# THE MODELLING OF NOISE PROCESSES IN STOCHASTIC DIFFERENTIAL EQUATIONS: APPLICATION TO BIOTECHNOLOGICAL PROCESSES

**S.G.T. Kops, K. Gernaey, O. Thas and P.A. Vanrolleghem** *

* *BIOMATH Department, University of Gent, Coupure Links 653*
*B-9000 Gent, Belgium. E-mail: Peter.Vanrolleghemrug.ac.be*

Abstract: In recent years, there has been a growing awareness of the ill-definedness of biotechnological processes, in particular the uncertainties attached to their models. Since both white-box and black-box models have their disadvantages, a mix between both formalisms is desirable. This formalism consists of grey-box models. In continuous time grey-box models often are described by stochastic differential equations (SDE's). In contrast with the increasing popularity of using SDE's, focus has rarely been upon the modelling of the noise characteristics. In this paper, a general framework in which the modelling of noise characteristics may proceed is presented. The framework will be illustrated by means of a practical biotechnological example.

Keywords: Biotechnology, Stochastic modelling, Differential equations

## 1. INTRODUCTION

In recent years, mathematical models have gained importance in describing, analyzing, optimising and controlling all kinds of systems. Along with the importance of models there is a growing awareness of the ill-definedness of certain systems (e.g. stock exchange or biotechnological processes).

In modelling ill-defined systems, two modelling approaches have been developed in the past (Spriet and Vansteenkiste, 1982). Deductive modelling ("white-box" models) uses the already available a priori knowledge, whereas inductive modelling ("black-box" models) is based on data.

Due to, for example a lack of knowledge, a lack of data, and the intrinsic unpredictability of ill-defined systems, there is a growing awareness of the importance of the concept of uncertainty. However, white-box (deterministic) models are not capable of including the uncertainties into the model. black-box models do take the uncertainties into account, but these models often lack a certain amount of predictive value and do not take advantage of substantial a priori knowledge that may be available. Therefore a mix between both the inductive and the deductive models is desirable. This can be achieved by including (stochastic) noise terms into the (deterministic) white-box models, resulting in stochastic *grey-box* models.

In continuous time grey-box models often are being described by *stochastic differential equations* (SDE's). The use of SDE's is becoming increasingly popular in, for example, water quality modelling (Finney *et al.*, 1982; Zielinski, 1988), in the modelling and control of the wastewater treatment processes (Carstensen, 1994; Tenno and Uronen, 1995) and biotechnological processes (Kinder and Wiechert, 1995). Some theoretical background on the subject of SDE's can, for example, be found in Bagchi (1993), Jazwinski (1970) or Kloeden and Platen (1992).

This paper will focus upon the modelling of the noise characteristics in grey-box SDE's. In contrast with the growing popularity of using SDE's, focus has rarely been on finding the adequate complexity or characteristics of the noise term. Often, different noise characteristics will yield different results in, for example, control (filtering) or uncertainty analysis. Moreover, inclusion of a noise term with particular characteristics might yield undesired results. For instance, a concentration of a substance in a bioreactor smaller than zero, may be obtained if the mean of the concentration is close to zero and the variance is large. Another motivating example is the desire to develop large systems of SDE's which can be used for a more robust control or risk analysis. Including noise terms without much consideration might result in unrealistic or very complex systems of stochastic differential equations.

In the next section SDE's will briefly be discussed. Section 3 will describe a general framework for modelling noise characteristics. This framework will be illustrated in section 4 with an example. Finally some conclusions will be given in section 5.

