in this talk we will present a formula for the Geometric Dilution of Precision (GDOP) in the case of $k \geq 4$ emitters (satellites). In Minkowski space-time, a RPS can be thought of as a set of at least four emitters A ($A = 1, 2, 3, 4$) of world-lines $\gamma_A(\tau^A)$, broadcasting their respective proper times τ^A by means of electromagnetic signals. The term GDOP is used for a matrix constructed from the positions of the satellites on the sky, the GDOP matrix, as well as for the square root of its trace, the so called GDOP scalar, a metric indicator of the goodness in the determination of the user location solely from the geometry of the satellite configuration in view, leaving aside any kind of measurement errors.

2 Previous work

Massat and Rudnick [10] seem to be the first authors who obtained a geometric formula for the GDOP scalar in the case of four emitters, by expressing it in terms of volumes and areas defined by the relative position of the emitters on the user's unit celestial sphere (see Fig. 1). More recently, in [11], Santerre, Geiger and Banville generalized Massatt and Rudnick's results for an arbitrary number of emitters and arrived at the following result:

$$GDOP = \sqrt{\frac{\sum_{i=1}^{\binom{k}{3}} [(U_i)^2 + \frac{1}{9}(\mathcal{A}_i)^2]}{\sum_{i=1}^{\binom{k}{4}} (\mathcal{V}_i)^2}}, \tag{1}$$

where the involved quantities are volumes and areas associated to the relative positions of the quads and triads of emitters on the unit celestial sphere S_U of the user U (see Fig. 1), that is:

[1] rasemon@alumni.uv.es

- $\mathcal{V}_i$ is the volume of the inscribed tetrahedron defined by a i-quad of emitters,
- U_i is the volume of inner tetrahedron defined by a i-triad of emitters and the user,
- $\mathcal{A}_i$ is the area defined by a i-triad of emitters.

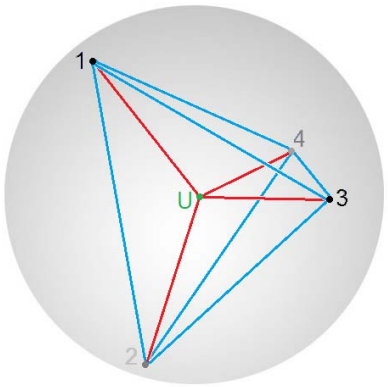


Figure 1: User's unit celestial sphere S_U and *inscribed* tetrahedron defined by the relative positions of four emitters on it. Four *inner* tetrahedrons with common vertex on U and with a sole inscribed face on S_U are also shown.

3 Results

A covariant formulation of the GDOP matrix is presented here in the framework of a RPS. The *GDOP* scalar is computed in terms of observable quantities, the number of emitters in view, k, and their mutual angles, θ_{AB}, from this formula

$$GDOP = \sqrt{\frac{\frac{\lambda}{3} + (2-k)\sigma + 2\chi}{(k-3)\frac{\lambda}{3} + s\frac{\sigma}{2} - \xi}}. \tag{2}$$

with s, σ, χ, λ and ξ given by

$$s \equiv 2\sum_{A<B}^{k} \mathcal{G}_{AB}, \tag{3}$$

$$\sigma \equiv 2\sum_{A<B}^{k} (\mathcal{G}_{AB})^2 = \operatorname{tr}\mathcal{G}^2, \tag{4}$$

$$\chi \equiv 2\sum_{A<B}^{k} (\mathcal{G}^2)_{AB} \tag{5}$$

$$\lambda \equiv \sum_{A,B,C=1}^{k} \mathcal{G}_{AB}\,\mathcal{G}_{BC}\,\mathcal{G}_{CA} = \operatorname{tr}\mathcal{G}^3, \tag{6}$$

$$\xi \equiv 2\sum_{A<B}^{k} (\mathcal{G}^3)_{AB}, \tag{7}$$

$$\tag{8}$$

where $\mathcal{G}_{AB}$ is the Gram matrix defined as follows. If g the space-time metric with signature $(-+++)$ and u the user four-velocity ($u^2 \equiv g(u,u) \equiv u \cdot u = -1$), the unit frequency parameterization is taken for every null vector $l_A = u + \vec{n}_A$, with $u \cdot \vec{n}_A = 0$, $\vec{n}_A$ being the unit vector along the direction of propagation of the signal from emitter A to the user (that is, $-\vec{n}_A$ is the unit vector on the line of sight, from the user to emitter A). The entries of the Gram matrix $\mathcal{G}_{AB}$ are the scalar products of the k future oriented null vectors, $\{l_A\}_{A=1}^{k}$, that is

$$\mathcal{G}_{AB} = l_A \cdot l_B = -1 + \cos\theta_{AB} = -2\sin^2\frac{\theta_{AB}}{2}, \quad A, B = 1, ..., k, \tag{9}$$

with $0 \leq \theta_{AB} \leq \pi$.

The geometric interpretation of the GDOP scalar is here recovered by direct substitution of the following identities in Equation 2:

$$\frac{1}{4}(\omega - \frac{\sigma^2}{2}) = \sum_{i=1}^{\binom{k}{4}} (6\mathcal{V}_i)^2 = (k-3)\,\frac{\lambda}{3} + s\,\frac{\sigma}{2} - \xi. \tag{10}$$

$$\sum_{i=1}^{\binom{k}{3}} [(6U_i)^2 + (2\mathcal{A}_i)^2] = \frac{\lambda}{3} + (2-k)\,\sigma + 2\chi \tag{11}$$

Acknowledgments

We would like to thank the support from the Conselleria d'Educació, Universitats i Ocupació, Generalitat Valenciana, Project CIAICO/2022/252.

References

[1] Coll, B., Ferrando, J. & Morales-Lladosa, J. Positioning systems in Minkowski spacetime: from emission to inertial coordinates. *Classical And Quantum Gravity*. 27: 065013, 2010, https://doi.org/10.1088/0264-9381/27/6/065013

[2] Coll, B., Ferrando, J. & Morales-Lladosa, J. Positioning systems in Minkowski space-time: bifurcation problem and observational data. *Phys. Rev. D*. 86: 084036, 2012, https://link.aps.org/doi/10.1103/PhysRevD.86.084036

[3] Coll, B. Relativistic positioning systems: perspectives and prospects. *Acta Futura*. 7: 35-47, 2013

[4] Coll, B. Epistemic Relativity: An Experimental Approach to Physics. *Relativistic Geodesy: Foundations And Applications*. 291-315, 2019, https://link.springer.com/chapter/10.1007/978-3-030-11500-5_8

[5] Coll, B. & Pozo, J. Relativistic positioning systems: the emission coordinates. *Classical And Quantum Gravity*. 23: 7395 (2006,11), https://dx.doi.org/10.1088/0264-9381/23/24/012

[6] Delva, P., Kostić, U. & Čadež, A. Numerical modeling of a Global Navigation Satellite System in a general relativistic framework. *Advances In Space Research*. 47: 370-379, 2011, https://www.sciencedirect.com/science/article/pii/S0273117710004643, Sci-entific applications of Galileo and other Global Navigation Satellite Systems - I

[7] Puchades, N. & Sáez, D. Approaches to relativistic positioning around Earth and error estimations. *Advances In Space Research*. 57, 499-508, 2016, https://www.sciencedirect.com/science/article/pii/S0273117715007504

[8] Serrano Montesinos, R. & Morales-Lladosa, J. Minkowskian Approach to the Pseudorange Navigation Equations. *Universe*. **10** (2024), https://www.mdpi.com/2218-1997/10/4/179

[9] Serrano Montesinos, R., Ferrando, J. & Morales-Lladosa, J. Location Problem in Relativistic Positioning: relative Formulation. *Universe*. 10, 2024, https://www.mdpi.com/2218-1997/10/7/299

[10] Massatt, P. & Rudnick, K. Geometric Formulas for Dilution of Precision Calculations. *Navigation*. 37: 379-391, 1990, https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2161-4296.1990.tb01563.x

[11] Santerre, R., Geiger, A. & Banville, S. Geometry of GPS dilution of precision: revisited. *GPS Solut.* 21: 1747-1763, 2017, https://doi.org/10.1007/s10291-017-0649-y

Relativistic Diatomic Molecules in an Exponential-Type Radial Potential

Michael M. Tung[♭1] and Guillem Gómez i Blanch[♮]

(♭)(♮) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, 46022 València, Spain.

1 Introduction

Diatomic molecules consist of two atoms that vibrate periodically around their shared center of mass. These vibrations occur without rotational motion in their lowest energy state and involve the bond stretching or compressing along a single axis. Initially, this vibrational behaviour can be approximated as harmonic, i.e the potential energy varies quadratically with the displacement from the equilibrium bond length [**?**]. In this idealized scenario, the molecule produces a fundamental vibrational frequency, and all higher vibrational modes occur at exact integer multiples of this base frequency.

In practice, however, molecular vibrations deviate from perfect harmonic behaviour as the energy increases and do not follow the predicted pattern precisely. Instead, their higher-energy levels become increasingly compressed. To account for this effect, the Morse potential [**?**, **?**, **?**] was introduced, providing a more accurate model. It realistically captures the anharmonic nature of molecular vibrations and effectively describes both the gradual softening of vibrational levels and the eventual breaking of the molecular bond. The Morse potential is not simply an empirical model but is related to the theory of atomic orbitals.

While the Morse potential remains a widely used and historically significant model for the theoretical description of diatomic molecules, particularly in capturing anharmonic vibrational behaviour, it is by no means the only available option. In fact, a broader and more flexible class of models exists, known as multiparametric exponential-type radial (MER) potentials. These MER potentials offer enhanced versatility and precision in modeling molecular interactions, as they include a greater number of parameters to more accurately reflect the underlying physics of molecular vibrations. Indeed, the Morse potential can be viewed as a very specific, simplified case within the wider framework of MER potentials. The development of MER potentials is based on rigorous quantum mechanical principles, providing also a robust theoretical foundation for analyzing the vibrational spectra of diatomic molecules [**?**]. This increased generality makes MER potentials particularly useful for high-precision spectroscopic studies. For instance, in the case of hydrogen chloride isotopologue HCl^{35}, the use of the MER potential allows for a significantly more accurate determination of vibrational constants compared to estimates established solely within the Morse model [**?**].

The primary aim of the present study is to investigate and analyze the behavior and characteristics of multiparametric exponential-type radial potential models when subjected to strong

[1] mtung@imm.upv.es

gravitational fields. While MER potentials have been extensively studied in non-relativistic contexts for their ability to model anharmonic effects in diatomic molecules with high accuracy, their application in environments influenced by significant gravitational forces—such as those near massive astrophysical objects—remains largely unexplored. This work, therefore, seeks to extend the scope of MER models by incorporating relativistic corrections that account for the influence of curved spacetime on molecular dynamics.

Specifically, we focus on developing a theoretical framework that applies a semiclassical approximation to describe the vibrational motion of the two nuclei within a diatomic molecule when exposed to high-impact gravitational potentials. In this approach, the nuclear dynamics are treated using a semiclassical formalism, while the background gravitational field is incorporated through general relativistic modifications. By doing so, we aim to assess how strong gravitational effects alter the vibrational energy levels and spectroscopic behavior of such molecular systems. This investigation not only contributes to the refinement of MER models but also opens new possibilities for understanding molecular interactions under extreme physical conditions, such as those found near neutron stars or black holes.

2 Methods

Our approach is built on recent work for the relativistic harmonic oscillator in a strong gravitational field [**?**] by developing a more comprehensive model describing the vibrational motion of rotationless diatomic molecules under relativistic conditions. The original mathematical framework is rooted on an analysis of the relativistic brachistochrone [**?**], which has since then been adapted to various related systems such as the relativistic pendulum [**?**, **?**].

More precisely, this procedure will be using the Euler-Lagrange formalism—a fundamental method in classical mechanics used to derive the equations of motion for a physical system from a scalar quantity called the Lagrangian, generally given by $L = T - V$, where T is the relativistic kinetic term and V the potential in question, *viz.* [**?**]. We will present the underlying principle of stationary action, deriving the trajectories which the physical system's path will follow by minimising and maximising the corresponding action integral (which is a functional related to the system's energy).

Here, we incorporate the MER potential given by a function of the interatomic distance r:

$$V(r) = V_0 \left[\left(\frac{1 - q^2 e^{-2\beta r_e}}{(1 - q e^{-\beta r})^2} \right) e^{-2\beta(r - r_e)} - 2 \left(\frac{1 - q e^{-\beta r_e}}{(1 - q e^{-\beta r})^2} \right) e^{-\beta(r - r_e)} \right],$$

where $V(r_e) = -V_0$ is the potential depth at equilibrium point r_e, and further β and q are the two remaining parameters for characterising the underlying model of the diatomic molecule. Then, the associated variational principle is

$$\delta \int dt \left[T - V(r) - V_g(r) \right] = 0.$$

Here, V_g is the yet to be determined gravitational potential, where, for proper analysis, it makes generally sense to assume that the gravitational force field is fully aligned with the internuclear axis of the diatomic molecule.

This formalism provides a systematic and powerful framework for deriving the equations which govern the vibrational motion of diatomic systems in relativistic context, thus incorporating the approximate effects of curved spacetime and strong gravitational fields into the dynamics of this system. However, a significant challenge in this approach arises from the complicated nature of the relativistic gravitational potential. Unlike the mostly simple potentials encountered in classical

mechanics, here the gravitational potential in a relativistic setting is not easily expressed in a straightforward closed form, introducing substantial mathematical difficulties when attempting to solve the resulting equations of motion analytically. In fact, due to these complications, obtaining a complete and exact closed-form solution for the vibrational dynamics of such a system proves to be analytically intractable. Consequently, numerical integration methods must be employed for producing important physical predictions for this model.

Our numerical estimates are based on models for well-studied gaseous compounds whose traces can be found in interstellar space, or, as in the case of hydrogen chloride gas, in the atmosphere of planet Mars. There, hydrogen chloride (HCl) was first identified by the Atmospheric Chemistry Suite (ACS) and subsequently confirmed by the *Nadir and Occultation for Mars Discovery* (NOMAD) instruments, both onboard the European Space Agency's ExoMars Trace Gas Orbiter (TGO) mission, see e.g. [**?**]. In interstellar space, diatomic hydrogen (H_2)is the most abundant molecule, although difficult to detect due to the weakness of its rotational and vibrational modes. However, in hotter regions it is easier to detect. Other molecules like carbon monoxide (CO) are much easier to observe than molecular hydrogen and are therefore frequently used as proxies to trace H_2 studying the physical conditions within such molecular clouds.

Summing up, the ability of these numerical models to make accurate predictions could prove highly valuable in the field of radiation astronomy. In this domain, scientists often rely on the detection and analysis of vibrational features within molecular spectra to gain insights into astrophysical processes. These spectral signatures, while not direct indicators themselves, can serve as indirect evidence pointing to the presence and influence of extreme gravitational phenomena, such as those found near black holes, neutron stars, or other compact astrophysical objects. By improving our understanding of how molecular vibrations behave under such conditions, these models make it easier to interpret observational data and give scientists a clearer picture of how matter and energy act in some of the most extreme environments in space.

3 Conclusions

Through the implementation of numerical integration techniques, it becomes possible to simulate the relativistic behaviour of a diatomic molecular system with a high degree of precision. By computing the system's evolution over time and analyzing the results across a range of representative initial conditions and parameter values, we are able to give numerical estimates that highlight the model's most significant physical features. Furthermore, these simulations will allow us to visualize the vibrational modes of the molecule within a strong gravitational field. Additionally, they might serve to illustrate the distinct effects of anharmonicity and the apparent bond dissociation due to the underlying potential deviating from the very simple harmonic model. Before bond dissociation takes place, phase-space diagrams will expect to exhibit closed trajectories, characteristic of periodic vibrational motion. These results will be used to compare the behaviour predicted by the relativistic model in a strong, uniform gravitational field with that of the corresponding classical predictions.

Computational modelling of this kind proves especially valuable in both theoretical investigations and observational applications, particularly in situations where direct experimentation is either highly impractical or altogether impossible. In environments with extreme physical conditions—such as those found in the vicinity of compact astrophysical objects—the ability to observe diatomic dynamics and then compare it to theoretical predictions becomes a powerful tool for studying and testing scenarios that are beyond the reach of laboratory settings, especially since the presence of such strong gravitational fields will impact the molecular dynamics in appreciable and measurable

Therefore, the predictive power of these numerical models may hold relevance for radiation

astronomy, where the detection and interpretation of vibrational signatures in molecular spectra could offer indirect evidence of intense gravitational phenomena.

Acknowledgments

This work has been supported by "Optimized Computation of Matrix Functions and Applications to Artificial Intelligence" (COFMIA), R+D Projects for Consolidated Research Groups 2024–2027, funded by the Conselleria de Educación, Universidades y Empleo, Generalitat Valenciana (Spain).

References

[1] Goldstein, H., Poole, C., and Safko, J., Classical Mechanics, San Francisco, Addison Wesley, 2001.

[2] Morse, P.M., Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels, *Phys. Rev.*, 34(1):57–64, 1929.

[3] Varshni, Y.P., Comparative Study of Potential Energy Functions for Diatomic Molecules, *Rev. Mod. Phys.*, 29:664–682, 1957.

[4] Mirzanejad, A., Varganov, S., Derivation of Morse Potential Function, *ChemRxiv*, 2023:1–13, 2023.

[5] Peña, J.J., Morales, J., García-Ravelo, J., and Avendaño, J., On the one-dimensional Morse potential as limit of a class of multiparameter exponential-type radial potential, *Int. J. Quantum Chem.*, 121(7):1–12, 2021.

[6] Tung, M.M., The Relativistic Harmonic Oscillator in a Uniform Gravitational Field, *Mathematics 2021*, 9(4):1–12, 2021.

[7] Goldstein, H.F., Bender, C.M., Relativistic brachistochrone, *J. Math. Phys.*, 27(2):507–511, 1986.

[8] Erkal, C., The simple pendulum: a relativistic revisit, *Eur. J. Phys.*, 21:377–384, 2000.

[9] Torres, P.J., Periodic oscillations of the relativistic pendulum with friction, *Phys. Lett. A*, 372:6386–6387, 2008.

[10] Christodoulides, C., The Special Theory of Relativity: Foundations, Theory, Verification, Applications, Berlin, Springer, 2016.

[11] Korablev, O. *et al.*, Transient HCl in the atmosphere of Mars, *Sci. Adv.*, 7:1–8, 2021.

Modeling the Gravitational-Wave Forest from Galactic and Extragalactic source Populations

[1]Shao-Dong Zhao[a,b,c], Pau Amaro Seoane[a]

a Universitat Politècnica de València, València, Spain ,

b Institute of Theoretical Physics & Research Center of Gravitation, Lanzhou University, Lanzhou 730000, China

c Lanzhou Center for Theoretical Physics, Key Laboratory of Theoretical Physics of Gansu Province, Key Laboratory of Quantum Theory and Applications of MoE, Gansu Provincial Research Center for Basic Disciplines of Quantum Physics, Lanzhou University, Lanzhou 730000, China

1 Introduction

The gravitational capture of a compact object by a supermassive black hole (SMBH) represents one of the most promising sources for upcoming space-based gravitational wave observatories, such as LISA [5, 1, 3], Taiji [15], and TianQin [12]. Although the extreme-mass-ratio inspiral (EMRI) event rate in a single galaxy like the Milky Way is low (on the order of 10^{-5} to 10^{-6} per year), the expected cumulative cosmological rate suggests that a few events per year will be detectable, potentially out to redshifts $z \sim 1$–4 [7, 2].

Recent studies [6] suggest that several tens of early-EMRIs (E-EMRIs) may currently be present in their early stages in the Galactic Center. These E-EMRIs can maintain signal-to-noise ratios (SNR) above 10 for up to 10^5–10^6 years before final plunge. During the relatively short observational window of LISA (a few years), such systems appear nearly monochromatic, dominated by fixed peak frequencies despite exhibiting rich harmonic structure. Over time, as their orbital frequency gradually increases, these sources transition into an oligochromatic regime. In contrast, EMRIs in their final stages—just weeks or months from merger—enter a polychromatic phase, sweeping rapidly across the LISA band [6].

Substellar objects like brown dwarfs can also inspiral into the SMBH without undergoing tidal disruption, producing gravitational waves in the extreme mass-ratio regime. These extremely large mass ratio (XMRI) systems can reach very high SNRs in their late stages [2, 11], and even in earlier phases, they can enable extremely precise measurements of SMBH parameters [17].

The models in [2, 6] assume a steady-state source population. Provided that SgrA* satisfies two-body relaxation conditions and has a mass between $10^5 M_\odot$ and a few $10^7 M_\odot$—sufficient to remain dynamically centered—there will always be a persistent population of several tens of E-EMRIs and XMRIs present near the Galactic Center.

In [16], we analyzed the combined gravitational wave signal from E-EMRIs and XMRIs at the Galactic Center using a realistic population catalog and conservative assumptions. We demon-

[1]sdzhao@doctor.upv.es

strate that the incoherent superposition of these numerous weak sources—each with a randomly distributed phase—produces a confusion signal we refer to as a "forest". This forest has both background and foreground components. It can obscure other signals in the LISA band, including massive black hole binaries, verification binaries, and individually resolvable EMRIs.

The dominant contributors to the confusion foreground are high-eccentricity, slowly evolving sources with $\Delta f = \dot{f}\, T_{\rm obs} \ll 1/T_{\rm obs}$, which overlap densely in frequency and are effectively unresolved. These form a stochastic confusion foreground, analogous to the unresolved Galactic white dwarf binary population. Conversely, sources with lower eccentricity, measurable frequency drift ($\Delta f \sim 1/T_{\rm obs}$), and higher SNR form a more structured foreground. These may be partially resolvable over the mission lifetime, although prior to identification, they contribute excess power that biases background estimation.

The relative contribution of each population class evolves with the observation time $T_{\rm obs}$. As $T_{\rm obs}$ increases, some monochromatic sources accumulate enough SNR to become oligochromatic, while others transition into fully resolvable polychromatic systems. A source is considered resolvable if it satisfies both SNR $> \rho_{\rm thresh}$ (typically $\rho_{\rm thresh} \approx 10$) and $\Delta f \cdot T_{\rm obs} > 1$.

From a data analysis perspective, this classification implies that truly monochromatic sources can be treated as part of an incoherent background, while oligochromatic sources may require specialized detection strategies or hierarchical subtraction techniques to prevent contamination of noise models. The forest's non-Gaussian and time-modulated features—especially due to LISA's time-varying antenna response toward the Galactic Center—underscore the need for careful modeling to separate it from instrumental and astrophysical backgrounds.

In this short abstract, we first examine the *minimum* contribution of the forest signal from the Galactic Center. We then extend our analysis to investigate the potential gravitational-wave forest signals originating from a nearby galaxy.

2 Catalog and strain

2.1 GW source catalog

In this study, we examine two distinct classes of GW sources: XMRIs, which involve brown dwarfs as the inspiraling objects, and E-EMRIs, which are formed by stellar-mass black holes orbiting SMBH. For the XMRI population, we consider companion masses in the interval 0.01–0.08 $M_\odot$, a range encompassing and slightly extending the conventional brown dwarf regime ($\sim$ 13–80 Jupiter masses). For E-EMRIs, we adopt a stellar-mass black hole range of 10–100 $M_\odot$, consistent with mass distributions derived from gravitational-wave observations [9, 10].

To explore the potential for detecting such systems beyond the Milky Way, we turn our attention to NGC 4303 (Messier 61), a nearby galaxy hosting an SMBH with estimated mass $M_\bullet \sim 5 \times 10^6\, M_\odot$ [13], located at a distance of approximately 16.1 Mpc. This selection is motivated by prior work on detectability from extragalactic sources [2, 6, 16] and represents a realistic scenario for LISA-band gravitational-wave signals originating outside the Galactic center.

The initial conditions for each binary system are defined using the critical semi-major axis $a_{\rm crit}$, which marks the transition between two dynamical regimes. At larger separations, stellar relaxation and other environmental interactions govern orbital evolution, while within $a_{\rm crit}$, the dynamics become dominated by gravitational-wave emission [1, 3]. Starting the inspiral at this threshold ensures we focus on the GW-driven phase, which contributes most significantly to the LISA-band signal and allows for a consistent treatment of relativistic dynamics.

To construct synthetic populations of XMRIs and E-EMRIs, we sample secondary masses from power-law distributions favoring higher-mass companions. Eccentricities are drawn from a logarithmic distribution, in line with expectations that compact objects begin their GW-dominated inspirals

with relatively high eccentricity, which then gradually decreases due to radiation reaction [4, 2, 6]. These parameters, together with the initial orbital radius $a_{\rm crit}$, define the configuration of each binary system for our numerical modeling of gravitational-wave emission across the LISA sensitivity range.

We calculate evolutionary trajectories for each unique initial configuration, secondary mass and critical eccentricity, using the approximation of [14]:

$$\frac{df_{\rm orb}}{dt} = \frac{96}{10\pi}\left(\frac{\mathcal{M}\,R_\odot}{c}\right)^{5/3}\frac{(2\pi f_{\rm orb})^{11/3}}{(1-e^2)^{7/2}}\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right), \tag{1}$$

$$\frac{de}{dt} = -\frac{e}{15}\left(\frac{\mathcal{M}\,R_\odot}{c}\right)^{5/3}\frac{(2\pi f_{\rm orb})^{8/3}}{(1-e^2)^{5/2}}\left(304+121e^2\right), \tag{2}$$

where $\mathcal{M}$ represents the chirp mass of the binary system, defined as $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$, m_1 and m_2 are the component masses. $R_\odot$ is the Schwarzschild radius of one solar mass. These individual evolution trajectories are then combined to create complete XMRI and E-EMRI source banks used for the forest signal estimation.

2.2 Strain determination

To evaluate the gravitational-wave forest (GWF) signal produced by populations of XMRIs and E-EMRIs, we compute the strain amplitudes contributed by in-band sources across the LISA frequency range. Each individual binary emits radiation across multiple harmonics of its orbital frequency. The effective signal from the n-th harmonic is quantified using the root-mean-square (rms) strain amplitude [8]:

$$h_{{\rm rms},n} \equiv \sqrt{\langle h_{n+}^2 + h_{n\times}^2\rangle} = \frac{1}{\pi d f_n}\sqrt{\frac{G\dot{E}_n}{c^3}}, \tag{3}$$

where $f_n = n f_{\rm orb}$ is the frequency of the n-th harmonic, $\dot{E}_n$ denotes the power radiated in that harmonic, and d is the distance to the source. This quantity characterizes the contribution of each unresolved binary to the total background, in analogy with the flux from unresolved sources in astronomical imaging.

A key aspect of signal detectability is whether the source is effectively monochromatic (negligible frequency drift) or displays measurable frequency evolution over the mission duration. The accumulated phase drift, $\Delta\varphi = \pi\dot{f}t^2$, plays a critical role in distinguishing coherent sources from noise. Sources that undergo significant frequency evolution over the observation time benefit from coherent integration, thereby enhancing their detectability.

For systems exhibiting noticeable frequency drift, we define the *characteristic strain* to reflect the signal's evolution:

$$h_{c,n} \equiv h_{{\rm rms},n}\sqrt{\frac{2f_n^2}{\dot{f}_n}} = \frac{1}{\pi d}\sqrt{\frac{G\dot{E}_n}{c^3\dot{f}_n}}, \tag{4}$$

which estimates the total amplitude accumulated as the signal sweeps through the frequency domain, analogous to a chirping burst.

Explicit expressions for $h_{{\rm rms},n}$ and $h_{c,n}$ can be obtained by inserting the corresponding energy flux and frequency evolution equations:

$$h_{{\rm rms},n}^2 = \frac{2^{10/3}\cdot 32(G\mathcal{M})^{10/3}(\pi f_{\rm orb})^{4/3}g(n,e)}{5n^2d^2c^8}, \tag{5}$$

$$h_{c,n}^2 = \frac{2(2\pi)^{2/3}(G\mathcal{M})^{5/3}g(n,e)}{3c^3 f_n^{1/3}\mathcal{F}(e)\pi^2 d^2}, \tag{6}$$

where $g(n, e)$ is the spectral enhancement factor from eccentricity, and $\mathcal{F}(e)$ is the total gravitational-wave flux factor.

To ensure accurate treatment of rapidly evolving sources, we apply a cycle-based normalization. The number of wave cycles spent near f_n during an observation of duration $T_{\rm obs}$ is given by $\mathcal{N}_{\rm cyc} = f_n T_{\rm obs}$. We define an effective strain for such sources as:

$$h_{c,n}^{\rm eff} = \frac{h_{c,n}}{f_n T_{\rm obs}}, \tag{7}$$

down-weighting the contribution of transient signals to avoid overestimation in narrow frequency bins.

The aggregate GWF signal, expressed as the power spectral density $h_{\rm gwf}^2(f)$, depends on whether the source evolves significantly in frequency:

$$h_{\rm gwf}^2(f) = \begin{cases} \displaystyle\int \mathrm{d}\mathcal{M}\mathrm{d}e \left[\sum_n h_{{\rm rms},n}^2(f) \frac{\mathrm{d}^3 N}{\mathrm{d}\mathcal{M}\,\mathrm{d}e\,\mathrm{d}\ln f_{\rm orb}} \right], & \dot{f}_n T_{\rm obs} < f_n, \\ \displaystyle\int \mathrm{d}\mathcal{M}\mathrm{d}e \left[\sum_n \frac{h_{c,n}^2(f)}{2 f T_{\rm obs}} \frac{\mathrm{d}^3 N}{\mathrm{d}\mathcal{M}\,\mathrm{d}e\,\mathrm{d}\ln f_{\rm orb}} \right], & \dot{f}_n T_{\rm obs} \geq f_n. \end{cases} \tag{8}$$

The first case applies to sources whose frequency remains nearly constant during the observation; their rms strain accurately represents their contribution. The second regime includes inspirals whose frequency evolution must be fully resolved using $h_{c,n}$, scaled by the observation bandwidth. In both scenarios, the integration over $\mathcal{M}$, e, and harmonic index n ensures a statistically complete estimate of the total forest background, informed by the differential source distribution $\mathrm{d}^3 N/(\mathrm{d}\mathcal{M}\mathrm{d}e\mathrm{d}\ln f_{\rm orb})$ across parameter space.

3 Results

3.1 Complement results for the forest signal from galactic center

The detailed configuration of the SNR region divisions and the corresponding number of sources in each region is described in [16]. To quantify the maximum contribution of this population to the gravitational-wave forest, we define a cumulative normalized measure as

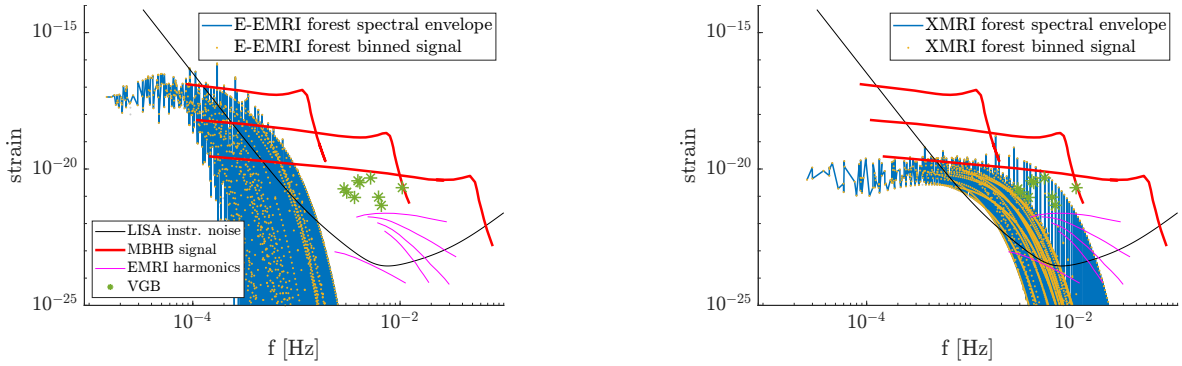
$$V_{\rm norm} = \sum_i \sum_s f_i \, h_{{\rm bkg},\,s}(f_i), \tag{9}$$

where s indexes individual sources, i labels frequency bins, f_i is the central frequency of bin i, and $h_{{\rm bkg},s}(f_i)$ denotes the strain amplitude of source s at that frequency. We perform $N_{\rm real} = 10^3$ independent realizations to obtain a statistically robust estimate of $V_{\rm norm}$.

Figures 4 and 7 in [16] illustrate the *maximum* case of our XMRI and E-EMRI forest analysis, shown in comparison with representative gravitational-wave sources expected in the LISA band. Here, "maximum" refers to the case corresponding to the highest value of the normalized signal measure.

We also compute the *minimal* case of the forest signal from the Galactic Center, using the same SNR-based source populations. The results are shown in Figure 1. Overlaid for comparison are characteristic signals from massive black hole binaries (MBHBs), Galactic verification binaries, and typical EMRI harmonics, along with the LISA instrumental sensitivity curve. Yellow dots represent the summed XMRI strain amplitudes binned by frequency, while the blue curve connects the non-zero strain values to indicate the overall spectral shape of the forest.

Notably, the XMRI background partially overlaps with the harmonic structure of polychromatic EMRIs near $\sim 5\,\mathrm{mHz}$, potentially contributing confusion noise in this region. However, due to the low companion masses in XMRI systems, the background remains modest and is unlikely to significantly impede the detection of primary sources such as MBHBs or Galactic binaries. Even in this minimal configuration, the forest signal introduces a non-negligible effect on LISA's sensitivity to other gravitational-wave sources.



(a) 1-year E-EMRI forest minimum configuration

(b) 1-year XMRI forest minimum configuration

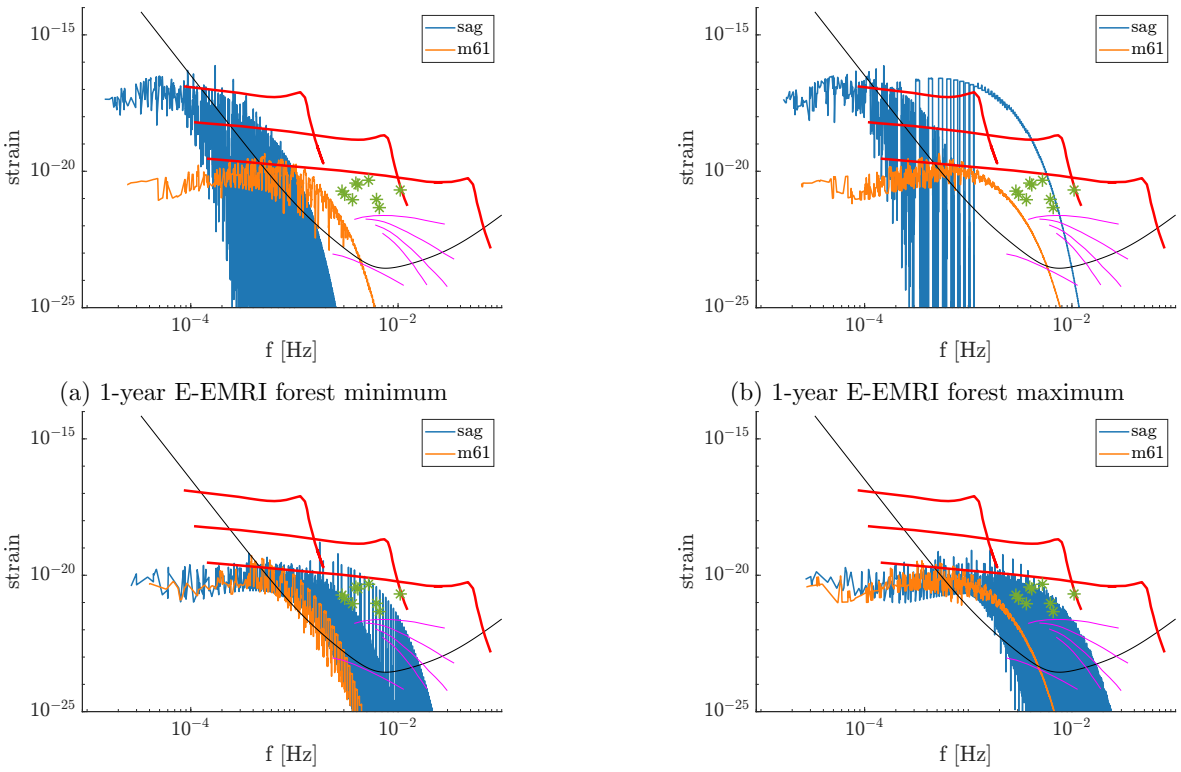
Figure 1: Spectral characterization of the XMRI and E-EMRI forests for 1-year (top row) and 10-year (bottom row) observation periods. Grey points show individual harmonics from distinct sources. Yellow dots represent the superposed strain within frequency bins. The blue curve connects all non-zero strain values to trace the overall spectrum shape. Overlaid for comparison are representative signals from massive black hole binaries (red), Galactic verification binaries (green stars), EMRI harmonics (magenta), and LISA noise (black).

3.2 Signal from neighborhood galaxy

For our analysis of the NGC4303 system, we sample approximately $N \approx 12$ in-band sources from the XMRI catalog and $N \approx 18$ in-band sources from the E-EMRI catalog. These numbers are based on the population models of [2, 6], but are lower than those considered in [16] due to the significantly greater distance to NGC4303 compared to the Galactic Center.

To ensure representative coverage across the range of signal strengths, we apply a stratified selection based on SNR. For the XMRI population over a 1-year observation, we select $n_1 = 8$ sources from the low-SNR interval $[10, 50)$, $n_2 = 3$ from the moderate range $[50, 500)$, and $n_3 = 1$ from the intermediate range $[500, 1000)$. For the E-EMRI population, the 1-year sample includes $n_1 = 10$ sources from $[10, 50)$, $n_2 = 5$ from $[50, 500)$, and $n_3 = 3$ from $[500, 1000)$. Using these stratified samples, we conduct $N_{\mathrm{real}} = 10^3$ independent realizations for each population to obtain a statistically robust characterization of the gravitational-wave forest originating from NGC4303.

In Figure 2, we compare the forest signals from the Galactic Center and NGC 4303 for both E-EMRIs and XMRIs. Despite the weaker strain from NGC 4303 due to its distance, its contribution remains non-negligible. The Galactic Center forests are significantly stronger and exhibit more spectral structure, particularly near the EMRI band, potentially introducing notable confusion noise for LISA.



(a) 1-year E-EMRI forest minimum

(b) 1-year E-EMRI forest maximum

(c) 1-year XMRI forest minimum

(d) 1-year XMRI forest maximum

Figure 2: Spectral characterization of the E-EMRI (top row) and XMRI (bottom row) forests from different galaxies over a 1-year observation period. The blue curve shows the overall spectral shape of the forest signal originating from the Galactic Center, while the orange curve represents the corresponding signal from NGC 4303.

Conclusion

In this work, we investigated the gravitational-wave forest (GWF) originating from XMRIs and E-EMRIs, focusing on their spectral imprint in the LISA frequency band. Even under conservative assumptions, the minimum expected GWF signal from the Galactic Center remains a significant contributor, potentially acting as a source of confusion noise near the $\sim 1\,\mathrm{mHz}$ region. Moreover, while the forest signals originating from the nearby galaxy NGC4303 are comparatively weaker due to its larger distance, they are not negligible and may still contribute to the overall low-frequency background. These results highlight the importance of incorporating both local and extragalactic extreme mass-ratio inspirals when modeling the unresolved gravitational-wave background in the LISA era.

Acknowledgments

This work was supported by National Key Research and Development Program of China (Grant No. 2021YFC2203003), the National Natural Science Foundation of China (Grants No. 12247101), the 111 project under (Grant No. B20063). We gratefully acknowledge the use of high perfomance computers at the Supercomputing Cneter of Lanzhou University and State Key Laboratory of Scientific and Engineering Computing, CAS. The support provided by China Scholarship Council

(CSC) during a visit to València is acknowledged.

References

[1] Amaro-Seoane, P., Relativistic dynamics and extreme mass ratio inspirals. *Living Reviews in Relativity*, 21(1):4, 2018.

[2] Amaro-Seoane, P., Extremely large mass-ratio inspirals. *Physical Review D*, 99(12):123025, 2019.

[3] Amaro-Seoane, P., Gair, J.R., Freitag, M., Coleman Miller, M., Mandel, I., Cutler, C.J., and Babak, S., Intermediate and extreme mass-ratio inspirals—astrophysics, science applications and detection using lisa. *Classical and Quantum Gravity*, 24(17):R113, 2007.

[4] Amaro-Seoane, P., Sopuerta, C.F., and Dewi Freitag, M., The role of the supermas-sive black hole spin in the estimation of the emri event rate. *Monthly Notices of the Royal Astronomical Society*, 429(4):3155–3165, 2013.

[5] Amaro-Seoane, P., Audley, H., Babak, S., Baker, J., Barausse, E., Bender, P., Berti, E., Binetruy, P., Born, M., Bortoluzzi, D., et al. Laser Interferometer Space Antenna. *arXiv preprint arXiv:1702.00786*, 2017.

[6] Amaro Seoane, P., Lin, Y., and Tzanavaris, K., Mono-and oligochromatic extreme-mass-ratio inspirals. *Physical Review D*, 110(6):064011, 2024.

[7] Babak, S., and Petiteau, A., LISA Data Challenge Manual. Technical Report LISA-LCST-SGS-MAN-002, APC, Paris, 2020.

[8] Finn, L.S., and Thorne, K.S., Gravitational waves from a compact star in a circular, inspiral orbit, in the equatorial plane of a massive, spinning black hole, as observed by lisa. *Physical Review D*, 62(12):124021, 2000.

[9] Fishbach, M. and Holz, D.E., Where are ligo's big black holes? *arXiv preprint arXiv:1709.08584*, 2017.

[10] Fishbach, M., Doctor, Z., Callister, T., Edelman, B., Ye, J., Essick, R., Farr, W.M., Farr, B., and Holz, D.E., When are ligo/virgo's big black hole mergers? *The Astrophysical Journal*, 912(2):98, 2021.

[11] Gourgoulhon, E., Le Tiec, A., Vincent, F.H., and Warburton, N., Gravita-tional waves from bodies orbiting the galactic center black hole and their detectability by lisa. *Astronomy & Astrophysics*, 627:A92, 2019.

[12] Luo, J., Chen, L.S., Duan, H.Z., Gong, Y.G., Hu, S., Ji, J., Liu, Q., Mei, J., Milyukov, V., Sazhin, M., et al. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, 2016.

[13] Pastorini, G., Marconi, A., Capetti, A., Axon, D.J., Alonso-Herrero, A., Atkinson, J., Batcheldor, D., Marcella Carollo, C., Collett, J., Dressel, L., et al. Supermassive black holes in the sbc spiral galaxies ngc 3310, ngc 4303 and ngc 4258. *Astronomy & Astrophysics*, 469(2):405–423, 2007.

[14] Carl Peters, P.. Gravitational radiation and the motion of two point masses. *Physical Review*, 136(4B):B1224, 1964.

[15] Ruan, W.H., Guo, Z.K., Cai, R.G., and Zhang, Y.Z., Taiji program: Gravitational-wave sources. *International Journal of Modern Physics A*, 35(17):2050075, 2020.

[16] Amaro Seoane, P. and Zhao, S.D., A forest of gravitational waves in our galactic centre. *arXiv preprint arXiv:2504.20147*, 2025.

[17] Vázquez-Aceves, V., Lin, Y., and Torres-Orjuela, A. Sgr a* spin and mass estimates through the detection of an extremely large mass-ratio inspiral. *The Astrophysical Journal*, 952(2):139, 2023.

Part VIII

Uncertainty Quantification and Modelling

A Study of a Random Fractional Differential Equation with jumps

C. Burgos[♭],[1] A. Navarro-Quiles, [♭],A. Sánchez-Sánchez, [♭] and L. Villafuerte[♮]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) Department of Mathematics, University of Texas at Austin,
2515 Speedway, PMA 8.100 Austin, TX 78712.

1 Introduction

Differential equations with jumps is powerful tool for modeling systems that exhibit sudden changes or discontinuities, which commonly arise in fields such as finance, physics, biology, and engineering. These jumps can represent abrupt events like shocks, failures, or external interventions that cannot be captured by classical differential equations.

In many realistic scenarios, these systems also exhibit memory effects or anomalous diffusion, which are better described using fractional derivatives. The combination of jumps and fractional dynamics provides a rich modeling framework capable of capturing both discontinuous behaviors and long-range dependencies.

In realistic scenarios, the input data or parameters involved in these systems are often uncertain or subject to randomness. This motivates the study of fractional differential equations with jumps considering that its parameters are described by random variables instead deterministic values. Incorporating randomness into the model allows for a deeper understanding of the solution behavior and the impact of uncertainty on system dynamics.

This work addresses the study of the following fractional differential equation with jumps, considering random variables as inputs:

$$\left.\begin{array}{rcl} \left({}^{C}D^{\alpha}X\right)(t) + AX(t) & = & \displaystyle\sum_{j=1}^{M} C_j \delta(t-t_j), \quad 0<\alpha<1, \quad t>0, \\ x(t_0) & = & X_0, \end{array}\right\} \tag{1}$$

where X_0, A , and C_j $(j = 1, \ldots, M)$ are random variables (RVs) defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The right-hand side represents a sequence of impulses given at discrete time instants, satisfying $t_0 < t_1 < \cdots < t_M$. In particular, $\delta(t - t_i)$ represents the Dirac delta function centered at $t - t_i$, defined as:

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ \infty, & x = 0. \end{cases}$$

[1] clabursi@mat.upv.es

The term $\left({}^{C}D^{\alpha}X\right)(t)$ denotes the mean square fractional derivative in Caputo sense of order $\alpha \in]0,1[$ of the stochastic process (sp) $x(t)$ defined by [1].

$$\left({}^{C}D^{\alpha}X\right)(t) = \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-\tau)^{-\alpha}X'(\tau,\omega),\mathrm{d}\tau, t>0$$

In this work, we will obtain the solution of the fractional IVP (1) using the Laplace Transform (LT) for second order sp, which is studied in [2]. Given a second order stochastic process, $X(t)$, its LT is given by:

$$\mathcal{L}\{X(t);s\} = \int_0^{\infty} e^{-st}X(t)\mathrm{d}t, \quad s \in \mathcal{S} \subset \mathrm{R}. \tag{2}$$

The existence of this integral is guaranteed if $\exists s_0 \geq 0$ and $\exists M>0:\ ||X(t||_2 \leq Me^{s_0 t}, \forall t \geq T_0 \geq 0$, [2, Proposition 2.4].

2 Obtaining the solution

To obtain the solution of (1), it is necessary to obtain the LT of the Caputo derivative where $\alpha \in]0,1[$. Taking into account the same procedure described in [2], one gets:

$$\mathcal{L}\{\left({}^{C}D^{\alpha}_{0^+}X\right)(t);s\} = s^{\alpha}\mathcal{L}\{X(t);s\} - s^{\alpha-1}X(0) \tag{3}$$

Now, we apply the LT in the differential equation and we isolate $\mathcal{L}\{X(t);s\}$ obtaining:

$$\mathcal{L}\{X(t);s\} = X_0\frac{s^{\alpha-1}}{s^{\alpha}+A} + \sum_{j=1}^{M} C_j\frac{e^{-t_j s}}{s^{\alpha}+A} \tag{4}$$

Taking into account that the geometric series $\sum_{n=0}^{\infty}x^n$ converges to $\frac{1}{1-x}$ when $|x|<1$ and assuming that A is a bounded RV i.e. exist positive constants $M_A>0$ such that $|A| \leq M_A$, for $s > M_a^{1/\alpha}$ we can rewrite (4) as

$$\mathcal{L}\{X(t);s\} = X_0\sum_{n=0}^{\infty}(-1)^n\frac{A^n}{s^{\alpha(n+1)}} + \sum_{j=1}^{M}C_j e^{-t_j s}\sum_{n=0}^{\infty}(-1)^n\frac{A^n}{s^{\alpha(n+1)}} \tag{5}$$

It is well known that the solution of the IVP (1) has to verify that its Laplace transform match with (5). The following Theorem guarantees it:

Theorem 2.1 *Let be A a bounded random variable. Let us consider that X_0 and C_j (j=1,...,M) are second order random variables, i.e. $||X_0||_2 < \infty$ and $||X_j||_2 < \infty$. We will assume that A, x_0 and C_j $(j=1,\ldots,M)$ are mutually independent. Then, the stochastic process given by*

$$X(t) = X_0\sum_{n=0}^{\infty}\frac{(-1)^n A^n}{\Gamma(\alpha n+1)}t^{\alpha n} + \sum_{j=1}^{M}C_j H(t-t_j)\sum_{n=0}^{\infty}\frac{(-1)^n A^n}{\Gamma(\alpha(n+1))}(t-t_j)^{\alpha(n+1)-1}, \tag{6}$$

is mean square solution of the RIVP (1) for all $t>0$. The function $H(t-t_j)$ is the Heaviside step function defined as

$$H(t-t_j) := \begin{cases} 0, & t < t_j, \\ 1, & t \geq t_j. \end{cases}$$

3 Mean and Second Order Moment

Once a convergent solution in m.s. sense is obtained, we can obtain convergent approximations for the mean and for the second order moment. It is well known, that if a sequence of RVs converge in m.s. its mean and its variance also converge to the mean and second order moment of the limit, [3, Theorem 4.2.1]. Defining that sequence as a truncation of the solution,

$$X_N(t) = X_0 \sum_{n=0}^{N} \frac{(-1)^n A^n}{\Gamma(\alpha n+1)} t^{\alpha n} + \sum_{j=1}^{M} C_j H(t-t_j) \sum_{n=0}^{N} \frac{(-1)^n A^n}{\Gamma(\alpha(n+1))} (t-t_j)^{\alpha(n+1)-1}, \tag{7}$$

we can guarantee that the mean and the second order moment of (7) converges to the mean and the second order moment of (6). They are given by

$$\begin{aligned} \mathbb{E}[X_N(t)] = \mathbb{E}[X_0] & \sum_{n=0}^{N} \frac{(-1)^n \mathbb{E}[A^n]}{\Gamma(\alpha n+1)} t^{\alpha \cdot n} \\ & + \sum_{j=1}^{M} \mathbb{E}[C_j] H(t-t_j) \sum_{n=0}^{N} \frac{(-1)^n \mathbb{E}[A^n]}{\Gamma(\alpha(n+1))} (t-t_j)^{\alpha(n+1)-1} \end{aligned} \tag{8}$$

$$\begin{aligned} E[X_N^2(t)] = & \sum_{n=0}^{N} \sum_{p=0}^{N} \frac{(-1)^{n+p}}{\Gamma(\alpha n+1)\Gamma(\alpha p+1)} t^{\alpha(n+p)} E[A^{n+p}] E[X_0^2] \\ & + 2 \sum_{j=1}^{M} H(t-t_j) \sum_{n=0}^{N} \sum_{p=0}^{N} \frac{(-1)^{n+p}}{\Gamma(\alpha n+1)\Gamma(\alpha(p+1))} t^{\alpha n} (t-t_j)^{\alpha(p+1)-1} E[A^{n+p}] E[C_j] E[X_0] \\ & + \sum_{k=1}^{M} \sum_{j=1}^{M} H(t-t_k) H(t-t_j) \sum_{n=0}^{N} \sum_{p=0}^{N} \frac{(-1)^{n+p}}{\Gamma(\alpha(n+1))\Gamma(\alpha(p+1))} \\ & \cdot (t-t_k)^{\alpha(n+1)-1} (t-t_j)^{\alpha(p+1)-1} E[A^{n+p}] E[C_k, C_j] \end{aligned} \tag{9}$$

4 Examples

This section is devote to illustrate the theoretical findings with a numerical example. Let us consider that, the order of the derivative is $\alpha = 0.5$, X_0 is a Gaussian random variable $N(0; 0.1)$, $C_1 \sim Be(2,3)$ and $C_2 \sim Be(3,4)$. Figure 1 represents approximations for the mean and for the standard deviation considering different order of truncations. We can see that as M increases the mean and the standard deviation converges.

