I. Troch, F. Breitenecker, eds. ISBN 978-3-901608-35-3

STOCHASTIC MODELING IN THE TASK OF BIOCHEMICAL OXYGEN DEMAND MONITORING

Darya Filatova[1], M. Grzywaczewski[2], M. Zili[3]
[1]Analytical Center, Russian Federation, [2] Politechnika Radomska, Poland, [3] Preparatory Institute for Military Academies, Tunisia

Corresponding author: Daria Filatova, Analytical Center of RAS
119991 Moscow, Vavilova 40, Russian Federation, daria_filatova@rambler.ru

Abstract. The environmental monitoring, analysis, forecast and control are very important and highly complicated problems as far as ecological systems show stochastic nature and high dimensionality. So, the choice of a suitable mathematical model depends on the available data and consequently on an estimation method. The paper presents the new analytical method for controlling pollutants, namely for the monitoring of biochemical oxygen demand.

1 Introduction

Assume that the mathematical model of some nonlinear dynamic system with random noises when multiple observations are available can be presented by the one-dimensional, time-homogeneous stochastic differential equation (SDE), which in a filtered probability space has the form

$$dX_t^j = f(t, X^j, \theta_1)dt + g(t, X^j, \theta_2)dW_i^j, \tag{1}$$

where $t \in [t_0, T]$ is the independent variable with the fixed final value $T \in (t_0, \infty)$ and the initial value $t_0 \in [0, \infty)$, $\theta = [\theta_1, \theta_2]$ is a $(p_1 + p_2)$ - dimensional vector of unknown parameters, $X = X(t, \theta)$ is a state variable vector depending on t and θ, and the coefficient function maps $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^{p_1}$ into $\mathbb{R}$ and the coefficient function maps into $\mathbb{R}$, dW_t is the increment of A-adapted Wiener process W and $J = 1, 2, ..., M$ $(M \in \mathbb{N})$ is a number of a sample path.

For the SDE (1) are defined initial values and parameters restrictions

$$c_q(X(t_0), \theta) \leq 0 \tag{2}$$

where $q = 1, 2, ..., Q,\ Q \in \mathbb{N},$, is the number of restrictions.

In addition, we suppose that the SDE (1) has a strongly unique solution $X = \{X_t, t \in [t_0, T]\}$ on $t \in [t_0, T]$ with $\sup \mathbb{E}\left[|X_t|^2\right]$ ($\mathbb{E}[\cdot]$ is the mathematical expectation operator) if

I) the coefficient functions f and g are assumed to be jointly - L_T^2 measurable in $(t, x) \in [t_0, T] \times \mathbb{R}$;

II) there exists a finite constant $K_1 > 0$ such that $|f(t, x, \theta_1) - f(t, y, \theta_1)| \leq K_1 |x - y|$ and $|g(t, x, \theta_2) - g(t, y, \theta_2)| \leq K_1 |x - y|$ for all $t \in [t_0, T]$ and $x, y \in \mathbb{R}$;

III) there exists a constant $K_2 > 0$ such that $|f(t, x, \theta_1)|^2 \leq K_2(1 + |x|^2)$ and $|g(t, x, \theta_2)|^2 \leq K_2(1 + |x|^2)$ for all $t \in [t_0, T]$ and $x \in \mathbb{R}$;

IV) X_{t_0} is A_{t_0}-measurable with $\mathbb{E}\left[|X_{t_0}|^2\right] < \infty$.

The model (1) presents the panel data and therefore one has to take into account this fact constructing the estimation procedure. There are only a few identification methods which deal with panel data [4]. The main idea of these methods is to find the parameters of the stochastic process probability density function using Chi-square or Kolmogorov-Smirnov criterion functions.

If one takes into account only one sample path of the stochastic process (1), then the method of moments is the simplest computationally way to get estimators of SDE parameters. The theoretical justification for this method relies on the fact that under appropriate assumptions the sample moments are consistent estimators of the respective theoretical moments. However, the unbiasedness of the method of the moment estimators cannot be easily guaranteed. In a contrary, the maximum likelihood estimation method assumes that there is no prior information available on the parameters θ [1]. What is needed for the maximum likelihood estimation is the probability density function of the i.i.d. observations X. The maximum likelihood estimators θ have highly desirable asymptotic optimality properties when the sample size of observed data is large [2]. Unfortunately it is not suitable for BOD time series. Thus the different method has to be constructed. Our goal is to find parameters estimates θ of SDE (1) coefficient functions which satisfy the conditions I) - IV).

