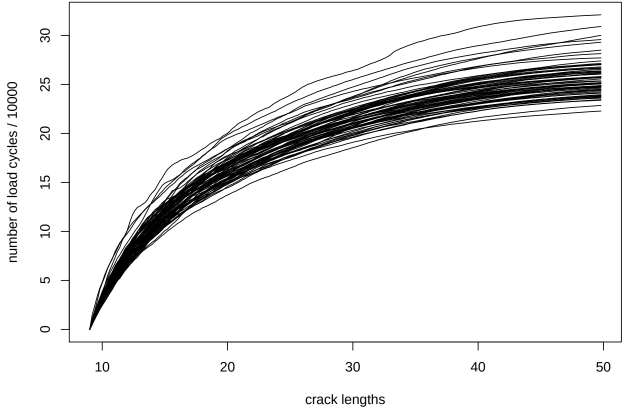
Prediction for Stochastic Growth Processes in Fatigue Experiments

Katja Ickstadt, TU Dortmund University

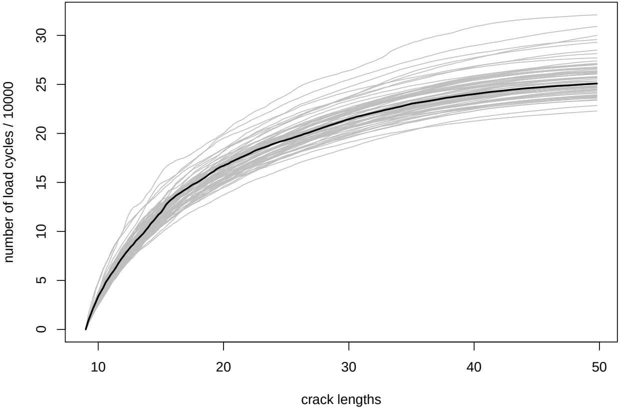
13th March 2018

with Simone Hermann and Christine H. Müller

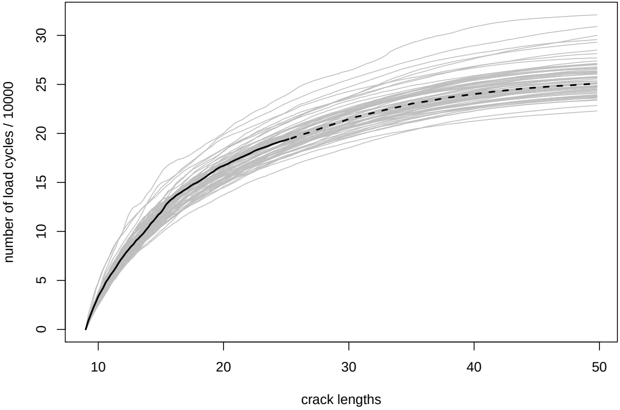
Motivation


number of load cycles / 10000
0
5
10
15
20
25
30
10
20
30
40
50
crack lengths

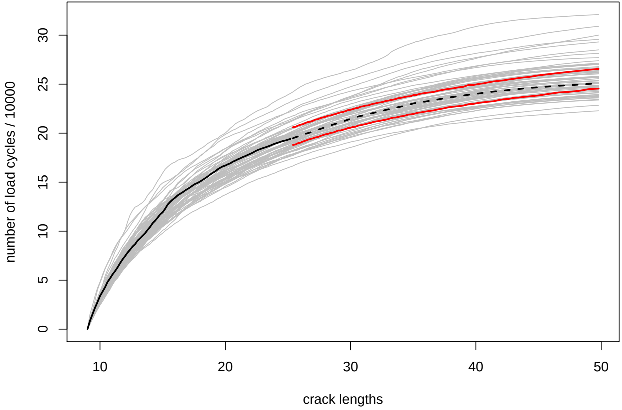
Motivation


number of load cycles / 10000
0
5
10
15
20
25
30
10
20
30
40
50
crack lengths

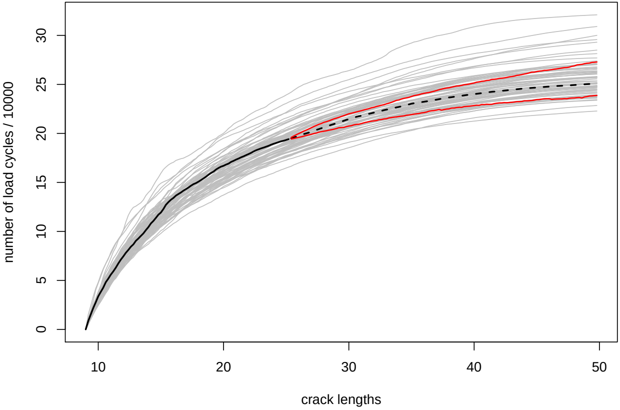
Motivation


number of load cycles / 10000
0
5
10
15
20
25
30
10
20
30
40
50
crack lengths

Motivation


number of load cycles / 10000
0
5
10
15
20
25
30
10
20
30
40
50
crack lengths

Motivation



number of load cycles / 10000
0
5
10
15
20
25
30
10
20
30
40
50
crack lengths

Outline

The Beginning...

Starting point of modeling: function f that describes the fundamental behavior of the series
But:

- a deterministic function will not describe measuring errors
- for a prediction, uncertainties are very important

$\hookrightarrow$ in the following: comparison of two approaches which include a stochastic behavior

Nonlinear Regression versus Stochastic Process

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

Nonlinear Regression versus Stochastic Process

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

↙

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

Nonlinear Regression versus Stochastic Process

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

$\swarrow$ $\searrow$

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

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