## 2. STOCHASTIC DIFFERENTIAL EQUATIONS

A stochastic differential equation, which is appropriate for biological growth processes (Kloeden and Platen, 1992), is defined by

$$d\mathbf{X}_t = \mathbf{f}(t, \mathbf{X}_t)dt + \mathbf{G}(t, \mathbf{X}_t)d\mathbf{W}_t, t \geq t_0 \quad (1)$$
$$\mathbf{X}_{t_0} \textit{ is a given random vector}$$

where $\mathbf{X}_t$ is a $d$-dimensional vector, $\mathbf{f}$ is a $d$-dimensional vector-valued function, $\mathbf{G}$ is a matrix-valued function of order $d \times m$ and $\mathbf{W}_t$ is a $m$-dimensional *Wiener process* or *Brownian motion*. A Wiener process is heuristically defined by

$$W_t = \int_0^t N_s ds, \; t \geq 0 \quad (2)$$

where $N_s$ is a Gaussian white noise process. The functions $\mathbf{f}(t, \mathbf{X}_t)$ and $\mathbf{G}(t, \mathbf{X}_t)$ are called drift and diffusion, respectively.

### 2.1 *Solutions of Stochastic Differential Equations: The Fokker-Planck Equation*

The solution of a SDE is a Markov process. It is a stochastic process with the property that, given the value of $\mathbf{X}_t$, the values of $\mathbf{X}_s$, $s > t$ do not depend on the values of $\mathbf{X}_\tau$, $\tau < t$. Therefore the probability density function (pdf) $p = p(t, \mathbf{x}; t_0, \mathbf{x}_0)$, defined as the density of the transition probability $P(\mathbf{X}_t = \mathbf{x} | \mathbf{X}_{t_0} = \mathbf{x}_0)$, satisfies the *Fokker Planck equation*

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x^i}\{f^i(t, \mathbf{x})\} - \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x^i \partial x^j}\{d^{i,j}(t, \mathbf{x})p\} = 0 \quad (3)$$

with the initial condition

$$\lim_{t \downarrow t_0} p(t, \mathbf{x}; t_0, \mathbf{x}_0) = \delta(\mathbf{x}_0 - \mathbf{x}) \quad (4)$$

where $\delta$ is the Dirac delta function on $\mathcal{R}^d$. The matrix-element $d^{i,j}$ is given as the $i, j^{th}$-element of the matrix $\mathbf{D} = \mathbf{G}\mathbf{G}^T$. For simple problems this equation can be used to provide the pdf. However, for more complex problems it is not possible to solve the Fokker Planck equation. Since one is often only interested in the central moments (mean, variance, skewness) of the solution of a SDE, the moment equations might provide a good alternative.

### 2.2 *The Moment Equations*

The *moment equations* are given by

$$\frac{dE[\phi(\mathbf{X}_t)]}{dt} = \sum_{i=1}^{d} E[f^i(t, \mathbf{X}_t)\frac{\partial \phi}{\partial x^i}(\mathbf{X}_t)] + \frac{1}{2}\sum_{i,j=1}^{d} E[d^{i,j}(t, \mathbf{X}_t)\frac{\partial^2 \phi}{\partial x^i \partial x^j}(\mathbf{X}_t)] \quad (5)$$

where $\phi(x_t)$ is equal to $\{x_t^i\}$ for calculation of the first moment of the $i^{th}$-element and $\{\prod_{j=1}^{n} \delta x_t^{i_j}$, $1 \leq i_j \leq d\}$ or $\{\prod_{j=1}^{n} x_t^{i_j}$, $1 \leq i_j \leq d\}$ for calculation of the $n^{th}$-order central moments or $n^{th}$-order (non-central) moments, respectively. Here, $\delta x_t^i = x_t^i - E[x_t^i]$. If both the drift and diffusion of the SDE are linear these equations will provide ordinary differential equations exactly describing the moments. However, for highly complex and nonlinear systems, approximation of the moments equations is possible but might result a system of non-stable ODE's. In this case numerical methods for solving SDE's will turn out to be valuable tools.

### 2.3 *Numerical Solutions of Stochastic Differential Equations*

Numerical methods for SDE's can be used in direct simulations, *i.e.*, simulating one trajectory (realisation), such as filtering or testing estimators. Another application of numerical methods for SDE's is the approximation of $E[\phi(\mathbf{X}_t)]$, where $\phi$ is equal to, for example, $\mathbf{X}_t$ or $\mathbf{X}_t\mathbf{X}_t^T$ (Kloeden and Platen, 1992).

