A general framework for statistical inference on discrete event systems

Robin P. Nicolai*
Tinbergen & Econometric Institute
Erasmus University Rotterdam

Alex J. Koning†
Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report 2006-45

Abstract

We present a framework for statistical analysis of discrete event systems which combines tools such as simulation of marked point processes, likelihood methods, kernel density estimation and stochastic approximation to enable statistical analysis of the discrete event system, even if conventional approaches fail due to the mathematical intractability of the model.

The approach is illustrated with an application to modelling and estimating corrosion of steel gates in the Dutch Haringvliet storm surge barrier.

Keywords: parameter estimation; discrete event systems; optimization via simulation; marked point process; likelihood methods; stochastic approximation; kernel density estimation;

*Corresponding author. Tinbergen & Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. E-mail address: rnicolai@few.eur.nl. Tel. +31 10 4082524; fax: +31 10 4089162.

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. E-mail address: koning@few.eur.nl.

1 Introduction

Uncertainty is an important part of life, and hence statistical modelling is appropriate for many real life phenomena. Unfortunately, the complexity of statistical models is a trade-off between the mathematical tractability of the model on one hand, and insight into the structure of the phenomenon of interest on the other hand. On the one side of the spectrum there are black box models, which are easily analyzed statistically, but do not provide any deep structural insights. On the other side of the spectrum there are white box models, which carefully detail structure, but are hardly amenable for statistical analysis.

In this paper we focus on discrete event systems, dynamic systems which result in the occurrence of events at particular points in time; each event has a certain discrete type. In the context of discrete event systems, modelling the time to occurrence of a certain event by choosing a distribution from some catalogue of distributions (for instance, Weibull, gamma, lognormal) is an example of the black box approach, whereas a full-fledged simulation model is an example of a white box model.

Our primary motivation for undertaking this study comes from experience in the field of degradation modelling, see Nicolai *et al.* (2006). In this field, there is an abundance of grey box models which are motivated by but do not fully capture the underlying physical process of degradation. Examples are degradation models derived from Brownian motion, gamma or compound Poisson processes, see Doksum and Hoyland (1992); Çinlar *et al.* (1977); Sobczyk (1987). These models have been successfully applied to quantify degradation such as wear, corrosion, crack growth and creep.

As black box models are available for purely descriptive purposes, the abundance of grey box degradation models underlines the need for incorporating structural insights into the statistical models. Nevertheless, white box models, the pinnacles of structural thinking, are not used in the statistical analysis. The reason for this is that conventional approaches to statistical analysis are not feasible due to the mathematical intractability of the white box models. For instance, applying likelihood methods is not feasible since computing the likelihood becomes too complex.

The objective of this paper is to develop a likelihood based methodology for statistical analysis of discrete event systems by means of white box models, in which the mathematical intractability of the white box model

is circumvented by using computer intensive methods. In particular, likelihood is "maximized" using simulation, density estimation and stochastic approximation.

The remainder of this paper is as follows. In Section 2 we introduce our methodology, in Section 3 we illustrate the approach with an application to modelling and estimating corrosion, and in Section 4 we draw conclusions.

2 Methods

2.1 Model building

Capturing the white box model In this section we present a methodology for statistical analysis of discrete event systems given a white box model. As the methodology is likelihood based, we first need to capture the white box model in mathematical equations, which allows us to identify the parameters to be estimated.

We achieve this by formulating the discrete event systems in terms of a marked point process, a stochastic process which consists of events of a certain type (the mark) taking place at certain points of time. To each marked point process belongs a counting process $N^* = \{N^*(t) : t \geq 0\}$, which counts the number of occurrences of events in the interval $[0, t]$. Remark that the counting process increases by 1 at every occurrence of any event. Thus, every "jump time" of the counting process coincides with an occurrence of some event, and vice versa. In addition, to each marked point process there belongs a mark sequence $S_1, S_2, \ldots$, which accommodates the event type: the i^{th} mark S_i gives the type of the event occurring at the i^{th} jump time. The marks S_i take values in some mark space $\mathcal{S}$. For discrete event systems, the mark space $\mathcal{S}$ is discrete.

As for every possible mark value $s \in \mathcal{S}$ there exists a counting process $N_s = \{N_s(t) : t \geq 0\}$, which counts the number of occurrences of events of type s in the interval $[0, t]$, we may alternatively view the marked point process as a multivariate counting process $\{N_s\}_{s \in \mathcal{S}}$. Moreover, we may write $N^* = \sum_{s \in \mathcal{S}} N_s$.

