AIMS Mathematics, 9(10): 27687–27703.
DOI: 10.3934/math.20241344
Received: 14 July 2024
Revised: 02 September 2024
Accepted: 19 September 2024
Published: 25 September 2024

https://www.aimspress.com/journal/Math

Research article

A new stochastic diffusion process based on generalized Gamma-like curve: inference, computation, with applications

Safa' Alsheyab and Mohammed K. Shakhatreh*

Department of Mathematics and Statistics, Jordan University of Science and Technology, P.O.Box 3030, Irbid 22110, Jordan

* **Correspondence:** Email: mkshakhatreh6@just.edu.jo; Tel: +96227201000; Fax: +96227095123.

Abstract: This paper introduces a novel non-homogeneous stochastic diffusion process, useful for modeling both decreasing and increasing trend data. The model is based on a generalized Gamma-like curve. We derive the probabilistic characteristics of the proposed process, including a closed-form unique solution to the stochastic differential equation, the transition probability density function, and both conditional and unconditional trend functions. The process parameters are estimated using the maximum likelihood (ML) method with discrete sampling paths. A small Monte Carlo experiment is conducted to evaluate the finite sample behavior of the trend function. The practical utility of the proposed process is demonstrated by fitting it to two real-world data sets, one exhibiting a decreasing trend and the other an increasing trend.

Keywords: generalized Gamma distribution; maximum likelihood estimate; prediction; stochastic diffusion process; trend function
Mathematics Subject Classification: 62M86, 60H30, 65C30

1. Introduction

Modelling real data, particularly in fields like finance and biology, often involves latent randomness. Consequently, using deterministic models, such as those based on ordinary or partial differential equations, to represent this data may lead to inaccuracies. Alternatively, stochastic diffusion processes (SDPs) can be used, as they are designed to account for the random behavior inherent in these data sets. Typically, a diffusion process $X(t)$ is a solution of the stochastic differential equation (SDE)

$$dX(t) = a(t, X(t), \theta)dt + b(t, X(t), \theta)dW(t), \tag{1.1}$$

where W is a Wiener process, $a(t, X(t), \theta)$, $b(t, X(t), \theta)$ are called the drift and diffusion functions respectively, are known functions, and θ are unknown indexed parameters. Recently, (1.1)

has been employed to develop various new homogeneous/non-homogeneous SDPs, including the Log-Logistic [1], Schumer [2], Lomax [3], Weibull [4], Square-Brennan-Schwartz [5], Brennan-Schwartz [6], Gamma [7], Gompertz [8], and Rayleigh [9] diffusion processes. Most of these studies have focused on modeling growth data, such as microorganism culture growth [1], the evolution of electricity net consumption [2], population growth [5], and the growth of the total stock of private car-petrol [7]. On the other hand, modeling data with a declining trend, such as mortality rates, unemployment rates, and infectious diseases, remains of great interest. However, there are few SDPs available for such data. For example, in [3] proposed the Lomax SDP to model the adolescent fertility rate in Morocco, and in [4] introduced the Weibull SDP to model the age dependency ratio in Morocco.

It is worth mentioning that homogeneous SDPs assume a constant rate for the phenomenon under consideration, whereas non-homogeneous SDPs allow for variability in the rate. The latter is more reasonable in many situations. For example, interest rates are typically a function of time, making it impractical to assume a constant rate. Similarly, while mortality rates can sometimes be constant over a short period, they often change over time due to various factors, notably education and health care improvements. Therefore, modeling such phenomena requires non-homogeneous SDPs to provide insightful analysis and more accurate trend forecasting.

The SDPs introduced in the above-cited references are solved using stochastic calculus methods. Moreover, statistical inference techniques, particularly the maximum likelihood (ML) estimates, are employed to estimate the parameters involved in these models. Consequently, the trend function of the process which is a function in these parameters, can be easily estimated immediately due to the invariance property. While the ML estimates are satisfactory, the derivation of these estimates requires the functional form of the processes, which can be summarized through the transition probability density function. Unfortunately, the transition probability density function is quite difficult to obtain in many processes, which makes obtaining the estimates of these parameters impossible, and approximation methods are required in these cases, for example, [10–12], among others.

In this article, we introduce a novel stochastic model related to a particular type of the generalized Gamma-like stochastic diffusion model. The generalized Gamma probability distribution, proposed in [13], is particularly useful for modeling diverse types of data, especially lifetime data. The probability density function of the generalized Gamma distribution is

$$f(t; a, \alpha, p) \propto t^{\alpha-1} \exp\left\{-\left(\frac{t}{a}\right)^{p}\right\}, \tag{1.2}$$

where $a,\ \alpha,\ p > 0$. Note that a is a scale parameter, while the other two parameters are shape parameters. Additionally, this distribution includes several other common distributions, such as the exponential, gamma, and Weibull distributions, as special cases. However, we shall consider the following version of the generalized Gamma distribution with one parameter:

$$f(t; \alpha) \propto t^{\alpha} \exp\{-b_1(\alpha)\, t^{b_2(\alpha)}\},$$

where $b_1(\alpha),\ b_2(\alpha) > 0$.

We are motivated to introduce a novel SDP that is designed to model various types of trend data, particularly decreasing and increasing trends. This contrasts with many existing SDPs in the literature, which typically model only a single type of trend. Interestingly, the proposed SDP, based primarily on a generalized gamma-like distribution with a single parameter, effectively accommodates both

decreasing and increasing trend data. It is important to note that adding more parameters does not necessarily improve the modeling of such data. For instance, the two-parameter Weibull stochastic diffusion, as discussed in [4], does not significantly enhance the analysis of the data described in the data section. Moreover, introducing additional parameters can complicate the process of obtaining maximum likelihood estimates.

Our main contributions in this paper are as follows: Firstly, we define a novel SDP capable of modeling decreasing and increasing trend data through careful selection of its drift function, which involves a single parameter. Secondly, we thoroughly study the main characteristics of the proposed process by demonstrating its existence and uniqueness, determining the transition probability density function, calculating the moments of the process, and analyzing both the trend and conditional trends. Notably, the trend function is proportional to the probability density function given in (1.1). Thirdly, the SDP parameters are estimated using the maximum likelihood method. While the likelihood function appears intractable, which is common in various SDPs, it becomes manageable due to the kernel function being a log-normal distribution. By applying the log-likelihood function, the process simplifies after some elementary algebra. Fourthly, we conduct simulation experiments to demonstrate the consistency of the maximum likelihood estimates. Finally, two real-world applications exhibiting decreasing and increasing trends are analyzed, and the SDP outperforms several existing processes.

The paper is organized as follows: Section 2 provides a description of the novel SDP and its characteristics, including the solution of the process, the transition probability density function (TPDF), and the moments. In Section 3, the process parameters are obtained using the maximum likelihood estimation method with discrete sampling of the process. A small Monte Carlo experiment is conducted in Section 4. The performance of the proposed process is applied to a real-world data set in Section 5. Finally, some concluding remarks are given in Section 6.

