Journal of Theoretical Biology 305 (2012) 78–87

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Development of a restricted state space stochastic differential equation model for bacterial growth in rich media

Jan Kloppenborg Møller*, Kirsten Riber Bergmann, Lasse Engbo Christiansen, Henrik Madsen

DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark

ARTICLE INFO

Article history:
Received 24 November 2010
Received in revised form
11 April 2012
Accepted 16 April 2012
Available online 24 April 2012

Keywords:
Stochastic differential equations
Logistic diffusion
Monod growth
Rich media
Enterococcus
Salmonella

ABSTRACT

In the present study, bacterial growth in a rich media is analysed in a Stochastic Differential Equation (SDE) framework. It is demonstrated that the SDE formulation and smoothened state estimates provide a systematic framework for data driven model improvements, using random walk hidden states. Bacterial growth is limited by the available substrate and the inclusion of diffusion must obey this natural restriction. By inclusion of a modified logistic diffusion term it is possible to introduce a diffusion term flexible enough to capture both the growth phase and the stationary phase, while concentration is restricted to the natural state space (substrate *and* bacteria non-negative). The case considered is the growth of Salmonella and Enterococcus in a rich media. It is found that a hidden state is necessary to capture the lag phase of growth, and that a flexible logistic diffusion term is needed to capture the random behaviour of the growth model. Further, it is concluded that the Monod effect is not needed to capture the dynamics of bacterial growth in the data presented.

1. Introduction

Stochastic Differential Equations (SDEs) have shown great potential for modelling in various areas of application, e.g. Tornøe et al. (2004), Pedersen et al. (2008) and Madsen et al. (1987). In general such models are based on formulations in which a constant diffusion term is added to a non-linear drift term, which in turn is derived from known theory. In many cases this is a fruitful approach, but for systems with a restricted state space, this might not be the best approach, since additive diffusion will allow the system to leave the physically feasible state space, referred to as the natural state space. A main theme of this paper is to provide a proper description of diffusion terms in restricted bacterial growth models.

It is well known that SDEs give more correct state estimation because filtering techniques allow the information given by observations to be incorporated into the state estimate. In addition to improved state estimation, an SDE approach provides a framework for identification of model deficiencies (Kristensen et al., 2004b). In this paper the method is applied to bacterial growth data and allows us to identify and argue for possible model improvements.

Parameter estimation in SDEs is a difficult task and approximate methods have to be applied for complex systems. In this study we apply a statistical method based on the Extended Kalman Filter (EKF) (Jazwinski, 1970), implemented in the statistical software CTSM[1] (Kristensen and Madsen, 2003; Kristensen et al., 2004a). CTSM has proven powerful in many applications (e.g. Philipsen et al., 2010); it is, however, not possible to include state dependent diffusion in CTSM. One way to overcome this restriction is by including some input in the description of the diffusion term (e.g. state observations Philipsen et al., 2010). A more correct description can be obtained by applying Lamperti type transformations (e.g. Iacus, 2008) to obtain a system with state independent diffusion from a system with state dependent diffusion. Under the restriction that Lamperti transformations are available, different parametric representations of the diffusion term are derived and evaluated with respect to a case study of bacterial growth. In particular a novel modified logistic diffusion term will be introduced and evaluated.

The case study in this paper is a bacterial growth model, and an adequate description of bacterial growth kinetics, i.e. the relationship between bacterial growth and substrate concentration, is important for many applications in microbiology, for instance for fermentation processes (Kompala et al., 1984; Patnaik, 1999). The Monod equation was the first suggestion for a mathematical description of the growth curve. It has been extensively discussed since its introduction in 1949 (Monod, 1949). When originally proposed, it seemed to be a "convenient and logical" (Monod, 1949) choice for a mathematical expression to fit the growth curve. Several attempts have been made to formulate a mechanistic

* Corresponding author. Tel.: +45 453375.
E-mail address: jkm@imm.dtu.dk (J.K. Møller).

[1] www2.imm.dtu.dk/~ctsm

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.04.015

background for the Monod equation (Lobry et al., 1992; Liu, 2006). Even though the Monod equation is a good description of the growth on a single substrate, it fails to model growth in rich media (Bajpai-Dikshit et al., 2003). Therefore several attempts have been made to find another equation for this growth (Doshi et al., 1997). According to Kovárová-Kovar and Egli (1998) these studies can be divided into three methods: (i) extending the Monod model with additional constants, (ii) developing an empirical or mechanistic model from kinetic concepts and (iii) describing how the Monod growth parameters are influenced by physiochemical factors.

A problem with determining other expressions for the growth rate has been the lack of high quality reproducible data which relates the growth rate to the substrate concentration (Kovárová-Kovar and Egli, 1998). The method proposed here makes it possible to extract these data from bioscreen measurements, and thus limit the time and resources used on experiments significantly.

The transition from modelling using Ordinary Differential Equations (ODEs) to Stochastic Differential Equations (SDEs) paves the way for many strong statistical tools for model development and inference (Kristensen et al., 2004a). In this paper a model development method based on SDEs (and the applicability of the Lamperti transform) will be introduced to examine and subsequently describe the relation between growth rate and substrate concentration. The use of SDEs enables the separation of measurement noise and system noise, and this is used in the method. First the SDE framework is introduced, followed by a short presentation of the data. Then a thorough analysis of the diffusion term is performed to determine the best way to introduce diffusion into the model. In the last part we develop the drift term and present a simulation study of the final model.

2. Methodology

The systematic framework in this study is a continuous-discrete time stochastic state-space formulation

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t,\boldsymbol{u}_t,t,\boldsymbol{\theta})\,dt + \boldsymbol{\sigma}(\boldsymbol{X}_t,\boldsymbol{u}_t,t,\boldsymbol{\theta})\,d\boldsymbol{w}_t, \tag{1}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k,\boldsymbol{u}_k,t_k,\boldsymbol{\theta}) + \boldsymbol{e}_k, \tag{2}$$

where Eq. (1) is the continuous time system equation and Eq. (2) is the discrete time observation equation. $\boldsymbol{X}_t$ is a vector of state variables, $\boldsymbol{Y}_k$ is a vector of measured output variables at times t_k, $\boldsymbol{u}_t$ is a vector of known input variables, $\boldsymbol{e}_k$ is an l-dimensional white noise process with $\boldsymbol{e}_k \in \mathcal{N}(0,\boldsymbol{S}(\boldsymbol{u}_k,t_k,\boldsymbol{\theta}))$, $\boldsymbol{w}_t$ is a standard Wiener process with zero mean and independent Gaussian time increments, and $\boldsymbol{\sigma}(\boldsymbol{X}_t,\boldsymbol{u}_t,t,\boldsymbol{\theta})$ is the diffusion coefficient. The first part of the system equation is called the drift term and the second part is called the diffusion term. Everywhere in this article the Itô interpretation of the SDEs is used (see e.g. Øksendal, 2003).

2.1. Estimation

The estimation procedure is a maximum likelihood procedure in which the EKF is used to evaluate the likelihood function. A full account is available in Kristensen and Madsen (2003), and we will not present the details here. We will, however, give a few remarks on the output from the software and the restrictions of the procedure.

In addition to the parameter estimates and the parameter covariance matrix (estimated by the inverse Hessian, e.g. Madsen (2008)), the implementation allows us to calculate k-step predictions of both the state and the observations, and the smoothened state (state estimate given all observations).

Of special interest is the one-step prediction. The standardised residuals for one dimensional observations are given by

$$r_k = \frac{Y_k - \hat{Y}_{k|k-1}}{\sqrt{\Sigma^{yy}_{k|k-1}}}, \tag{3}$$

where $\hat{Y}_{k|k-1}$ is the one-step prediction of the observations and $\Sigma^{yy}_{k|k-1}$ is the variance of the one-step prediction. If the model has captured the mechanistic behaviour of data, then r_k is a white noise process. Furthermore, if the sampling intervals are equidistant, then r_k can by analysed using traditional residual analysis, like the autocorrelation function and the partial autocorrelation function.

