Stochastic modeling for characterisation of biofilm development with discrete detachment events (sloughing)

A. Bohn*, B. Zippel**, J.S. Almeida**** and J.B. Xavier*

*Biomathematics Group, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, R Quinta Grande 6, 2780-156 Oeiras, Portugal (E-mail: abohn@itqb.unl.pt; jxavier@itqb.unl.pt)

**Department of River Ecology, UFZ Centre for Environmental Research Leipzig-Halle, Brückstr. 3a, 39114 Magdeburg, Germany (E-mail: barbara.zippel@ufz.de)

***Department of Biostatistics and Applied Mathematics, University of Texas MDAnderson Cancer Center – unit 447, 1515 Holcombe Blvd, Houston, TX 77030-4009, USA (E-mail: jalmeida@mdanderson.org)

Abstract The monitoring of biofilm development at a small-scale is often observed to be a stochastic process. This raises important issues concerning the reproducibility of biofilm growth monitoring experiments. By realising that there are limits to the latter, a model of biofilm accumulation curves that takes into account the dynamics of seemingly random fluctuations resulting from sloughing events is proposed. The model is derived from a stochastic differential equation (SDE) based on the logistic equation, adding a stochastic term for the sloughing events and measurement noise. Experimental light absorbance data that correlate with biofilm biomass obtained from the development of phototrophic biofilms are analysed to illustrate the use of SDE modeling.

Keywords Phototrophic biofilms; sloughing; stochastic differential equations

Introduction

The issue of reproducibility of biofilm structure is addressed frequently when biofilm is monitored at relatively small scales (Heydorn et al., 2000; GrayMerod et al., 2005). It has been noted that biofilm development is, to some extent, a stochastic process and that therefore independent experiments, even executed under the same conditions, never produce the same exact structure (Heydorn et al., 2000). This is certainly to be expected from the random nature of several of the multiple physical and biological processes involved in biofilm formation. For instance, the process of inoculation of flowcells for biofilm cultivation, i.e. the transport of bacterial cells to an initially clean surface by convection or Brownian motion, is a stochastic process by itself. Other processes include phenotypic variance in the development of bacterial cells and the influence of environmental conditions that are themselves non-deterministic, such as hydrodynamic shear from turbulent flow. Sloughing, i.e. discrete detachment events apparently occurring at random times in which significant amounts of the biomass are lost, is one of the manifestations of the stochastic nature of biofilm development (Lewandowski et al., 2004).

In spite of such randomness, large scale features of biofilm development that result from averaging over large areas can be reasonably conserved. Owing to this, deterministic models, e.g. by Wanner and Gujer (1986), have been used together with empirical parameters, such as maximum specific growth rates ($\mu_{\text{max}}$) or detachment rates, to successfully describe experimental observations, as for instance performed by Horn and Hempel (1997). A more difficult task is at hand, however, when the purpose is to interpret data monitored at a small scale, for which the stochastic nature does not average out.
The present work proposes a method to interpret data from apparently erratic time series from online measurements. The method combines a deterministic model of the growth curves with a stochastic term to describe sloughing events. Model parameters may be estimated from time series such as growth curves using maximum likelihood evaluation. Here, the application of the method is illustrated by analysing experimental time series of light absorbance measured online during the development of phototrophic biofilms.

Methods

Time series of light absorbance from phototrophic biofilms

Time series originate from phototrophic freshwater biofilms cultivated in laboratory incubators in the context of the EU-project PHOBIA, as described in detail by Zippel and Neu (2005). Irradiance was set to 120 μmol m⁻² s⁻¹, temperature to 20°C, and flow rate to 100 L/h. Monitoring of biofilm development was carried out by online measurements of light absorbance at nine locations in each incubator lane. The biomass accumulated on the incubator surface at a given location is related to absorbance in an exponential fashion (Zippel et al., submitted). The data analysed in this work were produced by temporally averaging the online data over each day, up to 36 days. They thus consist of 36 time points at a sampling frequency of 1 day⁻¹, with nine replicates for a single experimental run.