2. The model and its probabilistic characteristics

In this section, we introduce a new one-dimension stochastic diffusion process. Some features of the process, such as the existence of the solution, transition probability distributions, and moments, are explained and derived.

2.1. The generalized Gamma-like curve (GGC) stochastic diffusion model

The GGC process is defined as the non-homogeneous diffusion process depending on $\{X(t), t \in [t_1, T], t_1 > 0\}$ assuming values in $(0, \infty)$ with infinitesimal moments given by

$$a(t, X(t), \theta) = \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha}\right) X(t), \quad b^{1/2}(t, x(t), \theta) = \sigma X(t), \tag{2.1}$$

where t_1 is the initial time and T is the last time. The above-described process can be formally viewed as a solution to the following SDP:

$$dX(t) = \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha}\right) X(t)dt + \sigma X(t)dW(t); \qquad x(t_1) = x_1, \tag{2.2}$$

where $\sigma > 0$, and $\alpha \in \mathbb{R} \setminus \{0\}$ is a non-zero real constant. Clearly, the solution of (2.2) can be obtained via the one-dimensional Itô's integral as follows:

$$X(t) = X_1 + \int_{t_1}^{T} \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right) X(t)dt + \sigma \int_{t_1}^{T} X(t)\mathrm{d}W(t).$$

2.2. Existence and uniqueness

Here, we show the existence and uniqueness of the solution for the GGC process given via in (2.2). Toward this goal, it is enough to verify that the infinitesimal moments satisfy uniform Lipschitz and linear growth conditions; see [11].

Theorem 2.1. *The SDE in* (2.2) *possesses a unique solution.*

Proof. First, we show that the GGC process satisfies a uniform Lipschitz. To do so, consider $x, y \in \mathbb{R}^+$ and $t \in [t_1, T]$. It then follows that

$$\begin{aligned}
|a(t,x) - a(t,y)| + |\sqrt{b(t,x)} - \sqrt{b(t,y)}| &= |a(t,x-y)| + |\sqrt{b(t,x-y)}|, \\
&= \left| \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right)(x-y) \right| + |\sigma(x-y)|, \\
&= \left(\left| \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right) \right| + |\sigma| \right) |x-y|, \\
&\leq \left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right) \right| + |\sigma| \right) |x-y|.
\end{aligned}$$

On the other hand, the process satisfies linear growth since for $y = 0$, we have that

$$\begin{aligned}
|a(t,x)|^2 + |\sqrt{b(t,x)}|^2 &\leq \left(|a(t,x)| + |\sqrt{b(t,x)}| \right)^2, \\
&\leq \left[\left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right) \right| + |\sigma| \right) |x| \right]^2, \\
&\leq \left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{\alpha}{t} - \frac{10^3}{\alpha} t^{-10^2/\alpha} \right) \right| + |\sigma| \right)^2 (1 + |x|)^2.
\end{aligned}$$

Thus, there exists an almost surely (a.s.) continuous process $\{x(t), t \in [t_1, T]; t_1 > 0\}$ that is the unique solution of the SDE (2.2) with probability 1. □

2.3. Probability distribution of GGC process

The determination of the probability distribution for the solution of the GGC process plays a central role in studying the fundamental characteristics of the proposed process, especially its mean function, which serves as a basis for trend analysis. Moreover, since the process parameters are unknown, estimating these quantities using methods like maximum likelihood requires knowledge of the probability distribution of the sample path. The explicit solution of the SDE (2.2) can be obtained by considering the transformation $Y(t) = \log(X(t))$. Upon applying Itô's formula to y, we have the following:

$$dY(t) = \frac{1}{X(t)}dX(t) - \frac{1}{2X^2(t)}\sigma^2 X^2(t)dt,$$
$$= \left(\frac{\alpha}{t} - \frac{10^3}{\alpha}t^{-10^2/\alpha} - \frac{\sigma^2}{2}\right)dt + \sigma dW(t),$$

where $Y(t_1) = \log(X_1)$. On integrating the above equation, we obtain,

$$Y(t) - Y(t_1) = \int_{t_1}^{t}\left(\frac{\alpha}{s} - \frac{10^3}{\alpha}s^{-10^2/\alpha} - \frac{\sigma^2}{2}\right)ds + \sigma(W(t) - W(t_1)),$$

and hence

$$Y(t) = Y(t_1) + \alpha\log(t/t_1) - \frac{10^3/\alpha}{1 - \frac{10^2}{\alpha}}(t^{\frac{-10^2}{\alpha}+1} - t_1^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t - t_1) + \sigma(W(t) - W(t_1)).$$

Therefore, the solution in terms of the original GGC process.

$$X(t) = X_1\left(\frac{t}{t_1}\right)^{\alpha}\exp\left(-\frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - t_1^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t - t_1)\right)e^{\sigma(W(t)-W(t_1))}. \tag{2.3}$$

Observe that the process $Y(t)$ is a Gaussian process if and only if the initial condition Y_1 is a constant or is normally distributed. Since the initial condition is constant a.s, i.e., $P(Y_1 = y_1) = 1$, and $Y(t)$ is a Markovian process, it then follows that the finite-dimensional distribution of $Y(t)$ is normal. Consequently, the finite dimensional distribution of $X(t)$ is log-normal distribution. Additionally, the TPDF of $X(t)$ given $X(s)$ where $s < t$ follows a log-normal distribution denoted by $\Lambda_1(\mu(s,t,x_{ts}), \sigma^2(t-s))$, where $\mu(s,t,x_s)$ is given by

$$\mu(s,t,x) = \log(x) + \alpha\log(t/s) - \frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - s^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t-s).$$

Therefore, the TPDF of the process considered has the following form:

$$f(y,t|x_s,s) = \frac{1}{y}\left[2\pi\sigma^2(t-s)\right]^{-1/2}\exp\left(-\frac{[\log(y) - \mu(s,t,x_s)]^2}{2\sigma^2(t-s)}\right). \tag{2.4}$$

2.4. Mean function of the process

Since $X(t)$ is distributed according to $\Lambda_1(\mu(s,t,x_s), \sigma^2(t-s))$, it then follows from the properties of the lognormal distribution that the n^{th} conditional moment of $X(t)$ given $X(s)$ is

$$E[X^n(t)|X(s) = x_s] = \exp\left(n\mu(s,t,x_s) + \frac{n^2\sigma^2}{2}(t-s)\right).$$

The conditional mean which considers as the trend function can be obtained using ($n = 1$). That is

$$E[X(t)|X(s) = x_s] = x_s\left(\frac{t}{s}\right)^{\alpha}\exp\left(-\frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - s^{\frac{-10^2}{\alpha}+1})\right). \tag{2.5}$$