The Pros and Cons

nonlinear regression:

- mathematically simple model (likelihood is product of normal densities)
- for prediction, the information of the last observation cannot be used (no Markov property)

stochastic process:

- exact likelihood only known with explicit solution (approximation required in many cases)
- for prediction of the process, the Markov property can be used

Bayesian Estimation

Hierarchical Models

Observation variables: $y_{i0}, ..., y_{iN_i}$, $i = 1, ..., I$ for I individuals.
Assumed models (with $\{W_t,\ t \in [0, \infty)\}$ Wiener process):

nonlinear regression:

$$y_{in} = f(t_{in}, \phi_i) + \epsilon_{in}$$
$$\phi_i \sim \mathcal{N}(\mu, \Omega) \text{ iid.}$$
$$\epsilon_{in} \sim \mathcal{N}(0, s_1^2(\gamma^2, t_{in}))$$

stochastic process:

$$y_{in} = Y_{t_{in}}(\phi_i), n = 0, ..., N_i, i = 1, ..., I$$
$$\phi_i \sim \mathcal{N}(\mu, \Omega) \text{ iid.}$$
$$dY_t(\phi) = b(\phi, t, Y_t)\, dt + s_2(\gamma^2, t, Y_t)\, dW_t$$

$\Rightarrow \theta = (\phi_1, ..., \phi_I, \mu, \Omega, \gamma^2)$ large vector

$\Rightarrow$ Gibbs Sampler

Gibbs Sampler

Choose starting values and sample from the full conditional posteriors for $k = 1, ..., K$:

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

$$\Rightarrow (\phi_{1k}^*, ..., \phi_{Ik}^*, \mu_k^*, \Omega_k^*, \gamma_k^{2*}) \sim$$

$$p(\phi_1, ..., \phi_I, \mu, \Omega, \gamma^2 | \{y_{in}\}_{n=0,...,N_i; i=1,...,I}),$$

$k = 1, ..., K$ (neglecting a burn-in phase and a thinning rate)

Full Conditional Posterior $p(\mu|\ \phi_1, ..., \phi_I, \Omega)$

With the prior $\mu \sim \mathcal{N}(m, V)$ (which is conjugate to the normal likelihood) we get the posterior distribution:

