Application of an extension of the random Verhulst model

J.-C. Cortés[♭] A. Navarro-Quiles[♮] and S.-M. Sferle[♭],[1]

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

(♮) Departamento de Estadística e Investigación Operativa, Universitat de València
Dr. Moliner, 50, Burjassot, Spain.

1 Introduction

Differential equations serve as an essential mathematical tool for modeling a wide range of dynamical systems, offering significant insights into their progression and underlying mechanics. Our focus lies on equations that model systems related to growth patterns, particularly those whose solutions generate sigmoidal-type curves. The study of these patterns is important in several fields, including population biology (understanding the growth of populations organisms), epidemiology (modeling the spread of diseases), economics and business (analyzing market penetration and adoption rates of new products), technology and innovation (studying the life cycle of technological innovations), ecology (understanding the carrying capacity of ecosystems), sociology (analyzing the spread of information, behaviors or cultural phenomena), etc. The aforementioned curve followed by these applications encompasses three phases: a lag phase marked by slow or negligible growth, an exponential growth phase and a plateau phase. In the latter, growth stabilizes when the system reaches its maximum potential under existing conditions.

The pioneer in this field is Verhulst, who proposed the logistic equation that responds to the characteristics just described. His model describes how populations grow when limited resources are taken into account. Due to some limitations of his model, other authors such as Richards, von-Bertalanffy, Gompertz or Blumberg have proposed extensions and modifications of the model to overcome them [1]. These offer greater flexibility, enabling them to more precisely fit a variety of real-world growth patterns.

In this work we will focus on an extension of the logistic equation, called by Turner the hyper-logistic equation [2]. The particularity of this equation, with respect to the Verhulst model, lies in the introduction of the additional parameter, p. This makes it more flexible and adaptable to various growth scenarios. It is formulated by the following differential equation

$$\begin{cases} n'(t) = \frac{r}{K} n(t)^{1-p}(K - n(t))^{1+p}, \\ n(t_0) = n_0. \end{cases} \tag{1}$$

The parameters in the equation (1) are defined as follows: $r > 0$ is the growth rate constant, $K > 0$ is the carrying capacity, and $\frac{r}{K}$ forms the proportionality constant. The shape parameter is

[1] smsferle@doctor.upv.es

within the range $0 < p < 1$. Additionally, the function $n(t)$ represents the quantity of interest at time t, with $n_0 \geq 0$ denoting the initial quantity at initial time $t_0 \geq 0$, and $n'(t)$ representing the rate of change of n with respect to t.

At this point, it is important to note that when this model is applied to the real world, uncertainty comes into play. This can come from several sources, such as the intrinsic randomness of the system itself and its fluctuations from the environment, measurement errors such as in the collection of data on the quantity of interest, or factors that influence the different phases of the sigmoidal curve. To address this problem, one way would be to introduce uncertainty directly into the model through the inputs, considering them as random variables. This fact gives rise, therefore, to the following random differential equation, which would be the randomization of (1)

$$\begin{cases} n'(t,\omega) = \frac{r(\omega)}{K(\omega)} n(t,\omega)^{1-p(\omega)} (K(\omega) - n(t,\omega))^{1+p(\omega)}, \\ n(t_0) = n_0(\omega), \end{cases} \tag{2}$$

where $r(\omega)$, $K(\omega)$, $p(\omega)$ and $n_0(\omega)$ are absolutely continuous random variables defined on a common complete probability space $(\Omega, F_\Omega, \mathbb{P})$, with a known joint probability density function (PDF) denoted as $f_0(r, K, p, n_0)$. Thus, the uncertainty and variability of real-world conditions are incorporated into the model, making the equation (2) a more realistic representation of the growth scenario compared to the deterministic (1). Hence, its solution now represents a stochastic process of the following form

$$n(t,\omega) = K(\omega) - \frac{K(\omega)}{1 + \left(p(\omega) r(\omega)(t - t_0) + \left(\frac{K(\omega)}{n_0(\omega)} - 1 \right)^{-p(\omega)} \right)^{\frac{1}{p(\omega)}}}. \tag{3}$$

One of the objectives of this study is to conduct a thorough probabilistic analysis of the solution (3) for each time. This will involve determining the first probability density function (1-PDF) and analyzing its main statistical characteristics such as mean, variance and probabilistic intervals.