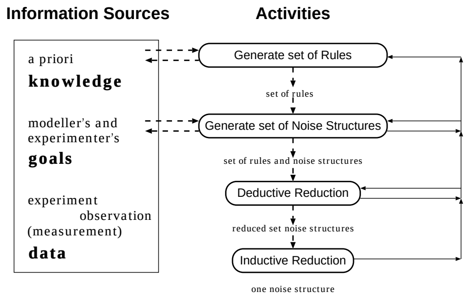


Fig. 1. *The process of noise characterisation*

The most simple discrete approximation of a SDE is the *Euler approximation* which is given by

$$X^i_{n+1} = X^i_n + f^i(t_n, \mathbf{X}_n)\Delta + \sum_{j=1}^{m} G^{i,j}(t_n, \mathbf{X}_n)\Delta W^j \qquad (6)$$

where $\Delta = t_{n+1} - t_n$ and $\Delta W = W(t_{n+1}) - W(t_n)$ is the $N(0;\Delta)$ increment of the Wiener process $W$.

In order to approximate, for example, the first and second central moments one needs to apply Monte Carlo methods for generating a batch of random trajectories of the SDE. Subsequently, the first and second moments (at each time instant) can be estimated by using the standard statistical formulas.

## 3. A FRAMEWORK FOR MODELLING NOISE CHARACTERISTICS

The modelling of noise characteristics is defined as characterising the structure of the diffusion function $\mathbf{G}$ in equation (1). As in modelling ill-defined systems (Kops *et al.*, 1997; Spriet and Vansteenkiste, 1982), the modelling of noise characteristics consists of constant interactions between *information sources* and *activities*. A schematic representation of the process of noise characterisation is given in figure 1.

### 3.1 *Information Sources*

Three major information sources can be identified: (a) goals and purposes, (b) a priori knowledge and (c) experimental data.

The goals and purposes of the modeller will orient the noise characterisation process. It will, for example, determine the complexity of the noise structure. The a priori knowledge reflects the knowledge already gathered (e.g. physical “laws”). The experimental data are the observations of the systems behaviour. For noise characterisation special focus is upon the noisy fluctuations of the observed behaviour.

### 3.2 *Activities*

From Figure 1 it may be concluded that all activities have to be performed top down. However, during the generation of both *rules* and *structures* there exist constant interactions between the activities and information sources.

The *rules* are a reflection of all three information sources. Since there will often exist a one-to-one mapping between the a priori knowledge (laws) and the rules, generating rules from the a priori knowledge is rather straightforward. Most rules resulting from the goals of the user reflect the complexity of the desired noise structure. For example, on-line control requires less complex noise structures than analysis of the system.

Generation of rules using the available data is more complex. If enough data (of repeated experiments) is available, the distribution (of all these data sets) can be approximated and used as a guideline. However, often only few data is available. In this case, a guideline for the variance characteristics can be provided by fitting the corresponding deterministic model to the data and examining the remaining residuals.

After rule generation, different candidate *noise structures* must be generated. Noise structures can be generated randomly or with use of the available information sources. Two examples of noise structures are additive noise terms, *i.e.*, adding a constant noise term to the deterministic model, and “parameter noise”. Parameter noise is generated by assuming the (deterministic) model parameters to be stochastic variables and rewriting (with or without approximations) the model into an SDE form. Since parameter noise often leads to rather complex structures, it is also advised to generate more simple structures such as linear or square root dependency (implying linear dependency of the noise matrix $\mathbf{D} = \mathbf{G}\mathbf{G}^T$) on the state variables.

Having generated a candidate set of rules and candidate noise structures, an appropriate structure must be selected. This will be done by *deductive* and *inductive* reduction. Deductive reduction will eliminate all models of which can be theoretically shown to violate one or more rules. No data will be used in the evaluation. Inductive reduction will use data to find the structure which obeys the rules as close as possible.