Stochastic model for growth curve interpretation

The stochastic model used here models signal growth assuming a logistic curve in the absence of sloughing. The logistic dynamics is used here as a “black-box” representation for biofilm accumulation, i.e. a non-mechanistic model, and was chosen primarily for reasons of simplicity, since its differential form requires only two parameters, μ and b.

\[
\frac{dX}{dt} = \mu X - bX^2
\]  

(1)

The variable X may represent any observable that is correlated to the quantity of biofilm accumulated, such as biomass, biofilm thickness or, as in the case of the experimental data used here, light absorbance. Equation (1) represents first order growth (μX) and second order decay (−bX²) of the signal being analysed. The value \(X_{\text{max}} = \mu/b\) is usually termed the carrying capacity (note that the right hand side of equation 1 becomes 0 for \(X = X_{\text{max}}\)). μ is a net specific growth rate that describes increase in absorbance signal resulting from all biomass accumulation (growth and attachment), and b is a loss rate coefficient that describes a slowing down of the absorbance signal increase. The latter may result from the achievement of a maximum biofilm accumulation resulting from a balance of accumulation (from growth and attachment) and losses such biomass decay and detachment by erosion. The term “erosion” is commonly used for biomass-detachment events that occur with a large frequency and, therefore, for modelling purposes, may be considered a continuous biomass loss process. This is in contrast to “sloughing”, which describes biomass detachment events that occur less frequently, and hence are considered discrete events for modeling purposes. The distinction between erosion and sloughing is arbitrary (Stewart, 1993). Here, events that occur frequently in relation to the data sampling frequency \(\Delta t^{-1}\) (where \(\Delta t\) is the sampling time interval), such that they cannot be discriminated from the growth curve, are considered erosion, and hence their effect is included in parameters μ and b. Sloughing, in turn, is modeled using a stochastic term \(\xi(t)\). Extending equation (1) with this term produces the following...
stochastic differential equation (SDE):
\[
\frac{dX}{dt} = \mu X - bX^2 - \xi(t)X,
\]
which is solved with the initial condition
\[
X(t = 0) = X_0
\]
Similar to the case of “white noise”, which is a normally distributed random variable that is the time derivative of the Wiener process, ref. (Higham, 2001), the stochastic term \(\xi(t)\) is here defined as the time derivative of a process \(S\) that causes the sloughing. A discrete form of equation (2) is used here to solve the SDE numerically using an Euler integration scheme with step size \(\delta t\):
\[
X(t + \delta t) = X(t) + \delta t(\mu X(t) - bX(t)^2) - sX(t),
\]
where \(s\) is a random variable that represents the integral:
\[
\int_t^{t+\delta t} X(t) \, dS = sX(t).
\]
Rather than defining the process \(S\) directly, the discrete form, \(s\), is here defined through the expression:
\[
s = u^\alpha,
\]
with \(u\) being a uniformly distributed random number. Equation (6) implements the assumption that sloughing events causing the loss of large fractions of the biofilm are highly unlikely, whereas smaller sloughing events are more frequent. This is a central assumption of our model. The physical significance of the term \(-sX(t)\) in equation (4) is the loss of biomass from detachment in the interval of time that goes from \(t\) to \(t + \delta t\). Since \(s\) can have any value in the \([0,1]\) interval, \(-sX(t)\) means that any amount of biofilm up to the value of \(X(t)\) may be lost in that interval. Solving equation (4) also requires the definition of an integration time step, \(\delta t\). The existence of two relevant time scales in SDE solutions, the integration step (\(\delta t\)) and the sampling step (\(\Delta t\)), is a detail of extreme importance, as noted by Timmer (2000). In particular, a value of \(\delta t\) that is significantly smaller than \(\Delta t\) should be used. For comparison with the experimental time series sampled at \(\Delta t\) intervals, \(\delta t\) should be chosen such that \(\Delta t\) is an integer multiple of \(\delta t\). Each time point of interest is therefore obtained from the time series:
\[
X_n = X(n\Delta t) + N(0, \sigma),
\]
which results from the superposition of \(X\) at points \(n\Delta t\) and the experimental noise, here modelled as a normal distributed random variable with zero mean and standard deviation \(\sigma\).

