Failure prediction for monitored systems

B. H. Lindqvist & V. Slimacek

Department of Mathematical Sciences Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: We consider a monitored system with observation Y(t) at time t modeled by a stochastic process, and where system failure is connected to the exceedance of some threshold for this process. Typically, the threshold is not exceeded under normal conditions, where the process Y(t) is supposed to show some kind of stationarity. However, due to unexpected events the process may leave the stationary behavior, in which case an early detection of an increasing Y(t) is necessary to avoid costly failures of the system. The prediction of the time T of future exceedance of a given threshold for the process Y(t) will hence be an important issue, and is the basic problem studied in the paper. We assume that Y(t) is a stochastic process with a probability mechanism depending on an unobservable underlying process S(t). The latter process has a finite state space, $\{0, 1, \dots, k\}$, where state 0 corresponds to the normal stationary conditions for the process Y(t), while states 1,2,... in increasing order means an increasing severity of the underlying conditions which eventually will lead to system failure. In particular we consider the case when Y(t) is modeled as a Wiener process. It is then natural to assume that the drift parameter of the process equals 0 when S(t) = 0, while the drift is positive and increasing with S(t), if S(t) > 1. A special case with k = 1 will be considered in detail. In this case estimation of the unobservable "switching" time τ at which the underlying process S(t) changes from state 0 to state 1 is of particular interest. A Bayesian approach will be used, where a Markov Chain Monte Carlo approach will be needed for doing the computations in the most general cases.

1 INTRODUCTION

The motivating example for this paper is the modeling of failure development of a bearing in an offshore wind turbine, and the prediction of ultimate failure based on continuous monitoring of temperature (Brurok, Valland, Lindqvist, & Slimacek 2012).

Assume that the normal temperature of the bearing under production is around 60 degrees Celsius. If temperature increases from this level, it is therefore an indication that a failure is developing. Such an increase may for example be caused by impurities in the oil, which in turn will lead to mechanical wear and in the end to bearing breakdown and turbine shut down. Early detection of the failure development is therefore of great importance, and the aim of this paper is to present a simple first approach to a prediction model for such failure developments.

Several models have been devised for the situation where the failure of a system is connected to a certain stochastic process exceeding some threshold. The probably most popular type of stochastic processes in such applications is the Wiener process, which may be realistic in cases when measurements are done continuously and when the path of the measurements is continuous. Some examples of theoretical treatments in this connection are Whitmore (1986), Whitmore & Schenkelberg (1997), Whitmore, Crowder, & Lawless (1998), Lindqvist & Skogsrud (2008). Alternatively, one may use gamma processes, which for some applications may seem more reasonble due to paths being increasing functions of time, see Lawless & Crowder (2004) for an application and the comprehensive review paper Van Noortwijk (2009).

In the present paper, motivated by the continoously monitored bearing temperature, we will stick to the Wiener process as the basic modeling tool. A brief comparison to a gamma process approach will be given in the final section of the paper.

The rest of the paper is organized as follows. In Section 2 we give the definition and some relevant properties of the Wiener process. Section 3 presents the probabilistic model for the development of the monitored process, both under normal conditions and after the initiation of a failure development. A simplified model is considered in the final subsection, and the practical problem of predicting the failure is considered in Section 4, using a Bayesian approach. Some concluding remarks are given in the final Section 5.

2 THE WIENER PROCESS AND THE INVERSE GAUSSIAN DISTRIBUTION

Definition 1 A stochastic process $\{W(t), t \ge 0\}$ is a Wiener process with drift coefficient ν and variance parameter σ^2 if

- 1. W(0) = 0,
- 2. $\{W(t), t \ge 0\}$ has stationary and independent increments,
- 3. for every t > 0, W(t) is normally distributed with mean νt and variance $\sigma^2 t$.

A special feature that makes the Wiener process mathematically tractable is that if the drift parameter ν is positive, then the first passage time to a level a > 0 is inverse Gaussian distributed with density

$$f(t;\nu,\sigma,a) = \frac{a}{\sqrt{2\pi\sigma}} t^{-\frac{3}{2}} \exp\left\{-\frac{(a-\nu t)^2}{2t\sigma^2}\right\}, t > 0,$$

(Aalen & Gjessing 2001, e.g.). We denote this distribution by $IG(\nu, \sigma, a)$. The corresponding survival function is given by

$$S(t;\nu,\sigma,a) = \Phi\left(\frac{a-\nu t}{\sigma\sqrt{t}}\right) - e^{2a\nu/\sigma^2}\Phi\left(\frac{-a-\nu t}{\sigma\sqrt{t}}\right)$$
(1)

where Φ is the standard normal cumulative distribution function. It should be noted that the above functions depend on the parameters only via a/σ and ν/σ . We shall, however, find it convenient to use the representation above using the three parameters ν, σ, a . We shall also later need the expected value of the corresponding distribution, which is a/ν .