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

Full Conditional Posterior $p(\Omega|\ \phi_1, ..., \phi_I, \mu)$

For $\phi_i = (\phi_{i,1}, ..., \phi_{i,p}) \in \mathbb{R}^p$ and $\Omega = \text{diag}(\omega_1^2, ..., \omega_p^2)$ choose the priors $\omega_j^2 \sim IG(\alpha_j^\Omega, \beta_j^\Omega)$. Then the posterior distribution is given by

$$\omega_j^2|\ \phi_{1,j}, ..., \phi_{I,j}, \mu_j \sim IG\left(\alpha_j^\Omega + \frac{I}{2}, \beta_j^\Omega + \frac{1}{2}\sum_{i=1}^{I}(\phi_{i,j} - \mu_j)^2\right),$$

$j = 1, ..., p.$

for the nonlinear regression model

Likelihood in the NL Regression Model

For the estimation of $\phi_1, ..., \phi_I$ with the MH algorithm, we only need the likelihood (the prior distribution is given by the model definition):

$$p(y_{i0}, ..., y_{iN_i} | \ \phi_i, \gamma^2) =$$

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

$$i = 1, ..., I$$

Metropolis-Hastings Step

Draw

$$\phi_i^{cand} \sim \mathcal{N}(\phi^*_{i(k-1)}, \text{"proposal variance"})$$

and set $\phi^*_{ik} = \phi_i^{cand}$ with probability

$$\min\left\{1, \frac{p(y_{i0}, ..., y_{iN_i} | \ \phi_i^{cand}, \gamma^{2*}_{k-1}) p(\phi_i^{cand} | \ \mu^*_{k-1}, \Omega^*_{k-1})}{p(y_{i0}, ..., y_{iN_i} | \ \phi^*_{i(k-1)}, \gamma^{2*}_{k-1}) p(\phi^*_{i(k-1)} | \ \mu^*_{k-1}, \Omega^*_{k-1})}\right\}$$

otherwise set $\phi^*_{ik} = \phi^*_{i(k-1)}$, $i = 1, ..., I$.

Full Conditional Posterior for γ^2

Assume $s_1^2(\gamma^2, t) = \gamma^2 \cdot \bar{s}^2(t)$ and $\gamma^2 \sim IG(\alpha_\gamma, \beta_\gamma)$ prior. It follows

$$\gamma^2 | \ \{y_{in}\}_{n=0,\dots,N_i; i=1,\dots,I}, \phi_1, \dots, \phi_I$$

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

For other functions $s_1^2(\gamma^2, t)$, nonlinear in γ^2, a MH-algorithm can be implemented.

for the diffusion model

Euler-Maruyama Approximation

For the estimation of $\phi_1, ..., \phi_I$ and γ^2 without an explicit solution of the SDE, an approximation method is needed. The Euler approximated variables are given by

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

Likelihood in the Diffusion Model

For the Euler variables, we have 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

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

(often called pseudo-likelihood).

Full Conditional Posterior for γ^2

For $s_2^2(\gamma^2, t, y) = \gamma^2 \cdot \widetilde{s}^2(t, y)$ and the prior $\gamma^2 \sim IG(\alpha_\gamma, \beta_\gamma)$, the posterior is given by

$$\gamma^2 | \; \{Y_{in}\}_{n=1,\dots,N_i; i=1,\dots,I}, \phi_1, \dots, \phi_I \sim IG(\alpha_\gamma^{\text{post}}, \beta_\gamma^{\text{post}})$$

$$\alpha_\gamma^{\text{post}} = \alpha_\gamma + \frac{1}{2} \sum_{i=1}^{I} N_i$$

$$\beta_\gamma^{\text{post}} = \beta_\gamma + \frac{1}{2} \sum_{i=1}^{I} \sum_{n=1}^{N_i} \frac{(Y_{in} - Y_{i(n-1)} - b(\phi_i, t_{i(n-1)}, Y_{i(n-1)})\Delta_{in})^2}{\widetilde{s}^2(t_{i(n-1)}, Y_{i(n-1)})\Delta_{in}}$$