In addition to the improved state estimates, smoothened state estimates of a modified system can help to identify model deficiencies (Kristensen et al., 2004b). A suspicion that the model is not sufficient to describe the variation in data can be analysed by considering smoothened state estimates of an extended state space, with the system equation

$$d\begin{bmatrix}\boldsymbol{X}_t \\ \theta^i_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{f}(\boldsymbol{X}_t,\boldsymbol{u}_t,t,\theta^i_t,\tilde{\boldsymbol{\theta}}) \\ 0\end{bmatrix} dt + \begin{bmatrix}\boldsymbol{\sigma}(\boldsymbol{X}_t,\boldsymbol{u}_t,t,\tilde{\boldsymbol{\theta}}) & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_\theta\end{bmatrix} d\boldsymbol{w}_t, \tag{4}$$

where θ^i_t is the parameter to be investigated and $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}\backslash\theta^i = \{\theta^1,\ldots,\theta^{i-1},\theta^{i+1},\ldots,\theta^p\}$, and p is the dimension of the parameter space. By plotting the smoothened state estimates of θ^i_t as a function of time and possible covariates, model improvements can be identified.

2.2. Transformation of the state space

As mentioned above it is not possible to include processes with state dependent diffusion in CTSM (i.e. the SDE models are restricted to the form $\boldsymbol{\sigma}(\cdot) = \boldsymbol{\sigma}(\boldsymbol{u}_t,t,\boldsymbol{\theta})$). For one dimensional diffusion processes with a state dependent diffusion term it is, however, always possible to find a transformation of the state space such that the diffusion is independent of the states (e.g. Baadsgaard et al., 1997). The transformation is often referred to as the Lamperti transform (Iacus, 2008)

$$Z_t = \psi(X_t,\cdot) = \int \frac{d\xi}{\sigma(\xi,\cdot)}\bigg|_{\xi = X_t}, \tag{5}$$

and by Itô's lemma (e.g. Øksendal, 2003), Z_t will again be an Itô process, given by Luschgy and Pagés (2006)

$$dZ_t = \left(\psi_t(\psi^{-1}(Z_t,\cdot),t) + \frac{f(\psi^{-1}(Z_t,\cdot),\cdot)}{\sigma(\psi^{-1}(Z_t,\cdot),\cdot)}\,dt - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t,\cdot),\cdot)\right)dt + dw, \tag{6}$$

where subscript refers to partial differentiation. The transformed SDE in (6) has state independent diffusion. The practical applicability of the Lamperti transform is limited to cases where we can find an explicit inverse ($\psi^{-1}(Z_t,\cdot)$). The existence of a Lamperti transform with an explicit inverse is the basis of the presented analysis.

3. Data

Optical density measurements for the growth of a *Salmonella* strain and an *Enterococcus* strain growing in BHI media are available for the analysis. For each bacteria culture a 9-fold dilution as well as the non-diluted strain are measured in duplicates. The measurements are made for 40 h with a sampling interval of 20 min in a bioscreen (Microbiology Reader Bioscreen C) at 16 °C under continuous shaking.

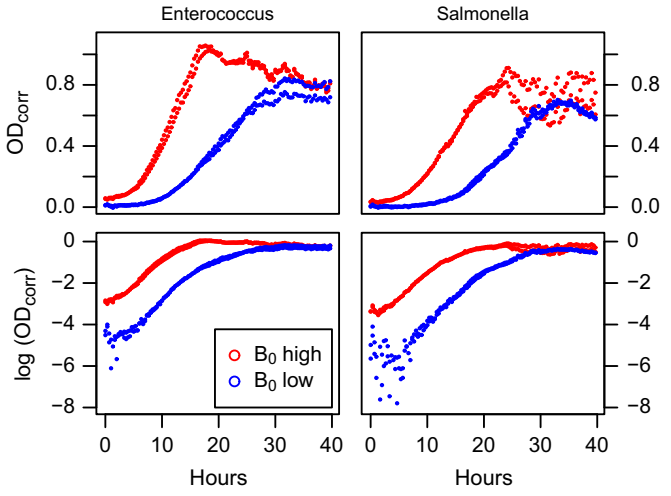

Fig. 1. Untransformed optical density measurements (top row) and log-transformed optical density measurements (bottom row).

The OD measurements are preliminarily corrected for background broth measurements and subsequently a correction is made for the shadow effect for high OD values by the relation (Philipsen et al., submitted for publication)

$$\mathrm{OD}_{\mathrm{corr}} = -\frac{1}{b}\log\left(1-\frac{\mathrm{OD}_{\mathrm{meas}}}{a}\right). \tag{7}$$

The constants a and b are found by fitting the relation to results from an experiment with *Salmonella* for which the OD for different concentrations has been measured. It is assumed that the same parameters can be used for *Enterococcus*. It should be stressed that Eq. (7) originates from the physical properties (shadow effects) of the measurements. The OD measurements after corrections, $\mathrm{OD}_{\mathrm{corr}}$, are shown in Fig. 1.

The models that are developed in this work are not suited for modelling the transition to the stationary phase (or the stationary phase itself). However, the transition is needed in order to be able to estimate the stoichiometric constant, η, (see Section 4 below). It has therefore been decided to include all data-points from the experiments and the aim is to capture the stationary phase partly by the diffusion term.

Even if the stationary phase can be partly captured by the diffusion term, it is evident that the fast variation in the stationary phase (see Fig. 1) cannot be captured by a standard growth model, and time varying observation noise will therefore also be included in the model.

The data set consists of two replications of each condition, and this is divided into two parts of equal sizes. The first part (the training set) consists of one of the replications of each of the conditions, and the second part (the test set) consists of the other replication of each of the same experimental conditions. The model development in Sections 4–7 is based on the training set, while the model results are compared with the test set in Sections 8 and 9.

4. A minimal stochastic growth model

A bacterial growth model should include mechanisms to ensure that the growth rate is zero when the substrate is depleted and that substrate and bacteria levels are both positive at all times. While these restrictions are simple in an ODE setting, it is more complicated in a SDE setting, hence the diffusion term must be handled with care. A first general formulation is

$$d\begin{bmatrix}B_t\\S_t\end{bmatrix} = \begin{bmatrix}\tilde{\mu}(S_t,B_t,\boldsymbol{\theta})B_t\\-\eta\tilde{\mu}(S_t,B_t,\boldsymbol{\theta})B_t\end{bmatrix}dt + \begin{bmatrix}\tilde{\sigma}_B(S_t,B_t,\boldsymbol{\theta})\\-\eta\tilde{\sigma}_B(S_t,B_t,\boldsymbol{\theta})\end{bmatrix}dw, \tag{8}$$

where η is a stoichiometric conversion from bacteria units to substrate units, and $\tilde{\mu}(\cdot)$ and $\tilde{\sigma}_B(\cdot)$ are functions to be determined.

Now consider $T_t = S_t + \eta B_t$. Clearly T_t is constant ($dT_t = 0$) and S_t can be expressed as

$$S_t = T_t - \eta B_t = S_0 + \eta(B_0 - B_t). \tag{9}$$

and (8) reduces to the one-dimensional diffusion process

$$dB_t = \tilde{\mu}(S_t,B_t,\theta)B_t\,dt + \tilde{\sigma}_B(S_t,B_t,\theta)\,dw, \tag{10}$$

with S_t given by (9).

S_t is implemented as standardised substrate such that $T_t = T_0 = 1$. The simplest model for the drift term that ensures substrate *and* bacteria concentration to be above zero at all times is $\tilde{\mu}(\cdot) = \mu S_t B_t$, (with μ constant). It is tempting to introduce a diffusion proportional to the state, i.e. $\tilde{\sigma_B}(S_t,B_t,\theta) = \sigma_B B_t$, with σ_B constant (corresponding to variance proportional to B_t^2). Such a formulation does not, however, include a mechanism to reduce the diffusion fast enough as S_t approaches zero, and will therefore hold a positive probability of $\eta B_t > T_0$ (implying $S_t < 0$), which is clearly a violation of the natural constraints for the system.