While the theoretical results to be obtained can be applied to various growth processes, such as the growth of multicellular tumor spheroids [3, 4], our other objective in this study is different. We intend to apply our model to a completely different context: cumulative data (reflect the accumulation of events or quantities over time). Cumulative data usually present a sigmoidal shape, similar to the solutions of our model. By fitting our model to these data, we aim to assess its ability to capture the underlying trend, demonstrating its wider applicability beyond traditional growth processes.

2 Methods

In order to address my first objective, which is to determine the 1-PDF, we will employ the Random Variable Transformation (RVT) technique. The theorem on which this technique is based is as follows, as can be found in [5].

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 Jacobian, defined as*

$$J_n = \det \begin{pmatrix} \frac{\partial s_1(\mathbf{y})}{\partial Y_1} & \cdots & \frac{\partial s_n(\mathbf{y})}{\partial Y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_1(\mathbf{y})}{\partial Y_n} & \cdots & \frac{\partial s_n(\mathbf{y})}{\partial Y_n} \end{pmatrix}.$$

Upon setting all the identifications of the Theorem 1, we obtain the following analytical expression for the 1-PDF of the solution (3).

$$f_1(n,t) = \int_{n_0}^{\infty} \int_0^1 \int_0^K f_0 \left(\frac{\left(\frac{K}{n} - 1\right)^{-p} - \left(\frac{K}{n_0} - 1\right)^{-p}}{p(t-t_0)}, K, p, n_0 \right) \frac{K}{n^2(t-t_0)} \left(\frac{K}{n} - 1 \right)^{-p-1} \mathrm{d}n_0 \, \mathrm{d}p \, \mathrm{d}K. \tag{4}$$

By definition, the carrying capacity K must be greater than the initial quantity n_0, $K > n_0 \geq 0$, since K is the upper limit that the quantity of interest can reach under given conditions.

To apply the theoretical results in practice, it is essential to assign an appropriate probability distribution to the joint PDF f_0 in order to accurately capture the uncertainty in the data. Although there are several methods to achieve this, in this paper we have chosen to use a Bayesian approach for this purpose. The PDF we are interested in will be obtained using Bayes' theorem [6], where f_0 represents the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{X})$

$$p(\boldsymbol{\theta}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\theta}$ represents the parameter vector (r, K, p, n_0) with the associated prior PDF $p(\boldsymbol{\theta})$, and $\boldsymbol{X}$ represents the random vector of the observed data with the likelihood function $p(\boldsymbol{X}|\boldsymbol{\theta})$.

In practice, the posterior distribution is estimated using the Gibbs sampling algorithm, which is a method within the Monte Carlo Markov Chains (MCMC) framework (see [7]). Based on the knowledge obtained from the available literature on our model, we have chosen a Normal distribution for the likelihood function, and for the prior distributions of the model parameters, we have opted for non-informative prior distributions, namely Uniform distributions.

Once the posterior distribution has been estimated, subsequent convergence analyses are performed to ensure that the Markov chain has sufficiently explored the parameter space and that the samples obtained are representative of the true posterior distribution. With the joint distribution f_0 now established, we can determine the 1-PDF and, subsequently, its statistical characteristics, including the mean, standard deviation and 95% probabilistic intervals (PI's)

$$\begin{gathered} \mathbb{E}[n(t,\omega)] = \int_{-\infty}^{\infty} n f_1(n,t) \, \mathrm{d}n, \quad \sigma_n(t) = \sqrt{\int_{-\infty}^{\infty} (n - \mathbb{E}[n(t,\omega)])^2 f_1(n,t) \, \mathrm{d}n}, \\ \mathbb{P}(\{\omega \in \Omega \ / \ n(\hat{t},\omega) \in [n_1(\hat{t}), n_2(\hat{t})]\}) \Rightarrow \int_0^{n_1(\hat{t})} f_1(n,\hat{t}) \, \mathrm{d}n = \frac{0.05}{2} = \int_{n_2(\hat{t})}^{K} f_1(n,\hat{t}) \, \mathrm{d}n, \end{gathered} \tag{6}$$

where $\hat{t} \geq 0$ represents an arbitrary fixed time.