On the other hand, the mean function or the unconditional trend of the process is given by

$$E[X(t)] = x_{t_1}\left(\frac{t}{t_1}\right)^{\alpha}\exp\left(-\frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1}-t_1^{\frac{-10^2}{\alpha}+1})\right), \tag{2.6}$$

where the above equation is obtained under the assumption that $P(X(t_1) = x_1) = 1$. Notice that when $\alpha > 0$, it then follows that the trend function is proportional to one-parameter generalized Gamma density. Similarly, other statistical measures such as the variance, Skewness, and Kurtois can be obtained. In particular, the variance of the process is

$$Var[X(t)] = E[X^2(t)] - (E[X(t)])^2 = x_{t_1}^2\left(\frac{t}{t_1}\right)^{2\alpha}\exp\left(-\frac{2(10^3)}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1}-t_1^{\frac{-10^2}{\alpha}+1})\right)\left(e^{\sigma^2(t-t_1)}-1\right).$$

3. Inference on the process

Once the process, along with its basic properties, is introduced and discussed, it becomes important to examine its significance in simulation and application. However, the presence of unknown parameters makes trend analysis in practice difficult. Consequently, these parameters need to be estimated, and the maximum likelihood estimate emerges as a suitable choice, given the known functional form of the TPDF. Furthermore, ML estimates possess advantageous properties such as invariance, efficiency, and asymptotic normality.

3.1. ML estimates

Here, the two parameters involved in the drift and diffusion functions are estimated using the ML method. We consider discrete sampling observations of the process $x(t_1), x(t_2), \ldots, x(t_n)$ at times $t_1, t_2, \ldots, t_n = T$. For simplicity, put $t_{j+1} - t_j = h$ and use x_i to refer to $x(t_i) = x_i$. The likelihood function of $\boldsymbol{\theta} = (\alpha, \sigma^2)^T$ can be obtained from Eq (2.4), taking into account that the initial condition taking $P(X(t_1) = x_1) = 1$, is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \prod_{j=1}^{n-1} f(x_{j+1}, t_{j+1}|x_j, t_j),\\ &= \prod_{j=1}^{n-1}\frac{1}{x_{j+1}}\left[2\pi\sigma^2(t_{j+1}-t_j)\right]^{-1/2}\exp\left(-\frac{[\log(x_{j+1})-\mu(j,j+1,x_j)]^2}{2\sigma^2(t_{j+1}-t_j)}\right).\end{aligned}$$

The log-likelihood equation is

$$\begin{aligned}\boldsymbol{\ell}(\alpha,\sigma^2) = &-\frac{n-1}{2}\log(2\pi h) - \frac{n-1}{2}\log(\sigma^2) - \sum_{j=1}^{n-1}\log(x_{j+1})\\ &-\frac{1}{2\sigma^2 h}\sum_{j=1}^{n-1}\left[\log(\frac{x_{j+1}}{x_j}) - \alpha\log(\frac{t_{j+1}}{t_j}) + \frac{10^3}{-10^2+\alpha}(t_{j+1}^{\frac{-10^2}{\alpha}+1} - t_j^{\frac{-10^2}{\alpha}+1}) + \frac{\sigma^2 h}{2}\right]^2.\end{aligned} \tag{3.1}$$

The $\boldsymbol{\ell}(\alpha, \sigma^2)$ can be maximized by solving the nonlinear likelihood equation obtained by differentiating with respect to $\boldsymbol{\theta} = (\alpha, \sigma^2)^T$. The first partial derivatives of $\boldsymbol{\ell}(\alpha, \sigma^2)$ are given by

$$\frac{\partial \boldsymbol{\ell}(\alpha,\sigma)}{\partial \alpha} = \sum_{j=1}^{n-1}(\mathcal{H}_{\alpha,j} + \frac{\sigma^2 h}{2})\Big[- \log(\frac{t_{j+1}}{t_j}) + \frac{10^5}{\alpha^2(-10^2+\alpha)}\Big(\log(t_{j+1})t_{j+1}^{\frac{-10^2}{\alpha}+1} - \log(t_j)t_j^{\frac{-10^2}{\alpha}+1}\Big)$$
$$- \frac{10^3}{(-10^2+\alpha)^2}\Big(t_{j+1}^{\frac{-10^2}{\alpha}+1} - t_j^{\frac{-10^2}{\alpha}+1}\Big)\Big], \tag{3.2}$$

$$\frac{\partial \boldsymbol{\ell}(\alpha,\sigma))}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{1}{2\sigma^4 h}\sum_{j=1}^{n-1}\mathcal{H}_{\alpha,j}^2 - \sum_{j=1}^{n-1}\frac{h}{8}, \tag{3.3}$$

where

$$\mathcal{H}_{\alpha,j} = \log(\frac{x_{j+1}}{x_j}) - \alpha\log(\frac{t_{j+1}}{t_j}) + \frac{10^3}{-10^2+\alpha}(t_{j+1}^{\frac{-10^2}{\alpha}+1} - t_j^{\frac{-10^2}{\alpha}+1}).$$

Let $S(\boldsymbol{\theta}) = (\partial\boldsymbol{\ell}(\boldsymbol{\theta})/\partial\alpha, \partial\boldsymbol{\ell}(\boldsymbol{\theta})/\partial\sigma^2)^T$ be the score function. Then, the MLE of $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\sigma^2})$ can be obtained by solving $S(\boldsymbol{\theta}) = 0$. Unfortunately, the MLEs cannot be obtained in closed form, and numerical methods are required to obtain these estimates. From Eq (3.3), we obtain

$$\frac{\hat{\sigma}^2}{2} = \frac{1}{h}\left[\left(1 + \frac{1}{n-1}\sum_{j=1}^{n-1}\mathcal{H}_{\hat{\alpha},j}^2\right)^{1/2} - 1\right]. \tag{3.4}$$

On substituting (3.4) in Eq (3.2) the following nonlinear equation is obtained for the estimator $\hat{\alpha}$:

$$\sum_{j=1}^{n-1}\Big(\mathcal{H}_{\hat{\alpha},j} + \frac{\hat{\sigma}^2 h}{2}\Big)\Big[- \log(\frac{t_{j+1}}{t_j}) + \frac{10^5}{\hat{\alpha}^2(\hat{\alpha}-10^2)}\Big(\log(t_{j+1})t_{j+1}^{1-\frac{10^2}{\hat{\alpha}}} - \log(t_j)t_j^{1-\frac{10^2}{\hat{\alpha}}}\Big) \tag{3.5}$$
$$- \frac{10^3}{(\hat{\alpha}-10^2)^2}(t_{j+1}^{1-\frac{10^2}{\hat{\alpha}}} - t_j^{1-\frac{10^2}{\hat{\alpha}}})\Big] = 0.$$

Let $g(\alpha)$ be the left-hand side equation (3.6). Therefore, the ML estimate of α can be achieved by solving the non-linear equation $g(\alpha) = 0$.