Evaluation of suitability of model parameters
The differential form of the SDE requires a set of three model parameters: \(\mu\), \(b\) and \(\alpha\). The suitability of a parameter set to represent time series data \(\{x_1, x_2, \ldots, x_N\}\) with \(N\) points will be assessed using the sum of the logarithm of the likelihood (Timmer, 2000):
\[
L(x_1, x_2, \ldots x_N|\mu, b, \alpha) = \sum_{i=1}^{N-1} \ln p(x_{i+1}|x_i, \mu, b, \alpha)
\]
Here, \(\{x_1, x_2, \ldots, x_N\}\) represent the actual measurements at times \(\{t, t + \Delta t, \ldots, t + N\Delta t\}\). The conditional densities, \(p(x_{i+1}|x_i, \mu, b, \alpha)\), represent in integral terms the probability of a data point of value \(x_{i+1}\) following a data point of value \(x_i\) evaluated for parameters \(\mu, b, \alpha\). These conditional density values are here estimated from a simulated distribution. For each data pair \((x_n, x_{n+1})\) the integration equation \((4)\) is used several times to create a distribution of simulated values for \(x_{i+1}\). The conditional density of \(x_{i+1}\) of the actual data point is then determined by kernel smoothing, e.g. using the \texttt{ksdensity}\ function in the \textsc{Matlab}\textsuperscript{TM} Statistical Toolbox. The L-value obtained from summing the logarithm of all experimental data pairs provides a measure for the quantitative assessment of the suitability of a given parameter set for the description of a particular experimental time series.

**Parameter estimation**

The estimation of parameters for non-linear SDEs is a recent field of research. In particular, most of the literature refers to systems in which the stochastic term follows a normal distribution, and uses simplifications that are not applicable to the system analysed here. Therefore, maximum likelihood estimation (MLE) is used for parameter estimation. Applying MLE for a non-linear SDE is a computationally intensive process that requires the combination of optimisation algorithms with the determination of the L as the function to be maximised. In order to simplify the procedure, the parameter space is here reduced to two parameters \(\mu, \alpha\). This is possible by coupling the value of \(b\) to \(\mu\) and \(x_{\text{max,obs}}\), the maximum value for \(X\) in the data being evaluated:

\[
b = \frac{\mu}{x_{\text{max,obs}}} \quad (9)
\]

The kernel smoothing for the estimation of the conditional probabilities required for equation \((8)\) was implemented with a normal distribution kernel with a bandwidth of 0.1. Its physical significance consists in the assumption that the measurement noise is normally distributed with standard deviation \(\sigma = 0.1\).

**Results and discussion**

**Figure 1** (panels a–i) shows nine light absorbance time series measured online by the nine sensors in a single phototrophic biofilm experiment. An apparently erratic time course of the light absorbance signal \((X)\) is evidenced together with a reduced reproducibility of the trends among the nine sensors. Light absorbance correlates exponentially with the biomass in the biofilm \((\text{Zippel et al., submitted})\), with increasing light absorbance indicating increased biofilm biomass. Typically, time series show periods indicative of biomass increase (increase in signal), interrupted by rare events of sudden loss resulting from sloughing. These losses are of variable size, ranging up to decreases in light absorbance signal of 40\% in a 1 day period. Since the sampling time is 1 d, it is conceivable that more than one sloughing event may occur during that period, meaning that observed losses may result from a summation of sloughing events. Furthermore, the change in signal \((\Delta X_n)\) between two consecutive time points \((\Delta X_n = X_{n+1} - X_n)\) combines effects from both growth and sloughing. It is therefore possible that small sloughing events still occur in a time interval for which \(\Delta X > 0\). In addition to this, measurement noise, i.e. uncertainty on the value of any recorded \(X_n\), must be considered.

The dynamics of biomass accumulation, sloughing and measurements noise are reflected in a combined fashion in the time series. This situation turns the decomposition of \(X\) into these principal parts into a challenging task. The stochastic model presented
above includes these three effects. By using an integration time step $\delta t = 0.01\Delta t$, simulations account for cumulative effects of sloughing and growth in each 1 day interval.

Figure 2 shows the results of the parameter fit of the stochastic model to the full experimental data set shown in Figure 1. Figure 2(a) shows the value of $L$ for a scanned parameter space, with $\mu$ in $[0.1, 0.5]$ and $\alpha$ in $[100, 10,000]$. The optimal parameter set ($\mu = 0.28, \alpha = 1130$) was obtained by maximising $L$.