5 Conclusions

This work investigates a random fractional differential equation with jumps. Approximations for the mean and standard deviation have been derived. The theoretical findings are supported by a numerical example that illustrates the applicability of the proposed approach.

Acknowledgments

This work has been supported by

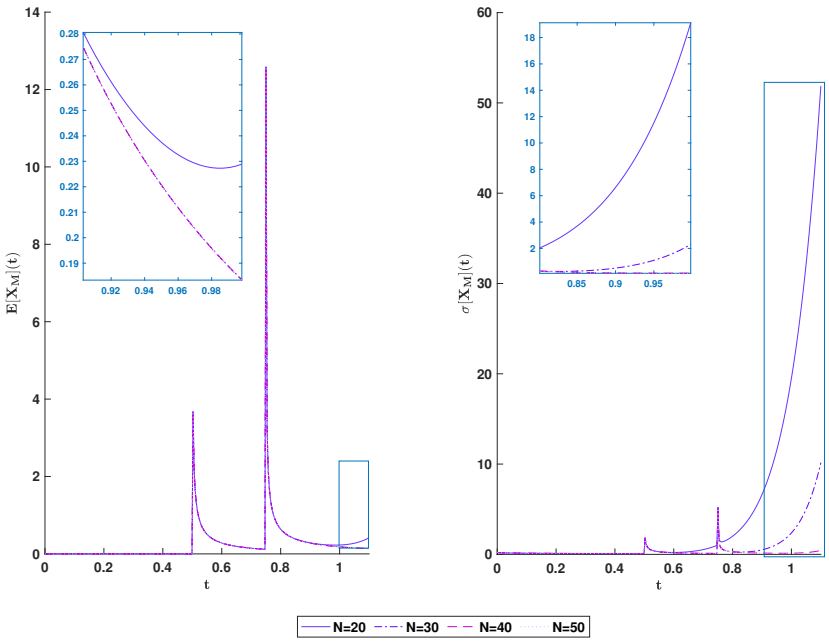


Figure 1: Mean and Standard distribution with the data described in Section 4

- The Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER UE) (Grant PID2020-115270GB-I00).
- Project CIGE/2023/79 (Generalitat Valenciana)

References

[1] Burgos, C., Cortés, J. C., Villafuerte, L. and Villanueva, R. J., Extending the deterministic Riemann–Liouville and Caputo operators to the random framework: A mean square approach with applications to solve random fractional differential equations. *Chaos, Solitons & Fractals,* 102, 305-318, 2017.

[2] Burgos, C., Cortés, J. C., Villafuerte, L. and Villanueva, R. J., Solving random fractional second-order linear equations via the mean square Laplace transform: Theory and statistical computing. *Applied Mathematics and Computation*, 418: 126846, 2022.

[3] Soong, T. T., Random Differential Equations in Science and Engineering. Academic Press, 1973.

An Approach to Simulating the Dynamics of Stochastic Urban Spatial Network Models

H. de la Cruz[♭],[1] and M. Munoz[♭]

(♭) School of Applied Mathematics. FGV-EMAp
Botafogo, Rio de Janeiro. Brazil

1 Introduction

Urban networks play a crucial role within urban systems, influencing their functionality and shaping the dynamics of urban areas. These networks are marked by their complexity, interdependence, and dynamic characteristics. Therefore, constructing, analyzing, and understanding urban network models is essential for tackling complex urban challenges, promoting sustainable development, and enhancing the overall quality of life in cities and regions. In this work, we present a stochastic spatial-interaction model of urban structures, examining its properties in detail. Additionally, we develop an effective time-stepping integrator that enables us to analyze the evolution of this stochastic system over large time intervals. This integrator provides valuable insights into the dynamic behavior of the system as it reaches equilibrium. We also carried out numerical simulation studies to demonstrate the practical effectiveness of the proposed methodology.

References

[1] de la Cruz, H., and M. Munoz. A conjugate method for simulating the dynamics of stochastic urban spatial network models. *Chaos, Solitons and Fractals*, 187: 115430, 2024.

[2] Ellam, L., et al. Stochastic modelling of urban structure. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474.2213: 20170700, 2018.

[3] Wilson, A. G. Complex spatial systems: Challenges for modellers. *Mathematical and computer modelling* 36.3: 379-387, 2002.

[4] White, R., Guy, E., and Inge, U., Modeling cities and regions as complex systems: From theory to planning applications. MIT Press, 2015.

[1] hugo.delacruz@fgv.br

Stochastic differential calculus for Gaussian and non-Gaussian white noises

R. Laudani[♭],[1] and G. Falsone[♭]

(♭) Dipartimento di Ingegneria, Università degli Studi di Messina,
C. da Di Dio, 98166, Messina, Italy.

1 Introduction

Impulsive events can be observed in a variety of natural and engineering systems. The need to use stochastic processes to model many engineering problems, such as loading resulting from earthquakes, turbulent winds, or ocean waves, has been recognized in recent years. Therefore, an adequate probabilistic characterization of the structural response must be performed. In these circumstances, however, the excitations are significantly non-normal, and also, in the simplest case of linear systems, the response is non-normal. This need inspired the development of stochastic differential calculus (SDC). In this sense, an important contribution was made by Itô [1], which introduced the Itô calculus with the famous Itô differential rule. The basic concept of the Itô calculus is the idea that the response of any nonlinear differential equation enforced by a normal white noise can be improved by the introduction of the second term in the Taylor expansion series. However, due to the fact that stochastic integration rules are involved, the solutions of the stochastic differential equations (SDE) may not be unique. Moreover, the Itô approach fails when the input is non-normal. Anyway, the Itô calculus is relatively simple to apply, and in the applications, it is preferred over another stochastic integral, called the Stratonovich integral. In this regard, it was demonstrated that it is always possible to use the application advantages of the Itô calculus, obtaining the same results as the Stratonovich one, by adding some corrective terms in the SDE, the so-called Wong–Zakai correction terms [2].

Over the years, the potential of applying the Itô differential calculus to the case of non-Gaussian white noise processes (such as delta-correlated processes or Lévy white noises) was investigated by several scholars [3, 4, 5, 6]. The generalization of the Gaussian white noise in the non-Gaussian field can be represented by delta-correlated Poisson processes. Similarly, the generalization of the Brownian motion is represented by the compound Poisson motion. Therefore, taking into account that the delta-correlated process is the formal derivative of the compound process, in 1993 Di Paola and Falsone [7] extended to these non-Gaussian processes, the concept of correction terms. This work will show the application of the correction term, which allows one to obtain very accurate results, even when the Itô-type integral is used.

[1] rlaudani@unime.it

2 Preliminary concepts and definitions

In this section, some peculiar concepts of the Brownian motion (Wiener process) and the compound Poisson process are recalled to better understand the application of the stochastic differential calculus. These motion processes have the characteristic of being connected to the white noise process through a generalized (formal) time derivative. Specifically, it is well-known that the normal white noise processes, $W(t)$, and the Poissonian one, $\Phi(t)$, can be considered as the formal derivative of the Brownian motion process $B(t)$, and of the Compound process $C(t)$, respectively. That is for normal white noise processes, $W(t)$:

$$W(t) = \frac{dB(t)}{dt}; \tag{1a}$$

$$B(b) - B(a) = \int_a^b W(t)\, dt. \tag{1b}$$

As well as, for the Poisson delta correlated processes, $\Phi(t)$, can be considered as the formal derivatives (in the generalized sense) of the Poisson compound processes, $C(t)$, that is:

$$\Phi(t) = \frac{dC(t)}{dt}; \tag{2a}$$

$$C(b) - C(a) = \int_a^b \Phi(t)\, dt. \tag{2b}$$

The common characteristic between $B(t)$, and $C(t)$ is that they have independent increments. Nonetheless, if increments of the process $B(t)$ and $C(t)$ are compared, it is possible to appreciate that they are of different orders. In fact, the sample functions of the Poisson process and normal white noise are very different because, in the first case, finite-amplitude pulses occur in a well-defined space, whereas in the second case, infinitesimal-amplitude pulses appear in an extensive temporal space. Specifically, the zero-mean normal white noise, $W(t)$ is characterized in probabilistic sense by the second-order moment, that is:

$$R_W^{(2)}(t_0, t_1) = k_2 [W(t_0) W(t_1)] = q_2(t_0)\, \delta(t_1 - t_0) \tag{3}$$

where $q_2(\cdot)$ is a definite positive function defining the intensity of the white noise and $\delta(\cdot)$ is the Dirac delta function. Thus, the increment of the Brownian motion $dB(t)$, is described by

$$E[dB(t)] = 0; \tag{4a}$$

$$E[dB(t_1)\, dB(t_2)] = q_2 dt \tag{4b}$$

Indeed, as the order of differentials in the first and second member of Eq. 4(b) must be equal, then the first differential $dB(t)$ has the order of $dt^{\frac{1}{2}}$. While, a sample of the compound Poisson process $C(t)$ has the following form:

$$C(t) = \sum_{j=0}^{N(t)} Y_j U(t - t_j) \tag{5}$$

where $N(t)$ is a Poisson process and Y_j, with $j = 1, 2, \ldots$, is a family of independent random variables with equal distribution. Then, an increment $C(t)$ is characterized by:

$$E[(dC(t))^k] = \lambda E\left[Y^k\right] dt \quad \forall k > 2 \tag{6}$$

From this last equation, it is clear that every moment of the increment $dC(t)$ are infinitesimal of the first order, dt. Hence, the Eq. 4(b) and Eq. 6 evidence a very important property of the stochastic differential calculus.

3 Stochastic differential calculus and correction terms

This section provides a summary of the key concepts of the SDC for a first-order oscillator that is excited by a parametric Gaussian white noise excitation and by a parametric Poisson delta-correlated process.

3.1 Gaussian white noise parametric excitation

Let the first-order dynamical system be given in the form:

$$du\left(t\right) = a\left(u,t\right)dt + g\left(u,t\right)dB(t) \tag{7}$$

where $a(\cdot)$ and $g(\cdot)$ are functions, in general non-linear, of the response $u\left(t\right)$, $B(t)$ is the Brownian motion process related to the parametric Gaussian white noise $W(t)$. It is possible to provide that the increment and the differential of a generic continuous and differentiable function, $\varphi\left(u\left(t\right),t\right)$ can be different (see [8]). Indeed, the following relationship can be written:

$$\Delta\varphi(u(t),t) = d\varphi\left(u\left(t\right),t\right) + \frac{1}{2!}d^2\varphi\left(u\left(t\right),t\right) + \frac{1}{3!}d^3\varphi\left(u\left(t\right),t\right) + \cdots \tag{8}$$

From the above expression, it is possible to appreciate that non-neglectable quantities are also given by the second term of the previous series, if the system is subject to Gaussian white noise parametric forces. Indeed:

$$d\varphi(u(t),t) = \frac{\partial\varphi(u(t),t)}{\partial t}dt + \frac{\partial\varphi(u(t),t)}{\partial u}\,du(t) \tag{9a}$$

$$d^2\varphi\left(u\left(t\right),t\right) = \frac{\partial^2\varphi\left(u\left(t\right),t\right)}{\partial u^2}\left(du(t)\right)^2 + \frac{\partial\varphi(u(t),t)}{\partial u}\,d^2u(t) \tag{9b}$$

From Eq. 7, the quantity $d^2u(t)$ can be considered as:

$$d^2u\left(t\right) = d\left(du(t)\right) = d\left(a\left(u,t\right)dt + g\left(u,t\right)dB(t)\right) \tag{10}$$

which provides the following not-neglectable term of order dt:

$$\frac{\partial g\left(u,t\right)}{\partial u}\left(dB(t)\right)^2 \tag{11}$$

Therefore, Eq. 8 can be simplified as follows:

$$\Delta\varphi(u(t),t) = \frac{\partial\varphi(u(t),t)}{\partial t}dt + \frac{\partial\varphi(u(t),t)}{\partial u}du + \frac{1}{2}\frac{\partial g(u,t)}{\partial u}(dB(t))^2 \tag{12}$$

This last equation represents the so-called Itô differential rule. In comparison with the differential rule of continuous functions, it is differentiated by the presence of the last term. It is important to remark that the other second-order terms of the series have been ignored because they are either zero or being infinitesimals of lesser order than dt. Now, by setting $\varphi\left(u(t)\right) = u\left(t\right)$ in Eq. 8 and neglecting the terms of greater order than the second one, the following expression is obtained:

$$\Delta u(t) = du(t) + \frac{1}{2}d^2u(t) = a(u,t)dt + g(u,t)dB(t) + \frac{1}{2}\frac{\partial g(u,t)}{\partial u}(dB(t))^2 \tag{13}$$

In literature, the last term of Eq. 13 is referred to as the Wong-Zakai correction term.

3.2 Poisson white parametric excitation

Now, let us consider the same system examined in the previous section, but driven by a parametric Poisson white process. Hence, the generalization of Eq. 7 can be written as follows:

$$du\left(t\right)=a\left(u,t\right)dt+g\left(u,t\right)dC(t) \tag{14}$$

where $C(t)$ is the Poisson compound process correspondent to the white Poisson input process, $\Phi(t)$. As the increments of $dC(t)$ are of the order of dt, this implies that all the terms of the series of Eq. 8 give an infinitesimal contribution of order dt to the increment $\Delta\varphi(u)$. Hence, even if, theoretically, the number of the series terms should be extended up to ∞, practically, a closure of order n can be applied, that is:

$$\Delta\varphi(u(t),t)=d\varphi\left(u\left(t\right),t\right)+\sum_{i=2}^{n}\frac{1}{i!}d^{i}\varphi\left(u\left(t\right),t\right) \tag{15}$$

Then, by neglecting the infinitesimal of order greater than dt and after some algebra, the following expression of the differential rule holds:

$$\Delta\varphi(u(t),t)=\frac{\partial\varphi(u(t),t)}{\partial t}dt+\sum_{i=1}^{n}\frac{\partial^{i}\varphi(u(t),t)}{\partial u^{i}}(\Delta u(t))^{i} \tag{16}$$

By particularizing Eq. 16, and after some algebra, the response increment $\Delta u(t)$ is obtained by:

$$\Delta u(t)=a\left(u\left(t\right),t\right)dt+g\left(u\left(t\right),t\right)dC(t)+\sum_{i=2}^{n}\frac{1}{i!}g^{(i)}\left(u\left(t\right),t\right)\left(dC(t)\right)^{i} \tag{17}$$

where:

$$g^{(1)}\left(u\left(t\right),t\right)=g\left(u\left(t\right),t\right);\quad g^{(i)}\left(u\left(t\right),t\right)=\frac{\partial g^{(i-1)}\left(u\left(t\right),t\right)}{\partial u}g\left(u\left(t\right),t\right) \tag{18}$$

From Eq. 17, it is possible to see the correction term, given by:

$$\Gamma_{C}^{(n)}(t)=\sum_{i=2}^{n}\frac{1}{i!}g^{(i)}\left(u\left(t\right),t\right)\left(dC(t)\right)^{i} \tag{19}$$

Importantly, every term in this sum is an infinitesimal of order dt. Additionally, it is simple to see that the first term of the sum ($i=2$) coincides with the Wong-Zakai one. At last, the response increment $\Delta u(t)$ can be rewritten as:

$$\Delta u(t)=du(t)+\Gamma_{C}^{n}(t) \tag{20}$$

Indeed, even if all the integrals are solved in the Itô sense, accurate results can be obtained by taking into account the terms $\Gamma_{C}^{(n)}$ given a sufficient value of n.

It is possible to state the correction term in Eq. 20 as the generalization to the Poisson white noise actions of the Wong-Zakai terms.

4 Numerical results

In this section, the following quasi-linear system driven by a parametric Poisson white noise process, $\Phi(t)$, is considered:

$$\dot{u}(t)=\zeta+\gamma u(t)\Phi(t);\quad u(0)=0\quad w.p.1 \tag{21}$$

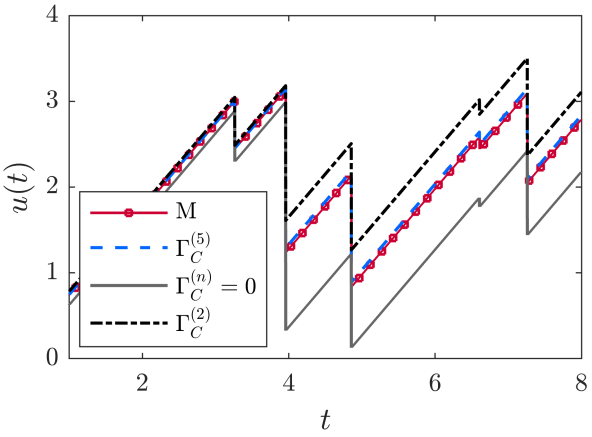


Figure 1: Response sample of the quasi-linear system to Poisson white noise input process. Marked line: Marcus integration solution (M); dashed line: Itô integration solution with corrective coefficients, $\Gamma_C^{(5)}$; solid line: Itô integration result ($\Gamma_C^{(n)} = 0$); dot-dashed line: Stratonovich integration result ($\Gamma_C^{(2)}$).

with deterministic initial conditions. The time-independent parameters ζ and γ, are assumed equal to 1 and -1, respectively. The Poisson white noise input process $\Phi(t)$, is characterized by the mean number of impulses per unit of time $\lambda = 1$, and the random variable Y_j is uniformly distributed in the range $[0, 1]$. The Eq. 21 can be formally replaced by the following form:

$$du(t) = \zeta dt + \gamma u(t) dC(t); \quad u(0) = 0 \tag{22}$$

where $C(t)$ is the Poisson compound processes. By applying the Eq. 19, the expression of the response increment $\Delta u(t)$ becomes:

$$\Delta u(t) = du(t) + \Gamma_C^{(n)} = \zeta dt + \gamma u(t) dC(t) + \Gamma_C^{(n)} \tag{23}$$

where the expression of the correction term, $\Gamma_C^{(n)}$, can be easily derived as following:

$$\Gamma_C^{(n)} = \sum_{i=2}^{n} \frac{1}{i!} g(u(t), t)^i \left(dC(t)\right)^i = \sum_{i=2}^{n} \frac{1}{i!} \gamma^j u(t) \left(dC(t)\right)^i \quad \text{with} \quad g(u(t), t)^i = \gamma^i u(t) \tag{24}$$

The results of applying Eq. 23 with $n = 5$ terms of $\Gamma_C^{(n)}(t)$ are illustrated in Fig. 1. In order to assess the accuracy, the sample example function of response, $u(t)$, is compared with those obtained by integrating Eq. 21 through the Marcus integral resolution (see more details for the application in [9]). These two obtained results perfectly overlap. The same figure also reports the Itô result, which means that the series terms are not taken into account ($\Gamma_C^{(n)} = 0$), and the situation when only the first term of the sum of the series is taken into consideration, i.e $\Gamma_C^{(2)}$. In particular, the outcomes of this latter method are consistent with those of Stratonovich integration. It is evident from this figure that these two last examples exhibit quite different response sample function results.

In addition, in Fig 2 the response moments up to the fourth order are shown and the above comparison has been made too. These analyses generally show that the results of an Monte Carlo simulation (with 15000 samples) obtained by the integration of Eq. 21 through the application of Marcus equation coincide with the moments' response results if the correction term series, given in Eq. 24, is taken into consideration.

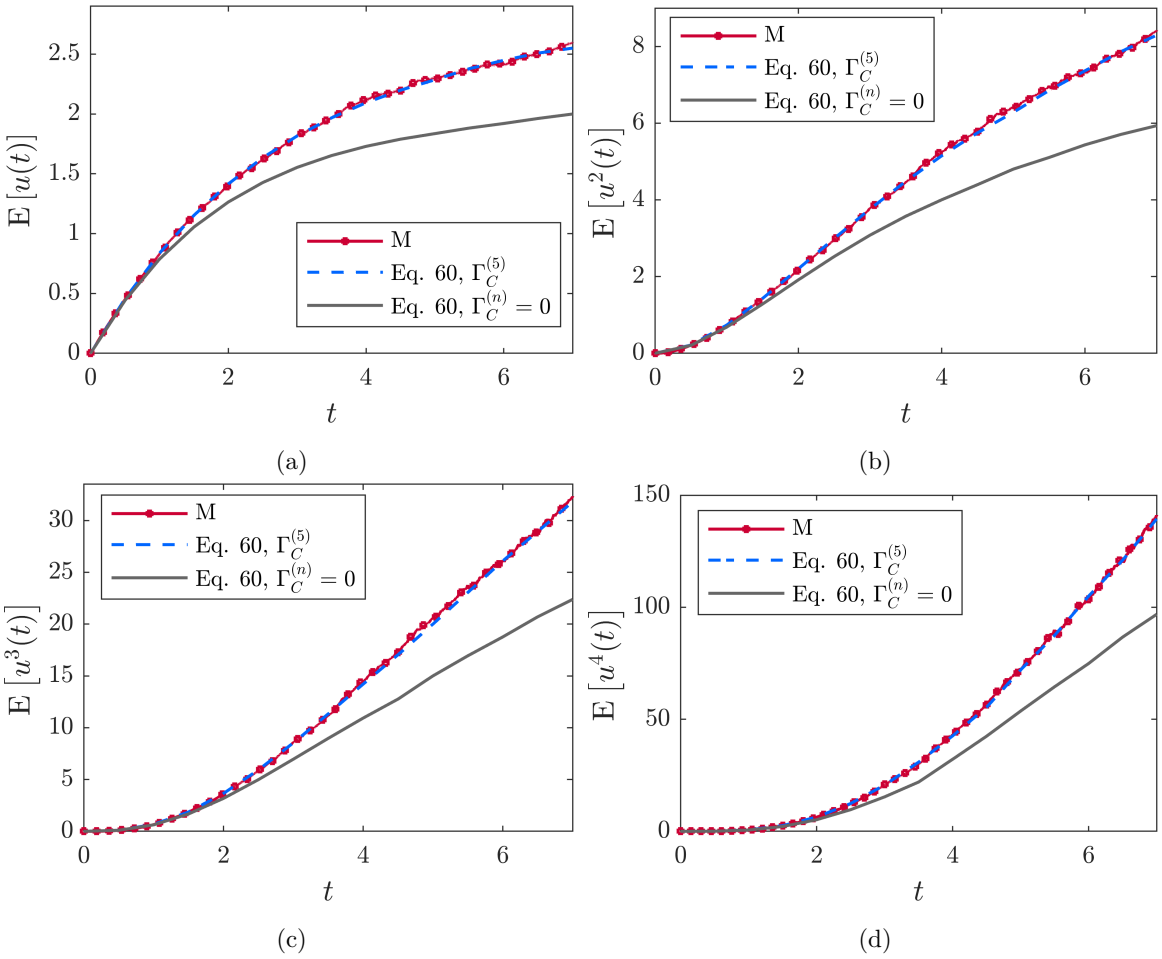


Figure 2: Response moments of the quasi-linear system to Poisson white noise parametric excitation. Marked line: Marcus integration solution (M); dashed line: Itô integration solution with corrective coefficients, $\Gamma_C^{(5)}$; Itô integration result, ($\Gamma_C^{(n)} = 0$); (a) Response mean; (b) Response second-order moment; (c) Response third order-moment; (d) Response fourth-order moment.

5 Conclusions

The application of the SDC for the evaluation of the response of the system driven by two types of parametric white noise processes (normal, Poissonian) has been reported. Both processes exhibit an impulse occurrence at each time interval, but the main difference between them is in the order of the increment. For Gaussian white noise, the increment is of the order of $dt^{\frac{1}{2}}$. While for the Poisson white noise, the increment order is of dt. In this work, the influence of the correction terms on the response by integration in the Itô-type sense has been addressed. Specifically, the single sample response system and the response moments up to the fourth order, obtained by means of MCS, confirm the validity of the use of the proposed correction terms for parametric non-normal input processes.

References

[1] Itô, K., 109. stochastic integral *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

[2] Wong, E. and Zakai, M., On the convergence of ordinary integrals to stochastic integrals *The Annals of Mathematical Statistics*, 36(5):1560–1564, 1965.

[3] Ken-Iti, S., Lévy processes and infinitely divisible distributions. Cambridge university press, 1999.

[4] Proppe, C., The Wong–Zakai theorem for dynamical systems with parametric Poisson white noise excitation *International journal of engineering science*, 40(10):1165–1178, 2002.

[5] Sun, X., Duan, J. and Li, X., An alternative expression for stochastic dynamical systems with parametric Poisson white noise *Probabilistic Engineering Mechanics*, 32:1–4, 2013.

[6] Song, Y., Xu, W. and Niu, L., Multiplicative Lévy noise-induced transitions in gene expression *Science China Technological Sciences*, 65(8):1700–1709, 2022.

[7] Di Paola, M. and Falsone, G., Stochastic dynamics of nonlinear systems driven by non-normal delta-correlated processes *Journal of Applied Mechanics*, 60:141-148, 1993.

[8] Di Paola, M. and Falsone, G., Itô and Stratonovich integrals for delta-correlated processes *Probabilistic engineering mechanics*, 8(3-4):, 1993.

[9] Marcus, S. I., Modeling and approximation of stochastic differential equations driven by semimartingales *Stochastics: An International Journal of Probability and Stochastic Processes*, 4(3):223–245, 1981.

Computation and convergence of the approximate 1-PDF for a randomized fractional differential equation with impulsive forcing

C. Burgos[♭], A. Navarro-Quiles[♮1] and A. Sánchez-Sánchez[♭]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) Department of Statistics and Operational Research,
Universitat de València, Dr. Moliner 50, 46100, Burjassot, Spain.

1 Introduction

In this contribution, we study the following fractional-order model:

$$\left.\begin{array}{rcl} \left({}^{C}D^{\alpha}x\right)(t,\omega)+a(\omega)x(t,\omega) & = & \displaystyle\sum_{j=1}^{M} c_j(\omega)\delta(t-t_j), \quad 0<\alpha<1, \quad t>t_0, \\ x(t_0,\omega) & = & x_0(\omega), \end{array}\right\} \tag{1}$$

where the initial condition $x_0(\omega)$, the coefficient $a(\omega)$ and the impulse amplitudes $c_j(\omega)$ ($j = 1, 2, \ldots, M$) are assumed to be absolutely continuous mutually independent random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with Probability Density Functions (PDFs) $f_{x_0}(x_0)$, $f_a(a)$ and $f_{c_j}(c_j)$, $j = 1, 2, ...M$, respectively. The right-hand side in Model (1) represents a finite sequence of impulses applied at discrete time instants $t_0 < t_1 < \cdots < t_M$, each carrying a random intensity given by $c_j(\omega)$, $j = 1, \ldots, M$.

2 Computing the 1-PDF of the approximated solution and convergence analysis

Let be $a(\omega)$ a bounded random variable. Let us consider that $x_0(\omega)$ and $c_j(\omega)$ ($j = 1, \ldots, M$) are second order random variables, i.e. $\mathbb{E}[|x_0(\omega)|^2]^{1/2} < \infty$ and $\mathbb{E}[|c_j(\omega)|^2]^{1/2} < \infty$. Then, the stochastic process given by

$$x(t,\omega) = x_0(\omega)S_0(t,\alpha;a(\omega)) + \sum_{j=1}^{M} c_j(\omega)H(t-t_j)S_j(t,\alpha;a(\omega)), \tag{2}$$

being for $j = 1, \ldots, M$ the series

$$S_0(t,\alpha;a(\omega)) = \sum_{n=0}^{\infty} \frac{(-1)^n a(\omega)^n}{\Gamma(\alpha n+1)}(t-t_0)^{\alpha n}, \quad S_j(t,\alpha;a(\omega)) = \sum_{n=0}^{\infty} \frac{(-1)^n a(\omega)^n}{\Gamma(\alpha(n+1))}(t-t_j)^{\alpha(n+1)-1},$$

[1] ana.navarro@uv.es

is a mean square solution of the randomized initial value problem (1) for all $t > t_0$. In (2), the function $H(t - t_j)$ denotes the Heaviside step function.

From a computational perspective, it is more practical to employ the truncated solution

$$x^N(t,\omega) = x_0(\omega)S_0^N(t,\alpha;a(\omega)) + \sum_{j=1}^{M} c_j(\omega)H(t-t_j)S_j^N(t,\alpha;a(\omega)),$$

where for $j = 1,\ldots,M$, the truncated series are defined as

$$S_0^N(t,\alpha;a(\omega)) = \sum_{n=0}^{N} \frac{(-1)^n a(\omega)^n}{\Gamma(\alpha n+1)}(t-t_0)^{\alpha n}, \quad S_j^N(t,\alpha;a(\omega)) = \sum_{n=0}^{N} \frac{(-1)^n a(\omega)^n}{\Gamma(\alpha(n+1))}(t-t_j)^{\alpha(n+1)-1}.$$

Applying the Random Variable Transformation method, we determine an expression for the 1-PDF of the truncated solution stochastic process

$$f_1^N(x,t) = \int_{\mathcal{D}(a,c)} f_{x_0}\left(\frac{x - \sum_{j=1}^{M} c_j H(t-t_j)S_j^N(t,\alpha,a)}{S_0^N(t,\alpha;a)}\right) f_a(a)f_c(c)\left|\frac{1}{S_0^N(t,\alpha;a)}\right| da dc. \quad (3)$$

In expression (3), $c(\omega)$ represents the collection of M impulsive amplitude random variables $c(\omega) = (c_1(\omega), c_2(\omega), ..., c_M(\omega))$ and given the independence of the random variables $c_1(\omega),\ldots c_M(\omega)$ the joint PDF can be written as the product of the marginals $f_c(c) = f_{c_1}(c_1)f_{c_2}(c_2)...f_{c_M}(c_M)$.

Theorem 1 *Consider the random initial value problem* (1)*, and suppose that:*

- *Random variables $x_0(\omega)$, $a(\omega)$, and $c_j(\omega)$, $j = 1,\ldots,M$ are mutually independent.*
- *The random variable $a(\omega)$ is bounded.*
- *The random variables $x_0(\omega)$ and $c_j(\omega)$, $j = 1,\ldots,M$ are second-order.*
- *The PDF $f_{x_0}(x_0)$ is Lipschitz continuous.*

Then, for each $(x,t) \in \mathbb{R} \times [0,\infty[$ fixed the approximation $f_1^N(x,t)$ converges to the 1-PDF of the exact solution, $f_1(x,t)$, i.e.,

$$\lim_{N\to\infty} f_1^N(x,t) = f_1(x,t), \quad \textit{for each } (x,t) \in \mathbb{R} \times [t_0,\infty[\textit{ fixed},$$

where $f_1(x,t)$ is given by the expression

$$f_1(x,t) = \int_{\mathcal{D}(a,c)} f_{x_0}\left(\frac{x - \sum_{j=1}^{M} c_j H(t-t_j)S_j(t,\alpha,a)}{S_0(t,\alpha;a)}\right) f_a(a)f_c(c)\left|\frac{1}{S_0(t,\alpha;a)}\right| da dc.$$

3 Numerical Example

In this section, we illustrate the theoretical findings with a numerical example. Let us consider that the order of the derivative is $\alpha = 0.5$ and the model has two jumps, i.e. $M = 2$. Random variables $x_0(\omega)$, $a(\omega)$, $c_1(\omega)$ and $c_2(\omega)$ are mutually independent with the following distributions:

$$X_0 \sim N(0;0.1), \quad a(\omega) \sim U([0,2]), \quad c_1(\omega) \sim Be(2,3), \quad c_2(\omega) \sim Be(3,4).$$

These random variables satisfy the assumptions in Theorem 1, then the convergence is assured. Specifically:

- Independence.
- $a(\omega)$ is bounded, since it follows a uniform distribution.
- $x_0(\omega)$ and $c_i(\omega)$, $i = 1, 2$ are second-orden.
- $f_{x_0}(x_0)$ is Lipschitz continuous. This result follows directly from the Mean Value Theorem combined with the observation that the derivative of the normal distribution's probability density function is uniformly bounded over $\mathbb{R}$.

Figure 1 illustrates the approximation of the 1-PDF given in Expression (3) at the instant time $t = 2$ for various truncation levels ($N = \{20, 30, 40, 50\}$). The approximations exhibit clear convergence as the number of terms in the truncated series increases.

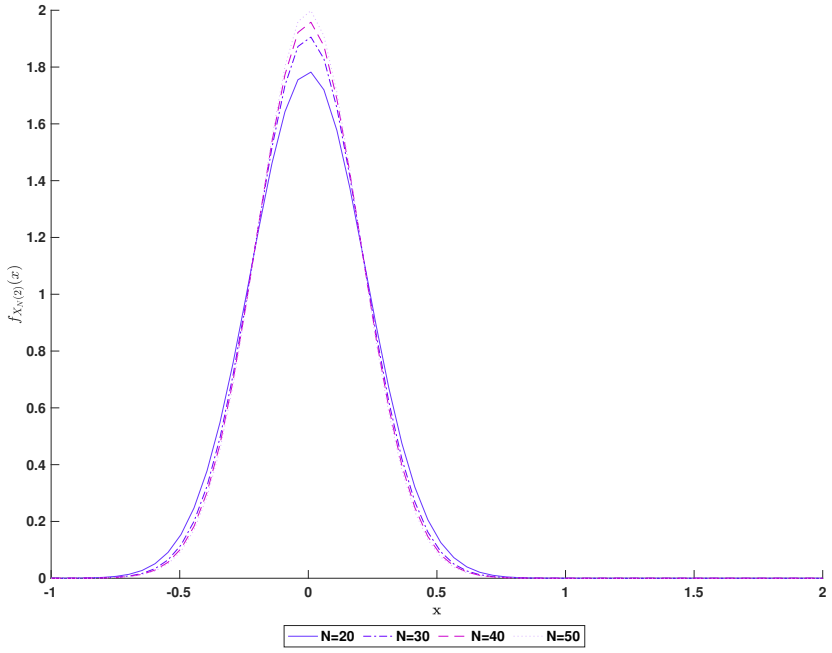


Figure 1: Approximation of the 1-pdf given in Expression (3) at the instant time $t = 2$ with the data described in Section 3, for various truncation levels, $N = \{20, 30, 40, 50\}$.

4 Conclusions

In this work, we addressed the problem of computing the 1-PDF of the solution stochastic process to a randomized fractional differential equation with impulsive forcing. By considering a Caputo-type fractional derivative of order $0 < \alpha < 1$ and assuming random initial condition, coefficient, and impulsive amplitudes, we proposed a methodology to approximate the 1-PDF using a truncated mean square solution. The Random Variable Transformation technique was employed to derive an explicit integral representation of the approximate 1-PDF.

It has been shown that, under certain regularity conditions and independence of the involved random variables, the approximation of the 1-PDF obtained from a truncated solution converges to the exact 1-PDF as the truncation order increases. The proposed methodology provides a practical and theoretically justified tool for analysing randomized fractional differential equations with impulsive effects.

Acknowledgments

This work has been supported by the research project CIGE/2023/79 (Generalitat Valenciana) and The Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER UE) (Grant PID2020-115270GB-I00).

References

[1] Burgos, C., Calatayud, J., Cortés, J.C., Navarro-Quiles, A. A full probabilistic solution of the random linear fractional differential equation via the random variable transformation technique *Mathematical Methods in the Applied Sciences*, 41 (18): 9037-9047, 2018.

[2] Cortés, J.C., Delgadillo-Alemán, S.E., Kú-Carrillo, R.A., Villanueva, R.J. Probabilistic analysis of a class of impulsive linear random differential equations forced by stochastic processes admitting Karhunen-Loève expansions. *Discrete and Continuous Dynamical Systems - S* 15(11): 3131-3153, 2022.

[3] Soong, A., Random Differential Equations in Science and Engineering. Academic Press, 1973.

A Random Differential Equation Approach for Detecting Variations in Growth Patterns of the Atlantic Herring

Jacob Burbank [♮], Juan Carlos Cortés López [♭], Cristina Luisovna Pérez[♭1], and Rafael Jacinto Villanueva[♭]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) Fisheries and Oceans Canada, Gulf Fisheries Centre,
S343 Université Avenue, Moncton, NB E1C 9B6, Canada.

1 Introduction

Understanding the life history of a species, including its growth trajectories, is crucial for comprehending population dynamics and informing fisheries management decisions [8]. The growth of a species is influenced by changing ecosystem conditions such as food availability, temperature, and habitat suitability. The examination and quantification of time-varying growth are often overlooked for many marine species, including small-bodied fish [6] like the Atlantic herring (*Clupea harengus*), which is the focus of this study.

The Southern Gulf of St. Lawrence (sGSL) Atlantic herring, divided into spring and fall spawners [7], is crucial for the ecosystem, serving as a key forage species for predators like Grey Seals, Tuna, and Cod [5, 2, 13] . It is also economically valuable for commercial fisheries. The spring spawning herring is critically low and closed for commercial fishing since 2022. Fall spawning herring is also declining and nearing critical levels [9]. There's a notable decline in growth and length-at-age [4], impacting reproduction and population recovery. Understanding growth and age structure through advanced examinations, supported by annual biological sample collection, can aid in identifying the causes of population decline.

Differential equations provide a valuable framework for the analysis of growth models that exhibit randomness. This study is dedicated to the fitting of growth curves for fall-spawning Atlantic herring. In order to augment the analytical process, we transition from a deterministic methodology to a probabilistic framework by regarding model parameters as random variables in the context of Random Differential Equations. The principle of maximum entropy [10] is employed to deduce the probability distributions of these model parameters from the data, with the objective of investigating temporal changes in their distributions guided by the data itself.

[1] cperdiu@upv.edu.es

2 Methods

2.1 Data

We investigate trends in the length and age of fall-spawning herring in the sGSL from 1988 to 2021 using data from commercial fishing and scientific sampling [12]. Analyzing data every four years, which corresponds to the set of years $G = \{1988, 1992, 1996, 2000, 2004, 2008, 2012, 2016, 2020\}$, our research focuses on capturing generational changes. From the raw data itself, illustrated in Figure 1, one can note a decline in the maximum recorded length of these fish over the specified period.

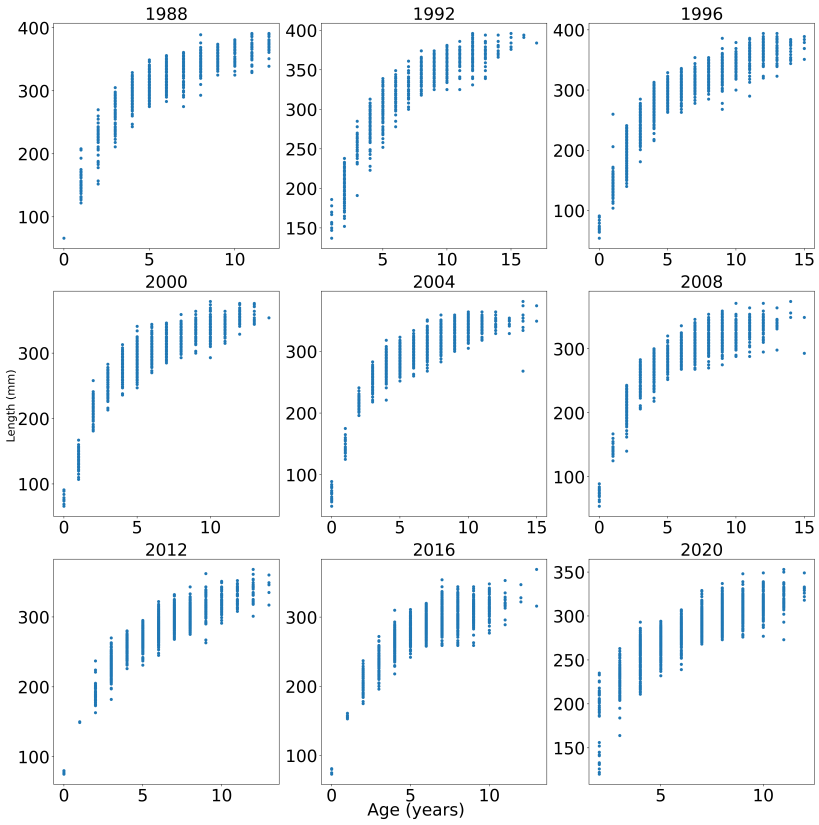


Figure 1: Length (mm) vs Age (years) of fall spawning herrings for years $g \in G$.

2.2 Model

To explore changes in growth patterns over time, we fit growth models to the data from all referenced years. This will help examine variations in model parameters, particularly changes in the upper limit of growth models. By randomizing the classical growth model, we seek to incorporate uncertainty into growth assessments over time.

Suppose that $\boldsymbol{U} = (A, L_0, K)$ is a vector of continuous random variables defined on a common complete probability space $(\Omega, \mathcal{F}_\Omega, \mathbb{P})$ with a joint probability density function $f_{\boldsymbol{u}}$. The model used is in is of the form

$$L(y) = A\left(1 - (1 - L_0/A)\mathrm{e}^{-Ky}\right), \tag{1}$$

Where $L(y)$ is the length of a y-year old fish, L_0 is its length at birth and A is its maximum length. We then obtain the 1-PDF of the stochastic process $L(y)$,

$$f_{L(y)}(l) = \int_{\mathbb{R}^2} f_{\boldsymbol{U}}(a, a - a\mathrm{e}^{ky} + l\mathrm{e}^{ky}, k)\mathrm{e}^{ky}\,\mathrm{d}a\mathrm{d}k. \tag{2}$$

We then use the PME [10] technique to calculate the PDF of $\boldsymbol{U}$. We propose the following PDFs for A, W_0, and K, respectively,

$$f_A(a) = \exp\left(-1 - \lambda_0^A - \lambda_1^A a - \lambda_2^A a^2\right), \quad a \in (200, 400), \tag{3}$$

$$f_{L_0}(l_0) = \exp\left(-1 - \lambda_0^0 - \lambda_1^0 l_0 - \lambda_2^0 l_0^2\right), \quad l_0 \in (4, 200), \tag{4}$$

$$f_K(k) = \exp\left(-1 - \lambda_0^K - \lambda_1^K k - \lambda_2^K k^2\right), \quad k \in (0, 1). \tag{5}$$

We assign to each year of data a non-parametric PDF estimate by leveraging kernel density estimation implemented in the *SciPy* library as *gaussian_kde* [14, 11].
To enhance readability and to underscore the implementation of a model fitted for each successive four-year dataset, we adopt the subsequent nomenclature. Let us denote the model solution for each $g \in G$ as

$$Y^g(t) = A^g(1 - W_0^g/A^g)\mathrm{e}^{-K^g t}, \tag{6}$$

where

$$f_{A^g}(a) = \exp\left(-1 - \lambda_0^{A,g} - \lambda_1^{A,g} a - \lambda_2^{A,g} a^2\right), \quad a \in (200, 400), \tag{7}$$

$$f_{L_0^g}(l_0) = \exp\left(-1 - \lambda_0^{0,g} - \lambda_1^{0,g} l_0 - \lambda_2^{0,g} l_0^2\right), \quad l_0 \in (4, 200), \tag{8}$$

$$f_{K^g}(k) = \exp\left(-1 - \lambda_0^{K,g} - \lambda_1^{K,g} k - \lambda_2^{K,g} k^2\right), \quad k \in (0, 1). \tag{9}$$

We fit the random model (1) with 1-PDF (2) for every $g \in G$ by minimizing the L^2-norm of the data PDF and the model PDF using the *minimize* algorithm from the *SciPy* library [14].

3 Results

After calibrating the model hyper-parameters, we obtain the set $\boldsymbol{\lambda}^g = \{\lambda_1^{A,g}, \lambda_2^{A,g}, \lambda_1^{0,g}, \lambda_2^{0,g}, \lambda_1^{K,g}, \lambda_2^{K,g}\}$ that solves the minimization problem described. We do so for each $g \in G$.

Figure 2 shows the calibration results in 2D for all $g \in G$. The data (blue) is compared to the model mean (orange), with 95% and 99% PIs. Nearly all data points fall within the 99%PI, which isn't excessively wide.

The random model fit allows us to see trends in the distribution of the random parameter A through the years. It also allows us to compare the deterministic and random fits of the models. Several important trends can be observed. Both random and deterministic models exhibit a general downward trend in asymptotic length A from the early 1990s to the 2010s, suggesting that fish are achieving smaller maximum lengths, with potential biological and ecological impacts.

The random model generally gives slightly higher average A estimates than the deterministic model, except in 2012 when both coincide. Notably, all deterministic calibrations of A are within the 95%PIs of the random model, indicating that the random model captures a broader range of outcomes by incorporating data variability and uncertainty, while the deterministic model provides a more limited estimate but both agree on the average trend. The increasing PIs in recent years imply greater uncertainty, possibly due to smaller sample sizes or more growth variability, suggesting the random model reflects growth pattern variability not seen in the deterministic model.

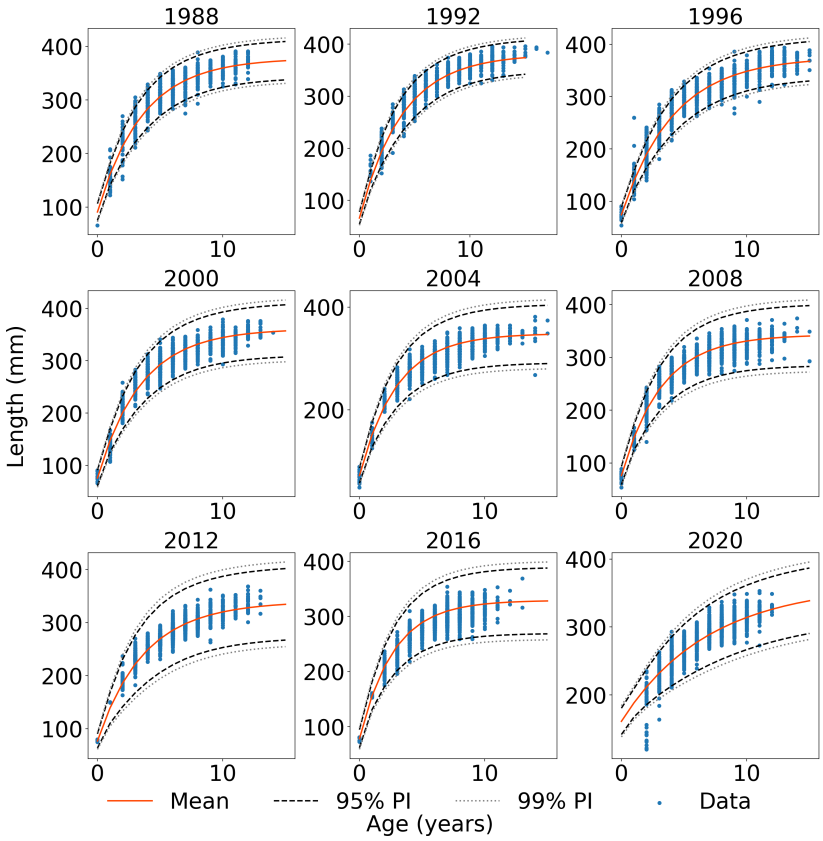


Figure 2: Data, model mean, 95% and 99% PIs, for all $g \in G$.

4 Conclusions

This study models nine years of age-length data for the sGSL Atlantic herring from 1988 to 2020 using a random differential equation, capturing growth trends and randomness. Using the Principle of Maximum Entropy, we estimated probability density functions for the solution and parameters treated as random variables, revealing trends in the asymptote parameter, A. Over time, the distribution of A became more symmetrical and shifted from the upper limit of A ($\text{support}(A) = (200, 400)$), indicating smaller fish sizes over generations. A limitation is treating each model year independently, ignoring temporal dependencies. Modeling A as a time-varying stochastic process could provide a more coherent representation of herring dynamics. Overall, the analysis shows fall-spawning sGSL herring are not growing as large, nor reaching older ages, impacting reproductive output. Increased predation—from Grey Seals and Bluefin Tuna—has raised natural mortality in fish aged 7+, peaking in 2011 [1, 12], driving the observed reduction in older fish. Larger, older herring are crucial for population growth [3], so their decline limits production. Modeling growth as a random process reveals it is less consistent with more variability recently, due to climate-induced temperature changes and increased predation.

Acknowledgments

This work has been supported by the grant PID2020-115270GB–I00 granted by MCIN/AEI/10.13039/501100011033.
Cristina-Luisovna Pérez has been supported by an FPU21/02343 grant (Formación de Profesorado Universitario 2021) from the Ministerio de Universidades.
Jacob Burbank is supported by funding from Fisheries and Oceans Canada. We thank all those involved in the collection of samples and biological data over the years. Acknowledgments of your work.

References

[1] Benoît, H.P. et al., Principal predators and consumption of juvenile and adult Atlantic Herring (*Clupea harengus*) in the southern Gulf of St. Lawrence, DFO Canadian Science Advisory Secretariat, 2016/065, 2016.

[2] Burbank, J. et al., Ontogenetic shifts in isotope-inferred trophic ecology of Atlantic bluefin tuna *Thunnus thynnus* in Atlantic Canada *Marine Ecology Progress Series*, Volume 744: 147–160, 2024. doi:10.3354/meps14669

[3] Burbank, J. et al., Temporal declines in fecundity: A study of southern Gulf of St. Lawrence Atlantic herring (*Clupea harengus*) and implications for potential reproductive output *Journal of Fish Biology*, Volume 105: 279–287, 2024. doi:10.1111/jfb.15784

[4] Burbank, J. et al., Temporal Variation in Von Bertalanffy Growth Curves and Generation Time of Southern Gulf of St. Lawrence Spring and Fall Spawning Atlantic Herring (*Clupea harengus*) *Fishes*, 8(4), 2023. doi:10.3390/fishes8040205

[5] Hammill, M.O. et al., Feeding by grey seals in the Gulf of St. Lawrence and around Newfoundland *NAMMCO Scientific Publications*, Volume 6: 135—152, 2007. doi:10.7557/3.2729

[6] Huang, M. et al., The impacts of climate change on fish growth: A summary of conducted studies and current knowledge *Ecological Indicators*, Volume 121, 2021. doi:10.1016/j.ecolind.2020.106976

[7] Lamichhaney, S. et al., Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean *Proceedings of the National Academy of Sciences*, 114(17): E3452–E3461, 2017. doi: 10.1073/pnas.1617728114

[8] Lorenzen, K., Toward a new paradigm for growth modeling in fisheries stock assessments: Embracing plasticity and its consequences. *Fisheries Research*, Volume 180: 4–22, 2016. doi:10.1016/j.fishres.2016.01.006

[9] Maynard, L.D. et al., Supporting material for stock assessment of NAFO Division 4TVn southern Gulf of St. Lawrence Atlantic Herring (*Clupea harengus*) in 2022-2023 *DFO Canadian Science Advisory Secretariat*, 2024/058, 2024.

[10] Michalowicz, J. V., Nichols, J. M., and Bucholtz, F., Handbook of Differential Entropy, CRC Press, Taylor & Francis Group, 2018.

[11] Python Software Foundation, Python: A Programming Language for Software Development, 2023. url: https://www.python.org/

[12] Rolland, N. et al., Assessment of the NAFO division 4TVn southern Gulf of St. Lawrence Atlantic herring (*Clupea harengus*) in 2020-2021, *Canadian Science Advisory Secretariat (CSAS)*, 2022.