3.2. Estimated trend functions and confidence bounds

Once the ML estimates of α and σ^2 are obtained, we proceed to provide estimates for the conditional mean and the mean of the process. Due to the invariance property of the ML estimates, see for example Theorem [5.28, 308] in [14], the ML estimates for the conditional mean and the unconditional mean can be obtained. Let $\widehat{\alpha}$ and $\widehat{\sigma^2}$ be the ML estimates of α and σ^2, respectively, then the estimated conditional mean of the process (ECMF) is

$$\widehat{E}(X(t)|X(s)) = x_s\left(\frac{t}{s}\right)^{\widehat{\alpha}}\exp\left(-\frac{10^3}{-10^2+\widehat{\alpha}}\Big(t^{\frac{-10^2}{\widehat{\alpha}}+1} - s^{\frac{-10^2}{\widehat{\alpha}}+1}\Big)\right). \tag{3.6}$$

Similarly, the estimated mean function (EMF) of the process is

$$\widehat{E}(X(t)|X(t_1)) = x_1\left(\frac{t}{t_1}\right)^{\widehat{\alpha}}\exp\left(-\frac{10^3}{-10^2+\widehat{\alpha}}\Big(t^{\frac{-10^2}{\widehat{\alpha}}+1} - t_1^{\frac{-10^2}{\widehat{\alpha}}+1}\Big)\right), \tag{3.7}$$

where in the above equation we used the assumption that $P(X(t_1) = x(t_1)) = 1$.

At time, we would like to obtain a confidence band for the CMF and MF of the process. From Eq (2.4), we have that for $t > s$, $X(t)|X(s)$ follows $\Lambda_1(\mu(s,t,x_{xs}), \sigma^2(t-s)$.) Therefore, we have that

$$Z = \frac{\ln(X(t)) - \mu(s,t,x_s)}{\sigma\sqrt{t-s}} \sim \mathcal{N}(0,1).$$

Consequently, a $(1-\alpha)100\%$ confidence band for z is determined by $P(-Z_{\gamma/2} \leq Z \leq Z_{\gamma/2}) = 1-\gamma$, for $\gamma \in (0,1)$.

$$\mathbb{P}\left[-Z_{\gamma/2} \leq \frac{\log\left(\frac{x(t)}{x_1}\right) - \alpha\log\left(\frac{t}{t_1}\right) + \frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - t_1^{\frac{-10^2}{\alpha}+1}) + \frac{\sigma^2}{2}(t-t_1)}{\sigma\sqrt{t-t_1}} \leq Z_{\gamma/2}\right] \approx 1-\gamma.$$

Therefore, the $(1-\gamma)100\%$ confidence bound (CB) for $X(t)$ is

$$x_{lower}(t) \leq x(t) \leq x_{upper}(t), \tag{3.8}$$

where,

$$x_{lower}(t) = x_1 \exp\left[-Z_{\gamma/2}\sigma\sqrt{t-t_1} + \alpha\log\left(\frac{t}{t_1}\right) - \frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - t_1^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t-t_1)\right],$$

$$x_{upper}(t) = x_1 \exp\left[Z_{\gamma/2}\sigma\sqrt{t-t_1} + \alpha\log\left(\frac{t}{t_1}\right) - \frac{10^3}{-10^2+\alpha}(t^{\frac{-10^2}{\alpha}+1} - t_1^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t-t_1)\right].$$

4. Simulation experment

Here, our primary goal is to illustrate the pattern of the GGC process by simulating sample trajectories. Additionally, we use these sample paths to evaluate the frequentist performance of the ML estimates for the trend function of the GGC process. The following algorithm describes the procedure for simulating a trajectory of the GGC process computed at N time points.

Algorithm

1: Initialize, $t_1, T, N,\ x_1, \alpha, \sigma$,
2: Compute $\Delta := \frac{T-t_1}{N}$,
3: **For** $j = 2, \ldots, N$, **do**
   (a) Discretize the time interval $[t_1, T]$, into N time points with each point computed as $t_i = t_{i-1} + (i-1)\Delta$,
   (b) Generate a standard normal observation, i.e., $Z : N(0,1)$, and the increments are computed as $W_j = W_{j-1} + \sqrt{\Delta}\, Z_i$,
   (c) The sample path of the GGC process can be obtained upon using (2.3), i.e., $X_j = X_{j-1}\left(\frac{t_j}{t_{j-1}}\right)^{\alpha} \exp\left\{-\frac{10^3}{-10^2+\alpha}(t_j^{\frac{-10^2}{\alpha}+1} - t_{j-1}^{\frac{-10^2}{\alpha}+1}) - \frac{\sigma^2}{2}(t_j - t_{j-1})\right\} \exp\left(\sigma(W_j - W_{j-1})\right)$.
4: **End do**
5: Repeat step 3 m times to obtain m sample paths.

6: For $i = 1, \ldots, m$, use the sample path $X_i(t_j)$ in step 5 to compute the ML estimates, using Eqs (3.2) and (3.3), $\widehat{\alpha}_i$ and $\widehat{\sigma}_i$, respectively.
7: The ML estimates are then given by: $\widehat{\alpha} = m^{-1}\sum_{i=1}^{m}\widehat{\alpha}_i$ and $\widehat{\sigma} = m^{-1}\sum_{i=1}^{m}\widehat{\sigma}_i$ with mean square errors respectively, $MSE(\widehat{\alpha}) = (m-1)^{-1}\sum_{i=1}^{m}(\widehat{\alpha}_i - \alpha)^2$ and $MSE(\widehat{\sigma^2}) = (m-1)^{-1}\sum_{i=1}^{m}(\widehat{\sigma^2}_i - \sigma^2)^2$.

We generate $m = 50$ training sample paths using the algorithm described above, varying the parameters in the graphs. All computations in this paper were performed using R software [15].

(1) In the first scenario, we set the time interval $[t_1, T] = [1000, 1010]$, $N = 100$, and $X(t_1) = 9000$. Figure 1 shows that the trajectories vary in direction based on the sign of the parameter α. The dispersion of the process is influenced by the value of σ; smaller values result in trajectories that closely follow a single curve, while larger σ values lead to more dispersed yet similarly trending curves. Using the 50 sample paths, we obtain the maximum likelihood (ML) estimates of the process parameters and subsequently the ML estimates of the trend function. Figure 1 demonstrates that the estimated trend closely matches the actual trend function, indicating the reliability of our estimates.
(2) In the second scenario, we consider $[t_1, T] = [500, 510]$, $N = 100$, and $X(t_1) = 3000$. Similarly, Figure 2 reveals that the process exhibits either a decreasing or increasing trend. The sample paths tend to either closely follow a single curve or scatter depending on whether the dispersion parameter σ is small or large, respectively. Additionally, using 50 sample trajectories, we obtain the ML estimates of α and σ, and consequently the estimate of the trend function. A closer examination shows that the ML estimates behave consistently across different scenarios.