2 Dynamic moment equations

To find the parameters of (1) we will transform and present initial panel data by dynamic moment equations. Denote the k^{th} state moment of the stochastic process $X = \{X_t, t \in [t_0, T]\}$ by $m_k(t) = \mathbb{E}\left[(X_t)^k\right]$. To derive dynamic moment equations for SDE (1) we apply the Itô formula

$$u(t,X_t) = \left(f(t,X,\theta_1)\frac{\partial u(t,X_t)}{\partial x} + \frac{1}{2}(g(t,X,\theta_2)^2\frac{\partial^2 u(t,X_t)}{\partial x^2}\right) dt + g(t,X,\theta_2)\frac{\partial u(t,X_t)}{\partial x} dW \tag{3}$$

(where a function $u : [0,\infty) \times \mathbb{R} \twoheadrightarrow \mathbb{R}$ is twice continuously differentiable with respect to the spatial component x), to obtain an SDE for $u(X_t) = (X_t)^k$

$$d\,(X_t)^k = k\,(X_t)^{k-1}\left[f(t,X,\theta_1)dt + g(t,X,\theta_2)dW_1\right] + \frac{1}{2}k(k-1)\,(X_t)^{k-2}\,(g(t,X,\theta_2))^2\,dW_t \tag{4}$$

and then take the expectation on the integral form of this equation having in mind the martingale property of an Itô integral

$$\frac{dm_k(t)}{dt} = k\overline{f}(t,X,\theta_1) + \frac{1}{2}k(k-1)\left[\overline{g}(t,X,\theta_2)\right]^2 dW \tag{5}$$

where $m_k(t_0) = \mathbb{E}\left[(X_{t_0})^k\right]$, $\overline{f}(t,X,\theta_1)$ and $\overline{g}(t,X,\theta_2)$ are transformed coefficient functions.

In particular case, for linear SDE with $f(t,X,\theta_1) = \theta_1 X_t$ and $g(t,X,\theta_2) = \theta_2 X_t$ first and second moment dynamic equations (4) are

$$dm_1(t) = \theta_1 m_1(t)dt \tag{6}$$

$$dm_2(t) = 2\theta_1 m_2(t)dt + \theta_2^2 m_2(t)dt \tag{7}$$

for nonlinear SDE with $f(t,X,\theta_1) = \left(\theta_{11}X_t - \theta_{12}X_t^2\right)$ and $g(t,X,\theta_2) = \theta_2 X_t$ the same moments are

$$dm_1(t) = \left[\theta_{11}m_1(t) - \theta_{12}\,(m_1(t))^2\right] dt \tag{8}$$

$$dm_2(t) = 2\left[(\theta_{11} + \theta_2^2)m_2(t) - \theta_{12}\,(m_2(t))^{3/2}\right] dt \tag{9}$$

where $m_1(t_0) = \mathbb{E}\left[X_{t_0}\right]$ and $m_2(t_0) = \mathbb{E}\left[(X_{t_0})^2\right]$.

3 The identification method

In order to estimate the unknown parameters $\theta = [\theta_1, \theta_2]$, a number of measurements, say n_T, are available for each of trajectories of the stochastic process (1). These $\{Y_n\}_{n=1}^{n_T=1}$ measurements are used in order to estimate the first k^* moments $\widehat{m}_k(t) = \mathbb{E}\left[Y^k\right]$, $k = 1,2,...,k^*$, but, due to the inherent errors in the observations, estimates $\widehat{m}_k$ do not coincide with theoretical values

$$\widehat{m}_k(t_n) = m_k(t_n,\theta) + \varepsilon_n, \qquad n = 1,2,...,n_T \tag{10}$$

where ε_n are independently and identically normally distributed errors due to properties of Wiener process.

As one can see equation (5) is an ordinary differential equation (ODE), so that we can use ideas of estimation procedures for ODEs [3]. For this purpose we introduce a discretization $t_0 < t_1 < ... < t_n < ... < t_{n_T} = T$ of the time interval $[t_0, T]$ with some integer $n_T \geq 2$, denote the linear combination of right-hand side function of ODE (5) for intermediate arguments values by $G(t_n, m_k(t_n,\theta))$and rewrite this equation as

$$0 = \overline{m}_k(t_{n+1},\theta) - \overline{m}_k(t_n,\theta) - \Delta_n G(t_n, \overline{m}_k(t_n,\theta)) \tag{11}$$

where $\overline{m}_k(\cdot,\cdot)$ is an approximation to the solution $m_k(\cdot,\cdot)$ of ODE (5), $\overline{m}_k(t_n,\theta) = \mathbb{E}\left[\left(X_{t_0}\right)^k\right]$, $\Delta_n = t_{n+1} - t_n$ is the length of the time discretization subinterval $[t_n, t_{n+1}]$ (for the simplicity we will consider equidistant time discretization denoting it by Δ), the function $G(t_n, \overline{m}_k(t_n,\theta)) : [0,\infty) \times \mathbb{R} \times \mathbb{R}$ is twice continuously differentiable and $n = 0,1,...,n_{T-1}$.

