

A diffusion model with cubic drift: statistical and computational aspects and application to modelling of the global CO₂ emission in Spain

R. Gutiérrez*[†], R. Gutiérrez-Sánchez, A. Nafidi and E. Ramos

*Department of Statistics and Operations Research, University of Granada, Facultad de Ciencias,
Campus de Fuentenueva s/n, 18071 Granada, Spain*

SUMMARY

The aim of this work is the study of a new stochastic diffusion model with a cubic-type drift coefficient. The model is considered as the solution of an Ito stochastic differential equation. Using the Ito's stochastic calculus and properties of the Kummer function, the trend functions and steady-state distribution for the process are obtained. Statistical estimation and corresponding computational methodology are established. Finally, the model is applied to modelling and prediction of the global CO₂ emission in Spain. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: cubic drift; Rayleigh and logistic diffusion processes; Kummer function; statistical inference; trend functions; simulation; CO₂ emission in Spain

1. INTRODUCTION: BACKGROUND AND AIMS OF THE STUDY

In this paper we propose a stochastic diffusion model with a cubic-type drift coefficient, containing terms involving powers three and one of the process. This model is original and novel in comparison with other cubic-type diffusions that have been suggested as theoretical models within the theory of population growth. Among other advantages, this model can be statistically fitted to real cases of growth within a stochastic environment. Moreover, it enables us to fit the underlying trends in the phenomena being considered and to analyse possible anomalies affecting such trends. Below, we examine in greater detail the context in which this model is proposed, its technical background and the new possibilities it offers for application to real data.

1.1. Stochastic modelling of growth phenomena in a random environment

As is well known, various authors have introduced stochastic versions of classical deterministic growth models (Gompertz, bass, logistic, etc.) that are used, among other purposes, for modelling

*Correspondence to: R. Gutiérrez, Department of Statistics and Operations Research, University of Granada, Facultad de Ciencias, Campus de Fuentenueva s/n, 18071 Granada, Spain.

[†]E-mail: rgjaimez@ugr.es

secular trends in phenomena of interest in a wide variety of fields, including growth in animal or cell populations, economics, energy, hydrology and environmental studies. In particular, stochastic versions of lognormal, Gompertz and logistic models have been considered in Gutiérrez *et al.*, (1991); Gutiérrez *et al.*, (2005b, 2006a) and Skiadas and Giovani (1997).

The behaviour of growth phenomena, in general, is affected by environmental fluctuations that are responsible for the discrepancies between experimental data and the corresponding theoretical predictions. Different authors have introduced the idea of 'growth in a random environment'. It has been proposed, for example, by Capocelli *et al.* (1974) that such fluctuations could be taken into account by replacing the 'intrinsic fertility' in the growth equation with a 'Gaussian stationary delta-correlated random process', the mean of which is identified with the fertility of the population. This situation leads, technically speaking, to describe the growth process in such a way that the number of individuals present at each time is identified with a continuous Markov process, that is a diffusion process.

From a technical point of view, the first step is to establish the stochastic versions of the deterministic growth models. For this purpose, an interesting and technically appropriate approach is to take the deterministic differential equations whose solution, under appropriate initial or boundary conditions, is the growth curve being considered, and from these to construct the corresponding Ito or Stratonovich stochastic differential equation (SDE) of the behaviour of the dynamic variable X_t under the randomisation hypotheses of the characteristics of the phenomenon studied. For example, on the basis of the logistic differential equation (DE) given in terms of the 'intrinsic fertility' of the population, we can formulate the corresponding stochastic version, randomising this fertility, as shown below.

After having determined the SDE of the randomised version of the behaviour of the variable X_t , whose solution is a stochastic diffusion process under given technical hypotheses, the Ito or Stratonovich stochastic calculus can be applied to the probabilistic analysis of the randomised model. An alternative, complementary methodology is that of analysis based on the backward and forward Kolmogorov equations corresponding to the SDE of the model. By means of either methodology we may approach different problems of a probabilistic nature concerning the randomised growth model, such as the obtention of the model transition density and moments, the density of first-passage time by barriers of specific interest in practice, among other characteristics, as can be seen for example in Gutiérrez *et al.* (1997), Ricciardi *et al.* (1999).

1.2. Technical background: logistic growth and cubic-type diffusion processes

As the stochastic process introduced in this paper is addressed in relation with other cubic-type diffusion processes, which are generated in the context of the randomisation of deterministic logistic curves, we shall limit ourselves, henceforth, to the logistic case.

When the logistic curve is used, for example, to describe secular trends in demography, economics or environmental sciences, the continuous logistic curve is fitted with non-statistical or asymptotic statistical criteria, based on data that are often recorded at equally spaced time intervals (months, quarters, natural years). Various methods have been used to carry out these fits. A paradigmatic example of application of the deterministic logistic model, extensions and chronological evolution of the estimation methodology is the case of the 'tractors in Spain' introduced by Mar-Molinero (1980) and continued by Oliver (1981), Harvey (1984), Meade (1985), Oliver (1987), Gamerman and Migon (1991) and Franses (2002).

When these fits are performed, often the fitted logistic curve does not reflect variations or certain anomalies in the trends which may be present in the data observed due to the influence of exogenous factors on the endogenous variable under consideration. As a consequence, there tend to occur hard-to-explain discrepancies between the fitted model and the observed data.

For example, in Modis (1988), the three-parameter logistic model

$$P(t) = M[1 + \exp(-\alpha(t - t_0))]^{-1} \quad t \geq t_0 \quad (1)$$

where $P(t)$ is the number of Nobel prizes accumulated during the year t in USA, is fitted by nonlinear least squares and by using numerical computation methods. The methodology and discussion of corresponding discrepancies between real and fitted data can be seen in Golden and Zantek (2004).

