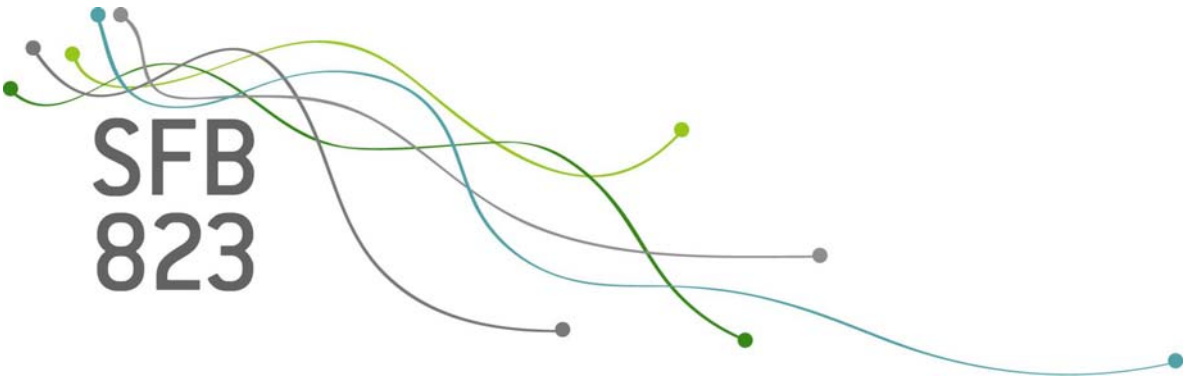
SFB 823

Discussion Paper

Prediction of crack growth based on a hierarchical diffusion model

Simone Hermann, Katja Ickstadt, Christine H. Müller

Nr. 4/2015



Prediction of Crack Growth Based on a Hierarchical Diffusion Model

Simone Hermann, Katja Ickstadt and Christine H. Müller
Faculty of Statistics, TU Dortmund University

January 28, 2015

Abstract

A general Bayesian approach for stochastic versions of deterministic growth models is presented to provide predictions for crack propagation in an early stage of the growth process. To improve the prediction, the information of other crack growth processes is used in a hierarchical (mixed-effects) model. Two stochastic versions of a deterministic growth model are considered. One is a nonlinear regression setup where the trajectory is assumed to be the solution of an ordinary differential equation with additive errors. The other is a diffusion model defined by a stochastic differential equation (SDE) where increments have additive errors. Six growth models in the two versions are compared with respect to their ability to predict the crack propagation in a large data example. Two of them are based on the classical Paris-Erdogan law for crack growth, and four are other widely used growth models. It turned out that the three-parameter Paris-Erdogan model and the Weibull model provide the best results followed by the logistic model. Suprisingly, the SDE approach has no advantage for the prediction compared with the nonlinear regression setup.

Keywords: Fatigue propagation, Paris-Erdogan equation, stochastic differential equation, Euler-Maruyama approximation, Bayesian estimation and prediction

1 Introduction

In many research areas of engineering, material fatigue plays an important role. Experiments are often very expensive because they take a long time and the constructions are costly. To extract as much information as possible from the existing experiments and to predict fatigue, statistical models are a valuable tool. One of these existing experiments was conducted by Virkler et al. (1979). Sixty-eight replicate constant amplitude tests in aluminum alloy were carried out to investigate the fatigue crack propagation. In each of these tests, the number of cycles that leads to fixed crack lengths was observed. Against the natural assumption that something is observed at fixed times, here the time is the dependent variable and the crack length is the independent variable.

Many contributions to the literature consider crack growth modeling. We will restrict on the articles that developed statistical methods for the data set of Virkler et al. (1979) being aware that this will not provide a comprehensive overview. The here presented papers all build their ideas on the Paris model

$$\frac{dL}{dT} = CL^{\frac{m}{2}}, \tag{1}$$

where L denotes the crack length, T the time in cycle counts and C and m are material dependent constants. Their approach for a statistical model is to multiply the equation's right hand side by a stochastic process. Chiquet et al. (2009) include this idea in the framework of piecewise deterministic Markov processes. The used Markov process is a stationary jump process with finite state space and between these jumps, a deterministic behaviour is assumed. Lin and Yang (1983) approximate the crack length process by a diffusion process and, additionally, determine the probability for a random arrival time at a fixed crack length. The model parameters are estimated with standard regression methods which are plugged into the process formula and then the process is simulated. From the simulations they derive predictions and heuristic prediction intervals. Ortiz and Kiremidjian (1988) multiply a zero mean stationary Gaussian process on the right hand side of equation (1) and apply the logarithm on both sides. The result is a lognormally distributed, additive error process for the logarithm of the differences. Additionally, they calculate a probability distribution for the amount of cycle counts that is needed to arrive

at a fixed crack length. However, their application to the data reveals that a distinction between short, medium and long crack lengths is needed. Ray and Tangirala (1996) propose a similar model with lognormally distributed crack growth rates. They generate expected values from the extended Kalman filter and compare the first two moments of the crack length with the model results. Any further estimation or prediction theory is missing. A newer approach to predict crack growth is proposed by Zárate et al. (2012). The L on the right hand side of equation (1) is replaced by a polynomial of L, whose parameters are estimated with Bayesian methods. In Wu and Ni (2004), we mainly find a comparison of two experimental data sets, namely the one of Virkler et al. (1979) also used in the present paper, as well as one under random loading produced by themselves. In addition, a comparison of three models is made, a Markov chain model, a polynomial model, and a model based on the Paris equation called Yang's power law model. But, just like Chiquet et al. (2009) and Zárate et al. (2012), they ignore the fact that the independent variable in the Virkler data is the crack length. A good overview of possible approaches modeling fatigue processes in general is given in Sobczyk and Spencer (1992).

In the above mentioned works, either each series is modeled on its own, or all series are taken together for modeling. The disadvantages are, that either the information of the other series cannot be used in the estimation or a very high variance is implied. Both can be avoided by using a hierarchical mixed model, also called mixed-effects model in frequentist statistics. That means, each series is taken as an observation of one individual. Every single individual follows the same law, for example, a nonlinear parametric regression model or a diffusion process where the model parameter can vary with each individual. The various parameters are assumed to be distributed according to a known family whose parameters are estimated. This approach gives the opportunity to analyze intra and interindividual variances separately. As a result we have a lower intraindividual variance which helps to get a more precise prediction for each individual.

For nonlinear parametric regression, mixed models are well-known. For example, Demidenko (2013) gives a good overview whereas Zimmermann and Núñez-Antón (2001) explicitly focus on growth curve data, both with frequentist estimation methods. Ditlevsen and de Gaetano (2005) and Picchini et al. (2010) apply maximum likelihood estimation to a

mixed-effect model defined by stochastic differential equations. The present work is based on the ideas of Oravecz and Tuerlinckx (2011) and Donnet et al. (2010). The first one compares a linear mixed model with the hierarchical Ornstein-Uhlenbeck model. The second one investigates a hierarchical nonlinear mixed model that is extended by replacing the deterministic regression function by a stochastic process. This leads to a two error model, one error resulting from the stochastic process and the other one is additive. The special case of the hierarchical Ornstein-Uhlenbeck model has been investigated by Oravecz et al. (2009) with Bayesian methods.

None of these approaches deals with prediction of future observations for one of the available series in an early stage of this series. However, the prediction of the crack propagation in an early stage is a very important problem. Therefore, this paper provides a Bayesian estimation and prediction procedure to derive distributions for future observations, and thus prediction intervals, in a hierarchical model. We propose and compare two models with respect to their qualification to predict the material fatigue process. Firstly, a nonlinear mixed regression model is used, whose regression function is the solution of an ordinary differential equation (ODE), and secondly, the closely related mixed diffusion model with matching drift function is considered. The aim of our work is to provide a tool box for growth processes in a general setup. Therefore, we will restrict to growth curves but the estimation and prediction procedure can also be used for any other function that leads to a uniquely existing solution of the SDE. Conservative crack growth models presented in the above mentioned engineering literature will be compared to famous growth functions, i.e. the Gompertz, logistic, Richards and Weibull function.

The remainder of this article is structured as follows. The next section introduces the models and the third the corresponding Bayesian estimation. The fourth section presents the prediction theory and the fifth section includes the application on the data set as well as a simulation study to validate the proposed procedure. Finally the results are summarized and an outlook for future work is given.