Using the least squares method, the parameter estimation problem for SDE (1) can be formulated as follows:

$$\arg\min \Phi(\theta) = \frac{1}{2n_T}\sum_{k=1}^{k^*}\sum_{n=1}^{n_T}\left[\overline{m}_k(t_{n_T},\theta) - \widehat{m}_k(t_{n_T})\right]^2 \tag{12}$$

Replications, N	Scheme	$\widehat{\theta}_{11}$	$\widehat{\theta}_{12}$	$\widehat{\theta}_2$
		$std\left(\widehat{\theta}_{11}\right)$	$std\left(\widehat{\theta}_{12}\right)$	$std\left(\widehat{\theta}_2\right)$
25	EU	0.5662	0.5250	0.1632
		0.0254	0.0204	0.0078
	RK	0.5772	0.5097	0.1832
		0.0194	0.0141	0.0066
100	EU	0.5664	0.5101	0.1718
		0.0248	0.0130	0.0062
	RK	0.5964	0.4948	0.1918
		0.0148	0.0109	0.0041
500	EU	0.5711	0.4977	0.1869
		0.0238	0.0131	0.0051
	RK	0.6092	0.5003	0.2007
		0.0143	0.0076	0.0036

Table 1: Results of Monte Carlo experiment: the scheme selection

subjected to (11) and

$$|\overline{m}_k(t_{n_T},\theta) - \widehat{m}_k(t_{n_T})| \leq \Delta^{\gamma} \tag{13}$$

where $\gamma \geq 1$ is the order of the approximation $\overline{m}_k$ convergence to m_k.

Problem (11) - (13) is a typical nonlinear optimization problem, which can be solved by Gauss-Newton method; however, the final results may strongly depend on the structure of the equation (11).

4 Monte Carlo simulation

The objective of this section is to show stability and efficiency properties of our estimation method. The methodology described above has been tested on Workstation in double precision Matlab version 7.0 under Vista Microsoft operation system. In our experiments, the stopping tolerance is set to 10^{-6}.

In biology the following model is often considered

$$dX_t = \left(\theta_{11}X_t - \theta_{12}X_t^2\right)dt + \theta_2 X_t dW_t \tag{14}$$

where X_t is the continuous-time single-species population. Many of the term structure models found in the literature may be nested within this model class by imposing appropriate parameter constraints for a survey.

To generate the observed data, SDE (14) with actual coefficient values $\theta_{11} = 0.6$, $\theta_{12} = 0.5$, and $\theta_2 = 0.2$ with starting value $X(t_0) = 0.3$ was solved numerically by Milstein scheme [5]:

$$Y_{n+1}^j = Y_n^j + Y_n^j\left(\theta_{11} - \theta_{12}Y_n^j\right)\Delta + \theta_2 Y_n^j \Delta W_n^j + \frac{1}{2}\theta_2^2 Y_n^j\left[\left(\Delta W_n^j\right)^2 - \Delta\right] \tag{15}$$

with $Y_0 = X(t_0)$, $j = 1,2,...,M$ (in this experiment $M = 25$) and $n = 0,1,...,n_{T-1}$. The sampling interval $t \in [0,6]$ was divided into $n_T = 48$ time steps of length $\Delta = 2^{-3}$. The Marsaglia method was used in order to simulate the increments ΔW_n of Wiener process [5]. The theoretical values of the moments were also calculated using ODEs (8) - (9) with starting values $m_1(t_0) = X(t_0)$ and $m_2(t_0) = X^2(t_0)$. To study the influence of the equation (11) structure on the final results the explicit Euler scheme (EU) with $\gamma^{EU} = 1.0$ and explicit two-steps Runge-Kutta scheme (RK) with $\gamma^{RK} = 2.0$ were used. Optimization of the problem (11) - (13) with starting values $\theta_{11}^{st} = 0.4$, $\theta_{12}^{st} = 0.4$ and $\theta_2^{st} = 0.4$ were done by Matlab Optimization Toolbox. Table 1 reports the findings for the estimates. As we expected, the estimators are close to the true parameter values. Better results were gotten for Runge-Kutta method. So, we can conclude that the scheme with higher values of γ gives more precise results.