[13] Turcotte, F., Hanke, A., and McDermid, J.L., Atlantic bluefin tuna diet variability in the southern Gulf of St. Lawrence, Canada *Marine Environmental Research*, Volume 187, 2023. doi:10.1016/j.marenvres.2023.105949

[14] Virtanen, P. et al., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python *Nature Methods*, 17(3): 261–272, 2020. doi:10.1038/s41592-019-0686-2

The use of d-Choquet integrals for differential privacy: experimental evidence

José Carlos R. Alcantud [♭1]

(♭) BORDA Research Group and IME,
University of Salamanca, Spain.

Abstract. This work is a contribution to the knowledge of the recently defined differentially private d-Choquet integrals. Experiments accompany theoretical results. They illustrate the performance of this novel mechanism for privacy preserving disclosure of aggregate data.

Keywords: Choquet integral, d-Choquet integral, fuzzy measure, dissimilarity, data privacy.
MSC Codes: 26A39, 68P27.

1 Introduction

This contribution extends the investigation of differentially private d-Choquet integrals initiated by Alcantud [2]. Two fundamental references justify this study. First, Torra [10] proposed a differentially private Choquet integral and studied its performance, inclusive of experimental evidence. Alcantud [1] extended the theoretical basis of Torra's pioneering inspection. Secondly, restricted dissimilarity functions [5] allowed Bustince *et al.* [6] to generalize discrete Choquet integrals to discrete d-Choquet integrals. With these components, Alcantud [2] put forward an axiomatic analysis of a new concept, namely, differentially private d-Choquet integrals. Here we supplement this analysis with new results and experimental evidence that compares to that given by Torra [10] for the performance of differentially private Choquet integrals.

2 Fundamental concepts

Henceforth we work with an arbitrary set $N = \{e_1, \ldots, e_T\}$ with T elements.

For any $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}^T$, a vector $\mathbf{a}_{\nearrow} = (a_{\sigma(1)}, \ldots, a_{\sigma(T)})$ is presented with $a_{\sigma(1)} \leqslant \ldots \leqslant a_{\sigma(T)}$ for a bijective mapping $\sigma : \{1, \ldots, T\} \longrightarrow \{1, \ldots, T\}$ (i.e., a permutation of the indices). Associated with $\mathbf{a}$ and j we refer to $L_j^{\mathbf{a}} = \{\sigma(j), \ldots, \sigma(T)\}$, the set of indices corresponding to the largest $T - j + 1$ components of $\mathbf{a}$.

As in Torra [10] and Alcantud [1], henceforth we consider the case of the space of databases $\mathcal{D} = [0,1]^T = \{f | f : N \longrightarrow [0,1]\}$. We abbreviate $f(e_i) = f_i \in [0,1]$ for simplicity. When $f, \bar{f} \in \mathcal{D}$ are neighboring databases (i.e., they differ in at most 1 element), we write $f \sim \bar{f}$.

We denote by $rg(M)$ the range of any mapping M.

2.1 Fundamental definitions concerning capacities and Choquet integrals

To define Choquet integrals, the following notion is needed:

Definition [4, Def. 2.75]. A discrete fuzzy measure (also, a discrete capacity) on N is a set function $\mu : 2^N \longrightarrow [0,1]$ that satisfies $\mu(\varnothing) = 0$ and $\mu(N) = 1$, plus $\mu(A) \leqslant \mu(A')$ for each $A \subseteq A' \subseteq N$ (monotonicity).

[1] jcr@usal.es

The smallest (or null, sometimes degenerate) fuzzy measure $\mu_{\mathbb{S}}$ assigns $\mu_{\mathbb{S}}(N) = 1$, and $\mu_{\mathbb{S}}(N') = 0$ when $N' \subsetneq N$. The largest (also called universal or maximal) fuzzy measure $\mu_{\mathbb{L}}$ assigns $\mu_{\mathbb{L}}(\varnothing) = 0$, and $\mu_{\mathbb{L}}(N') = 1$ when $\varnothing \neq N' \subseteq N$.

Using (discrete) fuzzy measures on N, Choquet integrals are defined as follows:

Definition. Fix a discrete fuzzy measure μ on N. The discrete Choquet integral $C^{\mu} : \mathbb{R}_+^T \longrightarrow \mathbb{R}_+$ associated with μ evaluates each $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}_+^T$ by

$$C^{\mu}(\mathbf{a}) = \sum_{i=1}^{T} \left(a_{\sigma(i)} - a_{\sigma(i-1)}\right) \mu(L_i^{\mathbf{a}}), \tag{1}$$

where the permutation σ ensures $\mathbf{a}_{\nearrow} = (a_{\sigma(1)}, \ldots, a_{\sigma(T)})$ and the convention $a_{\sigma(0)} = 0$ applies.

It is well known that the discrete Choquet integral associated with $\mu_{\mathbb{S}}$, resp., $\mu_{\mathbb{L}}$, concides with the maximum, resp., minimum, function.

The Choquet integral on $\mathcal{D} = \{f | f : N \longrightarrow [0,1]\}$ –the space of databases– associated with μ is defined by Torra [10] as follows: for any $f \in \mathcal{D}$, $C^{\mu}(f) = C^{\mu}(f_1, \ldots, f_T)$. This formula always produces a number in $[0,1]$.

Beliakov *et al.* [3, section 5.2] report on a reformulation of Eq. (1) that sums up over all evaluations by the fuzzy measure. It evaluates each $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}_+^T$ by

$$C^{\mu}(\mathbf{a}) = \sum_{S \subseteq N} g_S(\mathbf{a}) \mu(S), \tag{2}$$

where $g_S(\mathbf{a}) = \max\{\min_{j \in S} a_j - \max_{j \notin S} a_j, 0\}$. The convention that minimum/maximum over the empty set is zero applies.

2.2 Restricted dissimilarity functions and d-Choquet integrals

A recent extension of the discrete Choquet integral replaces the arithmetic subtraction of values in Eq. (1) with a restricted dissimilarity function. This concept originates with an axiomatic approach by Bustince *et al.* [5]:

Definition [5]. The mapping $\delta : [0,1]^2 \to [0,1]$ is a *restricted dissimilarity function* when for all $a_1, a_2, a_3 \in [0,1]$:

a) $\delta(a_1, a_2) = \delta(a_2, a_1)$,

b) $\delta(a_1, a_2) = 0 \Leftrightarrow a_1 = a_2$, and $\delta(a_1, a_2) = 1 \Leftrightarrow \{a_1, a_2\} = \{0, 1\}$,

c) $a_1 \leqslant a_2 \leqslant a_3 \Rightarrow \delta(a_1, a_2) \leqslant \delta(a_1, a_3)$ and $\delta(a_2, a_3) \leqslant \delta(a_1, a_3)$.

A direct manipulation on the definition of discrete Choquet integral generates the discrete d-Choquet integral:

Definition [6]. Fix a discrete fuzzy measure μ on N and $\delta : [0,1]^2 \to [0,1]$, a restricted dissimilarity function. The *discrete d-Choquet integral* (with T variables) with respect to μ and δ evaluates each $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}_+^T$ by

$$C^{\mu,\delta}(a_1, \ldots, a_T) = \sum_{i=1}^{T} \delta(a_{\sigma(i)}, a_{\sigma(i-1)}) \mu(L_i^{\mathbf{a}}), \tag{3}$$

where the permutation σ ensures $\mathbf{a}_{\nearrow} = (a_{\sigma(1)}, \ldots, a_{\sigma(T)})$ and the convention $a_{\sigma(0)} = 0$ applies.

Remark 1. When $\delta(x, y) = |x - y|$, the d-Choquet integral boils down to the usual Choquet integral. In addition, [6, Example 3.48] proves that $C^{\mu_{\mathbb{S}},\delta}(a_1, \ldots, a_T) = \delta(\min\{a_1, \ldots, a_T\}, 0)$ and $C^{\mu_{\mathbb{L}},\delta}(a_1, \ldots, a_T) = \sum_{i=1}^{T} \delta(a_{\sigma(i)}, a_{\sigma(i-1)})$, for any restricted dissimilarity function δ.

Note $C^{\mu,\delta} : [0,1]^T \longrightarrow [0,T]$, i.e., we cannot in general assure that the range of $C^{\mu,\delta}$ is contained in $[0,1]$. It is generally not true that $C^{\mu,\delta}$ is averaging, resp., monotonic.

Following Alcantud [2], we define d-Choquet integrals on $\mathcal{D}$ by

$$C^{\mu,\delta}(f) = C^{\mu,\delta}(f(e_1), \ldots, f(e_T)) \in [0,T] \text{ for each } f \in \mathcal{D}. \tag{4}$$

Alcantud [2] demonstrates that in fact, $C^{\mu,\delta}(f) \in [0,1]$ for all $f \in \mathcal{D}$ whenever $C^{\mu,\delta}$ is monotonic.

2.3 Sensitivity of a d-Choquet integral: importance for data privacy

Dwork [7] pioneered the study of differential privacy. It is defined in terms of a "privacy budget" ϵ. This budget determines how much privacy is enforced: smaller values of ϵ convey better privacy, and absolute privacy is achieved when $\epsilon = 0$. The implementation of this concept by the Laplace mechanism uses the sensitivity of a query or function [7]. We are concerned with disclosure of the output of an aggregation of databases by a d-Choquet integral $C^{\mu,\delta}$, and consequently with the literature, this figure is denoted by $\Delta_{\mathcal{D}}(C^{\mu,\delta})$.

Using these tools and the techniques developed by Torra [10] and Alcantud [1], Alcantud [2] showed that the sensitivity of the differentially private d-Choquet integral defined from a fuzzy measure μ and a restricted dissimilarity function δ is

$$\Delta_{\mathcal{D}}(C^{\mu,\delta}) = \max\{|C^{\mu,\delta}(f) - C^{\mu,\delta}(\bar{f})| : f, \bar{f} \in \mathcal{D} \text{ and } f \sim \bar{f}\}. \tag{5}$$

Therefore $L_{C^{\mu,\delta}}(f) = C^{\mu,\delta}(f) + L\Big(0, \frac{\Delta_{\mathcal{D}}(C^{\mu,\delta})}{\epsilon}\Big)$ is the differentially private d-Choquet integral defined from μ and δ. This expression stems from the Laplace mechanism [7] adapted for $C^{\mu,\delta}$ when it operates on databases from $\mathcal{D}$. Thus, $L\Big(0, \frac{\Delta_{\mathcal{D}}(C^{\mu,\delta})}{\epsilon}\Big)$ is the Laplacian noise with 0 mean and scale parameter $\frac{\Delta_{\mathcal{D}}(C^{\mu,\delta})}{\epsilon}$.

Altogether, this design guarantees $Pr(L_{C^{\mu,\delta}}(f) \in S) \leqslant e^{\epsilon} \cdot Pr(L_{C^{\mu,\delta}}(\bar{f}) \in S)$ whenever $f \sim \bar{f}$ and $S \subseteq rg(L_{C^{\mu,\delta}})$, i.e., ϵ-differential privacy for $L_{C^{\mu,\delta}}$ as established by [7].

Computations become easier under specific conditions. One of the restrictions (defined by Bustince *et al.* [6]) that help is

$$\text{Property (P1): } \sum_{j=1}^{T} \delta(a_{j-1}, a_j) \leqslant 1 \text{ when } 0 \leqslant a_1 \leqslant \cdots \leqslant a_T \leqslant 1. \tag{6}$$

Let us recall the computation of sensitivities in a few interesting cases.

Example 1 [2, Examples 7 and 8]. For each restricted dissimilarity function δ, $\Delta_{\mathcal{D}}(C^{\mu_{\mathbb{S}},\delta}) = 1$. If δ satisfies condition (P1) then $\Delta_{\mathcal{D}}(C^{\mu_{\mathbb{L}},\delta}) = 1$.

Proposition 1 [2, Proposition 7]. Suppose that δ is a restricted dissimilarity function that satisfies (P1). Then the sensitivity of the differentially private d-Choquet integral defined from a fuzzy measure μ and δ equals the sensitivity of the differentially private Choquet integral defined from μ, when $C^{\mu,\delta}$ is monotonic.

3 A new expression for d-Choquet integrals

In this section we provide a new compact expression for the d-Choquet integral. It is motivated by Eq. (2) that recalculates the Choquet integral.

The proof of the following result is a matter of direct calculations. Its formulation uses the standard sign (or signum) piecewise function:

Proposition 2. In the conditions of section 1.2.2, the discrete d-Choquet integral defined by Eq. (3) is expressed as: for each $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}_+^T$,

$$C^{\mu,\delta}(a_1, \ldots, a_n) = \sum_{S \subseteq N} \gamma_S(\mathbf{a})\mu(S), \tag{7}$$

where $\gamma_S(\mathbf{a}) = \max\{\text{sgn}(\min_{j\in S} a_j - \max_{j\notin S} a_j) \cdot \delta(\min_{j\in S} a_j, \max_{j\notin S} a_j), 0\}$ for all $S \subseteq N$.

The arguments in Beliakov *et al.* [3, section 5.2] apply here as a justification of the usefulness of this expression e.g., in learning fuzzy measures. Having an index that runs over all subsets is very convenient to set up the corresponding programming problems.

A more restricted version is given below with the help of the median function:

Proposition 3. In the conditions of Proposition 2, suppose that δ is such that for each $a, b \in [0,1]$, $\delta(a,b) \leqslant |a-b|$. Then the discrete d-Choquet integral is expressed as: for each $\mathbf{a} = (a_1, \ldots, a_T) \in \mathbb{R}_+^T$,

$$C^{\mu,\delta}(a_1, \ldots, a_n) = \sum_{S \subseteq N} \gamma'_S(\mathbf{a})\mu(S), \tag{8}$$

where $\gamma'_S(\mathbf{a}) = \text{med}\{\min_{j\in S} a_j - \max_{j\notin S} a_j, \delta(\min_{j\in S} a_j, \max_{j\notin S} a_j), 0\}$ for all $S \subseteq N$.

Proof. When S is not a subset of indices of t largest components of $\mathbf{a}$, one has $\min_{j\in S} a_j - \max_{j\notin S} a_j \leqslant 0$. And in this case $\gamma'_S(\mathbf{a}) = 0$ because $\delta(\min_{j\in S} a_j, \max_{j\notin S} a_j) \geqslant 0$.

When $S = L_t^{\mathbf{a}}$ for some t, $\gamma'_S(\mathbf{a}) = \delta(a_{\sigma(t)}, a_{\sigma(t-1)})$ because the assumption on δ guarantees $0 \leqslant \delta(a_{\sigma(t)}, a_{\sigma(t-1)}) \leqslant |a_{\sigma(t)} - a_{\sigma(t-1)}| = \min_{j\in S} a_j - \max_{j\notin S} a_j$.

In conclusion, we obtain the calculation produced by Eq. (7). □

4 Experimental evidence

In this section we report on computational simulations of two different types. Both compare the rankings produced from the aggregation by a d-Choquet integral with the ranking produced from aggregation with privacy preserving d-Choquet integrals (i.e., executions of an appropriate random modification of it described in the previous section). The benchmark set of evaluations that is aggregated remains constant: it is the experts' assessments for ten candidates reproduced in Torra [10, Table 2].

4.1 Experimental comparison with Torra [10, Section 4]

This section takes advantage of Alcantud [2, Proposition 7] to compare the results of the first experiment in Torra [10, Section 4] with the case where other dissimilarities are used. In his experiment, the fuzzy measure μ_0 is defined in [10, Table 1]. Torra calculates $\Delta_{\mathcal{D}}(C^{\mu_0}) = 0.7049$ using the expression in [10, Proposition 12].

Note that we can assure $\Delta_{\mathcal{D}}(C^{\mu,\delta}) = \Delta_{\mathcal{D}}(C^{\mu})$ when μ is any fuzzy measure and either $\delta_1(a,b) = |\sqrt{a} - \sqrt{b}|$ or $\delta_2(a,b) = |a^2 - b^2|$. The reason is that both dissimilarities satisfy (P1) by [6, Remark 3.13], and that [6, Example 4.13] established that $C^{\mu,\delta}$ must be monotonic for each μ. Hence Proposition 1 applies.

In conclusion, we are ready to replicate the experiment for the comparison of the rankings provided by C^{μ_0} and its privacy preserving modification using the Laplacian mechanism. To this purpose we modify Torra's freely available Python software https://mdai.cat/code/. In this way we reproduce his simulation with the two privacy preserving d-Choquet integrals defined from μ_0 and either δ_1 (first case) or δ_2 (second case). We justified above that both are obtained from the sensitivity 0.7049. To make comparisons meaningful, we consider the same values of the privacy budget ϵ and 50 executions. The rankings produced in each case by the d-Choquet integral and its privacy preserving version are compared by means of Spearman's correlation coefficient as well.

The next Figure 1 represents the results of the two computational experiments.

For each value of ϵ, the figures represent both the average and standard deviation of Spearman's correlation coefficient. It is worth mentioning that a Spearman correlation near to zero indicates that there is almost no tendency for the modified aggregate assessment to either increase or decrease when the d-Choquet aggregate evaluation increases. A Spearman correlation of one indicates that the rank of the candidates is identical in both assessments.

We can observe the influence of the choice of the dissimilarity. The experimental evidence suggests that δ_2 is associated with a better performance. These experiments are comparable with the first case of Torra [10, Section 4]: this is the exercise where we replace the dissimilarity with $\delta_3(a,b) = |a-b|$ in the previous experiments, all other factors remaining unchanged. Figure 2 shows the results of that simulation.

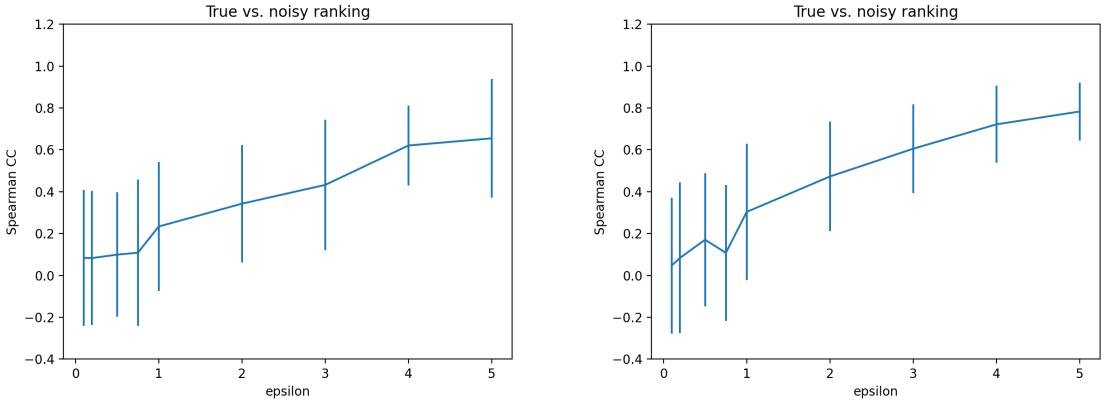


Figure 1: Spearman's correlation coe icients for the ranking derived from two d-Choquet integrals versus 50 simulations with its Laplacian modification ensuring ϵ-differential privacy. We use δ_1 (left) and δ_2 (right).

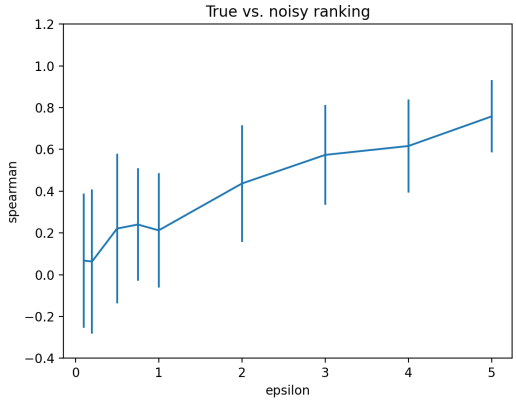


Figure 2: Spearman's correlation coe icients for the ranking derived from Choquet integral versus 50 simulations with its Laplacian modification e nsuring ϵ -differential pr ivacy (r eplication of th e experiment producing Figure 1 of Torra [10]).

4.2 Experimental comparison with extreme d-Choquet integrals

Now we replicate the experiments changing the fuzzy measure. This will allow us to compare the performance of new privacy preserving d-Choquet integrals, with the results obtained with the dissimilarities δ_1 and δ_2 in the case of the μ_0 measure of section 1.4.1. The gist of these simulations is Example 1, that cites conclusions from [2]. As both δ_1 and δ_2 satisfy (P1), the sensitivity of the d-Choquet integrals that they define jointly with the smallest/largest fuzzy measures is 1: $\Delta_{\mathcal{D}}(C^{\mu_{\mathbb{S}},\delta_1}) = \Delta_{\mathcal{D}}(C^{\mu_{\mathbb{L}},\delta_1}) = \Delta_{\mathcal{D}}(C^{\mu_{\mathbb{S}},\delta_2}) = \Delta_{\mathcal{D}}(C^{\mu_{\mathbb{L}},\delta_2}) = 1$.

The next Figure 3 represents the results of two new computational experiments. Both stick to the comparison made in the previous section. Now the d-Choquet integral is respectively defined from δ_1 and the smallest and largest fuzzy measures, and its privacy preserving modifications are defined with the appropriate sensitivity (namely, 1).

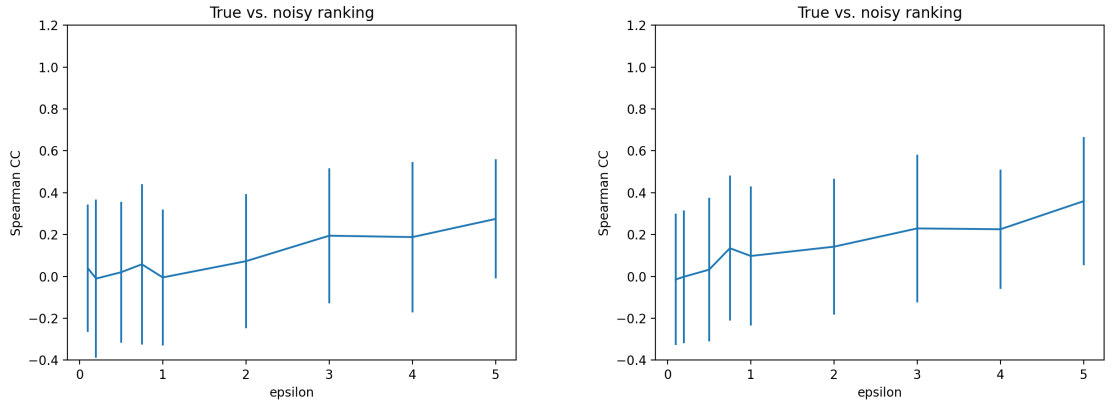


Figure 3: Spearman's correlation coefficients for the ranking derived from d-Choquet integral versus 50 simulations with its Laplacian modification ensuring ϵ-differential privacy. We use δ_1 with smallest (left) and largest (right) fuzzy measure.

Similarly, Figure 4 represents the results of new computational experiments comparing the d-Choquet integral respectively defined from δ_2 and the smallest and largest fuzzy measures, with its privacy preserving modifications (defined with sensitivity 1).

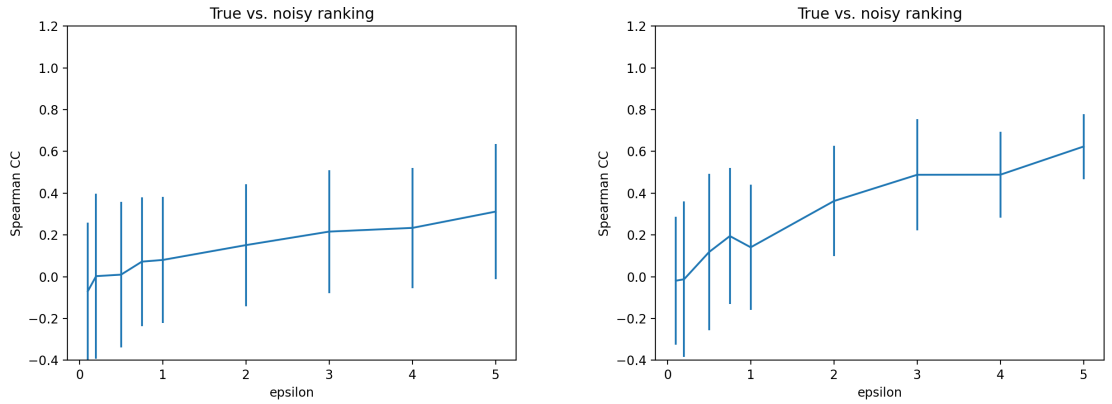


Figure 4: Spearman's correlation coefficients for the ranking derived from d-Choquet integral versus 50 simulations with its Laplacian modification ensuring ϵ-differential privacy. We use δ_2 with smallest (left) and largest (right) fuzzy measure.

As expected, the utilization of the extreme cases of fuzzy measures results into worse performances in comparison with the case of the fuzzy measure μ_0, irrespectively of whether the d-Choquet integral (and its privacy preserving modification) is defined with δ_1 or δ_2. Among the four cases discussed here, the combination of the largest fuzzy measure with δ_2 outperforms its

combination with δ_1. Remark 1 hints at the origin of this low performance of the smallest fuzzy measure $\mu_{\mathbb{S}}$: the d-Choquet integrals that it defines provide "naive" evaluations that discard most of the informational content.

5 Conclusions and future work

This contribution demonstrates that the investigation of differential privacy in the context of generalized forms of the Choquet integral is a promising avenue for further research. Note that in addition to d-Choquet integrals, other extensions of the discrete Choquet integral have been recently proposed by e.g., Sartori *et al.* [8], Takáč *et al.* [9], and Wieczynski *et al.* [11]. Our work provides further motivation for the investigation of data privacy issues in these new flexible frameworks for aggregation.

References

[1] Alcantud, J.C.R., New classes of differentially private Choquet integrals for privacy preservation, *submitted* (February 2025)

[2] Alcantud, J.C.R., The differentially private d-Choquet integral: an extension of differentially private Choquet integrals, *Submitted to the 22nd International Conference on Modeling Decisions for Artificial Intelligence* (March 2025)

[3] Beliakov, G., James, S., Wu, J., Discrete fuzzy measures. Computational Aspects. Springer Nature Switzerland AG, 2020.

[4] Beliakov, G., Pradera, A., Calvo, T., Aggregation Functions: A Guide for Practitioners. Heidelberg, Springer-Verlag, 2007.

[5] Bustince, H., Barrenechea, E., Pagola, M., Relationship between restricted dissimilarity functions, restricted equivalence functions and normal EN-functions: Image thresholding invariant *Pattern Recognition Letters*, 29(4):525–536, 2008.

[6] Bustince, H., Mesiar, R., Fernández, J., Galar, M., Paternain, D., Altalhi, A., Dimuro, G., Bedregal, B., Takáč, Z., *d*-Choquet integrals: Choquet integrals based on dissimilarities *Fuzzy Sets and Systems*, 414:1–27, 2021.

[7] Dwork, C., Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) International colloquium on automata, languages, and programming. Lecture Notes in Computer Science, vol. 4052, pp. 1–12. Springer, 2006.

[8] Sartori, J., *et al.*, d-CC integrals: generalizing CC-integrals by restricted dissimilarity functions with applications to fuzzy-rule based systems. In: Naldi, M. C., Bianchi, R. A. C. (eds.) Intelligent Systems, pp. 243–258. Springer Nature Switzerland, 2023.

[9] Takáč, Z., Uriz, M., Galar, M., Paternain, D., Bustince, H., Discrete IV dG-Choquet integrals with respect to admissible orders *Fuzzy Sets and Systems*, 441:169–195, 2022.

[10] Torra, V., Differentially private Choquet integral: extending mean, median, and order statistics *International Journal of Information Security*, 24:68, 2025

[11] Wieczynski, J., *et al.*, *d*-XC integrals: On the generalization of the expanded form of the Choquet integral by restricted dissimilarity functions and their applications *IEEE Transactions on Fuzzy Systems*, 30(12): 5376–5389, 2022.

Modeling Uncertainty in Linear Homogeneous Neutral Delay Differential Equations

J.-C. Cortés[♭], G. Gonzalez-Parra[♮] and S.-M. Sferle[♭1]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

(♮) Department of Mathematics, New Mexico Tech,
New Mexico, 87801, USA.

1 Introduction

Neutral delay differential equations (NDDEs) arise naturally in systems where the rate of change of a state variable depends not only on its current and past values, but also on the past values of its derivative. These models are especially relevant in disciplines such as engineering, biology and control theory, where time delays and memory effects are intrinsic [1].

The purpose of this work is to focus on a specific class of NDDEs, namely linear and homogeneous equations, and study them from a random perspective that better aligns with real-world applications. While various strategies exist for incorporating uncertainty into delay differential models, our approach introduces randomness through the model coefficients, which are treated as continuous random variables. This reformulation transforms the original deterministic NDDE into a random differential equation, and consequently, its solution becomes a stochastic process.

This probabilistic framework enables us to go beyond the deterministic analysis by deriving the one-dimensional probability density function (1-PDF) of the solution at any fixed value of the independent variable, here denoted as t, representing time. From the 1-PDF, one can compute all statistical moments, so providing a richer and more informative description of the system's behavior under uncertainty.

Let us consider the following random linear, homogeneous neutral delay differential equation with a fixed delay $h > 0$,

$$\begin{aligned} z'(t) = A_1 z(t) + A_2 z(t-h) + A_3 z'(t-h), \quad t \in [t_0, t_1], \\ z(t) = A_4 \phi(t), \quad t \in [-h + t_0, t_0], \end{aligned} \tag{1}$$

where t is the independent variable representing time, $z(t)$ is the state variable, $z(t-h)$ is the delayed state, $z'(t-h)$ is the derivative of the solution with delay h, and $\phi(t)$ is the so called *history* function. The coefficients $A_1(\omega), A_2(\omega), A_3(\omega)$ and $A_4(\omega)$ are assumed to be absolutely continuous random variables defined on a common complete probability space $(\Omega, \mathcal{F}_\Omega, \mathbb{P})$, with a known joint PDF denoted by $f_{A_1A_2A_3A_4}(a_1, a_2, a_3, a_4)$. Since the present study is primarily theoretical and no empirical data are available, the random variables are assumed to be independent to facilitate analytical tractability. Under this assumption, the joint PDF factorizes as the product of the marginal densities, $f_{A_1}(a_1) f_{A_2}(a_2) f_{A_3}(a_3) f_{A_4}(a_4)$.

[1] smsferle@doctor.upv.es

Here, $\omega \in \Omega$ denotes a generic outcome, and the solution should be written as $z(t,\omega)$ to reflect its stochastic nature. However, for notational simplicity, we omit the explicit dependence on ω throughout the paper.

The inclusion of the random variable A_4 in the initial condition reflects a general case in which the history function is scaled by a random amplitude, accounting for uncertainty in the initial state. For concreteness, we fix $t_0 = 0$, $t_1 = h$, and $h = 2$, and take the deterministic history function as $\phi(t) = \sin(\pi t)$.

2 Methods

Due to their higher degree of complexity compared to ordinary differential equations, NDDEs can exhibit instability by time delays or even admit periodic solutions. As a consequence, numerous studies have focused on developing both analytical and numerical techniques for solving such equations.

In this work, we adopt the classical method of steps (MoS) to solve the NDDE, which yields piecewise-defined solutions and, in many cases, enables an analytical expression for the solution [2]. This method is particularly suitable for our study, as it provides a natural bridge between the deterministic and random settings. Once the deterministic solution is obtained, the random formulation can be constructed by treating the model coefficients as absolutely continuous random variables, since the underlying theory remains valid under randomization.

Although this method has the drawback of being limited to short time intervals (as the procedure tends to accumulate complexity or numerical error with each additional step), our objective here is not to improve numerical techniques, but rather to illustrate how uncertainty in the model parameters propagates through the system. Consequently, computing the solution over just two intervals is sufficient to demonstrate the stochastic framework.

Applying the MoS to the deterministic version of equation (1), we obtain explicit expressions for the solution over two consecutive intervals. On the first interval $(0, 2]$, the solution takes the form

$$z_1(t) = \frac{a_4((a_2 + a_1a_3)\,\mathrm{e}^{a_1 t}\pi - (a_2 + a_1a_3)\,\pi\cos(\pi t) + \left(\pi^2 a_3 - a_1a_2\right)\sin(\pi t))}{a_1^2 + \pi^2}, \quad 0 < t \le 2, \tag{2}$$

and on the second interval $(2, 4]$, the solution continues is

$$\begin{aligned} z_2(t) = \frac{a_4}{(a_1^2+\pi^2)^2}\Big((a_2 + a_1a_3)\mathrm{e}^{a_1(t-2)}\pi\left(a_1^2\left(-1 + \mathrm{e}^{2a_1} + a_2(t-2)\right) + \pi^2\left(-1 + 2a_3 + \mathrm{e}^{2a_1} + a_2(t-2)\right) + \right. \\ + a_1\left(-2a_2 + a_3\pi^2(t-2)\right) + a_1^3a_3(t-2)\big) + 2\pi(a_2 + a_1a_3)(a_1a_2 - a_3\pi^2)\cos(\pi t) + \\ \left. + \left(a_1^2a_2^2 - (a_2^2 + 4a_1a_2a_3 + a_1^2a_3^2)\pi^2 + a_3^2\pi^4\right)\sin(\pi t)\right), \quad 2 < t \le 4. \end{aligned} \tag{3}$$

For notational simplicity and to facilitate the application of the Random Variable Transformation (RVT) technique, that will be introduced later, we rewrite both expressions as

$$z_1(t) = a_4\ \frac{g_1(a_1, a_2, a_3, t)}{a_1^2 + \pi^2}, \quad z_2(t) = \frac{a_4}{(a_1^2 + \pi^2)^2}\ g_2(a_1, a_2, a_3, t).$$

These closed-form expressions depend explicitly on the model parameters a_1, a_2, a_3, a_4, which is essential for the stochastic extension.

As previously indicated, once the coefficients are modeled as absolutely continuous random variables, the solution becomes a stochastic process, and we aim to quantify the resulting uncertainty by computing its 1-PDF at a fixed time. To this end, we employ the RVT technique, which

requires an explicit expression for the deterministic solution as a function of the random inputs. This requirement motivated the use of the MoS, as it yields piecewise analytical solutions suitable for transformation. A version of the RVT technique is stated below [3].

Theorem 1 *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_n)$ *be* n*-dimensional random vectors. Suppose* $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^n$ *is a one-to-one transformation of* $\mathbf{X}$ *into* $\mathbf{Y}$*, such that* $\mathbf{Y} = \mathbf{r}(\mathbf{X})$*. Assume that* $\mathbf{r}$ *is continuous in* $\mathbf{X}$ *and has continuous partial derivatives with respect to* $\mathbf{X}$*. If* $f_{\mathbf{X}}(\mathbf{x})$ *denotes the known joint PDF of vector* $\mathbf{X}$*, and* $\mathbf{s} = \mathbf{r}^{-1}$ *represents the inverse mapping of* $\mathbf{r}$*, the joint PDF of vector* $\mathbf{Y}$ *is given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{s}(\mathbf{y})\right) |J_n| \,,$$

where $|J_n|$ *is the absolute value of the Jacobian determinant, defined as*

$$J_n = \det \begin{pmatrix} \dfrac{\partial s_1(\mathbf{y})}{\partial Y_1} & \cdots & \dfrac{\partial s_n(\mathbf{y})}{\partial Y_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial s_1(\mathbf{y})}{\partial Y_n} & \cdots & \dfrac{\partial s_n(\mathbf{y})}{\partial Y_n} \end{pmatrix}.$$

Let $\mathbf{X} = (A_1, A_2, A_3, A_4)$ be the random vector with known marginal PDFs $f_{A_1}(a_1)$, $f_{A_2}(a_2)$, $f_{A_3}(a_3)$, $f_{A_4}(a_4)$, t a fixed value and $\mathbf{r} : \mathbb{R}^4 \to \mathbb{R}^4$ the one-to-one transformation of $\mathbf{X}$ into $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$, whose components are

$$Y_1 = r_1(A_1, A_2, A_3, A_4) = A_1, \quad Y_2 = r_2(A_1, A_2, A_3, A_4) = A_2, \quad Y_3 = r_3(A_1, A_2, A_3, A_4) = A_3,$$

$$Y_4 = r_4(A_1, A_2, A_3, A_4) = z_1(t) = A_4 \, \frac{g_1(A_1, A_2, A_3, t)}{A_1^2 + \pi^2},$$

where the inverse mapping $\mathbf{s} : \mathbb{R}^4 \to \mathbb{R}^4$ is given by

$$A_1 = s_1(Y_1, Y_2, Y_3, Y_4) = Y_1, \quad A_2 = s_2(Y_1, Y_2, Y_3, Y_4) = Y_2, \quad A_3 = s_3(Y_1, Y_2, Y_3, Y_4) = Y_3,$$

$$A_4 = s_4(Y_1, Y_2, Y_3, Y_4) = \frac{Y_4(Y_1^2 + \pi^2)}{g_1(Y_1, Y_2, Y_3, t)}.$$

The absolute value of the Jacobian determinant is

$$|J_4| = \left| \frac{Y_1^2 + \pi^2}{g_1(Y_1, Y_2, Y_3, t)} \right|.$$

Based on the previous identifications, we now apply Theorem 1 to obtain the joint PDF of the random vector $\mathbf{Y}$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(y_1, y_2, y_3, \frac{y_4(y_1^2 + \pi^2)}{g_1(y_1, y_2, y_3, t)}\right) \left| \frac{y_1^2 + \pi^2}{g_1(y_1, y_2, y_3, t)} \right|.$$

Taking t arbitrary on the first interval $(0, 2]$ and marginalizing the above joint PDF with respect to the first three components, and using the fact that the input variables are independent taking into account the marginals PDFs, we obtain the 1-PDF of the solution, given by

$$f_{Z_1(t)}(z_1) = \int_{\mathbb{R}^3} f_{A_1}(a_1) f_{A_2}(a_2) f_{A_3}(a_3) f_{A_4}\left(\frac{z_1(a_1^2 + \pi^2)}{g_1(a_1, a_2, a_3, t)}\right) \left| \frac{a_1^2 + \pi^2}{g_1(a_1, a_2, a_3, t)} \right| \, \mathrm{d}a_1 \, \mathrm{d}a_2 \, \mathrm{d}a_3. \quad (4)$$

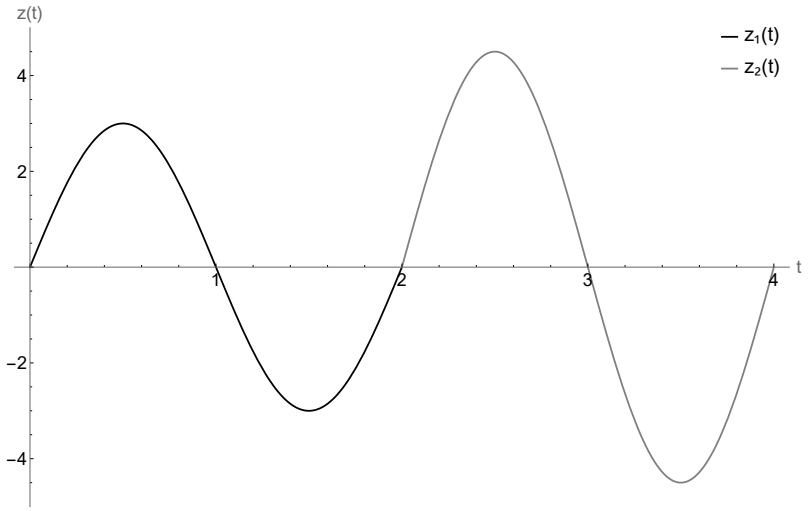


Figure 1: Deterministic solution, as defined in (7), corresponding to the deterministic NDDE (6).

The denominator in the argument of the function f_{A_4} vanishes only on a set of measure zero, since A_1, A_2 and A_3 are absolutely continuous random variables. Therefore, the transformation is well-defined almost surely.

Analogously, we obtain the 1-PDF of the solution on the second interval $(2, 4]$, whose expression is given by

$$f_{Z_2(t)}(z_2) = \int_{\mathbb{R}^3} f_{A_1}(a_1) f_{A_2}(a_2) f_{A_3}(a_3) f_{A_4}\left(\frac{z_2(a_1^2+\pi^2)^2}{g_2(a_1,a_2,a_3,t)}\right) \left|\frac{(a_1^2+\pi^2)^2}{g_2(a_1,a_2,a_3,t)}\right| \, \mathrm{d}a_1 \, \mathrm{d}a_2 \, \mathrm{d}a_3. \quad (5)$$

3 Results

Let us first consider the deterministic NDDE

$$\begin{aligned} z'(t) &= -2z(t) + 3z(t-2) + 1.5z'(t-2), \\ z(t) &= 2\sin(\pi t), \quad t \in [-2, 0], \end{aligned} \quad (6)$$

whose solution is given by

$$\begin{aligned} z_1(t) &= 3\sin(\pi(t-2)), \quad 0 < t \leq 2, \\ z_2(t) &= 4.5\sin(\pi(t-4)), \quad 2 < t \leq 4, \end{aligned} \quad (7)$$

and is plotted in the Figure 1. As expected from the sinusoidal history function, the solution exhibits oscillatory behavior over time. In addition, due to the scaling factor on the history function, the solution reflects a corresponding amplification, where the amplitude appears slightly higher in the second interval.

To illustrate the theoretical results obtained in the Section 2, we compute and visualize the 1-PDF of the solution at selected time points within the first and second intervals. Given that the marginal PDFs of the random variables are known, we assume that A_1, A_2, A_3, and A_4 follow independent Uniform distributions centered around the corresponding deterministic values in (6)

$$A_1 \sim \mathcal{U}[-3, -1], \quad A_2 \sim \mathcal{U}[2, 4], \quad A_3 \sim \mathcal{U}[1, 2], \quad A_4 \sim \mathcal{U}[1, 3].$$

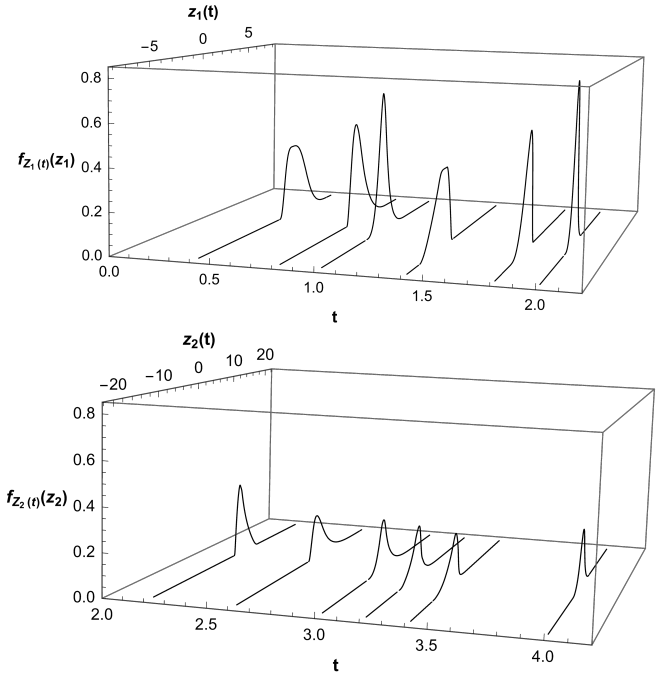


Figure 2: Visualization of the 1-PDF of the stochastic solution to the random NDDE. The top panel shows $f_{Z_1(t)}(z_1)$, defined in (4), corresponding to the solution $z_1(t)$ in (2) over the interval $(0, 2]$. The bottom panel shows $f_{Z_2(t)}(z_2)$, defined in (5), associated with the solution $z_2(t)$ in (3) over the interval $(2, 4]$.

Figure 2 displays the 1-PDF of the stochastic solution across the first two intervals. The upper panel corresponds to $f_{Z_1(t)}(z_1)$, defined in (4), representing the 1-PDF of the solution $z_1(t)$ at several time points within the interval $(0, 2]$, while the lower panel shows $f_{Z_2(t)}(z_2)$, defined in (5), for $z_2(t)$ in the interval $(2, 4]$. In both cases, the center of the PDFs shifts over time in an oscillatory manner, reflecting the behavior of the deterministic solution. In addition, the shape of the PDFs varies across time, particularly with respect to kurtosis. Notably, the second interval exhibits a more platykurtic behavior, associated with increased variance, which is consistent with the higher amplitude of the underlying deterministic oscillations.

4 Conclusions

In this work, we derived an explicit expression for the 1-PDF of the solution to a linear homogeneous neutral delay differential equation with random input parameters. The method of steps was employed to construct an exact, piecewise solution over a finite interval, which served as the foundation for the application of the RVT technique. Furthermore, an illustrative example was presented to validate the theoretical developments, showing that the mean behavior of the stochastic solution closely follows its deterministic counterpart.

Acknowledgments

This work has been supported by the Spanish grant PID2020-115270GB-I00 funded by MCIN/AEI/10.13039/501100011033 and the grant CIGE/2023/79 (Generalitat Valenciana). Sorina Madalina Sferle has been suported by the grant PRE2021-101090 funded by

MCIN/AEI/10.13039/501100011033 and by FSE+.

References

[1] Hale, J. K., Verduyn Lunel, S. M., Introduction to Functional Differential Equations. New York, Springer, 1993.

[2] Bauer, R. J., Mo, G., Krzyzanski, W., Solving delay differential equations in SADAPT by method of steps *Computer methods and programs in biomedicine*, 111(3):715–734, 2013.

[3] Soong, T. T., Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

Probabilistic study of random oscillators with cross-nonlinearities in position and velocity

Juan-Carlos Cortés[†], José-Vicente Romero[†], María-Dolores Roselló[†] and Joaquin F. Valencia Sullca[†],[1]

(†) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

It is well known that in Physics and Engineering, a significant number of vibratory systems have been studied, in particular nonlinear oscillators [1]-[2], that can be described by differential equations with nonlinearities as shown following,

$$\ddot{X}(t) + 2\beta\dot{X}(t) + \epsilon g(X(t), \dot{X}(t)) + \omega_0^2 X(t) = Y(t), \quad t > 0, \tag{1}$$

where $X(t)$ is the position of the oscillatory system, $\beta > 0$ is the damping constant, $\omega_0 > 0$ is the undamped angular frequency, ϵ is the perturbation that affects to the nonlinear function $g(X(t), \dot{X}(t))$ that depends on both the position, $X(t)$, and velocity, $\dot{X}(t)$; and $Y(t)$ is an external forcing term acting upon the system.

The stochastic perturbation technique has allowed to obtain reliable approximations of the Equation (1) [3]. For example, in [4] has been proposed a family of nonlinear oscillators with small nonlinearities where the nonlinear function is given by $g(t) = X^2\dot{X}(t)$; and excited by stationary Gaussian stochastic processes. However, one of the limitations of this technique is that it is restricted to problems with small perturbation. To overcome this limitation, the stochastic linearization technique [5], can allow to obtain reliable approximations of the Equation (1). More recently, in [6] this technique has been applied to a class of nonlinear oscillators, where the nonlinear function depends only on the spatial position, namely, $g(t) = \sin X(t)$.

Therefore, the objective of this paper is to study the stochastic equivalent linearization technique in a class of random nonlinear oscillators whose nonlinear function depends both the position and velocity, and driven by stationary Gaussian stochastic process, in order to obtain the approximations of the main statistics of the steady state, including the first moments, and the correlation function.

2 Stochastic Equivalent Linearization

The stochastic linearization equivalent technique is based on the construction of an equivalent linear equation from a nonlinear differential equation, as shown below,

$$\ddot{X}(t) + f\left[X(t), \dot{X}(t)\right] = Y(t), \quad t \in T, \tag{2}$$

[1] joavasu2@teleco.upv.es

$$\ddot{X}(t) + 2\beta_e \dot{X}(t) + \omega_e^2 X(t) = Y(t). \tag{3}$$

To minimize the error, $N(t)$, in mean-square sense, it is required that the parameters β_e and ω_e^2 be chosen such that

$$\mathbb{E}\left\{N^2(t)\right\} = \mathbb{E}\left\{\left[2\beta_e \dot{X} + \omega_e^2 X - f\left(X, \dot{X}\right)\right]^2\right\}, \tag{4}$$

is minimized for $t \in T$.

2.1 Probabilistic Model Study

Our study consider that the nonlinear function $g(X(t), \dot{X}(t))$ of the Equation (1), depending on both the position, $X(t)$, and velocity, $\dot{X}(t)$, has the form

$$g(X(t), \dot{X}(t)) = \sum_{\substack{n=0 \\ (n,m)\neq\{(0,0)\}}}^{N} \sum_{m=0}^{M} a_{n,m} (X(t))^n (\dot{X}(t))^m, \tag{5}$$

and the value of $f\left[X(t), \dot{X}(t)\right]$ in Equation (2) has the form

$$f\left[X(t), \dot{X}(t)\right] = 2\beta \dot{X}(t) + \omega_0^2 \left(X(t) + \epsilon g(X(t), \dot{X}(t))\right). \tag{6}$$

Then, the value of β_e and ω_e^2 are the solutions of the following equations

$$\begin{cases} 2\beta_{\mathrm{e}} \mathbb{E}\{(\dot{X}(t))^2\} + k_{\mathrm{e}}^2 \mathbb{E}\{X(t)\dot{X}(t)\} - \mathbb{E}\{\dot{X}(t) f(X(t), \dot{X}(t))\} = 0, \\ k_{\mathrm{e}}^2 \mathbb{E}\{(X(t))^2\} + 2\beta_{\mathrm{e}} \mathbb{E}\{X(t)\dot{X}(t)\} - \mathbb{E}\{X(t) f(X(t), \dot{X}(t))\} = 0. \end{cases} \tag{7}$$

So, from the first equation of (7), one obtains

$$\begin{aligned} 2\beta_{\mathrm{e}} \mathbb{E}\{(\dot{X}(t))^2\} + k_{\mathrm{e}}^2 \mathbb{E}\{X(t)\dot{X}(t)\} - 2\beta \mathbb{E}\{(\dot{X}(t))^2\} - \omega_0^2 \mathbb{E}\{X(t)\dot{X}(t)\} \\ - \epsilon\omega_0^2 \sum_{\substack{n=0 \\ (n,m)\neq\{(0,0)\}}}^{N} \sum_{m=0}^{M} a_{n,m} \mathbb{E}\{(X(t))^n (\dot{X}(t))^{m+1}\} = 0. \end{aligned} \tag{8}$$

Next, from the second equation of (7), one obtains

$$\begin{aligned} k_{\mathrm{e}}^2 \mathbb{E}\{(X(t))^2\} + 2\beta_{\mathrm{e}} \mathbb{E}\{X(t)\dot{X}(t)\} - 2\beta \mathbb{E}\{X(t)\dot{X}(t)\} - \omega_0^2 \mathbb{E}\{X(t)^2\} \\ - \epsilon\omega_0^2 \sum_{\substack{n=0 \\ (n,m)\neq\{(0,0)\}}}^{N} \sum_{m=0}^{M} a_{n,m} \mathbb{E}\{(X(t))^{n+1} (\dot{X}(t))^m\} = 0. \end{aligned} \tag{9}$$

At this point, we consider the stochastic processes $X(t)$ and $\dot{X}(t)$ as stationary, independent, and with zero mean. Also, we will consider, in this paper, that the nonlinear function is given by $f(X(t), \dot{X}(t)) = X^2 \dot{X}(t)$, which is a particular case of expression (5). From these considerations the values of β_e and ω_e^2 are

$$\beta_{\mathrm{e}} = \frac{\omega_0^2 \epsilon(\sigma_X^2)}{2} + \beta, \tag{10}$$

and

$$\omega_{\mathrm{e}}^2 = \omega_0^2. \tag{11}$$

2.2 Variance Computation

To obtain the solution of the equivalent linear equation [7], from the value of auxiliary parameters β_e and ω_e, it is necessary the calculation of the variance. The variance, σ_X^2, can be computed from the spectral density function, $S_{YY}(\omega)$, of $Y(t)$ by

$$\sigma_X^2 = \int_0^\infty S_{XX}(\omega)\mathrm{d}\omega = \int_0^\infty |y(i\omega)|^2 S_{YY}(\omega)\mathrm{d}\omega, \tag{12}$$

where $y(i\omega)$ is the frequency response associated with the equivalent linear system described by Equation (3). It has the form

$$y(i\omega) = \frac{1}{\omega_e^2 - \omega^2 + 2i\beta_e\omega}. \tag{13}$$

The frequency response function, $y(i\omega)$, can be approximated by setting $\epsilon = 0$, with the objective to facilitate the computation of β_e and ω_e. Then $\beta_e = \beta$ and $\omega_e = \omega_0$, and $\sigma_{X_0}^2$ is equal to

$$\sigma_{X_0}^2 = \int_0^\infty |y_0(i\omega)|^2 S_{YY}(\omega)\mathrm{d}\omega = \int_0^\infty \left[\frac{1}{\left(\omega_0^2 - \omega^2\right)^2 + 4\omega^2\beta^2}\right] S_{YY}(\omega)\mathrm{d}\omega. \tag{14}$$

Finally, the relation between the power spectral density of $Y(t)$ and the correlation function Γ_{YY} is given by

$$S_{YY}(\omega) = \frac{1}{\pi}\int_{-\infty}^\infty e^{-i\omega\tau}\Gamma_{YY}(\tau)\mathrm{d}\tau. \tag{15}$$

2.3 Steady-State Solution

The analysis of the steady-state solution of the Equation (3), can be obtained by using the convolution integral as follow

$$X(t) = \int_0^\infty h(s)Y(t-s)\mathrm{d}s, \tag{16}$$

where

$$h(t) = \begin{cases} \left(\omega_e^2 - \beta_e^2\right)^{-\frac{1}{2}} e^{-\beta_e t}\sin\left[\left(\omega_e^2 - \beta_e^2\right)^{\frac{1}{2}} t\right], & \text{if } t > 0, \\ 0, & \text{if } t \leq 0, \end{cases} \tag{17}$$

is the impulse response function for the underdamped case $\left(\dfrac{\beta_e^2}{\omega_e^2} < 1\right)$.