## 4. AN EXAMPLE: THE SINGLE MONOD MODEL

The simple process to which this modelling framework is applied for illustrative purposes consists

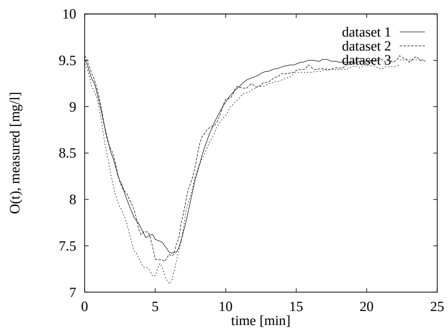


Fig. 2. *Measured $O_t$ data.*

of a simple biological growth system in which one organism is aerobically growing on a single substrate. The system described below in fact is a subsystem of a larger system characterised by a population of micro-organisms aerobically growing on a mixture of substrates. In the project the possible usefulness of SDE models is evaluated in the framework of uncertainty analysis, robust control and basic understanding of such bioprocesses.

The simple growth system studied is subjected to pulse substrate additions, allowing biokinetic characterisation of the growth process.

The culture with a certain initial biomass concentration $X_0$ is confronted with a sudden increase of the limiting substrate concentration from zero to $S_0$. Growth with a yield $Y_{sx}$ is assumed to occur according to Monod kinetics and endogenous metabolism is assumed to occur at a rate $(-bX_t)$, while maintenance is assumed absent. Oxygen consumption is proportional to substrate oxidation and endogenous metabolism and oxygen supply occurs through continuous aeration of the bioreactor $(K_La(O^s - O))$ leading to the oxygen mass balance. The only measurement made on the bioreactor consists of a dissolved oxygen electrode. Typical data sets collected during such pulse experiments are depicted in Figure 2. Experiment duration is approximately 20 minutes and measuring frequency is about 10 seconds. Initial substrate and biomass concentrations are known from off-line analysis.

The model is given by equation (1) where the "deterministic" part, the drift, is given by

$$\mathbf{f}(\mathbf{X}_t)^T = \Big[\frac{\mu_{max}S_t}{K_s+S_t}X_t - bX_t, -\frac{1}{Y_{sx}}\frac{\mu_{max}S_t}{K_s+S_t}X_t, K_La(O^s - O_t) - \frac{1-Y_{sx}}{Y_{sx}}\frac{\mu_{max}S_t}{K_s+S_t}X_t - (1-f_I)bX_t\Big] \quad (7)$$

where $\mathbf{X}_t = [X_t\ S_t\ O_t]^T$ with $X_t$ the biomass concentration, $S_t$ the substrate concentration, and $O_t$ the oxygen concentration. The noise structure $\mathbf{G}$ will be characterised in the following subsections.

### 4.1 *Generation of Rules*

Below some (first) rules are defined

1. *All concentrations cannot be smaller than zero.* This rule is relevant since in this experiment the substrate concentration approach zero. Whenever it is highly improbable, during a different type of experiment, that the concentrations will be close to zero, this rule may be omitted in order to reduce the noise structure complexity.
2. *The noise structure should be as simple as possible.* The simplicity (or complexity) desired depends highly on the modellers goal. However, as a general rule it is recommended to model the noise structure as being constant, linear or as a square root (implying $\mathbf{G}\mathbf{G}^T$ to be linear) if this does not imply violation of other rules.

These rules were defined using mostly a priori knowledge and the modeller goals. The next rule is based on the available data (figure 2).

3. *The noise structure should reflect the noisy fluctuations of the oxygen concentration just before (and during) it reaches a minimum and after it has (almost) stabilised. It is expected that the variance will increase when the mean oxygen concentration approaches its minimum. Thereafter it will decrease until an almost constant value.* This rule can be deduced from figure 2 and from Vanrolleghem *et. al.* (1994) where a residual analysis is done after fitting a deterministic model to oxygen uptake rate (OUR) data.