Next we study the property of the estimates, namely if the estimates are asymptotically unbiased. The estimator is called asymptotically unbiased if the bias tends to zero as the number M of observations increases, that is

$$\lim_{M\to\infty} \mathbb{E}\left[\overline{\theta}_i(M)\right] = 0 \tag{16}$$

where $\overline{\theta}_i(M) = \theta_i - \widehat{\theta}_i$ is the estimator error.

Staying with we same conditions for data generation as in previous experiment, using the two-steps Runge-Kutta scheme and $N = 250$ replications on integration interval $t \in [0,6]$ with $\Delta = 0.125$, we use $M = \{10,25,50,100\}$ to check (16). Table 2 reports the mean values of the estimated statistics.

M	$\widehat{\theta}_{11}$ $bias\left(\widehat{\theta}_{11}\right)$ $std\left(\widehat{\theta}_{11}\right)$	$\widehat{\theta}_{12}$ $bias\left(\widehat{\theta}_{12}\right)$ $std\left(\widehat{\theta}_{12}\right)$	$\widehat{\theta}_{2}$ $bias\left(\widehat{\theta}_{2}\right)$ $std\left(\widehat{\theta}_{2}\right)$
10	0.5596 -0.0404 0.0501	0.5115 0.0115 0.0281	0.2253 0.0257 0.0057
25	0.5731 -0.0269 0.0312	0.5090 0.0091 0.0186	0.1868 -0.0135 0.0052
50	0.5867 -0.0133 0.0211	0.4974 -0.0026 0.0119	0.1912 -0.0109 0.0046
100	0.5967 -0.0033 0.0113	0.5014 0.0015 0.0108	0.1940 -0.0061 0.0035

Table 2: Results of Monte Carlo experiment: the unbiasedness

As we can see the estimates of SDE (1) are asymptotically unbiased. This allows concluding that estimation method can be successfully used. We checked the method on the field data received for BOD in Omega Bay in Sevastopol, Ukraine. Finally we got data only from one layer of sensors. The estimation method gave the model

$$dX_t = \left(0.9916X_t - 1.2127X_t^2\right)dt + 0.01276X_t dW_t \tag{17}$$

where X_t is BOD index. Figure 1 illustrates the data from the sensors and forecast results for BOD.

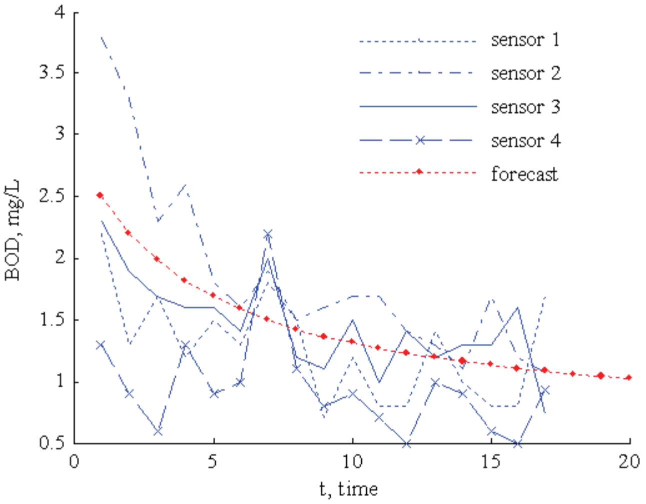

Figure 1: The data for Omega Bay of Sevastopol (Ukraine) and forecast.

5 References

[1] Jang M.-J., Wang J.-S., and Liu Y.-C.: *Applying differential transformation method to parameter identification problems*. Applied Mathematics and Computation, 139 (2003), 491 – 502.

[2] Hansen J.A., and Penland C.: *On stochastic parameter estimation using data assimilation*. Physica D, 230 (2007), 88 – 98.

[3] Li Z., Osborn M.R., and Prvan T.: *Parameter estimation of ordinary differential equations*. IMA Journal of Numerical Analysis, 25 (2005), 264 – 285.

[4] McDonald A.D., and Sandal L.K.: *Estimating the parameters of stochastic differential equations using a criterion function based on the Kolmogorov-Smirnov statistic*. J. Statist. Comput. Simul., 64 (1999), 235 –250.

[5] Platen E., and Heath D.: *A benchmark approach to quantitative finance*. Springer, Berlin, 2007.

[6] Prvan T., and M.R. Osborne: *Model selection in a stochastic setting*. ANZIAM Journal, 45 E (2004), C787 – C799.