3 PROBABILISTIC MODELING OF THE MONITORED PROCESS

Let Y(t) be the monitored process, which for illustrative purposes we assume is the temperature of the windturbine bearing. Thus Y(t) is the measured temperature at time t, observed from a starting time t = 0, with time unit given as days, say. In order to make probabilistic predictions we shall consider Y(t) as a stochastic process, where the aim is to predict its future behavior from observation of Y(t) from time 0 to some given time t_0 . The object of main interest is the time when the temperature exceeds some threshold, a.

3.1 A latent Markov process

Assume that the failure development can be considered as a sequence of distinguished stages, numbered as Stage 1, Stage 2, etc. In our motivating example, these stages could be (Brurok, Valland, Lindqvist, & Slimacek 2012); Stage 1: Impurities in oil; Stage 2: Mechanical wear; Stage 3: Micropitting, pitting; Stage 4: Chipping; Stage 5: Bearing breakdown; Stage 6: Turbine shut down.

The failure development defined through stages may be modeled as a latent discrete state process S(t). Then we let S(0) = 0 (normal condition) and let S(t)be the failure stage number at time t. A possible probabilistic model defines S(t) to be a continuous time Markov process with finite state space $\{0, 1, \ldots, k\}$, where k = 6 in the example, and time-homogeneous transition rates. This process is particularly simple to describe and handle under the natural assumption that S(t) is non-decreasing.

The latent process S(t) is assumed to affect the probability mechanism of the temperature process Y(t) in a way such that different probability regimes are valid under each possible state of S(t). This will be considered in more detail next.

3.2 The Wiener process based model for Y(t)

The stochastic process Y(t) will basically be modeled as a piecewise Wiener process. The idea is that the parameters ν and σ of the underlying Wiener process (Section 2) are allowed to vary with the state of the latent process, S(t). It is then natural to assume that the drift parameter ν equals 0 when S(t) = 0 (normal conditions). Under the failure development through Stages 1 to k, on the other hand, the drift is assumed to be positive, with values ν_i when S(t) = i, and with ν_i increasing with i. Likewise, the variance parameter σ may depend on the state S(t).

It should here be noted that the independent increment property of the Wiener process makes the above described change of parameter values particularly easy to handle. More precisely, by properties of the Wiener process, for the process Y(t) will hold that for any s < t is Y(t) - Y(s) normally distributed with expected value $\nu(t - s)$ and variance $\sigma^2(t - s)$. Further, the differences Y(t) - Y(s) obtained from pairwise disjoint time intervals are stochastically independent.

As already mentioned in Section 2, a feature that makes the Wiener process mathematically tractable is that the first passage time to a given level, assuming positive drift, is inverse Gaussian distributed. This property is unfortunately violated for the more general model considered for Y(t) due to its dependence on the latent process S(t). We shall, however, below consider a simplified, though realistic, version, in which the inverse Gaussian distribution can still be identified.

3.3 The Wiener process with a single change point

Assume that the temperature process Y(t) from time t = 0 follows a Wiener process started at Y(0) = 60. Further, assume that until a "switching" time τ (the time of entrance to Stage 1 in the bearing application)



Figure 1: The Wiener process with a single change point: Simulation of the temperature process Y(t) for $0 \le t \le 5000$ with $\tau = 1500$, $\sigma = 0.05$, $\nu = 0.003$. Horizontal axis is time (days); vertical axis is temperature.

there is no drift. Then from time τ on, there is a positive drift ν . Let T be the time when the temperature crosses a given level, say a > 0. We shall make the simplifying assumption that even if there may be a small probability that a is reached before time τ , T is always larger than τ . More precisely we define T as

$$T = \inf\{t > \tau | Y(t) \ge a\}.$$

Conditional on τ and $Y(\tau)$, $T - \tau$ has an inverse Gaussian distribution as already described. This corresponds to the time of hitting the threshold $a - Y(\tau)$ for a Wiener process with drift ν , starting at state 0. In particular, the conditional expected value of T is then $\tau + (a - Y(\tau))/\nu$, as follows from the result at then end of Section 2.

Figure 1 shows a simulation of the process Y(t) with $\tau = 1500$, $\sigma = 0.05$, $\nu = 0.003$. If a = 65, say, then the expected value of T (i.e. time for crossing 65 degrees) will be

$$1500 + \frac{65 - 58}{0.003} = 3833$$

which seems reasonable from the figure.