Summary Gibbs Sampler

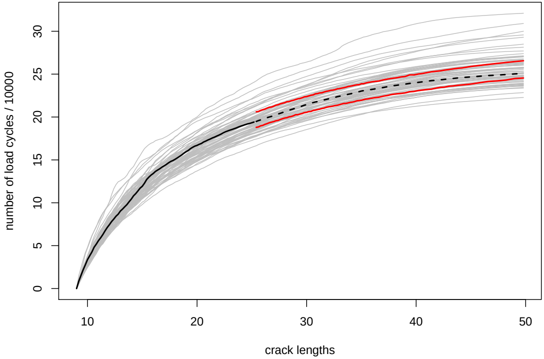
For $k = 1, ..., K$ draw iteratively:

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

$\Rightarrow p(\phi_1, ..., \phi_I, \mu, \Omega, \gamma^2 | \{y_{in}\}_{n=0,...,N_i; i=1,...,I})$ stationary distribution of Markov chain

Bayesian Prediction

Bayesian Prediction for the Hierarchical Regression Model



$$y_{in} = f(t_{in}, \phi_i) + \epsilon_{in}$$
$$\phi_i \sim \mathcal{N}(\mu, \Omega) \text{ iid.}$$
$$\epsilon_{in} \sim \mathcal{N}(0, s_1^2(\gamma^2, t_{in})),$$
$$n = 1, ..., N_i; \;\; i = 1, ..., I$$

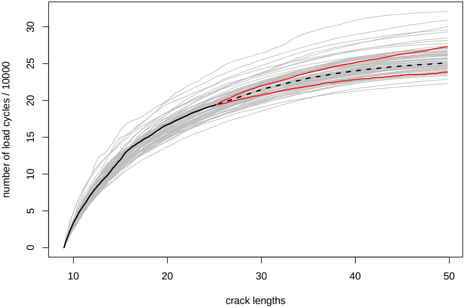
Predictive Distribution, NL Regression

Prediction for the Ith series in $t_I^* > t_{IN_I}$:

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

with $(\phi_{Ik}^*, \gamma_k^{2*}) \sim p(\phi_I, \gamma^2 | \{y_{in}\}_{n=0,\dots,N_i; i=1,\dots,I})$ coming from the Gibbs sampler.

Bayesian Prediction for the Hierarchical Diffusion Model



$$
\begin{aligned}
y_{in} &= Y_{t_{in}}(\phi_i) \\
\phi_i &\sim \mathcal{N}(\mu, \Omega) \text{ iid.} \\
dY_t &= b(\phi, t, Y_t)\, dt + s_2(\gamma^2, t, Y_t)\, dW_t, \\
n &= 0, ..., N_i; \;\; i = 1, ..., I
\end{aligned}
$$

Diffusion Model, First for a Single Series

Assume that the process $\{Y_t,\ t \in [0, \infty)\}$ defined by

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

is discretely observed in $t_0, ..., t_N$. The observation variables can be approximated by

$$\begin{aligned} Y_0 &= y_0 \\ Y_n &= Y_{n-1} + b(\phi, t_{n-1}, Y_{n-1})\Delta_n + s_2(\gamma^2, t_{n-1}, Y_{n-1})\sqrt{\Delta_n}\xi_n \\ \Delta_n &:= t_n - t_{n-1} \\ \xi_n &\sim \mathcal{N}(0,1) \text{ iid.},\ \ n = 1, ..., N. \end{aligned}$$

Prediction with Euler

Possible problems:

- the distribution of the approximated variables is normal, but this will not be the case for the true underlying distribution in general
- for $t^* \gg t_N$ the approximation can get very poor

$\hookrightarrow$ Propose a stepwise prediction procedure

Iterative Prediction Procedure

At first choose a sampling partition $t_N = \tau_0 < \tau_1 < ... < \tau_M = t^*$, for example

$$\tau_m = m \cdot \Delta^* + t_N, \; m = 0, ..., M, \quad \Delta^* := \frac{t^* - t_N}{M}$$

$\Rightarrow$ for $m = 1, ..., M$ and $Y_0^* = Y_N \approx Y_{t_N}$

$$Y_m^* | \; Y_{m-1}^*, \phi, \gamma^2 \sim \mathcal{N}(Y_{m-1}^* + b(\phi, \tau_{m-1}, Y_{m-1}^*)\Delta^*, s_2^2(\gamma^2, \tau_{m-1}, Y_{m-1}^*)\Delta^*)$$

builds an Euler approximation for Y_{τ_m}.