2 Models

As mentioned in the introduction we will have a further look on the differential equation (1) called Paris or Paris-Erdogan equation. Parameterized another way we get

$$\frac{dL}{dT} = \theta_1 L^{\theta_2}.$$

Since the cracks grow explosively over time, we assume $\theta_1 > 0, \theta_2 \geq 1$. Therefore, this differential equation is solved by

$$L(T) = \begin{cases} \{\theta_1(\theta_2 - 1)(\theta_0 - T)\}^{\frac{1}{1-\theta_2}}, & \theta_2 > 1 \\ \widetilde{\theta}_0 \exp(\theta_1 T), & \theta_2 = 1 \end{cases}$$

for some θ_0 with $T < \theta_0$ and $\widetilde{\theta}_0 > 0$. This can easily be seen by differentiation. Since in the data of Virkler et al. (1979) the length L is the explanatory variable and the time T is the dependent variable, we have to invert $L(T)$. This yields

$$T(L) = \begin{cases} \theta_0 - \frac{1}{\theta_1(\theta_2-1)} L^{1-\theta_2}, & \theta_2 > 1 \\ \frac{1}{\theta_1}\left(\log(L) - \log(\widetilde{\theta}_0)\right), & \theta_2 = 1 \end{cases}$$

with corresponding derivations that lead to the differential equations

$$\frac{dT(L)}{dL} = \begin{cases} \frac{1}{\theta_1} L^{-\theta_2} = \frac{\theta_2 - 1}{L} \cdot \frac{1}{\theta_1(\theta_2-1)} L^{1-\theta_2} = \frac{\theta_2-1}{L} \cdot (\theta_0 - T(L)), & \theta_2 > 1 \\ \frac{1}{\theta_1 L} = \frac{1}{L\log(L)} \cdot \frac{1}{\theta_1} \log(L) = \frac{1}{L\log(L)} \cdot (T(L) + \frac{1}{\theta_1}\log(\widetilde{\theta}_0)), & \theta_2 = 1. \end{cases}$$

By reparameterization we obtain

$$\text{Paris 1 } (\theta_2 > 1): \qquad f(t,\phi) = A - Bt^{-C}; \qquad b(\phi,t,y) = \frac{C}{t}(A - y),$$

$$\text{Paris 2 } (\theta_2 = 1): \qquad f(t,\phi) = A\log(t) - B; \qquad b(\phi,t,y) = \frac{1}{t\log(t)}(y + B),$$

with $\phi = (A, B) \in (0,\infty) \times \mathbb{R}$, and $\phi = (A, B, C) \in (0,\infty)^3$ respectively, allowing for a unified representation of the differential equation

$$\frac{df}{dt}(t,\phi) = b(\phi, t, f(t)), \;\; f(t_0) = y_0.$$

In addition to the growth models evolved above we can also regard other models resulting from such an ODE. Here, we particularly consider the following functions:

Gompertz:	$f(t,\phi) = Ae^{-Be^{-Ct}};$	$b(\phi,t,y) = BCe^{-Ct}y;$	$y_0(\phi) = A \cdot e^{-B},$
Logistic:	$f(t,\phi) = \frac{A}{1+Be^{-Ct}};$	$b(\phi,t,y) = Cy(1-\frac{1}{A}y);$	$y_0(\phi) = \frac{A}{1+B},$
Richards:	$f(t,\phi) = \frac{A}{(1+Be^{-Ct})^D};$	$b(\phi,t,y) = \frac{BCDe^{-Ct}}{1+Be^{-Ct}}y;$	$y_0(\phi) = \frac{A}{(1+B)^D},$
Weibull:	$f(t,\phi) = A - Be^{-Ct^D};$	$b(\phi,t,y) = CDt^{D-1}(A-y);$	$y_0(\phi) = A - B.$

To get a stochastic model, two different types of errors will be introduced. The first and simplest possibility to obtain a stochastic model is a regression model with additive error. Here, the solution of an ODE is taken as the regression function in the nonlinear regression model given by

$$y_n = f(t_n,\phi) + \epsilon_n,\ \epsilon_n \sim \mathcal{N}(0, s_1^2(\gamma^2, t_n)),\ n = 0, ..., N.$$

We here assume a time dependent error variance for a better comparison to the SDE model. In the simplest case, s_1^2 will be identical to a constant γ^2, but other time-dependent cases are possible. De la Cruz-Mesía and Marshall (2006), e.g., assume an autoregressive error structure in a hierarchical nonlinear regression model. Here, we will restrict to the time-dependent case and will assume a constant variance in the application.

Another way to define a stochastic model is extending the ODE to the corresponding SDE defined by

$$dY_t(\phi) = b(\phi, t, Y_t)\, dt + s_2(\gamma^2, t, Y_t)\, dW_t,\ Y_0 = y_0,$$

where $\{W_t,\ t \in [0,\infty)\}$ denotes a Brownian motion. The observations are assumed to be a discretization of one path of the process, i.e. $y_n = Y_{t_n}(\phi),\ n = 0, ..., N$. The most appealing property of the SDE model for our purpose of prediction is the fact that prediction intervals will increase over time. Due to the Markov property the prediction is based on the last observation point and, therefore, has a low variance close to the last observation which becomes bigger with growing distance.

As mentioned, for some choices of b and s_2 we can calculate a solution for the SDE defined above. For example, we get a solution for the Gompertz function with $s_2(\gamma^2, t, y) =$

$\gamma y,\ \gamma = \sqrt{\gamma^2} > 0$ that is given by

$$Y_t(\phi) = \exp\left(\log(A) - Be^{-Ct} - \frac{1}{2}\gamma^2 t + \gamma W_t\right). \tag{2}$$

In this special case the process has the nice feature that its logarithm is normally distributed. That means, on the one hand we get a likelihood for estimation that is the multivariate normal distribution (for the logarithms) and on the other hand we have a density for the predictive distribution. This is not self-evident, see, for example, the Weibull function. To get a solution we decide for $s_2(\gamma^2, t, y) = \gamma(A - y)$. In this case the process $Y_t(\phi)$ is given by

$$Y_t(\phi) = A - \frac{B}{\exp(Ct^D + \frac{1}{2}\gamma^2 t + \gamma W_t)}.$$

The calculations for the two cases are based on Itô's formula, see, for example, Protter (2005). Unfortunately, we do not obtain a normally distributed process by such a simple transformation as given in (2). Furthermore, these calculations cannot be done for all choices of b or s_2. To keep the following calculations as general and programming codes as flexible as possible we will use the SDE for the estimation and prediction scheme presented in the following. Therefore, we need to use some approximation for the discretely observed process. We will take the Euler-Maruyama scheme or in short Euler scheme, see, for example, Fuchs (2013), as a frequently used scheme. We are aware that this will introduce an approximation error but the application to the data of Virkler et al. (1979) shows that the resulting prediction intervals are quite good.

The underlying data set for this article contains measurements of several individuals as it is often the case for growth curve data. That means, we have observations $y_{i0}, ..., y_{iN_i}$ at points $t_{i0}, ..., t_{iN_i}$, $i = 1, ..., I$. If we assume each of the individuals to follow the same curve with varying parameters up to an error term, we in turn assume that the data result from one of the two hierarchical models:

model (1): *nonlinear regression*	model (2): *stochastic process*
$y_{in} = f(t_{in}, \phi_i) + \epsilon_{in}$	$y_{in} = Y_{t_{in}}(\phi_i)$
$\phi_i \sim \mathcal{N}(\mu, \Omega)$ iid.	$\phi_i \sim \mathcal{N}(\mu, \Omega)$ iid.
$\epsilon_{in} \sim \mathcal{N}(0, s_1^2(\gamma^2, t_{in}))$,	$dY_t(\phi) = b(\phi, t, Y_t)\, dt + s_2(\gamma^2, t, Y_t)\, dW_t$,

$n = 0, ..., N_i; i = 1, ..., I$. Models (1) and (2) are mixed models, model (1) will also be referred to as mixed regression model, model (2) as mixed diffusion model. Let us first note that, strictly, we would have to include another index in the notation of the Brownian motion because it is one Brownian motion for each individual and not the same for all. Secondly, ϕ_i and the Brownian motion of the ith individual as well as the I Brownian motions among themselves have to be stochastically independent. Models (1) and (2) will each be applied for six growth curves which leads to twelve models altogether.

The variables μ, Ω and γ^2 are parameters we need to estimate. As already mentioned, in the following we will use the Euler approximation given by

$$
\begin{aligned}
&Y_{i0} = y_0(\phi_i) \\
&Y_{in} = Y_{i(n-1)} + b(\phi_i, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in} + s_2(\gamma^2, t_{i(n-1)}, Y_{i(n-1)})\sqrt{\Delta_{in}}\xi_{in} \\
&\Delta_{in} := t_{in} - t_{i(n-1)} \\
&\xi_{in} \sim \mathcal{N}(0,1) \text{ iid.}, \ n = 1, ..., N_i; \ i = 1, ..., I.
\end{aligned}
$$

In the remainder of this work we will use the following notations. Our data are taken to be realisations from Y_{in}, if we assume the diffusion model, and from y_{in}, if we assume the regression model, being aware that both are random variables. The main goal of this article is the prediction of some y_I^*, respectively $Y_{t_I^*}$, in $t_I^* > t_{IN_I}$, but to obtain some distribution for that we first have to estimate the unknown parameters.