Alternatively, one may be interested in computing a value t_p , say, such that, for some p close to 1, $P(T > t_p) = p$. This may be computed using (1).

4 A PRACTICAL ILLUSTRATION

To make a simple practical illustration, assume now that it is known that $\sigma = 0.05$, $\nu = 0.003$ (these could be based on a long time experience with similar turbines). Assume also that the process has been observed from time 0 to some given time t_0 . In the example we will consider $t_0 = 2500$. Let the problem be to estimate the value of τ based on the observation of the process from time 0 to time t_0 . From the estimate $\hat{\tau}$, say, we can then estimate the expected time of crossing the threshold a to be

$$\hat{\tau} + (a - Y(\hat{\tau}))/\nu,$$

using the earlier result.

It follows from the above that the estimation of τ is of primary interest, and that interesting predictions can be made from this.

4.1 A discrete time solution

The following is a way to solve the problem by a Bayes approach. It is now assumed that the temperature Y(i) is observed at discrete time points, here *days*, i = 1, 2, ..., n. The independent increment property of the Wiener process means that the differences $X_i = Y(i+1) - Y(i)$ are independent and normally distributed. The assumptions made above moreover imply that, for an unknown positive integer time $\tau, X_1, ..., X_{\tau-1}$ are $N(0, \sigma^2)$, while $X_{\tau}, X_{\tau+1}, ..., X_n$ are $N(\nu, \sigma^2)$. If $\tau > n$, then accordingly, $X_1, ..., X_n$ are $N(0, \sigma^2)$.

Following Shiryaev (1963) we shall give τ a prior distribution which is geometric with

$$\pi(\tau) = q(1-q)^{\tau-1}$$
 for $\tau = 1, 2, \dots$ (2)

where q is a given number. Note that then the prior expectation of τ is 1/q. Note also that the geometric distribution is the discrete analogue to the exponential distribution, and has the memoryless property. In our case it has furthermore a reasonable interpretation if we think of q as the probability of a Stage 1 condition starting at random with the same probability q at any given day.

In order to obtain the posterior distribution for τ we shall need the likelihood function for our data, $L(x_1, \ldots, x_n | \tau)$, which is proportional to

$$\exp\{-\frac{1}{2\sigma^2} (\sum_{i=1}^{\tau-1} x_i^2 + \sum_{i=\tau}^n (x_i - \nu)^2)\} \text{ if } \tau \le n; \qquad (3)$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} \text{ if } \tau > n.$$
(4)

The posterior distribution for τ is hence proportional to $\pi(\tau)L(x_1,\ldots,x_n|\tau)$. Multiplying L by $\exp\{-(1/2\sigma^2)\sum_{i=1}^n x_i^2\}$ we get the posterior distribution for τ , $\pi(\tau|x_1,\ldots,x_n)$, proportional to $q(1-q)^{r-1}$ times the function which is

$$\exp\{\frac{1}{2\sigma^2}[2\nu\sum_{i=\tau}^n x_i - (n-\tau+1)\nu^2]\}$$
(5)

if $\tau \leq n$ and 1 if $\tau > n$.

Figure 2 shows the non-normalized posterior distributions computed at time n = 2500. For the prior distribution for τ in (2) was chosen q = 1/5000. It is seen that the maximum posterior value for τ is not far from the true value $\tau = 1500$ used to simulate the data. If we are interested in the posterior expectation for τ , we need to normalize (5) to ensure the sum over τ to be 1. Having done this, we computed the value



Figure 2: Posterior distribution (non-normalized) for τ (horizontal axis) in the simulated example in Figure 1, estimated at time n = 2500. The prior distribution (2) for τ has q = 1/5000.

 $\hat{\tau} = 2275$ for n = 2500. We did this computation also for other values of n, with the following results; for n = 3000: $\hat{\tau} = 2158$; for n = 5000: $\hat{\tau} = 1380$. As is natural, the estimates become better as the observation time n increases. When considering these values it should be noted that the prior expectation for τ was set to 5000.

4.2 The continuous time solution

Suppose now that the time parameter is continuous and that $\tau > 0$ is a continuous parameter. This case can be seen as the limiting case of the discrete time case just considered, when the discrete time unit tends to 0. Then the geometric prior distribution for τ becomes the exponential distribution,

$$\pi(\tau) = \lambda e^{-\lambda \tau}$$
 for $\tau > 0$,

where $\lambda > 0$ is a parameter to be given. Note that the expected value of the prior distribution is now $1/\lambda$.

It follows, furthermore, from (5) that the posterior distribution for an observation of Y(t) from t = 0 to $t = t_0$, $\pi(\tau | Y(t), 0 \le t \le t_0)$, is proportional to $\lambda e^{-\lambda \tau}$ times the function which is

$$\exp\{\frac{1}{2\sigma^2}[2\nu(Y(t_0) - Y(\tau)) - (t_0 - \tau)\nu^2]\}$$
(6)

if $\tau \leq t_0$ and 1 if $\tau > t_0$.