On the other hand, the statistical moments of odd order of the solution $X(t)$ are null

$$\mathbb{E}\left\{X^n(t)\right\} = 0, \quad n = 1, 3, 5.. \tag{18}$$

Next, the second-order moment for $X(t)$ is given by

$$\mathbb{E}\left\{X^2(t)\right\} = \int_0^\infty h(s)\int_0^\infty h(s_1)\Gamma_{YY}(s - s_1)\mathrm{d}s_1\mathrm{d}s. \tag{19}$$

From the previous consideration $\mathbb{E}\left\{X(t)\right\} = 0$, the covariance and correlation functions of $X(t)$ coincide, namely, $\mathbb{C}\mathrm{ov}\left\{X(t_1)X(t_2)\right\} = \Gamma_{XX}(\tau),\ \tau = |t_1 - t_2|$, where the correlation function of $X(t)$ is given by

$$\Gamma_{XX}(\tau) = \int_0^\infty h(s)\int_0^\infty h(s_1)\Gamma_{YY}(\tau - s_1 + s)\mathrm{d}s_1\mathrm{d}s. \tag{20}$$

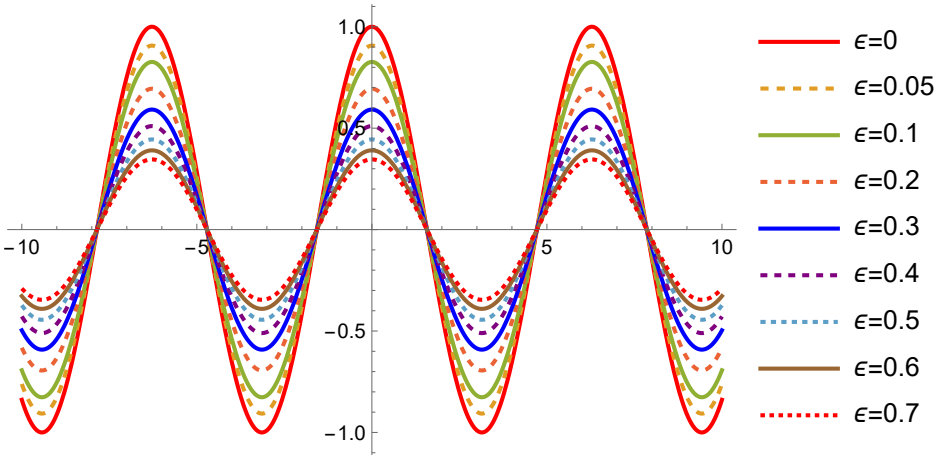


Figure 1: Correlation function $\Gamma_{XX}(\tau)$ of $X(t)$ for different values of ϵ.

2.4 Results and Discussion

To validate the theoretical results, we have chosen as input excitation the trigonometric stochastic process defined by $Y(t) = \xi_1 \cos(t) + \xi_2 \sin(t)$, where ξ_1 and $\xi_2 \sim N(0,1)$ are independent variables. Observe that $Y(t)$ satisfies the hypotheses, i.e., $\mathbb{E}\{Y(t)\} = 0$, is Gaussian and stationary, being its correlation $\Gamma_{YY}(\tau) = \cos(\tau)$. In this case, Equation (1) becomes

$$\ddot{X}(t) + 2\beta_e \dot{X}(t) + \omega_e^2 X(t) = Y(t). \tag{21}$$

For this model, we consider the following parameters $\beta = 0.5$, $\omega_0^2 = \omega_e^2 = 1$ and $\beta_e = \frac{\epsilon\omega_0^2}{2}\{\sigma_{X_0}^2\} + \beta$, where $\sigma_{X_0}^2 = 1$.

Next, using expression (19), the second-order moment is determined by

$$\mathbb{E}\{X^2(t)\} = \frac{1}{(1+\epsilon)^2}. \tag{22}$$

Notice that the condition corresponding to the underdamped case (assumed in the theoretical development) is fulfilled, since

$$\begin{aligned} \beta_{\mathrm{e}} &= \frac{(\epsilon\sigma_X^2 + 2\beta)}{2} < \omega_0^2 = \omega_e^2 = 1, \\ & (\epsilon\sigma_X^2 + 1) < 2, \\ & \left[\frac{\epsilon}{(1+\epsilon)^2}\right] < 1, \quad \forall \epsilon \geq 0. \end{aligned} \tag{23}$$

Now, applying (20), we obtain the following approximation of the correlation function,

$$\Gamma_{XX}(\tau) = \frac{\cos(\tau)}{(1+\epsilon)^2}. \tag{24}$$

In Figure 1, we show the graphical representation of the correlation function, $\Gamma_{XX}(\tau)$, given in the expression (24) for different values of ϵ.

3 Conclusions

In this paper, we have studied random nonlinear oscillators subject to perturbations on the nonlinear term that depend on both the position and velocity, that have been excited by a zero-mean stationary Gaussian process. Additionally, we have obtained reliable approximations of the main statistics of the stationary solution by means of the stochastic equivalent linearization technique. Finally, we have illustrated the case in that the input excitation is a trigonometric stochastic process.

Acknowledgments

This work has been supported by the grant PID2024-155597NB-I00 granted by MCIN/AEI/10.13039/501100011033.

References

[1] Yildirim, A., Askari, H., Saadatnia, Z., KalamiYazdi, M., Khan, Y., Analysis of nonlinear oscillations of a punctual charge in the electric field of a charged ring via a Hamiltonian approach and the energy balance method. *Computers & Mathematics with Applications*, 62(1): 486–490, 2011.

[2] Khan, Y., Vazquez-Leal, H., Hernandez-Martinez, L., Removal of noise oscillation term appearing in the nonlinear equation solution. *Journal of Applied Mathematics*, 2012(1): 387365, 2012.

[3] Crandall, S. H., Perturbation techniques for random vibration of nonlinear systems. *The Journal of the Acoustical Society of America*, 35(11): 1700-1705, 1963.

[4] Cortés, J.-C., López-Navarro, E., Romero, J.-V., Roselló, M.-D., Approximating the density of random differential equations with weak nonlinearities via perturbation techniques. *Mathematics*, 9(3): 204, 2021.

[5] Caughey, T. K., Equivalent linearization techniques. *The Journal of the Acoustical Society of America*, 35(11): 1706–1711, 1963.

[6] Cortés, J.-C., Romero, J.-V., Roselló, M.-D., Valencia Sullca, J.-F, Probabilistic analysis of the steady state of weakly perturbed linear oscillators subject to a class of Gaussian inputs. *Chaos, Solitons & Fractals*, 187: 115451, 2024.

[7] Soong, T. T., Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

Stability Concepts for Random Semi-dynamical Systems

O. Albescu[♭],[1] L.E. Biris[♮] T. Ceausu[♮] and I.-L. Popa[⋆,♭]

(♭) Faculty of Mathematics and Computer Science,
Transilvania University of Brasov, Brasov, Romania

(♮) Department of Mathematics,
West University of Timisoara, Timisoara, Romania

(⋆) "1 Decembrie 1918" University of Alba Iulia,
Gabriel Bethlen Nr. 5, 510109 Alba Iulia, Romania.

(♭) Faculty of Mathematics and Computer Science,
Transilvania University of Brasov, Brasov, Romania.

1 Introduction

Let $\mathbb{R}_+$ denotes the set of positive real. $(X, ||\cdot||)$ denotes a Banach space. By $\mathcal{B}(X)$ we denote the Banach algebra of all bounded linear operators acting from X into X. Also, by $(\Omega, \mathfrak{F}, \mathbb{P})$ we denote a probability space. We consider measurable application $\theta^t : \Omega \to \Omega$, for all $t \in \mathbb{R}_+$. This application preserves the probability measure, i.e. $\mathbb{P} \circ \theta^t = \mathbb{P}$, thus $\mathbb{P}(\theta^t B) = \mathbb{P}(B)$, for all $B \in \mathfrak{F}$. Moreover, we have that $\theta^0 = I_\Omega$ and $\theta^t \circ \theta^s = \theta^{t+s} = \theta^{s+t} = \theta^s \circ \theta^t$, for all $t, s \in \mathbb{R}_+$. Finally, we point out that the application $\mathbb{R}_+ \times \Omega \ni (t, \omega) \to \theta^t\omega \in \Omega$ is measurable for all $t \in \mathbb{R}_+$.

For simplifications, the family of applications $(\theta^t)_{t\in\mathbb{R}_+}$ will be denoted by θ. In this case, for the measurable family $\theta : \Omega \to \Omega$ we consider the metric semi-dynamical system which is denoted by $(\Omega, \mathfrak{F}, \mathbb{R}_+, \theta)$. The application $\varphi : \Omega \to (0, +\infty)$ is called $\theta-$invariant if verifies the property $\varphi(\theta^t\omega) = \varphi(\omega)$, for all $t \in \mathbb{R}_+$ and $\omega \in \Omega$.

Finally, we consider the measurable application $\phi : \mathbb{R}_+ \times \Omega \to \mathcal{B}(X)$ with satisfies the properties:

(a) $\phi(0, \omega) = I_X$, for all $\omega \in \Omega$;

(b) $\phi(t + s, \omega) = \phi(t, \theta^s\omega)\phi(s, \omega)$, for all $t, s \in \mathbb{R}_+$ and $\omega \in \Omega$.

Obvious, relation (b) implies that

$$\phi(t, \theta^s\omega)\phi(s, \omega) = \phi(t + s, \omega) = \phi(s + t, \omega) = \phi(s, \theta^t\omega)\phi(t, \omega),$$

for all $t, s \in \mathbb{R}_+$ and for all $\omega \in \Omega$. Throughout this work the notation (θ, ϕ) will be considered for a linear random one-sided system.

[1] oana.albescu@uab.ro

2 Stability Concepts

Definition 1 *We say that the semi-dynamic system (θ, ϕ) is nonuniformly exponentially stable if there exists a random variable θ-invariant, $\alpha : \Omega \to (0, +\infty)$ and a function N: $\Omega \to [1, +\infty)$ such that*

$$||\phi(t, \omega)x|| \leq N(\omega)e^{-\alpha(\omega)t}||x||, \tag{1}$$

for all $(t, \omega, x) \in \mathbb{R}_+ \times \Omega \times X$.

Definition 2 *We say that the semi-dynamic system (θ, ϕ) has nonuniformly exponential growth if there exists a random variable θ-invariant, $\xi : \Omega \to (0, +\infty)$ and a function $M : \Omega \to [1, +\infty)$ such that*

$$||\phi(t, \omega)x|| \leq M(\omega)e^{\xi(\omega)t}||x|| \tag{2}$$

for all $(t, \omega, x) \in \mathbb{R}_+ \times \Omega \times X$. If we consider the particular case when then function $M : \Omega \to [1, +\infty)$ to be constant then we obtain the uniform exponential growth concept.

Theorem 1 *The semi-dynamic system (θ, ϕ) which has uniformly exponential growth is nonuniformly exponentially stable if and only if there exists a θ- invariant random $\beta : \Omega \to (0, +\infty)$ and two functions $N, D : \Omega \to [1, +\infty)$ such that*

$$\int_s^{+\infty} e^{\beta(\omega)(t-s)}N(\theta^s\omega)^{-1}||\phi(t, w)x||dt \leq D(\omega)||\phi(s, \omega)x|| \tag{3}$$

for all $(s, \omega, x) \in \mathbb{R}_+ \times \Omega \times X$.

Definition 3 *We say that the semi-dynamic system (θ, ϕ) is nonuniformly polinomially stable if there exists a random variable θ-invariant, $\alpha : \Omega \to (0, +\infty)$ and a function $K : \Omega \to [1, +\infty)$ such that*

$$||\phi(t, \omega)x|| \leq K(\omega)t^{-\alpha(\omega)}||x||, \tag{4}$$

for all $(t, \omega, x) \in \mathbb{R}_+^* \times \Omega \times X$.

Theorem 2 *The semi-dynamic system (θ, ϕ) which has uniformly exponential growth is nonuniformly polinomially stable if and only if there exists a θ-invariant random $\beta : \Omega \to (0, +\infty)$ and tow functions $S, K : \Omega \to [1, +\infty)$ such that*

$$\int_s^{+\infty} (t-s)^{\beta(\omega)}K(\theta^s\omega)^{-1}||\phi(t, \omega)x||dt \leq S(\omega)||\phi(s, \omega)x|| \tag{5}$$

for all $(s, \omega, x) \in \mathbb{R}_+ \times \Omega \times X$.

References

[1] Arnold L., Random dynamical systems, Springer, New York, 1998.

[2] Cong N. D. , Topological dynamics of random dynamical systems, Clarendon, Oxford, 1997.

[3] I.-L. Popa, Lyapunov functions for random semi-dynamical systems in terms of tempered exponential splitting, *Math. Meth. Appl. Sci.*, **44**(15): 11923–11932, 2021.

Part IX

Complex networks graphs and applications

The Plurality of Consensus: Because One Ranking Cannot Rule Them All

Juan A. Aledo[♭],[1] José A. Gámez[♭] and Alejandro Rosete[♮]

(♭) Escuela Superior de Ingeniería Informática,
Universidad de Castilla-La Mancha, Spain.

(♮) Universidad Tecnológica de La Habana Jose Antonio Echeverría, Cuba

1 Introduction

In rank aggregation problems (RAP), the goal is typically to determine a consensus ranking that best represents a given set of input orderings. Various approaches exist, differing not only in the types of rankings they take as input and produce as output but also in the criteria used to assess the quality of the final ranking. However, in certain machine learning tasks, such as subgroup discovery, or in multimodal optimization, the focus shifts from finding a single optimal solution to generating multiple solutions that capture the diversity present in the input data or across the search space.

With this aim, we propose an alternative approach to RAP where, instead of a single consensus ranking, a set of rankings is provided to better reflect the preferences embedded in the input orderings. Thus, this approach integrates a fairness perspective [4], ensuring that diverse priorities and viewpoints are adequately represented in the aggregation process while also preventing the emergence of biases inherent in the input preferences [3].

Specifically, for an RAP with input set of preferences Π, a distance measure d, and a natural number $b \geq 2$, we can define the generalized RAP, which consists in finding b rankings $\pi_1, \pi_2, \dots \pi_b$ and b weights $w^1, w^2, \dots w^b \in [0,1]$, $\sum_{k=1..b} w^k = 1$, achieving the best fitness value

$$\sum_{k=1}^{b} w^k d(\Pi, \pi_k).$$

We illustrate this idea through the Optimal Bucket Order Problem (OBOP) [5, 1, 2], a rank aggregation problem that seeks to compute a consensus ranking (allowing ties) from a precedence matrix encoding the input rankings. To extend this framework, we introduce the Optimal Set of Bucket Orders Problem (OSBOP), which generalizes OBOP by producing not just one, but a collection of consensus rankings.

We carry out an experimental study which reveals the benefits of this approach, showing that offering multiple consensus rankings leads to a significant improvement in solution quality compared to OBOP, while maintaining interpretability.

[1] juanangel.aledo@uclm.es

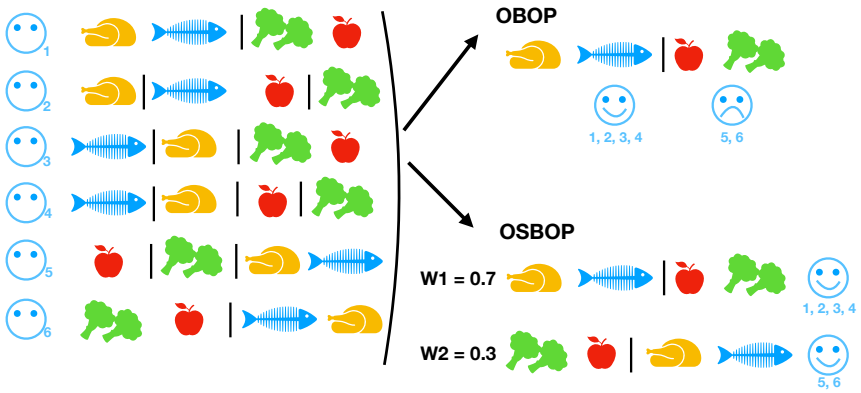


Figure 1: OBOP vs OSBOP

Acknowledgments

This work has been partially funded by the Government of Castilla-La Mancha and "ERDF A way of making Europe" under project SBPLY/21/180225/000062; by MCIU/AEI/10.13039/501100011033 and "ESF Investing your future" through project PID2022-139293NB-C32; and by the Universidad de Castilla-La Mancha and "ERDF A Way of Making Europe" under project 2022-GRIN-34437.

References

[1] Aledo, J.A., Gámez, J.A., Rosete, A., Utopia in the solution of the bucket order problem. *Decision Support Systems*, 97:69–80, 2017.

[2] Aledo, J.A., Gámez, J.A., Rosete, A., A highly scalable algorithm for weak rankings aggregation. *Information Sciences*, 570:144–171, 2021.

[3] Dong, Q., Yuan, Q., Shi, Y.B., Alleviating the recommendation bias via rank aggregation. *Physica A: Statistical Mechanics and its Applications*, 534:122073, 2019.

[4] Kuhlman, C., Rundensteiner, E., Rank aggregation algorithms for fair consensus. *Proceedings of VLDB Endowment*, 13(12):2706–2719, 2020.

[5] Ukkonen, A., Puolamaki, K., Gionis, A., Mannila, H., A randomized approximation algorithm for computing bucket orders. *Information Processing Letters*, 109(7):356 – 359, 2009.

First steps to predict extreme events in complex systems

K. Alfaro-Bittner[♭],[1] J. Tredicce[♭], L. Sarrazin[♮], M. Marconi[♮], and M. Giudici[♮]

(♭) MACIMTE – Applied Mathematics, Universidad Rey Juan Carlos
Calle Tulipán s/n, 28933 Móstoles, Madrid, Spain.

(♮) Institut de Physique de Nice, Université Côte d'Azur,
CNRS UMR 7010, Nice, 06200, France.

Initial conditions play an important role in classical systems because they determine their time evolution. Examples are: the projectile motion, pendulums, and why not, lasers. The latter gives us the capability of predicting what will happen in a $t + \Delta t$. Even chaotic solutions follow this "rule", however the chaotic behavior cannot be predicted because of the extreme sensitivity to initial conditions. One phenomenon observed in chaotic systems is extreme events whose occurrence is often studied statistically. They have a major impact on transport, cybersecurity, climate, and natural catastrophes (earthquakes and floods, among others). In this presentation, we want to shed light on the question, can we predict extreme events? We study both theoretically and experimentally the dynamical behavior of a solid-state laser with modulated cavity losses. This system exhibits a rich dynamical behavior going from the typical relaxation oscillations of this kind of lasers, coexistence of different stable periodic solutions, coexistence between periodic solutions with chaotic solutions, and chaos. Within the chaotic regime, we analyze the necessary conditions for the appearance of extreme events and, in particular, we identify a precursor that appears 20 periods of the modulation of the cavity losses before the extreme event occurs. The latter allows us to determine the initial conditions that give rise to this phenomenon. Furthermore, we study extreme events and synchronization in a network of these lasers.

Acknowledgments

This work has been supported by INCIBE/URJC Agreement M3386/2024/0031/001 within the framework of the Recovery, Transformation and Resilience Plan funds of the European Union (Next Generation EU).

References

[1] Sarrazin, L., Marconi, M., *et al.* Garden of bifurcating paths in a nonlinear optical system. *Physica D*, 475(134553):1–6, 2025.

[1] karin.alfaro@urjc.es

Ω-vector spaces

Patricia Ferrero[♭],[1] Jorge Jiménez[♭] María Luisa Serrano[♭] Andreja Tepavčević[♮] and Braminir Šešelja[♮]

(♭) Departamento de matemáticas, Universidad de Oviedo
Periurbano - Rural, 33203 Gijón, Asturias, Spain.

(♮) Faculty of Sciences, University of Novi Sad
Trg Dositeja Obradovića 3, Novi Sad 21000, Serbia.

1 Introduction

The topic of this research belongs to the field of lattice valued (fuzzy) algebraic structures equipped with generalized equality relation, replacing the classical equality. A basic source for this is the theory of Ω-sets. These structures appeared in 1979, by M.P. Fourman and D.S. Scott [1]. An Ω-set is a nonempty set A equipped with an Ω-valued equality E, with truth-values in a particular complete lattice. Another source is the concept of algebras with a fuzzy equality, introduced by Bělohlávek and Vychodil [2]. On this basis Ω-algebras are introduced by B. Šešelja and A. Tepavčević as algebras with generalized equalities [3].

An Ω-algebra is an algebraic structure equipped with the mentioned Ω-equality, being a particular map from the square of the algebra into a complete lattice. Classical identities are fulfilled as special lattice formulas. The quotient structures obtained by the cut of this generalized equality are classical algebras of the same type fulfilling these identities in the classical sense [4], [5].

Our aim is to introduce and investigate the new structure in this framework, namely Ω-vector spaces.

2 Preliminaries

An **Ω-valued set** μ on a nonempty set A is a function $\mu : A \to \Omega$, where $(\Omega, \leqslant)$ is a complete lattice. An **Ω-valued relation** E **on** A is an Ω-valued set on A^2.

E is **symmetric** if $E(x, y) = E(y, x)$ and E is **transitive** if $(x, z) \wedge E(z, y) \leqslant E(x, y)$ for all $x, y, z \in A$. A symmetric and transitive Ω-valued relation E on A is called an **Ω-valued equality** on A.

Let $\mu : A \to \Omega$ be an Ω-valued set and $E : A^2 \to \Omega$ an Ω-valued relation on A. If for $x, y \in A$, E and μ satisfy

$$E(x, y) \leqslant \mu(x) \wedge \mu(y),$$

then we say that E is an **Ω-valued relation on** μ.

The Ω-valued relation E is **reflexive on** μ if

$$E(x, x) = \mu(x) \text{ for every } x \in A.$$

[1] ferreropatricia@mail.com

If Ω is a complete lattice and $\mathcal{R} = (R, F)$ is an algebra, then the function $\mu : R \to \Omega$ is said to be **compatible** with the operations in F if for any $f \in F$, $f : R^n \to R, n \in \mathbb{N}$, for all $a_1, \ldots, a_n \in R$ and for a nullary operation $o \in F$, {equation* $\bigwedge_{i=1}^{n} \mu(a_i) \leqslant \mu(f(a_1, \ldots, a_n)), \quad \mu(o) = 1.$

Analogously, an Ω-valued relation $E : R^2 \to \Omega$ on $\mathcal{R}$ is **compatible** with the operations in F if the following holds: for every n-ary operation $f \in F$, for all $a_1, \ldots, a_n, b_1, \ldots, b_n \in R$, and for constant $o \in F$, {equation* $\bigwedge_{i=1}^{n} E(a_i, b_i) \leqslant E(f(a_1, \ldots, a_n), f(b_1, \ldots, b_n)); \quad E(o, o) = 1.$

A pair $\overline{\mathcal{A}} = (\mathcal{A}, E)$ is an **Ω-algebra** if $\mathcal{A} = (A, F)$ is an algebra, (A, E) is an Ω-set and E is compatible with the operations in F. We assume that if c is the nullary operation, we have $E(c, c) = 1$. $\mathcal{A}$ is the **underlying, basic algebra** of $\overline{\mathcal{A}}$.

If we do not fix Ω, we say that $\overline{\mathcal{A}} = (\mathcal{A}, E)$ is a **lattice-valued algebra**.

Let $u(x_1, \ldots, x_n) \approx v(x_1, \ldots, x_n)$ (briefly $u \approx v$) be an identity in the type of an Ω-algebra $\overline{\mathcal{A}} = (\mathcal{A}, E)$. We assume, as usual, that variables appearing in terms u and v are from $x_1, \ldots, x_n$. Then, $\overline{\mathcal{A}}$ **satisfies identity** $u \approx v$ (i.e., this identity **holds** on $\overline{\mathcal{A}}$) if for all $a_1, \ldots, a_n \in A$ {equation* $\bigwedge_{i=1}^{n} \mu(a_i) \leqslant E(u(a_1, \ldots, a_n), v(a_1, \ldots, a_n))$.

If Ω-algebra $\overline{\mathcal{A}} = (\mathcal{A}, E)$ satisfies an identity, then this identity doesn't necessarily hold on $\mathcal{A}$. On the contrary, if the supporting algebra fulfills an identity then also the corresponding Ω-algebra does (see [6]).

3 Results

We announce the following results:

Definition 1. Let $\overline{\mathcal{G}} = (\mathcal{G},\ E^\mu)$ be an Ω-group ($\mathcal{G} = (G, *)$ is the underlying groupoid) and let $\mathcal{P} = (P, +, \cdot)$ be a field. We assume the existence of a map $\circ : P \times G \to G$, denoted by the superposition, fulfilling for all $a, b \in G, \alpha \in P$, being α a constant. We also assume

{equation*E(a,b)$\leqslant E(\alpha \circ a, \alpha \circ b)$.

Under these conditions, if the following known formulas hold for $a, b \in G$, $\alpha, \beta \in P$ and ι - the unit under $\cdot$ in P,
$(i)\ \alpha \circ (a * b) \approx \alpha \circ a * \alpha \circ b$;
$(ii)\ (\alpha + \beta) \circ a \approx \alpha \circ a * \beta \circ a$;
$(iii)\ \alpha(\beta \circ a) \approx (\alpha \cdot \beta) \circ a$;
$(iv)\ \iota \circ a \approx a$.
we say that $\overline{\mathcal{G}}$ is an **Ω-vector space over $\mathcal{P}$**.

The above formulas are valid in $\overline{\mathcal{G}}$ if and only if for all $a, b \in G$, $\alpha, \beta \in P$, the following lattice-theoretic formulas hold, respectively:

$(i')\ \ \mu(a) \wedge \mu(b) \leqslant E(\alpha \circ (a * b), \alpha \circ a * \alpha \circ b)$;
$(ii')\ \ \mu(a) \leqslant E((\alpha + \beta) \circ a, \alpha \circ a * \beta \circ a)$;
$(iii')\ \ \mu(a) \leqslant E(\alpha \circ (\beta \circ a), (\alpha \cdot \beta) \circ a)$;
$(iv')\ \ \mu(a) \leqslant E(\iota \circ a, a)$.

If $(\mathcal{G},\ E)$ is an Ω-vector space over a field $\mathcal{P}$, then the map $\mu : G \to \Omega$ such that $\mu(x) = E(x, x)$ is the *domain* of $\overline{\mathcal{G}}$. If $p \in \Omega$, then the *p-cuts* (cuts) of μ and E are subsets of G and G^2 respectively, defined by:

$\mu_p := \{x \in G \mid \mu(x) \geqslant p\}\ \ ;\ \ E^\mu_p := \{(x, y) \in G^2 \mid E^\mu(x, y) \geqslant p\}$.

We also define $\mu_p / E_p = \{[x]_{E_p}$ such that $x \in \mu_p\}$, where $[x]_{E_p} = \{y$ such that $(x, y) \in E_p\}$

Theorem 2. *Let $(\mathcal{G}, E^\mu)$ be an Ω-group, where $(G, *)$ is a groupoid and Ω a complete lattice. Let also $(P, +, \cdot)$ be a field. Then $(\mathcal{G}, E^\mu)$ is an Ω-vector space over P if and only if for every $p \in \Omega$, μ_p / E_p is a*

classical vector space over P.

Definition 3. Let $\overline{\mathcal{G}} = (\mathcal{G}; E^{\mu})$ and $\overline{\mathcal{G}_1} = (\mathcal{G}; E^{\mu_1})$ be Ω-vector spaces over the field $\mathcal{P} = (P, +, \cdot)$. We say that $\overline{\mathcal{G}_1}$ is an Ω-vector subspace of Ω-vector space $\overline{\mathcal{G}}$, if E^{μ_1} is a restriction of E^{μ} to the omega-valued subalgebra μ_1 of $\mathcal{G}$, determined by E^{μ_1}.

Proposition 4. *Let $\overline{\mathcal{G}} = (\mathcal{G}; E^{\mu})$ be an Ω-vector space, and $\{V_i = (\overline{\mathcal{G}}; E^{\mu_i}) | i \in I\}$ a family of Ω-vector subspaces of $\overline{\mathcal{G}}$. The intersection $V = \bigcap_{i \in I} V_i = (\overline{\mathcal{G}}, E^{\mu_W})$, where $E^{\mu_W}(x, y) = \bigwedge_{i \in I} E^{\mu_i}(x, y)$, is an Ω-vector subspace.*

Theorem 5. *Let $\overline{\mathcal{G}} = (G, E^{\mu})$ be an Ω-vector space. The collection of all Ω-vector subspaces of $\overline{\mathcal{G}}$ is a complete lattice under $\leq$.*

4 Conclusions

We define Ω-vector spaces, study properties related to cut-quotients which reduce the framework of Ω-structures to the classical algebra. We also define Ω-vector subspaces and we analyse operations amongst them, such as the intersection. We show some examples to illustrate this.

References

[1] M.P. Fourman, D.S. Scott, Sheaves and logic, in: M.P. Fourman, C.J. Mulvey D.S. Scott (Eds.), *Applications of Sheaves, Lecture Notes in Mathematics*, 753: 302-401,1979.

[2] R. Bělohlávek, V. Vychodil, Algebras with fuzzy equalities, *Fuzzy sets and systems*, 157: 161-201, 2006.

[3] B. Šešelja, A. Tepavčević, Ω-algebras, *Proceedings of ALHawai'i, A conference in honor of Ralph Freese, William Lampe, and J.B*. Nation, 96-106, 2018.

[4] B. Šešelja, A. Tepavčević, Ω-groups in the language of Ω-groupoids, *Fuzzy sets and systems*, 397: 152–167, 2020.

[5] J.Jimenez, M.L.Serrano, B. Šešelja, A. Tepavčević, Omega-rings, *Fuzzy sets and systems*, 455: 183-197, 2023.

[6] B. Budimirović, V. Budimirović, B. Šešelja, A. Tepavčević, *E*-fuzzy groups, *Fuzzy Sets and systems*, 289: 94-112, 2016.

Personalization of Page Rank centrality vector: a fixed-point approach

David Aleja[♭,♮,*], Julio Flores[♭], [1] Eva Primo[♭,♮] Daniel Rodríguez[♭] and Miguel Romance[♭,♮,*]

([♭]) Departamento de Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica, Universidad Rey Juan Carlos
28933 Móstoles (Madrid), Spain

([♮]) Laboratory of Mathematical Computation on Complex Networks and their Applications, Universidad Rey Juan Carlos,
28933 Móstoles (Madrid), Spain

(*) Data, Complex networks and Cybersecurity Research Institute, Universidad Rey Juan Carlos, 28028 Madrid, Spain

1 Introduction

Let $\mathcal{G} = (V, E)$ be a directed graph where $V = \{1, 2, \ldots, N\}$ is the set of nodes and $N \in \mathbb{N}$. The *adjacency matrix* of $\mathcal{G}$ is an $N \times N$-matrix

$$A = (a_{ij})_{i,j} \quad \text{where} \quad a_{ij} = \begin{cases} 1, \text{ if } (i,j) \text{ is a link of } \mathcal{G} \\ 0, \text{ otherwise.} \end{cases}$$

We denote $k_{out}(i)$ the *outdegree* of node i, i.e, the number of outlinks of a node i. The graph $\mathcal{G} = (V, E)$ may have *dangling nodes*, which are nodes $i \in V$ with zero outdegree. Dangling nodes are characterized by a vector $\mathbf{d} \in \mathbb{R}^{N\times 1}$ with components d_i defined by

$$d_i = \begin{cases} 1, \text{ if } i \text{ is a dangling node of } \mathcal{G} \\ 0, \text{ otherwise} \end{cases} .$$

Let $P_A = (p_{ij})_{i,j} \in \mathbb{R}^{N\times N}$ be the row-stochastic matrix associated to the directed graph $\mathcal{G} = (V, E)$ defined as follows:

1. If i is a dangling node, then $p_{ij} = 0$ for all $j = 1, \ldots, N$.
2. Otherwise, $p_{ij} = a_{ij}/k_{out}(i) = a_{ij}/\sum_k a_{ik}$.

Roughly speaking, each coefficient p_{ij} is the probability of moving from the node i to the node j. The matrix P_A will be referred to as the *row-normalization* of A.

In addition to the links given by the adjacency matrix that allow for (topologically) traveling between nodes, the well-known PageRank algorithm contemplates the possibility of randomly

[1] julio.flores@urjc.es

jumping from one node to another according to some probability distribution vector, or *personalization vector*, $\mathbf{v} \in^{N\times 1}$ that collects the probability of teleportation between nodes. If dangling nodes exist, an additional probability vector $\mathbf{u} \in \mathbb{R}^{N\times 1}$ is also considered to provide an extra probability of jumping from these nodes. Finally, $\lambda \in (0,1)$ is a decision parameter that determines how likely it is that we move topologically throughout the graph or take leaps between nodes instead, with the probability given by $\mathbf{v}$.

Formally, let $G = G(\lambda, \mathbf{u}, \mathbf{v})$ the (row-stochastic) Google matrix, with $\lambda \in (0,1)$ defined as

$$G = \lambda(P_A + \mathbf{d}\mathbf{u}^T) + (1-\lambda)\mathbf{e}\mathbf{v}^T \in \mathbb{R}^{N\times N}. \tag{1}$$

The personalization vector $\mathbf{v} \in \mathbb{R}^{n\times 1}$ has all its entries positive (i.e. $\mathbf{v} > 0$) and satisfies $\mathbf{v}^T\mathbf{e} = 1$. Similarly, the vector $\mathbf{u} \in \mathbb{R}^{N\times 1}$ is such that $\mathbf{u} > 0$ and $\mathbf{u}^T\mathbf{e} = 1$.

The *PageRank vector* $\pi = \pi(\lambda; \mathbf{u}; \mathbf{v})$ is the unique positive vector which is an eigenvector of the matrix G^T with the associated eigenvalue 1 such that $\pi^T\mathbf{e} = 1$, i.e., $\pi > 0$, $\pi^T\mathbf{e} = 1$ and $G^T\pi = \pi$. In this sense, the *PageRank vector* $\pi > 0$ satisfies the conditions $\pi^T\mathbf{e} = 1$ and $\pi^T G = \pi^T$.

From equation (1) and the properties of the PageRank vector $\pi \in^{N\times 1}$, a straightforward calculation shows that π satisfies the equation

$$\pi^T = \lambda\pi^T P_A + (1-\lambda)\mathbf{v}^T$$

or equivalently

$$\pi^T = (1-\lambda)\mathbf{v}^T(I_N - \lambda P_A)^{-1} = \mathbf{v}^T X(\lambda), \tag{2}$$

where $X(\lambda) = (1-\lambda)R(\lambda)$ and $R(\lambda) = (I_N - \lambda P_A)^{-1}$ is the resolvent of P_A defined for all suitable damping factor $\lambda \in (0,1)$.

In this talk we ask ourselves if, among all possibilities, there is some particularly suitable personalization vector to choose. And perhaps the most natural answer is to look for a personalization vector that coincides with its corresponding PageRank vector, in other words to look for a fixed point in the equation above. Once this point of view is adopted it is equally natural to expect that an iteration process will lead to the sought vector. Iteration here means of course that we start with some initial personalization vector, compute the associated PageRank and then take this PageRank as a new personalization vector to produce another PageRank that will, in turn, be used as a new personalization vector ...

2 Results

Of course some important questions need to be addressed. Is there a limit for this sequence of PageRank vectors obtained by iteration? If so, does the limit depend of the initial personalization vector $\mathbf{v} \in^{N\times 1}$? In this talk we will show that this iterative process under the action of the matrix $X(\lambda)$ indeed converges when the row-normalization matrix P_A is irreducible, and in fact will show that the limit vector is the left-hand Perron vector associated to P_A. We recall that the non negative square matrix P_A is irreducible if the graph represented by A is connected, and conversely.

Let A be a non-negative irreducible square matrix of order N and let P_A be its row-normalization. For a fixed $\lambda \in (0,1)$, and any positive vector $\mathbf{x}_0 \in^{N\times 1}$ with $\|\mathbf{x}_0\|_1 = 1$, the recursive sequence $\{\mathbf{x}_k\}_{k\geq 0} \subset^{N\times 1}$

$$\mathbf{x}_k^T = \mathbf{x}_{k-1}^T X(\lambda) = \mathbf{x}_0^T X(\lambda)^k, (k \geq 1)$$

converges to the left-hand Perron vector $\mathbf{c} \in^{N\times 1}$ of the matrix P_A.

The case of a non-irreducible matrix is, as expected, algebraically quite more involved and it will also be addressed in this talk. More importantly it admits a very natural interpretation in terms of what we call *dangling clusters* which are but groups of nodes that share links exclusively among them with no outlinks to nodes outside the group.

3 Conclusions

In the talk we take a fixed-point approach in the quest of a distinguished canonical personalization vector for the Google matrix associated to a directed graph. I will be emphasized how the theoretical results of purely algebraic nature perfectly match the empirical results obtained in the examples provided.

Acknowledgement

This work has been supported by INCIBE/URJC Agreement M3386/2024/0031/001 within the framework of the Recovery, Transformation and Resilience Plan funds of the European Union (Next Generation EU).

References

[1] Aleja, D., Flores, J., Primo, E. and Romance, M. Time-dependent personalized PageRank for temporal networks: Discrete and continuous scales. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(8), 2024.

[2] Boldi, P., Santini, M. and Vigna, S. PageRank: functional dependencies.*ACM Transactions on Information Systems (TOIS)*, 27(1), 2009.

[3] Berman, A., and Plemmons, R.J. Nonnegative Matrices in the Mathematical Sciences, New York, Academic Press, 1979.

[4] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.U., Complex Networks: Structure and Dynamics, *Phys. Rep*, 424: 175, 2006.

[5] Boccaletti, S. Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z. and Zanin, M. The structure and dynamics of multilayer networks, *Physics Reports*, 544(1): 1-122, 2014.

[6] Gantmacher, F.R., The Theory of Matrices (Vol. II). Rhode Island, American Mathematical Soc., 2000.

[7] Gockenbach, M.S., Finite-Dimensional Linear Algebra. CRC Press Taylor& Francis Group, 2011.

[8] Horn, R.A. and Johnson C.R., Matrix Analysis. Cambridge, Cambridge University Press, 1985.

[9] Meyer, C.D.. Matrix Analysis and Applied Linear Algebra. Philadelphia, Society for Industrial and Applied Mathematics, 2000.

[10] Schwarz, S.. A new approach to some problems in the theory of non-negative matrices. *Czechoslovak Mathematical Journal*, 16: 274-284, 1966.

[11] Schwarz, S. New kinds of theorems on non-negative matrices. *Czechoslovak Mathematical Journal*, 16: 285-295, 1966.

[12] Varga, R.S. Matrix Iterative Analysis. New Jersey, Prentice-Hall, 1962.

Analysis of the Evolution of Student Enrollment Preferences in Degree University Programs Through Directional Community Analysis (2014–2023)

José-Miguel Montañana[♭] Antonio Hervás[♮],[1] and Pedro-Pablo Soriano-Jiménez[♯]

(♭) inpeek GmbH,
lda-Rhodes-Straße 1, 64295 Darmstadt, Germany.
jmmontanana@gmail.com, jose.montanana@inpeek.de

(♮) IUMM, Universitat Politècnica de València, Spain,
Camino de Vera, s/n 46022 Valencia, Spain.
ahervas@mat.upv.es,

(♯)SEPyC. Universitat Politècnica de València, Camino de Vera, s/n 46022 Valencia, Spain.
psoriano@upvnet.upv.es.

1 Introduction

The access procedure to the Spanish Public University System (SUPE) is a complex, multi-step process that differs significantly from both the Spanish private system and other international university systems. Each year, public universities across Spain offer a set number of places for new first-year students in each degree program. After completing high school, students must pass an entrance exam, known as the University Entrance Exam (EBAU/PAU). Their final placement in a degree program is determined by a weighted formula combining exam scores, high school grades, and performance in subjects related to the degrees they wish to apply for [1, 3, 4].

During the application process, students rank their degree preferences in order, from their most desired option to their less preferred alternatives.

In our previous study, the development of a structural model allowed us to construct a graph representing student mobility between university programs. This model offered deeper insights into how student mobility unfolds. Now, a decade later, university offerings have changed, and while the model may have helped shape these new offerings, it certainly allows us to study their evolution. By analyzing the behavior of this graph over the last 10 years, we can observe shifts in student preferences and program demand [10, 11, 12, 13].

We have chosen the Valencian Community as our regional focus within the SUPE system, which consists of 5 public universities (UA, UPV, UV, UMH, UJI), 1 semi-public online university (VIU), and 4 private universities (UCH-CEU, UCV, UEV)[2, 5].

[1] correspondingauthor@mail.com

2 Background

In the context of the Public University System (SUPE and SUPCV), object of interest in this paper, the nodes represent different degree programs, and the edges represent student transitions between programs, indicating the flow of students from one degree to another.

A community within a graph refers to a subset of nodes that are more densely connected to each other than to the rest of the graph. These communities often reveal underlying patterns or groupings in the data, such as clusters of degree programs that share strong student mobility. Community detection is a critical tool for understanding the structure of large networks, as it helps identify meaningful groupings that can reveal insights into the relationships between entities represented in the graph [7, 9].

Furthermore, tracking how this graph evolves over time, especially with the introduction of new degree programs and the rise of dual degrees, provides a more comprehensive view of the system's evolution.

Dual degrees, which allow students to obtain two qualifications through a predefined combination of courses, introduce new pathways in the graph and thus change its structure, making community detection even more crucial for understanding these shifts.

3 Methods

In this section, we outline the process and methodology used to analyze student mobility within the Spanish Public University System (SUPE), focusing on the public universities in the Comunidad Valenciana. Our approach revolves around the construction and analysis of directed graphs, where the nodes represent degree programs and the edges reflect student transitions based on their ranked preferences during the university admissions process.

We first describe how we construct the graphs from the input data, which includes student preferences and degree program information. According with [8], we build a graph for each academic year and assign weights to the edges based on student enrollment probabilities. This allows us to model the flow of students between degree programs based on their academic scores and admission likelihood.

To refine the graph and improve the accuracy of community detection, we propose a pruning technique inspired by edge-detection methods used in image processing [13].

In the case of student preferences for university degree selection, it is crucial to prune edges that represent improbable transitions between degree programs, as these may occur due to errors in data collection or outliers in student behavior. Since the mobility patterns between degree programs can reflect strong preferences or incidental selections, pruning irrelevant or noisy edges is essential to ensure that community detection algorithms can accurately identify meaningful clusters of degree programs.

The purpose of this pruning process is to refine the graph by removing weak, irrelevant edges, particularly those that connect different communities of degrees. By doing so, we ensure that the community detection algorithms are able to operate more effectively on a cleaner, more representative graph, which captures the true student mobility patterns without distortion caused by noise or outliers.

4 Results

4.1 Analysis of Student Preferences Through Pruned Graphs (2014-2023)

The results of applying our proposed method alongside the previously obtained optimal results through manual adjustments [8], show that the graphs appear nearly identical. This similarity demonstrates the consistency and effectiveness of our approach. Notably, all edges remaining after pruning with our method are also present in the previous method. However, our method successfully removes a few additional edges that the older method retained, highlighting its precision in eliminating unnecessary connections.

Regarding modularity and community detection, the results from the previous methods show some variation: the "directed-in" version identified 57 communities with a modularity of 0.716, while the "directed-out" version detected 30 communities with a modularity of 0.838. In contrast, our method identifies the same number of communities as the "directed-in" version (57) but achieves a significantly higher modularity of 0.971. This higher modularity suggests a much stronger and more coherent community structure.

Overall, this demonstrates that our pruning technique not only reduces manual intervention required in the previous method, but also enhances the graph's quality by improving modularity. The higher modularity achieved by our method reflects communities that are more clearly defined, providing a more accurate representation of student mobility patterns compared to previous methods.

Notice that pruning method is based solely on the weights of the links, and those values are not used in the calculation of modularity. Therefore, it cannot be stated in advance that the proposed pruning method will necessarily improve or worsen modularity.

Table 1: Modularity, number of communities, and total number of edges after applying the pruning algorithm to each year graph

Year	Number of Vertices	Number of Edges before pruning	Number of Edges after pruning	Number of communities	Modularity
2014	215	9720	161	62	0.9750
2015	222	10100	173	59	0.9684
2016	227	10906	171	65	0.9723
2017	230	10903	174	62	0.9715
2018	240	11410	181	67	0.9717
2019	244	11378	182	73	0.9747
2020	249	12009	186	74	0.9724
2021	256	12487	194	75	0.9752
2022	274	12275	211	72	0.9769
2023	279	12508	197	88	0.9812

Implications of the Stability of Communities

The stability of communities as defined provides a metric for assessing community dynamics over time, offering insights into both the resilience and the transformation of community structures [6].

Lower stability values for a community indicate shifts in community composition, suggesting either the attraction of new members or the migration of existing ones to other communities, which may due to changing interests and emerging trends. In contrast, higher stability values reflect a structure where members consistently retain their associations within the community across consecutive years.

Table 2: Stability of Communities

	2014–15	2015–16	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22	2022–23
Average Stab.	0.6438	0.6611	0.6539	0.6451	0.6486	0.6357	0.6529	0.6442	0.6442
WAS_NV	0.6950	0.7276	0.6916	0.7007	0.6923	0.6858	0.7095	0.6882	0.6876
WAS_FC	0.6835	0.7198	0.6963	0.7034	0.6881	0.6393	0.6648	0.6473	0.6757

These results, together with the graphs, show that approximately two-thirds of the degree programs and the students' first-choice preferences remain within the same communities from one year to the next, while the remaining third tends to move between each community and its neighboring ones.

This pattern can be of primary interest to those responsible for promoting degree programs, as it suggests that efforts should first focus on attracting students who have an interest in the fields within the same community, and secondarily on students interested in programs in neighboring communities.

5 Evolution of Student Demand Across Academic Disciplines

This section reviews the structural changes in student demand across various academic disciplines in the SUPE between 2014 and 2023. By examining the pruned graphs for each discipline, we identify how the introduction of new degrees, the establishment of double-degree programs, and the restructuring of existing academic offerings have influenced community formation based on student preferences.

Next, different disciplines are considered, highlighting notable shifts in demand, the emergence or dissolution of academic communities, and variations in inter-university mobility patterns.