In the same way that a continuous growth curve, a logistic one in the above case, is fitted to a phenomenon in which the variable studied $P(t)$, obviously, takes only integer values, the stochastic models of diffusion processes that are solutions to certain Ito or Stratonovich SDEs corresponding to stochastic versions of growth phenomena, with such solutions having almost surely continuous sample paths, are normally used in modelling growth phenomena that in many cases take integer values due to the very nature of the variable under consideration.

In particular, Nobile and Ricciardi (1980) and Ricciardi *et al.* (1999) consider an extension of the classical deterministic model of logistic growth proposed by Verhulst. Instead of considering that the growth process $x = x(t)$ is the solution of the DE

$$\frac{dx}{dt} = \alpha x - \beta x^2, \quad x(0) = x_0 \quad (2)$$

where α and $\beta > 0$ are arbitrary parameters and x_0 denotes the number of individuals present at the initial time, the above-mentioned authors consider $x(t)$ to be the solution to the DE

$$\frac{dx}{dt} = -\alpha x^2(x - \gamma), \quad \alpha > 0, \quad \gamma > 0 \quad (3)$$

In both models, α is the ‘intrinsic growth parameter’. Note that $x = \gamma$ is a globally stable equilibrium point.

Nobile *et al.* (1984a) consider different stochastic versions of model (3), based on SDEs obtained by identifying the intrinsic growth rate α with the stochastic process $\alpha + \Lambda(t)$, where $\Lambda(t)$ is white noise, that is $\Lambda(t)$ is a stationary Gaussian process with zero mean and delta-type correlation function: $\mathbb{E}(\Lambda(t)) = 0$ and $\mathbb{E}(\Lambda(t_1)\Lambda(t_2)) = \sigma^2\delta(t_1 - t_2)$.

Under the latter hypothesis, in fact, these authors propose various SDEs whose solutions in X_t are stochastic diffusion processes that can be characterised (see, for example, Wong and Hajek (1985)) by their infinitesimal moments (the infinitesimal mean or drift and the diffusion parameter). Specifically, they consider the diffusion for which the infinitesimal moments are of the following type:

$$A_1(x) = \xi x^2 - \alpha x^3, \quad A_2(x) = \sigma^2 x^4 \quad (4)$$

corresponding to the SDE

$$dX_t = (\xi X_t^2 - \alpha X_t^3)dt + X_t^2 d\Lambda(t) \quad (5)$$

This cubic-type diffusion encounters a number of problems which are an important drawback in relation to fitting and predicting the process. As observed by Nobile and Ricciardi (p. 184, 1984a, 1984b), it is not possible to calculate the transition density of the process, and neither are its moments known, especially the trend function of the process and the conditional trend function. As these expectations are the ones that, once estimated from observed data on the phenomenon studied, constitute the basis for statistical fitting and for the prediction methodology, respectively, this cubic-type diffusion model is not suitable for practical applications. On the other hand, it is possible to obtain interesting theoretical results for this process, for instance, concerning first-passage time densities. In the light of these considerations, we conclude the process to be mainly suited to theoretical modelling, for example in neurology or theoretical biology.

To avoid the above-described problems, the present paper introduces a stochastic diffusion model that is also of a cubic type (the infinitesimal mean is of cubic order), and such that it is possible to explicitly obtain its stationary density and the expectation functions (see Section 2). In this situation, hence, it is possible to obtain the maximum likelihood (MC) estimate of its parameters and therefore obtain the estimated expectations (see Section 3). Consequently, we are able to establish a statistical methodology for fitting and prediction on the basis of the observations made of the process X_t . This statistical methodology is applied (see Section 4) to a real case in order to highlight the practical possibilities of the diffusion model described in this paper.

2. THE STOCHASTIC CUBIC DIFFUSION MODEL

We consider a time-homogeneous diffusion process $\{X_t; t \in [t_0, T]\}$, with values in $(0, \infty)$ and infinitesimal moments (drift and diffusion coefficient, respectively) given by

$$A_1(x) = Ax^3 + Bx, \quad A_2(x) = c^2x^4 \quad (6)$$

It can be easily proved that the functions $A_1(x)$ and $A_2(x)$, $0 < x < \infty$, are Borel measurable and satisfies the uniform Lipschitz and the growth conditions (see, for example, Wong and Hajek (1985), proposition 4.1). Consequently, exists a separable, measurable and almost surely sample continuous process $\{X_t; t \in [t_0, T]\}$ which is the unique (a.s.) solution of the Ito differential equation

$$dX_t = (AX_t^3 + BX_t)dt + cX_t^2dW_t \quad (7)$$

with the initial condition $P[X_{t_0} = x_{t_0}] = 1$ and where W_t is a standard Wiener process.

By other hand, the functions (6) satisfies the growth and γ -Holder conditions (see, for example, Wong and Hajek (1985), proposition 7.1). Consequently, the transition density function $f(y, t|x, s)$ of the homogeneous diffusion process unique solution of the equation (7), it is the unique fundamental solution of the forward (Fokker–Planck) equation

$$\frac{\partial f}{\partial t} = -\frac{\partial[A_1(y)f]}{\partial y} + \frac{1}{2}\frac{\partial^2[A_2(y)f]}{\partial y^2}$$

with the initial condition $f(y, t|x, t) = \delta(y - x)$.

Given the forms of the infinitesimal moments equation (6), the above-defined process is of cubic type.