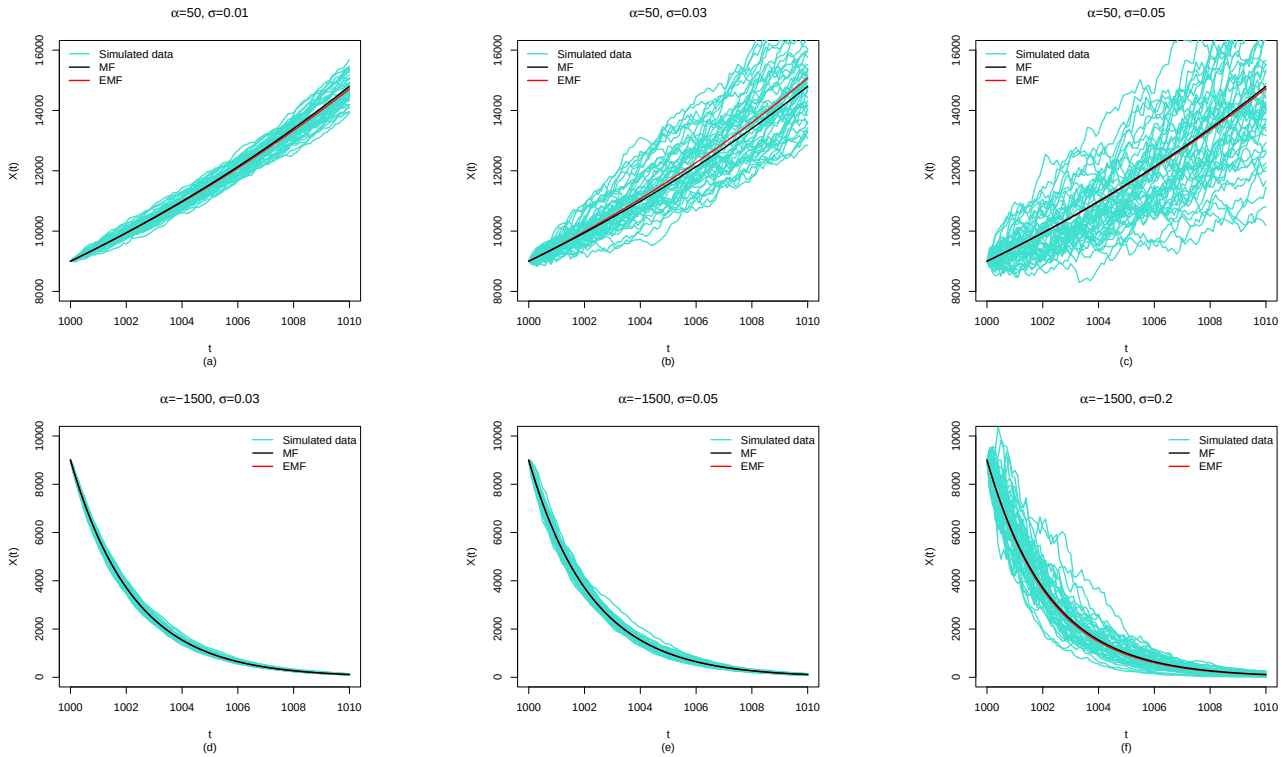

Figure 1. Simulated trajectories of the GGC process along with MF and EMF.

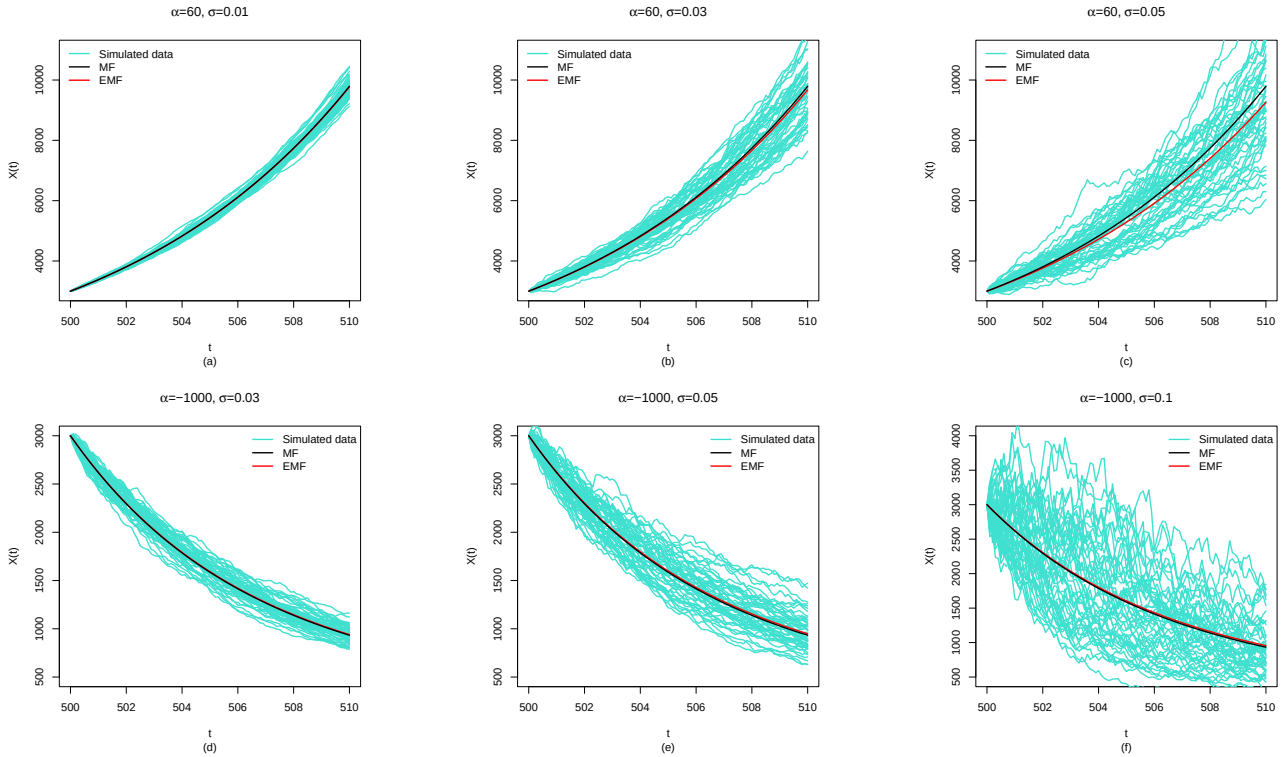

Figure 2. Simulated trajectories of the GGC process along with MF and EMF.