As a counting process is a submartingale, it follows as a result of the Doob-Meyer decomposition that there exists a corresponding compensator, that is, a nondecreasing predictable process such that the difference of the counting process and its compensator is a martingale, see (Andersen *et al.*,

1993, p. 73). Predictability may be characterized as follows: the value of a predictable process at time t is completely determined just before t, see (Andersen *et al.*, 1993, p. 67).

The derivative of the compensator with respect to time is called the intensity rate function. Let $\Lambda^* = \{\Lambda^*(t) : \ t \geq 0\}$ and $\Lambda_s = \{\Lambda_s(t) : \ t \geq 0\}$ respectively denote the compensator, and let $\lambda^* = \{\lambda^*(t) : \ t \geq 0\}$ and $\lambda_s = \{\lambda_s(t) : \ t \geq 0\}$ respectively denote the intensity rate function belonging to N^* and N_s. We may write

$$\Lambda^*(t) = \int_0^t \lambda^*(s)ds, \quad \Lambda_s(t) = \int_0^t \lambda_s(s)ds.$$

It is known that the structure of the compensator completely describes the probabilistic behaviour of the corresponding counting process, and thus we may model a counting process by imposing structure on its compensator, or equivalently, imposing structure on its intensity rate function. As

$$\lambda^* = \sum_{s \in \mathcal{S}} \lambda_s, \tag{1}$$

it follows that a discrete event system may be modelled by imposing structure on each of the mark-specific intensity rate functions λ_s. We shall assume that $\lambda_s(t)$ takes the form $\lambda(t; \theta_s, \mathbf{z}_s)$, where θ_s is a vector of structural parameters, $\mathbf{z}_s = \{\mathbf{z}_s(t) : \ t \geq 0\}$ is a vector of possibly time dependent covariates, and $\lambda(\cdot; \theta_s, \mathbf{z}_s)$ is the generic form of the intensity rate functions λ_s. The covariate vector $\mathbf{z}_s$ contains all environmental factors determining the "risk" of an event of type s. Remark that λ_s should be predictable, that is, $\lambda_s(t)$ is completely determined just before t; hence, we shall tacitly assume that a time dependent covariate $\mathbf{z}_s$ only has effect on $\lambda_s(t)$ through its "historical part" $\{\mathbf{z}_s(t')\}_{0<t'<t}$.

The covariate vector may include global as well as local environmental factors. An often useful way to handle local environmental factors is to incorporate the state of a neighbourhood system as covariate in the model. A neighbourhood system defines a neighbourhood $\mathcal{B}_s$ for each $s \in \mathcal{S}$. For example, if $\mathcal{S}$ is a two-dimensional grid, then the neighbourhood of s may contain the four sites directly north, west, south and east of s, see Figure 1. Note that s itself is not an element of $\mathcal{B}_s$.

The distinction between global and local environmental factors implies the existence of three types of structural parameters: (i) local coefficients,

parameters which act as regression coefficients for local environmental factors; (ii) global coefficients, parameters which act as regression coefficients for global environmental factors; (iii) baseline parameters, parameters which do not act as regression coefficients for environmental factors.

In model building, one typically starts with a baseline model which does not include any covariate. Then, the baseline model is extended by including covariates. This can be conveniently done by multiplying a baseline parameter by $r\left(\beta^T \mathbf{z}_s\right)$, where β is a vector of unknown regression coefficients, and r is a known scalar function. Typical choices are $r(v) = e^v$ and $r(v) = v$, yielding multiplicative and additive covariate effects, respectively.

Finally, we present a scheme for simulating a marked point process, which extends Algorithm 7.4.III in (Daley and Vere-Jones, 2003, p. 260), an algorithm for simulating counting processes, by including the assignment of event types by the inverse transform method. This scheme involves $p_s(t)$, the probability that the next event of N^* is of type s given that it occurs at time t. Observe that $p_s(t) = \lambda(t; \theta_s, \mathbf{z}_s)/\lambda^*(t)$.

Algorithm 1 (Simulation scheme marked point process $\{N_s\}_{s\in\mathcal{S}}$) *Initially, let t_0 be the starting time of the simulation.*

(1) Draw E, an exponential random variable with mean 1.