4.3 Markov Chain Monte Carlo (MCMC) solution

For the discrete time case considered above, we argued that a normalization is necessary for finding the expected posterior value for τ . It should then be noted that the normalization can not in general be done analytically. In particular it is necessary to set a largest possible value for τ .

Alternatively, one may use an MCMC solution by for example the Metropolis-Hastings method (Gilks, Richardson, & Spiegelhalter 1996, e.g.). There are several possible ways here, for example using an independence sampler which draws proposals from the prior distribution. (More efficient algorithms can of course be thought of).

Suppose now that also ν is an unknown parameter. The posterior distribution for (τ, ν) is then given by replacing the prior distribution for τ (5) (or the continuous time version (6) by the joint prior of (τ, ν) , which might be the product of their marginal priors if they are assumed independent.

If, on the other hand, also σ is assumed unknown, then the likelihood function in (3)-(4) shows that the expressions (5) and (6) can no longer be used as shortcuts for the posterior distribution. Still (3)-(4) multiplied by the prior for τ , ν , σ can be used via MCMC. In practice it may be, however, reasonable to assume that the value of σ is known from long experience. Another reasonable possibility is that while σ has a known value under normal conditions, this value may change beyond the start of Stage 1. This may also be taken into account in the model.

5 CONCLUDING REMARKS

Guérin, Barreau, Demri, Cloupet, Hersant, & Hambli (2010) perform a Bayesian analysis of a degradation model based on a Wiener process, related to the approach presented here.

While the Wiener process and its modifications may be useful to model temperatures, as in the presented example, the gamma process may be a useful alternative for the modeling of other types of monitored components. A number of examples are given by Van Noortwijk (2009).

As is the case for the Wiener process, a gamma process X(t) has independent increments, but the increments X(t) - X(s) are gamma-distributed with shape parameter $\alpha(t-s)$ and scale parameter β , for positive parameters α, β .

Fouladirad, Grall, & Dieulle (2008) use a gamma process model for a deteriorating process, where the aim is to detect the point in time, τ , where the deterioration process reaches a more severe level, defined by new values for both α and β . This is similar to the problem considered in the present paper. The cited article considers the detection of the change as a means for controlling condition based maintenance.

In some cases, the states of an underlying discrete state process like the process S(t) considered in Section 3, are directly observable, for example at certain inspection times. Predictions of the future behavior of a component may then be based on these observations. The models considered by Welte & Kile (2011) may be of interest in this connection.

REFERENCES

Aalen, O. O. & H. Gjessing (2001). Understanding the shape of the hazard rate: A process point of view (with comments and a rejoinder by the authors). *Statistical Science* 16(1), 1–22.

- Brurok, T., A. Valland, B. H. Lindqvist, & V. Slimacek (2012). Probabilistic modeling of wind farm using TCI framework. Manuscript draft.
- Fouladirad, M., A. Grall, & L. Dieulle (2008). On the use of online detection for maintenance of gradually deteriorating systems. *Reliability Engineering & System Safety 93*(12), 1814– 1820.
- Gilks, W. R., S. Richardson, & D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. Chapman Hall/CRC.
- Guérin, F., M. Barreau, A. Demri, S. Cloupet, J. Hersant, & R. Hambli (2010). Bayesian estimation of degradation model defined by a Wiener process. In *Mathematical and Statistical Models and Methods in Reliability*, pp. 345–357. Springer.
- Lawless, J. & M. Crowder (2004). Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis 10*(3), 213–227.
- Lindqvist, B. H. & G. Skogsrud (2008). Modeling of dependent competing risks by first passage times of wiener processes. *IIE Transactions* 41(1), 72–80.
- Van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety 94*(1), 2–21.
- Welte, T. M. & H. Kile (2011). Parameter estimation in degradation modelling: A case study using conditional monitoring data from wood pole inspections. In *Power Tech*, 2011 IEEE *Trondheim*, pp. 1–7. IEEE.
- Whitmore, G., M. Crowder, & J. Lawless (1998). Failure inference from a marker process based on a bivariate wiener model. *Lifetime Data Analysis* 4(3), 229–251.
- Whitmore, G. A. (1986). First-passage-time models for duration data: regression structures and competing risks. *The Statistician*, 207–219.
- Whitmore, G. A. & F. Schenkelberg (1997). Modelling accelerated degradation data using wiener diffusion with a time scale transformation. *Lifetime Data Analysis* 3(1), 27–45.