Practically

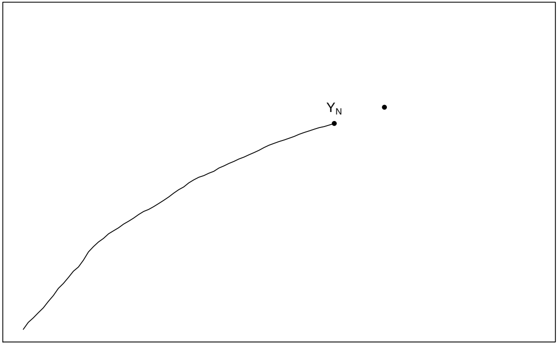
Take the samples ϕ_k^* and γ_k^{2*}, $k = 1, \dots, K$ resulting from the Gibbs sampler.

1. Set $m = 1$, $Y_0^{*(k)} := Y_N, k = 1, \dots, K$
2. Draw K samples $Y_m^{*(1)}, \dots, Y_m^{*(K)}$ from

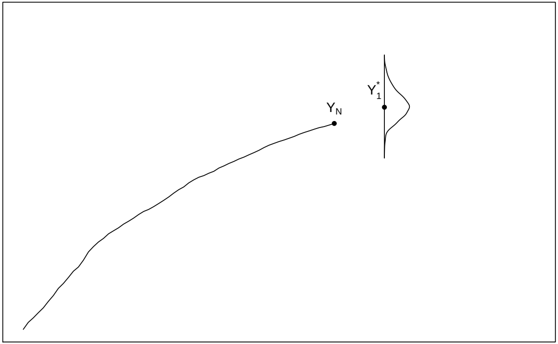
$$\frac{1}{K} \sum_{k=1}^{K} p\left(\cdot \middle| \, \phi_k^*, \gamma_k^{2*}, Y_{(m-1)}^{*(k)}\right)$$

3. If m=M stop
   else $m = m + 1$ and go to step 2

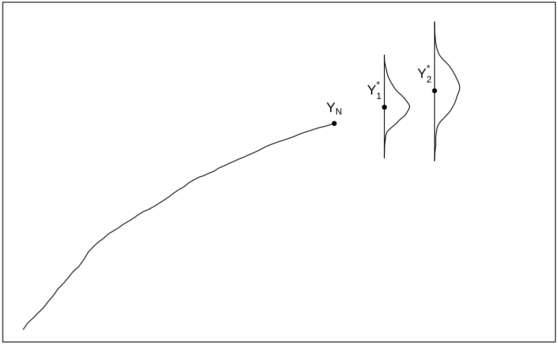
Motivation



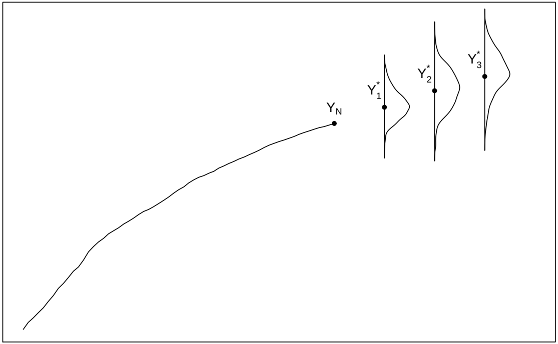
Motivation



Motivation



Motivation



Properties

- Advantage: with rejection sampling, for each $m = 1, ..., M$ the distribution is calculated once $\Rightarrow$ the algorithm gets very fast in contrast to drawing trajectories
- whole distribution of the last time point goes in, not only one sample ($\rightarrow$ more information)
- Disadvantage: one only gets the pointwise distribution, no trajectories of the process

Simulation Study

- For an SDE with explicit solution we compared our proposed prediction method with a pointwise prediction based on the explicit solution in a simulation study.
- For different simulation settings we compared the prediction of the explicit solution with Euler approximations with different partitions (i.e. different Δ^*).
- The proposed iterative prediction procedure works nearly as good as the one based on the explicit solution and much better than the one step Euler prediction.
- Δ^* must not be too small.
- Prediction with Euler for $t^* \gg t_N$ in one step leads to approximation errors.