(2) The next event time t_1 is the solution of $\Lambda^*(t) = \Lambda^*(t_0) + E$ with respect to t.

(3) Select event type s with probability $p_s(t_1)$. Specifically,

- *(3.1) Compute $p_s(t_1)$ for every $s \in \mathcal{S}$.*
- *(3.2) Let $F = 0$ and set s equal to the first element of $\mathcal{S}$. Draw U, a random variable distributed uniformly on $[0, 1]$.*
- *(3.3) Let $F = F + p_s(t_1)$.*
- *(3.4) If $U < F$, then return event type s and go to step (4). Otherwise, set s equal to the next element of $\mathcal{S}$, and return to step (3.3).*

(4) Set $t_0 = t_1$, update Λ^*, and return to step (1).

The observation scheme Algorithm 1 allows us to provide a model for a phenomenon under study. The next step involves fitting this model to sampled data, observations obtained from the phenomenon. As it is usually not feasible to fully observe all aspects of the phenomenon, we have to resort to partial observation in practice: we focus on one or more particular quantitative aspects of the phenomenon, and measure them at one or more points in time. Thus, on top of the phenomenon there is some observation scheme which ultimately produces the sampled data, a finite sequence of measurements, say $(X_1, X_2, \ldots, X_n)$. In general, the random variables $X_1, X_2, \ldots, X_n$, will be neither independent nor identically distributed. Thus, we should view $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ as a random vector drawn from some multivariate distribution.

In order to be able to simulate from the multivariate distribution of $\mathbf{X}$, we should mimic the observation scheme in simulations by some n-dimensional function $\mathbf{w}$ which assigns to each marked point process $\{N_s\}_{s\in\mathcal{S}}$ simulated by Algorithm 1 a random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ of measurements $w_1\left(\{N_s\}_{s\in\mathcal{S}}\right)$, $w_2\left(\{N_s\}_{s\in\mathcal{S}}\right)$, $\ldots$, $w_n\left(\{N_s\}_{s\in\mathcal{S}}\right)$.

2.2 Statistical inference

Likelihood Next, we take the multivariate distribution of the random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ as a model for $\mathbf{X}$. As the random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ is directly obtained from the marked point process $\{N_s\}_{s\in\mathcal{S}}$, it follows that the multivariate distribution of $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ has the same parameters as $\{N_s\}_{s\in\mathcal{S}}$, say $\theta_1, \theta_2, \ldots, \theta_k$. Collect these parameters in the k-dimensional parameter vector θ. Remark that θ contains all distinct elements of θ_s for every $s \in \mathcal{S}$. Let $\Theta \subset \mathbb{R}^k$ be the set of all possible values of θ.

Since we now have a statistical model for $\mathbf{X}$ which depends on a parameter vector θ, we are ready for parametric statistical inference on θ. We shall limit ourselves to likelihood methods. Although there are many approaches to statistical inference, likelihood methods have always been popular following their proposal more than eighty years ago, and consequently have become well understood.

The likelihood of a statistical model evaluated in an abitrary element ϑ of parameter space Θ is given by $L(\vartheta) = f_{\mathbf{X}}(\mathbf{X}; \vartheta)$, where $f_{\mathbf{X}}(\cdot; \vartheta)$ is the joint density of $\mathbf{X}$ under this model. In simpler models we are able to derive a closed mathematical expression for $L(\vartheta)$. Unfortunately, as we are dealing

with white box models, expressing $L(\vartheta)$ in this way is not feasible due to the mathematical intractability of the white box models.

However, we do have a simulation model, which allows us to generate data from the joint distribution of $\mathbf{X}$ under the statistical model. As the likelihood $L(\vartheta)$ is in essence a joint density evaluated in $\mathbf{X}$, we may resort to density estimation on basis of the simulated data to obtain an approximation to $L(\vartheta)$. Therefore, we independently run the simulation m times for the parameter value ϑ. As the j^{th} simulation run yields a n-dimensional simulated vector of observations $\mathbf{X}^{(j)}$, we obtain independent random vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}$, all obeying the common unknown density $f_{\mathbf{X}}(\cdot; \vartheta)$.