3 Bayesian Estimation

First we present the posterior distributions that are equal in both models, namely for μ and Ω. We assume a conjugate normal prior distribution of μ with mean m and variance matrix V. Then it can be seen by direct calculation (cf. Carlin and Louis (2009), p. 168)

that

$$\begin{aligned}\mu|\ \phi_1, ..., \phi_I, \Omega &\sim \mathcal{N}(m^{\text{post}}, V^{\text{post}}) \\ V^{\text{post}} &= (V^{-1} + I \cdot \Omega^{-1})^{-1} \\ m^{\text{post}} &= V^{\text{post}} \cdot \left(V^{-1} m + \sum_{i=1}^{I} \Omega^{-1} \phi_i \right).\end{aligned}$$

In the following, we assume Ω to be a diagonal matrix $\text{diag}(\omega_A^2, \omega_B^2)$, $\text{diag}(\omega_A^2, \omega_B^2, \omega_C^2)$ or $\text{diag}(\omega_A^2, \omega_B^2, \omega_C^2, \omega_D^2)$ for the respective model. Choosing the inverse Gamma distribution for the diagonal elements $\omega_A^2 \sim IG(\alpha_A, \beta_A)$, i.e. $\omega_A^{-2} \sim \text{Gamma}(\alpha_A, \beta_A)$, we get the conditional posterior distribution

$$\omega_A^2|\ A_1, ..., A_I, \mu_A \sim IG\left(\alpha_A + \frac{I}{2}, \beta_A + \frac{1}{2} \sum_{i=1}^{I} (A_i - \mu_A)^2 \right)$$

with $\mu = (\mu_A, \mu_B)$, $\mu = (\mu_A, \mu_B, \mu_C)$ or $\mu = (\mu_A, \mu_B, \mu_C, \mu_D)$ for the respective model. Analogously we proceed with ω_B^2, ω_C^2 and possibly ω_D^2 as can also be seen in Carlin and Louis (2009), p. 35.

The posteriors of μ and Ω depend on the unobserved $\phi_1, ..., \phi_I$. Therefore, we also need to calculate the full conditional posterior of ϕ_i for each $i = 1, ..., I$, first in model (1), the mixed *nonlinear regression* framework.

Although we are investigating nonlinear models we indeed could separate the components and estimate a few of them with a conjugate prior explicitly. However, some components have to be estimated with some sampling method like the Metropolis Hastings (MH) algorithm (see Carlin and Louis (2009), p. 130). Therefore, we refrain from calculating the posteriors explicitly for a part of the vector ϕ_i and use the MH algorithm for the whole vector ϕ_i. Given by the model definition, the likelihood in the mixed regression model (1) is

$$p(y_{i0}, ..., y_{iN_i}|\ \phi_i, \gamma^2) = \prod_{n=0}^{N_i} \frac{1}{\sqrt{2\pi} s_1(\gamma^2, t_{in})} \exp\left(-\frac{1}{2 s_1^2(\gamma^2, t_{in})} (y_{in} - f(t_{in}, \phi_i))^2 \right),$$

$i = 1, ..., I$. Furthermore, a normal prior is assumed by the model definition. The resulting posterior also depends on γ^2, for whose estimation we assume the structure $s_1^2(\gamma^2, t, y) = \gamma^2 \cdot \bar{s}^2(t)$ allowing for a conjugate prior. Of course a sampling method like the MH algorithm

could also be employed for other nonlinear functions. $\bar{s}^2$ can be for example $\bar{s}^2(t) = 1$ or $\bar{s}^2(t) = t$. $\gamma^2 \sim IG(\alpha_\gamma, \beta_\gamma)$ is a sensible choice for the variance parameter and we obtain

$$\gamma^2 |\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I}, \phi_1, ..., \phi_I \\ \sim IG\left(\alpha_\gamma + \frac{1}{2}\sum_{i=1}^{I}(N_i+1), \beta_\gamma + \frac{1}{2}\sum_{i=1}^{I}\sum_{n=0}^{N_i}\frac{(y_{in} - f(t_{in}, \phi_i))^2}{\bar{s}^2(t_{in})}\right).$$

With the full conditional posteriors, we can use the Gibbs sampler to simulate the posterior distribution $p(\phi_1, ..., \phi_I, \mu, \Omega, \gamma^2 |\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I})$ by drawing iteratively from the conditional marginals. After choosing starting values $\phi_{i0}^*,\ i = 1, ..., I$, for the MH step and $\mu_0^*,\ \Omega_0^*$ and γ_0^{2*} for the Gibbs sampler, we draw for $k = 1, ..., K$ from

$$\begin{aligned}
\phi_{ik}^* &\sim p(\phi_i |\ y_{i0}, ..., y_{iN_i}, \gamma_{k-1}^{2*}, \mu_{k-1}^*, \Omega_{k-1}^*),\ i = 1, ..., I \\
\mu_k^* &\sim p(\mu |\ \phi_{1k}^*, ..., \phi_{Ik}^*, \Omega_{k-1}^*) \\
\Omega_k^* &\sim p(\Omega |\ \phi_{1k}^*, ..., \phi_{Ik}^*, \mu_k^*) \\
\gamma_k^{2*} &\sim p(\gamma^2 |\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I}, \phi_{1k}^*, ..., \phi_{Ik}^*).
\end{aligned}$$

This procedure is called Metropolis-within-Gibbs or univariate Metropolis algorithm because drawing from the full conditional posterior of ϕ_i is executed through a Metropolis step, see Carlin and Louis (2009), p. 141. With this iterative procedure we obtain K drawings $(\phi_{1k}^*, ..., \phi_{Ik}^*, \mu_k^*, \Omega_k^*, \gamma_k^{2*}),\ k = 1, ..., K$, from the joint posterior distribution $p(\phi_1, ..., \phi_I, \mu, \Omega, \gamma^2 |\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I})$.

For the posterior of ϕ_i in the mixed *diffusion* model we get an approximated likelihood through the Euler approximation. We first obtain the transition distribution

$$Y_{in} |\ Y_{i(n-1)}, \phi_i, \gamma^2 \sim \mathcal{N}\left(Y_{i(n-1)} + b(\phi_i, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in}, s_2^2(\gamma^2, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in}\right),$$

$n = 1, ..., N_i$, that builds up the joint distribution

$$p(Y_{i1}, ..., Y_{iN_i} |\ \phi_i, \gamma^2) = \prod_{n=1}^{N_i} p(Y_{in} |\ Y_{i(n-1)}, \phi_i, \gamma^2),$$

which is the approximated likelihood for ϕ_i and γ^2. Of course, the true transition density of the process might not be a normal one. A good overview of possibly resulting problems can be seen in Sørensen (2004) or Beskos et al. (2006). The estimation methods presented

in the following can be transferred to other approximation schemes, because the likelihood function can easily be exchanged in the MH step. Pedersen (1995) proved that the likelihood of the Euler approximated process asymptotically tends to the likelihood of the true process for growing sample size. Cano et al. (2006) worked this idea out for a posterior density built up on the likelihood of the approximated process.

Looking at the functions b, all but one of the components of ϕ_i occur which makes it impossible to estimate the non-occuring component with the differences. One could calculate some estimation with the first observation point and the estimation of the other components but for the prediction this is not necessary and will, therefore, not be done in the application. Similar to the regression model, we decide for the structure $s_2(\gamma^2, t, y) = \gamma \cdot \widetilde{s}(t, y)$ without any further unknown variable for the estimation of γ^2. In the application section we will use $\widetilde{s}(t, y) = 1$ but other choices are possible. E.g. $\widetilde{s}(t, y) = t$ would yield a variance structure dependent on time or $\widetilde{s}(t, y) = y$ would lead to an autoregressive variance dependent on the actual process. With $\gamma^2 \sim IG(\alpha_\gamma, \beta_\gamma)$ the conditional posterior is given by

$$\gamma^2 |\ \{Y_{in}\}_{n=1,...,N_i;i=1,...,I}, \phi_1, ..., \phi_I$$
$$\sim IG\left(\alpha_\gamma + \frac{1}{2}\sum_{i=1}^{I} N_i, \beta_\gamma + \frac{1}{2}\sum_{i=1}^{I}\sum_{n=1}^{N_i} \frac{(Y_{in} - Y_{i(n-1)} - b(\phi_i, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in})^2}{\widetilde{s}^2(t_{i(n-1)}, Y_{i(n-1)})\Delta_{in}}\right).$$

This can be seen by direct calculation because

$$\frac{Y_{in} - Y_{i(n-1)} - b(\phi_i, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in}}{\widetilde{s}(t_{i(n-1)}, Y_{i(n-1)})\sqrt{\Delta_{in}}} \sim \mathcal{N}(0, \gamma^2),\ n = 1, ..., N_i;\ i = 1, ..., I.$$

The resulting Gibbs sampler is analogous to the regression model.