Furthermore, there is a relationship between this type of diffusion and Rayleigh's homogeneous diffusion process. Indeed (see Giorno *et al.* (1986), Gutiérrez *et al.* (2006b)) by taking Ito's SDE for a Rayleigh diffusion process $\{Y_t, t \in [t_0, T]\}$ of the form

$$dY_t = \left(\frac{a}{Y_t} + bY_t \right) dt + \sigma dW_t \tag{8}$$

with the initial condition $P[Y_{t_0} = y_{t_0}] = 1$.

We could consider the inverse process $X_t = \frac{1}{Y_t}$ and, by applying Ito's formula, show that the process X_t is the stochastic cubic diffusion process (SCDP) defined in equation (7) and that the parameters of this process are related to the Rayleigh ones by $A = \sigma^2 - a$, $B = -b$ and $c = -\sigma$.

2.1. Theoretical trend functions of the SCDP model

It is not possible to obtain, in an explicit form, the transition density of the SCDP, as is also the case with the Rayleigh process and with the cubic-type process introduced by Giorno *et al.* (1986). On the other hand, we can calculate trend and conditional trend functions, that is the mathematical expectations (CTFs) $\mathbb{E}(X_t)$ and $\mathbb{E}(X_t|X_s = x_s)$, respectively. The calculation of these trend functions is fundamental for the analysis of the trends of real phenomena which are statistically fitted by the SCDP model proposed in this paper (see Section 3).

Making use of the above transform, $X_t = \frac{1}{Y_t}$, we obtain

$$\mathbb{E}(X_t|X_s = x_s) = \mathbb{E}(1/Y_t|1/Y_s = 1/y_s) = \mathbb{E}(1/Y_t|Y_s = y_s) \tag{9}$$

Additionally, taking conditional expectations in Equation (8), we obtain

$$\frac{d}{dt} [\mathbb{E}(Y_t|Y_s = y_s)] = a\mathbb{E}(1/Y_t|Y_s = y_s) + b\mathbb{E}(Y_t|Y_s = y_s) \tag{10}$$

From Equations (9) and (10), it is deduced that

$$\mathbb{E}(X_t|X_s = x_s) = \frac{1}{a} \frac{d}{dt} [\mathbb{E}(Y_t|Y_s = y_s)] - \frac{b}{a} \mathbb{E}(Y_t|Y_s = y_s) \tag{11}$$

The expression of the CTF of the process Y_t (see Gutiérrez *et al.*, 2006) is given by

$$\mathbb{E}(Y_t|Y_s = y_s) = \frac{\Gamma(\alpha + 3/2)}{\Gamma(\alpha + 1)} \left(\frac{b}{\sigma^2(e^{2b(t-s)} - 1)} \right)^{-1/2} \Phi \left(-\frac{1}{2}, \alpha + 1, \frac{-by_s^2}{\sigma^2(1 - e^{-2b(t-s)})} \right)$$

where Φ is the Kummer function and $\alpha = \frac{a}{\sigma^2} - \frac{1}{2} = \frac{1}{2} - \frac{A}{c^2}$.

To simplify computation of these trends, let us apply the following notations:

$$k(t) = \frac{b}{\sigma^2} (e^{2b(t-s)} - 1)^{-1}, \text{ and } r(\alpha) = \frac{\Gamma(\alpha + 3/2)}{\Gamma(\alpha + 1)}$$

The above expression can then be rewritten as

$$\mathbb{E}(Y_t|Y_s = y_s) = r(\alpha)k^{-1/2}(t)\Phi\left(-\frac{1}{2}, \alpha + 1, -y_s^2z(t)\right)$$

where $z(t) = k(t) + b/\sigma^2$, and deriving this function with respect to t and using the relation (Lebedev, 1972, p. 261: 9.9.4), we obtain

$$\frac{d}{dz}[\Phi(\mu, \nu, z)] = \frac{\mu}{\nu}[\Phi(\mu + 1, \nu + 1, z)].$$

Thus, we have

$$\begin{aligned} \frac{d}{dt}[\mathbb{E}(Y_t|Y_s = y_s)] &= -\frac{1}{2}r(\alpha)k'(t)k^{-3/2}(t)\Phi\left(-\frac{1}{2}, \alpha + 1, -y_s^2z(t)\right) \\ &\quad - \frac{y_s^2}{2(\alpha + 1)}r(\alpha)k'(t)k^{-1/2}(t)\Phi\left(\frac{1}{2}, \alpha + 2, -y_s^2z(t)\right) \end{aligned}$$

Since $k'(t) = -2\sigma^2k(t)z(t)$, we find that

$$\begin{aligned} \frac{d}{dt}[\mathbb{E}(Y_t|Y_s = y_s)] &= \sigma^2r(\alpha)k^{-1/2}(t)z(t)\Phi\left(-\frac{1}{2}, \alpha + 1, -y_s^2z(t)\right) \\ &\quad - \frac{\sigma^2y_s^2}{(\alpha + 1)}r(\alpha)k^{1/2}(t)z(t)\Phi\left(\frac{1}{2}, \alpha + 2, -y_s^2z(t)\right) \end{aligned}$$