For ease of exposition, we shall in first instance focus on the univariate case $n = 1$, in which the sequence of random vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}$ coincides with a sequence of independent random variables $X^{(1)}, X^{(2)}, \ldots, X^{(m)}$. Define the kernel density estimator $\hat{f}(x)$ by

$$\hat{f}(x) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{X^{(j)} - x}{b}\right).$$

Here the kernel K is some probability density, symmetric around zero, and satisfying $\int (K(x))^2 \, dx < \infty$. The smoothing index b is often referred to as the bandwidth. Selecting the optimal bandwidth is an important topic in density estimation. A popular choice is the direct plug-in (DPI) bandwidth selector described in (Wand and Jones, 1995, paragraph 3.6.1). A variant of the DPI method is the "solve the equation" plug-in bandwidth proposed in Sheather and Jones (1991). The Sheather-Jones plug-in bandwidth has been implemented in most statistical packages, and is widely recommended due to its overall good performance, see Sheather (2004).

In the general case we will be faced with the situation where $n > 1$, which forces us to resort to multivariate density estimation, a straightforward extension of univariate density estimation. The general expression for the multivariate kernel estimator is

$$\hat{f}(\mathbf{x}) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{\mathbf{X}^{(j)} - \mathbf{x}}{b}\right),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, b is the bandwidth, and the kernel K is some n-variate square integrable probability density, symmetric around the origin. Work on the selection of the optimal bandwidth in multivariate density estimation is currently in progress, see Duong and Hazelton (2005).

One of the problems in multivariate density estimation is the "curse of dimensionality", see (Scott, 1992, Section 7.2). This relates to the fact that the convergence of any estimator to the true value of a smooth function defined on a space of high dimension is very slow. In fact, the number of simulations m required to attain a specified amount of accuracy grows at least exponentially as n increases. Thus, one should estimate the density of the vector $\mathbf{X}$ via estimating the densities of independent components of $\mathbf{X}$, if present. One may also resort to other forms of data reduction, such as principal components and projection pursuit, see (Scott, 1992, Section 7.2). Finally, one may take the "curse of dimensionality" into account in the choice of the observation scheme.

Stochastic approximation By evaluating the density estimator $\hat{f}(\mathbf{x})$ for $\mathbf{x} = \mathbf{X}$, we obtain an approximation

$$\hat{L}(\vartheta) = \hat{f}(\mathbf{X}) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{\mathbf{X}^{(j)} - \mathbf{X}}{b}\right).$$

for the likelihood function $L(\vartheta)$. We will refer to this approximation as the *simulated* likelihood. The method of maximum likelihood dictates that the parameters of the white box model should be estimated by maximizing $L(\vartheta)$. Unfortunately, we are unable to evaluate $L(\vartheta)$, and we have to make use of $\hat{L}(\vartheta)$, a noisy version of $L(\vartheta)$.

This kind of optimization problem can be handled in various ways, which differ in the way the gradient of $L(\vartheta)$ is estimated. We shall focus on finite differences (FD) and simultaneous perturbation (SP) techniques, since they only make use of the input and output of the underlying model.

In the two-sided FD approximation the ith component of the estimated gradient $\hat{\mathbf{g}}(\vartheta)$ is given by

$$\hat{g}_i(\vartheta) = \frac{\hat{L}(\vartheta + ce_i) - \hat{L}(\vartheta - ce_i)}{2c}, \ i = 1, \ldots, k, \tag{2}$$

where e_i a unit vector in the ith direction and $c > 0$ is sufficiently small. Approximating the gradient in this way requires $2k$ evaluations of the simulated likelihood.

The SP method for estimating the gradient was proposed in Spall (1987), and only requires a constant number of evaluations of the simulated likelihood. The idea is to perturb the parameters in separate random directions.

In the two-sided SP approximation the ith component of the estimated gradient $\hat{\mathbf{g}}(\vartheta)$ is given by

$$\hat{g}_i(\vartheta) = \frac{\hat{L}(\vartheta + c\boldsymbol{\Delta}) - \hat{L}(\vartheta - c\boldsymbol{\Delta})}{2c\Delta_i}, \ i = 1, \ldots, k, \tag{3}$$

where $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_k)^T$ is a vector of user-specified random variables satisfying certain conditions and $c > 0$ is a constant. Since the numerator in (3) is the same for each component i, only two function evaluations are required to obtain one SP gradient, that is, it does not depend on k.