Prediction in the Hierarchical Model

For the prediction of $Y_{t_I^*}(\phi_I)$ in t_I^*:

- take samples ϕ_{Ik}^* and $\gamma_k^{2*}, k = 1, ..., K$ from the Gibbs sampler
- choose sampling partition $t_{IN_I} = \tau_0 < \tau_1 < ... < \tau_M = t_I^*$
- start presented prediction procedure in the last Euler approximated observation variable, i.e., $Y_0^{*(k)} := Y_{IN_I}, k = 1, ..., K$
- resulting samples $Y_M^{*(1)}, ..., Y_M^{*(K)}$ simulate the predictive distribution of the Euler approximation of $Y_{t_I^*}(\phi_I)$

Application

The Data

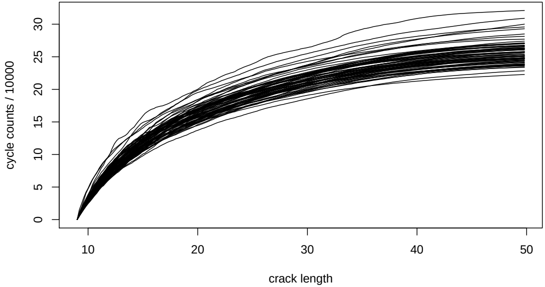


Figure: 68 series resulting of experiments with aluminum alloy, Virkler et al. (1979), independent variable: crack length

Paris Erdogan law

The Paris Erdogan law for crack length a:

$$\frac{da(t)}{dt} = \theta_1 a(t)^{\theta_2}.$$

For $\theta_2 > 1$:

$$a(t) = \{\theta_1(\theta_2 - 1)(\theta_0 - t)\}^{\frac{1}{1-\theta_2}}$$

for some θ_0 with $t < \theta_0$.

For the specific data set of Virkler et al. (1979), the inverse function is given by:

$$f(t) = a^{-1}(t) = \theta_0 - \frac{1}{\theta_1(\theta_2 - 1)} t^{1-\theta_2}$$

with corresponding derivation that leads to the differential equation

$$\frac{df(t)}{dt} = \frac{1}{\theta_1} t^{-\theta_2} = \frac{\theta_2 - 1}{t} \cdot \frac{1}{\theta_1(\theta_2 - 1)} t^{1-\theta_2} = \frac{\theta_2 - 1}{t} \cdot (\theta_0 - f(t)).$$

Other Functions in Comparison

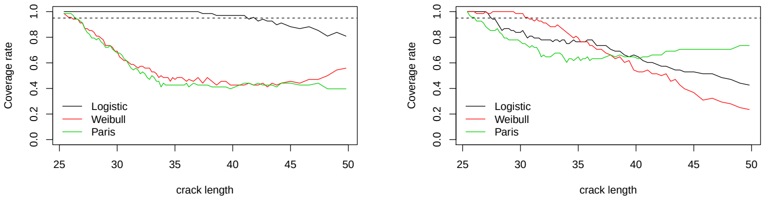
Logistic:

$$f(t,\phi) = \frac{A}{1 + Be^{-Ct}}; \quad b(\phi, t, y) = Cy(1 - \frac{1}{A}y); \quad y_0(\phi) = \frac{A}{1+B}$$

Weibull:

$$f(t,\phi) = A - Be^{-Ct^D}; \quad b(\phi, t, y) = CDt^{D-1}(A - y); \quad y_0(\phi) = A - B$$