By substitution in Equation (11), we obtain

$$\begin{aligned} \mathbb{E}(X_t|X_s = x_s) &= \frac{\sigma^2}{a}r(\alpha)k^{-1/2}(t)z(t)\Phi\left(-\frac{1}{2}, \alpha + 1, -\frac{z(t)}{x_s^2}\right) \\ &\quad - \frac{\sigma^2y_s^2}{a(\alpha + 1)}r(\alpha)k^{1/2}(t)z(t)\Phi\left(\frac{1}{2}, \alpha + 2, -\frac{z(t)}{x_s^2}\right) \\ &\quad - \frac{b}{a}r(\alpha)k^{-1/2}(t)\Phi\left(-\frac{1}{2}, \alpha + 1, -\frac{z(t)}{x_s^2}\right) \\ &= \frac{\sigma^2}{a}r(\alpha)k^{1/2}(t)\left[\Phi\left(-\frac{1}{2}, \alpha + 1, -\frac{z(t)}{x_s^2}\right) \right. \\ &\quad \left. - \frac{z(t)/x_s^2}{(\alpha + 1)}\Phi\left(\frac{1}{2}, \alpha + 2, -\frac{z(t)}{x_s^2}\right)\right] \end{aligned}$$

By using the recurrence formula (Lebedev, 1972, p. 262: 9.9.12)

$$\Phi(\mu, \nu, z) = \Phi(\mu + 1, \nu, z) - \frac{z}{\nu} [\Phi(\mu + 1, \nu + 1, z)]$$

then

$$\mathbb{E}(X_t | X_s = x_s) = \frac{\sigma^2}{a} r(\alpha) k^{1/2}(t) \Phi\left(\frac{1}{2}, \alpha + 1, -z(t)/x_s^2\right)$$

Finally, the CTF of the SCDP can be expressed as follows:

$$\begin{aligned} \mathbb{E}(X_t | X_s = x_s) &= \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha + 1)} \left(\frac{B}{c^2(1 - e^{-2B(t-s)})} \right)^{1/2} \\ &\quad \times \Phi\left(\frac{1}{2}, \alpha + 1, \frac{-B/x_s^2}{c^2(e^{2B(t-s)} - 1)}\right) \end{aligned} \quad (12)$$

With the initial distribution $P(X_{t_0} = x_{t_0}) = 1$, the expression of the trend function of the SCDP is

$$\begin{aligned} \mathbb{E}(X_t) &= \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha + 1)} \left(\frac{B}{c^2(1 - e^{-2B(t-t_0)})} \right)^{1/2} \\ &\quad \times \Phi\left(\frac{1}{2}, \alpha + 1, \frac{-B/x_{t_0}^2}{c^2(e^{2B(t-t_0)} - 1)}\right) \end{aligned} \quad (13)$$

The right-continuity of $\mathbb{E}(X_t)$ at t_0 can be proven as follows: from the Kummer transform, $\Phi(\mu, \nu, z) = e^z \Phi(\nu - \mu, \nu, z)$, and the relationship 47-9.6 in Sepanier and Oldham (1987), for very large, positive z and $a \neq 0, -1, -2, \dots$, the approximation $\Phi(a, b, z) \sim \frac{\Gamma(b)}{\Gamma(a)} z^{a-b} \exp(z)$ holds, and thus

$$\lim_{t \rightarrow t_0^+} \mathbb{E}(X_t) = x_{t_0}$$

2.2. Steady-state distribution of the process

We consider in this Section that the time-domain of the process $\{X_t \in [t_0, T]\}$, is $t \in [t_0, +\infty)$. The steady-state distribution $S(x)$ of the SCDP is given by

$$S(x) = \frac{k}{A_2(x)} \exp \left[2 \int_z^x \frac{A_1(y)}{A_2(y)} dy \right]$$

where z is an arbitrary point in the interval $(0, +\infty)$, $A_1(x)$ and $A_2(x)$ are the drift and the diffusion coefficient of the process (Equation (6)), and k is a normalising constant that is given by

$$k = \left[\int_0^{+\infty} \frac{1}{A_2(x)} \exp \left(2 \int_z^x \frac{A_1(y)}{A_2(y)} dy \right) dx \right]^{-1}.$$

One can show, using the identity (Gradshteyn and Ryzhik, 1988, p. 317)

$$\int_0^{+\infty} x^{\nu-1} e^{-\mu x} dx = \mu^{-\nu} \Gamma(\nu),$$

for $\nu > 0$ and $\mu > 0$, that the steady-state distribution of the SCDP is

$$S(x) = \frac{2 \left(\frac{B}{c^2}\right)^{\alpha+1} x^{-2\alpha-3} e^{-\frac{B}{c^2 x}}}{\Gamma(\alpha+1)}, \text{ for } B > 0 \text{ and } \alpha + 1 > 0.$$

Let X be the random variable with density function $S(x)$. It can be easily proved that the random variable $Z = X^2$ is distributed according to Gamma distribution with parameters $(\alpha + 1; c^2/B)$.

The higher-order asymptotic moments of the SCDP can be proved to be expressed as

$$\mathbb{E}(X_{\infty}^j) = \frac{\Gamma(\alpha + 1 - \frac{j}{2})}{\Gamma(\alpha + 1)} \left(\frac{B}{c^2}\right)^{j/2}, \text{ for } B > 0 \text{ and } j < 2\alpha + 2.$$

The asymptotic trend function of the model is

$$\mathbb{E}(X_{\infty}) = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha + 1)} \left(\frac{B}{c^2}\right)^{1/2}, \quad (14)$$

for $B > 0$ and $2\alpha + 1 > 0$, and the asymptotic variance is

$$\text{Var}[X_{\infty}] = \frac{B}{c^2} \left(\frac{1}{\alpha} - \frac{\Gamma^2(\alpha + \frac{1}{2})}{\Gamma^2(\alpha + 1)} \right), \text{ for } B > 0 \text{ and } \alpha > 0.$$

It can be seen that the limit of the trend function in Equation (13) (when t tends to ∞) coincides with the asymptotic trend function in Equation (14).

3. ESTIMATION AND COMPUTATION OF PARAMETERS

We shall now obtain the ML estimators of the drift parameters (A and B) on the basis of a continuous sampling of the SCDP, indicating the computational aspects involved in this calculation. The diffusion parameter c is computed by an indirect method.