The stochastic approximation algorithm The estimated gradient can be used in a steepest ascent procedure, where in each step additional simulation runs are performed. This method is generally referred to as stochastic approximation. We shall focus on stochastic approximation where the gradient is either estimated by means of finite differences (FDSA) or simultaneous perturbation (SPSA).

Stochastic approximation builds upon the iteration formula $\hat{\theta}^{[\ell+1]} = \hat{\theta}^{[\ell]} + a_\ell \hat{\mathbf{g}}(\hat{\theta}^{[\ell]})$, where $\hat{\theta}^{[\ell]} \in \mathbb{R}^k$ represents the estimate of θ after the ℓth iteration, $a_\ell > 0$ is a coefficient and $\hat{\mathbf{g}}(\hat{\theta}^{[\ell]})$ is the approximation of the gradient in $\hat{\theta}^{[\ell]}$. The constant c defined in equation (3) may depend on the iteration number ℓ. Now, under conditions on the numbers a_ℓ and c_ℓ and the likelihood, $\hat{\theta}^{[\ell]}$ will converge almost surely to the "true value" θ, for instance see Spall (2003).

In Kleinman *et al.* (1999) it is shown that using the same sequence of random numbers for the simulation runs required to compute $\hat{L}(\vartheta + c\boldsymbol{\Delta})$ and $\hat{L}(\vartheta - c\boldsymbol{\Delta})$ has a positive effect of the rate of convergence of both the FDSA and SPSA procedures. There is some evidence that the common random number version of SPSA should be preferred over its FDSA counterpart.

The scaled stochastic approximation algorithm described by Andradottir (1996) may be used to prevent the algorithm to diverge when the estimated gradient is either very steep or very flat. Finally, parameter transformations or projection based versions of the SA algorithm may be used to deal with parameter restrictions, for instance see Sadegh (1997).

3 Application to steel corrosion

3.1 Data

As an illustration, we consider the Haringvliet corrosion data presented in Heutink *et al.* (2004). The Dutch Haringvliet storm surge barrier was built in 1970, and closes the sea arm Haringvliet off from the sea. In the period from 1988 to 1996, the original coatings were replaced by new ones. In 2002, half of these coatings were inspected resulting in different inspection intervals. In Table 1 the percentage of the seaside surface of a steel gate that has been corroded due to ageing of the coating is given for five steel gates of the Haringvliet barrier. Note that $n = 5$. In Nicolai and Koning (2006) a more extensive analysis of the Haringvliet data is presented.

3.2 A white box model

To model the corrosion process of a steel gate as a discrete event system, we partition the rectangular surface of the steel gate in small square areas, which we will refer to as sites. Each of these sites can be in two states; initially, the square is "uncorroded", but after some time it may become corroded. Once the site is corroded, it cannot return to the uncorroded state. This discrete event system is described by a marked point process which records every time instance at which a site becomes corroded, and assigns to this time instance the location of the site as a mark. Thus, the mark space $\mathcal{S}$ is in fact a product of a finite "horizontal" space $\mathcal{S}_x$ and a finite "vertical" space $\mathcal{S}_y$. It follows from (1) that we may specify the discrete event system by formulating the structure of each site-specific intensity rate function λ_s.

As a site s cannot return to the uncorroded state once the site is corroded, the corresponding counting process N_s can only jump once. In other words, the counting process is randomly stopped after the first jump. We may incorporate this into the model by including a factor $Y_s(t) = 1 - N_s(t-)$ in the intensity rate function λ_s. Here $N_s(t-)$ is shorthand notation for $\lim_{\epsilon \downarrow 0} N_s(t - \epsilon)$. In statistical modelling, the process Y_s is usually referred to as the "number at risk", see (Andersen *et al.*, 1993, p. 128). Observe that the process Y_s is predictable.

Next, we shall assume that corrosion is the interplay of two physical processes: the initiation process and the propagation process.

In the initiation process the surface is constantly threatened by attacks.

If an uncorroded site is hit, the site may or may not become corroded. However, as time progresses, the uncorroded sites become more and more vulnerable. We assume that the initiation process may be modelled by a (randomly stopped) inhomogeneous Poisson process with an intensity proportional to some power of t. That is, the initiation process contributes a term $\lambda_s^{[\text{initiation}]}(t) = q\nu t^{q-1} Y_s(t)$ to $\lambda_s(t)$.