4 Bayesian Prediction

The main aim of this article is the prediction for y_I^* in t_I^* with $t_I^* > t_{IN_I}$. Of course we could predict parts of any of the I series, but to simplify notations let us say we want to predict the further development of the last series. In the real data application we would have observed $I - 1$ series already and want to use their information for the prediction of

the new one. Therefore, in the mixed regression model the predictive distribution is

$$p(y_I^*|\ \{y_{in}\}_{n=0,...,N_I;i=1,...,I}) = \int p(y_I^*|\ \phi_I, \gamma^2) \cdot p(\phi_I, \gamma^2|\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I})\, d(\phi_I, \gamma^2)$$
$$\approx \frac{1}{K} \sum_{k=1}^{K} p(y_I^*|\ \phi_{Ik}^*, \gamma_{1k}^{2*}),$$

with $(\phi_{Ik}^*, \gamma_{1k}^{2*}) \sim p(\phi_I, \gamma^2|\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I})$ coming from the Gibbs sampler. We will approximate this distribution in the application section by simulations. We use rejection sampling, see Carlin and Louis (2009), p. 116, with the indicator function on some candidate area C^* multiplied with $\max_{y_I^* \in C^*} (p(y_I^*|\ \{y_{in}\}_{n=0,...,N_i;i=1,...,I}))$ as the envelope function. This candidate area can be some neighborhood around the last observation y_{IN_I} as chosen in the application section. The point prediction, taken as the expected value of the predictive distribution, can then be approximated by the mean of the drawn sample. In the same way $(1-\alpha)$-prediction intervals can be obtained with the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles as interval boundaries.

Analogously we can predict $Y_{t_I^*}$ in t_I^* with $t_I^* > t_{IN_I}$ in the mixed diffusion model. Here, the whole issue becomes a bit more complicated, since the Euler approximation gets imprecise with big time differences. Therefore, we would get a bad prediction when t_I^* is much bigger than t_{IN_I} because of the approximation scheme even if the model is correct and the estimation is good. To avoid this problem we simulate the following process with some sampling partition $t_{IN_I} = \tau_0 < \tau_1 < ... < \tau_L = t_I^*$. By enlarging L, one gets a more precise approximation. For simplification we take

$$\tau_l = l \cdot \Delta^* + t_{IN_I},\ l = 0, ..., L,$$

where $\Delta^* := \frac{t_I^* - t_{IN_I}}{L}$ is the distance between the equidistant time points. Then we iteratively build the predicted Euler approximation of $Y_{t_I^*}(\phi_I)$ by

$$Y_{t_I^*}(\phi_I) \approx Y_{IL}^* = Y_{IN_I} + \sum_{l=0}^{L-1} \{b(\phi_I, \tau_l, Y_{Il}^*)\Delta^* + \gamma \widetilde{s}(\tau_l, Y_{Il}^*)\xi_l^*\}$$
$$\xi_l^* \sim \mathcal{N}(0, \Delta^*) \text{ iid},\ l = 0, ..., L-1$$

with $Y_{I0}^* := Y_{IN_I}$. Hence we obtain a distribution for Y_{IL}^* dependent on $\phi_I, \gamma^2, Y_{IN_I}$ and the

intermediate points

$$Y^*_{IL}|\ \phi_I, \gamma^2, Y_{IN_I}, Y^*_{I1}, ..., Y^*_{IL-1} \\ \sim \mathcal{N}\left(Y_{IN_I} + \Delta^* \sum_{l=0}^{L-1} b(\phi_I, \tau_l, Y^*_{Il}),\ \gamma^2 \Delta^* \sum_{l=0}^{L-1} \widetilde{s}^2(\tau_l, Y^*_{Il})\right).$$

Practically, we set starting values $Y^{*(k)}_{I0} = Y_{IN_I},\ k = 1, ..., K$ and iteratively draw

$$Y^{*(k)}_{Il} \sim \frac{1}{K} \sum_{r=1}^{K} p\left(Y^*_{Il}|\ \phi^*_{Ir}, \gamma^{2*}_r, Y^{*(r)}_{I(l-1)}\right),\ k = 1, ..., K$$

for $l = 1, ..., L$ with ϕ^*_{Ir} and $\gamma^{2*}_r,\ r = 1, ..., K$ resulting from the Gibbs sampler as derived in Section 3.

5 Simulations and Application to Crack Growth

All calculations in this section are made with the programming language R, R Core Team (2013).

In the following we compare the predictions made with the two models (1) and (2) each for the six presented growth curves. We will first present simulations to validate the estimation and prediction procedure and later on apply the models to the data of Virkler et al. (1979).

5.1 Simulation Study

For the simulation study we assume equally long series except the one to predict, the Ith series, that will be half as long. Then this series will be predicted in the missing time values. In particular, we will simulate 1000 times a set of $I = 20$ series, each with 101 observation points. For the estimation we will cut the Ith series after the 51st observation in each simulation set, so that $N_1 = ... = N_{I-1} = \widetilde{N} = 101$ and $N_I = 51$. For the Ith series, each point of the second half, i.e. $t^*_I \in \{t_{52}, ..., t_{101}\}$, will be predicted, so that $L = 50$. Considering the time-homogeneity of the stochastic process in model (2), we construct a finer grid, i.e., for each interval $[t_{i-1}, t_i]$, we insert nine equidistant points, $i = 1, ..., \widetilde{N}$. Then the process is simulated on this finer partition. For the observation points, we take

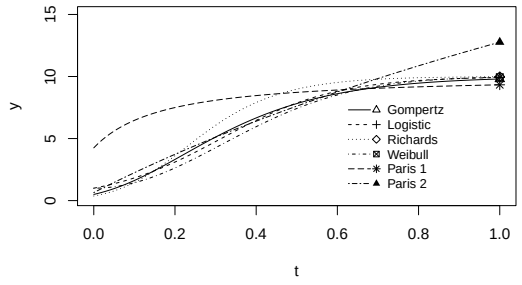
every tenth value of the simulated process. All series in model (2) are simulated with the starting value $y_0 = 0.5$.

For the Gompertz, Logistic, Richards and Weibull curves we take time points $t_0 = 0, t_1 = 0.01, ..., t_{101} = 1$, for the first Paris model $t_0 = 0.2, t_1 = 0.21, ..., t_{101} = 1.2$ and for the second Paris model $t_0 = 1.2, t_1 = 1.21, ..., t_{101} = 2.2$. As we have seen in the second section, the two Paris models may not start in 0 resp. 1. We conduct the simulations with the parameter μ equal to

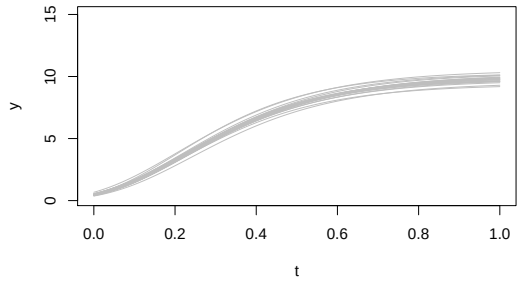
Gompertz	Logistic	Richards	Weibull	Paris 1	Paris 2
$(10, 3, 5)$	$(10, 9, 7)$	$(10, 2, 8, 3)$	$(10, 9, 5, 2)$	$(10, 1, 1.2)$	$(20, 3)$.

The resulting simulated data are represented in parts in Figure 1 where Figure 1(a) shows the six growth curves for the chosen parameters. For a better comparison the Paris derived curves are also plotted in $t \in [0, 1]$, i.e., for the Paris 1 model in $t - 0.2$ and for the Paris 2 model in $t - 1.2$.

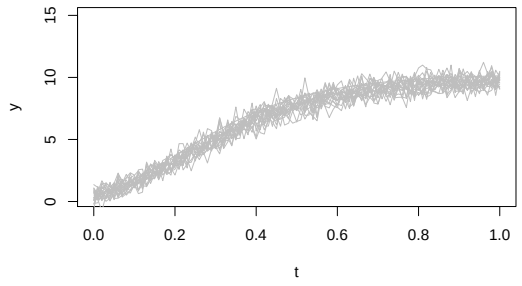
For the hyperparameter Ω we choose $\frac{1}{100} \cdot \text{diag}(\mu)$ for each model. With $\phi_i \sim \mathcal{N}(\mu, \Omega)$, $i = 1, ..., I$, we get I curves, for example see Figure 1(b) for the Gompertz function. For the variance, we choose $s_1^2(\gamma^2, t) \equiv \gamma^2 = \frac{1}{4}$ and with the resulting errors we obtain, for example, for the Gompertz function the points in Figure 1(c) for model (1). With $\widetilde{s} \equiv 1$ and $\gamma^2 = \frac{1}{4}$ a simulation set resulting from model (2) with the Gompertz function is displayed in Figure 1(d). In the comparison of Figures 1(c) and 1(d) one can see the effect of the different variance structures of model (1) and model (2) with otherwise equal parameters.