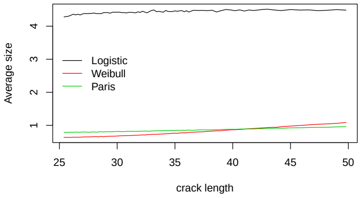
Results



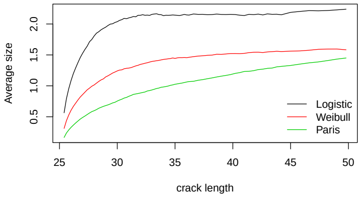
(a) hierarchical regression model   (b) hierarchical diffusion model

Figure: Coverage rates

Results



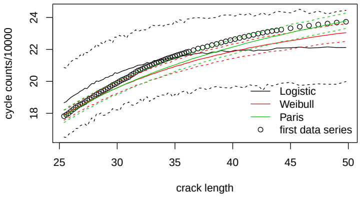
(a) hierarchical regression model



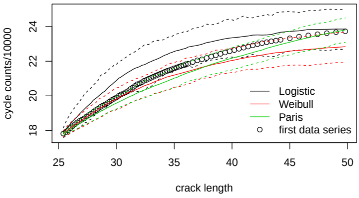
(b) hierarchical diffusion model

Figure: Average sizes

Prediction for the First Series



(a) hierarchical regression model



(b) hierarchical diffusion model

Figure: Prediction intervals for the first series

Summary

- Non-linear regression models for predicting crack growth have been used before
- Our Bayesian prediction approach for diffusion models based on the Euler approximation is new
- For the data set of Virkler et al. (1979) the advantage of the SDE model turned out to be very small
- Alternatives: other variance functions s_1 and s_2
- Possible extensions: transfer of theory to other approximation schemes than Euler

References

Hermann, S., Ickstadt, K. and Müller, C. H. (2016). Bayesian prediction of crack growth based on a hierarchical diffusion model. *Applied Stochastic Models in Business and Industry* 32(4), 494–510.

Donnet, S., Foulley, J.-L. and Samson, A. (2010) Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics 66, 733–741*.

Virkler, D. A., Hillberry, B. M. and Goel, P. K. (1979). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology* 101(2), 148–153.

Simulation Study for the Gompertz Process

With $\phi = (A, B, C)$ and the example SDE

$$dY_t = BCe^{-Ct} Y_t \, dt + \gamma Y_t \, dW_t, \;\; Y_0 = A \cdot e^{-B}$$

we get the explicit solution

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

and can compare the proposed method with the pointwise prediction based on the explicit solution.

Prediction Based on the Explicit Solution

It is

$$\log(Y_{t^*})|\ \log(Y_{t_N}), \phi, \gamma^2 \sim$$
$$\mathcal{N}\left(\log(Y_{t_N}) - B(e^{-Ct^*} - e^{-Ct_N}) - \frac{1}{2}\gamma^2(t^* - t_N), \gamma^2(t^* - t_N)\right)$$

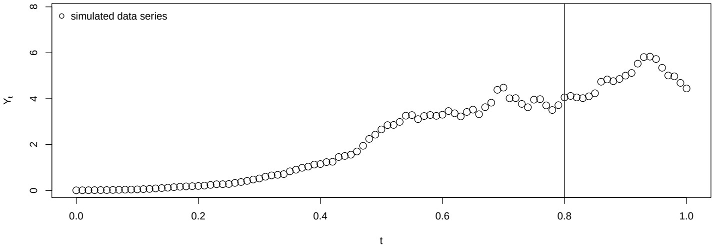
$\Rightarrow$ predictive distribution

$$\begin{aligned} & p(\log(Y_{t^*})|\ Y_{t_1}, ..., Y_{t_N}) \\ = & \int p(\log(Y_{t^*})|\ \log(Y_{t_N}), \phi, \gamma^2) p(\phi, \gamma^2|\ Y_{t_1}, ..., Y_{t_N})\, d(\phi, \gamma^2) \\ \approx & \frac{1}{K}\sum_{k=1}^{K} p(\log(Y_{t^*})|\ \log(Y_{t_N}), \phi_k^*, \gamma_k^{2*}) \end{aligned}$$