To include the mass balance in the diffusion term, we consider a logistic diffusion term (Schurz, 2007), i.e. $\tilde{\sigma}_B$ is modelled as

$$\tilde{\sigma}_B(S_t,B_t,\theta) = \sigma_B B_t^{\alpha} S_t^{\beta}, \tag{11}$$

where σ_B, α and β are real constants. Schurz (2007) analyses a class of logistic SDEs (including a logistic drift term) and proves that for $\alpha \geq 1$ and $\beta \geq 1$ the logistic SDE will stay within the boundary defined by the logistic ODE.

In Section 4.1 a minimal model with the desired characteristics is defined and in Section 6 we define a more flexible diffusion term, which gives a better representation of data.

4.1. Initial estimation model

Based on the reasoning above, the minimal drift term ($\tilde{\mu}(\cdot) = \mu B_t S_t$) and a logistic diffusion with $\alpha = \beta = 1$ the initial model is then given by

$$dB_t = \mu S_t B_t\,dt + \sigma_B S_t B_t\,dw, \tag{12}$$

where $S_t = T_0 - \eta B_t$.

In the following we will extend the model assuming that μ is not constant, but we will use (12) as a starting point for developing the diffusion term further before we turn to the drift term. Formulated in terms of B_t Eq. (12) is given by

$$dB_t = \mu(T_0 - \eta B_t)B_t\,dt + \sigma_B(T_0 - \eta B_t)B_t\,dw. \tag{13}$$

The Lamperti transform is

$$Z_t^B = \int \frac{d\xi}{\xi(T_0 - \eta\xi)}\bigg|_{\xi = B_t} = \frac{1}{T_0}\log\left(\frac{B_t}{T_0 - \eta B_t}\right), \tag{14}$$

with the inverse

$$B_t = T_0\frac{e^{Z_t^B T_0}}{1 + \eta e^{Z_t^B T_0}}, \tag{15}$$

and the transformed system is given by

$$dZ_t^B = \left(\mu + \frac{1}{2}\sigma_B^2(2\eta B_t - T_0)\right)dt + \sigma_B\,dw. \tag{16}$$

The system Eq. (16) is intentionally not transformed to unit diffusion, because we want to be able to control the initial conditions independently of the parameters (see Section 5).

The state space of Z_t^B is the entire real axis, and for $B_t \to 0$ and $\eta B_t \to T_0$ the drift term in dZ_t^B is a constant and for finite t, Z_t^B is finite *a.s.*. We will therefore use the formulation (16) as a starting point for the model development.

5. Estimation issues

As mentioned earlier there are two experiments for each bacteria, one with high starting concentration and one with low starting concentration, and two replications of each experiment (one in the training set and one in the test set). Low starting concentration is reported to be a 9-fold dilution of the high starting concentration, this relation is not, however, evident from the data, and it was decided to estimate the dilution.

The estimation of dilution is performed by starting the process at time $t=-1$ min, with B_{-1} close to zero (in practice $Z_{-1}^B=-10$) and letting the estimation procedure estimate the starting concentration by integrating a modified growth model

$$dB_t = (\mu S_t B_t + C_1 u_t^h + C_2 u_t^l)\,dt + \sigma_B S_t B_t\,dw, \tag{17}$$

where $u_t^l = 1$ for $t<0$ and low starting concentration and zero otherwise, and $u_t^h = 1$ for $t<0$ and high starting concentration and zero otherwise. In the following we will suppress these inputs, but note that they will be present throughout the analysis.

The formulation (17) as well as the further development presented below do not contain any mechanisms to capture the deterministic dynamics of the stationary phase, and the diffusion does not contain explicit information about the transition. It is therefore expected that the observation noise in the stationary phase will be larger than the observation noise in the growth phase. The observation equation is therefore formulated as

$$OD_{corr,k} = B_{t_k} + e_k, \tag{18}$$

where $e_k \sim N(0, s_k^2)$. As discussed above and in Section 3, the models we consider here are not well suited for modelling the stationary phase, this is the case for the SDE-models as well as the deterministic model considered as a benchmark model. One choice could be to cut off data in the stationary phase. Here we have chosen to down-weight data in the stationary phase by selecting s_k^2 as

$$s_k^2 = s_0^2 + s_1^2 \max(0, t_k - t_{k_s}). \tag{19}$$

with $k_s = \arg\max_k (OD_{corr,k})$. This formulation ensures that the influence of observations in the stationary phase is small while still maintaining the important transition to the stationary phase in the observations.

6. Model development

The shape of the diffusion term determines the random behaviour of the bacterial growth process. The drift terms of the models that will be analysed here are well suited for describing the growth part of the process, while the transition to the stationary phase is not included. The stationary phase should therefore be captured by the diffusion term. However, the diffusion term introduced above takes its maximum at $\eta B_t = \frac{1}{2}$, while the transition to the stationary phase should be close to 1, which is where the growth stops. It is therefore reasonable to assume that the maximum of the diffusion is at a value in the interval $(\frac{1}{2}, 1)$.

In order to introduce this kind of random behaviour, we propose a modified logistic diffusion term for modelling random bacterial diffusion

$$\tilde{\sigma}_B(B_t, S_t, \theta) = \sigma_B B_t (T_0 - (\eta B_t)^\gamma), \tag{20}$$

where γ is a positive constant. The Lamperti transform for this diffusion is

$$Z_t^B = \int \frac{d\xi}{\xi(T_0 - (\eta\xi)^\gamma)}\bigg|_{\xi = B_t} = \frac{1}{\gamma T_0}\log\left(\frac{(\eta B)^\gamma}{T_0 - (\eta B)^\gamma}\right), \tag{21}$$

with the inverse

$$B_t = \frac{T_0}{\eta}\frac{e^{T_0 Z_t^B}}{(1 + e^{\gamma T_0 Z_t^B})^{1/\gamma}}. \tag{22}$$

The state space of Z_t^B is the real line and the Itô process for Z_t^B is given by

$$dZ_t^B = \left(\mu\frac{T_0 - \eta B_t}{T_0 - (\eta B_t)^\gamma} - \frac{1}{2}\sigma_B^2(T_0 - (\gamma+1)(\eta B_t)^\gamma)\right)dt + \sigma_B\,dw_t, \tag{23}$$

where B_t is given by the inverse transformation (22).

Applying L'Hôpital's rule (with $T_0 = 1$) we get

$$\lim_{\eta B \to 1} dZ_t^B = \left(\frac{\mu}{\gamma} - \frac{\gamma}{2}\sigma_B^2\right)dt + \sigma_B\,dw_t, \tag{24}$$

implying that Z_t^B is finite *a.s.* for finite t.

The introduction of γ improved the likelihood significantly (Table 1) for both Enterococcus and Salmonella. The stoichiometric constant, η, also changes significantly (Table 1). The consequence of the small values of η in the initial model setup (12), compared to the values of η estimated by the modified logistic diffusion (20) is that transition to the stationary phase is estimated to be close to the maximum of the diffusion term (i.e. $\frac{1}{2}$ when $\gamma = 1$, Fig. 2).

Based on the likelihood results presented in Table 1 and the conventional knowledge that bacteria will consume the available substrate, the model analysis in the following includes γ, and we refer to the model (17) with the diffusion (20) as Model 0. The large values of γ imply that the diffusion is close to state proportional diffusion, except very close to the upper bound for bacteria concentration, where the diffusion is killed off very rapidly. However, the key difference is that the presented model guarantees that $S_t > 0$, whereas state proportional (or constant) diffusion will allow for $S_t < 0$.

Table 1
Estimation results for models for both $\gamma = 1$ and γ part of the estimation, respectively. The value of the key parameters (γ and η) are marked in bold face and standard deviations are given in parentheses.

Bacteria	γ	η	DF	log (L)	AIC	BIC
Enterococcus	**1** (NA)	**0.78** (0.024)	6	694.3	−1377	−1356
Enterococcus	**26.6** (12)	**0.93** (0.0059)	7	720.8	−1428	−1403
Salmonella	**1** (NA)	**0.87** (0.049)	6	687.9	−1364	−1343
Salmonella	**79.9** (11)	**1.09** (0.0013)	7	711.6	−1409	−1385

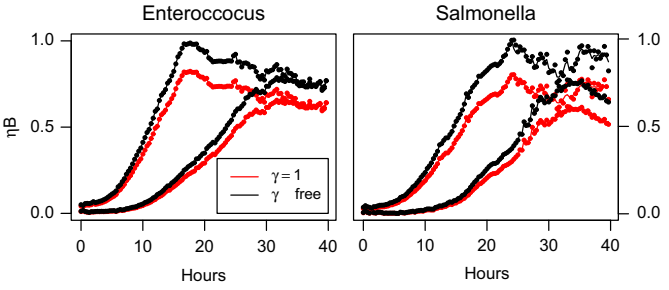

Fig. 2. Smoothened states of the bacteria concentration measured in substrate units, for Model 0 for both $\gamma = 1$ and γ estimated. Lines are smoothened values, while dots are measured values.