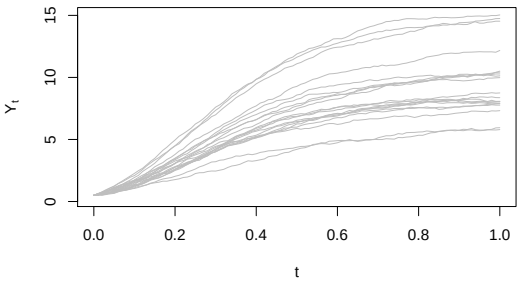
(a) Curve shapes for the six growth functions



(b) Curve shapes for the Gompertz function for $\phi_1, ..., \phi_I$



(c) Simulation points for model (1) with the Gompertz function



(d) Simulated process Y_t for model (2) with the Gompertz function

Figure 1: Simulation Example, (a) growth curves with chosen parameters, (b) Gompertz curves with drawn $\phi_i \sim \mathcal{N}(\mu, \Omega)$, $i = 1, ..., I$, iid., (c) curves from (b) with additive regression error, (d) Euler approximated diffusion processes following model (2)

For the estimation of the parameters, we assume $\gamma^2 \sim IG(4, 1)$, $\omega_a^2 \sim IG(\frac{100}{m}, 1)$, $a \in \{A, ..., D\}$ and $\mu \sim \mathcal{N}(m, V = \text{diag}(m))$, where in each case m is chosen as the true, in a simulation study known, value. Note that the aim is not a sensitivity analysis of the models which would be very elaborate for the twelve models altogether. We start with the true values of μ for μ_0^* and for $\phi_{10}^*, ..., \phi_{I0}^*$ and with $\gamma_0^{2*} = 1$. We simulate for each model 101000 Markov chain iterations, drop a burn-in of 1000 iterations, take a thinning rate of 100 and obtain a sample of $K = 1000$ from the posterior. In the MH-step for sampling ϕ_i we need to choose a proposal density. Here, the normal density with the last sample as

mean and $\frac{m}{50}$ as standard deviation is taken.

In Tables 1 and 2, the evaluation of the parameter estimations is presented. Deviation names the mean squared error of the point estimation, i.e. the mean of the posterior, and the true value. The second column is the coverage rate, i.e. the rate of 95%-credibility intervals that include the true value. Score refers to the scoring rule of Gneiting and Raftery (2007) for interval forecasts with $\alpha = 0.05$, i.e.

$$S(l, u, y) = (u - l) + \frac{2}{\alpha}(l - y)\mathbb{1}\{y < l\} + \frac{2}{\alpha}(y - u)\mathbb{1}\{y > u\}, \tag{3}$$

where l is the lower interval boundary, i.e. the $\frac{\alpha}{2}$ quantile, u the upper boundary, i.e. the $1 - \frac{\alpha}{2}$ quantile of the predictive distribution, and y the true value. The deviation and score values in the table each are averaged over all series. In Table 1 all parameters for model (1) have coverage rates around 95%. In Table 2 we mostly see the same picture for model (2), only for the Weibull curve, the variance parameter is slightly underestimated.

	Gompertz			Logistic			Paris 1		
	deviation	rate	score	deviation	rate	score	deviation	rate	score
A	0.0074	0.98	0.44	0.0063	0.98	0.41	0.0087	0.99	0.49
B	0.0056	0.99	0.41	0.0384	1.00	1.20	0.0044	1.00	0.37
C	0.0092	1.00	0.56	0.0117	0.99	0.59	0.0017	0.99	0.23
γ^2	0.000063	0.95	0.038	0.000064	0.95	0.038	0.000058	0.95	0.036
	Weibull			Richards			Paris 2		
	deviation	rate	score	deviation	rate	score	deviation	rate	score
A	0.0060	0.98	0.40	0.0050	0.97	0.35	0.0048	0.99	0.40
B	0.0133	0.98	0.59	0.0949	0.94	1.26	0.0171	0.99	0.75
C	0.0239	1.00	0.91	0.0307	0.98	0.88			
D	0.0022	0.99	0.24	0.0678	0.96	1.19			
γ^2	0.000069	0.94	0.038	0.000068	0.94	0.038	0.000064	0.95	0.037

Table 1: Evaluation of the parameters in model (1)

For the predictions in model (1) we sample from the predictive distribution with rejection sampling, as explained in the third section. The candidate area C^* is chosen as $[y_{IN_I} - 2, y_{IN_I} + 10]$ with a grid of 0.001. For the prediction in model (2) we also

	Gompertz			Logistic			Paris 1		
	deviation	rate	score	deviation	rate	score	deviation	rate	score
A				0.0115	0.97	0.54	0.0330	0.99	1.01
B	0.0077	0.99	0.46						
C	0.0216	0.98	0.75	0.0226	1.00	0.88	0.0051	0.88	0.31
γ^2	0.000081	0.92	0.042	0.000084	0.90	0.046	0.000116	0.83	0.057
	Weibull			Richards			Paris 2		
	deviation	rate	score	deviation	rate	score	deviation	rate	score
A	0.0068	0.97	0.41						
B				0.0490	0.94	1.04	0.0442	0.96	0.98
C	0.0221	0.99	0.86	0.0361	0.97	0.93			
D	0.0033	0.96	0.28	0.0862	0.92	1.29			
γ^2	0.000236	0.58	0.122	0.000090	0.91	0.044	0.000061	0.95	0.037

Table 2: Evaluation of the parameters in model (2)

sample with rejection sampling in each step $l = 1, ..., 50$. For each l a candidate area $[\bar{Y}^*_{I(l-1)} - 2, \bar{Y}^*_{I(l-1)} + 2]$ with $\bar{Y}^*_{I(l-1)} = \frac{1}{K}\sum_{k=1}^{K} Y^{*(k)}_{I(l-1)}$ and a grid of 0.001 is selected.

To validate the predictions of the models we compare four key figures. First we calculate the mean squared error of the point prediction, i.e. the mean of the predictive distribution, from the true value. Second, the sizes of the 1000 $1 - \alpha = 0.95$ prediction intervals will be averaged. Third, we calculate the scoring rule $S(l, u, y)$ in (3). If the interval covers the true value, the score is just the size. The score punishes intervals that do not cover the true value with the amount of the deviation from the boundary multiplied with $\frac{2}{\alpha}$. Fourth, the coverage rate will be calculated, i.e. the rate of prediction intervals that include the true value. The prediction results of the two models can be seen in Figure 2. On the left side, each of the above mentioned values for the mixed regression model (1) and on the right side the corresponding values for the mixed diffusion model (2) are displayed. We can see that the growth models from model (1) behave similarly with respect to the average deviation and average score. There are differences between the growth models of the diffusion model (2). However, they are small for predictions shortly after the last observed value. Looking at the coverage rate (bottom row of Figure 2) we can conclude that the used estimation

methods work well for all models.

5.2 Real Data Application

The main part of the present work is the application of our models to the data set of Virkler et al. (1979). In Figure 3 we can see the number of cycle counts devided by 10000 plotted against the crack lengths. There are $I = 68$ observation series each with $\widetilde{N} = 164$ observation points. The independent variable, the crack length, has a range of [9,49.8]. As mentioned in the second section, the models have partially different properties in terms of the range of t. Therefore, in the case of the usual growth curves, we transform the crack length to start in 0 and devide by 40 to obtain small differences. In the three-parameter Paris model, we only devide the crack length by 40 and further add 1 in the two-parameter Paris model.

For the estimation we choose $\gamma^2 \sim IG(1,1)$, $\omega_a^2 \sim IG(100,1)$, $a \in \{A,...,D\}$ and $\mu \sim \mathcal{N}(m, V = \text{diag}(m))$ with m equal to

Gompertz	Logistic	Richards	Weibull	Paris 1	Paris 2
$(30,5,7)$	$(30,9,7)$	$(25,1,5,23)$	$(28,20,8,5)$	$(25,2,1.8)$	$(50,10)$.

Employing the estimation procedure with the same adjustments as in the simulations study, it turned out that the proposal variance is too high and therefore the acceptance rate for ϕ_i gets very small. Therefore, we choose $\frac{m}{100}$ as proposal standard deviation and additionally iterate the MH-step inside the Gibbs sampler as long as one candidate is accepted or maximal 10 times, in the Gompertz and Weibull curves in the regression model maximal 20 times. With the starting values $\mu_0^* = \phi_{10}^* = ... = \phi_{I0}^* = m$ and $\gamma_0^{2*} = 1$ we simulate a Markov chain of 110000 iterations with a burn-in of 10000 and a thinning rate of 100 to obtain 1000 samples from the posterior. The exceptions are the Richards curve in both models and the Paris 1 regression model, where we sample 130000 iterations and drop a burn-in of 30000 due to a slower chain convergence.