3.1. Estimation of A and B

A similar proof to the one developed in Gutiérrez *et al.* (1991) allows us to ensure that the process defined by Equation (7) satisfies the conditions (see Lipster and Shiryaev (1978) p. 202) under which the measures corresponding to continuous observation of the solution of Equation (7) for different values of the parameter (A, B) are equivalent and then the continuous-time log-likelihood function is

$$\mathbb{L}(A, B) = \frac{1}{c^2} \int_{t_0}^T \frac{AX_t^3 + BX_t}{X_t^4} dX_t - \frac{1}{2c^2} \int_{t_0}^T \frac{(AX_t^3 + BX_t)^2}{X_t^4} dt$$

By applying the ML methodology, that is deriving with respect to the parameters and setting these derivatives equal to zero, and after various operations, the resulting ML estimators are found to be

$$\hat{A} = \frac{\int_{t_0}^T \frac{dt}{X_t^2} \int_{t_0}^T \frac{dX_t}{X_t} - (T - t_0) \int_{t_0}^T \frac{dX_t}{X_t^3}}{\int_{t_0}^T X_t^2 dt \int_{t_0}^T \frac{dX_t}{X_t^2} - (T - t_0)^2}$$

$$\hat{B} = \frac{\int_{t_0}^T X_t^2 dt \int_{t_0}^T \frac{dX_t}{X_t^3} - (T - t_0) \int_{t_0}^T \frac{dX_t}{X_t}}{\int_{t_0}^T X_t^2 dt \int_{t_0}^T \frac{dX_t}{X_t^2} - (T - t_0)^2}$$

From Ito's formula, the stochastic integrals in the expressions of the parameters are transformable into Riemann integrals, and thus we have

$$\int_{t_0}^T \frac{dX_t}{X_t} = \log(X_T) - \log(x_{t_0}) + \frac{c^2}{2} \int_{t_0}^T X_t^2 dt$$

$$\int_{t_0}^T \frac{dX_t}{X_t^3} = \frac{1}{2x_{t_0}^2} - \frac{1}{2X_T^2} + \frac{3c^2}{2} (T - t_0)$$

From the above expressions, the MLM estimators are found to be

$$\hat{A} = \frac{\left(\log\left(\frac{X_T}{x_{t_0}}\right) + \frac{c^2}{2} \int_{t_0}^T X_t^2 dt \right) \int_{t_0}^T \frac{dt}{X_t^2} - \frac{(T-t_0)}{2} \left(\frac{1}{x_{t_0}^2} - \frac{1}{X_T^2} + 3c^2(T - t_0) \right)}{\int_{t_0}^T X_t^2 dt \int_{t_0}^T \frac{dX_t}{X_t^2} - (T - t_0)^2} \tag{15}$$

$$\hat{B} = \frac{\frac{1}{2} \left(\frac{1}{x_{t_0}^2} - \frac{1}{x_T^2} + 3c^2(T - t_0) \right) \int_{t_0}^T X_t^2 dt - (T - t_0) \left(\log \left(\frac{X_T}{x_0} \right) + \frac{c^2}{2} \int_{t_0}^T X_t^2 dt \right)}{\left(\int_{t_0}^T X_t^2 \right) dt \int_{t_0}^T \frac{dX_t}{X_t^2} - (T - t_0)^2} \quad (16)$$

In practice, as full observations of a continuous sample path of the process is not available, we must consider approximations based on the discrete observations of the process at times $t(0) = t_0, \dots, t(n) = T$ (discrete sampling). A suitable computational method for this situation consists of approximating the Riemann integrals in Equations (15) and (16) by, for example, the trapezoidal method.

3.2. Approximation of the parameter c

The parameter c in the diffusion coefficient $A_2(x)$ of the model can be estimated by an indirect method based on an extension of the procedure described by Chesney and Elliot (1995) for the case of an SDE with linear drift and multiplicative noise (as previously used, for example, by Gutiérrez *et al.* (2006)). This extension is described as follows. From Ito's formula, we have

$$d\left(\frac{1}{X_t}\right) = -\frac{dX_t}{X_t^2} + c^2 X_t dt$$

Let us use the following approximations between $t - 1$ and t of these differentials:

$$d\left(\frac{1}{X_t}\right) \simeq \frac{1}{X_t} - \frac{1}{X_{t-1}}, \text{ and } dX_t \simeq X_t - X_{t-1}$$

Therefore, a first numerical approximation of c , based on the pair (X_{t-1}, X_t) , is

$$\hat{c}_{(t-1,t)} = \frac{|X_t - X_{t-1}|}{X_t \sqrt{X_t X_{t-1}}}$$

We thus deduce that for $n + 1$ observations of a given process sample path, an estimator of c is given by the following expression:

$$\hat{c} = \frac{1}{n} \sum_{t=1}^n \frac{|X_t - X_{t-1}|}{X_t \sqrt{X_t X_{t-1}}} \quad (17)$$

4. SIMULATION AND APPLICATION

4.1. Simulation

The realisations of this process can be obtained by using Taylor's algorithm in the order 1.5 strong Taylor scheme in Kloeden and Platen (1992) (p. 351). In this case, the algorithm is given by

$$\begin{aligned}
 x_{n+1} = x_n + [(A - c^2)x_n^3 + Bx_n]h + \frac{h^2}{2} [(Ax_n^3 + Bx_n)(3Ax_n^2 + B) + 3Ac^2x_n^5] \\
 + cx_n^2\Delta W \left[1 + cx_n\Delta W + c^2x_n^2(\Delta W)^2 \right] + cx_n[(A - c^2)x_n^3 - Bx_n]\Delta Z \\
 + 2cx_n[(A - c^2)x_n^3 + Bx_n]h\Delta W
 \end{aligned}$$

where $\Delta W = \sqrt{h}U_1$ and $\Delta Z = \frac{h^{3/2}}{2}(U_1 + U_2/\sqrt{3})$, with U_1 and U_2 being two standard normal independent random variables, and h is the step of discretisation.