Finally, since our intention is to apply the model to a completely different case of a typical growth process, we use the posterior predictive distribution for new data $\tilde{\boldsymbol{X}}_{new}$ given observed data $\boldsymbol{X}$, i.e. $p(\tilde{\boldsymbol{X}}_{new}|\boldsymbol{X})$, to assess whether the model plausibly explains the observed data, as follows

$$p(\tilde{\boldsymbol{X}}_{new}|\boldsymbol{X}) = \int p(\tilde{\boldsymbol{X}}_{new}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{X})d\boldsymbol{\theta}, \tag{7}$$

where $p(\tilde{\boldsymbol{X}}_{new}|\boldsymbol{\theta})$ is the likelihood of the new data given the parameters, and $p(\boldsymbol{\theta}|\boldsymbol{X})$ is the posterior distribution of the parameters given the observed data. The integral is taken over the entire parameter space.

3 Results

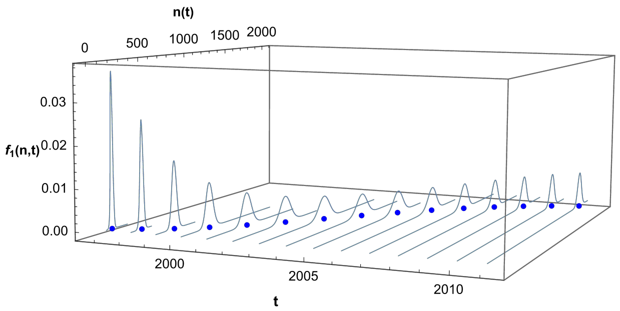
In this part, we apply the theoretical results from the previous section to real-world data. Specifically, we analyze the cumulative data on the number of annual publications of Kucharavy's and De Guio's work [8]. This dataset includes the total number of publications each year during the period 1996 – 2011 in the TRIZ journal, TRIZCON conferences, and ETRIA conferences. The data are summarized in Table 1.

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Publications	5	37	107	201	298	439	617	773	936	1121	1316	1451	1563	1629	1722	1788

Table 1: Cumulative number of TRIZ publications per year from 1996 to 2011.

After obtaining the joint PDF f_0 using the Gibbs sampling algorithm within the Bayesian framework described in equation (5), we can determine the 1-PDF defined in equation (4), the plot of which is shown in the left panel of Figure 1. It is evident that for each time the PDF is systematically centered around the recorded publication number, resembling the shape of a Normal distribution.

This behavior is most clearly illustrated in the right panel of Figure 1, where the observed data are juxtaposed to the mean of the 1-PDF, representing the probabilistic fit. The evolution of this fit over time shows a stabilization pattern characteristic of an S-shape, while the 95% probabilistic intervals effectively capture the uncertainties in the data.



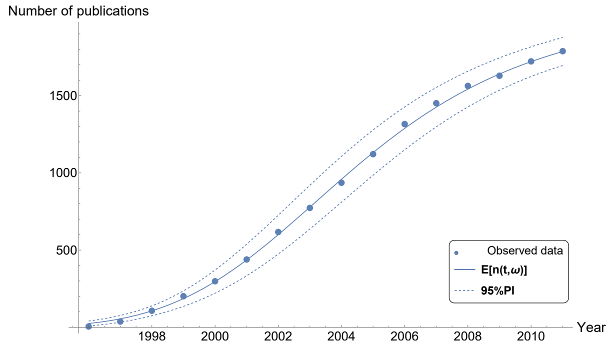


Figure 1: **Left side**: 1-PDF representation of the stochastic process described in (3) of the random differential equation (2) at different time instants. **Right side**: Visualization of the probabilistic fit to the observed data (dots), including the expectation (solid line) and 95% PI's (dashed lines) of the 1-PDF described in (4) associated of the solution presented in (3), utilizing the equations defined in (6).