We take each of the series as the one to be predicted and cut it after the first half of the observations, that means $N_I = 82$. Hence, we obtain $I = 68$ mixed model estimations and predictions.

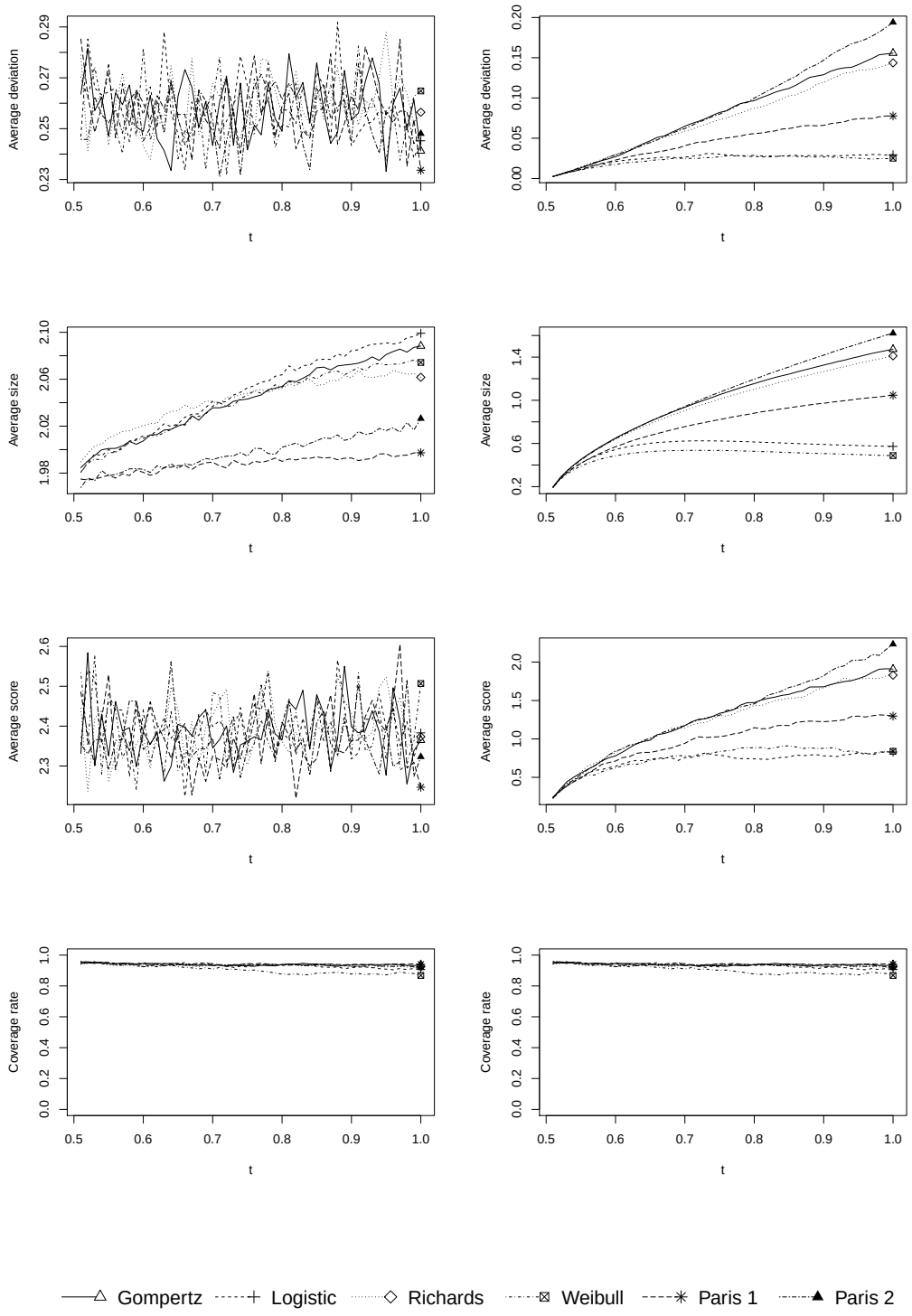


Figure 2: Prediction results in the simulation study for model (1) (left) and model (2) (right) for the four values average deviation (top row), average size (second row), average score (third row) and coverage rate (bottom row)

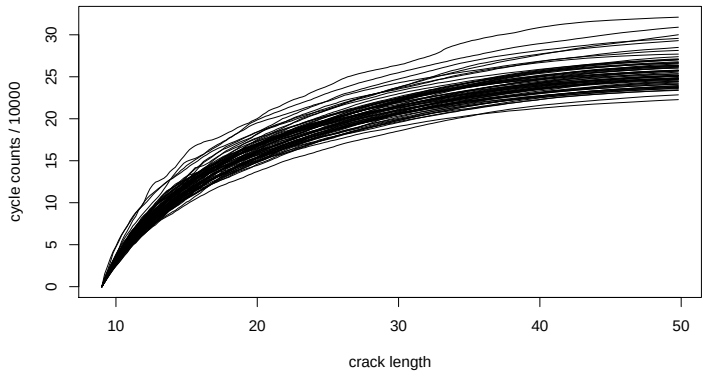


Figure 3: Observations of the 68 experiments of Virkler et al. (1979)

In Figure 4 the point prediction results of model (1) for the last series generated by taking the mean of the simulations from the predictive distribution are displayed for all six growth functions. Presented on the x-axis is the crack length, but we have calculated the estimations and predictions with the above mentioned transformation for the crack length range.

Eye-catching is the non-fitting Paris 2 model. For the estimation of μ_C in the Paris 1 model, we obtain $\widehat{\mu}_C = 0.76$ with a credibility interval [0.74,0.78]. With $\theta_2 = C + 1 = 1.76$ following from the calculations in the beginning, it is clear that a model with $\theta_2 = 1$ is not a good choice for these data.

On the left side in Figure 5 we see each the average over all 68 predictions for the several crack lengths for model (1), the mean squared error of the point prediction and the true value $\frac{1}{68}\sum_{i=1}^{68}(y_{in} - \widehat{y}^*_{in})^2$ for $n = 83, ..., 168$, denoted with average deviation, plotted against the predicted crack lengths in the top row, the average of the prediction interval sizes in the second row and the scoring rule for interval forecasts as explained above in the third row, finally, the amount of prediction intervals that cover the true value, denoted with coverage rate in the bottom row. The Weibull and the Paris 1 model perform about equally well in the point prediction and their interval sizes. However, the rate of covering intervals decreases rapidly with increasing crack length. The Gompertz and the Richards model may be considered equal in their average performance. Due to high estimated variance values the logistic and the Paris 2 model lead to the biggest intervals and the biggest deviations

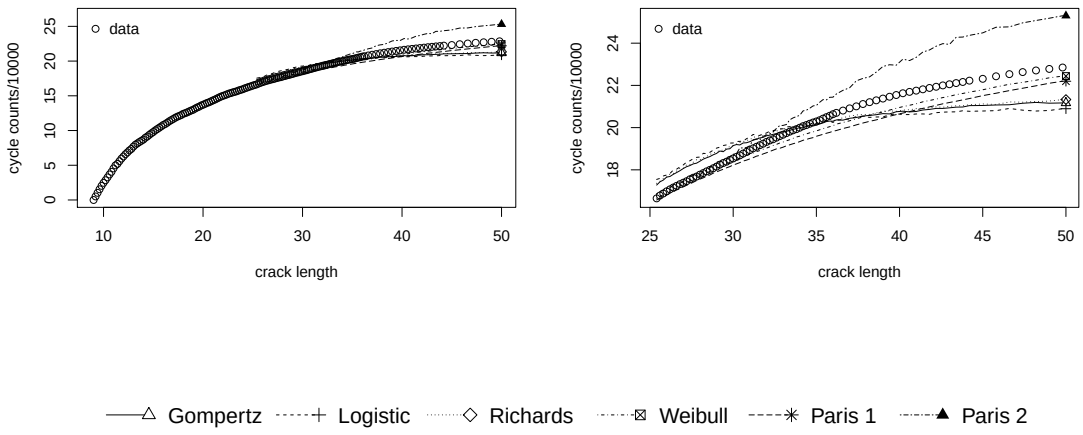


Figure 4: Point prediction for the last series of Virkler et al. (1979) with the mixed regression model (1), left: all crack lengths, right: the predicted crack lengths

of the point prediction to the true values but the highest coverage rate and are the worst fitting growth curves of all. For informative, i.e. precise, predictions from a model that describes the data we could take the scoring rule (3) into account. Up to a crack length of 30 mm, the Weibull and the three-parameter Paris model are the best predicting ones, for larger crack lengths the other models perform better because of the few noncovering intervals that are highly penalized with the term $\frac{2}{\alpha} = 40$.