6.1. Random walk identification of the growth process

By formulating μ as a pure random walk hidden state adapted to data it is possible to analyse the growth rate and then suggest further model improvements. The model formulation in the transformed domain is

$$d\begin{bmatrix} Z_t^B \\ \mu_t \end{bmatrix} = \begin{bmatrix} \mu_t \frac{T_0-\eta B_t}{T_0-(\eta B_t)^\gamma} - \frac{1}{2}\sigma_B^2(T_0-(\gamma+1)\eta B_t) \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_B & 0 \\ 0 & \sigma_\mu \end{bmatrix} dw, \tag{25}$$

where B_t is given by the inverse transformation (22).

The smoothened state (state estimate given all observations) of μ_t can now be analysed to get an idea of the possible model improvements. The shape of the μ_t-curves varies substantially depending on starting concentrations and type of bacteria (Fig. 3). It is also evident that simple functions of time or substrate (or bacteria) will not be sufficient to capture these shapes. The growth parameter is linearly dependent on the substrate in the growth phase (but with different slopes in the different situations), implying a second order interaction in the growth rate of bacteria. Furthermore, the autocorrelation function of the residuals from Model 0 (Fig. 4) has many significant values. Such behaviour suggests that an additional state could improve the model.

Following Bajpai-Dikshit et al. (2003) a possible extension of Model 0 is

$$\mu_t = \nu E_t, \tag{26}$$

with

$$\frac{dE_t}{dt} = \frac{\nu+\beta}{\kappa+S_t} S_t - E_t \frac{d\log(B_t)}{dt} - \beta E_t, \tag{27}$$

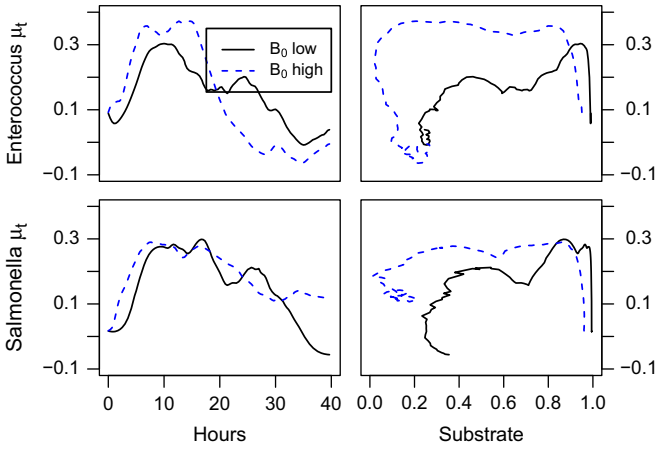

Fig. 3. Growth rates estimated from the model presented in Eq. (25).

where ν is the maximum bacterial growth rate, and β is the first-order degradation of proteins inside the cell. See Bajpai-Dikshit et al. (2003) for more specific details of the formulation.

To keep derivations simple and preserve flexibility of the enzyme diffusion, we consider only the deterministic part of dB_t in the derivations of dE_t. For special cases of the enzyme diffusion, a full stochastic inclusion of $d\log(B_t)$ is possible, however (see Appendix A). The deterministic formulation of dE_t is

$$dE_t = ((\nu+\beta)\lambda_t - E_t^2\nu\lambda_t - \beta E_t)\,dt, \tag{28}$$

with $\lambda_t = S_t/\kappa + S_t$. Following the modelling procedure in Bajpai-Dikshit et al. (2003) the bacteria growth rate should also follow the Monod term given in λ_t. Adding a diffusion term to dE_t gives the full SDE description

$$d\begin{bmatrix} B_t \\ E_t \end{bmatrix} = \begin{bmatrix} E_t\nu\lambda_t B_t \\ (\nu+\beta)\lambda_t - E_t^2\nu\lambda_t - \beta E_t \end{bmatrix} dt + \begin{bmatrix} \sigma_B B_t(T_0-(\eta B_t)^\gamma) & 0 \\ 0 & \tilde{\sigma}_E(E_t) \end{bmatrix} d\boldsymbol{w}. \tag{29}$$

Eq. (28) describes the relative (to the maximum) level of enzymes and E_t should therefore be restricted to the interval $[0,1]$. Even though this is theoretically the case, there is no break down of the dynamics of the equation, even if E_t exceeds 1 (but ν cannot be interpreted as maximum growth rate). However the dynamics do break down if E_t is allowed to be below 0. The remainder of the paper presents results for different choices of λ_t and $\tilde{\sigma}_E(E_t)$.

7. Estimation results

The model presented in Eq. (29) is estimated for different specifications of λ_t and $\tilde{\sigma}_E(E_t)$. The combinations are λ_t equals S_t or including Monod growth and enzyme diffusion proportional to the enzyme level ($\tilde{\sigma}_E(E_t) = \sigma_E E_t$) or enzyme diffusion which ensures enzyme levels to stay within the interval $(0,1)$ ($\tilde{\sigma}_E(E_t) = \sigma_E E_t(1-E_t)$). The likelihood results are summarised in Table 2, while parameter estimates for selected models are given in Table 3 (Model D refers to a deterministic ODE model defined in Eq. (31)).

Likelihood ratio testing cannot be applied, since the models are not nested, and the evaluation is therefore based on AIC and BIC values. The inclusion of an enzyme state without the Monod term and with state proportional enzyme diffusion (Model 1) gives highly significant improvements in AIC and BIC for both Enterococcus and Salmonella and all parameters (not shown) are significant (except enzyme diffusion for Salmonella) for both data sets.

The further inclusion of the Monod term (Model 2) does not give further improvement of AIC and BIC for any of the two datasets (Table 2). The parameters ν and κ are both estimated to very high values (not shown), and there is evidence that the Monod term should not be included.

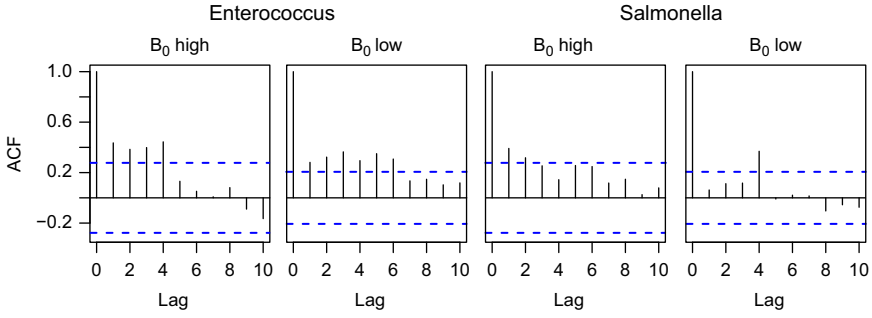

Fig. 4. Autocorrelation function for the standardised residuals in the growth phase of Model 0 (for the training set).

Table 2
Likelihood table for Enterococcus, Δ_0^{AIC} and Δ_0^{BIC} refer to the difference in AIC and BIC compared to Model 0.

Bacteria	λ_t	$\sigma_E(\cdot)$	log (L)	DF	Δ_0^{AIC}	Δ_0^{BIC}
Enterococcus						
Model 1	S_t	$\sigma_E E_t$	766	10	−86.6	−79.6
Model 2	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t$	766	11	−84.6	−74.2
Model 3	S_t	$\sigma_E E_t(1-E_t)$	757	10	−68.5	−61.5
Model 4	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t(1-E_t)$	770	11	−93.1	−82.7
Model 5[a]	S_t	$\sigma_E E_t(1-E_t)$	770	10	−94.6	−87.6
Salmonella						
Model 1	S_t	$\sigma_E E_t$	727	10	−27.5	−20.5
Model 2	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t$	728	11	−27.1	−16.7
Model 3	S_t	$\sigma_E E_t(1-E_t)$	729	10	−30.3	−23.4
Model 4	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t(1-E_t)$	733	11	−36.4	−26.0
Model 5[a]	S_t	$\sigma_E E_t(1-E_t)$	733	10	−38.4	−31.5

[a] The modifications in Model 5 are defined in Eq. (30).