Evolution on the Engineering degrees for the years 2014 and 2023

Table 3: Changes on edges reveal shifts in the enrollment demand

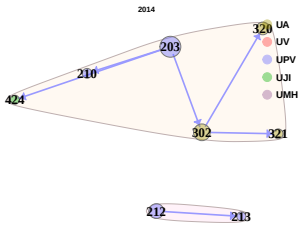
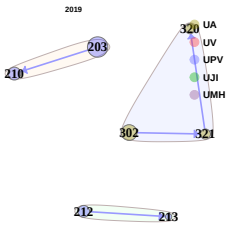
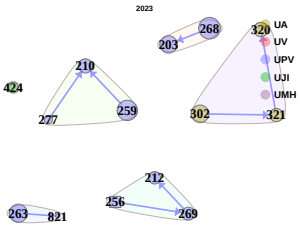
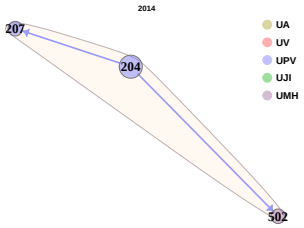
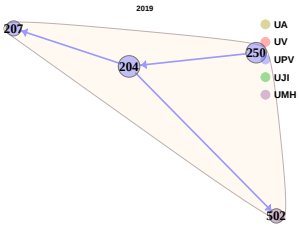
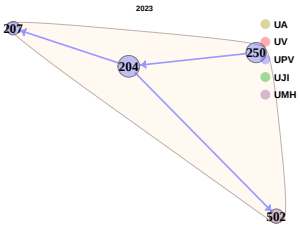
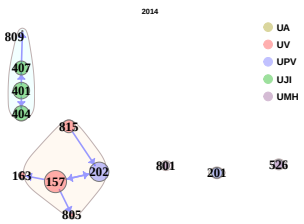
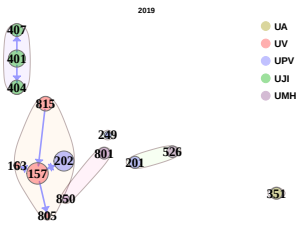
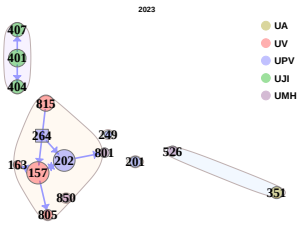
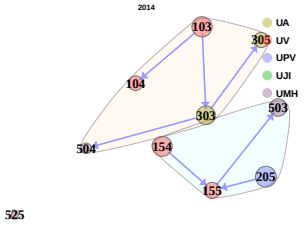
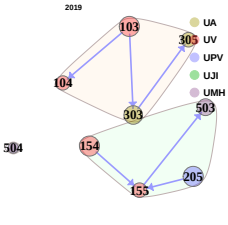
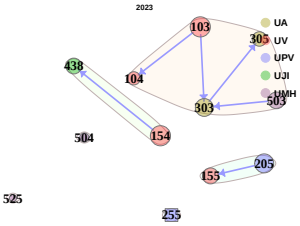
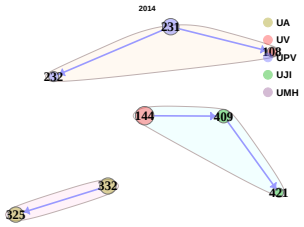
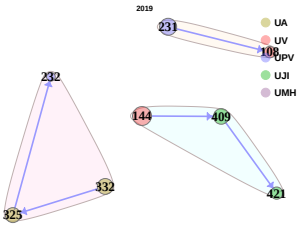
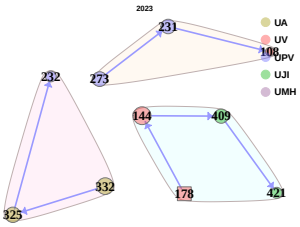
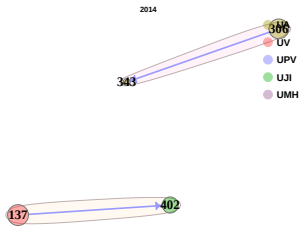
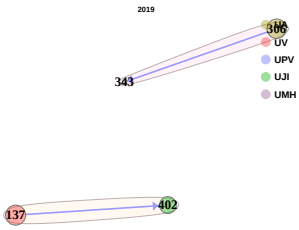
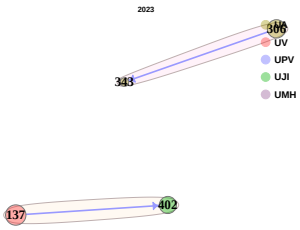
Degrees	Weight(2014)	Weight(2023)	Weight difference	Change
270 → 223	-	32.08	32.08	New Flow
229 → 228	-	7.27	7.27	New Flow
229 → 219	-	5.39	5.39	New Flow
817 → 214	-	1.64	1.64	New Flow
810 → 221	-	0.49	0.49	New Flow
221 → 151	3.74	4.66	0.92	Increased
229 → 811	0.56	1.09	0.53	Increased
420 → 417	5.04	5.33	0.29	Increased
217 → 218	5.53	5.64	0.11	Increased
228 → 220	2.17	1.45	-0.72	Decreased
418 → 420	6.16	4.14	-2.02	Decreased
221 → 219	8.91	5.19	-3.72	Decreased
223 → 238	9.51	2.65	-6.86	Decreased
229 → 221	14.60	6.60	-8.00	Decreased
417 → 420	5.04	-	-5.04	Discontinued
223 → 229	14.02	-	-14.02	Discontinued
209 → 223	12.60	-	-12.60	Discontinued
223 → 214	12.68	-	-12.68	Discontinued

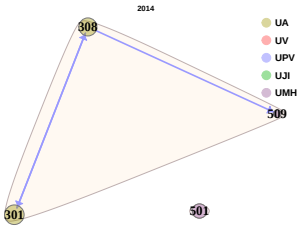
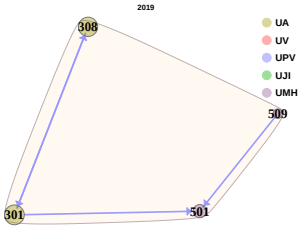
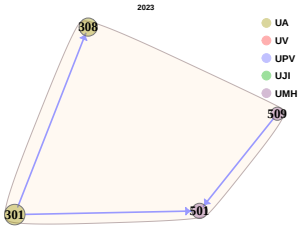
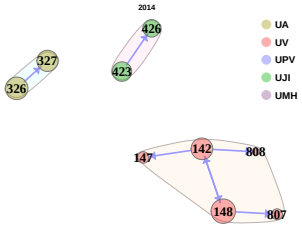
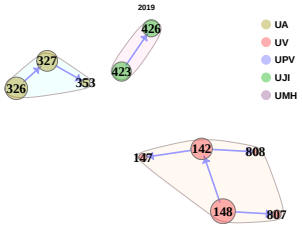
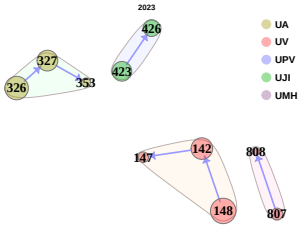
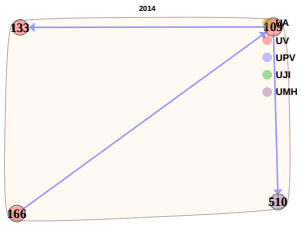
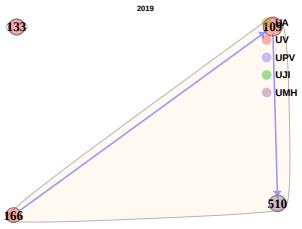
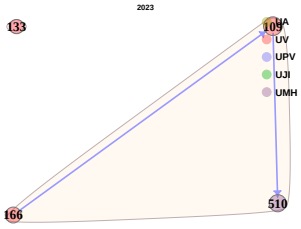
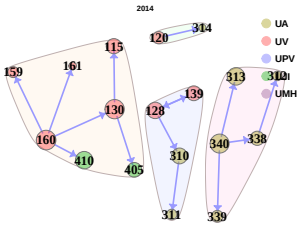
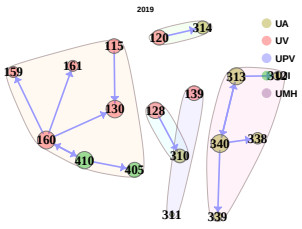
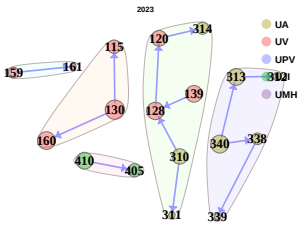
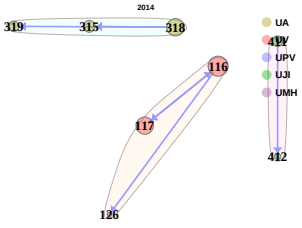
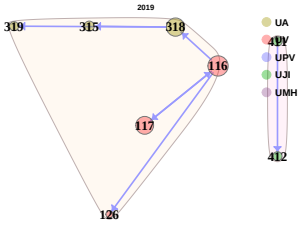
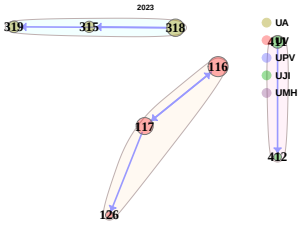
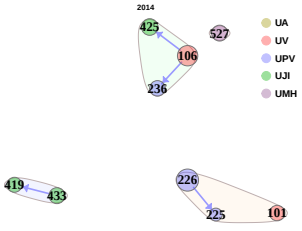
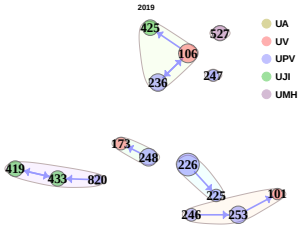
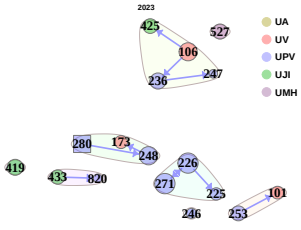
- **New Degrees are incorporated** 270 and 817.

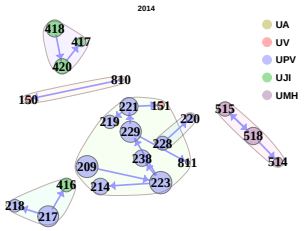
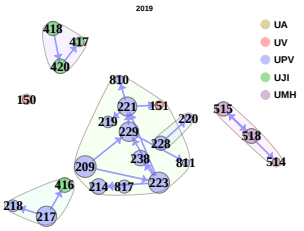
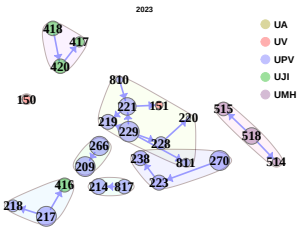
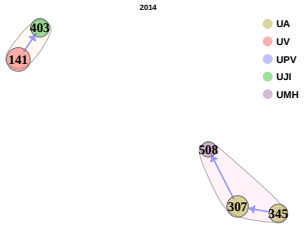
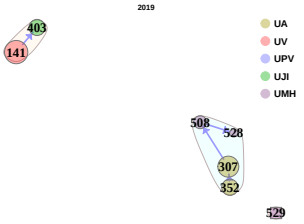
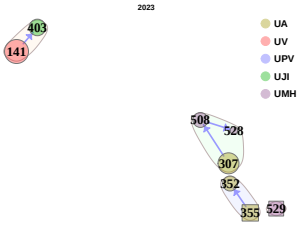
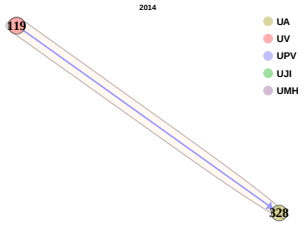
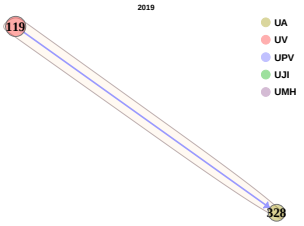
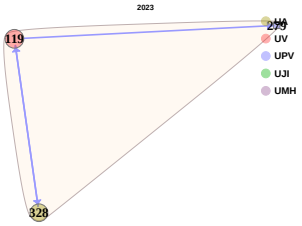
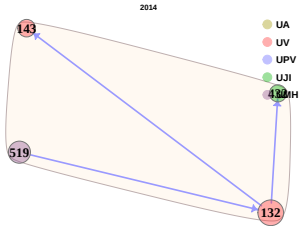
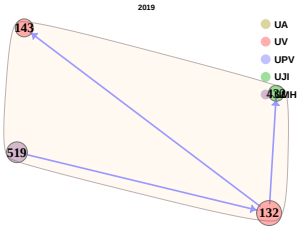
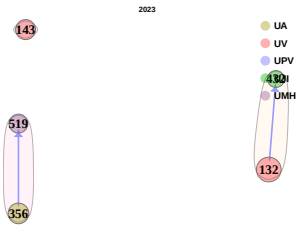
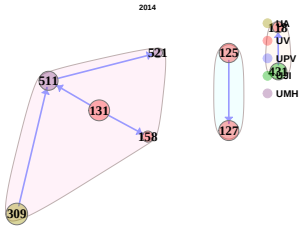
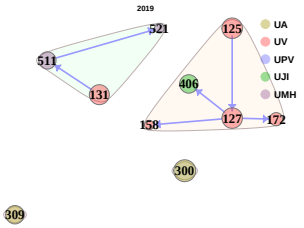
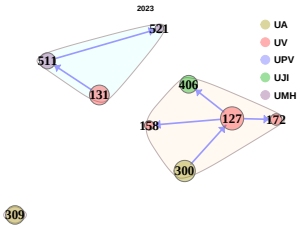
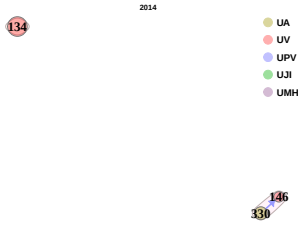
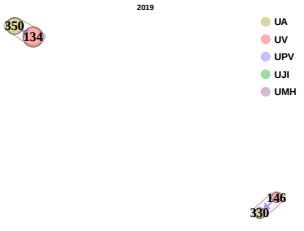
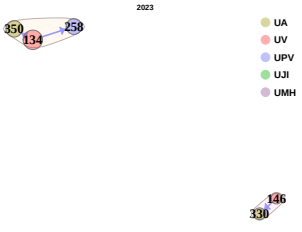
- **Changes on the Enrollment Demand Paths:** Changes on paths reflect evolving student preferences and shifting academic focus areas. Particularly the notable edge from **270 to 223** with a strong new connection in enrollment interest.
- **Community Consolidation:** We can see that amount of communities increased, suggesting a clustering of the demand.

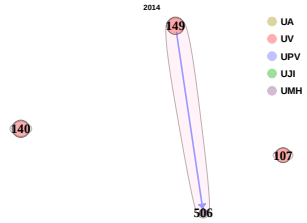
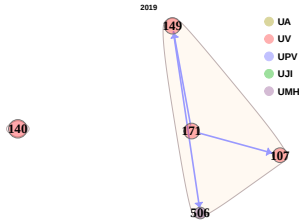
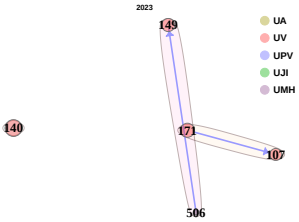
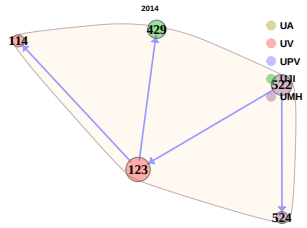
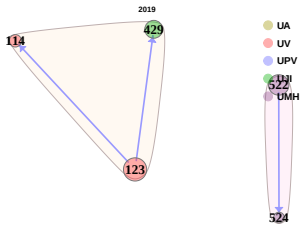
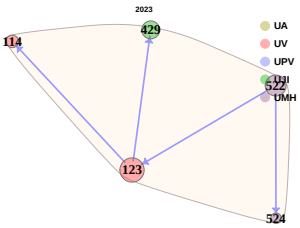
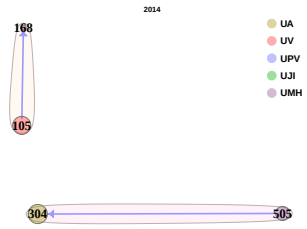
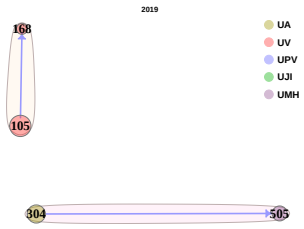
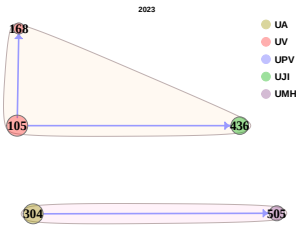
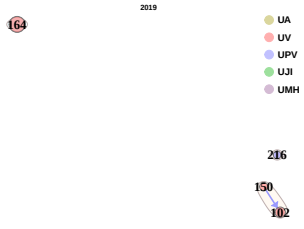
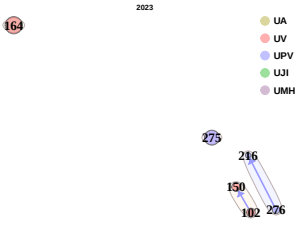
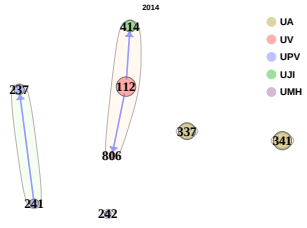
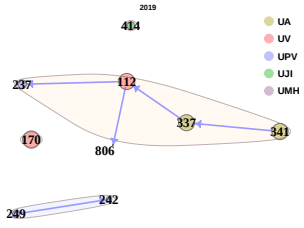
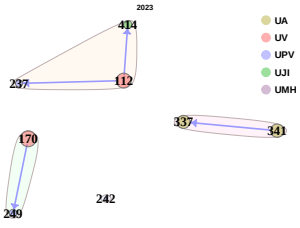
Discipline	Key Observations
Biology and Environmental & Forestry Sciences	The introduction of the double degree in Forestry Engineering and Natural Environment & Environmental Sciences (UPV) has merged three communities, making it a key attraction point, particularly for students from other universities.
Law	By 2024, the integration of double degrees with Law has strengthened the cluster, reducing fragmentation from two to one while increasing the number of nodes. Law remains the main attractor in the community.
Physics and Mathematics	In 2015: Only five vertices across two communities. In 2024: A single large community with 18 closely related degrees. The creation of double degrees in Physics & Mathematics and Mathematics & Engineering has significantly boosted interest in this area.
Biotechnology and Agri-Food Sciences	This field transitioned from four communities with five vertices to a single, larger community with seven vertices. The introduction of the double degree in Biotechnology and Agri-Food Sciences has unified the group and increased student demand.
Telecommunications	The emergence of double degrees in Telecommunications has improved the field's overall situation. However, this effect is localized, only impacting the university that introduced the program.
Computer Science	**Valencia:** A similar effect as Telecommunications, where double degrees significantly boost demand, but mobility between universities remains low, likely due to widespread availability. **Alicante:** Internal cluster structure changes, but effects remain localized to the offering university.
Nursing	Since 2015, all nursing programs have become part of a single community, with an overall increase in demand. This consolidation has stimulated student mobility between universities.
Humanities (History)	In 2015: Five separate communities. In 2024: A single, unified community. However, student mobility between universities remains relatively low.

The analysis of student preferences across disciplines reveals clear trends in the evolution of demand. The introduction of double-degree programs has played a crucial role in shaping community structures, often leading to greater integration and consolidation within fields. While some disciplines have experienced increased inter-university mobility (e.g., Nursing), others remain more localized (e.g., Computer Science, History). The observed transformations suggest that strategic academic offerings significantly impact student choices and institutional attractiveness over time.

	2014	2019	2023
Architecture			
Art			
Business			
Biology			
Chemistry			
Criminology			

	2014	2019	2023
Economy			
Education			
Pharmacy			
Philology			
History			
Computing			

	2014	2019	2023
Engineering			
Law			
Math			
Medicine			
Nursing			
Physics			

	2014	2019	2023
Politics			
Psychology			
Sports			
Telecom			
Tourism			

6 Discussion

The results obtained after applying the pruning method to the student preferences within the Spanish Public University System (SUPE), particularly in the Comunidad Valenciana, highlights several key findings and insights into how degree offerings have evolved over the past decade. One major trend observed is that the introduction of double-degree programs has significantly altered student movement patterns. Instead of shifting between entirely different academic fields, students now tend to choose double degrees that expand their qualifications while staying within closely related disciplines. In particular, in disciplines such as Biology and Environmental Sciences, Physics and Mathematics, and Biotechnology, the creation of new double-degree options has not only led to the merging of previously distinct communities but has also reinforced certain academic clusters as strong attractors for prospective students.

Moreover, institutional policies influence student mobility in significant ways.

The emergence of new degree offerings, particularly in interdisciplinary fields, has driven changes in preference structures. The evolution of these patterns suggests that future academic strategies should prioritize adaptability and responsiveness to emerging student interests. Future research should further investigate the long-term effects of these mobility trends, including their impact on student outcomes, employability, and overall higher education system efficiency.

References

[1] Universidad Politécnica de Valencia. Información sobre preinscripción. Enlace: `https://wiki.upv.es/confluence/pages/viewpage.action?pageId=96927778`. Last accessed October 15, 2024.

[2] Portal de la Transparencia, Administración General del Estado. Datos y cifras del sistema universitario español 2024. Enlace: `https://www.universidades.gob.es/wp-content/uploads/2024/05/SIIU_DatosCifras2024.pdf`. Last accessed October 15, 2024.

[3] Carina Farreras, "La selectividad tendrá el mismo número de exámenes que antes y se elegirá entre historia de España y filosofía," *La Vanguardia*, 2023, `https://www.lavanguardia.com/vida/20231011/9269182/selectividad-2023-2024-mismo-numero-examenes-historia-filosofia.html`.

[4] "La nueva selectividad: 5 cambios recientes a tener en cuenta," *Educaweb*, 2023, `https://www.educaweb.com/noticia/2023/10/11/nueva-selectividad-5-cambios-recientes-tener-cuenta-20380/`.

[5] Servicio de Regulación Universitaria, Dirección General de Universidad, Investigación y Ciencia. "Preinscripción Universitaria." Available online: `https://ceice.gva.es/es/web/universidad/preinscripcion` (accessed on 15 October 2024).

[6] Delvenne, J-C. Yaliraki, S. N., and Barahona, M. Stability of graph communities across time scales *Proc Natl Acad Sci (PNAS)*. Jun 30;107(29): 12755-12760, 2010.

[7] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5), 175–308, 2006.

[8] Hervás, A., Soriano-Jiménez, P. P., Jiménez, A., Peinado, J., Capilla, R., and Montañana, J. M. A Case Study of Directional Communities in a Directd Graph: The Accessing Procedure to the Spanish Public University System. Modeling Human Behavior, 9. Nova Science Publishers, Inc. Hauppauge, NY. USA. 2015.

[9] Lancichinetti A, and Fortunato S. Community detection algorithms: A comparative analysis. *Phys. Rev. E* 80(5):056117. 2009.

[10] Montañana, J. M., Hervás, A., and Soriano Jiménez, P. P. A Proposal for a Benchmark Generator of Weakly Connected Directed Graphs. *Open Journal of Modelling and Simulation*, 8(1): 18-34. 2020.

[11] Montañana, J. M., Hervás, A., and Méndez, A. Communities' detection in weakly connected directed graphs. In Proceedings of the Mathematical Modelling in Engineering & Human Behaviour, 72-77. ISBN 978-84-09-25132-2 SPUPV, Valencia, Spain. 2020.

[12] Montañana, J. M., Hervás, A., Morillas, S., and Méndez, A. A procedure for detection of border communities using convolution techniques. In Proceedings of the Mathematical Modelling in Engineering & Human Behaviour, 267-271. ISBN 978-84-09-36287-5 SPUPV, Valencia, Spain. 2021.

[13] Montañana, J. M., Hervás, A., Morillas, S., and Méndez, A. Detection of border communities using convolution techniques. In Proceedings of the Mathematical Modelling in Engineering & Human Behaviour, 133-138. ISBN 978-84-09-47037-2 SPUPV, Valencia, Spain. 2022.

PageRank with Time-Varying Personalization for temporal networks

D. Aleja [♭] [♮], J. Flores[♭], E. Primo[♭] [♮] [1] and M. Romance[♭] [♮]

([♭]) Department of Applied Mathematics, Science and Engineering of Materials, and Electronic Engineering, ESCET, Universidad Rey Juan Carlos, Móstoles (Madrid, Spain).

([♮]) Laboratory of Computational Mathematics on Complex Networks and their Applications, Universidad Rey Juan Carlos, Móstoles (Madrid, Spain).

Abstract

Complex networks ubiquitously model systems in nature, society, and technology, often exhibiting time-evolving interactions captured by temporal networks. Although much work focuses on discrete-time interactions, the continuous nature of time necessitates studying systems with continuously varying interactions. Understanding centrality in such dynamic systems is critical, and PageRank is a fundamental measure (see [5]). Existing adaptations of PageRank to temporal networks often assume a static personalization vector [2]. This work addresses these limitations by providing a comprehensive analysis of personalized PageRank in temporal networks with discrete and continuous time scales, crucially incorporating time-varying personalization vectors [1]. Furthermore, it explores the interplay between these scales through discretization and establishes localization results for the proposed PageRank definitions.

The primary objectives are twofold: 1) to provide robust definitions of the PageRank eigenvector for temporal networks in both discrete and continuous time scales, and to demonstrate the approximation of the continuous temporal network's PageRank vector by those of its discrete counterparts; 2) to highlight the significance of considering time-varying personalization vectors in temporal PageRank and to derive localization results that bound the influence of these varying preferences on individual node PageRank.

For a discrete sequence of time instants $I = \{t_1, \ldots, t_N\}$, the PageRank vector $\pi^T(t_k)$ at time t_k is defined as the dominant positive eigenvector of a time-dependent Google matrix

$$G(t_k) = \lambda(t_k)P_B(t_k) + (1 - \lambda(t_k))ev^T(t_k).$$

$P_B(t_k)$ is derived from the accumulated adjacency matrices $A(t_l)$ for $l \leq k$, weighted by $\omega_{lk} \geq 0$. Different time-varying personalization vectors $v(t)$ (Uniform, Input, Inverse-Input) are analyzed. For continuous interactions on $I = [t_0, t_1]$, the PageRank vector $\pi(t)$ is defined via a continuous analogue. A weight function $\omega(s,t) \geq 0$ for $s \leq t$ defines $B(t)$ by integrating $A(s)$, leading to

$$G(t) = \lambda(t)P(t) + (1 - \lambda(t))ev^T(t),$$

where $P(t)$ normalizes $B(t)$. $\pi(t)$ satisfies $\pi(t)^T G(t) = \pi(t)^T$.

[1] eva.primo@urjc.es

In conclusion, this research provides a comprehensive framework for analyzing PageRank in temporal networks, emphasizing the crucial roles of time-varying personalization and the underlying time scale. The ability to approximate continuous PageRank via discrete sampling offers practical insights for analyzing real-world networks. The localization results deepen the understanding of personalization's influence in dynamic settings. The study underscores the necessity of accounting for the temporal evolution of both network structure and personalization preferences in centrality analysis.

Acknowledgments

This work has been partially supported by projects M2978 and M3033 (URJC Grants) and PID2019-107701G (Ministerio de Ciencia e Innovación Grant, Spanish Government). The data that support the findings of this study are openly available in http://konect.cc/networks/, reference number Ui (link-dynamic-itwiki).

References

[1] Aleja, D., Flores, J., Primo, E., and Romance, M., Time-dependent personalized PageRank for temporal networks: Discrete and continuous scales. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(8),2024.*

[2] Ahmad, W., Porter, M. and Beguerisse-Díaz, M., Tie-Decay Networks in Continuous Time and Eigenvector-Based Centralities, *IEEE Trans. Networks Science and Eng.*, 8(2)1759—1771, 2021.

[3] Flores, J. and Romance, M., On eigenvector-like centralities for temporal networks: discrete vs. continuous time scales, *J. Comp. and Appl. Math.*, 330:1041–1051, 2018.

[4] García, E., Pedroche, F. and Romance, M., On the localization of the personalized PageRank of complex networks, *Lin. Alg. Appl.*, 439(3):640–652, 2013.

[5] Page, L. at al., The PageRank Citation Ranking: Bringing Order to the Web, *Technical Report* SIDL-WP-1999-0120, Stanford Digital Library Technologies Project, 1999.

Part X

Financial Mathematics

Bounded versus unbounded good deal indices in asset pricing

Alejandro Balbás[1] *

(1) University Carlos III of Madrid,
C/ Madrid, 126, 28903 Getafe (Madrid, Spain).

1 Introduction

Portfolio choice problems involving both continuous-time arbitrage free pricing models and coherent risk measure often present "pathologies" such as the existence of sequences of strategies whose expected risk premium tends to plus infinity whereas the risk tends to minus infinity. Former studies have created "a good deal index" with two different interpretations, namely, the maximum price change of a portfolio whose initial price equals one that prevents the existence of the previous sequence, and the minimum risk of a portfolio whose price equals one. This paper focuses on those cases where the index above becomes unbounded and equals plus infinity. Comparisons between risk measures with bounded and unbounded index will be implemented. In particular, the pricing rule modifications will be compared, as well as the portfolios whose price must reflect the maximum modification. Implications for practitioners will be given, in the sense that under- and over-priced strategies will be computed. Once again, the bounded or unbounded nature of the good deal index will play an essential role in the design of these investment strategies.

2 Methods

The good deal index is defined by means of an infinite-dimensional convex programming problem maximizing the expected risk premium and simultaneously minimizing the level of risk. This problem will be often unbounded, and consequently, additional constraints will be required. A dual approach will allow us to provide a new interpretation of the index in terms of "minimum pricing rule modification". The primal problem will provide us with high yield portfolios, whereas the dual one will generate alternative stochastic discount factors. Thinks will become quite different if the good deal index still equals plus infinity after the additional primal constraints. In such a case the initial "pricing errors" become huge and the potential (but improbable) capital losses are bounded. Accordingly, the dependence of the under- and over-priced portfolios with respect to the selected risk measure will play a critical role in this paper. Practical criteria will be given for constructing high-return strategies that are independent of the risk measure chosen.

* alejandro.balbas@uc3m.es

3 Results

High-yield investment strategies that are independent of the chosen risk measure will be explicitly presented.

4 Conclusions

There are important caveats/pathologies affecting the usual stochastic continuous-time pricing models. In particular, there are sequences of strategies whose expected risk premium tends to plus infinity whereas their coherent risks tends to minus infinity. The study of such strategies is very important because it should have two significant effects. On the one hand, high-yield strategies should be found. On the other hand, such strategies should suggest the way to modify the classical pricing models so as to prevent the existence of the pathology above.

References

[1] Artzner, P., Delbaen, F., Eber, J.M., Heath, D. Coherent measures of risk. *Mathematical Finance*, 9, 203 – 228, 1999.

[2] Balbás, A., Garrido, J., OKhrati, R. Good deal índices in asset pricing: actuarial and financial implications. *International transactions in operational research*, 26, 1475 – 1503, 2019.

[2] Duffie, D. Security markets: stochastic models. Academic Press, Boston, MA, 1988.

Pricing FX contracts with time-delay stochastic processes

L. Gómez-Valle[♭],[1] J. Martínez-Rodríguez[♭] and M. Zani[♮]

([♭]) Departamento de Economía Aplicada e IMUVA, Facultad de Ciencias Económicas y Empresariales, Universidad de Valladolid, Valladolid, Spain.

([♮]) DEAGLO Inc., Jersey City, New York, USA.

1 Introduction

Emerging markets offer high returns to investors, but come with challenges such as political instability, economic volatility, and currency fluctuations. In particular, currency risk affects significantly investors in their local currency. Fund managers in developed markets use effective hedging strategies, but those in emerging markets face constraints such as regulatory barriers and the interest rate differential.

To mitigate risk and optimize returns, fund managers are increasingly turning to advanced technologies such as machine learning, constrained optimization, and Monte Carlo simulations. These tools help to evaluate hedging strategies and improve portfolio efficiency. However, a proper modelization of the dynamics of the currency exchange rate and interest rates of each country is essential for predicting their future behavior and, therefore, being able to anticipate any investment risk caused by any change in them.

The foreign exchange (FX) market is one of the most liquid financial markets in the world. There are different currency derivatives, such as forwards, swaps and options. The study of this type of derivatives dates back to 1983, when Biger and Hull in [3] and Garman and Kohlhagen in [5] obtained an analytical solution when the exchange rate follows a geometric Brownian and interest rates (domestic and foreign) are considered constant.

Shastri and Wethyavivorn in [11] considered a jump-diffusion model, and Melino and Turnbull in [8] a stochastic volatility model for pricing FX options. Bates in [2] described a jump stochastic volatility model. However, the assumption of constant interest rates is very restrictive, and Abudy and Izhakian in [1] considered stochastic interest rates to improve the model. Ma, Yue, Wu and Ma in [7] and Wu, Jia, Yang and Liu in [12] considered a stochastic model with no correlation between interest rates and exchange rate variables and provided closed-form solutions for pricing options and swaps, respectively. Cao, Roslan and Zhang in [4] introduced the correlation between stochastic variables. Recently, He and Lin in [6] introduced two new factors into the Heston-Hull-white model when evaluating FX options and provided a closed-form solution.

In this paper, we propose a model that aims to accurately represent the dynamics of the exchange rate. To this end, stochastic processes are used not only for the exchange rate variable, but also for the two interest rates: domestic and foreign. In particular, we propose a stochastic process that incorporates memory into the volatility for modeling the exchange rate. In other

[1] correspondingauthor@mail.com

words, a delay is introduced in the process. This makes possible to take into account not only what is happening at a given moment but also what happened in a previous period for an accurate pricing. We value the FX forwards with this new model and compare them with prices obtained from stochastic model without delay.

2 Stochastic model and valuation

In this section, we describe the processes that the state variables of this FX model follow.

Suppose that $(\Omega, \mathcal{F}, \mathcal{P})$ is a given probability space with filtration $\{\mathcal{F}_t\}_{0\leq t\leq T}$ satisfying the usual conditions; see [10]. We assume that the change rate follows, under the risk-neutral measure $\hat{\mathcal{P}}$, a stochastic delay differential equations (SDDE):

$$dQ(t) = (r_d(t) - r_f(t))\, Q(t)\, dt + \sigma_Q(r_d(t), r_f(t), Q(t), Q(t-d))\, Q(t)\, dW_Q, \quad t \in (0, T], \tag{1}$$

$$Q(t) = \phi(t), \quad t \in [-d, 0], \tag{2}$$

where d is a positive constant, i.e. a fixed delay, and μ, σ are deterministic continuous functions and, are, respectively, the drift and volatility of the process, satisfying suitable regularity conditions; see [10]; W_Q is a Wienner process and, ϕ is a deterministic continuous function, $\phi \in \mathcal{C}([-d, 0])$, which represents past change rate information.

We consider that both the domestic, r_d, and foreign, r_f, interest rates follow under the domestic risk-neutral measure $\hat{\mathcal{P}}$, the stochastic differential equations (SDE) for $t \in (0, T]$:

$$dr_d(t) = \mu_d(r_d(t))\, dt + \sigma_d(r_d(t))\, dW_d, \tag{3}$$

$$dr_f(t) = \big(\mu_d(r_f(t)) - \sigma_f(r_f(t))\, \rho_{f,Q}\, \sigma_Q(r_d(t), r_f(t), Q(t), Q(t-d))\big)\, dt + \sigma_f(r_f(t))\, dW_f, \tag{4}$$

where, μ_d, σ_d and μ_f, σ_f are deterministic continuous functions, satisfying suitable regularity conditions; see [10]. Moreover, W_d and W_f are Wienner processes and $\rho_{f,Q}$ is the correlation between r_f and Q.

The FX forward is calculated by

$$f(t, Q, r_d, r_f; T) = \frac{B_f(t, r_f; T)}{B_d(t, r_d; T)}\, Q(t), \tag{5}$$

where B_f and B_d are the zero-coupon bonds (discount functions) on the foreign and domestic market, respectively; see [9]. To price the FX forward, we simulate the trajectories of $Q(t)$, with Euler discretizations of equations (1) and (3)-(4) and Monte Carlo method, and then we use the expression (5) to obtain the forward prices.

3 Empirical application

In the empirical application, we implement the proposed model in the previous section for pricing FX forwards in the Euro/Dollar market with data obtained from Bloomberg. More precisely, we use the forward contract prices, the change rate, and the interest rates of the USA and Europe.

We compare the prices obtained with the delay in the change rate process, with those obtained by the classical models without delay, and analyze the results.

References

[1] Abudy, M., Izhakian, Y., Pricing stock options with stochastic interest rate *International Journal of Portfolio Analysis and Management*, 1(3), 250-277, 2013.

[2] Bate, D.S., Jump and stochastic volatility: Exchange rate processes implicit in deutsche mark options *The Review of Financial Studies*, 9(1): 69-107, 1996.

[3] Biger, N.,Hull, J., The valuation of currency options *Financial Management*, 24-28, 1983.

[4] Cao, J., Roslan, T.R., Zhang, W., The valuation of variance swaps under stochastic volatility, stochastic interest rate and full correlation structure *Journal of the Korean Mathematical Society*, 57(5): 1167-1186, 2020.

[5] Garman, M.B., Kohlhagen, S.W., Foreign currency option values *Journal of International Money and Finance*, 2(3): 231-237, 1983.

[6] He, X.J., Lin, S., Analytical pricing foreign exchange options under a three-factor stochastic volatility and interest rate model: A full correlation structure *Expert Systems with Applications*, 246, 2024.

[7] Ma, C., Yue, S., Wu, H., Ma, Y., Pricing vulnerable options with stochastic volatility and stochastic interest rate *Computational Economics*, 56: 391-429, 2020.

[8] Melino, A., Turnbull, S.M., Pricing foreign currency options with stochastic volatility *Journal of Econometrics*, 45(I-2): 239-265, 1990.

[9] Musiela, M., Rutkowski, M., Martingale Methods in Financial Modelling, New York, Springer-Verlag, 1997.

[10] Oksendal, B., Sulem, A., Stochastic Differential Equations: An Introduction with Applications, Heidelberg, Springer-Verlag, 2000.

[11] Shastri, R., Wethyavivorn, K., The valuation of currency options for alternate stochastic processes *Journal of Financial Research*, 10(4): 283-293, 1987.

[12] Wu, H., Jia, Z., Yang, S., Liu, C., Pricing variance swaps under double Heston stochastic volatility model with stochastic interest rate *Probability in the Engineering and Informational Sciences*, 36(29), 564-580, 2021.

Stochastic Log-Normal Model for price prediction with two cryptocurrencies as underlying assets

C. Burgos-Simón[1] , J.F. Sánchez-Jiménez[1,*] and A. Sánchez-Sánchez[1]

(1) Universitat Politècnica de València,
Camí de Vera, s/n, 46022 València.

1 Introduction

The volatility and rapid evolution of cryptocurrency markets present unique challenges and opportunities for financial modeling. Digital assets such as Bitcoin (BTC) and Ethereum (ETH) exhibit high-frequency fluctuations and nonlinear dependencies, making traditional linear models insufficient for reliable forecasting (Liu & Tsyvinski, 2021). Unlike traditional assets, cryptocurrencies are prone to rapid price surges and crashes, frequently exhibiting speculative dynamics that challenge classical modeling techniques. Several studies have documented the formation and collapse of price bubbles in major cryptocurrencies, including precise identification of their timing and duration (Corbet el at., 2018). These findings underscore the need for predictive models that are sensitive to extreme behaviors and capable of providing actionable forecasts under high uncertainty. In this study, we extend a stochastic log-normal model with two correlated underlying assets, originally applied to traditional stocks (Gascó Ferrer, 2020), to the context of cryptocurrencies. This adaptation allows us to account for co-movements and correlation structures between major digital assets.
By estimating key parameters such as mean return, volatility, and correlation from historical log-returns, the model delivers point forecasts with associated confidence intervals. These outputs support dynamic portfolio strategies in environments where speculative and nonlinear behaviors dominate.

2 Methods

The model is based on two coupled differential equations. Each asset price $S_i(t)$, where i=1,2 follows a stochastic differential equation (SDE) of the form:

$$dS_1(t) = \mu_1 S_1(t)dt + \sigma_1 S_1(t)dW_i(t) + \rho\sigma_1 S_1(t)dW_2(t)$$
$$dS_2(t) = \mu_2 S_2(t)dt + \sigma_2 S_2(t)dW_1(t)$$

where:

* jfsanjim@doctor.upv.es

- μ_i is the mean of the expected return of cryptocurrency i,
- σ_i is the volatility,
- W_i (t) is an independent Wiener stochastic process

The stochastic processes $W_1(t)$ and $W_2(t)$ are correlated such that:

$$Cov(dW_1, dW_2) = \rho dt$$

with $\rho \in [-1,1]$ representing the instantaneous correlation between BTC and ETH.

Model parameters (μ_1, μ_2, σ_1, σ_2, ρ) are estimated via historical data as follows:

- The mean parameters μ_1 and μ_2 are estimated as the average of the log-returns for each asset.
- The volatility parameters σ_1, σ_2 are calculated as the standard deviations of the log-returns.
- The instantaneous correlation ρ between the two assets is computed using the Pearson correlation coefficient between their log-returns

We compute point forecasts and 95% confidence intervals and assess accuracy with MAPE and RMSE.

3 Results

The model has been implemented using closing prices of BTC and ETH over a 30-day window. The parameterized system successfully captured the individual volatilities and the correlation structure between the two cryptocurrencies. The model's short-term forecasts aligned closely with actual market behavior, particularly in periods of high stability. Accuracy metrics (MAPE and RMSE) were within acceptable thresholds, and confidence intervals covered the observed values in most instances.

Table 1 shows the input data used for the model with the prices and log-returns of BTC and ETH over a 30-day windows

Table 1. 30-day window prices and log-returns for bitcoin and Ethereum

Date	BTC price	ETH price	BTC log price	BTC log return	ETH log price	ETH log return
05-Jan-2025	98.347,70	3.635,11	11,49626444	0,038898922	8,198394651	0,014377144
06-Jan-2025	102.248,70	3.687,75	11,53516336	-0,053421076	8,212771795	-0,08697514
07-Jan-2025	96.929,80	3.380,56	11,48174228	-0,019409307	8,125796655	-0,0159283
08-Jan-2025	95.066,60	3.327,14	11,46233298	-0,026938876	8,109868355	-0,033306416

09-Jan-2025	92.539,80	3.218,15	11,4353941	0,023329633	8,076561939	0,015089959
10-Jan-2025	94.724,10	3.267,08	11,45872373	-0,001345864	8,091651899	0,005034626
11-Jan-2025	94.596,70	3.283,57	11,45737787	-0,000667265	8,096686524	-0,004967289
12-Jan-2025	94.533,60	3.267,30	11,45671061	7,40475E-06	8,091719235	-0,040464347
13-Jan-2025	94.534,30	3.137,73	11,45671801	0,021016451	8,051254888	0,027141826
14-Jan-2025	96.542,10	3.224,06	11,47773446	0,040176555	8,078396714	0,067961755
15-Jan-2025	100.499,80	3.450,79	11,51791102	-0,004940552	8,146358469	-0,042561756
16-Jan-2025	100.004,50	3.307,00	11,51297046	0,039991727	8,103796713	0,048931269
17-Jan-2025	104.084,90	3.472,84	11,55296219	0,004333207	8,152727982	-0,048405251
18-Jan-2025	104.536,90	3.308,74	11,5572954	-0,031261229	8,104322731	-0,028199028
19-Jan-2025	101.319,50	3.216,74	11,52603417	0,008630282	8,076123703	0,020291738
20-Jan-2025	102.197,70	3.282,68	11,53466445	0,038011842	8,096415441	0,013801501
21-Jan-2025	106.157,20	3.328,30	11,57267629	-0,023565487	8,110216942	-0,026086186
22-Jan-2025	103.684,80	3.242,60	11,54911081	0,002189822	8,084130756	0,029107187
23-Jan-2025	103.912,10	3.338,37	11,55130063	0,009102711	8,113237943	-0,008217304
24-Jan-2025	104.862,30	3.311,05	11,56040334	-0,001139285	8,105020639	0,002139025
25-Jan-2025	104.742,90	3.318,14	11,55926406	-0,02030249	8,107159664	-0,025628967
26-Jan-2025	102.637,80	3.234,18	11,53896156	-0,005334857	8,081530697	-0,017060902
27-Jan-2025	102.091,70	3.179,47	11,53362671	-0,007298463	8,064469795	-0,032489713
28-Jan-2025	101.349,30	3.077,83	11,52632825	0,023141277	8,031980082	0,011445605
29-Jan-2025	103.722,00	3.113,26	11,54946952	0,009641766	8,043425688	0,042128837
30-Jan-2025	104.726,90	3.247,22	11,55911129	-0,022232997	8,085554525	0,015726174
31-Jan-2025	102.424,20	3.298,69	11,53687829	-0,017493732	8,101280699	-0,055971303
01-Feb-2025	100.648,00	3.119,13	11,51938456	-0,029882289	8,045309396	-0,083329199
02-Feb-2025	97.684,90	2.869,75	11,48950227	0,014827355	7,961980197	-0,048678662
03-Feb-2025	99.144,10	2.733,40	11,50432963		7,913301535	

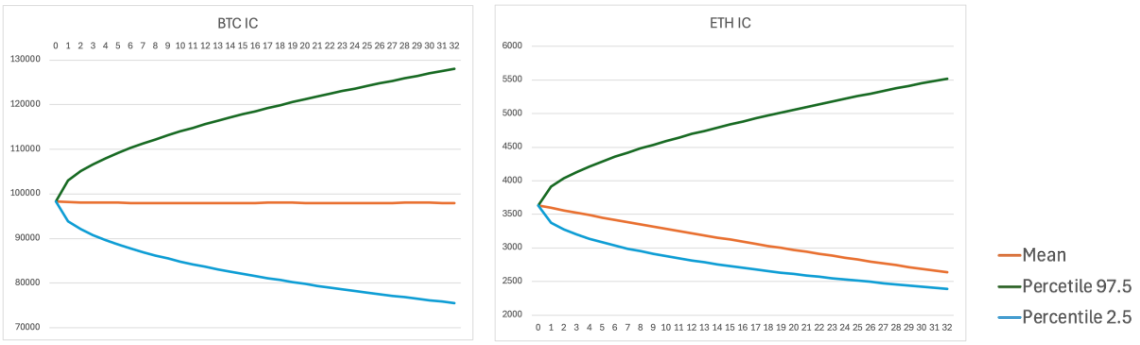


Figure 1: Theoretical confidence intervals for the prices of BTC and ETH after applying the stochastic log-normal model with two underlying assets.

4 Conclusions

Adapting a log-normal model with two underlying assets to cryptocurrencies provides a robust and flexible tool for modeling price dynamics. By incorporating the correlation between assets, the model offers a more realistic representation of market behavior than traditional univariate approaches.

From a practical standpoint, this modeling approach supports applications in dynamic portfolio optimization, where understanding the co-movement between assets can influence asset allocation decisions. It also holds relevance for automated trading systems that must operate in environments characterized by non-linear dependencies and sudden market shifts.

Furthermore, this framework opens the door to future extensions, such as generalizing the model to handle baskets of more than two correlated cryptocurrencies.

References

[1] Corbet, S., Lucey, B., & Yarovaya, L. Datestamping the Bitcoin and Ethereum bubbles. *Finance Research Letters*, 26: 81--88, 2018.

[2] Gascó Ferrer, R. Modelo Estocástico Log-Normal con dos subyacentes correlacionados y su aplicación a la construcción de carteras dinámicas: Aplicación a la modelización de los subyacentes cotizados Ferrovial y Repsol (Master's thesis, Universitat Politècnica de València). Universitat Politècnica de València Repository, 2020.

[3] Liu, Y., & Tsyvinski, A. Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6): 2689--2727, 2021.

A coupled stochastic model for cryptocurrency prediction: Mathematical modelling and implementation

C. Burgos[♭],[1] J.F. Sánchez-Jiménez[♭]
and A. Sánchez-Sánchez[♭]

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

The cryptocurrency market has evolved from a small technical experiment to an important part of global finance, creating challenges for traditional financial models. Conventional methods often fail to capture the extreme price volatility, unusual patterns, and complex structures found in cryptocurrencies [1]. Cryptocurrencies display unique statistical traits such as heavy-tailed returns, volatility clustering, and leverage effects that violate classical financial assumptions [2], with correlations between coins shifting rapidly under market changes.

Bitcoin, launched in 2009 by Satoshi Nakamoto [3], is the first decentralized cryptocurrency and a dominant market reference due to its large capitalization [4]. Its fixed supply and issuance schedule create scarcity, while its returns show extreme volatility and complex time-series behavior [5, 6], making it difficult to model traditionally.

Ethereum, launched in 2015 by Vitalik Buterin [7], differs by supporting smart contracts and decentralized apps, with Ether serving multiple ecosystem functions. Its shift to proof-of-stake (Ethereum 2.0) changed supply dynamics and introduced staking rewards [8]. Ethereum shares some statistical features with Bitcoin but is sensitive to network-specific factors like gas fees and developer activity [2, 9].

Bitcoin and Ethereum often move together during major market trends, but their correlation varies due to technical, regulatory, and ecosystem-specific factors [4]. Traditional univariate models fail to capture such dynamic interdependencies, risking underestimation of joint risks and missed diversification opportunities [10]. Coupled stochastic differential equation models better represent these relationships by accounting for volatility, asymmetric influences, and both deterministic and stochastic elements.

2 Methodology

Our approach consists of four key steps: data acquisition, stochastic modeling, simulation, and validation.

We collect daily BTC and ETH price data from January 1st to 30th, 2025. These series serve for both model calibration and validation.

[1] alsncsnc@mat.upv.es

We consider a coupled system of stochastic differential equations (SDEs) to model the joint dynamics of BTC and ETH:

$$dx(t) = \mu_1 x(t)\,dt + \sigma_1 x(t)\,dw_1(t) + \rho\sigma_1 x(t)\,dw_2(t), \tag{1}$$
$$dy(t) = \mu_2 y(t)\,dt + \sigma_2 y(t)\,dw_2(t), \tag{2}$$

where $x(t)$, $y(t)$ are BTC and ETH prices, μ_i and σ_i their respective drift and volatility terms, and $w_1(t)$, $w_2(t)$ are independent Wiener processes. The parameter ρ captures asymmetric coupling: BTC responds to both its own and ETH's shocks, while ETH evolves independently.

Discretizing the model to daily steps, we simulate price paths and evaluate performance using a cumulative absolute relative error metric, which accounts for scale differences. We implement a batch system to simulate thousands of parameter sets defined via JSON files, ensuring reproducibility through fixed seeds and structured outputs.

For parameter estimation, we apply a custom Particle Swarm Optimization (PSO) algorithm [11]. Each particle represents a parameter set, updated through cognitive, social, and random influences. A mutation operator promotes exploration, while reflective bounds and Pareto dominance preserve diversity and constraint compliance [6].

Validation combines multiple criteria: in-sample and out-of-sample error, confidence interval width, and coverage. Rather than selecting a single best solution, we identify a representative Pareto-optimal model closest to the ideal.

We compute 95% confidence intervals from repeated simulations and evaluate them in terms of coverage and relative width. RMSE complements the analysis for point forecast accuracy.

3 Results

The optimization process yielded multiple high-quality parameter sets within the Pareto front. Table 1 presents the parameter values for the knee point solutions, showing their mean values along with 2.5th and 97.5th percentiles to indicate estimation uncertainty.

Table 1: Descriptive analysis model parameters for the coupled stochastic model

Parameter	Symbol	2.5th percentile	Mean	97.5th percentile
Bitcoin growth rate	μ_1	-0.005757	0.000835	0.005378
Ethereum growth rate	μ_2	-0.016068	-0.006698	-0.000541
Bitcoin volatility	σ_1	0.000083	0.014178	0.029608
Ethereum volatility	σ_2	0.007731	0.019685	0.035205
Coupling coefficient	ρ	-0.328414	0.163894	0.681456

These parameter values reveal several important characteristics of cryptocurrency market dynamics. The estimated growth rates suggest that while Bitcoin shows a small positive average daily return, Ethereum's mean growth rate is negative over the study period, though both have wide confidence intervals that include positive values. Ethereum exhibits higher volatility than Bitcoin, consistent with its status as a smaller, more dynamic asset. The coupling coefficient shows substantial variation but has a positive mean, indicating co-movement between the two cryptocurrencies.

The calibrated model demonstrates strong fit to historical price data and robust predictive performance. Figure 1 illustrates a sample path simulation alongside actual price data, showing how the model captures both the trends and volatility characteristics of Bitcoin and Ethereum while providing uncertainty quantification through confidence intervals.

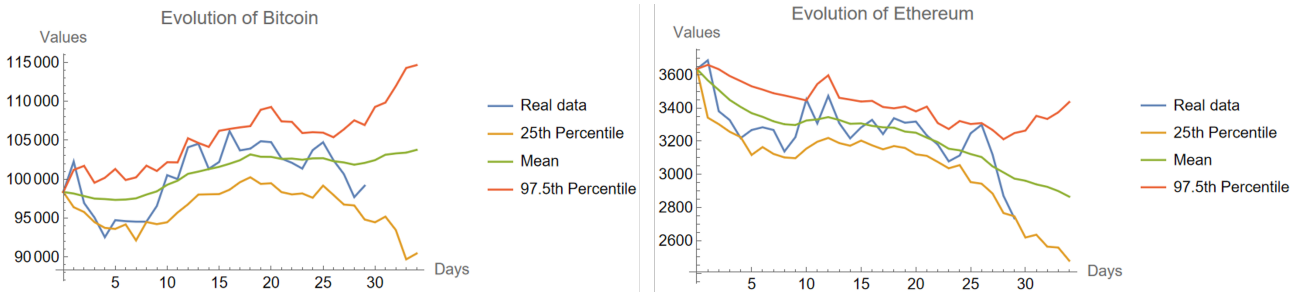


Figure 1: Comparison of model simulation (with 95% confidence intervals) and actual cryptocurrency prices.