Before we look at the right hand side of Figure 5, we consider the prediction for the last series. Analogously to the plots in Figure 4 for the regression model (1), we look at plots for the diffusion model (2) in Figure 6. Again, the Paris 2 model behaves worst. However, the Richards and the Gompertz model also perform poorly. For an explanation we take a look at Figure 7. There, the data points are realisations of $Y_{I(j+1)} - Y_{Ij}$ and the lines each represent $b(\widehat{\phi}_I, t_j, Y_{Ij}) \cdot (t_{j+1} - t_j),\ j = 1, ..., N_I - 1$ with $\widehat{\phi}_I$ the mean of the simulations from the posterior. The Richards and the Gompertz functions are estimated to 0 for the crack lengths that we want to predict leading to constant predictions. In contrast, the three-parameter Paris model performs quite well and the Weibull model can compete with it. On the right hand side of Figure 5 we notice that the mean squared deviation of the point predictions to the true values are very small for both growth curve

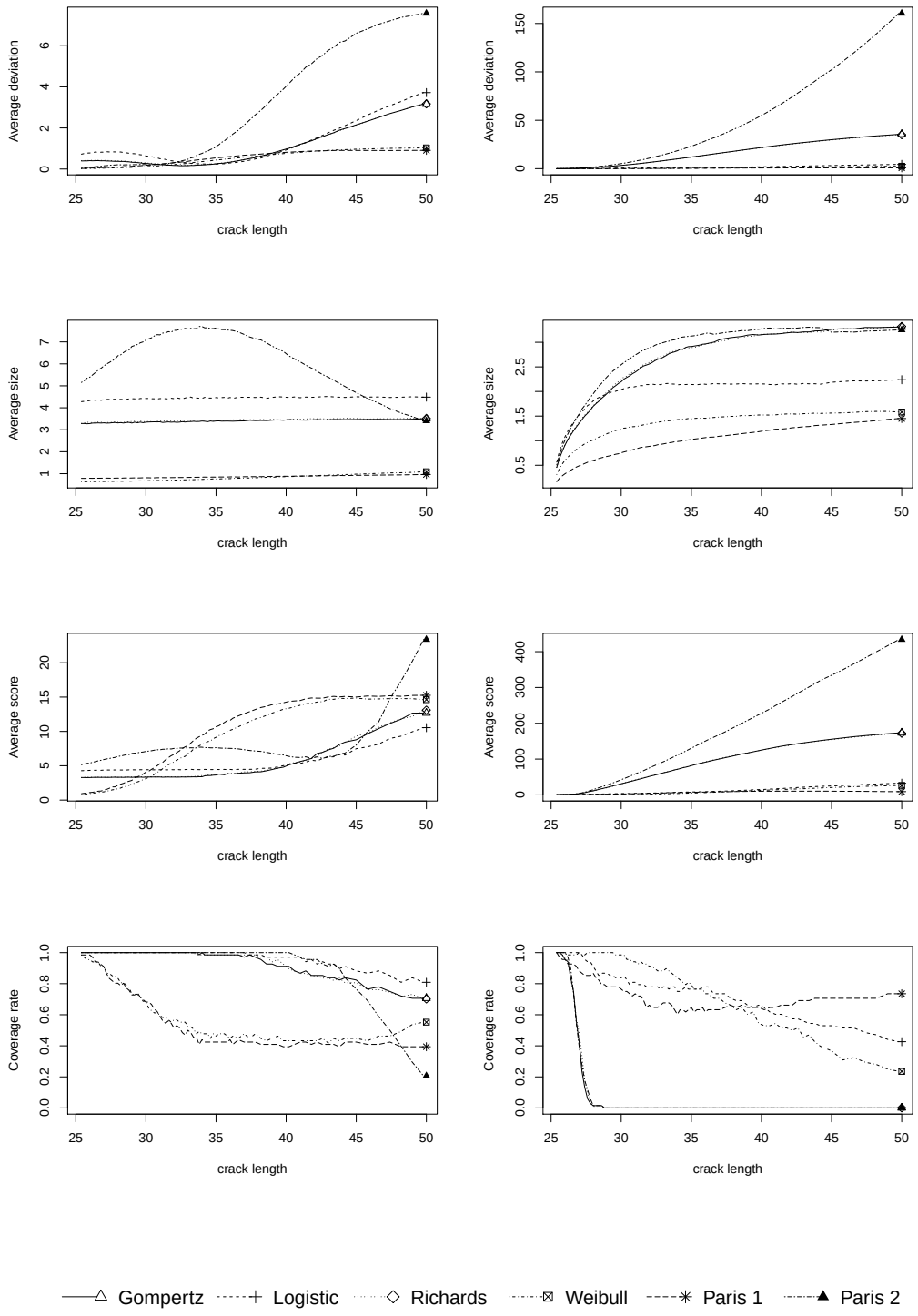


Figure 5: Analysis of data set of Virkler et al. (1979) for the mixed regression model (1) (left) and for the mixed diffusion model (2) (right) for the four values average deviation (top row), average size (second row), average score (third row) and coverage rate (bottom row)

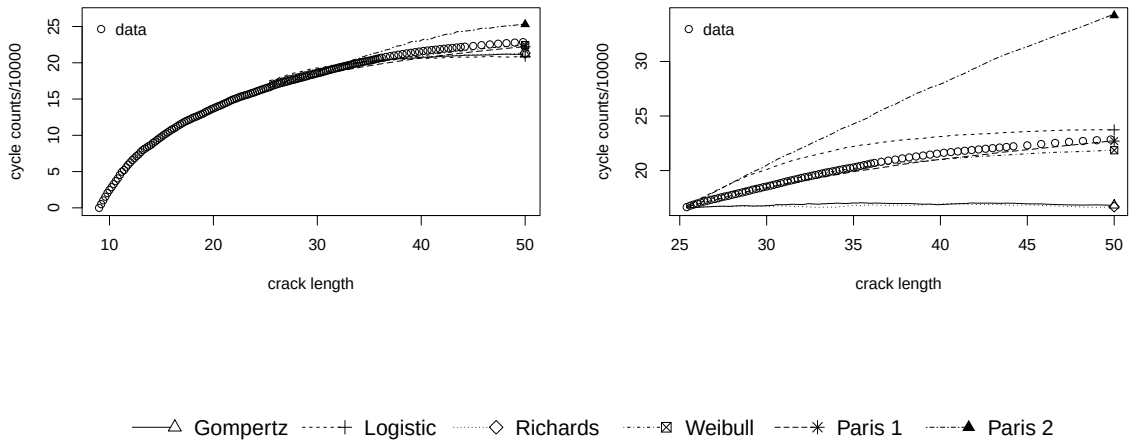


Figure 6: Point prediction for the last series of Virkler et al. (1979) with the mixed diffusion model (2), left: all crack lengths, right: the predicted crack lengths

functions. The Richards, Gompertz and the two-parameter Paris functions perform badly. The logistic model performs suprisingly well. Although it did not really fit the data in the mixed regression model, it has a low deviation of the point prediction and the coverage rate can compete with the Paris 1 model with medium prediction interval sizes for the diffusion model.

For a further comparison of the two models we take a look at Figure 8 that shows the consolidated plots from Figure 5 for the three best fitting growth curves. The mixed regression models are shown in solid lines, whereas the mixed diffusion models are plotted in dashed lines. For the logistic model the diffusion model is the only suitable if the classification is not only based on the coverage rate. For the Weibull and the three-parameter Paris curves the choice for the best model is not that easy. In the regression model, both perform equally well. For the diffusion model one can recognize differences in the distance of predictions from the last observed time point. Whereas the Weibull is the better choice for crack lengths up to 37 or 38 mm, for bigger crack lengths the Paris 1 model performs better, see, e.g., the average interval score picture where the dotted lines intersect. In summary one can say, all three functions in both models (1) and (2) can be used to describe and predict the data set of Virkler et al. (1979), where the Paris 1 and the Weibull model

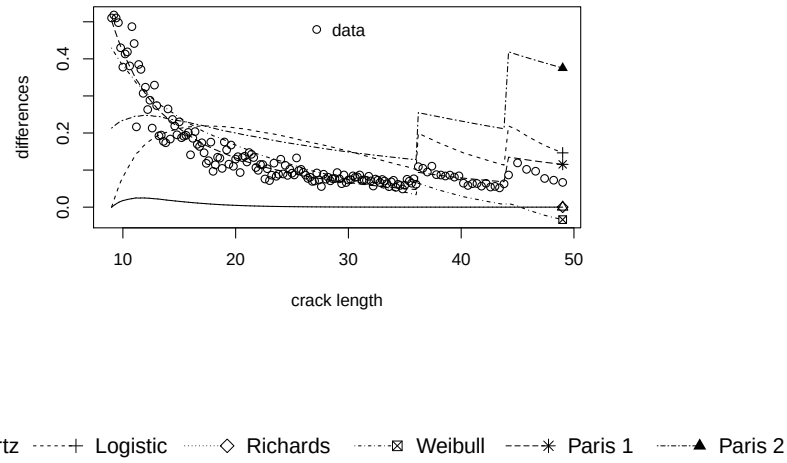


Figure 7: Comparison of the estimated differences $b(\widehat{\phi}_I, t_j, Y_{Ij}) \cdot (t_{j+1} - t_j),\ j = 1, ..., N_I - 1$ of the last series for model (2)

are a little more preferable.