5. Applications

This section presents the application of our model to two real-world datasets: the number of infant deaths in the United Kingdom from 1977 to 2020 and CO_2 emissions (in kilotons) in Morocco from 1990 to 2020. These annualized data are sourced from the World Bank database. The first dataset is fitted using the GGC process and compared with two existing models: the two-parameter Weibull SDP [4] and the Lomax SDP [3], both of which are suitable for modeling data with decreasing trends. Similarly, the second dataset is fitted using the GGC process and compared with the log-logistic SDP [1], which is appropriate for modeling data with increasing trends.

5.1. Application to number of infant deaths in the United Kingdom

According to the Royal College of Paediatrics and Child Health (RCPCH), infant death rates in all countries of the UK have significantly fallen over the past 40 years. Most childhood deaths occur during the first year of life, particularly in the first month (neonatal period). Newborn deaths account for 70% to 80% of infant deaths. The vast majority of newborn deaths are due to perinatal causes, especially preterm birth, and are strongly linked to maternal health and congenital anomalies. The remaining infant deaths occur due to a wide variety of causes, including sudden unexplained death in infancy (SUDI). The overall decline in infant death rates since 1980 likely reflects general improvements in health care, specifically prenatal and newborn care. Breastfeeding and safe sleeping positions are protective factors for infant survival, especially for premature babies. This study aims to use a stochastic diffusion process to model the number of infant deaths and to forecast the

number of infant deaths in the United Kingdom. We use the data from 1977 to 2018, available at https://databank.worldbank.org/source/world-development-indicators, as training data to estimate the model parameters: $\hat{\alpha} = -1779.057$ and $\hat{\sigma} = 0.02208178$. Consequently, the estimated trend functions (unconditional EMF and conditional ECMF) are obtained immediately due to the invariance of ML estimates. Additionally, we use the years 2019 and 2020 to forecast the number of infant deaths in the UK. The results for the forecasted values for these two years are presented in Table 1 and shown in Figure 3. These results demonstrate that the GGC process is quite effective in predicting the values for these years, particularly when employing ECMF. Figure 3 (left panel) displays the observed data, estimated trend function (EMF), and estimated confidence bands, revealing that the GGC process fits the current data well and provides accurate predictions. Figure 3 (right panel) further illustrates the conditional trend function estimation along with forecasts. Moreover, we compare the proposed diffusion process with the Weibull [4] and Lomax [3] diffusion processes using the current data. Table 2 lists the estimated parameters and the Akaike information criterion (AIC), showing that the GGC process has the lowest AIC value and therefore outperforms the two-parameter diffusion processes. This is also well demonstrated in Figure 3. Moreover, Figure 4 shows the fits made using the methods mentioned in Table 2, revealing that the GGC process performed much better than the Weibull [4] and Lomax [3] diffusion processes in modelling the current data.

The accuracy of the forecast can be quantified using measures such as the mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE), defined as follows: $\text{MAE} = \frac{1}{42}\sum_{i=1}^{42}|x_(t_i) - \widehat{x}(t_i)| = 257.9876$, $\text{RMSE} = \sqrt{\frac{1}{42}\sum_{i=1}^{42}|x(t_i) - \widehat{x}(t_i)|^2} = 330.2669$, and $\text{MAPE} = \frac{1}{42}\sum_{i=}^{42}\frac{|x(t_i)-\widehat{x}(t_i)|}{x(t_i)} \times 100 = 5.1237$. Table 3 shows the observed data along with their corresponding forecast. The value obtained for MAPE is less than 10, and this indicates that we obtained a high-accuracy prediction.

Table 1. Predictions from EMF and ECMF of the GGC process.

Year	Real Data	EMF	ECMF
2019	2703	2790.843	2763.366
2020	2571	2738.968	2653.739

Table 2. ML estimates of the parameters along with AIC.

Model	α	β	σ	AIC
GGC	-1779.057	NA	0.02208178	500.9154
Weibull	-58.01113386	7.200000	0.02290836	505.927
Lomax	-59.05057463	-0.29692752	0.02290831	505.924

Table 3. Observed data, fits and forecast using the EMF.

Year(t)	$x(t)$	$\widehat{x(t)}$	Year(t)	$x(t)$	$\widehat{x(t)}$
1977	9353	9353.000	1998	4121	4606.083
1978	9011	8999.828	1999	3979	4476.542
1979	9041	8664.142	2000	3814	4352.688
1980	8942	8344.971	2001	3675	4234.245
1981	8433	8041.403	2002	3600	4120.956
1982	7849	7752.580	2003	3619	4012.576
1983	7479	7477.700	2004	3681	3908.874
1984	7264	7216.006	2005	3680	3809.631
1985	7183	6966.78	2006	3676	3714.641
1986	7096	6729.379	2007	3700	3623.708
1987	6991	6503.152	2008	3696	3536.649
1988	6858	6287.515	2009	3614	3453.288
1989	6539	6081.914	2010	3515	3373.459
1990	6189	5885.826	2011	3424	3297.008
1991	5831	5698.758	2012	3324	3223.784
1992	5415	5520.246	2013	3179	3153.648
1993	5031	5349.855	2014	3051	3086.466
1994	4724	5187.171	2015	3010	3022.114
1995	4492	5031.808	2016	2988	2960.472
1996	4342	4883.399	2017	2925	2901.426
1997	4239	4741.599	2018	2817	2844.871

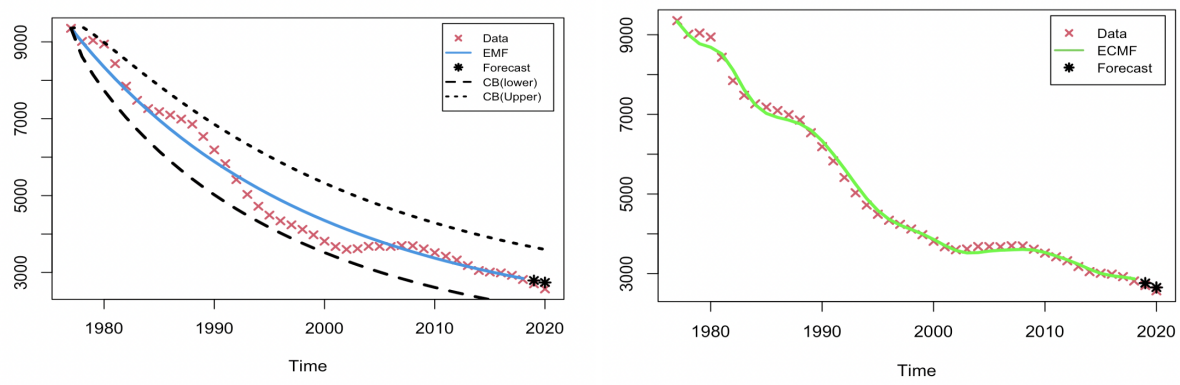

Figure 3. Left panel: observed data against EMF. Right panel: observed data versus ECMF.