Table 3
Estimation results for Models 0, 5, and D. Standard deviations are given in parenthesis and bold face numbers indicate significant estimates (on a 5% confidence level).

Bacteria	**Enterococcus**			**Salmonella**		
	Model 0	Model 5	Model D	Model 0	Model 5	Model D
Initialisation						
B_0^{high}	**0.051**	**0.055**	**0.099**	**0.033**	**0.036**	**0.044**
	(0.002)	(0.006)	(0.012)	(0.002)	(0.002)	(0.001)
B_0^{low}	**0.007**	**0.011**	**0.014**	**0.002**	**0.005**	**0.004**
	(0.006)	(0.001)	(0.002)	(0.0004)	(0.001)	(0.001)
Drift parameters						
μ	**0.222**			**0.226**		
	(0.010)			(0.012)		
η	**0.934**	**0.921**	**1.143**	**1.091**	**1.092**	**1.146**
	(0.006)	(0.064)	(0.010)	(0.001)	(0.001)	(0.040)
β		**0.204**	1.206		**0.164**	0.774
		(0.077)	(1.558)		(0.054)	(1.200)
ν		**0.366**	**0.183**		**0.302**	**0.416**
		(0.026)	(0.019)		(0.020)	(0.108)
κ			0.056			0.568
			(0.031)			(0.358)
Diffusion parameters						
σ_B	**0.049**	**0.029**		**0.061**	**0.051**	
	(0.003)	(0.003)		(0.004)	(0.003)	
γ	**26.61**	35.891		**79.858**	**98.677**	
	(11.844)	(47.934)		(10.904)	(8.4056)	
$\log(\sigma_E)$		0.086			0.293	
		(0.191)			(0.218)	
Variance of observation noise						
$\log(s_0^2)$	**−11.996**	**−12.138**	**−5.368**	**−11.95**	**−12.254**	**−6.834**
	(0.238)	(0.208)	(0.084)	(0.200)	(0.202)	(0.124)
$\log(s_1^2)$	**−12.588**	**−11.262**	−21.505	**−9.514**	**−9.354**	**−6.203**
	(1.251)	(0.318)	(401.95)	(0.378)	(0.304)	(0.289)

Removing the Monod term and including a logistic diffusion term for the Enzyme level (Model 3) gives improvements in both AIC and BIC for Salmonella, while AIC and BIC increase for Enterococcus (Table 2). The parameter estimate of σ_E also changes significantly (not shown), reflecting that E_t is well determined by the deterministic equation when E_t is close to 0 or 1, while it is dominated by diffusion for E_t away from the boundary.

The further inclusion of the Monod term (Model 4) gives improvements of AIC and BIC for both Enterococcus and Salmonella (Table 2). The parameters κ, ν, γ and $\log(\sigma_E)$ are not significant (not shown). The fact that $\log(\sigma_E)$ is insignificant does not imply that $\sigma_E = 0$, but rather that $\sigma_E = 1$, which is not a reasonable alternative hypothesis. The alternative hypothesis for γ is $\gamma = 1$, which has been tested with a resulting large increase in AIC and BIC for both Salmonella and Enterococcus. The large and insignificant values of κ and ν imply that $\beta+\nu \simeq \nu$ and the consequence is a decoupling of ν and β in the enzyme equation, while there is no real Monod term in the equation for bacteria levels. This implies that a simplified drift term not containing β in the first term might be appropriate, i.e. an enzyme process is given by

$$dE_t = (\nu\lambda_t - E_t^2\nu\lambda_t - \beta E_t)\,dt + \sigma_E E_t(1-E_t)\,dw_E, \tag{30}$$

with $\lambda_t = S_t$ and the bacteria process (29). This model is referred to as Model 5 in the estimation Table 3. AIC and BIC improved for both Enterococcus and Salmonella, and on the basis of the information criteria, the best choice is Model 5.

Rather than providing the estimates of (C_1, C_2), the resulting starting concentrations are given in Table 3, the standard deviations of the starting concentrations are calculated from the variance of C_i and the state variance. In general parameters with the same impact across models $(\eta, \gamma, \sigma_B, s_i^2)$ are relatively constant across models. This implies that estimated starting concentration (C_1 and C_2), onset of the transition to the stationary phase (η), random behaviour of the stationary phase (γ and σ_B) and observation noise (s_i^2) are constant across models. While the (random) bacteria drift term, determined by β, ν, κ and σ_E, varies across models.

For reference we use the deterministic version of the most complex model (Model 4), i.e.

$$d\begin{bmatrix} B_t \\ E_t \end{bmatrix} = \begin{bmatrix} E_t\nu\lambda_t B_t \\ (\nu+\beta)\lambda_t - E_t^2\nu\lambda_t - \beta E_t \end{bmatrix} dt \tag{31}$$

with

$$\lambda_t = \frac{S_t}{\kappa+S_t} = \frac{T_o-\eta B_t}{\kappa+T_o-\eta B_t} \tag{32}$$

and the observation equation given by (18) and (19). This is included in Table 3. It is worth noting that ν and κ are also not significant in the deterministic model. The stochometric constant, η, also changes between models, and the consequence of this change can be seen in Figs. 6 and 7.

8. Cross validation

The models fitted in the previous sections are compared with data from the test set. The performance measures we use in the current section are the autocorrelation function of the one-step predictions (Fig. 5) and the residual sum of squares for each of the models (Table 4).

Comparing autocorrelation functions of the residuals shows large improvements for Enterococcus when comparing Model 0 and the more complex model (Fig. 5), while the improvements between Models 1 and 5 are small (even though evident). These improvements are not evident for Salmonella, one reason might be that the assumption of constant variance of the standardised residuals does not apply. For high starting concentrations the variance increases with time, while the variance decreases for low starting concentrations.

As suggested by the likelihood results, the differences between Models 4 (not shown) and 5 are not visible in an autocorrelation plot. The difference between ODE modelling and SDE modelling is clear from Fig. 5. While the autocorrelation in the residuals for the stochastic models is weak, it is very strong in the deterministic

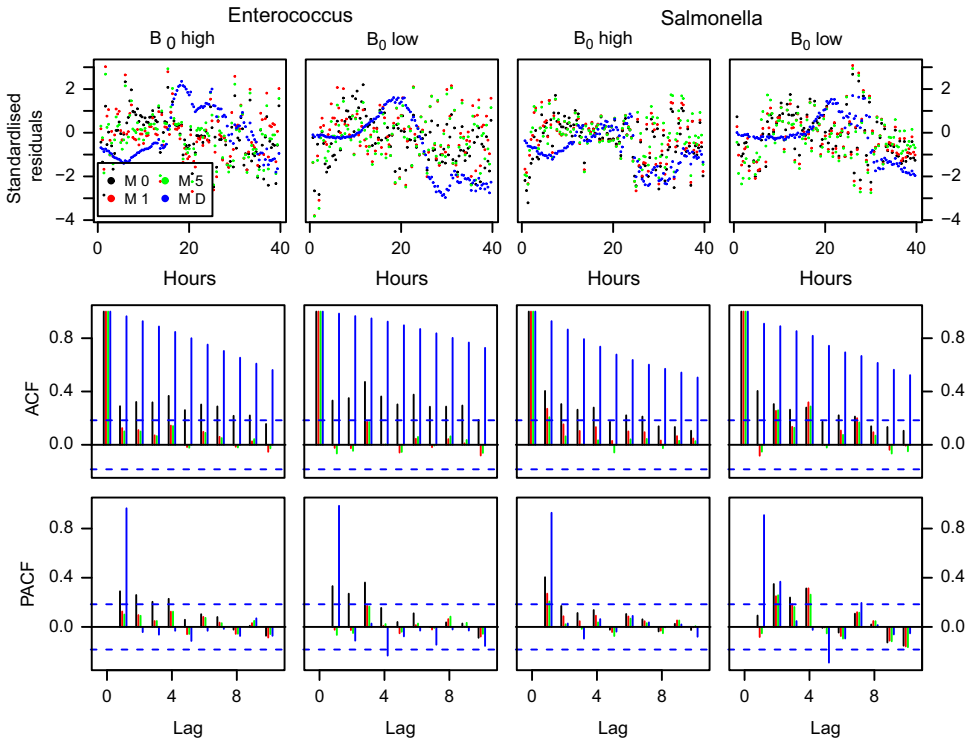

Fig. 5. Standardised residuals from Models 0, 1, and 5 (top row). Bottom rows show autocorrelation functions and partial autocorrelation functions for residuals of each of the models.