6 Conclusion

In this paper a novel Bayesian prediction approach for a mixed nonlinear regression and a mixed diffusion model has been developed. Both have been implemented for four commonly used growth curves and for two curves resulting from the Paris-Erdogan law that is widely used in the engineering literature to model crack growth. The appealing property of the diffusion model is the increasing size of the prediction intervals that start very small and depend on the last observation point. However, our results show that there is no advantage of the diffusion to the regression model. We therefore recommend to use the nonlinear regression model.

Of course, one could try other functional variance structures than the constant one used in this paper.

For the six growth curves used in this paper we can conclude that the three-parameter Paris-Erdogan and the Weibull curve are the best fitting ones for the data of Virkler et al. (1979). With the disadvantage of huge prediction intervals, the logistic curve can also be

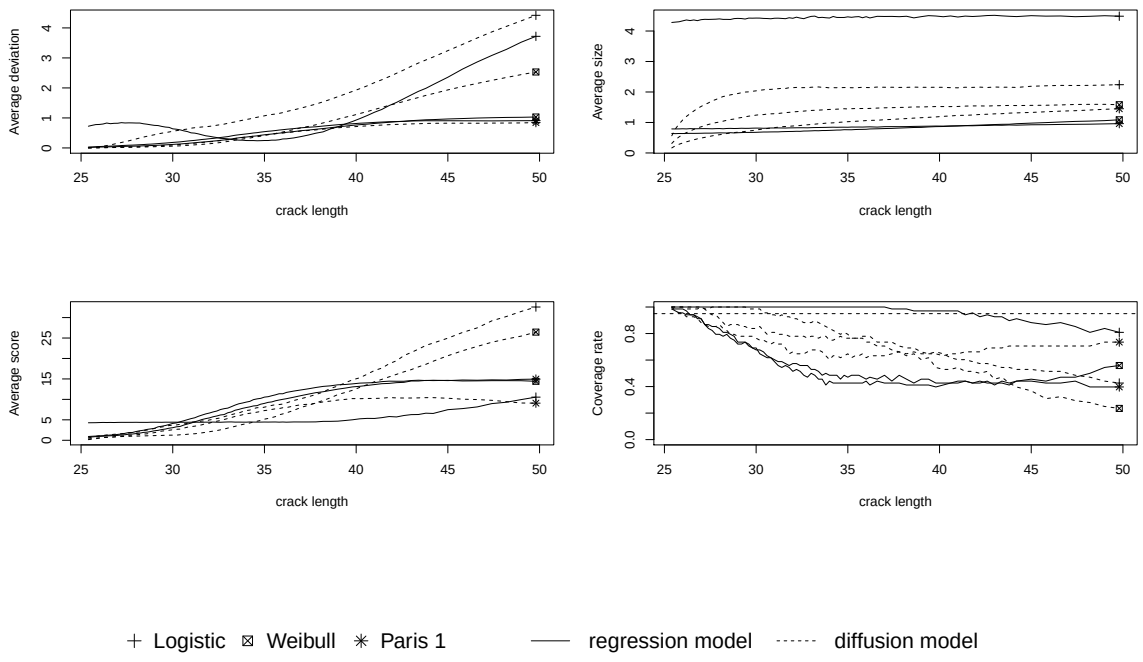


Figure 8: Comparison of the regression model (1) and the diffusion model (2) for the logistic, the Weibull and the Paris 1 growth curves, for the four values average deviation (top left), average size (top right), average score (bottom left) and coverage rate (bottom right)

used if one mainly is interested in a rough prediction with a high coverage rate.

In our approach, we used the common Euler-Maruyama approximation of the SDE. However, other approximation schemes are possible.

ACKNOWLEDGEMENTS

We want to thank Professor B. M. Hillberry for his experiments and Eric J. Tuegel for providing us the data. This work has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) in project B5 "Statistical methods for damage processes under cyclic load".

References

Beskos, A., O. Papaspiliopoulos, G. O. Roberts and P. Fearnhead (2006). Exact and Computationally Efficient Likelihood-Based Estimation for Discretely Observed Diffusion Processes. *J. R. Statist. Soc. B* **68**, 333–382.

Cano, J. A., M. Kessler and D. Salmerón (2006). Approximation of the Posterior Density for Diffusion Processes. *Statistics & Probability Letters* **76**, 39–44.

Carlin, B. P. and T. A. Louis (2009). *Bayesian Methods for Data Analysis*. 3nd ed. Boca Raton: CRC Press.

Chiquet, J., N. Limnios and M. Eid (2009). Piecewise Deterministic Markov Processes Applied to Fatigue Crack Growth. *Journal of Statistical Planning and Inference* **139**, 1657–1667.

De la Cruz-Mesía, R. and G. Marshall (2006). Non-linear Random Effects Models with Continuous Time Autoregressive Errors: A Bayesian Approach. *Statistics in Medicine* **25**, 1471–1484.

Demidenko, E. (2013). *Mixed Models: Theory and Application with R*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Ditlevsen, S. and A. de Gaetano (2005). Mixed Effects in Stochastic Differential Equation. *REVSTAT - Statistical Journal* **3**, 137–153.

Donnet, S., J.-L. Foulley and A. Samson (2010). Bayesian Analysis of Growth Curves Using Mixture Models Defined by Stochastic Differential Equations. *Biometrics* **66**, 733–741.

Fuchs, C. (2013). *Inference for Diffusion Processes*. 1st ed. Berlin Heidelberg: Springer.

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction and Estimation. *Journal of the American Statistical Association* **102**, 359–378.

Lin, Y. K. and J. N. Yang (1983). On Statistical Moments of Fatigue Crack Propagation. *Engineering Fracture Mechanics* **18**, 243–256.

Oravecz, Z., F. Tuerlinckx and J. Vandekerckhove (2009). A Hierarchical Ornstein-Uhlenbeck Model for Continuous Repeated Measurement Data. *Psychometrica* **74**, 395–418.

Oravecz, Z. and F. Tuerlinckx (2011). The Linear Mixed Model and the Hierarchical Ornstein-Uhlenbeck Model: Some Equivalences and Differences. *British Journal of Mathematical and Statistical Psychology* **64**, 134–160.

Ortiz, K. and A. S. Kiremidjian (1988). Stochastic Modeling of Fatigue Crack Growth. *Engineering Fracture Mechanics* **29**, 317–334.

Pedersen, A. R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics* **22**, 55–71.

Picchini, U., A. de Gaetano and S. Ditlevsen (2010). Stochastic Differential Mixed-Effects Models. *Scandinavian Journal of Statistics* **37**, 67–90.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations*. 2nd ed. Berlin Heidelberg: Springer.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ray, A. and S. Tangirala (1996). Stochastic Modeling of Fatigue Crack Dynamics for On-Line Failure Prognostics. *IEEE Transactions of Control Technology* **4**, 443–451.

Sobczyk, K. and B. F. Spencer (1992). *Random Fatigue: From Data to Theory*. 1st ed. London: Academic Press Limited.

Sørensen, H. (2004). Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: A Survey. *International Statistical Review* **72**, 337–354.

Virkler, D. A., B. M. Hillberry and P. K. Goel (1979). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology* **101**, 148–153.

Wu, W. F. and C. C. Ni (2004). Probabilistic Models of Fatigue Crack Propagation and their Experimental Verification. *Probabilistic Engineering Mechanics* **19**, 247–257.

Zárate, B. A., J. M. Caicedo, J. Yu and P. Ziehl (2012). Bayesian Model Updating and Prognosis of Fatigue Crack Growth. *Engineering Structures* **45**, 53–61.

Zimmermann, D. L. and V. Núñez-Antón (2001). Parametric Modelling of Growth Curve Data: An overview. *Sociedad de Estadística e Investigación Operativa* **10**, 1–73.