4.2. Application to a real case: analysis of the trend of global CO₂ emission in Spain

In this application, we examine the variable X_t defined by ‘global CO₂ emission from fossil-fuel burning, cement manufacture, and gas flaring in Spain’, in the period 1986–2002. The global data for each year have been extracted from the historical series (1751–2002) available by year, country and region. The methodological information on the estimation of these series can be consulted in Boden *et al.* (1995) and Andres *et al.* (1999), among others.

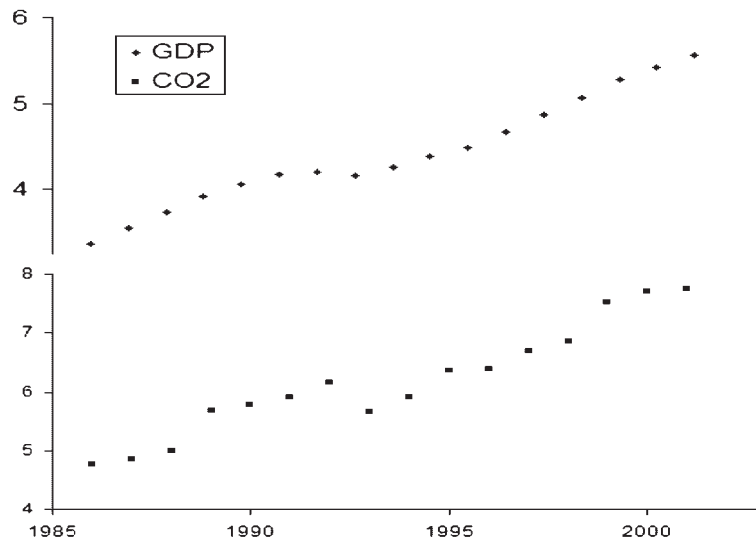
In this study, we analyse the trend of the above-mentioned global CO₂ emission in Spain, by fitting the observed data to the trend function (TF) and the CTF of a cubic-type stochastic diffusion model, as introduced in the above paragraphs.

The data in Table 1 are expressed in thousand millions of metric tons of carbon and are taken from Marland *et al.* (2005); these data may be consulted at <http://cdiac.esd.ornl.gov/ftp/ndp030/global.1751-2002.ems>.

Figure 1 shows the data corresponding to the period 1986–2002 of the variable X_t , together with the observed data for gross domestic product (GDP) in Spain. The computation of the estimators of the drift parameters given in Equations (15) and (16), together with the approximation of the estimator of the diffusion coefficient given by Equation (17), was carried out using programs developed in

Table 1. Fit and prediction

Year	Real value	ETF	ECTF
1986	4.7610	4.7610	4.7610
1987	4.8570	4.8690	4.8690
1988	4.9856	4.9837	4.9707
1989	5.6841	5.1058	5.1074
1990	5.7814	5.2362	5.8569
1991	5.9096	5.3760	5.9624
1992	6.1657	5.5264	6.1018
1993	5.6504	5.6891	6.3819
1994	5.9089	5.8660	5.8205
1995	6.3672	6.0593	6.1012
1996	6.3732	6.2723	6.6037
1997	6.6961	6.5090	6.6103
1998	6.8645	6.7749	6.9687
1999	7.5230	7.0779	7.1571
2000	7.7099	7.4282	7.9043
2001	7.7470	7.8336	8.1197
2002	8.2998	8.2811	8.1626

Figure 1. GDP versus CO₂ emission

Mathematica 5.1. The estimators calculated were as follows: $\hat{A} = 0.0007422$, $\hat{B} = 0.0052121$ and $\hat{c} = 0.0053286$.

Table 1 shows the observed values and those estimated for the TF and the CTF (i.e. the ETF and the ECTF, respectively) for the corresponding years.

The original data and the corresponding data fitted by the ETF and the ECTF are shown in Figures 2 and 3, respectively.