In the propagation process, an uncorroded site s may become "infected" by neighbouring corroded sites, that is, by corroded sites in its neighbourhood $\mathcal{B}_s$. We assume that $\mathcal{B}_s$ is as in Figure 1. The propagation process contributes a term $\lambda_s^{[\text{propagation}]}(t) = \delta z_s(t) Y_s(t)$ to $\lambda_s(t)$, where $z_s(t) = \sum_{s' \in \mathcal{B}_s} N_{s'}(t-)$ counts the number of corroded sites in $\mathcal{B}_s$ just before time t. It follows that

$$\lambda_s(t) = \lambda_s^{[\text{initiation}]}(t) + \lambda_s^{[\text{propagation}]}(t) = \left\{ q\nu t^{q-1} + \delta z_s(t) \right\} Y_s(t).$$

3.3 Results

In the previous subsection, we have formulated a model for the corrosion of a single steel gate. Assuming that the corrosion of one gate evolves independently of the corrosion of the other gates, this model is readily extended to the complete system of five steel gates. If we in addition assume that the five steel gates share the values of their parameters, we have $\theta = (\nu, q, \delta)^T$.

In order to approximate the likelihood function, the method of kernel density estimation is applied to the outcomes of $m = 1,000$ simulations runs of $\{N_s\}_{s \in \mathcal{S}}$, generated according to Algorithm 1. Each outcome consists of the percentage of corrosion at the five inspection times.

We have applied the stochastic approximation algorithm introduced in section 2.2 to find the maximum likelihood estimates of the parameters of the corrosion process. In particular, we used the SPSA approach in combination with common random numbers. Table 2 summarizes the results of the parameter estimation. The value of the maximized likelihood function is based on $m = 10,000$ simulation runs.

Figure 2 shows a realization of the corrosion process at time $t = 15$ years. The green pixels are uncorroded sites. The red pixels are corroded sites due to initiation, and the white pixels are corroded sites due to propagation.

4 Conclusion

In this article we have developed a general methodology for statistical analysis of white box simulation models by likelihood methods. The mathematical intractability of the white box model is circumvented by using computer intensive methods, including simulation, multivariate density estimation and stochastic approximation.

The objective in developing the methodology was complete generality. As a consequence, the methodology is indeed computer intensive, and may become very demanding on the available computing resources. For instance, due to the "curse of dimensionality", multivariate density estimation may well become a time consuming affair.

As reducing the computer intensiveness is a necessary condition for successful application to large problems, our methodology does not acquit the researcher from understanding the white box model at a detailed level. For instance, a cleverly devised observation scheme is instrumental in averting the "curse of dimensionality".

Also, knowledge of the model may make solving $\Lambda^*(t) = \Lambda^*(t_0) + E$ in step (2) of Algorithm 1 simpler. For instance, in Section 3 we could have made use of the fact that corrosion is in fact the interplay of two physical processes. Indeed, this is the way in which Figure 2 was simulated, hence the distinction between corroded sites due to initiation and corroded sites due to propagation.

References

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Andradottir, S. (1996), A scaled stochastic approximation algorithm, *Management Science*, **42(4)**, 475–498.

Çinlar, E., Z. P. Bazant, and E. Osman (1977), Stochastic process for extrapolating concrete creep, *Journal of the Engineering Mechanics Division*, **103(EM6)**, 1069–1088.

Daley, D. J. and D. Vere-Jones (2003), *An Introduction to the Theory of Point Processes. Vol. I*, second edn., Springer-Verlag, New York.

Doksum, K. A. and A. Hoyland (1992), Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution, *Technometrics*, **34**, 74–82.

Duong, T. and M. L. Hazelton (2005), Cross-validation bandwidth matrices for multivariate kernel density estimation, *Scandinavian Journal of Statistics*, **32**, 485–506.

Heutink, A., A. Van Beek, J. M. van Noortwijk, H. E. Klatter, and A. Barendregt (2004), Environment-friendly maintenance of protective paint systems at lowest costs, in *XXVII FATIPEC Congres; 19-21 April 2004, Aix-en-Provence*, Paris: AFTPVA.

Kleinman, N., J. C. Spall, and D. Q. Naiman (1999), Simulation-based optimization with stochastic approximation using common random numbers, *Management Science*, **45**, 1570–1578.

Nicolai, R. P., R. Dekker, and J. M. van Noortwijk (2006), A comparison of models for measurable deterioration: an application to coatings on steel structures, Tech. Rep. EI 2006-28, Econometric Institute (forthcoming in Reliability Engineering & System Safety).