**System dynamics reflected in the conditional probability**

The conditional probability density function (CPDF) is of chief importance for understanding the dynamics of the system. It reflects the probability that a biofilm in the state $X_n$ at
t + nΔt proceeds to a certain $X_{n+1}$ at $t + (n + 1)\Delta t$, i.e. on the next day. Figure 2(b) shows the CPDF for $X_{n+1}$ given $X_n = 1.5$. This CPDF was computed from the optimal parameter set and shows an asymmetrical shape where the mode (most probable value), in this case $X_{n+1} = 1.73$, differs slightly from the expect value, here $E(X_{n+1}) = 1.69$, because of sloughing. The extended tail for low values of $X_{n+1}$ reflects the fact that, although sloughing events of any specific size are highly improbable, some probability of sloughing events occurring within 1 day exists. In this particular case, a loss of biomass by sloughing of around 2% is expected over the next day. Note that the fact that CPDF >0 for some values where $X_{n+1} < 0$ does not mean that the model used here allows negative biomass values, but is rather a consequence of measurement errors which are also accounted for in the model.

Simulations using the stochastic model

Once the model parameters are determined, an arbitrary number of simulated time series may be generated. Figure 3(a) shows a typical time series (continuous line) obtained by solving the model with $X(0) = 0.01$ and, in the same plot, a deterministic curve (dashed line) defined by the same parameters (shown time shifted), but in the absence of sloughing events and measurement noise.

It must be stressed that the approach presented here is a minimal description using the simplest mathematical framework possible. Future modeling endeavours could achieve a closer match with experimental data, for instance by using specific mechanistic models for the deterministic growth equation instead of the logistic model, and comprising customised descriptions of the sloughing size-distribution.

The phase-space representation

Figure 3(b) shows the phase-space representation of the full experimental data set from Figure 1, in the continuous lines. The phase-space plots each data pair $(X_n, X_{n+1})$, allowing a comparison of the systems dynamics that is independent of the initial state, i.e. $X(t = 0)$. Also represented in Figure 3(b) (dashed line) is the phase-space representation of the deterministic logistic curve obtained from the optimal parameter set, representing
the system dynamics in the absence of sloughing events and measurement noise. As a consequence of the sloughing events, the experimental data show mainly negative deviations from deterministic logistic growth. Positive deviations and a small fraction of the negative deviations are possibly due to measurement noise.

Conclusions
The stochastic model and parameter estimation procedure presented here assists in the interpretation of time series of biofilm biomass accumulation that show seemingly random perturbations, allowing the extraction of parameters quantifying the underlying biofilm accumulation and detachment. This method may be used for the characterisation of biofilm dynamics directly from time series such as growth curves measured online and possibly as an expedite alternative to other methods for quantifying detachment, such as measuring detached particles collected at reactor effluent, e.g., Choi and Morgenroth (2003), whenever online growth curve data with high spatial and temporal resolution are available.

Although the model is based on the logistic equation to describe continuous biomass accumulation and makes specific assumptions on the size distribution of sloughing events and measurement noise, the same principle may be used with alternative formulations. At present, 1-d mechanistic models with purely deterministic dynamics are widely used tools in biofilm studies. Some of these approaches used thus far take into account the effect of sloughing events by either using average net detachment rates, as, e.g., Stewart (1993), or by defining explicit sloughing events using predefined lookup-tables (Morgenroth and Wilderer, 2000) that state the occurrence of sloughing events at specified times. In the case of multidimensional mechanistic models, at least some stochastic component using random number generators is implemented (Picioroaeu et al., 2001; Hunt et al., 2003; Xavier et al., 2005, among others). The phenomenological stochastic modelling approach proposed here is a prototype tool to describe these stochastic events, together with a method for parameter estimation, which should complement and improve the established mechanistic modeling approaches in providing further insight on the characteristics of sloughing events and the stochastic nature of biofilm development.

Acknowledgements
The authors thank Francisco Pinto for fruitful discussions on the conditional probability distribution functions and the kernel smoothing. This work was funded within the project PHOBIA by the European Commission under the 5th Framework Programme “Quality of Life, Key Action 3: The Cell Factory” (QLK3-CT-2002-01938). AB and JBX were financially supported by the FCT/MCTES, Portugal, through the grants SFRH/BPD/25967/2005 and SFRH/BPD/11485/2002, respectively.

References