### 4.2 *Generation of Noise Structures*

Three noise structures are proposed. The first two are given by

$$\mathbf{G}_1(\mathbf{X}_t) = \begin{bmatrix} \sigma_x & 0 & 0 \\ 0 & \sigma_s & 0 \\ 0 & 0 & \sigma_o \end{bmatrix} \quad (8)$$

$$\mathbf{G}_2(\mathbf{X}_t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_{os}K_La & \sigma_{kla}(O^s - O_t) \end{bmatrix} \quad (9)$$

and the third is shown in Figure 3, where $\mathbf{W}_{1,t} = [W_t^x\ W_t^s\ W_t^o]^T$, $\mathbf{W}_{2,t} = [W_t^{os}\ W_t^{kla}]^T$, and $\mathbf{W}_{3,t} = [W_t^{\mu}\ W_t^b\ W_t^{ks}\ W_t^y]^T$. The first noise structure $\mathbf{G}_1$ can be interpreted as *additive noise*. Noise structure $\mathbf{G}_2$ can be interpreted as noise on the parameters $K_La$ and $O^s$. The most complex noise

$$\mathbf{G}_3(\mathbf{X}_t) = \begin{bmatrix} \sigma_\mu \frac{S_t X_t}{K_s + S_t} & -\sigma_b X_t & -\sigma_{ks} \frac{\mu_{max} S_t X_t}{(K_s + S_t)^2} & 0 \\ -\sigma_\mu \frac{1}{Y_{sx}} \frac{S_t X_t}{K_s + S_t} & 0 & \sigma_{ks} \frac{1}{Y_{sx}} \frac{\mu_{max} S_t X_t}{(K_s + S_t)^2} & -\sigma_Y \frac{\mu_{max} S_t X_t}{K_s + S_t} \\ -\sigma_\mu \frac{1 - Y_{sx}}{Y_{sx}} \frac{S_t X_t}{K_s + S_t} & -\sigma_b (1 - f_I) X_t & \sigma_{ks} \frac{1 - Y_{sx}}{Y_{sx}} \frac{\mu_{max} S_t X_t}{(K_s + S_t)^2} & -\sigma_y \frac{\mu_{max} S_t X_t}{K_s + S_t} \end{bmatrix} \quad (10)$$

Fig. 3. Noise structure $\mathbf{G}_3$.

structure $\mathbf{G}_3$ can be interpreted, after some approximations (Kops, 1997), as noise on the kinetic parameters $\mu_{max}$, $b$, $K_s$ and the stoichiometric parameter $Y_{sx}$.

### 4.3 *Deductive Reduction*

Using the moment equations, it can analytically be proven that if one of the variables in $\mathbf{X}_t$ approaches zero, its variance will diverge if the noise structure $\mathbf{G}$ contains a constant term (Kops, 1997). Therefore, structures $\mathbf{G}_1$ ($\sigma_x$, $\sigma_s$ and $\sigma_o$ are constant) and $\mathbf{G}_2$ ($\sigma_{kla}$ is constant) violate rule (1), and $\mathbf{G}_3$ will be chosen in favour of $\mathbf{G}_1$ and $\mathbf{G}_2$. Note that most applications described in literature (*e.g.* Carstensen (1994) and Tenno and Uronen (1995)) do use an additive noise term, making them unfit for the description of system behaviour where one of the state variables approaches zero.