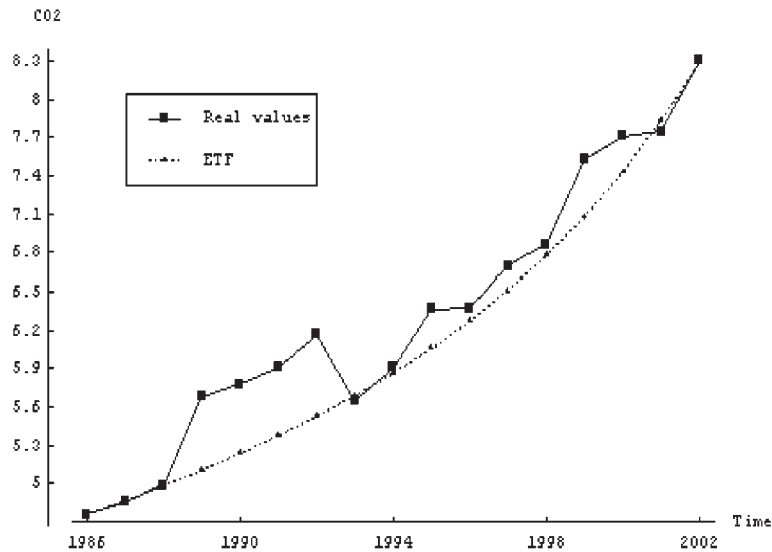


Figure 2. Real data versus ETF

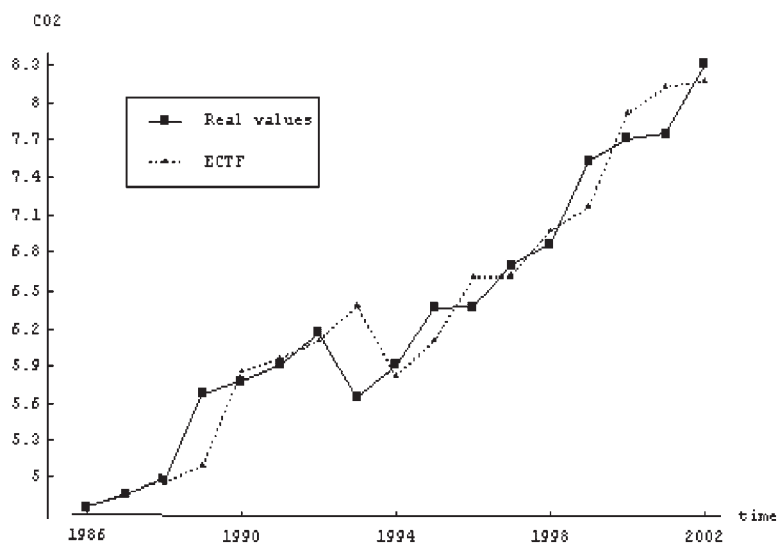


Figure 3. Real data versus ECTF

Figure 4 shows:

- (1) The simulation of 10 sample paths of the SCDP with parameters taken from the neighbourhood of the estimators obtained for the real case studied, i.e. $A = 0.00074$, $B = 0.0052$, $c = 0.0053$, with a discretisation step of $h = 0.1$ and an initial value $x_0 = 4.7$. This simulation was calculated according to the procedure described in Section 4.1.
- (2) The estimated trend function (ETF) fitted to the data for global CO₂ emission in Spain.

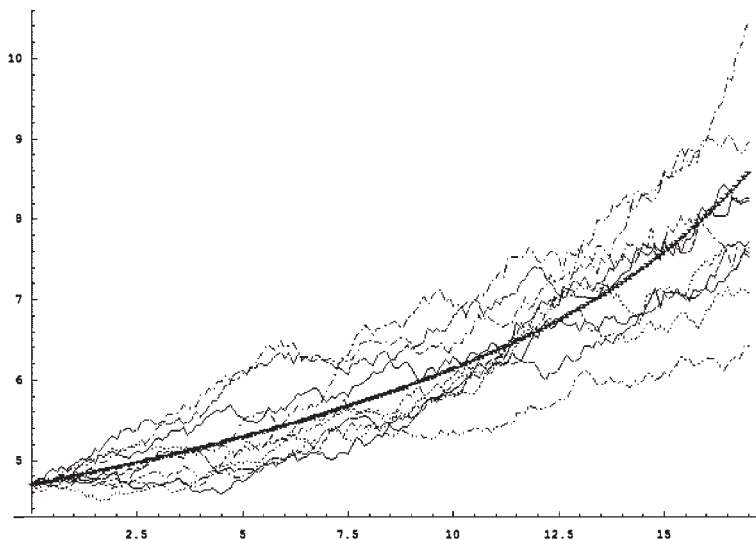


Figure 4. Simulated sample paths

5. DISCUSSION AND CONCLUSIONS

The SCDP model introduced in this paper is characterised, firstly, by its flexibility, in that its sample paths may be increasing or decreasing, as are their corresponding trends (see Subsection 4.1). Growth, for example, is at a more restrained rate than that of other increasing diffusions such as lognormal or Gompertz. Secondly, the model makes it possible to calculate the steady-state distribution, and thus to analyse the long-term behaviour of the trends (t tends to infinity). Moreover, it presents the particular feature that the trend functions (TF and CTF) can be calculated either theoretically (Subsection 2.1), estimated and computed in practice (Section 3).

The SCDP model introduced in this paper comprises a suitable alternative for real cases in which lognormal or Gompertz diffusions do not provide a good fit to the observed trends. Indeed, the respective fits performed for the global CO₂ emission in Spain data, following the latter methodology (see Gutiérrez *et al.* (1991, 2005a,b), Skiadas and Giovani (1997), are not acceptable. The SCDP model, furthermore, enables us to highlight groups of outliers in the trend for CO₂ emission in Spain, which are explicable in terms of abnormally high annual increases in the GDP in Spain for the corresponding years. This latter fact suggests that there is a high degree of correlation between the GDP and the emission of CO₂ in Spain.

The ECTF, which provides a fairly accurate fit to the real data (see Figure 3), constitutes an appropriate mechanism for the short- and medium-term prediction of the values of global emission of CO₂ in Spain. Indeed, if the fit carried out in Section 4, which utilises all the data observed for the period 1986–2002, was performed with the data for 1986–2001, a prediction for 2002 would be obtained by the ECTF of $X_{2002} = 8.1626$, while the real value observed for this year was 8.2998.

In principle, a nonparametric approach to the problem of statistical fitting of the model considered in this work is also possible (see, Fan, 2005). The authors are currently investigating this alternative methodology.

ACKNOWLEDGEMENTS

The authors are very grateful to the Editor and referees for constructive comments and suggestions. The authors are grateful to Dr. Jose M. Angulo (University of Granada) for his careful reading and excellent comments. This work was supported partially by research project MTM 2005-09209, Ministerio de Educación y Ciencia, Spain.