4 Conclusions

The coupled stochastic differential equation system framework successfully captures the dynamics of the evolution of values of Bitcoin and Ethereum through the described technique. The computational implementation demonstrates how adapted metaheuristic optimization can effectively calibrate stochastic systems. Empirical results validate that this approach provides more accurate quantification than traditional methods, particularly in capturing joint tail behaviors during market crises. From a practical perspective, the model's ability to simultaneously forecast both assets with properly calibrated uncertainty intervals offers significant value for applications ranging from portfolio optimization to derivatives pricing in cryptocurrency markets.

Acknowledgments

The Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER UE) (Grant PID2020-115270GB-I00). This work has also been funded by the research project CIGE/2023/79 (Generalitat Valenciana).

References

[1] Dyhrberg, A. H., Bitcoin, gold and the dollar—A GARCH volatility analysis. *Finance Research Letters*, *16*: 85–92, 2016. https://doi.org/10.1016/j.frl.2015.10.008

[2] Bariviera, A. F., Basgall, M. J., Hasperué, W., & Naiouf, M., Some stylized facts of the Bitcoin market *Physica A*, 484: 82–90, 2017.

[3] Nakamoto, S., Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin whitepaper. 2008. `https://bitcoin.org/bitcoin.pdf`

[4] Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L., Exploring the dynamic relationships between cryptocurrencies and other financial assets *International Review of Financial Analysis*, 62: 182–199, 2018.

[5] Katsiampa, P., Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, *158*: 3–6. 2017. https://doi.org/10.1016/j.econlet.2017.06.023

[6] Caporale, G. M., & Plastun, A., Price overreactions in the cryptocurrency market *Journal of Economics and Business*, *100*: 81–93, 2018.

[7] Buterin, V. A., Next-Generation Smart Contract and Decentralized Application Platform. Ethereum whitepaper. 2014. `https://ethereum.org/en/whitepaper/`

[8] Schär, F., Decentralized Finance: On Blockchain- and Smart Contract-Based Financial Markets. *Federal Reserve Bank of St. Louis Review*, 103(2): 153-–174. 2021. https://doi.org/10.20955/r.103.153-74

[9] Cong, L. W., Li, Y., & Wang, N., Tokenomics: Dynamic adoption and valuation *The Review of Financial Studies*, *34*(3): 1105–1155. 2021. https://doi.org/10.1093/rfs/hhaa089

[10] Liu, Y., & Tsyvinski, A., Risks and returns of cryptocurrency *The Review of Financial Studies*, 34(6): 2689–2727, 2019.

[11] Ji, Q., Bouri, E., Lau, C. K. M., & Roubaud, D., Dynamic connectedness and integration in cryptocurrency markets *The Quarterly Review of Economics and Finance*, 70: 203–213, 2019.

Part XI

Mathematical models in population dynamics

Efficient natural grid numerical methods for cell population balance models with asymmetric division

L. M. Abia [♭], O. Angulo [♭],[1] J. C. López-Marcos [♭], and M. A. López-Marcos [♭]

([♭]) Universidad de Valladolid e IMUVa,
Dpto. de Matemática Aplicada. Facultad de Ciencias. Univ. de Valladolid. Pso. Belén 7, 47011 Valladolid, SPAIN,
lmabia@uva.es, lopezmar@uva.es, malm@uva.es.

1 Introduction

Cell population balance models (cell PBM) were introduced in the early sixties and quickly experimented both theoretically and numerically a great progress. A general survey of the main mathematical problems and techniques employed in the theoretical analysis of general population balance models can be found in [5, 7, 8]. In this work we present an efficient second-order numerical method for the approximation of the solution to a population model describing the evolution of a size-structured cell population with asymmetric division. Second-order methods maintain a good compromise between the required smoothness of the vital functions based on realistic biological data and the efficiency of the numerical schemes. Efficiency is obtained through a numerical discretization along a dynamic selection of the characteristic curves that evolve from the natural grid associated to the problem [2, 3]. The scheme is analyzed and optimal rate of convergence to the solution is derived. Numerical experiments will be also reported to demonstrate the predicted accuracy of the scheme and to show its efficiency.

2 The cell PBM model with asymmetric cell division

We study the numerical integration of a model describing the evolution of a size-structured cell population. It is composed by a partial differential equation, which represents a conservation law and includes a nonlocal term, a boundary condition and an initial size distribution. We describe it as presented in [4],

$$u_t(x,t) + (g(x)\,u(x,t))_x = -\mu(x)\,u(x,t) - b(x)\,u(x,t) + 2\int_x^1 b(s)\,P(x,s)u(s,t)\,ds, \qquad x_{min} < x < 1,\ t > 0, \tag{1}$$

$$u\,(x_{min},t) = 0, \quad t > 0, \tag{2}$$

$$u(x,0) = \varphi(x), \quad x_{min} \le x \le 1. \tag{3}$$

The independent variables x and t represent size and time respectively. The existence of a non-negative minimum cell-size is indicated by x_{min} and the finite maximal size is normalized to 1, so

[1] oscar.angulo@uva.es

$0 \leq x_{min} < 1$. Maximal size value represents the point at which every cell must divide or die. The dependent variable $u(x,t)$ is the size-specific density of cells with size x at time t. The size of any cell varies according to the following ordinary differential equation

$$\frac{dx}{dt} = g(x),$$

where $g(x)$ is the growth rate, with $g > 0$ on $[x_{min}, 1)$. This hypothesis means that cells are growing during their whole life-span. The nonnegative functions μ and b represent the mortality and division rate respectively. These three functions are usually called the vital functions and define the life history of the individuals. Note that all of them only depend on size x (the internal structuring variable), so that we also assume an unlimited environment and that all possible nonlinear mechanisms are ignored.

The dispersion of sizes at division amongst the two daughter cells (unequal division) is defined in terms of the partitioning function $P(x,y)$, a probability density function which gives the distribution of the size of a daughter-cell x when the size of the mother is equal to y. It satisfies the following conditions:

$$\int_{x_{min}}^{1} P(x,y)\,dx = 1, \quad P(x,y) = P(y-x,y), \quad P(x,y) = 0,\ x \geq y.$$

In a particular (and extreme) case, mother cells could always divide in two identical daughter (equal fission). Thus, the partitioning function reduces to the Dirac delta function $P(x,y) = \delta(x - y/2)$, and the model turns into Diekmann *et al.* proposal [5].

Finally, the existence of a maximal size is an accepted biological wisdom: cells must divide or die with probability one before reaching this value. Therefore $\Pi(1) = 0$, where

$$\Pi(x) = \exp\left(-\int_{x_{min}}^{x} \frac{\mu(s) + b(s)}{g(s)}\,ds\right), \quad x_{min} \leq x \leq 1,$$

$\Pi(x)$ represents the probability that an individual of size x_{min} reaches size x. If we assume that μ and b are positive and bounded functions, cells will not disappear at the maximal size unless we consider a growth function that satisfies $\lim\limits_{x\to 1^-} \int_{x_{min}}^{x} \frac{ds}{g(s)} = +\infty$.

3 The numerical scheme

In this section, we introduce a new efficient numerical scheme to obtain the solution to (1)-(3). The numerical integration is based on the discretization of the solution along the characteristic curves. Therefore, we define $\mu^*(x) = g'(x) + \mu(x) + b(x)$ and rewrite (1) as

$$u_t(x,t) + g(x)\,u_x(x,t) \quad = \quad -\mu^*(x)\,u(x,t) + 2\int_x^1 b(s)\,P(x,s)u(s,t)\,ds, \tag{4}$$

$x_{min} < x < 1$, $t > 0$. We denote by $x(t;t^*,x_*)$ the characteristic curve of the equation (4) (and (1)) that takes the value x_* at time t^*, and define $w(t;t^*,x_*) = u(x(t;t^*,x_*),t)$, $t \geq t_*$. Thus

$$\begin{cases} \dfrac{d}{dt}x(t;t^*,x_*) = g(x(t;t^*,x_*)), & t > t^*, \\ x(t^*;t^*,x_*) = x_*, \end{cases} \tag{5}$$

and the solution to (4) along a characteristic curve is given by

$$u(x(t;t_*,x_*),t) = u(x_*,t_*)\,\exp\left\{-\int_{t_*}^{t}\mu^*\left(x(\tau;t_*,x_*)\right)\,d\tau\right\}$$
$$+\,2\int_{t_*}^{t}\left[\exp\left\{-\int_{\tau}^{t}\mu^*\left(x(s;t_*,x_*)\right)ds\right\}\int_{x(\tau;t_*,x_*)}^{1}b(\sigma)P(x(\tau;t_*,x_*),\sigma)u(\sigma,\tau)d\sigma\right]d\tau,$$
$$t\geq t_*. \tag{6}$$

Hereafter, we provide discretization procedures to solve numerically (5)-(6) and, therefore (1)-(3) along the time interval $[0,T]$.

Thus, given a positive integer N, we define $k=\frac{T}{N}$ and introduce the discrete time levels $t^n=n\,k$, $0\leq n\leq N$. The integration of (5) provides the grid of the method on the space (size) variable. Thus, if x is the solution to (5), with $t^*=0$ and $x_*=x_{\min}$ we define $x_j=x(t^j;0,x_{\min})$, $t^j=j\,k$, $0\leq j\leq J$. This is called *the natural grid*. This is nonuniform and invariant with time due to the explicity independence on the time variable of the growth function $g(x)$. Its good properties are well described in [6, 1]. The numerical integration of (5) is obtained with the modified Euler method

$$\begin{aligned} &X_0=x_{min},\\ &X_{j+1}=X_j+\frac{k}{2}\left(g(X_j)+g(X_j+k\,g(X_j))\right),\quad 0\leq j\leq J-1. \end{aligned} \tag{7}$$

Integer J represents the index of the last grid point computed at the size interval. It is chosen to satisfy the condition

$$K_0\,k<1-X_J\leq K_1\,k, \tag{8}$$

with K_0 and K_1 suitable constants (we refer to [1] for further details). We fix the last grid point as $X_{J+1}=1$

However, opposed to the benefits of *the natural grid*, among we include its ease use for an analysis of the solution steady states, we are confronted with two great weaknesses. First, its lack of efficiency on the one hand because the use of an unnecessary large number of grid points in the natural grid and, second, the non quasiuniformity of the grid points in the natural grid, which is needed when using high order cuadrature rules [1]. Therefore, we consider a suitable selection of the grid nodes of the numerical natural grid when approximating the nonlocal terms, as was introduced in [4].

Thus, we define the set $I=\{0,1,\ldots,J,J+1\}$ as the indexes of the natural grid nodes (7) and we determine an initial subgrid $I^0=\{j_m^0\}_{m=0}^{M+1}$, $I^0\subset I$, with $j_0^0=0$, $j_{M+1}^0=J+1$ and,

$$C_3\,k\leq X_{j_{l+1}^0}-X_{j_l^0}\leq C_4\,k,\quad 0\leq l\leq M,$$

where C_3 and C_4 are positive constants independent on k (for more details we refer to [1]).

Next, we only consider advancing the numerical solution along the characteristic curves corresponding to nodes involved in the initial subgrid I^0 to minimize the impact of the full *natural grid* in the lack of efficiency. The integration along these characteristic curves allows us to determine the subset of nodes employed in the quadrature rule for approximating the nonlocal terms at each one of the following time discretization steps. Moreover, because at each time step a new node that comes from the boundary is added to the subgrid and the number of nodes (or characteristics) in the subgrid increases in one, we intend to keep constant the number of grid points at every time step by selecting a grid node to be eliminated from the subgrid at the next and following time steps.

Thus, for each time step t^n, let $I^n = \{j_m^n\}_{m=0}^{M+1} \subset I$ denotes the set of indexes of the nodes belonging to the full *natural grid* that we employ in the numerical integration of the nonlocal terms at that time level. And, for $n > 1$, if we have determined I^p, $0 \le p \le n-1$, the selection procedure chooses the index of the node to be removed. First, if x_{j_M} matches with the last node in the full *natural grid*, x_J, then it is not possible to follow the corresponding numerical characteristic curve, and we choose to remove it. In other case, we select a specific node as follow: let l be the first index such that

$$|X_{j_{l+1}^{n-1}+1} - X_{j_{l-1}^{n-1}+1}| = \min_{0\le m\le M} |X_{j_{m+1}^{n-1}+1} - X_{j_{m-1}^{n-1}+1}|.$$

where $X_{j_{-1}^{n-1}+1} = X_0$ and $X_{j_{M+1}^{n-1}+1} = X_{J+1}$. (For instance, in the case $l = 0$, the node $X_{j_0^{n-1}+1} = X_1$ does not take part in the numerical integration at the next time level.) Then, the new subset of indexes I^n is defined as

$$\begin{cases} j_0^n = 0, \\ j_m^n = j_{m-1}^{n-1} + 1, & 1 \le m \le l, \\ j_m^n = j_m^{n-1} + 1, & l+1 \le m \le M, \\ j_{M+1}^n = J+1, \end{cases} \tag{9}$$

which completely determine our selection procedure.

We define $u_j^n = u(x_j, t^n)$, $0 \le j \le J+1$, $0 \le n \le N$ and denote with U_j^n a numerical approximation to u_j^n. The solution at a new time level is only computed at the nodes defined by I^n, $1 \le n \le N$. It is described in terms of the previous one, as follows

$$\begin{aligned} U_{j_l^n}^n = \exp\left\{-\frac{k}{2}\left(\mu^*\left(X_{j_l^n-1}\right) + \mu^*\left(X_{j_l^n}\right)\right)\right\}\left(U_{j_l^n-1}^{n-1} + k\,\mathcal{Q}_{k,X}^{j_l^n-1,n-1}(\mathbf{b}_X \cdot \mathbf{P}_X^{j_l^n-1} \cdot \mathbf{U}^{n-1})\right) \\ + k\,\mathcal{Q}_{k,X}^{j_l^n,n}(\mathbf{b}_X \cdot \mathbf{P}_X^{j_l^n} \cdot \mathbf{U}^n), \quad 1 \le l \le M, \end{aligned} \tag{10}$$

and we set $U_0^n = U_{M+1}^n = 0$. In (10), $\mathcal{Q}_{k,X}^{j_l^n,n}(\mathbf{V})$ represents the quadrature rule employed at time level t_n to approximate the integral over the interval $[X_{j_l^n}, 1]$, $0 \le l \le M+1$ of the function with values $\mathbf{V} = (V_{j_0^n}, \ldots, V_{j_{M+1}^n})$ on the nodes $(X_{j_0^n}, \ldots, X_{j_{M+1}^n})$. In this case $\mathbf{b}_X$, $\mathbf{P}_X^{j_l^n}$ and $\mathbf{U}^n$, represent the vectors with components $(b(X_{j_0^n}), \ldots, b(X_{j_{M+1}^n}))$, $(P(X_{j_l^n}, X_{j_0^n}), \ldots, P(X_{j_l^n}, X_{j_{M+1}^n}))$ and $(U_{j_0^n}^n, \ldots, U_{j_{M+1}^n}^n)$, respectively, and products $\mathbf{b}_X \cdot \mathbf{P}_X^{j_l^n} \cdot \mathbf{U}^n$, $0 \le l \le M+1$, $0 \le n \le N$ must be interpreted component-wise. Here, a second order quadrature formula is appropriate, in this case, we propose the composite trapezoidal quadrature rule on the subgrid

$$\mathcal{Q}_{k,X}^{j_l^n,n}(\mathbf{V}) = \sum_{k=l}^{M} \frac{X_{j_{k+1}^n} - X_{j_k^n}}{2}\left(V_{j_k^n} + V_{j_{k+1}^n}\right). \tag{11}$$

Therefore, if we start from an approximation to the initial data (3) of the problem, for example, the grid restriction of the function φ, we have completely defined the numerical scheme.

The numerical procedure seems to be implicit. However, if we compute the approximations at the new time level t^{n+1} downwards, it results in an explicit procedure. The reason is that the right hand side values in (10) at time t^n are either zero or previously computed.

4 Convergence results

Second order convergence of the method is derived from the second order consistency of the numerical approximations in the integration along the selected characteristic curves and the second-order accuracy of the quadrature rules employed at each time step. If we define

$$\mathbf{u}^n = (u_{j_0^n}^n, u_{j_1^n}^n, \ldots, u_{j_M^n}^n),\ u_{j_l^n}^n = u(x_{j_l^n}, t^n),\ 0 \le j \le J+1,\ 0 \le n \le N,$$

and we denote the error produced by the numerical approximation to the grid nodes and the numerical solution as

$$\boldsymbol{\Delta}^n = (\Delta_{j_1^n}, \ldots, \Delta_{j_M^n}), \quad \Delta_{j_l^n} = x_{j_l^n} - X_{j_l^n}^n, \quad 1 \leq l \leq M, \quad \mathbf{E}^n = (E_{j_1^n}^n, \ldots, E_{j_M^n}^n), \quad E_{j_l^n}^n = u_{j_l^n}^n - U_{j_l^n}^n,$$

$1 \leq l \leq M$, $0 \leq n \leq N$.

Theorem 1 *Let g be three times continuously differentiable, functions $\mu, b \cdot P$ and u be two times continuously differentiable and $\|\boldsymbol{\Delta}^0\|_\infty = O(k^2)$, $\|\mathbf{E}^0\|_\infty = O(k^2)$, as $k \to 0$, then*

$$\|\boldsymbol{\Delta}^n\|_\infty = O(k^2), \quad \|\mathbf{E}^n\|_\infty = O(k^2), \quad 0 \leq n \leq N,$$

as $k \to 0$.

The main technical issue is to guaranteee the quasiuniformity of all the numerical subgrids with indexes in I^n, $0 \leq n \leq N$, employed by the numerical quadratures at each time step. This can be proved after checking that, for k sufficiently small, at every time level t^n, $0 \leq n \leq N$, the number of characteristic curves we have dismiss along the integration ending in the interval between any two consecutive grid points $X_{j_{l+1}^n}$ y $X_{j_l^n}$ is uniformly bounded by a constant K independent of the discretization parameter.

5 Conclusions

The study of cell populations by means of the use of structured population models, and their numerical simulation, is a current and important topic. In this work we consider a new numerical method to attain the solution to a size-structured population model describing the dynamics of a cell population when the reproduction process is by division into two unequal parts. The main contribution is that we get a second-order convergence to the solution even when the numerical method only consider the numerical integration along a subset of the characteristic curves evolving from the natural grid that are chosen dinamically at each time step. We corroborate experimentally the second convergence order of the numerical procedure we analyse. This new numerical scheme is shown as the most efficient one when compared with methods that discretize (6) using all the characteristic curves emaning from the natural grid [2, 3, 4].

References

[1] Angulo, O., López-Marcos, J. C., Numerical schemes for size-structured population equations. Math. Biosci. 157 (1-2): 169–188, 1999.

[2] Angulo, O. López-Marcos, J. C., López-Marcos, M. A., A second-order method for the numerical integration of a size-structured cell population model, Abstr. Appl. Anal., 549168, 1-8, 2015.

[3] Angulo, O., López-Marcos, J. C., López-Marcos, M. A., A numerical study on the estimation of the stable size distribution on a cell population balance model, Math. Meth. Appl. Anal. 41: 2894–2905, 2018.

[4] Angulo, O., López-Marcos, J. C., López-Marcos, M. A., A second-order numerical method for a cell population model with asymmetric division, J. Comp. Appl. Math. 309: 522–531, 2017.

[5] Diekmann, O., Heijmans, H. J. A. M., Thieme, H. R., On the stability of the cell size distribution, J. Math. Biol. 19: 227–248, 1984.

[6] Ito, K., Kappel, F., Peichl, G., A fully discretized approximation scheme for size-structured population models., SIAM J. Numer. Anal. 28(4): 923-954, 1991

[7] Metz, J. A. J., Diekmann, O. (Eds.), The Dynamics of Physiologically Structured Populations, In: Lect. Notes Biomath. 68, New York, Springer-Verlag, 1986.

[8] Perthame, B., Transport Equations in Biology. Basel, Switzerland, Birkhäuser, 2007.

Multi-serotype nested immuno-epidemiological model for dengue hemorrhagic fever involving backward bifurcation and serotype invasion

Mostafa Adimy[♭],[1], Charlotte Dugourd-Camus[♭] and Ruben Taieb[♭,♮]

(♭) Inria Centre de Lyon, Villeurbanne, France.

(♮) Université Le Havre Normandie, Le Havre, France.

1 Introduction

Reinfection with the same dengue serotype is generally benign, as individuals develop protective immunity. On the other hand, in the case of reinfection with a different serotype, pre-existing antibodies can increase the risk of developing Dengue Hemorrhagic Fever (DHF), by inducing Antibody-Dependent Enhancement (ADE). To model this dynamic, we introduce a multi-scale immuno-epidemiological system. The immunological part is described by a system of ODEs representing the interaction between two antibodies (from previous and current infection) and the virus. The epidemiological part is represented by an infection-age structured SIRS system (for both the first and second infections) and a recovery-age structured equation (for the first infection). A detailed mathematical analysis of the equilibrium points of the multi-scale reinfection model, including disease-free, mono-endemic and bi-endemic states, is performed. We establish necessary and sufficient conditions for the existence of backward bifurcations and derive an expression for the invasion reproduction number, which shows that the second serotype can invade the population after a mono-endemic first serotype. We also investigate the dependence of the basic and invasion reproduction numbers on the immunological parameters of the first and second infections. This gives us a better understanding of the relationship between DHF and ADE during secondary infection.

2 Methods and results

In this work, we investigate a multi-scale immuno-epidemiological system for dengue reinfection. The immunological level is described by a system of ODEs representing the interaction between two antibodies, one from a previous infection and the other from the current infection, and the virus. The epidemiological component is represented by an SIRS system structured by time-since-last-infection (both primary and secondary) and time-since-last-recovery from the primary infection. The transition rates between these classes may depend on variables internal to the host, and the initial viral load of an infected individual may be influenced by the viral load of the individual who transmits the disease. We consider that the viral load at the beginning of the infection is independent of that transmitted by the infected individual. Indeed, as we are focusing

[1] mostafa.adimy@inria.fr

on dengue, which is a mosquito-borne disease, this hypothesis is consistent. On the other hand, as dengue reinfections beyond two are infrequent, we mainly consider the first and second infections. Nevertheless, the proposed model can be extended to the case of transmission of the viral load according to a relatively general probability law and also to multiple consecutive or simultaneous infections, and thus it could be extended to other diseases involving multi-serotypes (and multi-strains).

We analyze the existence and stability of the disease-free, mono-endemic and bi-endemic equilibrium of the multi-scale reinfection model. We establish a necessary and sufficient condition for the existence of a backward bifurcation. Then, we obtain an expression for the invasion reproduction number, which determines whether the second serotype can invade a population already endemic to the first serotype. We also study how the basic and invasion reproduction numbers depend on the immunological parameters associated with the first and second infection. In particular, we examine the role of the immune response - such as cross-protection and ADE - in the transmission dynamics of two successive infections. Finally, we complete our theoretical analysis with numerical simulations to highlight the potential of the proposed model. In particular, we use the backward bifurcation and the invasion reproduction number to explain the relationship between ADE phenomenon and the outbreak of the DHF epidemic during secondary infections.

Models similar to the one proposed in this work have been studied by Gandolfi *et al.* [1] and Gulbudak *et al.* [2]. Our model extends the study of [1] by accounting for reinfection and multi-serotypes (or multi-strains). Additionally, we complement the epidemiological analysis in [2] by explicitly identifying and proving a necessary and sufficient condition for backward bifurcation, and obtaining an explicit expression for the invasion reproduction number, which provides a condition for the two serotypes to coexist in the population. All results in this paper are proven within the age-structured immuno-epidemiological framework. Furthermore, unlike [2], we take into account the difference in within-host dynamics between primary and secondary infections, which is an important consideration for dengue, where secondary infections significantly increase the risk of DHF due to ADE phenomenon. The study [2] also assumes that an individual develops DHF when their viral load exceeds a certain threshold, which is a quantification of DHF cases that we did not use. Finally, unlike in [2], we only implicitly model the dynamics of vector populations, considering the large differences in the time scale between the dynamics of the host and that of the vector.

References

[1] Gandolfi, A., Pugliese, A. and Sinisgalli, C., Epidemic dynamics and host immune response: a nested approach, *Journal of Mathematical Biology*, 70: 399–435, 2015 `https://doi.org/10.1007/s00285-014-0769-8`.

[2] Gulbudak and H., Browne, C. J., Infection severity across scales in multi-strain immuno-epidemiological dengue model structured by host antibody level, *Journal of Mathematical Biology*, 80: 1803–1843, 2020, `https://doi.org/10.1007/s00285-020-01480-3`.

Boundary-value problems of functional differential equations with state-dependent delays

A. Andò[♭],[1] and J. Sieber[♮]

([♭]) CDLab – Computational Dynamics Laboratory,
Department of Mathematics, Computer Science and Physics,
University of Udine, 33100 Udine, Italy.

([♮]) Department of Mathematics and Statistics, University of Exeter,
Exeter EX4 4QF, UK.

1 Introduction

Phenomena such as maturation in population dynamics clearly indicate that natural processes do not happen instantaneously. In other words, there are delays between actions and their consequences. Moreover, in many models, these delays are not constant but rather dependent on the present state of the system (e.g., [3, 5]).

For delay differential equations (DDEs) of the form

$$y'(t) = G_{\mathrm{DDE}}(y_t, p) \tag{1}$$

with continuous $G_{\mathrm{DDE}} : C^0([-\tau_{\max}, 0]; \mathbb{R}^n) \times \mathbb{R} \to \mathbb{R}^n$ and some parameter $p \in \mathbb{R}$, we can formulate the problem of computing a periodic solution as a boundary-value problem in a suitable 1-periodic space. In the following we use C^ℓ_π for spaces of ℓ times continuously differentiable 1-periodic functions and $C^{0,1}_\pi$ for the space of Lipschitz continuous 1-periodic functions. The problem reads

$$y'(t) = TG_{\mathrm{DDE}}(y(t + (\cdot)/T), p) \quad \text{for } t \in [0, 1], \tag{2}$$

$$R_{\mathrm{aff}}[y, T, p] = 0, \tag{3}$$

where T is the (unknown) period, $y : \mathbb{R} \to \mathbb{R}^n$ is continuous and 1-periodic and $R_{\mathrm{aff}} : C^0_\pi \times \mathbb{R} \to \mathbb{R}$ defines the *phase condition*, making the system "square" when including the parameter T as an unknown. The BVP can also be formulated for problems including model parameters, in which case the phase condition has a higher dimension, accordingly. For an illustrative example, consider the DDE $y'(t) = -y(t - p - y(t))$, for some parameter $p \in \mathbb{R}$. The corresponding rescaled equation (2) reads

$$y'(t) = -Ty(t - p/T - y(t)/T). \tag{4}$$

[1] alessia.ando@uniud.it

2 Mild differentiability

G_{DDE} is not Lipschitz continuous as a functional on its whole domain C^0. This is attributable to the fact that unknowns appear inside the arguments of y, which is itself unknown. Walther [7] and Hartung et al. [4] observe this lack of continuous differentiability, but also that G_{DDE} satisfies a slightly stronger regularity condition when restricted to C^1, namely, the map

$$C^1 \times C^0 \ni (u, v) \mapsto \delta G_{\mathrm{DDE}}(u)v \in C^0$$

is continuous. This restricted differentiability can be generalized to $\ell > 1$: we say that G is $\ell_{\max}$ times *mildly differentiable* if its restriction to C^ℓ is ℓ times continuously differentiable and the map

$$C^\ell \times C^\ell \ni (u, \delta^u) \mapsto D^\ell G_{\mathrm{DDE}}(u)(\delta^u)^\ell \in \mathbb{R}^n$$

has a continuous extension to $C^\ell \times C^{\ell-1}$ for all $\ell \leq \ell_{\max}$. This mild differentiability implies mild local Lipschitz continuity properties such as

$$\|G_{\mathrm{DDE}}(u) - G_{\mathrm{DDE}}(v)\|_0 \leq \|u - v\|_0$$

for all u, v in a sufficiently small ball in C^1_π around some $u^* \in C^1_\pi$

3 Fixed point problem and discretization

Let $G(y_t, T) := TG_{\mathrm{DDE}}(y(t + (\cdot)/T))$. We define the following fixed point operator [2]:

$$\Phi(x) := \begin{bmatrix} t \mapsto & \alpha + \displaystyle\int_0^t G(v_s, \mu)\,\mathrm{d}s - t\int_0^1 G(v_s, \mu)\,\mathrm{d}s \\ & \alpha + \displaystyle\int_0^1 G(v_s, \mu)\,\mathrm{d}s \\ & \mu + R_{\mathrm{aff}}[v, \mu] \end{bmatrix} \quad \text{for } x = \begin{bmatrix} v(\cdot) \\ \alpha \\ \mu \end{bmatrix},$$

The first component of Φ is obtained by modifying (2) so that the problem can be posed in a space of 1-periodic functions. To enforce period 1, we subtract the average of the integrand and then impose that this average is zero as a separate equation. We also introduce the new variable $\alpha \in \mathbb{R}^n$, which will be equal to $y(0)$ in the solution. Note that the subtraction of the integrand's average and replacement of periodic boundary conditions by a condition stating that the average is zero modified the fixed-point problems considered in [1]. The fixed point problem

$$x = \Phi(x) \tag{5}$$

defined on $C^{0,1}_{\mathrm{e}} := C^{0,1}_\pi \times \mathbb{R}^n \times \mathbb{R}$ is then equivalent to the original problem of finding a periodic solution of DDE (1).

In order to formulate the corresponding discretized fixed-point problem, for $z \in C^0_\pi$ we define the interpolation projection $P_L z$ as the unique piecewise polynomial on mesh $(t_i)_{i=0}^L$ of degree $m-1$ where the piece on interval (t_{i-1}, t_i) equals z on collocation nodes of the form $t_{i,j} = t_{i-1} + (t_i - t_{i-1})t_{\mathrm{c},j}$, with $t_{\mathrm{c},j} \in [0, 1]$. Let $\mathcal{P}_L$ be the extension of P_L to the space of definition of the problem,

that is, $\mathcal{P}_L(v,\alpha,T) := (P_L v, \alpha, T)$. Moreover, observe that we can write $\Phi = \mathcal{L} \circ g$ for

$$g : C_{\mathrm{e}}^{0,1} \to C_{\mathrm{e}}^{0}, \quad \text{with} \qquad g\left(\begin{bmatrix} v \\ \alpha \\ \mu \end{bmatrix}\right) = \begin{bmatrix} t \mapsto G(v_t, \mu) \\ \alpha \\ \mu + R_{\mathrm{aff}}[v,\mu] \end{bmatrix}, \quad \text{and}$$

$$\mathcal{L} : L_{\mathrm{e}}^{\infty} \to C_{\mathrm{e}}^{0,1}, \quad \text{with} \qquad \mathcal{L}\begin{bmatrix} w \\ \alpha \\ \nu \end{bmatrix} = \begin{bmatrix} t \mapsto \alpha + \int_0^t w(s)\,\mathrm{d}s - t\int_0^1 w(s)\,\mathrm{d}s \\ \alpha + \int_0^1 w(s)\,\mathrm{d}s \\ \nu \end{bmatrix}.$$
$$\text{or } C_{\mathrm{e}}^{\ell} \to C_{\mathrm{e}}^{\ell+1}$$

Then, the discretized fixed-point problem approximating (5) reads

$$x = \mathcal{L}\mathcal{P}_L g(x) =: \Phi_L(x). \tag{6}$$

4 Convergence of the discretization

Solutions x_L of (6) are locally unique and converge to the solution x^* of (5) if the following assumptions are satisfied.

- G_{DDE} is mildly differentiable to order $\ell_{\max} \geq 1$.
- (2)-(3) has a solution (y^*, T^*, p^*).
- The linearization of (2)-(3) in (y^*, T^*, p^*) has only the trivial solution $(\delta^y, \delta^T, \delta^p) = 0$.

The result by Sieber [6] implies that the third assumption holds if and only if a finite-dimensional systems of algebraic equations has a regular solution.

To prove convergence, we formulate (6) in terms of $\delta^{x,L} := x_L - x^*$ as

$$\begin{aligned} &[I - D\Phi_L(x^*)]\delta^{x,L} = \\ =\ &\mathcal{L}(\mathcal{P}_L - I)g(x^*) && =: \epsilon_{\mathrm{c}}(L)) \\ &+ \mathcal{L}\mathcal{P}_L[g(x^* + \delta^{x,L}) - g(x^*) - Dg(x^*)\delta^{x,L}] && =: \epsilon_{\mathrm{nl}}(L, \delta^x). \end{aligned} \tag{7}$$

Note that the derivative of $\Phi(x)$ is well defined in $x^* = (v^*, \alpha^*, p^*)$ and equal to $D\Phi_L(x^*) = \mathcal{L}\mathcal{P}_L Dg(x^*)$ because v^* is in $C_\pi^2 \subset C_\pi^1$.

The stability of the discretized problem follows immediately from that of the original infinite-dimensional problem thanks to the following result.

Lemma 1 *There exists a monotone increasing continuous function $\omega^* : [0,\infty) \to [0,\infty)$ with $\omega^*(0) = 0$, such that*

$$\|D\Phi_L(x^*) - D\Phi(x^*)\|_{C_{\mathrm{e}}^{0,1} \leftarrow C_{\mathrm{e}}^{0,1}} = \|\mathcal{L}[I - \mathcal{P}_L]Dg(x^*)\|_{C_{\mathrm{e}}^{0,1} \leftarrow C_{\mathrm{e}}^{0,1}} \leq \omega^*(1/L).$$

To prove that the right-hand side of (7) is a contraction (and thus defines a well-posed problem) we need suitable bounds for $\epsilon_{\mathrm{c}}(L)$ and $\epsilon_{\mathrm{nl}}(L, \delta^{x,L})$.

Lemma 2 *The consistency error $\epsilon_{\mathrm{c}}(L) = \mathcal{L}(\mathcal{P}_L - I)g(x^*)$ satisfies*

$$\|\epsilon_{\mathrm{c}}(L)\|_{0,1} = O(L^{-\min\{\ell_{\max}, m\}}).$$

We now need a Lipschitz constant for ϵ_{nl} with respect to $\delta^{x,L}$ that is sufficiently small.

Lemma 3 *For all $\rho_{\rm nl} > 0$ there exists $r_{\rm nl} > 0$ such that all $\delta^{x,1}, \delta^{x,2} \in \mathcal{B}^{0,1}_{r_{\rm nl}}(0)$ and $L > 0$ satisfy*

$$\|\epsilon_{\rm nl}(L, \delta^{x,1}) - \epsilon_{\rm nl}(L, \delta^{x,2})\|_{0,1} \le \rho_{\rm nl} \|\delta^{x,1} - \delta^{x,2}\|_{0,1}.$$

The lemmata above lead to our final convergence result.

Theorem 1 *There exists a radius $r > 0$ such that the discretized fixed point problem $x = \Phi_L(x)$ with L polynomial pieces of degree m has a unique solution x^L in the ball $B^{0,1}_r(x^*)$ for all sufficiently large L. The error $\delta^{x,L} = x^L - x^*$ satisfies*

$$\|\delta^{x,L}\|_{0,1} = O(L^{-\min\{\ell_{\max}, m\}}).$$

5 Illustration — prototype problem

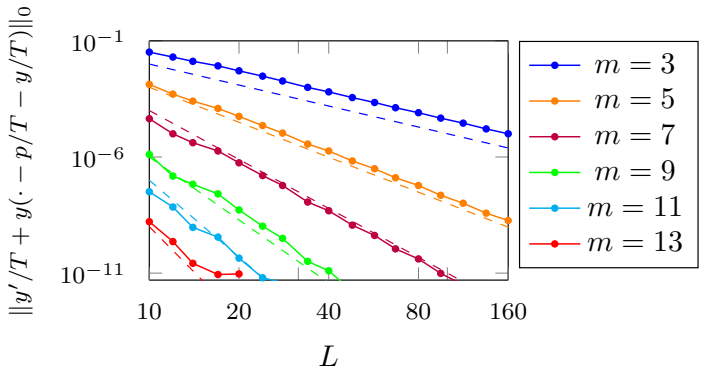


Figure 1: Maximum among the errors computed at 10001 equidistant points in $[0, T]$ for different values of m and L, compared to straight dashed lines having slope m.

We consider (4) with p a varying parameter. A possible choice for the affine conditions is

$$0 = R_{\rm aff}[y, T, p] := \begin{bmatrix} y(0) \\ 2\int_0^1 \sin(2\pi t) y(t)\,{\rm d}t - y_0 \end{bmatrix} \tag{8}$$

for a range of small y_0. We perform our tests on the BVP (4)-(8) for $y_0 = 0.75$. We compute the solution using different values of L and m and approximate the quantity

$$\|y'/T + y(\cdot - p/T - y/T)\|_0$$

by considering the maximum of the residuals on a uniform grid of 10001 points, as shown in Figure 1.

Acknowledgments

AA is a member of the INdAM group GNCS (Gruppo Nazionale di Calcolo Scientifico). Her work is partially supported by the MUR PRIN 2022 project No. 20229P2HEA "Stochastic numerical modelling for sustainable innovation" (CUP G53C24000710006). The research collaboration was supported by the Lorentz Center in Leiden (The Netherlands) via the workshop "Towards rigorous results in state-dependent delay equations", 4–8 March 2024.

References

[1] Andò, A., Breda, D., Convergence analysis of collocation methods for computing periodic solutions of retarded functional differential equations *SIAM Journal on Numerical Analysis*, 58(5):3010–3039, 2020.

[2] Andò, A., Sieber, J., Boundary-value problems of functional differential equations with state-dependent delays, submitted.

[3] Diekmann, O., Gyllenberg, M., Metz, J.A.J., Nakaoka, S., de Roos, A.M., Daphnia revisited: local stability and bifurcation theory for physiologically structured population models explained by way of an example, *Journal of Mathematical Biology*, 61:277-318, 2010.

[4] Hartung, F., Krisztin, T., Walther, H.-O., Wu, J., Functional differential equations with state-dependent delays: theory and applications. In P. Drábek, A. Cañada, and A. Fonda (eds.), *Handbook of Differential Equations: Ordinary Differential Equations*, 3, chapter 5, 435–545. North-Holland, 2006.

[5] Magpantay, F. M. G., Kosovalić, N., Wu, J., An Age-Structured Population Model with State-Dependent Delay: Derivation and Numerical Integration *SIAM Journal on Numerical Analysis*, 52(2):735–756, 2014.

[6] Sieber, J., Finding periodic orbits in state-dependent delay differential equations as roots of algebraic equations *Discrete and Continuous Dynamical Systems A*, 32(8):2607–2651, 2012.

[7] Walther, H.-O., The solution manifold and C^1-smoothness for differential equations with state-dependent delay *Journal of Differential Equations*, 195(1):46–65, 2003.

Numerical integration of a structured population model with two physiological structures

L. M. Abia [♭], O. Angulo [♭],[1] J. C. López-Marcos [♭], and M. A. López-Marcos [♭]

(♭) Universidad de Valladolid e IMUVa,
Dpto. de Matemática Aplicada. Facultad de Ciencias. Univ. de Valladolid. Pso. Belén 7, 47011 Valladolid, SPAIN,
lmabia@uva.es, lopezmar@uva.es, malm@uva.es.

1 Introduction

The study of populations dynamics can be made by using the properties of the individual level [6]. By using this hypothesis, individuals are distinguished by means of different physiological conditions such as age, maturation, size, stage, status, location, etc. These features will be use to determine their birth, growth and death rates, and their each-other or environment interactions. The resulting models are known as structured population ones and the aim of their study is to understand how these conditions affect the dynamics of the population.

With respect to the interest of the characteristic that structures the individuals in the population, the age is one of the most important. For instance, in epidemiology, we can structure individuals by different kind of age: chronological, infection, recovery, etc. The theoretical study of age-structured models is well developed and we refer to classical monographies [1, 3, 7].

However, in the literature, there are many models that combine the effects of two physiological features but there are very few theoretical studies of the main properties of such equations [2, 8, 4].

2 Model and Numerical Method

We consider the numerical integration of a model describing the evolution of a physiologically-structured population which is structured by two features. The following linear model takes the form

$$u_t(x,y,t)+u_x(x,y,t))+u_y(x,y,t)) = -\mu(x,y,t)\, u(x,y,t), \quad 0<x<x_M,\ 0<y<y_M,\ t>0, \tag{1}$$

$$u(0,y,t) = \int_0^{x_M}\int_0^{y_M} \alpha(y,\varepsilon,\sigma)\, u(\varepsilon,\sigma,t)\, d\sigma\, \varepsilon\ , \quad t>0, 0\le y\le y_m. \tag{2}$$

$$u(x,0,t) = \int_0^{x_M}\int_0^{y_M} \beta(x,\varepsilon,\sigma)\, u(\varepsilon,\sigma,t)\, d\sigma\, \varepsilon\ , \quad t>0, 0\le x\le x_m. \tag{3}$$

$$u(x,y,0) = \phi(x,y), \quad 0\le x \le\le x_m,\ 0\le y\le y_m. \tag{4}$$

where variable t represents the time, and x and y are two traits correlated with time. The dependent variable u is the density of individuals of traits x and t at time t. The boundary conditions (2)-(3) account for the input at time t of individuals at the initial level of each characteristic. The mortality

[1] oscar.angulo@uva.es

and recovering rates (μ, α and β) depend on the structuring variables and the time. Finally, function ϕ denots the initial distribution of the population with respect to both characteristics.

Goal of this work is to introduce a numerical method for the computation of the solution of this linear population model with two structures. As far as we know, there is currently no other numerical method available for this problem. The first order numerical method we introduce is based on an upwind discretization as was made in an age-structured population model [5]. We will demostrate the applicability of the numerical method to obtain the solution and a theoretical study about its convergence.

References

[1] Iannelli, M., Mathematical Theory of Age-Structured Population Dynamics, Pisa, Giardini editori e stampatori, 1995.

[2] Inaba, H., Endemic threshold analysis for the Kermack–McKendrick reinfection model, *Josai Math. Monogr.* 9: 105–133, 2016.

[3] Inaba, H., Age-Structured Population Dynamics in Demography and Epidemiology. New York Springer, 2017.

[4] Kang, H, Huo, X., Ruan, S., On first-order hyperbolic partial differential equations with two internal variables modeling population dynamics of two physiological structures, *Annali di Matematica Pura ed Applicata* 200: 403–452, 2021.

[5] Lopez-Marcos, J.C., An upwind scheme for a nonlinear hyperbolic integro-differential equation with integral boundary condition, *Computers & Mathematics with Applications* 22:15-28, 1991.

[6] Metz, J.A.J., Diekmann, O., The Dynamics of Physiologically Structured Populations. Lecture Notes in Biomathematics, vol. 86. Berlin, Springer, 1986.

[7] Webb, G.F., Theory of Nonlinear Age-Dependent Population Dynamics. New York, Marcel Dekker, 1984.

[8] Webb, G.F., Dynamics of populations structured by internal variables, *Math. Z.* 189: 319–335, 1985.

Harvesting Populations in Random Environments: Shortcomings of the Optimal Harvesting Policy, Proposition of Alternatives and their Assessment

Carlos A. Braumann[♭1] and Nuno M. Brites[♮]

(♭) Universidade de Évora, Centro de Investigação em Matemática e Aplicações (CIMA), Portugal; Universidade de Évora, Escola de Ciências e Tecnologia, Portugal.

(♮) ISEG – School of Economics and Management, Universidade de Lisboa; REM – Research in Economics and Mathematics, CEMAPRE, Portugal.

1 Introduction

In a randomly varying environment, we can describe the evolution of a harvested population size $X(t)$ using stochastic differential equations – SDE (see, for instance, [3]).

The yield $H(t) = qE(t)X(t)$, the amount of fish captured per unit time (p.u.t.), is proportional to the harvesting effort $E(t)$. The harvesting profit p.u.t. is $\Pi(t) = P(t) - C(t)$, where $P(t)$ is the revenue p.u.t. from selling the fish and $C(t)$ is the cost p.u.t. of fishing. We assume $P(t) = p(t)H(t)$, where the price per unit of harvested fish is $p(t) = p_1 - p_2 H(t)$ with $p_1 > 0, p_2 \geq 0$, and $C(t) = c(t)E(t)$, where the cost per unit effort is $c(t) = c_1 - c_2 E(t)$ with $c_1 > 0, c_2 > 0$ (to account for use of less efficient material and overtime payments when harvesting efforts are higher). Given an initial population $X(0) = x$, a social/financial discount rate $\delta > 0$, and a time horizon T, the *present value* is the expected accumulated discounted profit $V = \mathbb{E}\left[\int_0^T e^{-\delta t}\Pi(t)dt \mid X(0) = x\right]$.

We can use stochastic optimal control theory and the Hamilton-Jacobi-Bellman (HJB) equation (see, for instance, [4, 5]) to determine the *optimal harvesting policy*, i.e., the effort $E^*(t)$ that must be applied at each time t to maximise the present value. Let $\Pi^*(t)$ be the corresponding profit p.u.t. and V^* the corresponding optimal present value.

Unfortunately, for stochastic models, the optimal harvesting effort is incompatible with the logistics of fisheries and therefore inapplicable. This is due to the fast and abrupt random variations associated with the variations in population size induced by the random environmental fluctuations. This optimal policy also poses social problems, like the unemployment of fishermen, during periods of no or low harvesting. Furthermore, it requires knowledge of the population size at each instant, and estimating population size is an inaccurate, lengthy, and expensive task.

To overcome some or all of these shortcomings, we have proposed alternative sub-optimal harvesting policies and assessed them in terms of applicability, possible social problems, and profit. Profit is assessed by comparing the present values of the sub-optimal policies with the present value V^* of the optimal stochastic control policy, even though V^* is unreachable since the corresponding optimal policy is inapplicable.

[1] braumann@uevora.pt

2 Alternative harvesting policies proposed and their assessment

For the numerical comparison of the present values of the different harvesting policies proposed with the optimal policy present value V^*, we use extensive Monte Carlo simulations and an Euler scheme for time discretization of the SDE. V^* is obtained by numerically solving the HJB equation with a Crank-Nicholson scheme. For each alternative, we will also compare the curves showing the time evolution of the population size, the effort p.u.t., and the profit p.u.t.

We have used different population growth models and particular fisheries for which data on population and profit parameters is available. Here, we will focus on the logistic growth model and an application to the Pacific halibut with data provided in [13], using a discount factor $\delta = 0.05$/year and a time horizon $T = 50$ years.

2.1 Constant effort policies

We have first proposed as an alternative a constant effort policy $E(t) \equiv E$, which is very easy to apply and, since the effort is constant, social problems are avoided and there is no need to keep estimating the population size. This policy is a particular case of general autonomous effort policies, for which a qualitative theory was developed in [1, 2] for very general SDE population growth models, providing conditions for extinction and conditions for the existence of a stochastic equilibrium with a population stationary distribution. For the particular case of a constant effort policy satisfying the conditions for the existence of a stochastic equilibrium, the determination and assessment of the optimal effort E^{**} that maximises the expected profit per unit time under the stationary distribution can be seen in [4, 5, 6, 7]. For the Pacific halibut data, it turns out that the corresponding present value V^{**} is only 4.1% lower than the present value V^* of the optimal inapplicable policy. Furthermore, the profit p.u.t. has much milder random fluctuations than in the optimal inapplicable policy.

2.2 Stepwise effort policies

Since the optimal variable effort policy is not applicable, we have proposed *stepwise policies* (see [7, 8]), which are sub-optimal policies where the harvesting effort is determined at the beginning of each year (or of each biennium) according to optimal control theory but is kept constant throughout that year (or biennium). These policies are not optimal and pose similar social problems, but they have the advantage of being applicable since harvesting effort only changes once a year or once every two years. Furthermore, even though estimating population size is still required, that needs to be done at the beginning of each one- or two-year period. For the Pacific halibut data, the 1-year stepwise policy has a present value only 1.7% lower than V^*. This is better than the constant effort policy, but the same is not true for the 2-year stepwise policy.

2.3 Penalised effort policies

To avoid abrupt effort changes and periods of low or null harvesting efforts and the associated social problems, we have also considered optimal penalised policies, which incorporate a term that represents a running energy cost based on the effort (see, for instance, [14]). This extra cost term is not a real cost, just an artificial way of penalising the profit values when, at each instant, the effort takes abrupt changes from a reference effort value, which we may take as E^{**}. These sub-optimal policies have milder effort changes, the mildness of which depends on the strength of the penalty parameter. Still, changes, although milder, are still very frequent, and determination of the optimal effort still requires the estimation of population size at every instant, making these policies also inapplicable. The profit loss in comparison with the optimal policy increases with the strength of the penalty parameter. The comparisons can be seen in [6, 11, 12].

2.4 Stepwise penalised effort policies

We have also proposed applying the stepwise procedure to the penalised effort policies. In this way, we keep the advantage of penalised policies of avoiding social problems and the advantage of stepwise policies of applicability and of only needing to estimate population size at the beginning of each time period. In terms of present values, these policies can, for an appropriate choice of the penalty parameter, be even better than the stepwise policies with the same period (see [12]).

3 Conclusions

For SDE models of populations living in random environments, the optimal policies found using stochastic optimal control theory don't work in real life, cause social problems, and one needs to estimate the size of the population all the time, which is a difficult process. Therefore, we have proposed alternative sub-optimal policies and compared them in terms of applicability, social problems, and profit. As described in the previous Section, we conclude that there are quite good alternatives.

For general SDE models of non-harvested populations with Allee effects, conditions for extinction and conditions for the existence of a stochastic equilibrium have been established in [10]. For harvested populations and the logistic-like model with Allee effects, we have also obtained such conditions and have compared the different harvesting policies in [9, 11, 12]. These results are not reported in this presentation due to time constraints, but they show that even mildly weak Allee effects can considerably reduce profit and affect the choice and design of a good alternative sub-optimal policy.

Acknowledgments

C. A. Braumann is a member of the Centro de Investigação em Matemática e Aplicações, supported by Fundação para a Ciência e a Tecnologia (FCT/MCTES), Project UID/04674/2020, `https://doi.org/10.54499/UIDB/04674/2020`. N. M. Brites was partially supported by the Project CEMAPRE/REM - UIDB/05069/2020 - financed by FCT/MCTES through national funds.

References

[1] Braumann, C. A., Variable effort harvesting models in random environments. *Mathematical Biosciences*, 156: 1–19, 1999.

[2] Braumann, C. A., Variable effort fishing models in random environments: generalization to density-dependent noise intensities. *Mathematical Biosciences*, 177 & 178: 229–245, 2022.

[3] Braumann, C. A., Introduction to Stochastic Differential Equations with Applications to Modelling in Biology and Finance. Wiley, 2019.

[4] Brites, N. M., Stochastic Differential Equation Harvesting Models: Sustainable Policies and Profit Optimization. Ph. D. thesis, Universidade de Évora, Portugal, 2017.

[5] Brites, N. M., Braumann, C. A., Fisheries management in random environments: comparison of harvesting policies for the logistic model. *Fisheries Research*, 195: 238–246, 2017.

[6] Brites, N. M., Braumann, C. A., Fisheries management in randomly varying environments: comparison of constant, variable and penalized efforts policies for the Gompertz model. *Fisheries Research*, 216: 196–203, 2019.