Nicolai, R. P. and A. J. Koning (2006), Modelling and estimating degradation by simulating a physical process, Tech. Rep. EI 2006-46, Econometric Institute.

Sadegh, P. (1997), Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation, *Automatica*, **33(5)**, 889–892.

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons Inc., New York.

Sheather, S. J. (2004), Density estimation, *Statistical Science*, **19**, 588–597.

Sheather, S. J. and M. C. Jones (1991), A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society, Series B*, **53**, 683–690.

Sobczyk, K. (1987), Stochastic models for fatigue damage of materials, *Advances in Applied Probability*, **30(4)**, 652–673.

Spall, J. C. (1987), A stochastic approximation technique for generating maximum likelihood parameter estimates, in *Proceedings of the 1987 American Control Conference*, Minneapolis, MN.

Spall, J. C. (2003), *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley-Interscience, Hoboken, NJ.

Wand, M. P. and M. C. Jones (1995), *Kernel Smoothing*, Chapman and Hall Ltd., London.

Age (years)	**Corrosion (%)**
6	0.27
8	0.41
10	0.84
12	0.75
14	2.10

Table 1: Inspection results for the seaside surface of five steel gates of the Haringvliet storm surge barrier.

Parameter	**Estimate**
ν	2.01E-04
δ	7.88E-02
q	1.02

Likelihood	**Value**
$\hat{L}(\hat{\nu}, \hat{\delta}, \hat{q})$	4.25

Table 2: Estimation results for the degradation of the seaside surface of five steel gates of the Haringvliet storm surge barrier.