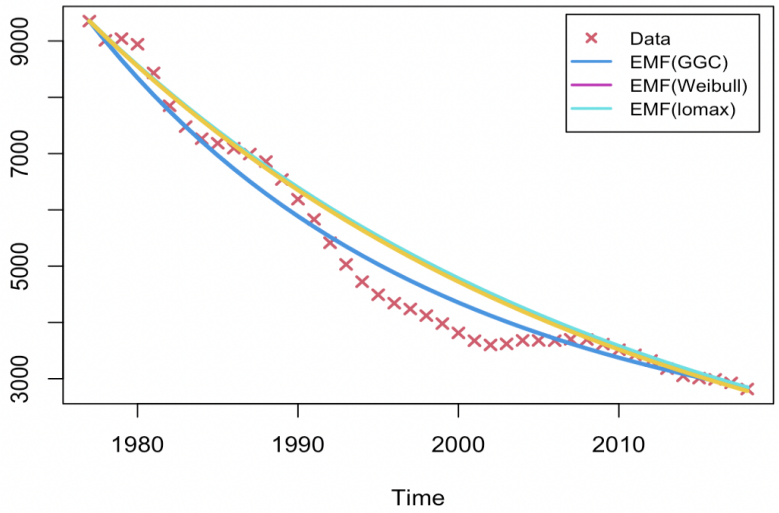

Figure 4. Observed data, EMF using GGC, Weibull, and Lomax diffusion processes.

5.2. Application to CO_2 emissions (kt) in Morocco

Another application of our model is to analyze Carbon Dioxide (CO_2) emissions in Morocco, measured in kilotons (kt). CO_2 emissions primarily result from burning fuels and cement manufacturing. Since the Industrial Revolution, the increasing combustion of carbon-based fuels has significantly raised atmospheric CO_2 concentrations, accelerating global warming and contributing to man-made climate change. Additionally, CO_2 dissolves in water to form carbonic acid, which leads to ocean acidification. This subsection examines CO_2 emissions (kt) in Morocco from 1990 to 2020, with data sourced from `https://databank.worldbank.org/source/world-development-indicators`. We use the data from 1990-2018 as training data to estimate the model parameters: $\hat{\alpha} = 81.55085457$ and $\hat{\sigma} = 0.02977168$. Therefore, the estimated trend functions both (the unconditional EMF and the conditional ECMF) are derived directly due to the invariance of the ML estimates. Similarly, we use the data from 2019 and 2020 to project CO_2 emissions in Morocco. The predicted values for these two years are reported in Table 4 and illustrated in Figure 5. Once again, the results demonstrate that the GGC process is capable of accurately predicting the values for these years using both EMF and ECMF. Figure 5 (left panel) depicts the observed data, the estimated trend function (EMF), and the estimated confidence intervals, indicating that the GGC process accurately fits the data and yields precise forecasts. Figure 5 (right panel) further demonstrates the conditional trend function estimation and projections. Furthermore, we compare the proposed diffusion process with the log-logistic diffusion process [1] using the current dataset. Table 5 presents the estimated parameters and the AIC criterion, showing that the GGC process has a lower AIC value and thus surpasses the two-parameter diffusion process. This is also clearly illustrated in Figure 5. Furthermore, Figure 6 displays the fits performed using the methods outlined in Table 5, demonstrating that the GGC process outperformed the Log-Logistic diffusion process in modeling this data.

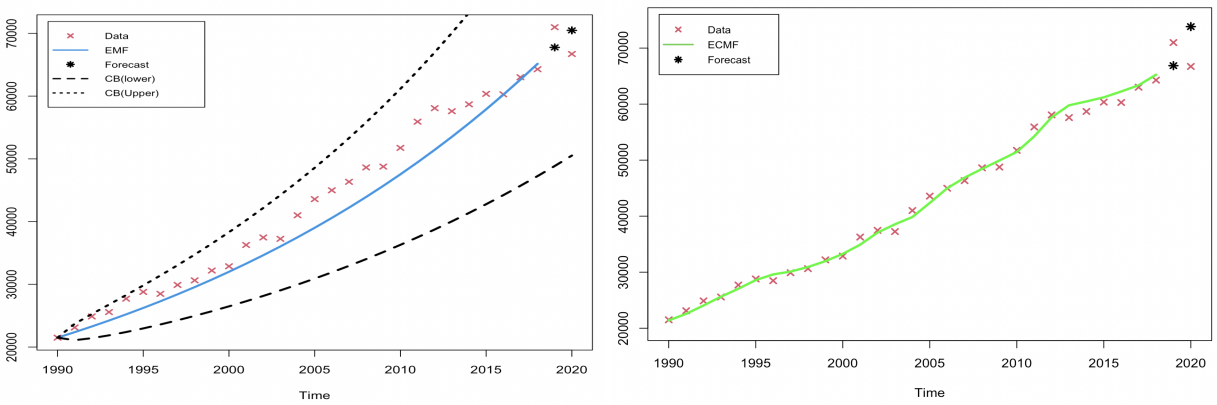

Figure 5. Left panel: observed data against EMF. Right panel: observed data versus ECMF.

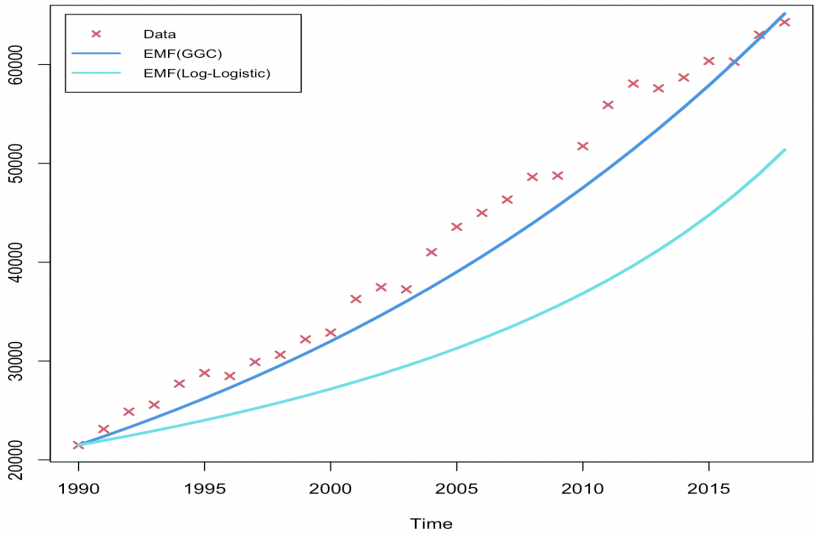

Figure 6. Observed data, EMF using GGC, and log-logistic processes.