Finally, we generated an empirical approximation of the posterior predictive distribution defined in (7) by simulating new data points for each set of posterior parameter samples. The result is shown in Figure 2, where it is indicated that the model successfully captures both the underlying trend and the variability of the data.

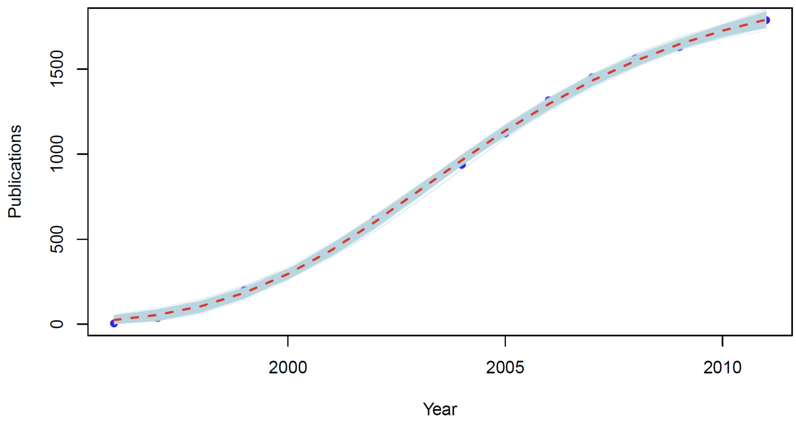


Figure 2: Representation of the posterior predictive distribution, where blue markers denote observed data points and dashed lines depict simulated data generated from model (3) using posterior parameter samples. The red dashed line represents the mean of the simulated data, summarizing the central tendency of the predictions.

4 Conclusions

We have studied an extension of the classical growth model from a probabilistic perspective, treating both the initial condition and all its parameters as continuous random variables. Taking advantage of the nature of the solution, we have employed the RVT technique to determine the 1-PDF. To calculate it in practice, a technique within the Bayesian framework has been applied to properly assign a joint distribution for the vector of model parameters.

These theoretical findings were subsequently applied to real-world data, specifically cumulative data exhibiting a sigmoidal curve pattern. These illustrates a completely different case of a typical growth process, where such a model is normally applied, making them an ideal candidate for testing the flexibility and robustness of the proposed model through empirical validation.

Acknowledgments

This work has been supported by the Spanish grant PID2020-115270GB-I00 granted by MCIN/AEI/10.13039/501100011033. S.M. Sferle has been suported by the grant PRE2021-101090 granted by MCIN/AEI/10.13039/501100011033 and by FSE+.

References

[1] Tsoularis, A., Wallace, J., Analysis of logistic growth model *Mathematical Biosciences*, 179:21–55, 2002.

[2] Turner, M. E. J., Bradley, E. L. J., Kirk, K. A., Pruitt, K. M., A theory of growth. *Mathematical Biosciences*, 29: 367–373, 1976.

[3] Marusic, M., Bajzer, Z., Freyer, J. P., Vuk-Pavlovic, S., Analysis of growth of multicellular tumour spheroids by mathematical models *Cell Proliferation*, 27:73–94, 1994.

[4] Cortés, J. C., Quiles-Navarro, A., Sferle, S. M., Extending the hyper-logistic model to the random setting: New theoretical results with real-world applications *Mathematical Methods in the Applied Sciences*, 1–25, 2024.

[5] Soong, T. T., Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

[6] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B, Bayesian Data Analysis Third Edition. New York, CRC Press, 2013.

[7] Smith, R., Uncertainty Quantification: Theory, Implementation, and Applications. New York, SIAM, 2014.

[8] Kucharavy, D., De Guio, R., Application of logistic growth curve. *Procedia Engineering*, 131: 280–290, 2015.