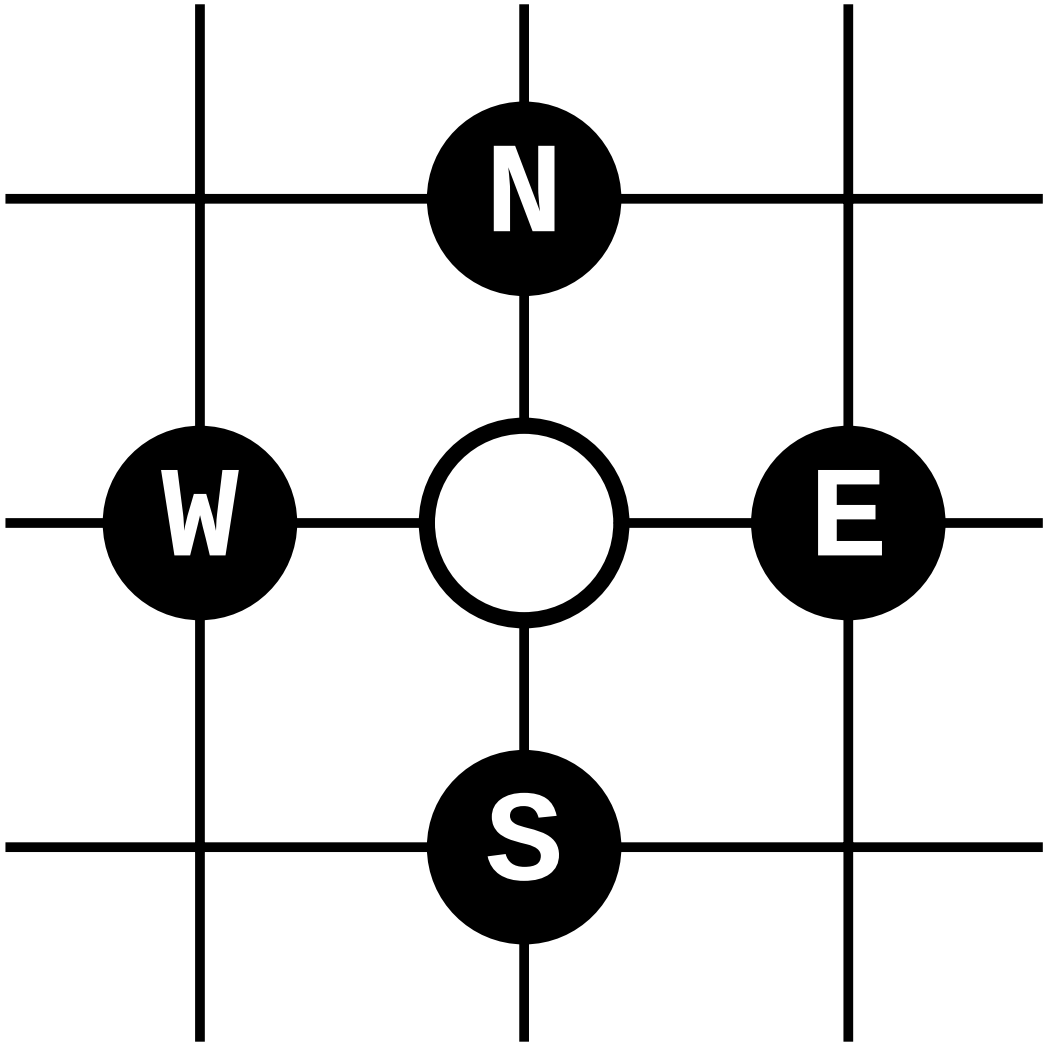

Figure 1: Example of a neighbourhood in a two-dimensional grid. The black dots N, W, S and E together form the neighbourhood of the white dot.

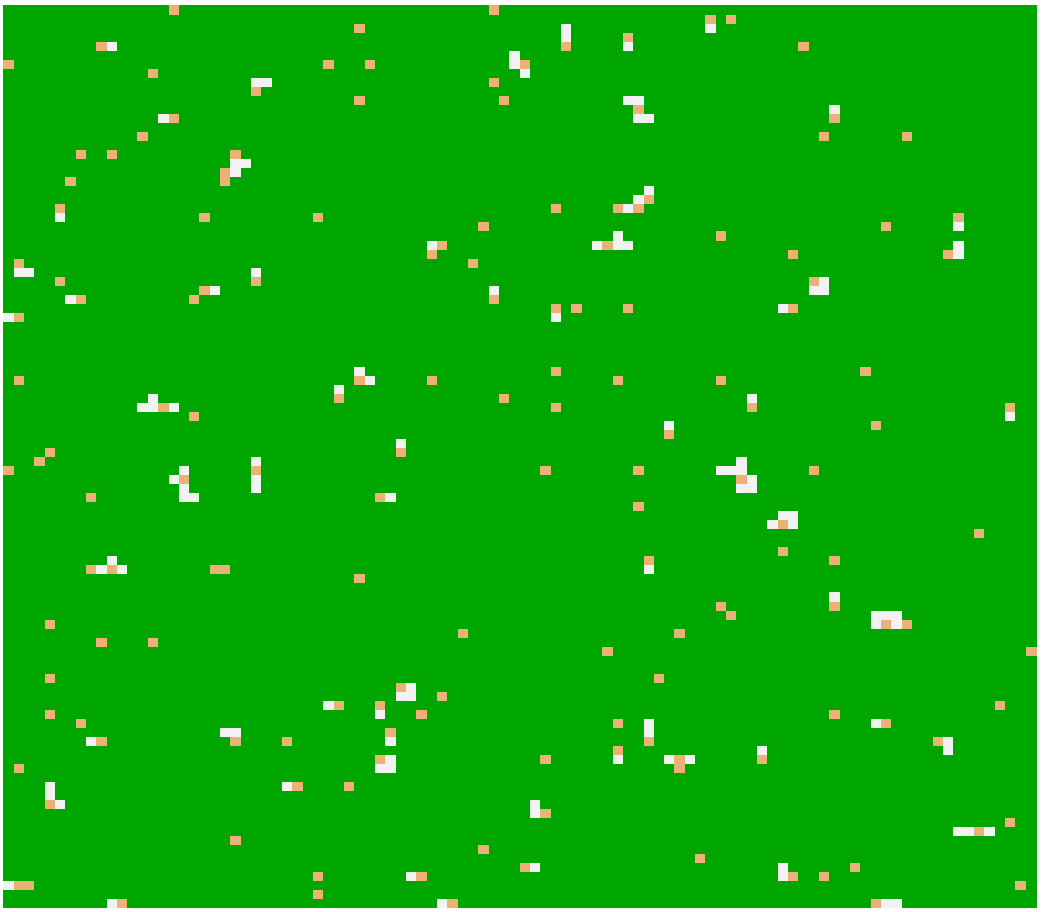

Figure 2: Simulated corrosion at $t = 15$ years.