REFERENCES

- Andres RJ, Fielding DJ, Marland G, Boden TA, Kumar N. 1999. Carbon dioxide emissions from fossil-fuel use, 1751–1950. *Tellus B* **51**: 759–765.
- Boden TA, Marland G, Andres RJ. 1995. Estimates of global, regional and national annual CO₂ emissions from fossil-fuel burning, hydraulic cement production, and gas flaring: 1950–1992. Technical Report: *ORNL/CDIAC-90,NDP-30/R6*. Oak Ridge National Laboratory, U.S. Department of Energy: Oak Ridge, Tennessee, USA.
- Capocelli RM, Ricciardi LM. 1974. A diffusion model for population growth in random environments. *Theoretical Population Biology* **5**(1): 28–41.
- Chesney M, Elliot J. 1995. Estimating the instantaneous volatility and covariance of risky assets. *Applied Stochastic Models and Data Analysis* **11**: 51–58.
- Fan J. 2005. A selective overview of nonparametric methods in financial econometrics. *Statistical Sciences* **20**(4): 317–337.
- Franses PH. 2002. Testing for residual autocorrelation in growth curve models. *Technological Forecasting and Social Change* **69**: 195–204.

- Gamerman D, Migon HS. 1991. Tractors in Spain—a dynamic reanalysis. *Journal of the Operational Research Society* **42**(2): 119–124.
- Giorno V, Nobile AG, Ricciardi LM, Sacerdote L. 1986. Some remarks on the Rayleigh process. *Journal of Applied Probability* **23**: 398–408.
- Golden BL, Zantek PF. 2004. Inaccurate forecasts of the logistic growth model for Nobel prizes. *Technological Forecasting and Social Change* **71**: 417–422.
- Gradshteyn IS, Ryzhik IM. 1988. *Table of Integrals, Series and Products*. Academic Press: San Diego.
- Gutiérrez R, Angulo JM, González A, Pérez R. 1991. Inference in lognormal multidimensional diffusion processes with exogenous factors: application to modelling in Economics. *Applied Stochastic Models and Data Analysis* **7**: 295–316.
- Gutiérrez R, Ricciardi LM, Román R, Torres P. 1997. First-passage-time densities for time-non-homogeneous diffusion processes. *Journal of Applied Probability* **34**(3): 623–631.
- Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A, Román P, Torres F. 2005a. Inference in Gompertz type non-homogeneous stochastic systems by means of discrete sampling. *Cybernetics and Systems* **36**: 203–216.
- Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. 2005b. Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. *Applied Energy* **80**(2): 115–124.
- Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. 2006a. Consumption electricity in Morocco: stochastic diffusion analysis with exogenous factors. *Applied Energy* (in press). Available on line at: www.sciencedirect.com/science/journal/03062619
- Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. 2006b. The Stochastic Rayleigh diffusion model: statistical inference and computational aspects. Applications to modelling of real cases. *Applied Mathematics and Computation* **175**: 628–644.
- Harvey AC. 1984. Time series forecasting based on the logistic curve. *Journal of the Operational Research Society* **32**: 641–646.
- Kloeden P, Platen E. 1992. *The Numerical Solution of Stochastic Differential Equations*. Springer: Berlin.
- Lebedev NN. 1972. *Special Functions and Their Applications*. Dover publications, Inc: New York.
- Lipter RS, Shirayev AN. 1978. *Statistics of Random Processes II. Applications*. Springer-Verlag: New York.
- Marland G, Boden TA, Andres RJ. 2005. Global, regional, and national fossil fuel CO₂ emissions. In *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information*. Analysis Center/Department of Energy: Oak Ridge National Laboratory, U.S. Oak Ridge, TN, USA.
- Mar-Molinero C. 1980. Tractors in Spain: a logistic analysis. *Journal of the Operational Research Society* **31**: 141–152.
- Meade N. 1985. Forecasting using growth curves. An adaptive approach. *Journal of Operational Research Society* **36**(12): 1103–1115.
- Modis T. 1988. Competition and forecasts for the Nobel prizes awards. *Technological Forecasting and Social Change* **34**(2): 95–102.
- Nobile AG, Ricciardi LM. 1980. Growth and extinction in random environments. In *Application of Information and Control Systems*, Lainiotis DG, Tzannes NS (eds). Reidel Pub. Comp.: Dordrecht; 455–465.
- Nobile AG, Ricciardi LM. 1984a. Growth with regulation in fluctuating environments. *Biological Cybernetics* **49**: 179–188.
- Nobile AG, Ricciardi LM. 1984b. Growth with regulation in fluctuating environments. II. Intrinsic lower bounds to population size. *Biological Cybernetics* **50**: 285–299.
- Oliver FM. 1981. Tractors in Spain: a further logistic analysis. *Journal of Operational Research Society* **32**: 499–502.
- Oliver FM. 1987. A Bayesian model to predict saturation and logistic growth. *Journal of Operational Research Society* **38**: 49–56.
- Ricciardi LM, Crescenzo A, Giorno V, Nobile AG. 1999. An outline of theoretical and algorithmic approaches to first passage time problems with applications to biological modeling. *Mathematical Japonica*. **50**(2): 247–322.
- Sepanier J, Oldham KB. 1987. *An Atlas of Functions*. Springer-Verlag: New York.
- Skias CH, Giovanni AN. 1997. A stochastic bass innovation diffusion model for studying the growth of electricity consumption in Greece. *Applied Stochastic Models and Data Analysis* **13**: 85–101.
- Wong E, Hajek B. 1985. *Stochastic Processes in Engineering Systems*. Springer-Verlag: New York.