Table 4
Square root of residual sum of squares, in the original domian and in the log domain (numbers in parentheses) for the test set and each model, best performing model in each row is marked in bold face.

Model	0	1	2	3	4	5	D
Enterococcus							
B_0 Low	0.232	0.209	0.209	0.203	**0.203**	0.203	0.830
	(0.382)	(0.341)	(0.328)	(0.345)	(0.326)	(**0.325**)	(2.803)
B_0 High	0.164	**0.129**	0.132	0.133	0.132	0.130	1.149
	(**2.033**)	(2.174)	(2.117)	(2.275)	(2.197)	(2.194)	(4.429)
Total	0.284	0.246	0.247	0.243	0.242	**0.241**	1.418
	(**2.069**)	(2.201)	(2.143)	(2.301)	(2.221)	(2.218)	(5.241)
Salmonella							
SB_0 Low	0.325	0.303	0.301	0.289	**0.284**	0.284	1.383
	(0.741)	(0.599)	(0.601)	(0.587)	(**0.577**)	(0.577)	2.766
SB_0 High	0.191	0.177	0.175	0.161	0.157	**0.157**	0.784
	(4.962)	(5.105)	(5.024)	(**4.909**)	(5.342)	(5.342)	(6.730)
Total	0.377	0.351	0.349	0.331	0.325	**0.325**	1.590
	(5.017)	(5.140)	(5.060)	(**4.944**)	(5.373)	(5.373)	(7.276)

model, reflecting that the ODE models are not able to adapt to local information given by data.

The models are compared by the one-step residual sum of squares on the original scale and on log-scale in Table 4, with the residuals given by

$$r_i = OD_{corr,k} - \hat{B}_{t_k|t_{k-1}} \tag{33}$$

$$r_i^{log} = \log(OD_{corr,k}) - \log(\hat{B}_{t_k|t_{k-1}}). \tag{34}$$

Comparing residual sums of squares (Table 4) gives essentially the same conclusion as comparing the likelihoods of the training sets, even though the differences between models are small. The best performing model on the original scale is still Model 5 for both Enterococcus and Salmonella. On the log-scale the conclusion is not clear cut, but it is also on a completely different scale than where the optimisation was performed. The differences between ODE and SDE models are very large on both scales, this is not surprising since the residuals are calculated locally, and local and global SDE residuals are very different, while local and global ODE residuals are equal. The conclusion from Table 4 is therefore that the local dynamics of the data are much better described by an SDE model than the ODE model. The global dynamics are not considered quantitatively, but the next section considers the global dynamics by simulation studies.

9. Simulation study

The purpose of this section is to analyse the distributional behaviour of the best model from Sections 7 and 8 (Model 5). The optimisation in Section 7 is based on the maximum likelihood estimation, which corresponds to one-step predictions (in this case 20 minutes), and therefore a comparison between data and simulated distribution will give information on our ability to predict the distribution of a future experiment. Simulations in this section are performed with an Euler scheme (Kloden and Platen, 1999) on the Lamperti-transformed process.

Since the scope of this section is a discussion of the distributional properties of the models, and not model selection, we will use only visual inspection. It should be noted that quantitative methods for comparing the distributional properties of models like quantile skill scores exist (e.g. Gneiting and Raftery, 2007).

Since the observation noise is small compared to the diffusion of the models, the smoothened state and data will be close and rather than comparing data and estimated distribution, we have chosen to compare the smoothened state and the distribution. A further advantage of this approach is that the smoothened state exists for both bacteria and enzyme concentration.

The unconditional distributions are compared with the smoothened states and the deterministic solution in Figs. 6 and 7. The mode of the unconditional distribution and the smoothened state is quite different for Enterococcus data (Fig. 6), while the deterministic

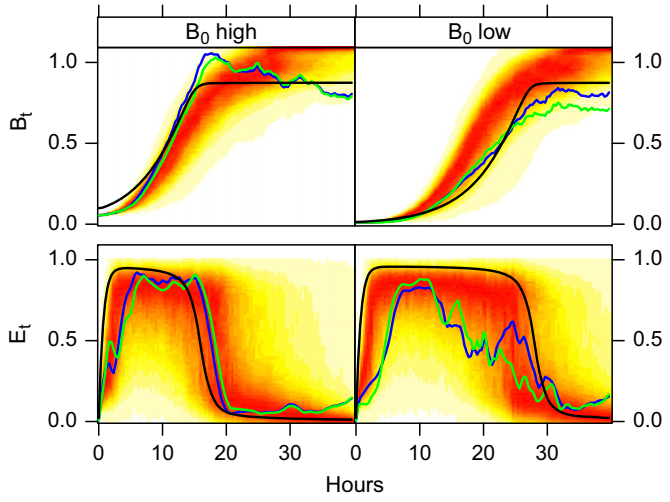

Fig. 6. Smoothened states of the stochastic Model 5 (blue and green lines) of bacteria levels in the test and the training set, and the deterministic solution (black lines) 5 estimated with the Enterococcus data. Background colours indicate the density of simulated data. Left column is high start concentrations and right column is low start concentrations. Parameters are given in Table 3. The density estimates are based on 1000 independent realizations of the process, simulated with the Euler scheme (on the transformed process) with $\Delta t = 1$ s. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

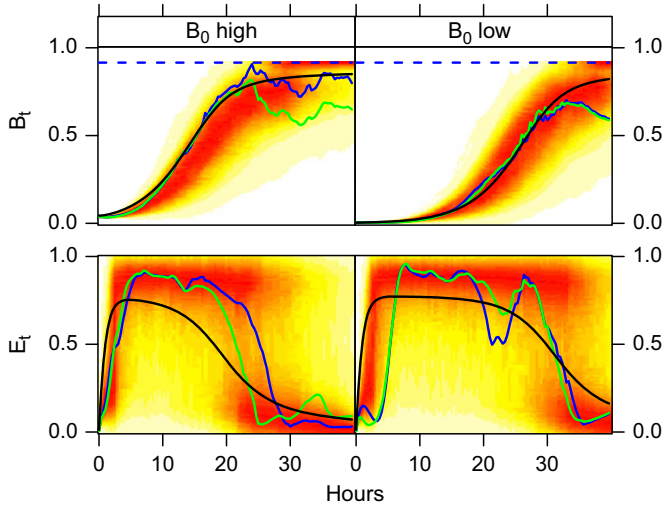

Fig. 7. Smoothened states of the stochastic Model 5 (blue and green lines) of bacteria levels in the test and the training sets, and the deterministic solution (black lines) 5 estimated with the Salmonella data. Background colours indicate the density of simulated data. Left column is high start concentrations and right column is low start concentrations. Parameters are given in Table 3. The density estimates are based on 1000 independent realizations of the process, simulated with the Euler scheme (on the transformed process) with $\Delta t = 1$ s. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

solution fits the first part of the growth phase quite well. It is, however, clear that there are many observations above the maximum for high starting concentrations, which correspond to negative values of the substrate. This problem is not present in the stochastic models. For the bacteria levels (based on Salmonella data) the smoothened state and the mode of the unconditional distribution differ substantially for high starting concentrations, while the performance is better for low starting concentrations (Fig. 7, top row).