An important tool in deductive reduction is *sensitivity analysis*. If the model output, in particular its first two moments, is not sensitive to changes in the noise parameter $\sigma_\cdot$, the noise term related to $\sigma_\cdot$ can be dropped. In figures 4 and 5 the sensitivity of the variance of $O_t$ (at two different time instants) towards the parameters $\sigma_\mu$, $\sigma_b$, $\sigma_{ks}$, and $\sigma_y$ is shown [1]. Indeed, focus is upon this characteristic of the noise model because the third rule of the exercise asks for evaluation of $Var(O_t)$. In these figures "100%" is, for each parameter, a reference value. Since $O_t$ is the most commonly measured variable and $EO_t$ does not show any sensitivity towards the noise parameters, the variance of $O_t$ is chosen as a criterion. The first time instant for sensitivity evaluation has been chosen in such a way that $Var(O_t)$ is at its maximum and the second time instant lies in the tail of both the mean and variance of $O_t$ (figure 6). As can been seen in figures 4 and 5, the variance of $O_t$ is not very sensitive towards the parameter $\sigma_{ks}$. This implies that the noise term related to this

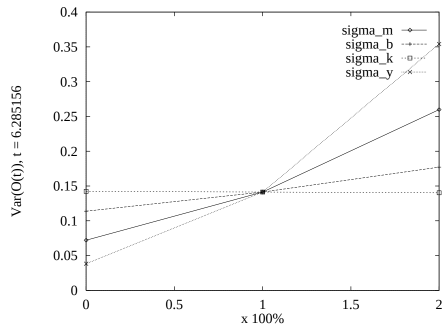


Fig. 4. *The sensitivity of* $Var(O_t)$, $t = 6.285156$

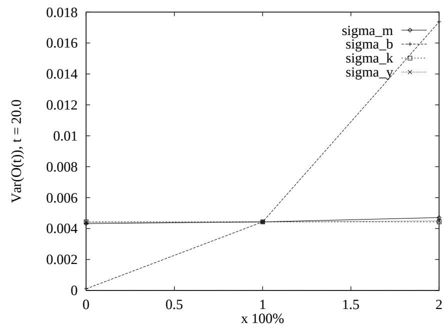


Fig. 5. *The sensitivity of* $Var(O_t)$, $t = 20.0$.

parameter ($\frac{\mu_{max} S_t X_t}{(K_s+S_t)^2}$ times a constant) only has a very small influence on the model output noise characteristics. Since this noise term is included in all three differential equations (only differing with some constant) this term can be dropped from noise structure $\mathbf{G}_3$.

### 4.4 *Inductive Reduction*

In figure 6 the variance of noise structure $\mathbf{G}_3$ is shown. Notice that the characteristics of the variance shown in this figure matches rule (3). Since structure $\mathbf{G}_3$ is rather complex, another more simple noise structure is proposed

$$\mathbf{G}_4(\mathbf{X}_t) = \begin{bmatrix} \sigma_\mu \sqrt{S_t} & -\sigma_b & 0 \\ -\sigma_\mu \frac{1}{Y_{sx}} \sqrt{S_t} & 0 & -\sigma_y \sqrt{S_t} \\ -\sigma_\mu \frac{1 - Y_{sx}}{Y_{sx}} \sqrt{S_t} & -\sigma_b (1 - f_I) & -\sigma_y \sqrt{S_t} \end{bmatrix} \quad (11)$$

[1] All means and variances in this subsection have been approximated using numerical methods for SDE's together with Monte Carlo simulation (as described in subsection 2.3).

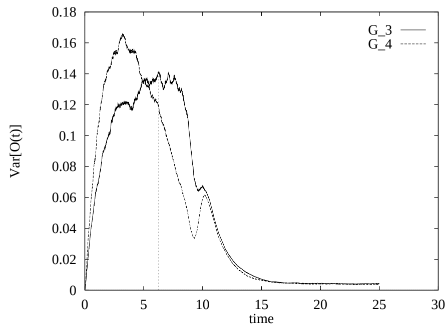


Fig. 6. *The variance of* $O_t$.

where $\mathbf{W}_{4,t} = [W_t^\mu \; W_t^b \; W_t^y]^T$. In this structure the Monod kinetics are replaced by $\sqrt{S_t}$. This implies that the variance of the noise terms is (linearly) related to $S_t$. The variable $X_t$ is dropped from the noise model since, *in this experiment*, it is far from zero and almost constant. Figure 6 shows the variance of the SDE's with both noise structures $\mathbf{G}_3$ and $\mathbf{G}_4$. These figures show that, roughly, the characteristics of the variance are the same, *i.e.*, $\mathbf{G}_4$ still obeys rule 3. If a simple noise structure is highly desirable (rule 2), structure $\mathbf{G}_4$ can be chosen in favour of $\mathbf{G}_3$.