Simulation Settings

- draw 1000 series with the explicit solution in $t_0 = 0, \ldots, t_{100} = 1$
- truncate the series, use the first 81 points for the estimation and the last 20 for prediction ($t^* \in \{t_{81}, \ldots, t_{100}\}$)
- estimate ϕ and γ^2 with the explicit solution ($K = 1000$ samples)
- predict
  - with the explicit solution
  - with Euler and $\Delta^* = 0.01$ as in the observation variables
  - with Euler and $\Delta^* = 0.001$ for comparison
  - with Euler and $M = 1$, i.e., no partition of the interval, to check whether there are any approximation errors

Example, one Series



Example, one Series

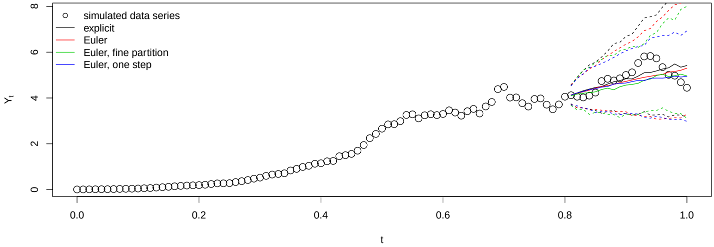


Figure: Pointwise prediction intervals for the four 9 different methods, first with the explicit solution, second with Euler and $\Delta^* = 0.01$, third with Euler and $\Delta^* = 0.001$ and fourth with Euler without any partition

Example, one Series

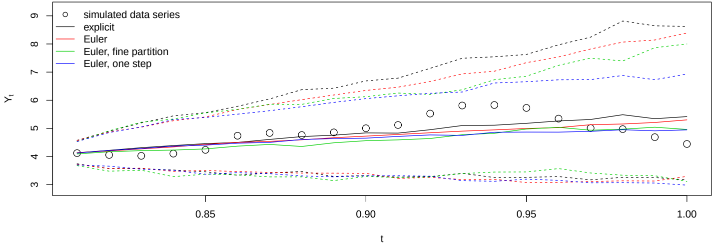


Figure: Pointwise prediction intervals for the four 9 different methods, first with the explicit solution, second with Euler and $\Delta^* = 0.01$, third with Euler and $\Delta^* = 0.001$ and fourth with Euler without any partition

Coverage Rates

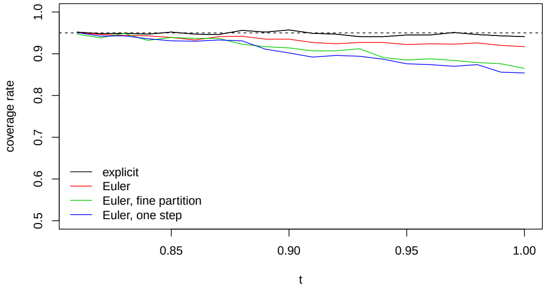


Figure: Coverage rates for the four different prediction methods, first with the explicit solution, second with Euler and $\Delta^* = 0.01$, third with Euler and $\Delta^* = 0.001$ and fourth with Euler without any partition

Mean Interval Sizes

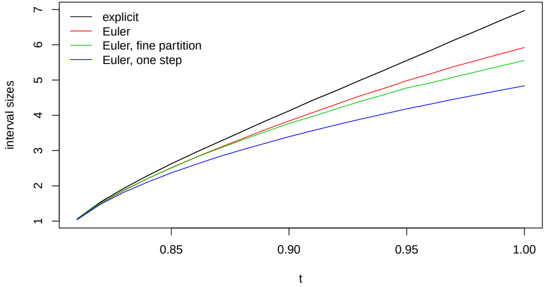


Figure: Interval sizes in average for the four different prediction methods, first with the explicit solution, second with Euler and $\Delta^* = 0.01$, third with Euler and $\Delta^* = 0.001$ and fourth with Euler without any partition