The behaviours of unconditional distributions are clearly not satisfactory and indicate that the model description is not fully adequate. Obviously one feature that is missing is the ability to describe the transition to the stationary phase. Data show a small decline in bacteria level after the maximum has been reached, while the model can only predict positive average growth rates. The consequence is that the expected value of bacteria levels from the model will converge to 1 (measure in substrate units). This is seen as the state, where all the substrate has been converted to bacteria acting as an effectively absorbing state.

For the smoothened enzyme levels, we see a very fast increase from zero to values close to one and then a rapid decline to values close to zero when the maximum bacteria levels are reached (Figs. 6 and 7, bottom row). The smoothened state is close to the mode of the unconditional distribution.

10. Discussion

This paper presents an analysis of two sets of bacteria growth data in an SDE setting; traditionally such data are analysed in a deterministic setting. SDE models are rather data-intensive and require data sets with high sampling frequency, such as the data sets presented in the present work. This is clearly a limitation regarding data sets where measurements are taken manually with low sampling frequency. However we believe that automatic measurement techniques similar to the optical density measurements presented here will become more widely used and therefore produce more high-resolution data sets, well suited for SDE based models.

The presented SDE formulation includes a logistic diffusion term and we argue that this inclusion is the most simple diffusion term that obeys the natural constraints of the system. By introducing the power γ, the shape of the diffusion function can be controlled. The large estimated values of γ indicate that the diffusion is proportional to the state, but state-proportional diffusion is not reasonable since such models can produce negative substrate concentrations. The formulation of the diffusion term is based on empirical reasoning and not mechanistic understanding of the cell division process. This is a reasonable approach since the deterministic model is also a lumped model.

Inclusion of a random walk growth rate indicates how the model could be improved by the inclusion of an additional state. The need for an extra state is supported by the analysis of the autocorrelation of the standardised residuals from a simple model with growth rate proportional to the available substrate. The resulting model, which includes an enzyme catalysing growth, is analysed with different enzyme diffusion terms. This analysis stresses that the formulation of enzyme diffusion is important for the conclusions we are able to draw from the analysis. The analysis does not support Monod growth, but rather a decoupling of parameters. In general the analysis emphasises the importance of proper inclusion of the diffusion term.

The setup analysed here does not include bacteria diffusion in the enzyme process. Such an approach gives a simpler and more flexible (regarding Lamperti transforms) setup, and it should be noted that the inclusion of bacteria diffusion in the enzyme process has been tested (with enzyme diffusion proportional to enzyme concentration). As this gives similar results, it was decided to include the simpler setup only, but the derivation of the system equation of the Lamperti transform is given in Appendix A.

Comparing the smoothened state (equivalent to data) and estimated densities show that the models are not very well suited for simulation studies and that further model development might be appropriate. The models lack a mechanistic description of the transitions to the stationary phase, and one possibility would be to exclude some of the data after this transition from the analysis.

In this case, however, we would be faced with the problem of deciding a proper truncation point. Another possibility is the inclusion of a mortality parameter, which seems reasonable from the data, but since knowledge of the mechanistic behaviour in the stationary phase is not available, this is not a trivial task. This drawback is shared with the deterministic model, although the deterministic model did capture the growth phase better than the densities of the stochastic models, but the transition to the stationary phase is not very well captured by the deterministic model either.

Furthermore, it has been considered to include more parameters in the random walk analysis to obtain data-driven hypotheses on the time development of the parameters. As the main scope of this article is proper introduction of the diffusion term, we have chosen to leave these analyses for future studies.

Appendix A. Including bacteria diffusion in the enzyme process

The SDE extension of enzyme level is given by

$$dE_t = (\nu+\beta)\lambda_t\,dt - E_t\,d\log(B_t) - \beta E_t\,dt + \tilde{\sigma}_E(t,\omega_t)\,dw_2 \tag{35}$$

in (28) bacteria levels are only included by the deterministic formulation, but by Itôs' lemma we can also include bacteria diffusion.

By Itôs' lemma we obtain

$$d\log(B_t) = (E\nu\lambda_t - \tfrac{1}{2}(T_0-\eta B_t)^2)\,dt + \sigma_B(T_0-(\eta B)^\gamma)\,dw_1 \tag{36}$$

inserting in (35) gives

$$\begin{aligned} dE_t &= ((\nu+\beta)\lambda_t - \beta E_t)\,dt - E_t(E_t\nu\lambda_t - \tfrac{1}{2}(T_0-(\eta B_t)^\gamma)^2)\,dt \\ &\quad + E_t\sigma_B(T_0-\eta B_t)\,dw_1 + \tilde{\sigma}_E(t,\omega_t)\,dw_2 \end{aligned} \tag{37}$$

$$\begin{aligned} dE_t &= ((\nu+\beta-\nu E_t^2)\lambda_t - E_t(\beta - \tfrac{1}{2}(T_0-(\eta B_t)^\gamma)^2))\,dt \\ &\quad - \sigma_B E_t(T_0-(\eta B_t)^\gamma)\,dw_1 + \tilde{\sigma}_E(t,\omega_t)\,dw_2 \end{aligned} \tag{38}$$

Now choose the transformation

$$Z_t^2 = \log(B_t) + \log(E_t) \tag{39}$$

then

$$dZ_t^2 = \frac{dB_t}{B_t} + \frac{dE_t}{E_t} - \frac{1}{2}\left(\frac{(dB_t)^2}{B_t^2} + \frac{(dE_t)^2}{E_t^2}\right) \tag{40}$$

$$\begin{aligned} dZ_t^2 &= \left(E_t\nu\lambda_t - \frac{1}{2}\sigma_B^2(T_0-(\eta B_t)^\gamma)^2\right)dt + \sigma_B(T0-(\eta B_t)^\gamma)\,dw_1 \\ &\quad + \left(\left(\frac{\nu+\beta}{E_t} - \nu E_t\right)\lambda_t - \beta + \frac{1}{2}\sigma_B^2(T_0-(\eta B_t)^\gamma)^2\right)dt \\ &\quad - \sigma_B(T0-(\eta B_t)^\gamma)\,dw_1 + \frac{\tilde{\sigma}_E(t,\omega_t)}{E_t}\,dw_2 \\ &\quad - \frac{1}{2}\left(\sigma_B^2(T_0-(\eta B_t)^\gamma)^2 + \frac{\tilde{\sigma}_E^2(t,\omega_t)}{E_t^2}\right)dt \end{aligned} \tag{41}$$

$$\begin{aligned} dZ_t^2 &= \left(\frac{\nu+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\left(\sigma_B^2(T_0-(\eta B_t)^\gamma)^2 + \frac{\tilde{\sigma}_E^2(t,\omega_t)}{E_t^2}\right)\right)dt \\ &\quad + \frac{\tilde{\sigma}_E(t,\omega_t)}{E_t}\,dw_2. \end{aligned} \tag{42}$$

If $\tilde{\sigma}_E(t,\omega_t)$ is chosen as proportional to E_t then dZ_2 has constant diffusion and estimation is available through CTSM. Now, we have already seen that a better approach is to choose $\tilde{\sigma}_E(t,\omega_t)$ as the logistic diffusion. Unfortunately the Lamperti transform with an explicit inverse cannot be derived in this case. It is however possible for $\tilde{\sigma}_E(t,\omega_t) = E_t f(z_2)$ where f is a function that is simple enough to allow an explicit inverse of the Lamperti transform. Choose e.g.

$$\tilde{\sigma}_E(t,\omega_t) = E_t(T_0 - \eta e^{Z_t^2}). \tag{43}$$

The state space of E_t now depends on B_t and actually E_t is only restricted to $(0,1)$ when $\eta B_t = 1$.