[7] Brites, N. M., Braumann, C. A., Harvesting in a random varying environment: optimal, stepwise and sustainable policies for the Gompertz model. *Statistics, Optimization and Information Computing*, 7(3): 533–544, 2019.

[8] Brites, N. M., Braumann, C. A., Harvesting policies with stepwise effort and logistic growth in a random environment. In: Current Trends in Dynamical Systems in Biology and Natural Sciences (Aguiar, M. *et al.*, eds.), SEMA SIMAI Springer Series, 95–110, 2020.

[9] Brites, N. M., Braumann, C. A., Stochastic differential equations harvesting policies: Allee effects, logistic-like growth and profit optimization. *Appl Stochastic Models Bus Ind.*, 36: 825–835, 2020.

[10] Carlos, C., Braumann, C. A., General population growth models with Allee effects in a random environment. *Ecological Complexity*, 30: 26–33, 2017.

[11] Brites, N. M., Braumann, C. A., Profit optimization of stochastically fluctuating populations under harvesting: the effects of Allee effects. *Optimization*, 71(11): 3277–3293, 2022.

[12] Brites, N. M., Braumann, C. A., Harvesting optimization with stochastic differential equations models: is the optimal enemy of the good? *Stochastic Models*, 39(1): 41–59, 2023.

[13] Hanson, F. B., Ryan, D., Optimal harvesting with both population and price dynamics. *Mathematical Biosciences*, 148: 129–146, 1998.

[14] Iolov, A., Ditlevsen, S., Longtin, A., Stochastic optimal control of single neuron spike trains. *J. Neural Eng. NIH*, 11: 1–22, 2014.

Modelling protective behaviour in response to mosquito-borne epidemics

A. Bizzotto[♮,♭], S. De Reggi[♮],[1] A. Pugliese[♮], M. Sabbatino[♮], and C. Soresina[♮]

(♮) Department of Mathematics, University of Trento, Trento, Italy
(♭) Center for Health Emergencies, Fondazione Bruno Kessler, Italy

Abstract

If humans protect themselves against mosquito bites, can we contain vector-borne epidemics? We examine a model for mosquito-borne epidemics, in which it is assumed that individuals may adopt behaviours that protect from mosquito bites when informed of an ongoing epidemic. The model is based on Ross-Macdonald assumptions, according to which mosquito biting rate is constant, and bites are distributed among protected humans, humans that do not use protective measures, and other mammals.

Assuming that host behaviour cannot change, it is found that the reproduction number of the epidemic (and thus its outbreak potential) may increase or decrease depending on the fraction that adopts protecting behaviour, on the effectiveness of protection, and on the ratio between other potential hosts and human population.

When information on the number of infected cases influences the adoption of protective measures, simulations show that the initial dynamics of the infection may change (relative to an epidemic without behavioural changes) in a way corresponding to what is found for the static case. However, in the longer term, multiple infection waves may be observed.

These theoretical results may help in designing public health communication strategies.

Acknowledgments

This work was supported by the MUR PNRR BaC PER INF-ACT S4 BEHAVE-MOD project (No. PE00000007, CUP: I83C22001810007).

References

[1] L. Bolzoni, A. Pugliese, R. Rosà, The role of heterogeneity on the invasion probability of mosquito-borne diseases in multi-host models, *Journal of Theoretical Biology* 377: 25-35, 2015.

[2] A. d'Onofrio, P. Manfredi, Information-related changes in contact patterns may trigger oscillations in the endemic prevalence of infectious diseases, *Journal of Theoretical Biology* 256(3): 473-478, 2009.

[1] simone.dereggi@unitn.it

[3] C. Dye, H. Günther Hasibeder, Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others, *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80(1):69-77, 1986.

[4] K. Roosa, N. H. Fefferman, A general modeling framework for exploring the impact of individual concern and personal protection on vector-borne disease dynamics, *Parasites & Vectors* 15(1): 361, 2022.

Pseudospectral approximation of Floquet multipliers for state-dependent delay differential equations

A. Andò, D. Breda, D. Liessi[1] and C. Tanase

CDLab – Computational Dynamics Laboratory,
Department of Mathematics, Computer Science and Physics,
University of Udine, via delle Scienze 206, 33100 Udine, Italy.

Delays often appear in mathematical models. Examples from biology include maturation in population models, incubation and latent periods in epidemics models, impulse transmission time in neurological models, and transcription delays in gene regulatory networks. Some of these delays may depend on the state of the system: as an example, the maturation age of an individual may depend on the availability of resources during the individual's life. Models with state-dependent delays appear also in other context, such as control theory, economics, lasers and climate (see [3, 4] and the references therein).

A common problem in the study of dynamical systems is the determination of the asymptotic stability of particular invariant sets of states, such as equilibria and periodic orbits. In both cases the stability properties can be investigated by analyzing the spectra of the evolution operators associated to linearizations of the system; in particular, for periodic orbits, a stability indicator is given by the largest modulus of the Floquet multipliers, which constitute the spectrum of the monodromy operators.

The method of [1] applies a pseudospectral collocation technique to the evolution operators of linear(ized) delay differential equations (DDEs) with constant delays, obtaining finite-rank approximating operators (converging in norm). These can be represented as matrices, and their eigenvalues, computed with standard methods, are taken as approximations of the desired spectrum. Our goal is to extend this approach to state-dependent DDEs, i.e. DDEs with state-dependent delays.

We consider the state-dependent DDE

$$u'(t) = \alpha u(t) + \beta u(t - 1 - u(t - b)) \tag{1}$$

proposed in [4]. After a brief summary of the relevant theoretical background [2, 5], we will describe the numerical method and its implementation. We will then present some experiments on (1) comparing the results with those obtained in [4] using DDE-BIFTOOL, and we will exemplify the convergence properties of the method.

Acknowledgments

AA and DB are members of INdAM research group GNCS. AA, DB and DL are members of UMI research group "Modellistica socio-epidemiologica". This work originates from CT's MSc thesis [6], supervised by AA, DB and DL. This work was partially supported by the Italian Ministry of

[1] davide.liessi@uniud.it

University and Research (MUR) through the PRIN 2022 project (No. 20229P2HEA) “Stochastic numerical modelling for sustainable innovation” (CUP: G53C24000710006).

References

[1] Breda, D., Maset, S., Vermiglio, R., Approximation of eigenvalues of evolution operators for linear retarded functional differential equations, *SIAM J. Numer. Anal.*, 50(3):1456–1483, 2012, DOI:10.1137/100815505.

[2] Hartung, F., Krisztin, T., Walther, H.-O., Wu, J., Functional differential equations with state-dependent delays: theory and applications, in *Handbook of Differential Equations: Ordinary Differential Equations*, vol. 3, Cañada, A., Drábek, P., Fonda, A., eds., Amsterdam, North-Holland, 2006.

[3] Humphries, A. R., DeMasi, O. A., Magpantay, F. M. G., Upham, F., Dynamics of a delay differential equation with multiple state-dependent delays, *Discrete Contin. Dyn. Syst.*, 32(8):2701–2727, 2012, DOI:10.3934/dcds.2012.32.2701.

[4] Humphries, A. R., Krauskopf, B., Ruschel, S., Sieber, J., Nonlinear effects of instantaneous and delayed state dependence in a delayed feedback loop, *Discrete Contin. Dyn. Syst. Ser. B*, 27(12):7245–7273, 2022, DOI:10.3934/dcdsb.2022042.

[5] Mallet-Paret, J., Nussbaum, R. D., Stability of periodic solutions of state-dependent delay-differential equations, *J. Differential Equations*, 250(11):4085–4103, 2011, DOI:10.1016/j.jde.2010.10.023.

[6] Tanase, C., *On discretizing monodromy operators of state-dependent delay differential equations*, MSc thesis, University of Udine, 2025.

A meta-population model for non-timber forest product harvesting: connectivity, plant viability, and sustainable yield

Maria C.A. Leite[♭1], Ruth I. Oliva-Zúniga[♮], and Pedro J. Torres[†]

(♭) Department of Applied Mathematics, Universidad de Oviedo, Oviedo, Spain.

(♮) Department of Mathematics, Universidad Nacional Autónoma de Honduras, Cortés, Honduras.

(†) Department of Applied Mathematics, Universidad de Granada, Granada, Spain.

1 Introduction

Fragmentation and habitat loss are viewed as important causes of known extinctions and the main threats to species persistence.

For plants, smaller habitats experience lower seed set, have declining populations, and are more likely to suffer from competition and displacement caused by invasive species. Additionally, habitat loss and fragmentation lead to both direct and indirect effects on plant species, including altered environmental conditions that further challenge their survival and persistence.

Harvesting non-timber forest products (NTFP) from species of wild plants is a long-standing issue in ecological research and has been widely recognised to affect the dynamics of the plant population, alongside other factors such as habitat fragmentation and environmental change [4].

The practice of harvesting NTFPs has increased mainly because of current high human population growth, income generation needs, and a clear shift in the non-lethal harvesting of plants from subsistence use to commercial sale. Thus, it has resulted in the over-exploitation of plants. Although it is clear that the harvesting of plants provides economic benefits to people both on an individual level and on a national scale, it is also critical to consider the long-term effects of the harvest on the plant population in fragmented habitats so that we continue to reap the benefits in the future.

Research on the population response to habitat loss and NTFP harvesting has evolved independently. Furthermore, each of these research fields has remained the focus of considerable study and debate. There are a few recent works addressing the combined effect of these two factors on population dynamics (e.g., [6, 7], and references therein).

In this work, we continue to use mathematical modelling to explore further these dynamics. The aim is to investigate what are the rational non-lethal harvesting strategies that maximize the yield in a fragmented ecosystem; and at the same time assuring the viability of the plant. At this point, it is necessary to introduce the precise definitions of two key concepts:

- **Maximum sustainable yield (MSY)**: The maximum sustainable yield is a key concept in ecology. It refers to the largest yield that can be taken from a specific harvestable resource over

[1] almeidamaria@uniovi.es

an indefinite period under constant environmental conditions without affecting the resource's long-term viability.

- **Optimal harvesting rate**: The optimal harvesting rate is simply the harvesting rate at which the MSY is reached. This term is frequently used in literature. To avoid misinterpretation, it is important to clarify that using this term does not imply that the particular harvesting strategy considered in our model is "optimal" with respect to other possibilities. Additionally, the model that we propose does not include economic factors such as market demand, costs, or prices.

2 Methods

In this paper, we propose a meta-population model within the existing framework of meta-population theory to better understand the effect of the interactions between the harvesting and fragmentation of NTFPs on the dynamics of the plant population. The model is applicable to plant population dynamics mediated by seed dispersal [10].

To the best of our knowledge, this is the first meta-population model crafted to the particular case of plant populations under NTFP harvesting. Due to the specificity of plant ecosystems, the term modeling the connectivity between patches is density-dependent, in contrast with most of the existing meta-population models. Specifically, the connectivity is reflected in the intrinsic growth rate of the logistic model, rather than through a standard linear removal term that is used in classical meta-population models. This aspect adds more complexity to the theoretical analysis of the model.

There is a large body of literature on meta-population models and they have been widely used to model heterogeneous ecosystems and to investigate distinct aspects of population dynamics (e.g., [2, 5, 1] and references therein). Since fragmented ecosystems are heterogeneous landscapes, they can be described as a network of discrete habitat patches within a (possibly hostile) matrix of habitats of other types. Consequently, fragmented ecosystems have been widely described by meta-populations models (e.g.[5, 1], and references therein). Furthermore, it has been indicated that meta-population models are especially suitable for plant populations (e.g., [2]) because plants, as opposed to animals, occupy fix sites. There no physical migratory flow of individuals among sites. The plants disperse by means of pollination or seed displacement to colonize empty sites.

Our model is built upon one of the standard meta-population model approaches, which consists in considering ordinary differential equations (ODEs) to model the dynamics of the population (e.g, [5]). The underlying assumption is an average or mean-field approximations of the true time and spatial scales of the object of study, which classically is used to investigate asymptotic dynamics. Furthermore, we consider that each patch is homogeneous, which is also another widely used assumption in meta-population modeling approach (e.g, [5]).

Alternative model formulation could involve, for example, individual based models (e.g., [8, 9] and references therein). However, any attempt in this direction would be mathematically challenging and would require the apriori knowledge of a multitude of parameters and, necessarily, data to validate it. These requirements may be difficult to satisfy.

Therefore, since the model is new we kept it simple and focus only on the feedback between fragmentation and non-lethal harvesting. The approach of using a simpler model enables us to provide an analytical study of the model and, therefore, gain theoretical insights.

Recent work by [5] using meta-population models focused on investigating how the topology of the network of patches maximizes the total population size to derive management guidelines for Marine Protected Areas and the construction of ecological corridors. Their results, consistent with well-established ecological and conservation biology principles, indicate that the optimal network

structure is strongly dependent on the degree of mobility of the populations. In [1], the authors investigate the effect of the connectivity among patches on the maximum sustainable yield (MSY) in the context of fishery models. These works do not apply directly to plant populations and do not investigate the effect of coupling non-lethal harvesting strategies and fragmentation disturbances, which is the focus of our paper. However, they were a source of motivation for carrying-on our study.

The new model is used to investigate the following specific research questions:

1. For a fragmented ecosystem with fixed characteristics, what is the suitable choice of harvesting rates that maximizes the harvesting yield and, at the same time guarantees the viability of the plant population? In particular, we consider all parameters of the model fixed except the NTFP harvesting rates, and investigate the rational harvesting strategies that guarantee a maximum harvesting yield and plant population survival.

2. What is the desirable patch connectivity that provides the maximum harvesting yield in a viable plant population? Specifically, we investigate the effect of connectivity on the maximum harvesting yield under the condition that the plant population does not go extinct.

We focus on the asymptotic dynamical behavior of the system (long-term dynamics) to ensure the survival of the plant population. Both questions are investigated in an ecosystem of two patches.

The first question is addressed by deriving a main result that provides explicit formulas for computing the optimal harvesting rates.

The second research question is explored under two scenarios: first, considering symmetric connectivity between patches, where the rate of seed transference is identical in both directions; and second, under fully asymmetric connectivity, where seeds disperse from one patch to the other but not in the reverse direction. Both cases are examined through a combination of analytical methods and numerical simulations to provide a comprehensive understanding of the effects of connectivity on harvesting yield and plant population viability.

The model we propose is given by the following system of equations:

$$\frac{db_1}{dt} = \Big(\beta_{11}(1-\lambda_{21})b_1 + \beta_{12}\lambda_{12}b_2\Big)\left(1-\frac{b_1}{K_1}\right) - \alpha_1 b_1; \tag{1a}$$

$$\frac{db_2}{dt} = \Big(\beta_{22}(1-\lambda_{12})b_2 + \beta_{21}\lambda_{21}b_1\Big)\left(1-\frac{b_2}{K_2}\right) - \alpha_2 b_2; \tag{1b}$$

The variables and parameters involved in the model are described in Table 1.

Table 1: Description of the variables and parameters in model (1).

Variable	Description	Units
b_i	Total biomass in patch i, where $i = 1, 2$	Mass
Parameter	**Description**	**Units**
β_{ij}	Growth rate contribution from patch j to patch i	1/time
λ_{ij}	Fraction of seeds produced in patch j that land in patch i	Dimensionless
K_i	Carrying capacity of patch i	Mass
α_i	Harvesting rate in patch i	1/time

3 Results

The primary objective of this study is to investigate the influence of harvesting rates and patch connectivity on the yield of non-timber forest products (NTFPs) in fragmented plant populations. Our

findings provide insights into the interactions between NTFP harvesting and habitat fragmentation and their effects on plant population dynamics.

We develop a novel deterministic meta-population model tailored to populations of plants subject to NTFPs harvesting with n patches. The model consists of a system of ordinary differential equations and is applicable to plant populations whose dynamics are driven by seed dispersal. A new feature is an explicit dependence of the intrinsic growth rate of the plant population on the connectivity between patches. We model the colonization of new sites by the movement of seeds. Consequently, dispersal of seeds modifies the natural growth rates of the patches, in contrast to classical meta-population models where there is a direct transfer of individuals between patches [5, 1]. This nonlinearity complicates the theoretical analysis of our model, but we successfully demonstrated results for the case of 2 patches without employing the fast scale hypothesis.

Our findings pertain to long-term dynamics. They were derived assuring that the dynamics of the ecosystem were in a steady state in which at least one patch maintains nonzero biomass, thereby ensuring the sustainability of plant populations over time. Additionally, we restrict the domain of the variables that represent biomass so that they have non-negative values and that the patches are never above their carrying capacity.

Concerning the first research question, we found sustainable harvesting rates that maximize harvesting yield without jeopardizing population viability. We identify four possible cases: Case 1, Cases 2.1-2.3. In each case, we provide formulas to compute the maximum harvesting rate explicitly. Thus, for specific plant populations with available data, a harvesting intensity can be determined. This is particularly relevant in the context of sustainable natural resource management. To the best of our knowledge, most of the existing studies do not provide such key explicit formulas.

In Case 1, optimal harvesting rates α_1, α_2 are positive and finite. In Case 2.1, the optimal harvesting choice is to harvest only in one patch. In Case 2.2, the optimal harvesting rate is finite in one of the patches. In the other patch, the MSY is approached asymptotically as the harvesting rate tends to infinite. Finally, in Case 2.3 the harvesting rate in one of the patches is zero while in the other patch, as in Case 2.2, MSY is approached asymptotically as the harvesting rate tends to infinite. Cases 2.2 and 2.3 are not realistic because it is not feasible to increase the NTFP harvesting rate indefinitely. This issue could potentially be addressed by introducing harvesting costs, a factor not considered in our work.

To investigate the second research question, we conducted an analytical and numerical study. We investigate how connectivity between the two patches affects the maximum sustainable yield (MSY) for two case studies (symmetric vs. asymmetric). In our work, the connectivity is measured by the rate at which seeds transfer between patches, i.e, λ_{12} and λ_{21}. In this way, we avoid the specification of the mechanisms underlying seeds' transference from patch to patch. This is a relevant aspect to be studied, which is not well understood. Nonetheless it is not the focus of our paper.

Our results show that the influence of patch connectivity on the MSY relies not only on the network topology, but also on the dispersal rate, λ_{ij}'s. Our model without harvesting keeps constant the total carrying capacity regardless of the topology of the network. However, the MSY is highly sensitive to variations in the topology or in the value of the dispersal rate. Thus, our findings suggest that small alterations in the network structure or dispersal parameters can have a significant impact on the sustainability of NTFPs harvesting, highlighting the need for adaptable management based on precise data. In contrast, in the family of models by [5, 1] the network's topology affects the total carrying capacity of the ecosystem. However, the MSY remains constant, at least for small values of the dispersal rate.

An important result of our study is that the the MSY, as a function of the dispersal rate, can exhibit non-monotonic behavior in both symmetric and asymmetric cases. Analytical results indicate that increasing patch connectivity does not always lead to an increase in sustainable yield.

4 Conclusions

The meta-population model developed in this paper belongs to a class of models that incorporate few ecological details but are suitable for mathematical analysis. Despite the ecological simplification inherent in our model, it captures the intricate and not yet fully understood nonlinear synergetic effects of coupled NTFP harvesting and fragmentation drivers on the dynamics of the plant population and, consequently, on MSY (e.g. [7, 9] and references therein). However, recent work such as [9] bridging meta-population ecology and landscape shows the importance of integrating a higher level of system description, which our current model does not account for. An extension of our model to overcome this limitation could involve incorporating heterogeneity within each patch and including meta-communities. The extension could also account for the variability in the ecosystem by integrating randomness. This would allow, for example, to explore how short- and long-term dynamics compare. It has been shown that these dynamics may differ across temporal scales (e.g.,[6]).

Several aspects warrant further exploration within the current model framework. For example, the concepts outlined for the toy model with two patches can be directly extrapolated to meta-populations with an arbitrary number of patches. Our theoretical and numerical study already provides initial insights into how optimal harvesting rates and MSY are impacted by changes in inter- and intra-patch contributions to the intrinsic growth rate of the plant population. A more comprehensive future study would provide a deeper understanding of the underlying processes. The simplest logistic law to describe the dynamics within each patch was used as a first step, but more complex consumer/resource models can be analyzed (see [1] for the case of fisheries). Another direction would be the integration of data to parametrize the model and test its performance against real-world scenarios.

Acknowledgments

RIOZ thanks the Faculty for the Future program of the Schlumberger Foundation for the scholarship.

References

[1] Auger, P., Kooi, B., Moussaoui, A., Increase of maximum sustainable yield for fishery in two patches with fast migration. *Ecological Modelling*, 467: 109898–109898, 2022.

[2] Cantrell, R. S., Cosner, C., Fagan, W. F., The implications of model formulation when transitioning from spatial to landscape ecology. *Mathematical Biosciences and Engineering*, 9(1):27–60, 2011.

[3] Leite, M. C. A., Sauchuk, R., Agusto, F. B., Gaoue, O. G., Chen-Charpentier, B., Modeling the persistence of plant populations in fragmented ecosystems. *Ecological Modelling*, 457: 109681–109681, 2021.

[4] Schmidt, I. B., Mandle, L., Ticktin, T., Gaoue, O. G., What do matrix population models reveal about the sustainability of non-timber forest product harvest? *Journal of Applied Ecology*, 48(4):815–826, 2011.

[5] Ruiz-Herrera, A., Torres, P. J., Optimal network architectures for spatially structured populations with heterogeneous diffusion. *The American Naturalist*, 196(1):29–44, 2020.

[6] Leite, M. C. A., Chen-Charpentier, B., Agusto, F. B., Gaoue, O. G., Hritonenko, N., Viability of *Pentadesma* in reduced habitat ecosystems within two climatic regions with fruit harvesting. *Journal of Biological Dynamics*, 16(1):207–235, 2022.

[7] Gaoue, O. G., Kouagou, M., Natta, A. K., Gado, C., Response of a tropical tree to non-timber forest products harvest and reduction in habitat size. *PLOS ONE*, 12(8):1–1, 2017.

[8] Cantrell, R. S., Cosner, C., *Spatial Ecology via Reaction-Diffusion Equations*. Chichester, John Wiley & Sons Ltd, 2003.

[9] Tao, Y., Hastings, A., Lafferty, K. D., Hanski, I., Ovaskainen, O., Landscape fragmentation overturns classical metapopulation thinking. *Proceedings of the National Academy of Sciences*, 121(20):1–1, 2024.

[10] Leite, M. C. A., Oliva-Zúniga, R. I., Torres, P. J., Plant viability: Modeling the impacts of harvesting non-timber forest products in a fragmented landscape. *Ecological Modelling*, 504: 111080–111080, 2025.

A Free-boundary Nonlocal Diffusion Logistic Model: Analysis and Computing

V. Egorova[♮],[1] M.-C. Casabán[♭] R. Company[♭] L. Jódar[♭]
and M. El Fakharany[♯],[♭♭]

(♮) Depto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria
Avda. de los Castros, s/n, 39005, Santander, Spain

(♭) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València
Camí de Vera s/n, 46022, València, Spain

(♯) Mathematics Department, Faculty of Science, Tanta University
31527, Tanta, Egypt

(♭♭) Mathematics and Statistics Department, College of Science, Taibah University
42353, Yanbu, Saudi Arabia

1 Introduction

The study in [3] examines a Fisher-KPP population model that includes nonlocal diffusion and free boundaries. This model offers a mathematical framework to understand how populations change as species spread in habitats where the edges (boundaries) can move or change. It is an extension of the standard Fisher-KPP equation. The addition of nonlocal diffusion allows the model to describe more complex patterns of species dispersal found in the real world. The main equation for this system is:

$$\partial_t u = \alpha^2 \left(\int_{g(t)}^{h(t)} J(x-y)u(t,y)dy - u(t,x) \right) + f(u), \quad t > 0,\ g(t) < x < h(t). \tag{1}$$

Here, $u(t,x)$ is the population density at time t and position x, within moving boundaries $g(t)$ and $h(t)$. The function $J(x-y)$ is the dispersal kernel, giving the probability density of movement from y to x. This kernel J is typically assumed to be symmetric, non-negative, and normalized such that $\int_{\mathbb{R}} J(z)\,dz = 1$. The growth function $f(u)$ dictates the local reproduction rate of the population, usually of logistic type

$$f(u) = ru(1-u/K), \tag{2}$$

where r is the growth rate, and K is the carrying capacity. The growth function $f(u)$ is positive for $u \in (0,K)$, and the growth rate $f(u)/u$ must be strictly decreasing for $u > 0$ [1, 2]. Moreover,

$$f(0) = f(K) = 0; \quad f'(u) \text{ exists, and } f'(0) > 0. \tag{3}$$

[1] vera.egorova@unican.es

The system also includes the following boundary conditions:

$$u(t, g(t)) = u(t, h(t)) = 0. \tag{4}$$

These conditions state that the population density is zero at the moving boundaries $g(t)$ and $h(t)$. In addition, the model requires initial conditions:

$$u(0, x) = u_0(x), \quad g(0) = -h_0, \quad h(0) = h_0, \quad h_0 > 0. \tag{5}$$

The dynamics of the boundaries are governed by Stefan conditions. These conditions depend on the outward population flux across these boundaries:

$$\frac{dh}{dt} = \mu \int_{g(t)}^{h(t)} \int_{h(t)}^{+\infty} J(x-y)u(t,x)dydx, \quad \frac{dg}{dt} = -\mu \int_{g(t)}^{h(t)} \int_{-\infty}^{g(t)} J(x-y)u(t,x)dydx. \tag{6}$$

The aim of this study is to propose a numerical method for solving this free-boundary nonlocal diffusion model. This helps to extend the previous workd [1, 4] in key qualitative characteristics of the model, such as the conditions of the spreading-vanishing dichotomy. Moreover, we aim to explore how different types of dispersal kernels influence the speed and nature of population spread, including the possibility of accelerated spreading.

2 Methods

2.1 Front-fixing transformation

To handle the complexities associated with two moving boundaries, $g(t)$ and $h(t)$, a front-fixing transformation is used. This technique transforms the original, time-varying spatial domain $[g(t), h(t)]$ into a fixed domain (e.g., $[0, 1]$). This fixed domain is much more suitable for applying numerical methods.

The transformation proceeds in two stages:

1. Fixing the left boundary: A new spatial variable $z = x - g(t)$ is introduced. This change of variable maps the original left boundary $x = g(t)$ to the fixed position $z = 0$. The population density function is redefined in terms of this new variable as $v(t, z) = u(t, x)$. This first transformation maps the original time-varying domain $[g(t), h(t)]$ to an intermediate domain $[0, l(t)]$, where the length $l(t) = h(t) - g(t)$ still depends on time.

2. Scaling to a fixed domain: A second variable transformation $r = \dfrac{z}{l(t)}$ is applied. This scales the intermediate domain $[0, l(t)]$ to the fixed unit domain $[0, 1]$. The population density function is again redefined as $w(t, r) = v(t, z)$. To express the equation in terms of $w(t, r)$, we use the cumulative distribution function (CDF) of the dispersal kernel $J(s)$, defined as

$$K(x) = \int_{-\infty}^{x} J(s)\, ds.$$

The transformed equation governing $w(t, r)$ in the fixed domain $[0, 1]$ can then be written as follows:

$$\begin{aligned} \frac{\partial w}{\partial t} =& \mu \left[r \int_0^1 w(t,\rho) \left(1 - K[l(t)\rho] + K[l(t)(\rho - 1)]\right) d\rho - \int_0^1 w(t,\rho) \left(1 - K[l(t)\rho]\right) d\rho \right] \frac{\partial w}{\partial r} \\ &+ \alpha^2 \left[l(t) \int_0^1 J(l(t)(r-\rho)) w(t,\rho)\, d\rho - w(t,r) \right] + f(w), \quad t > 0,\ 0 \le r \le 1. \end{aligned} \tag{7}$$

Stefan conditions are transformed as follows

$$g'(t) = -\mu\, l(t) \int_0^1 w(t,\rho)\,(1 - K[l(t)\rho])\; d\rho, \tag{8}$$

$$l'(t) = \mu\, l(t) \int_0^1 w(t,\rho)\,(1 - K[l(t)\rho] + K[l(t)(\rho-1)])\; d\rho. \tag{9}$$

The initial and boundary conditions from (4)–(5) are now rewritten as follows:

$$w(t,0) = w(t,1) = 0, \quad w(0,r) = u_0\left((h_0 - g_0)r + g_0\right), \quad l(0) = l_0 = h_0 - g_0. \tag{10}$$

2.2 Numerical algorithm

The transformed PIDE problem (7) with conditions given in (8), (9) and (10) is solved numerically by using a finite difference method (FDM).

Let t range from 0 to T with a time step $\Delta t = \dfrac{T}{N}$. Similarly, r ranges from 0 to 1 with a space step $\Delta r = \dfrac{1}{M}$. We define a uniform grid

$$r_i = i\Delta r,\ i = 0,1,2,\ldots,M, \quad t^n = n\Delta t,\ n = 0,1,2,\ldots,N. \tag{11}$$

The approximated numerical solution at each grid node is denoted by $w_i^n \approx w(t_n, r_i)$ and the distance between two free boundaries $g^n \approx g(t^n)$ and $h^n \approx h(t^n)$, is denoted by $l^n \approx l(t^n)$.

The integrals arising in the model must be approximated numerically. Several quadrature rules are suitable for this, such as the midpoint rule or the trapezoidal rule, which rely on piecewise constant or linear approximations of the integrand, respectively. For this study, we select Simpson's rule. This choice is motivated by its higher order of accuracy compared to the aforementioned methods. We now introduce the notation required for its application:

$$S(w^n) = \frac{\Delta r}{3}\left(w_0^n + 4\sum_{\text{odd } i} w_i^n + 2\sum_{\text{even } i} w_i^n + w_M^n\right), \tag{12}$$

where $w^n = [w_0^n, w_1^n, \ldots, w_M^n]$.

Using a forward Euler approximation for the time derivatives in (8)–(9), we get the updates for the boundary position g^n and the interval length l^n at time step $n+1$ based on values at time step n:

$$\frac{g^{n+1} - g^n}{\Delta t} = -\mu\, l^n S(p^n), \quad \frac{l^{n+1} - l^n}{\Delta t} = \mu\, l^n S(q^n), \tag{13}$$

where $S(p^n)$ and $S(q^n)$ represent the numerical values of the boundary flux integrals, approximated using Simpson's rule applied to the vectors p^n and q^n. The components of these vectors at time step n are defined as:

$$p_j^n = w_j^n(1 - K[l^n r_j]), \quad q_j^n = w_j^n(1 - K[l^n r_j] + K[l^n(r_j - 1)]),\ j = 0,\ldots,M. \tag{14}$$

Consequently, the advection coefficient is approximated at the discrete node (r_i, t^n) using the Simpson's rule approximations $S(p^n)$ and $S(q^n)$ as follows: $A^n(r_i) = \mu\left(r_i S(q^n) - S(p^n)\right)$.

Initial and boundary conditions (10) for all considered methods are given by

$$w_0^n = w_M^n = 0,\ n = 0,1,2,\ldots,N; \quad w_i^0 = u_0\left((h_0 - g_0)r_i + g_0\right),\ i = 0,1,2,\ldots,M; \quad l^0 = h_0 - g_0. \tag{15}$$

Table 1: Maximum Courant number $C^*(M)$ and corresponding CPU time (in seconds) for various spatial step counts M in Example 1, with $T = 50$ and a time step size of $\Delta t = 10^{-4}$.

M	50	100	200	400	800	1600
$C^*(M)$	7.72×10^{-4}	$1.54 \cdot 10^{-3}$	3.09×10^{-3}	6.17×10^{-3}	$1.23 \cdot 10^{-2}$	2.47×10^{-2}
CPU-time, s	14.36	32.36	73.97	209.95	598.72	1622.77

When solving the time-dependent transformed PDE numerically, a simple time-stepping method like the explicit Euler scheme could be considered due to its simplicity. However, explicit Euler is only first-order accurate in time and often suffers from severe stability restrictions. Furthermore, it can introduce significant numerical errors, primarily numerical diffusion or dispersion. To overcome these limitations, we employ the Lax-Wendroff method [5]. This is a well-established scheme specifically designed for hyperbolic partial differential equations (PDEs), which often govern advection-dominated phenomena. The time derivative is approximated as follows

$$\frac{\partial w}{\partial t} \approx \frac{w_i^{n+1} - w_i^n}{\Delta t} - \frac{\Delta t}{2} w_{tt} = \frac{w_i^{n+1} - w_i^n}{\Delta t} - \frac{\Delta t}{2} |A^n(r_i)|^2 w_{rr}. \tag{16}$$

Then, (7) is approximated as follows

$$w_i^{n+1} = w_i^n + \frac{C_i^n}{2}(w_{i+1}^n - w_{i-1}^n) + \frac{(C_i^n)^2}{2}(w_{i+1}^n - 2w_i^n + w_{i-1}^n) + \Delta t \alpha^2 \left[l^n S(y^n(r_i)) - w_i^n\right] + \Delta t f(w_i^n). \tag{17}$$

The Lax-Wendroff scheme is conditionally stable for linear convection equation if the Courant number is less or equal to one, which corresponds, in our case, to the condition $\Delta t < \dfrac{\Delta r}{\mu}$.

3 Results

Computations have been carried out by Matlab© software version R2024b for Windows 11 Home (64-bit) 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10GHz 3.11 GHz.

Example 1 *Let us consider the problem* (1) – (6) *with the following parameters*

$$h(0) = h_0 = 0.7 = -g(0),\ \ \mu = 1.5,\ \ \alpha^2 = 2,\ u(0,x) = h_0^2 - x^2,\ f(u) = 2u(1-u),\ J(x) = \frac{1}{\pi(1+x^2)}. \tag{18}$$

In Example 1, we set the time step size as $\Delta t = 1 \times 10^{-4}$ and investigate the spatial discretization using a set of values for M from the set $\{50, 100, 200, 400, 800, 1600\}$. The maximum Courant number is defined as $C^* = \max\limits_{i,n} |C_i^n| = \max\limits_{i,n} |A^n(r_i)| \dfrac{\Delta t}{\Delta r}$. Table 1 reports the maximum Courant number for different numbers of spatial nodes, alongside the corresponding CPU times (in seconds) for the Lax-Wendroff numerical scheme.

3.1 Spreading-vanishing dichotomy study

Now, let us examine the theoretical results regarding the spreading-vanishing dichotomy proposed in [1]. According to this theory, vanishing occurs if $f'(0) < \alpha^2$, $h_0 < \ell^*/2$, and $\mu \le \mu^*$, where ℓ^* and μ^* are some positive constants to be determined.

In the case of the parameters given by (18), where $f'(0) = \alpha^2 = 2$, spreading always occurs, independent of the choice of h_0 and μ. However, if we consider a family of functions $f_\gamma(u) = \gamma u(1-u)$ as the growth term, we expect to observe both spreading and vanishing scenarios. For this analysis, we extend the time horizon to $T = 25$, while the rest of the parameters remain the same as taken in (18).

Now, we set $f(u) = u(1-u)$ as a boundary case and vary h_0 to observe spreading-vanishing dichotomy. If $h_0 < 0.7$, the vanishing observed, however, for $h_0 > 0.7$, the spreading occurs. Note that here we do not pretend that $h_0 = 0.7$ is a threshold value $\ell^*/2$, but it can be considered an initial guess for the approximation.

Finally, we consider $f(u) = u(1-u)$ with $h_0 = 0.5$ (resulting in vanishing) and vary μ to examine the spreading-vanishing dichotomy. When $\mu \geq 5$, spreading occurs. This indicates that there exists a critical value μ^* such that if $\mu > \mu^*$, spreading will happen.

3.2 Impact of kernel function

Studies [1, 2, 4] analyze the long-term behavior (spreading vs. vanishing) of solutions to this nonlocal free boundary problem. The type of spreading fundamentally depends on the properties of the dispersal kernel $J(x)$. A critical condition is the following: if

$$\int_0^\infty xJ(x)\mathrm{d}x < \infty. \tag{19}$$

holds, spreading occurs linearly at a constant asymptotic speed $c_0 > 0$, meaning $\lim_{t\to\infty} h(t)/t = -\lim_{t\to\infty} g(t)/t = c_0$ [2].

If condition (19) fails (a "heavy-tailed" kernel), the population undergoes *accelerated spreading*, where $\lim_{t\to\infty} h(t)/t = \lim_{t\to\infty} -g(t)/t = \infty$. This phenomenon is examined in [3, 4].

A related, stricter condition involves the existence of exponential moments ("thin-tail"):

$$\exists\, \phi > 0 \,\text{such that} \int_{-\infty}^\infty J(x)e^{\phi x}\mathrm{d}x < \infty. \tag{20}$$

Condition (20) implies (19). If (19) holds but (20) fails, spreading remains linear, although its characteristics might differ from the thin-tailed case [4].

Let us analyze and compare several kernel functions listed in Table 2.

The evolution of the habitat, $h(t)$, over time, is shown for each kernel in Figure 1. The results are consistent with previous theoretical findings for nonlocal diffusion problems discussed in [2, 4]. Specifically, in the case of spreading, it has a finite speed if and only if condition (19) is satisfied; otherwise, spreading becomes accelerated.

Among the kernels examined, $J_1(x)$ and $J_4(x)$ fail to satisfy the first-moment condition (19) (see the third column of Table 2). Specifically, for $J_1(x) \sim |x|^{-2}$ as $|x| \to \infty$, we have:

$$\lambda_1 = \lim_{x\to\infty} J_1(x)x^2 = \lim_{x\to\infty} \frac{x^2}{\pi(1+x^2)} = \frac{1}{\pi}. \tag{21}$$

Therefore, the spreading rate is $\lim_{t\to\infty} \dfrac{h(t)}{t\ln t} = \dfrac{\mu}{\pi}$.

Numerically, by fitting the regression equation $LR(t\ln t) = c_0 t\ln t + c_1$, we obtain $c_0 = 0.4247$, which closely approximates the theoretical value $c_0^* = \dfrac{\mu}{\pi} = 0.4775$. This agreement validates the accuracy of our numerical algorithm in capturing the theoretical accelerated spreading rate.

Similarly, for $J_4(x) \sim |x|^{-3/2}$ as $|x| \to \infty$, we compute:

$$\lambda_4 = \lim_{x\to\infty} J_4(x)x^{3/2} = \lim_{x\to\infty} \frac{\Gamma\left(\frac{3}{4}\right)x^{3/2}}{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)(1+x^2)^{3/4}} = \frac{\Gamma\left(\frac{3}{4}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)} \approx 0.1907. \tag{22}$$

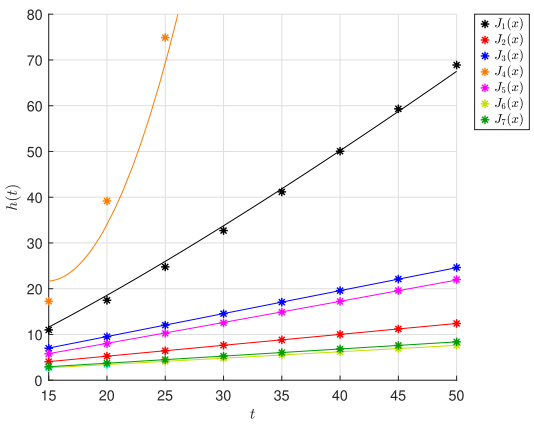


Figure 1: Free boundary $h(t)$ over time for various kernel functions of Table 2 and parameters given in (18). Dashed lines represent the corresponding regression curves.

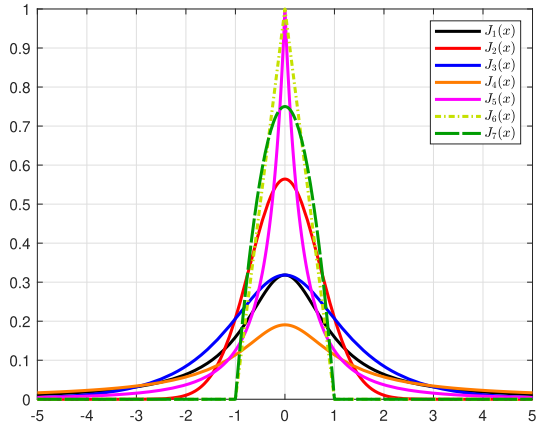


Figure 2: Kernel functions given in Table 2.

Thus, the spreading rate is $\lim_{t\to\infty} \dfrac{h(t)}{t^2} = \dfrac{\sqrt{2}}{1/2}\mu\lambda_4 = \dfrac{2\sqrt{2}\Gamma\left(\frac{3}{4}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)}\mu.$

Numerically, by fitting the parabolic regression equation $PR(t) = c_0t^2 + c_1t + c_2$ to the data for the range of values $T = \{30,\ 35,\ 40,\ 45\}$, we obtain $c_0 = 0.8535$, which approximates the theoretical value $c_0^* = 0.8091$ with high accuracy.

In contrast, kernels $J_2(x)$ and $J_3(x)$ satisfy the first-moment condition (19), meaning their asymptotic behavior is linear. To calculate the spreading rate, we performed a series of simulations with $T = \{5, 10, \ldots, 50\}$, and fitted a linear regression of the form $LR(t) = c_0t + c_1$ with $R^2 = 1$.

Finally, the kernel $J_5(x)$ satisfies the first-moment condition (19), but it does not satisfy the thin-tail condition (20). According to the results in [3], this behavior implies that the difference $c_0t - h(t)$ asymptotically behaves as $\ln t$. Consequently, the regression equation for the asymptotic growth takes the form $LogLR(t) = c_0t + c_1 \ln t$. Based on numerical fitting, we obtain $c_0 = 0.4759$.

In conclusion, our numerical simulations validate the theoretical framework concerning the role of the kernel's first moment. We observed that when the finite first-moment condition (19) is satisfied by the kernel (e.g., for thin-tailed kernels like $J_2(x)$ and $J_3(x)$), the population spreads linearly with a constant asymptotic speed, aligning with theoretical predictions. In contrast, heavy-tailed kernels like $J_1(x)$ and $J_4(x)$ exhibit logarithmic or quadratic spreading rates, reflecting the kernel's influence on the population's dispersal patterns. Figure 1 visually highlights these distinctions.

3.3 Impact of the shape of an initial population

Next, let us examine the impact of the initial conditions.

Example 2 *Let us consider the problem* (1) – (6) *with the following parameters*

$$h(0) = h_0 = 0.7 = -g(0), \quad \mu = 1.5, \quad \alpha^2 = 2, \ f(u) = 2u(1-u), \ J(x) = \frac{1}{\pi(1+x^2)}, \ T = 1, \quad (23)$$

and vary the initial population function $u_0(x)$ *given in Table 3.*

The initial population distributions from Table 3, denoted by $u_0^1(x)$, $u_0^2(x)$ and $u_0^3(x)$, as well as numerical solutions are illustrated in Figures 3 and 4, respectively. A key observation is that,

Table 2: Kernel functions and characteristics.

$J(x)$	$\int_0^\infty xJ(x)\mathrm{d}x$	Thin / Heavy tail	Spreading rate
$J_1(x) = \dfrac{1}{\pi(1+x^2)}$	∞	Heavy	$\lim\limits_{t\to\infty} \dfrac{h(t)}{t\ln t} = \dfrac{\mu}{\pi}$
$J_2(x) = \dfrac{e^{-x^2}}{\sqrt{\pi}}$	$\dfrac{1}{2\sqrt{\pi}}$	Thin	$\lim\limits_{t\to\infty} \dfrac{h(t)}{t} = c_0 = 0.2381$
$J_3(x) = \dfrac{\operatorname{sech} x}{\pi}$	$\dfrac{2G}{\pi}$, $G \approx 0.91597$	Thin	$\lim\limits_{t\to\infty} \dfrac{h(t)}{t} = c_0 = 0.5019$
$J_4(x) = \dfrac{\Gamma\left(\frac{3}{4}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)(1+x^2)^{3/4}}$	∞	Heavy	$\lim\limits_{t\to\infty} \dfrac{h(t)}{t^2} = \dfrac{2\sqrt{2}\Gamma\left(\frac{3}{4}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)}\mu$
$J_5(x) = \dfrac{1}{(1+\lvert x\rvert)^3}$	$\dfrac{1}{2}$	Heavy	$\lim\limits_{t\to\infty} \dfrac{h(t)}{t} = c_0 = 0.4759$

Table 3: Initial population density function and corresponding characteristics.

$u_0(x)$	$P(0) = \int_{-h_0}^{h_0} u_0(x)dx$	Shape characteristics
$u_0(x) = h_0^2 - x^2$	0.4573	Symmetric unimodal
$u_0^1(x) = \cos\left(\pi\dfrac{x}{h_0}\right) + \cos\left(3\pi\dfrac{x}{h_0}\right) + 2$	2.8000	Symmetric multimodal
$u_0^2(x) = \left(\dfrac{x}{h_0}\right)^3 - 1.2\left(\dfrac{x}{h_0}\right)^2 - \left(\dfrac{x}{h_0}\right) + 1.2$	1.1200	Asymmetric unimodal
$u_0^3(x) = -\left(\dfrac{x}{h_0}\right)^4 + 0.4\left(\dfrac{x}{h_0}\right)^3 + 0.71\left(\dfrac{x}{h_0}\right)^2 - 0.4\left(\dfrac{x}{h_0}\right) + 0.29$	0.4573	Asymmetric multimodal

regardless of which of these initial distributions is used, the solution profile eventually evolves into a symmetric, unimodal shape.

4 Conclusion

In this study, we developed and implemented a numerical method for solving the Fisher-KPP equation with nonlocal diffusion and free boundaries. Our approach employs a two-stage front-fixing transformation, effectively handling the complexities of moving boundaries by reformulating the problem onto a fixed computational domain. The resulting transformed system was solved using the second-order accurate Lax-Wendroff scheme for the advection-diffusion terms, coupled with Simpson's rule for precise evaluation of the nonlocal integral.

Numerical results validated the method's ability to accurately and efficiently capture the essential dynamics, including population profiles and boundary movement, for diverse initial conditions and dispersal kernels. Through simulations, we explored the conditions leading to the spreading-vanishing dichotomy, distinguishing between scenarios where the population expands indefinitely and those leading to extinction. Furthermore, we specifically investigated the phenomenon of accelerated spreading associated with heavy-tailed kernels, quantifying how these expansion rates differ from the constant speeds predicted for thin-tailed kernels. These findings underscore the significant

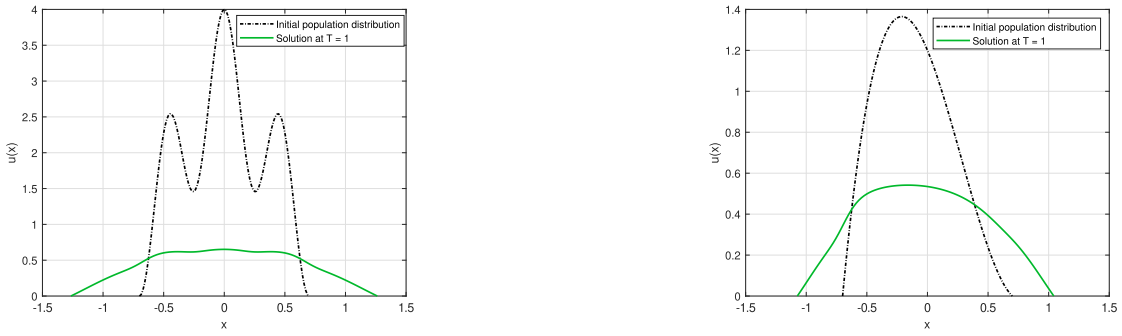


Figure 3: Initial and final population density distribution at $T = 1$ for $u_0^1(x)$ (left) and $u_0^2(x)$ (right).

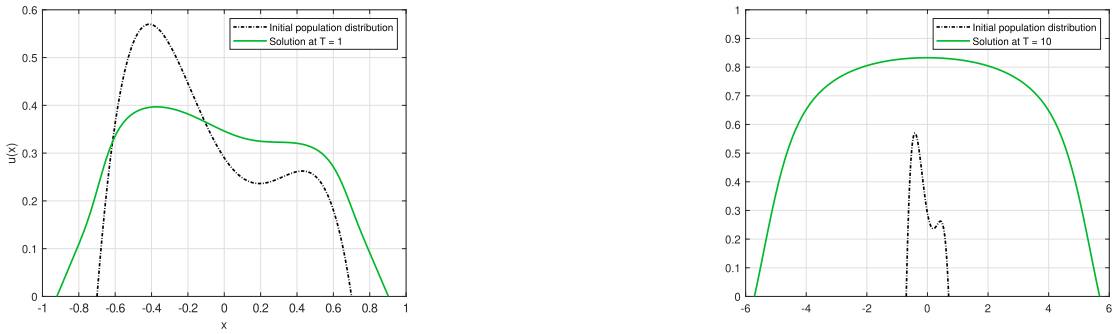


Figure 4: Initial and final population density distribution at $T = \{1,\ 10\}$ for $u_0^3(x)$.

impact of the dispersal kernel's tail properties on the long-term spatial dynamics and propagation speeds within this class of free boundary problems.

References

[1] Cao, J. F., Du, Y., Li, F., Li, W. T. The dynamics of a Fisher-KPP nonlocal diffusion model with free boundaries. *Journal of Functional Analysis*, 277(10):2772–2814, 2019.

[2] Du, Y., Li, F., Zhou, M. Semi-wave and spreading speed of the nonlocal Fisher–KPP equation with free boundaries. *Journal des Mathematiques Pures et Appliquees*, 154(10):30–66, 2021.

[3] Du, Y., Ni, W. Rate of propagation for the Fisher-KPP equation with nonlocal diffusion and free boundaries. *Journal of the European Mathematical Society*, 27(3): 1267–1319, 2025.

[4] Du, Y., Ni, W. Exact rate of accelerated propagation in the Fisher-KPP equation with nonlocal diffusion and free boundaries. *Mathematische Annalen*, 389(7):2931–2958, 2024.

[5] Fletcher, C. A. J. Computational Techniques for Fluid Dynamics 1. Berlin, Springer Berlin Heidelberg, 1998.

Vulnerability and risk of forest ecosystems in the face of global change: Design of an integrated observation and modelling system for adaptation

Miguel A. Zavala[1*], Julián Tijerin[1,2], Marina Rodes-Blanco[2], Mariano García[2], Inmaculada Aguado[2], and Paloma Ruiz Benito[1,2]

(1) Universidad de Alcalá, Departamento de Ciencias de la Vida, Grupo de Ecología y Restauración Forestal (FORECO),
28805 Alcalá de Henares, Madrid, Spain
(2) Universidad de Alcalá, Departamento de Geología, Geografía y Medio Ambiente. Environmental Remote Sensing Research Group (GITA).
Calle Colegios 2, 28801 Alcalá de Henares, Madrid, Spain.

1 Introduction

Iberian Mediterranean forests provide key ecosystem services for society and harbour high biodiversity levels. Yet forest ecosystems are dynamic in nature and are subject to various pressures such as land abandonment and climate change that jeopardize their future stability. Managers require monitoring and modelling tools that allow them to diagnose and anticipate forest responses to these processes so adaptation measures can be implemented. Here we present an integrated approximation of observation and modelling to identify and track global change impacts on forests, and we develop a forest management planning framework for forest adaptation under increasing uncertainty.