Table 4. Predictions from EMF and ECMF of the GGC process.

Year	Real Data	EMF	ECMF
2019	70986.3	67764.45	66863.89
2020	66719.5	70480.36	73831.33

Table 5. ML estimates of the parameters along with AIC.

Model	α	β	σ	AIC
GGC	81.55085457	NA	0.02977168	481.1204
log-logistic	-0.03638392	0.64735468	0.03437501	490.76

Similarly, the accuracy of the forecast can be assessed using the metrics: MAE = 2567.19, RMSE = 3115.123, and MAPE = 6.032141. Table 6 shows the observed data along with their corresponding forecast. The MAPE value obtained is less than 10, indicating a high-accuracy prediction.

Table 6. Observed data, fits and forecast using the EMF.

Year(t)	$x(t)$	$\widehat{x}(t)$	Year(t)	$x(t)$	$\widehat{x}(t)$
1990	21497.8	21497.80	2005	43579.9	39011.22
1991	23119.0	22372.13	2006	44991.0	40585.80
1992	24879.0	23281.55	2007	46341.0	42223.10
1993	25577.0	24227.46	2008	48630.2	43925.60
1994	27712.5	25211.30	2009	48765.7	45695.85
1995	28791.8	26234.57	2010	51749.5	47536.51
1996	28484.7	27298.83	2011	55923.5	49450.35
1997	29903.8	28405.71	2012	258076	51440.24
1998	30634.6	29556.88	2013	57595.5	53509.16
1999	32198.7	30754.09	2014	58691.7	55660.21
2000	32876.5	31999.17	2015	60362.5	57896.61
2001	36273.0	33294.00	2016	60289.9	60221.69
2002	37470.0	34640.54	2017	63014.9	62638.94
2003	37251.1	36040.83	2018	64286.1	65151.94
2004	41009.0	37496.99			

6. Conclusions

We introduce a new stochastic diffusion process based on generalized Gamma-like curves, referred to as the GGC process. The GGC process is capable of modeling both increasing and decreasing trend data. We investigate and derive several structural properties, including the explicit solution of the process, the transition probability density function, and the moments of the process. The maximum likelihood method is employed to estimate the GGC process parameters. Simulation studies demonstrate that the ML estimates are consistent even in small samples. The potential of the GGC process for analyzing data with a declining trend is illustrated by fitting it to infant mortality data in the UK. The proposed diffusion process outperforms two existing popular stochastic diffusion processes that are designed to fit decreasing trend data. Additionally, it is applied to CO_2 emissions in Morocco, which exhibit an increasing trend. The GGC process again showed satisfactory performance in modeling this data and outperformed the log-logistic SDP, which is suitable for modeling growth

data.

As with any stochastic diffusion process or model, the GGC process may be inappropriate for modeling unimodal or bathtub-shaped data. This can be addressed by selecting different infinitesimal moments or by incorporating additional suitable parameters.

For future work, we plan to model the first passage time using the GGC process with more applications. Additionally, we will explore other estimation methods, particularly the Bayesian approach, which may be more powerful in handling other application scenarios.

Author contributions

Conceptualization, M.K.S.; methodology, M.K.S.; writing-original draft preparation, M.K.S.; software, M.K.S and S.A.; validation, S.A.; formal analysis, M.K.S. and S.A. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We would like to thank the Editor, Associate Editor, and the three anonymous referees for their insightful comments and valuable suggestions, which greatly improved our paper.

Conflict of interest

No potential conflict of interest was reported by the authors.

References

1. A. El Azri, N. Ahmed, A stochastic log-logistic diffusion process: statistical computational aspects and application to real data, *Stoch. Models*, **40** (2024), 261–277. https://doi.org/10.1080/15326349.2023.2241070
2. A. Nafidi, A. El Azri, R. Gutiérrez-Sánchez, A stochastic Schumacher diffusion process: probability characteristics computation and statistical analysis, *Methodol. Comput. Appl. Probab.*, **25** (2023), 66. https://doi.org/10.1007/s11009-023-10031-4
3. A. Nafidi, I. Makroz, R. Gutiérrez-Sánchez, A stochastic Lomax diffusion process: statistical inference and application, *Mathematics*, **9** (2021), 100. https://doi.org/10.3390/math9010100
4. A. Nafidi, M. Bahij, R. Gutiérrez-Sánchez, B. Achchab, Two-parameter stochastic Weibull diffusion model: statistical inference and application to real modeling example, *Mathematics*, **8** (2020), 160. https://doi.org/10.3390/math8020160
5. A. Nafidi, G. Moutabir, R. Gutiérrez-Sánchez, E. Ramos-Ábalos, Stochastic square of the Brennan-Schwartz diffusion process: statistical computation and application, *Methodol. Comput. Appl. Probab.*, **7** (2020), 455–476. https://doi.org/10.1007/s11009-019-09714-8
6. A. Nafidi, G. Moutabir, R. Gutiérrez-Sánchez, Stochastic Brennan–Schwartz diffusion process: statistical computation and application, *Mathematics*, **7** (2019), 1062. https://doi.org/10.3390/math7111062

7. R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi, The trend of the total stock of the private car-petrol in Spain: stochastic modelling using a new gamma diffusion process, *Appl. Energy*, **86** (2009), 18–24. https://doi.org/10.1016/j.apenergy.2008.03.016
8. R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi, Modelling and forecasting vehicle stocks using the trends of stochastic Gompertz diffusion models: the case of Spain, *Appl. Stoch. Model. Bus. Ind.*, **25** (2009), 385–405. https://doi.org/10.1002/asmb.754
9. R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi, The stochastic Rayleigh diffusion model: statistical inference and computational aspects. Applications to modelling of real cases, *Appl. Math. Comput.*, **175** (2006), 628–644. https://doi.org/10.1016/j.amc.2005.07.047
10. B. M. Bibby, M. Sørensen, Martingale estimation functions for discretely observed diffusion processes, *Bernoulli*, **1** (1995), 17–39. https://doi.org/10.2307/3318679
11. P. E. Kloeden, E. Platen, *Numerical solution of stochastic differential equations*, Springer Berlin, Heidelberg, 1992. https://doi.org/10.1007/978-3-662-12616-5
12. B. L. S. Prakasa Rao, *Statistical inference for diffusion type processes*, Arnold, London, UK, 1999.
13. E. W. Stacy, A generalization of the Gamma distribution, *Ann. Math. Statist.*, **33** (1962), 1187–1192. https://doi.org/10.1214/aoms/1177704481
14. M. J. Schervish, *Theory of statistics*, Springer-Verlag, New York, USA, 1995. https://doi.org/10.1007/978-1-4612-4250-5
15. The R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. Available from: `https://web.mit.edu/r_v3.3.1/fullrefman.pdf`.

AIMS Press