In this case choose the transformation

$$Z^3 = \psi^3(Z^2) = \frac{1}{T_0}\log\left(\frac{e^{Z^2}}{T_0 - \eta e^{Z^2}}\right), \tag{44}$$

with the inverse given by

$$Z^2 = \log\left(\frac{T_0 e^{T_0 Z^3}}{1+\eta e^{T_0 Z^3}}\right). \tag{45}$$

The derivatives of ψ^3 are given by

$$\psi^3_{z_2} = \frac{1}{T_0 - \eta e^{z_2}} = \frac{1}{T_0 - \eta E_t B_t} \tag{46}$$

$$\psi^3_{z_2 z_2} = \frac{\eta e^{z_2}}{(T_0 - \eta e^{z_2})^2} = \frac{\eta E_t B_t}{(T_0 - \eta E_t B_t)^2} \tag{47}$$

and

$$\begin{aligned} dZ_t^3 &= \left\{\frac{\nu+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\left(\sigma_B^2(T_0-(\eta B_t)^\gamma)^2 + \sigma_E^2\frac{E_t^2(T_0-\eta B_t E_t)^2}{E_t^2}\right)\right\} \\ &\quad \times \frac{dt}{T_0 - \eta E_t B_t} + \sigma_E\,dw_2 + \frac{1}{2}\frac{\eta E_t B_t}{(T_0-\eta B_t E_t)^2}\sigma_E^2\frac{E_t^2(T_0-\eta E_t B_t)^2}{E_t^2}\,dt \end{aligned} \tag{48}$$

$$\begin{aligned} dZ_t^3 &= \left\{\frac{1}{T_0-\eta E_t B_t}\left(\frac{\nu+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\sigma_B^2(T_0-(\eta B_t)^\gamma)^2\right)\right. \\ &\quad \left. -\frac{1}{2}\sigma_E^2(T_0-\eta B_t E_t)\right\}dt + \sigma_E\,dw_2 + \frac{1}{2}\eta\sigma_E^2 E_t B_t\,dt \end{aligned} \tag{49}$$

$$\begin{aligned} dZ_t^3 &= \left\{\frac{1}{T_0-\eta E_t B_t}\left(\frac{\nu+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\sigma_B^2(T_0-(\eta B_t)^\gamma)^2\right)\right. \\ &\quad \left. +\frac{1}{2}\sigma_E^2(2\eta B_t E_t - T_0)\right\}dt + \sigma_E\,dw_2 \end{aligned} \tag{50}$$

with the inverse

$$E_t = \frac{e^{z_t^2}}{B} \tag{51}$$

$$E_t = e^{z_t^2}\frac{\eta}{T_0}\frac{(1+e^{\gamma T_0 Z_t^B})^{1/\gamma}}{e^{T_0 Z_t^B}} \tag{52}$$

$$E_t = \frac{\eta}{T_0}\frac{T_0 e^{T_0 Z_t^3}}{1+\eta e^{T_0 Z_t^3}}\frac{(1+e^{\gamma T_0 Z_T^B})^{1/\gamma}}{e^{T_0 Z_t^B}} \tag{53}$$

$$E_t = \eta\frac{e^{T_0 Z_t^3}}{1+\eta e^{T_0 Z_t^3}}\frac{(1+e^{\gamma T_0 Z_t^B})^{1/\gamma}}{e^{T_0 Z_t^B}}. \tag{54}$$

Actually the transformed process does not have a reasonable asymptotic behaviour unless some restrictions are imposed on the relation between β and ν. Also an attempt to estimate through CTSM (with no restriction on the relation between β and ν) leads to a break down of the estimation procedure, illustrating the importance of considering the asymptotic behaviour.

A.1. Required relation between β and ν

We consider the case where $\lambda_t = S_t$. Furthermore we set $T_0 = 1$, the critical point is $\eta E_t B_t = 1 \Rightarrow E_t = 1/\eta B_t$, if this should work for arbitrary σ_B, we get the restriction

$$\frac{\nu+\beta}{E_t}(1-\eta B_t) - \beta < 0, \tag{55}$$

inserting the critical point ($E_t = 1/\eta B_t$) gives

$$\eta B_t(\nu+\beta)(1-\eta B_t)-\beta < 0, \tag{56}$$

the left had side is a quadratic polynomial in ηB_t with its maximum at $\eta B_t = \frac{1}{2}$, and we get the restriction

$$\frac{\nu+\beta}{2E_t}-\beta < 0, \tag{57}$$

and the critical point is $E_t \to 2$ and we get

$$\frac{\nu+\beta}{4}-\beta < 0 \Rightarrow \beta > \frac{1}{3}\nu. \tag{58}$$

So in this case the required asymptotic behaviour is ensured when $\beta > \frac{1}{3}\nu$, of course the appropriateness of the model has not been discussed, and we have not attempted to incorporate these restrictions in the model.

References

Baadsgaard, M., Nielsen, J.N., Spliid, H., Madsen, H., Preisel, M., 1997. Estimation in stochastic differential equations with state dependent diffusion term. In: SYSID '97—11th IFAC Symposium on System Identification, IFAC.

Bajpai-Dikshit, J., Suresh, A.K., Venkatesh, K.V., 2003. An optimal model for representing the kinetics of growth and product formation by *Lactobacillus rhamnosus* on multiple substrates. J. Biosci. Bioeng. 96 (5), 481–486.

Doshi, P., Rengaswamy, R., Venkatesh, K.V., 1997. Modelling of microbial growth for sequential utilization in a multisubstrate environment. Process Biochem. 32 (8), 643–650.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.

Iacus, S.M., 2008. Simulation and Inference for Stochastic Differential With R Examples. Springer Series in Statistics.

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York, ASU.

Kloden, P., Platen, E., 1999. Numerical Solutions of Stochastic Differential Equations. Springer-Verlag.

Kompala, D.S., Ramkrishna, D., Tsao, G.T., 1984. Cybernetic modelling of microbial growth on multiple substrates. Biotechnol. Bioeng. 26 (11), 1272–1281.

Kovárová-Kovar, K., Egli, T., 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol. Mol. Biol. Rev. 62 (3), 646–666.

Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004a. Parameter estimation in stochastic grey-box models. Automatica 40, 225–237.

Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004b. A method for systematic improvement of stochastic grey-box models. Comput. Chem. Eng. 28, 1431–1449.

Kristensen, N.R., Madsen, H., 2003. Continuous Time Stochastic Modelling – CTSM 2.3 – Mathematics Guide. Technical University of Denmark.

Liu, Y., 2006. A simple thermodynamic approach for derivation of a general Monod equation for microbial growth. Biochem. Eng. J. 31, 102–105.

Lobry, J., Flandrois, J., Carret, G., Pave, A., 1992. Monod's bacterial growth model revisited. Bull. Math. Biol. 54, 117–122.

Luschgy, H., Pagés, G., 2006. Functional quantization of a class of Brownian diffusions: a constructive approach. Stoch. Proc. Appl. 116, 310–336.

Madsen, H., Holst, J., Thyregod, P., 1987. A Continuous Time Model for the Variations of Air Temperature, 10. In: Conference on Probability and Statistics in Atmospheric Science. American Meteorological Society, 52–58, Edmonton.

Madsen, H., 2008. Time Series Analysis. Chapman & Hall/CRC.

Monod, J., 1949. The growth of bacterial cultures. Ann. Rev. Microbiol. 3, 371–394.

Øksendal, B., 2003. Stochastic Differential Equations—An Introduction with Applications, 6th ed. Springer, Berlin.

Patnaik, P.R., 1999. Transient sensitivity analysis of a cybernetic model of microbial growth on two substrates. Bioprocess Eng. 21 (2), 135–140.

Pedersen, M.W., Righton, D., Thygesen, U.H., Andersen, K., Madsen, H., 2008. Geolocation of North Sea cod using Hidden Markov Models and behavioral switching. Can. J. Fish. Aquat. Sci. 65, 2367–2377.

Philipsen, K.R., Christiansen, L.E., Hasman, H., Madsen, H., 2010. Comparison of calibration curves for the relation between optical density and viable count bacteria data. J. Theor. Biol. 263, 134–142.

Philipsen, K.R., Christiansen, L.E., Mandsberg, L.F., Hasman, H., Madsen, H. Comparison of calibration curves for the relation between optical density and viable count bacteria data, submitted for publication.

Schurz, H., 2007. Modelling, analysis and discretization of stochastic logistic equations. Int. J. Numer. Anal. Modelling 4 (2), 178–197.

Tornøe, C.W., Jacobsen, J., Pedersen, O., Hansen, T., Madsen, H., 2004. Grey-box modelling of pharmacokinetic/pharmacodynamic systems. J. Pharmacokinet. Pharmacodyn. 31, 401–417.