2 Methods

Our support decision approximation includes three stages: i) Observation, ii) impact assessment and, iii) adaptation scenarios. i) Observation: We first identify the potential of widely available sensors to detect drought-induced forest decline in Mediterranean forests and quantify the influence of spectral resolution, spatial resolution and forest cover. ii) Impact assessment: We integrate remote sensing, ground observations and modelling to assess previous climate change impacts on productivity. iii) Adaptation scenarios: We use forest dynamics models to explore likely forest management adaptation scenarios to decrease forest vulnerability under increasing aridity.

* ma.zavala@uah.es

3 Results

i) Remote sensing monitoring is a valuable tool for die-back identification but integrated models including remotely sensed variables and ground data are needed to understand impacts and their causes. ii) The negative effects of temperature and drought on forest productivity have intensified over time, with hotter droughts exerting stronger impacts in recent years. Forest productivity evaluated at different spatial scales is declining by drought effects are confounded by structural and legacy effects. Several mechanisms can confer forest resistance and resilience: local adaptation and plasticity, demographic compensation, the importance of stand structure, and the stabilizing role of functional diversity. iii) Dynamical simulation or analytical models are needed to understand the role of these variables on forest response, and thus to identify best adaptation measures.

4 Conclusions

This complexity of social-ecological systems can be best approached by diagnostic and modelling approximations that integrate observations from different scales and levels of organization to define vulnerability and adaptation scenarios. A Resist-Accept-Direct (R.A.D) can assist administration and managers to allocate resources a more cost-effective way and include adaptation into multiple purpose forestry; including conservation, bioeconomy and fire prevention.

Acknowledgements

This study was supported by the National Park Autonomous Agency (OAPN) (Spanish Ministry for the Ecological Transition) under the VERDAT project ('Vulnerabilidad y Riesgo de los ecosistemas de pino silvestre frente al cambio climático: Diseño de un sistema de Alerta Temprana y Seguimiento' Ref: 794/2021).

References

[1] Helfenstein, I. S., Schneider, F. D., Schaepman, M. E., & Morsdorf, F., Assessing biodiversity from space: Impact of spatial and spectral resolution on trait-based functional diversity. Remote Sensing of Environment, 275, 2022.

[2] Lynch, A.J., Thompson, L.M., Beever, E.A., Cole, D.N., Engman, A.C., Hawkins Hoffman, C., Jackson, S.T., Krabbenhoft, T.J., Lawrence, D.J., Limpinsel, D. and Magill, R.T.. Managing for RADical ecosystem change: applying the Resist-Accept-Direct (RAD) framework. Frontiers in Ecology and the Environment, 19(8): 461-469, 2021.

[3] Tijerín-Triviño, J., Lines, E. R., Zavala, M. A., García, M., Astigarraga, J., Cruz-Alonso, V., ... & Ruiz-Benito, P., Forest Productivity Decreases in Response to Recent Changes in Vegetation Structure and Climate in the Latitudinal Extremes of the European Continent. *Global Ecology and Biogeography*, 34(2): e70011, 2025.

[4] Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., Beguería, S., Trigo, R., López-Moreno, J. I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J., & others., Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences*, 110(1): 52–57, 2013.

[5] Zavala, M.A., Angulo, Ó., de la Parra, R.B. et al. Scaling up tree growth to assess forest resilience under increasing aridity: the case of Iberian dry-edge pine forests. Landsc Ecol 39, 6, 2024.

Part XII

Student's Projects

Mathematical modelling of cranial parameters in Ugandan infants: quantitative analysis of normality indices and their implication for infant cranial deformation

A. Marta Muñoz Domingo[1], B. Marc Romeu Ferras[1], C. José Luis Lerna[2] and D. David Vivas Consuelo[1]
(1) Centro de Investigación en Economía y Gestión de la Salud, CIEGS.
Univesitat Politècnica de València, UPV.
Camí de Vera, s/n, Edificio 7J, 3ª planta, Algirós, 46022 València, Valencia.
(2) Grupo de Investigación en Fotogrametría y Láser Escáner, GIFLE.
Camí de Vera, s/n, Edificio 7I, Algirós, 46022 València, Valencia.

1 Introduction

Uganda, a country with a population of approximately 48.7 million, faces significant challenges in its healthcare system, particularly in rural areas, where over 75% of the population resides. The decentralised healthcare structure aims to improve accessibility, yet struggles with a severe shortage of specialised medical personnel, inadequate infrastructure, and limited financial resources. The healthcare system consists of national referral hospitals, regional facilities, and community health centres, with Village Health Teams (VHTs) playing a crucial role in primary healthcare and health education. However, maternal and child health indicators remain concerning, with high infant and maternal mortality rates and limited access to specialised paediatric care.

One of the emerging health concerns in Uganda is the prevalence of cranial deformities in infants, notably positional plagiocephaly. This condition, which affects skull development and symmetry, has increased globally due to the recommendation of supine sleeping to prevent sudden infant death syndrome (SIDS) [1], [2]. Although beneficial, this practice has led to a rise in skull asymmetries, especially in regions with limited access to medical guidance. In low-resource settings like Uganda, differentiating positional plagiocephaly from craniosynostosis — a condition that requires surgical intervention — is challenging due to the lack of specialised diagnostic tools.

This project focuses on developing and implementing accessible screening methods for cranial deformities, using mobile applications based on photogrammetry (e.g., PhotoMeDAS) [3], [4]. These digital tools allow early detection without expensive imaging equipment, making diagnosis more feasible in rural communities. Additionally, educational initiatives targeted at parents and healthcare workers aim to improve prevention strategies, such as proper infant positioning techniques. By integrating low-cost technology and community-driven healthcare

approaches, this initiative seeks to enhance paediatric healthcare accessibility and reduce the incidence of preventable cranial deformities in Uganda.

2 Methods

This study follows a Participatory Action Research (PAR) methodology, ensuring the active involvement of community members and healthcare professionals in preventing and investigating cranial deformation in infants.

2.1 Research and Prevention Design. Process and Materials.

Process mapping and flowcharts were developed to structure and streamline the campaign workflow, ensuring a systematic approach to data collection and intervention strategies.

Materials used: The study employs various technological and educational resources, including high-resolution smartphone cameras, nylon capes with QR codes, anatomical reference adhesive markers, and the PhotoMeDAS software for data capture and analysis. Additionally, educational materials and instructions are provided to families to raise awareness about preventing cranial deformities and the importance of correct positioning in the early months of life.

Data Collection Procedure

1. A patient profile is created in the application, recording basic demographic information.
2. Parents or legal guardians sign an informed consent form for participation.
3. The infant is fitted with a QR-coded cape, securely linking measurements to a unique patient identifier.
4. Three anatomical reference markers are placed: one on the forehead aligned with the eyes and two on the tragus of each ear, allowing precise cranial morphology analysis.
5. High-resolution images are captured using a smartphone camera via the PhotoMeDAS app, automatically scanning and transmitting the data to a central database.
6. The software generates a detailed patient report, analysing cranial indices and comparing them with reference values.
7. In clinically justified cases, follow-up radiographs are performed to confirm diagnoses, prioritising patient safety by minimising unnecessary radiation exposure.

2.2 Study Population

Due to the high cranial plasticity in early life, the study focuses on Ugandan infants from the central region, particularly those aged 0 to 2 years. This age range is crucial for assessing

standard cranial indices and implementing early intervention strategies. Older infants provide valuable natural cranial morphology evolution data within the Ugandan paediatric population. Data collection is conducted in collaboration with local healthcare centres and community medical teams, ensuring a representative sample that reflects the diverse environmental and cultural factors influencing cranial development.

Participants are recruited through medical consultations, home visits, and primary care programs in the central Ugandan region. Parents and guardians receive comprehensive information about the study's objectives, procedures, and the importance of regular follow-ups. Educational materials on infant cranial growth and development are distributed to raise awareness about proper positioning during sleep and breastfeeding to prevent cranial deformities.

Age Ranges: The study employs a dispersion plot using IBM SPSS Statistics to analyse cranial perimeter data by age. The following age ranges are established based on observed growth trends: 0-2 months, 3-6 months, 7-12 months, 13-20 months, and over 20 months (Pindrik et al., 2016).

Inclusion Criteria:
Infants aged 2 years or younger, with slight flexibility to include older children for normality studies. Signed informed consent from legal guardians.

Exclusion Criteria:
Severe congenital malformations that hinder cranial assessment and that could alter study results.

2.3 Descriptive Sample Analysis

Data analysis begins with univariate descriptive statistics using Excel, categorising the sample by sex, age, and ethnic group. Absolute and relative frequencies are calculated for categorical variables, while comparative analysis evaluates sex and ethnic group proportions.

Bar charts and distribution diagrams are used to visualise data distribution, identify potential biases, and exclude gross errors from the dataset. Additionally, SPSS software analyses key morphological variables, providing statistical measures such as mean, confidence intervals, median, variance, standard deviation, skewness, and kurtosis.

2.4 Statistical Sample Analysis

Given the significant age-related differences, data are analysed separately by sex and subsequently within each age range.

Key Variable Analyzed: Head circumference (HC) or cranial perimeter [5], [6] and Vault Asymmetry Index (CVAI) [7], [8], [9].

Statistical Procedure:

- Normality Assessment: Kolmogorov-Smirnov test (for samples >50) or Shapiro-Wilk test (for samples <50) determines data distribution.
- Homogeneity of Variances: Levene's test evaluates variance equality between groups.
- Mean Comparison:
    - If data follows a normal distribution with equal variances: Student's t-test (two groups) or ANOVA (multiple groups).
    - If variances are unequal: Welch's t-test.
    - If data does not follow a normal distribution: Mann-Whitney U test (two groups) or Kruskal-Wallis test (multiple groups).
    - Post-Hoc Analysis: If significant differences exist in ANOVA, Tukey's test identifies specific group differences.

2.5 Cranial Morphology Variability Analysis Using Aggregated Images by Age and Sex

Using medical images, a 3D photometry-based methodology is presented to analyse cranial perimeter variability in Ugandan infants. The two approaches calculate the mean cranial perimeters and standard deviations (±1, ±2, ±3 z-scores) for different age and sex groups, establishing reference patterns based on previous 3D photogrammetry studies [2][3].

For the first approach each skull is represented as a set of 3D points $X_i = \{x_{ij} \in R^3\}$ which are pre-oriented and centred by computing the centroid $c_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$resulting in the centred points $X_i' = \backslash\{x_{ij} - c_i\backslash\}$.

These centred points are projected onto canonical 2D planes, and the convex hull $H\left(\pi(X_i')\right)$ is computed to obtain a polygon with vertices $\{h_{ik}\}$. Each polygon is parametrised in polar coordinates with respect to a global centroid c_g, defining a radial function $r_i(\theta)$ for $\theta \in [0,2\pi)$, such that $c_g + t(cos\theta, sin\theta) \in H\left(\pi(X_i')\right)$. The radial functions $f_i(\theta)$ derived from $r_i(\theta)$ are aggregated across subjects by age and sex. In doing so, a global mean function and the standard deviation is calculated $\overline{f}(\theta) = \frac{1}{N}\sum_{i=1}^{N} f_i(\theta)\,, \sigma(\theta) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(f_i(\theta) - \overline{f}(\theta)\right)^2}$ where N represents the number of subjects.

In the second approach, the focus shifts to identifying the maximum cranial perimeter or maximum head circunference by realigning the ring of points outlining it. Let $X = \{x_i \in R^3 | i =$

$1, \ldots, n\}$ represent the point cloud corresponding to the maximum 3D perimeter. The procedure is to first translate the cloud to a reference point by setting the origin as default $x_i' = x_i - r$

Singular Value Decomposition on the centred data $X' = U\Sigma V^\top$ where the last row of $V^\top$corresponds to the normal vector n of the best fitting plane. To standardise the orientation, is determined the rotation matrix R that aligns n with the vertical unit vector $(0,0,1)^\top$. Define $\theta = arccos(\langle n, (0,0,1)^\top \rangle)$ and $u = \frac{n \times (0,0,1)^\top}{\|n \times (0,0,1)^\top\|}$.

Then, using Rodrigues' rotation formula, the rotation matrix is expressed as $R = I + sin\theta[u]_\times + (1 - cos\theta)[u]_\times^2$, where $[u]_\times$ is the skew-symmetric matrix associated with u.

Finally, the projection and contour extraction is carried out, where the rotation R is applied to the centred points and projects them onto the xy-plane $y_i = Rx_i'$,with$p_i = (y_{i1}, y_{i2})$ and the convex hull computed $H(\backslash\{p_i\backslash\})$ to derive the radial function $f(\theta)$, similar to before. The radial functions obtained here are processed in the same way as in the first approach.

3 Results

195 Ugandan infants aged 0-2 years were analysed, with 101 females (51.79%) and 94 males (48.21%). The study established normative cranial indices for this population, providing reference values for early diagnosis and preventive measures. Cranial perimeter values increased progressively with age, exhibiting statistically significant differences between sexes across all age groups.

- 0-2 months: 366.56 mm (±21.85) in females, 378.88 mm (±18.74) in males.
- 3-6 months: 405.24 mm (±18.37) in females, 416.15 mm (±19.65) in males.
- 7-12 months: 445.67 mm (±15.78) in females, 454.23 mm (±17.49) in males.
- 13-20 months: 476.34 mm (±12.56) in females, 483.87 mm (±14.32) in males.
- >20 months: 497 mm (±11.83) in females, 505.36 mm (±19.74) in males.

Descriptive and Statistical Analysis

The cranial perimeter analysis revealed that the Ugandan infants studied exhibit elongated cranial shapes, with some values falling into the scaphocephaly range, a condition characterised by an increased cephalic length-to-width ratio. These findings suggest that the studied population may present distinct cranial morphology patterns compared to global references.

The CVAI shows a progressive decrease in asymmetry with age.

CVAI values ranged from 5.21% (±2.14) in neonates to 2.87% (±1.32) in infants older than 12 months, indicating a natural tendency toward cranial remodelling.

A detailed comparison of cranial perimeter distributions using the Kolmogorov-Smirnov and Shapiro-Wilk tests confirmed that the data followed a non-normal distribution, necessitating non-parametric statistical tests for group comparisons. Levene's test indicated significant variance differences across age groups, further supporting the hypothesis that cranial growth patterns in this population deviate from standard paediatric norms.

Cranial Morphology Analysis

A 3D photogrammetry-based methodology was implemented to analyse cranial morphology using perimeter extractions. The data were processed using open3d, pandas, numpy, and scipy, allowing for the reconstruction of cranial shape profiles and extracting the mean and the standard deviation across different age groups and sexes.

Aggregated cranial contour images were generated for axial, coronal, and sagittal planes, visually representing cranial development trends.

The analysis revealed that infants in the youngest age groups exhibited the highest variability in cranial perimeter, particularly in the anteroposterior dimension, further supporting the presence of scaphocephaly-like growth tendencies.

The exclusion of outliers based on CVAI thresholds refined the dataset, revealing a clear progression toward cranial symmetry after 12 months.
These computational analyses validated photogrammetry as an effective tool for large-scale cranial assessment, demonstrating high accuracy in perimeter estimation and cranial asymmetry detection.

4 Conclusions

This first project, focused on researching and preventing cranial deformities in children in the Lugazi region, has provided crucial information on normal cranial indices in this paediatric population. Using photogrammetric technology and advanced statistical analysis, key morphological patterns have been established as a reference for future medical interventions and prevention strategies. Smartphones and PhotoMeDAS software enabled efficient data collection, reducing unnecessary radiation exposure and allowing early detection of cranial anomalies. Additionally, standard deviation analysis across different age groups revealed variations between sexes and age ranges, highlighting the need for specific screening and monitoring protocols.

One of the most significant achievements of the campaign was raising awareness among the local community about preventing cranial deformities. Training for parents and caregivers helped promote recommended practices to reduce positional plagiocephaly and other abnormalities. However, challenges such as language barriers required collaboration with local translators to ensure clear communication. Limited resources also hindered thorough follow-up for identified cases, emphasising the need to strengthen the paediatric healthcare infrastructure.

Despite these challenges, the findings lay the foundation for broader screening programs and collaboration with local health institutions to ensure project continuity. Future efforts should focus on integrating cranial measurement technology into diagnostic tools, adapting detection criteria to the Ugandan population, and developing personalised intervention protocols based on identified risk factors.

Acknowledgements

We would like to express our sincere gratitude to the local communities of Lugazi for their participation and collaboration in this project. Special thanks to the healthcare professionals and educators from the Kitega Community Centre, whose support was essential in organising awareness sessions and facilitating data collection. We are also grateful to the Village Health Teams for their active involvement in disseminating knowledge and assisting with communication in different local languages.

Additionally, we thank the team behind PhotoMeDAS for providing a valuable tool for non-invasive cranial measurements. This research would not have been possible without the cooperation of parents and caregivers, who showed great interest in learning preventive strategies for cranial deformities. Lastly, we acknowledge the continuous guidance and support from the academic and research institutions that contributed to the success of this study.

References

[1] Fishbein, M. M., Experts link infant sleep position to SIDS risk: AAP advises placing babies on sides, backs, *AAP news*, 8(5): 1–11, 1992.

[2] Santos, G. and Bernal, G, Sistema de Información Científica, *International Journal of Developmental and Educational Psychology*, 3(1): 59–65, 2012.

[3] Baselga, S., Mora-Navarro, G., and Lerma, J. L., Assessment of Cranial Deformation Indices by Automatic Smartphone-Based Photogrammetric Modelling, *Applied Sciences 2022,* 12(22): 11499, 2022.

[4] Barbero-García, I, Lerma, J. L., and Mora-Navarro, G., Fully automatic smartphone-based photogrammetric 3D modelling of infant's heads for cranial deformation analysis, *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 268–277, 2020.

[5] Poca, M. A. *et al.*, Head circumference: The forgotten tool for hydrocephalus management. A reference interval study in the Spanish population, *Clin Neurol Neurosurg*, 115(11): 2382–2387, 2013.

[6] Martini, M., Klausing, A., Lüchters, G., Heim, N. and Messing-Jünger, M., Head circumference - a useful single parameter for skull volume development in cranial growth analysis?, *Head Face Med*, 14(1), 2018.

[7] Sasaki, J., Hasegawa, S., Yamamoto, S., Watanabe, S., Miyachi, H., and Nagao, T., Relationship between facial asymmetry and positional plagiocephaly analyzed by three-dimensional computed tomography, *Journal of Cranio-Maxillofacial Surgery*, 48(3): 193–198, 2020.

[8] Foster, J. *et al.*, Pediatric cranial deformations: Demographic associations, *J Neurosurg Pediatr*, 26(4): 415–420, 2020.

[9] Loveday, B. P. T. and De Chalain, T. B., Active counterpositioning or orthotic device to treat positional plagiocephaly?, *J Craniofac Surg*, 12(4): 308–313, 2001.

Introduction to Bifurcations in Two-Dimensional Systems of Differential Equations in Artificial Neuron Models

Damián Oussa Vañó Fernández[♭],[1] Xuan Liu[♭] and Antonio Sineiro Abal[♭]

(♭) Universitat Politècnica de València
Camí de Vera s/n, València, Spain.

1 Introduction

The transmission of information in the nervous system is carried out through electrical impulses. In artificial neurons, this transmission is modeled using electrical circuits, which are represented by differential equation systems. Certain bifurcations of the associated dynamical system give rise to oscillatory behaviors that drive the transmission of this information.

The objective of this project is to develop a theoretical and applied introduction to bifurcations in planar systems of two differential equations that may arise in neuronal models, as well as the methods by which they can be computed.

2 Methods

2.1 Hodgkin-Huxley equations

Hodgkin and Huxley [2] were two scientists who, in 1952, published five papers that provided a physical model for the changes in permeability that occur in the excitable membrane of axons during an action potential. The action potential is a signal that is fundamental for the transmission of nerve impulses along neurons and for muscle contraction.

The model proposed by Hodgkin and Huxley [2] consists mainly of a nonlinear differential equation that describes the membrane current. This equation represents a balance between incoming and outgoing currents, as well as the possibility of leakage.

$$I = C_M \frac{dV}{dt} + \bar{g}_{leak}(V - E_{leak}) + \bar{g}_K n^4 (V - E_K) + \bar{g}_{Na} m^3 h (V - E_{Na}) \tag{1}$$

Within the first equation are the parameters n, m, and h, each governed by its own differential equation. These parameters correspond to the leak current and to the currents of Na^+ and K^+ ions, respectively.

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n \tag{2}$$

[1] dovaofer@cam.upv.es

$$\frac{dm}{dt} = \alpha_m(1-m) - \beta_m m \tag{3}$$

$$\frac{dh}{dt} = \alpha_h(1-h) - \beta_h h \tag{4}$$

They are of ohmic nature and can be quantified if the following parameters are known:

$$\alpha_n = 0.01\frac{(V+55)}{1-e^{\frac{-(V+55)}{10}}} \tag{5}$$

$$\beta_n = 0.0555e^{\frac{-V}{80}}\,\alpha_m = \frac{0.1(V+40)}{1-e^{\frac{-V+40}{10}}} \tag{6}$$

$$\beta_m = 0.108e^{\frac{-V}{18}} \tag{7}$$

$$\alpha_h = 0.0027e^{\frac{-V}{20}} \tag{8}$$

$$\beta_h = \frac{1}{1-e^{\frac{-V+35}{10}}} \tag{9}$$

It is observed that the model is a system of four equations. In this project, the authors have focused on studying the bifurcations that commonly appear in neuronal models, but they will work with two-dimensional systems, which are simplified models of the Hodgkin-Huxley model. To this end, the theory explained in the reference [6] has been followed.

A neuron is excitable because its resting state is near a bifurcation, this means it is close to a transition from a resting state to a periodic firing state. These types of bifurcations appear when a "ramp current" is injected, causing the model to enter a state of periodic spike generation.

Transitions from a resting state to an excited state occur through equilibrium bifurcations. Four types of equilibrium bifurcations are identified: saddle-node, saddle-node with a limit cycle, supercritical Andronov-Hopf bifurcation, and subcritical Andronov-Hopf bifurcation. Therefore, the explanation for bifurcations in two-equation systems will focus on these four types.

2.2 Saddle-Node Bifurcation

The saddle-node bifurcation in two-dimensional systems is very similar to that in one-dimensional systems. It is known that this type of bifurcation indicates that equilibrium points are created or destroyed. To explain this type of bifurcation, the following example [3] will be provided:

$$\begin{cases} \dot{x} = \mu - x^2 \\ \dot{y} = -y \end{cases} \tag{10}$$

The trace-determinant method will be used to classify the equilibrium points. From this system, the fixed points must be obtained by setting the derivatives equal to zero and solving for the values of x and y in terms of the parameter μ. Then,

$$\mu - x^2 = 0 \Rightarrow \mu = x^2 \Rightarrow x = \pm\sqrt{\mu}$$

- If $\mu < 0 \Rightarrow \nexists$ equilibrium points
- If $\mu = 0 \Rightarrow x = 0$
- If $\mu > 0 \Rightarrow x = \pm\sqrt{\mu}$.

In addition, $y = 0$. Eventually, the equilibrium points, if $\mu \geq 0$, are: $\left(\sqrt{\mu}, 0\right)$ and $\left(-\sqrt{\mu}, 0\right)$.

The Jacobian matrix is calculated:

$$\begin{bmatrix} -2x & 0 \\ 0 & -1 \end{bmatrix}$$

For $\mu > 0$ there are two fequilibrium points:

- The point $\left(\sqrt{\mu}, 0\right)$ is an attracting node, where the flow along the horizontal and vertical axes directs toward the point.
- The point $(\sqrt{-\mu}, 0)$ is locally a saddle point, where the flow along the horizontal axis directly tends toward it, while along the vertical axis and the rest of the space, the flow is repelled.

This is the flow diagram obtained in Matlab for $\mu = 1$ (Figure 1).

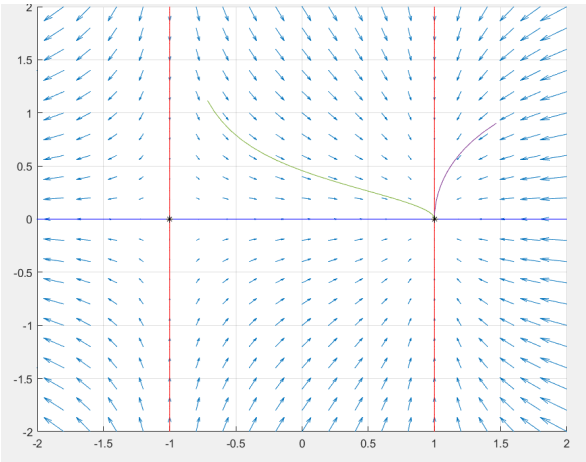


Figure 1: Solution $\mu = 1$

The diagram shows the x-nullclines and the y-nullclines. The two equilibrium points, (-1,0) and (1,0), are also marked with a black asterisk. Two solutions with different initial conditions have been drawn.

For $\mu = 0$, the determinant of the Jacobian matrix is zero, which indicates that the system is degenerate and, therefore, there exist infinitely many fixed points along a line.

When $(\mu = 0)$, there is a triple equilibrium point: $\left(\sqrt{\mu}, 0\right) = \left(-\sqrt{\mu}, 0\right) = (0, 0)$ where the saddle points and the node merge.

The Jacobian matrix at the point $(0, 0)$ is:

$$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

Its corresponding eigenvalues are $\lambda_1 = 0 \quad \text{y} \quad \lambda_2 = -1$.

Since $(\text{Re}(\lambda_1) = 0)$, the point is non-hyperbolic. Its associated eigenvectors are:

$$\lambda_1 = 0 \rightarrow v_1 = (1, 0)$$

As the real part of this eigenvalue is zero, the eigenvector belongs to the center manifold E^c.

$$\lambda_2 = -1 \rightarrow v_2 = (0, 1)$$

Since the real part of this eigenvalue is negative, the eigenvector belongs to the stable manifold E^s. Now, the phase portrait obtained in Matlab is shown (Figure 2).

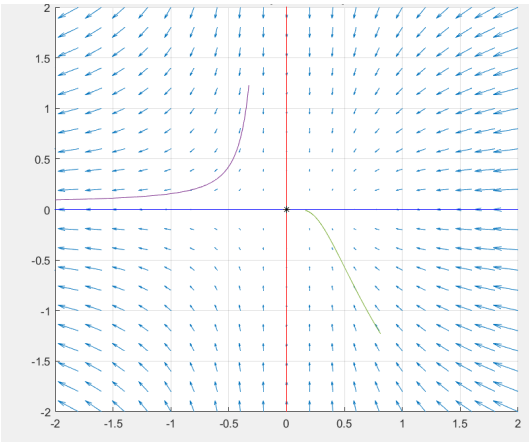


Figure 2: Solution $\mu = 0$

For $\mu < 0$ there are no equilibrium points, but valuable information can still be obtained from the flow diagram (Figure 3), such as the ghost regions.

A "ghost" or bottleneck region [3] is defined as a very small area near (0,0). When the value of the parameter μ is very small (less than 1), it is observed that the flow and the representation of the solution curves passing through this region slow down significantly, as if the fixed point were still there. In the phase diagram, when $\mu = 0$ it is observed that the solutions reach the horizontal axis. In the case of $\mu < 0$, it is seen that as the solutions approach (0, 0), they slow down, but deviate to the left parallel to the horizontal axis without touching it. It is important to note that the size of the ghost region cannot be quantified, but this phenomenon is very useful because if, while testing with phase diagrams, it is observed that some solutions slow down in a certain area, it indicates that a saddle-node bifurcation will occur there. It is also possible to estimate the approximate time required to cross the ghost region [3] :

$$t = \frac{1}{\sqrt{|\mu - \mu_c|}}$$

Where μ_c is the critical parameter, which is the value of the parameter at which the saddle-node bifurcation occurs. In this case, $\mu_c = 0$.

2.3 Saddle-Node Bifurcation with Invariant Circle

The saddle-node bifurcation on an invariant circle occurs in nonlinear dynamical systems when a change in a parameter causes the simultaneous appearance or disappearance of equilibrium points

of node and saddle type in an invariant circle or limit cycle. This invariant circle is a closed orbit that separates regions of phase space with qualitatively different behaviors. When a critical parameter value is crossed, the node and saddle points emerge on or are annihilated along it. This type of bifurcation is key to understanding transitions between periodic orbits and heteroclinic trajectories in two-dimensional systems.

An example of the Saddle-Node bifurcation on an invariant circle or limit cycle is now presented [5]. This type of bifurcation is commonly referred as a 'Saddle-Node Infinite Period Bifurcation on a Cycle,' or SNIPER.

$$\begin{cases} \dot{r} = r(1 - r^2) \\ \dot{\theta} = \mu - \sin(\theta) \end{cases} \tag{11}$$

This two-dimensional system is a decoupled system. To observe the saddle-node bifurcation on the invariant circle, it will be assumed that $\mu > 0$. At the phase diagram there is a limit cycle with unit radius and center at $(0, 0)$. It is known that equation $\dot{\theta} = \mu - \sin(\theta)$ is a sinusoidal equation with unit amplitude. In $\mathbb{R}^2$ the graph of $\dot{\theta}$ will be plotted, where θ is on the horizontal axis and $\dot{\theta}$ on the vertical axis. The domain of the horizontal axis is $(0, 2\pi)$ due to the polar coordinates. Now we plot the graph for $\mu = 1$.

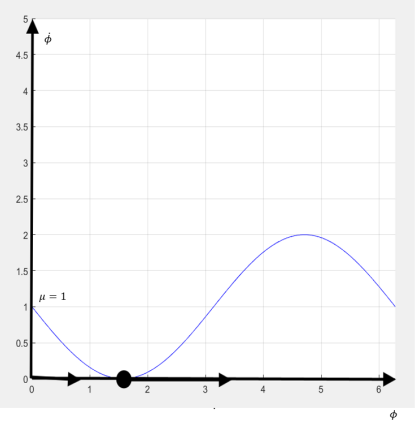


Figure 3: Solution $\mu = 1$

An intersection between the curve and the horizontal axis is obtained at $\theta = \frac{\pi}{2}$, which corresponds to a saddle point. This is consistent with the discussion of the saddle-node bifurcation, as the flow slows down in that region. At the saddle point, the flow will pass through it, but along the rest of the invariant circle, the flow will be faster.

The next case occurs when the curve is no longer tangent to the axis and intersects it, giving rise to two fixed points, a node and a saddle, for $0 < \mu < 1$. The graph is plotted for $\mu = 0.5$. Two intersections between the curve and the horizontal axis are obtained, where the left point corresponds to an attractor and the right to a repeller. In the phase diagram, these two points appear on the invariant circle and originate from the north pole at $\theta = \frac{\pi}{2}$. The attractor moves to the right along the circle, while in the 2D system, the unstable point is a saddle that moves to the left. At both the node and the saddle, since polar coordinates are being used, the radial line originating at $(0, 0)$ and passing through the two points (i.e., all points in space sharing the

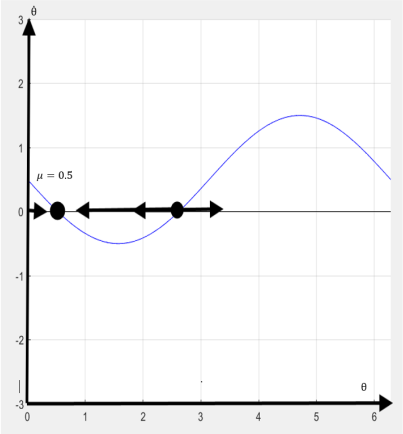


Figure 4: Solution $\mu = 0.5$

same angle as the stable point, but with varying radius) causes the flow passing through them to move directly toward the point. These two lines correspond to the stable manifolds of the two fixed points.The following phase diagrams were generated using Mathematica.

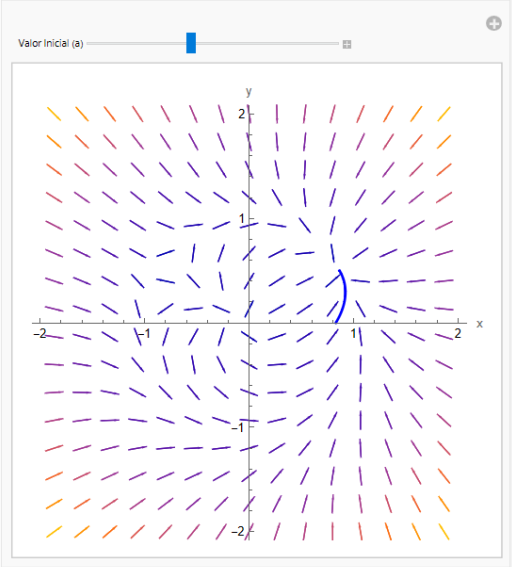


Figure 5: Solution $\mu = 0.5$

2.4 Hopf Bifurcation

The following example of a dynamical system is provided [4].

$$\begin{cases} \dot{x} = \mu x - y + \alpha x \left(x^2 + y^2\right) \\ \dot{y} = \mu y + x + \alpha y \left(x^2 + y^2\right) \end{cases} \tag{12}$$

Which, after applying the coordinate transformation, becomes:

$$r^2 = x^2 + y^2 \Rightarrow \dot{r} = \frac{x\dot{x}+y\dot{y}}{r} \Rightarrow$$
$$\dot{r} = \frac{x\left(\mu x - y + \alpha x\left(x^2+y^2\right)\right) + y\left(\mu y + x + \alpha y\left(x^2+y^2\right)\right)}{r} =$$
$$= \frac{\mu x^2 + \alpha x^2\left(x^2+y^2\right) + \mu y^2 + \alpha y^2\left(x^2+y^2\right)}{r} =$$
$$= \frac{\mu r^2 + \alpha x^2\left(r^2\right) + \alpha y^2\left(r^2\right)}{r} = \mu r + \alpha r^3$$

$$tan(\theta) = \frac{y}{x} \Rightarrow (1 + tan^2(\theta))\dot{\theta} = \frac{\dot{y}x + \dot{x}y}{x^2} \Rightarrow$$
$$\dot{\theta} = \frac{\dot{y}x - \dot{x}y}{x^2\left(1+\frac{y^2}{x^2}\right)} = \frac{\dot{y}x - \dot{x}y}{r}$$
$$= \frac{\left(\mu y + x + \alpha y\left(x^2+y^2\right)\right)x - \left(\mu x - y + \alpha x\left(x^2+y^2\right)\right)y}{r} =$$
$$= \frac{x^2+y^2}{r} = 1$$

Therefore, the resulting system is:

$$\begin{cases} \dot{r} = \mu r + \alpha r^3 \\ \dot{\theta} = 1 \end{cases} \tag{13}$$

That $\dot{\theta}$ equals 1 indicates that the angular motion is uniform with period $T = 2\pi$. Depending on the value of μ, the orbits will be subcritical or supercritical.

2.5 Neuronal Model with Hopf Bifurcation

In this section, a neuronal model will be studied, and the values of the parameter that produce a Hopf bifurcation will be sought, as it will lead to a periodic voltage in the artificial neuron. The starting point is the initial neuronal model from the article "Iontronic Nanopore Model for Artificial Neurons: The Requisites of Spiking" by Bisquert [1].

$$\left\{ \begin{array}{c} \frac{du}{dt} = \frac{I}{C_m} - \frac{f(u,p)}{C_m} \\ \frac{dp}{dt} = \frac{g(u,p)}{\tau_k} \end{array} \right\}$$

The system is expressed in matrix form:

$$\begin{bmatrix} u' \\ p' \end{bmatrix} = \begin{bmatrix} \frac{-1}{C_m} f_u\left(u_1, p_1\right) & \frac{-1}{C_m} f_p\left(u_1, p_1\right) \\ \frac{1}{\tau_k} g_u\left(u_1, p_1\right) & \frac{1}{\tau_k} g_p\left(u_1, p_1\right) \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix}$$

Where,

$$F' = \begin{bmatrix} \frac{-1}{C_m} f_u\left(u_1, p_1\right) & \frac{-1}{C_m} f_p\left(u_1, p_1\right) \\ \frac{1}{\tau_k} g_u\left(u_1, p_1\right) & \frac{1}{\tau_k} g_p\left(u_1, p_1\right) \end{bmatrix}$$

Using the trace-determinant theorem for the classification of equilibrium points (used in the previous examples), the following expressions for the trace and the determinant are obtained:

$$\operatorname{tr}\left(F'\right) = \frac{-1}{C_m} f_u\left(u_1, p_1\right) + \frac{1}{\tau_k} g_p\left(u_1, p_1\right)$$

$$f_u\left(u, p\right) = G\left(p\right)\left(F'\left(u\right)\left(u - E_p\right) + F\left(u\right)\right) + \frac{1}{R_L}$$

$$g_p\left(u, p\right) = -d_0$$

$$F'\left(u\right) = \frac{1}{2V_T\left(1 + \cosh\left(\frac{u - V_F}{V_T}\right)\right)} = \frac{e^{\left(\frac{u+V_F}{V_T}\right)}}{V_T\left(e^{\left(\frac{u}{V_T}\right)} + e^{\left(\frac{V_F}{V_T}\right)}\right)^2}$$

$$G(p) = \frac{G_0}{1 + \frac{p}{p_0}} = \frac{G_0}{1 + \frac{r_0}{d_0 p_0} K(u)}$$

$$\det\left(F'\right) = \frac{-1}{C_m} f_u(u, p_{ss}) \left(\frac{-d_0}{\tau_k}\right) + \frac{1}{C_m} f_p(u, p_{ss}) \frac{1}{\tau_k} g_u(u, p_{ss})$$

$$K'(u) = \frac{1}{2V_T\left(1 + \cosh\left(\frac{u - V_K}{V_T}\right)\right)}$$

$$f_p(u, p_{ss}) = \frac{-G_o P_o}{(rK(u) + P_o)^2} F(u)(u - E_p)$$

The parameter values used in Figure 2(d) of [1] are applied and using Mathematica, the values of I1 = 0.0092989 and I2 = 0.0169945 are obtained. With these solutions, the values of the current in the artificial neurons have been found, where the current would exhibit a steady-state behavior, and the transmission of potentials in real neurons could be simulated.

3 Conclusions

With this work, it has been possible to deepen the study of two-dimensional systems of partial differential equations. An introduction to saddle-node bifurcations, saddle-node on invariant circles, and Hopf bifurcations has been made. Through mathematical modeling and the analysis of Hopf bifurcations, it is possible to simulate more realistic neural networks and predict the behavior of neurons in response to various stimuli with greater accuracy. Furthermore, the potential of these techniques opens the door to the creation of artificial neural models that could, in the future, emulate the functionality of the human brain, offering innovative applications in both neuroscience and the development of neuroinformatics technologies. In conclusion, this work reaffirms that mathematics continues to be a fundamental tool in unraveling the mysteries of the brain and, potentially, in creating artificial and functional neural models.

References

[1] Bisquert, J., Iontronic Nanopore Model for Artificial Neurons: The Requisites of Spiking *The Journal of Physical Chemistry Letters*, 14(40):9027–9033, 2023.

[2] Hopper, A. J., Beswick-Jones, H., Brown, A. M., A color-coded graphical guide to the Hodgkin and Huxley papers *Advances in Physiology Education*, 46(4):580–592, 2022.

[3] Koon, W. S., Lo, M. W., Marsden, J. E., Ross, S. D., Dynamical Systems, the Three Body Problem and Space Mission Design. Marsden Books, 2022.

[4] García Soria, A., Atravesando lentamente una bifurcación de Hopf. Oviedo, Universidad de Oviedo, 2023.

[5] Bramburger, J. J., Data-Driven Methods for Dynamic Systems. Philadelphia, Society for Industrial and Applied Mathematics, 2024.

[6] L. Perko., Differential equations and dynamical sistems. Springer, 2001.

Machine Learning-Based Prediction of Percutaneous Nephrolithotomy Outcomes in a Large General Hospital

Laura Shalabayeva[1], Pilar Bahílo Mateu[2] , Marc Romeu Ferras [1], Javier Díaz Carnicero[1], Alberto Budía[2] and David Vivas-Consuelo [1]

(1) Centro de Investigación en Economía y Gestión de la Salud, CIEGS.
Univesitat Politècnica de València, UPV.
Camí de Vera, s/n, Edificio 7J, 3ª planta, Algirós, 46022 València, Valencia.
(2) Unidad de Litotricia y Endourología,
Av. Fernando Abril Martorell, nº 106, 46026 València, Valencia.

1 Introduction

Nephrolithiasis is a disease of the urinary system that causes the formation of kidney stones, which are solid deposits of substances present in urine, such as calcium, oxalate, uric acid, or phosphate. Patients with this condition may experience symptoms such as dizziness, sweating, nausea, vomiting, and colic, which refers to intense pain caused by the movement of the stones [1]. According to statistics, more than 12% of the population will experience colic at some point in their life, and the recurrence rate within the 5 years following the episode of colic is 50% [2].

One of the possible treatments of nephrolithiasis is Percutaneous nephrolithotomy (PCNL), which is especially recommended for patients with stone size larger than 20 mm and stag-horn calculi [3]. Besides being minimally invasive surgery, in some cases PCNL can result in some complications like bleeding, infection, sepsis or need of transfusion [4]. Thus, to be able to identify patients who would face complications and require a hospitalization stay, the objective of development of a Machine Learning (ML) model is established.

2 Methods

2.1 Patients and data set

A database composed of 901 patients who underwent PCNL treatment between March 2011 and December 2024 in the Department of Lithotripsy at Hospital Fe, Valencia, Spain was used for the study. Among them, 171 patients had complete data and 223 patients were labelled as ones that require hospitalization.

lshalab@upv.edu.es

2.2 Predictors and outcomes

The following predictive variables were analyzed: sex, diabetes, age, body mass index (BMI), ASA, culture result, microorganism, treatment type, side, major diameter, minor diameter, number, location, Guy's score, Hounsfield Units (UH), distance, position, access, catheterization, contrast, dilation method, multi-trajectory, sheath caliber, fragmentation source, localization ease, duration, drains, tubeless, procalcitonin, leukocytes, %neutrophils. And the outcome of need of hospitalization in base of possible complications was examined.

2.3 Missing data

Several options for treatment of missing data were evaluated: imputation by median value and moda for numerical and categorical variables, respectively ; imputation by means of Multiple Imputation by Chained Equations (MICE). Cases with undetermined label were excluded from the study.

2.4 Model development and evaluation

A series of different models like Logistic Regression, Decision Tree, Random Forest and Extreme Gradient Boosting were built via Python programming language. For evaluation of the models' performance the dataset was split into training, validation and test sets in proportion 70/15/15 and the following metrics were extracted from validation and test sets: Accuracy, Sensitivity, Specificity and Area Under the Curve (AUC).

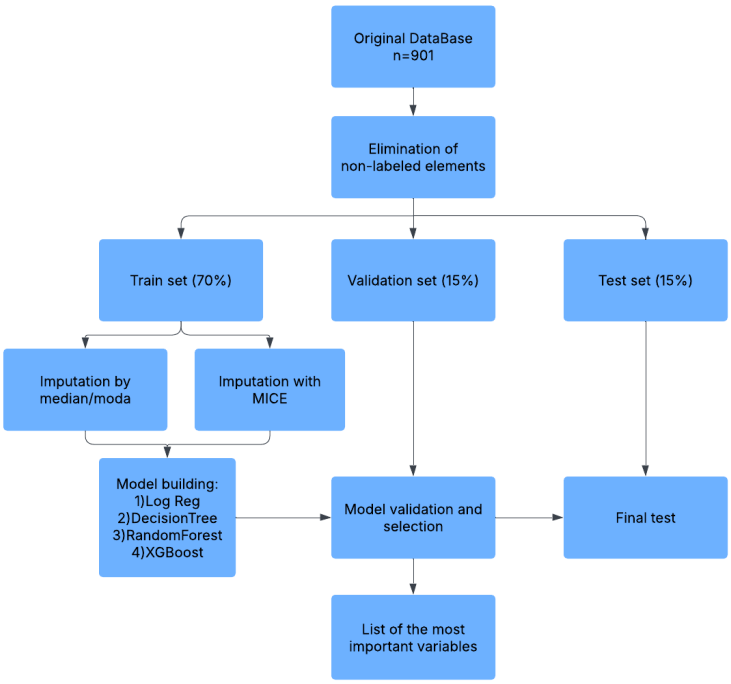


Figure 1. Flowchart of the study.

2.5 Statistical analysis

A statistical analysis of the data was performed by means of SPSS Software and Python programming language. In case of continuous variables, t-tests and the Mann–Whitney U test were used for normally and non-normally distributed variables, respectively, while for analysis of categorical variables Chi-square tests were applied.

3 Results

3.1 Demographics

The database was composed by 48,1% males and 51,9% women, whose overall mean age and BMI were 54,5 ± 15,4 years and 27,6 ± 5,8 kg/m², respectively. The proportion of patients with and without need of hospitalization was unbalanced , more specifically, 24,9% patients were labelled as ones that require hospitalization stay. The analysis of the rest of characteristics can be observed in Table 1 for cathegorical variables and in Table 2 for numerical variables.

Variable	Frequency	Percentage	Variable	Frequency	Percentage
Sex			**Access**		
Male	432	48,1	Fluoroscopy	24	2,7
Female	466	51,9	Ultrasound	9	1,0
Diabetes			Both	855	96,2
No	757	84,3	**Catheterization**		
Type 1	35	3,9	No	101	11,3
Type 2	106	11,8	Yes	793	88,7
ASA			**Contrast**		
ASA I	205	22,8	No	107	12,4
ASA II	545	60,7	Yes	755	87,6
ASA III	130	14,5	**Dilation method**		
ASA IV	18	2,0	Amplatz	68	7,7
Culture result			Dilation balloon	396	44,6
Negative	556	65,3	Both	394	44,4
Positive	287	33,7	Metallic	29	3,3
Contaminated	9	1,1	**Multi-trajectory**		
Microorganism			No	837	93,9
Negative	537	66,1	Yes	54	6,1
Urease negative	166	20,4	**Fragmentation source**		

Urease positive	109	13,4	Lithoclast	524	59,5
Treatment type			Ultrasound	3	,3
Primary	693	77,2	Lithoclast + ultrasound	5	,6
Post-ESWL	75	8,4	Holmium laser	148	16,8
Post-URS	52	5,8	Basket	25	2,8
Post-NLP	78	8,7	Fórceps	43	4,9
Side			Lithoclast + laser	128	14,5
Right	430	47,9	Irrigation	4	,5
Left	462	51,4	**Localization ease**		
Bilateral	4	,4	Easy	386	43,4
Transplant	2	,2	Medium	373	41,9
Location			Difficult	131	14,7
Pelvis	234	26,2	**Drains**		
Superior calyx	27	3,0	No	8	,9
Middle calyx	5	,6	Nephrostomy	159	17,8
Inferior calyx	84	9,4	Double-J stent	330	36,9
Pseudocoraliform	103	11,5	Double-J stent + nephrostomy	381	42,6
Coraliform	235	26,3	24h ureteral catheter	9	1,0
Renal and ureteral	26	2,9	24h ureteral catheter + nephrostomy	8	,9
Calyceal diverticulum	8	,9	**Tubeless**		
Pelvis + upper calyceal group	107	12,0	No	545	60,9
Multiple calyces	32	3,6	Yes	350	39,1
Proximal ureter	8	,9	**Procalcitonin**		
Distal ureter	6	,7	Normal range	800	88,8
Pelvis + lower calyceal group	19	2,1	Out of normal range	101	11,2
Guy's score			**Leukocytes**		
Guy Score I	218	24,5	Normal range	769	85,3
Guy Score II	283	31,9	Out of normal range	132	14,7
Guy Score III	230	25,9	**%neutrophils**		
Guy Score IV	157	17,7	Normal range	461	51,2
Position			Out of normal range	440	48,8
Prone	11	1,2	**Need of hospitalization**		
Supine	881	98,8	No	671	75,1
			Yes	223	24,9

Table 1. Frequencies of cathegorical variables.

	Mean ± SD
BMI	27,6 ± 5,8
major diameter	25 ± 13,3
minor diameter	15,2 ± 9,3
UH	1010,7 ± 347
distance	28,4 ± 36,9
sheath caliber	25,3 ± 3,8
duration	107,9 ± 38,9
age	54,5 ± 15,4

Table 2. Numerical characteristics of the patients.

3.2 Model selection

According to Table 3 and Table 4, it can be observed that models that used the database with missing data imputed by mean/moda and by MICE have similar metrics. Besides, among 4 studied models (Logistic Regression, Decision Tree, Random Forest and XGBoost), Random Forest models demonstrated the best behaviour with AUC larger than 0.7 in all cases. Moreover, Random Forest model trained with the mean/moda imputed data is considered the most sensitive and, thus, the most recommended for the established objective. The top ten most important factors for the model predictions are neutrophil percentage, use of a nephrostomy tube for urinary drainage, major diameter, patient's BMI and age, UH, culture result, number of stones, Guy's score, which indicates complexity of stone, and distance between patients skin and the stone.

MODEL	Subset	AUC	Accuracy	Sensitivity	Specificity
Logistic Regression	Validation	0.68	0.66	0.50	0.72
	Test	0.68	0.72	0.62	0.75
Decision Tree	Validation	0.67	0.75	0.29	0.94
	Test	0.59	0.73	0.19	0.91
Random Forest	Validation	0.73	0.71	0.61	0.75
	Test	0.74	0.74	0.62	0.78
XGBoost	Validation	0.73	0.71	0.07	0.97
	Test	0.66	0.75	0.10	0.97

Table 3. The metrics for different models, where Database missing data was imputed by mean/moda.

MODEL	Subset	AUC	Accuracy	Sensitivity	Specificity
Logistic Regression	Validation	0.70	0.70	0.54	0.76
	Test	0.65	0.68	0.43	0.77
Decision Tree	Validation	0.68	0.75	0.29	0.94
	Test	0.59	0.73	0.19	0.91
Random Forest	Validation	0.75	0.76	0.57	0.84
	Test	0.71	0.72	0.52	0.78
XGBoost	Validation	0.73	0.70	0.07	0.96
	Test	0.66	0.75	0.05	0.98

Table 4. The metrics for different models, where Database missing data was imputed by MICE.

4 Conclusions

Although final validation is ongoing, initial findings support the potential utility of machine learning techniques in assisting clinical decision-making for patients undergoing PCNL. By identifying high-risk patients, this model could contribute to more personalized management strategies, potentially reducing complications and optimizing hospital resource use. Further refinement and prospective validation of the model are necessary before clinical implementation.

Acknowledgements

We would like to thank the healthcare professionals from Department of Lithotripsy at Hospital Fe in Valencia for their continuous support, collaboration and data collection.

References

[1] Liangos, O., & Jaber, B. L., Kidney Stones. In *Humana Press eBooks*: 513–527, 2008. https://doi.org/10.1007/978-1-59745-032-4_21

[2] Del Pilar Alcoba García, M., Serrano, G. B., Jiménez, J. T., López, R. G., Ruíz, L. C., & Enguita, C. G., Is extracorporeal shock wave lithotripsy a treatment option for renal colic? *Archivos Españoles De Urología*, *76*(3): 175, 2023. https://doi.org/10.56434/j.arch.esp.urol.20237603.20

[3] Ghani, K. R., Andonian, S., Bultitude, M., Desai, M., Giusti, G., Okhunov, Z., Preminger, G. M., & De La Rosette, J., Percutaneous nephrolithotomy: Update, trends, and future directions. *European Urology*, *70*(2): 382–396, 2016. https://doi.org/10.1016/j.eururo.2016.01.047

[4] Baltar, C. F., Corral, M. E. M., Fentes, D. P., Predicting and Avoiding complications in percutaneous nephrolithotomy in the Era of Personalized Medicine: A scoping review. *Journal of Personalized Medicine*, *14*(9): 962, 2024 https://doi.org/10.3390/jpm14090962