## 5. CONCLUSIONS

In this paper a general framework for modelling noise characteristics in stochastic differential equations is given. Using this framework, an appropriate noise structure for the modelled system can be found. Including it into a (deterministic) white-box model, will result a grey-box model. This model is capable of both describing the uncertainties in the system and the available a priori knowledge.

A major consideration for modelling the noise structure is the generation of large systems of grey-box SDE's which can be used for more robust control, better analysis of the system or risk analysis. Here, too little consideration about the noise structure might lead to unrealistic or very complex SDE's. The framework provided in this paper, illustrated using a simple model, can be used as a guideline for obtaining grey-box SDE's from more complex and larger (more variables) deterministic models (such as the cell age model for *Penicillium chrysogenum* fed-batch fermentation proposed by Yuan (Yuan *et al.*, 1997)).

## 6. ACKNOWLEDGEMENTS

This work is supported (in part) by FWO-project G.0286.96 of the Fund for Scientific Research, Belgium

## 7. REFERENCES

Bagchi, A. (1993). *Optimal Control of Stochastic Systems*. Prentice-Hall. London, UK.

Carstensen, J. (1994). Identification of Wastewater Processes. PhD thesis. Technical University of Denmark.

Finney, B.A., D.S. Bowles and M.P. Windham (1982). Random differential equations in river water quality modeling. *Water Resources Research* **18**(1), 122–134.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Kinder, M. and W. Wiechert (1995). Stochastic simulation of biotechnical processes. In: *Proceedings of IMACS/IFAC First International Symposium on Mathematical Modelling and Simulation in Agriculture and Bio-industries*. Vol. 1. Brussels, Belgium. pp. 1–6.

Kloeden, P.E. and E. Platen (1992). *Numerical Solutions of Stochastic Differential Equations*. Springer-Verlag. Heidelberg.

Kops, S., H. Vangheluwe, F. Claeys, F. Coen, P. Vanrolleghem, Z. Yuan and G. Vansteenkiste (1997). The process of model building and simulation of ill-defined systems: Application to wastewater treatment. In: *Proceedings 2nd IMACS symposium on Mathematical Modelling (MATHMOD)* (I. Troch and F. Breitenecker, Eds.). Vienna, Austria. pp. 859–864.

Kops, S.G.T. (1997). Theoretical aspects of modelling the noise characteristics of a simple growth system with single monod kinetics. Technical report. Dep. of Applied Mathematics, Biometrics and Process Control (BIOMATH), University of Gent. Gent, Belgium.

Spriet, J.A. and G. Vansteenkiste (1982). *Computer-aided modelling and simulation*. Academic Press. London.

Tenno, R. and P. Uronen (1995). State and parameter estimation for wastewater treatment processes using a stochastic model. *Control Engineering Practice* **3**(6), 793–804.

Vanrolleghem, P.A., M. Van Daele, P. Van Overschee and G. Vansteenkiste (1994). Model structure characterization of nonlinear wastewater treatment systems. In: *Preprints 10th IFAC Symposium on System Identification* (M. Blanke and T. Soderstrom, Eds.). Vol. 1. Danish Automation Society. Copenhagen, DK. pp. 279–284.

Yuan, J.Q., S.R. Gou and K.-H. Bellgardt (1997). A cell age model for penicillin cultivation. In: *Proceedings of IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*. Berlin, Germany.

Zielinski, P.A. (1988). Stochastic dissolved oxygen model. *Journal of Environmental Engineering*.