Journal of the Operational Research Society (2003) 54, 255–262

www.palgrave-journals.com/jors

Kriging for interpolation in random simulation

WCM van Beers*[1] and JPC Kleijnen[2]

[1]*Department of Information Systems and Management, Tilburg University (UVT), Tilburg, The Netherlands; and*
[2]*Department of Information Systems and Management/Center for Economic Research (CentER), Tilburg University (UVT), Tilburg, The Netherlands*

Whenever simulation requires much computer time, interpolation is needed. Simulationists use different interpolation techniques (eg linear regression), but this paper focuses on Kriging. This technique was originally developed in geostatistics by DG Krige, and has recently been widely applied in deterministic simulation. This paper, however, focuses on random or stochastic simulation. Essentially, Kriging gives more weight to 'neighbouring' observations. There are several types of Kriging; this paper discusses—besides Ordinary Kriging—a novel type, which 'detrends' data through the use of linear regression. Results are presented for two examples of input/output behaviour of the underlying random simulation model: Ordinary and Detrended Kriging give quite acceptable predictions; traditional linear regression gives the worst results.
Journal of the Operational Research Society (2003) **54,** 255–262. doi:10.1057/palgrave.jors.2601492

Keywords: simulation; statistics; stochastic; regression; methodology

Introduction

A primary goal of simulation is *what if* or sensitivity analysis: What happens if inputs of the simulation model change? Therefore, simulationists run a given simulation program—or computer code—for (say) n different combinations of the k simulation inputs. We assume that these inputs are either parameters or quantitative input variables of the simulation model. Typically, Kriging assumes that the number of values per input variable is quite 'big', certainly exceeding two (two values are used in simulation experiments based on 2^{k-p} designs).

Given this set of n input combinations, the analysts run the simulation and observe the outputs. (Most simulation models have multiple outputs, but in practice these outputs are analysed *per* output type.)

The crucial question of this paper is: How to *analyse* this simulation input/output (I/O) data? Classic analysis uses linear-regression (meta) models; see Kleijnen.[1] A *metamodel* is an approximation of the I/O transformation implied by the underlying simulation program. (Many other terms are popular in certain disciplines: response surface, compact model, emulator, etc.) Such a metamodel treats the simulation model as a *black box*; that is, the simulation model's I/O is observed, and the parameters of the metamodel are estimated. This black-box approach has the following advantages and disadvantages.

**Correspondence: WCM van Beers, Department of Information Systems and Management, Tilburg University (UVT), PO Box 90153, 5000 LE Tilburg, The Netherlands.*
E-mail: wvbeers@uvt.nl
Web: http://www.tilburguniversity.nl/faculties/few/im/staff/kleijnen/

An *advantage* is that the metamodel can be applied to all types of simulation models, either deterministic or random, either in steady state or in transient state. A *disadvantage* is that it cannot take advantage of the specific structure of a given simulation model, so it may take more computer time compared with techniques such as perturbation analysis and score function.

Metamodelling can also help in optimization and validation of the simulation model. In this paper, however, we do not discuss these two topics, but refer to the references of this paper. Further, if the simulation model has hundreds of inputs, then special 'screening' designs are needed, discussed in Campolongo *et al*.[2] In our examples—but not in our methodological discussion—we limit the number of inputs to the minimum, namely a single input.

Whereas polynomial-regression metamodels have been applied extensively in discrete-event simulation (such as queueing simulation), *Kriging has hardly been applied to random simulation*: a search of International Abstracts of Operations Research (IAOR) gave only two hits. However, in deterministic simulation (applied in many engineering disciplines; see our references), Kriging has been applied frequently, since the pioneering article by Sacks *et al*.[3] In such simulation, Kriging is attractive because it can ensure that the metamodel's prediction has exactly the same value as the observed simulation output (as we shall see below). In random simulation, however, this Kriging property may not be so desirable, since the observed (average) value is only an estimate of the true, expected simulation output. Unfortunately, Kriging requires extensive computation, so adequate software is needed. We discovered that for random simulation no software is

available, so we developed our own software, in *Matlab*.

Note that several types of *random simulation* may be distinguished:

(i) *Deterministic simulation with randomly sampled inputs*. For example, in investment analysis we can compute the cashflow development over time through a spreadsheet such as Excel. Next, we sample the random values of inputs—such as the cashflow growth rate—by means of either Monte Carlo or Latin Hypercube Sampling (LHS) through an add-on such as @Risk or Crystal Ball; see Groenendaal *et al*.[4]

(ii) *Discrete-event simulation*. For example, classic queueing simulation is applied in logistics and telecommunications.

(iii) *Combined continuous/discrete-event simulation*. For example, simulation of nuclear waste disposal represents the physical and chemical processes through deterministic nonlinear difference equations and models the human interventions as discrete events.[5]

Our research contribution consists in the development of a novel (namely, detrended) Kriging type, and the exploration of how well this Kriging type performs compared with Ordinary Kriging and traditional polynomial-regression modelling. The main conclusion of our examples is: a perfectly specified detrending function gives best predictions; Ordinary Kriging is acceptable; the usual linear regression gives the worst results.

We organize the remainder of this paper as follows. First, we sketch the history of Kriging and its application in geology, metereology, and deterministic simulation. Then we describe the basics of Kriging, and give a formal Kriging model. Next, we introduce our novel model for detrending the I/O data through low-order polynomial regression, including a classic cross-validation test. We illustrate this Kriging through two simple examples. In a separate section, we give a third random simulation example to study the so-called nugget effect in Kriging. Finally, we present the conclusions and mention possible future research topics.

Kriging

History of Kriging

Kriging is an interpolation technique originally developed by DG Krige, a South African mining engineer. In the 1950s he devised this method to determine true ore-grades, based on samples. Next, he improved the method in cooperation with G Matheron, a French mathematician at the 'Ecole des Mines'. At the same time, in meteorology L Gandin (in the former Soviet Union) worked on similar ideas, under the name 'optimum interpolation'.[6]

Nowadays, Kriging is also applied to I/O data of *deterministic simulation* models; we refer again to Sacks *et al*'s[3] pioneering article. Many more publications followed; for example, Meckesheim *et al*[7] give 35 references. Also see Koehler and Owen,[8] and Jones *et al*.[9]

Basics of Kriging

Kriging is an approximation method that can give predictions of unknown values of a random function, random field, or random process. These predictions are *best linear unbiased estimators*, under the Kriging assumptions presented in the next subsection.

Actually, these predictions are *weighted* linear combinations of the observed values. Kriging assumes that *the closer the input data, the more positively correlated the prediction errors*. Mathematically, this assumption is modelled through a *second-order stationary covariance process*: the expectations of the observations are constant and do not depend on the location (the input values), and the covariances of the observations depend only on the 'distances' between the corresponding inputs. In fact, these covariances decrease with the distance between the observations. The prediction criterion is minimum mean-squared prediction errors. The result is an estimated metamodel such that observations closer to the prediction point get more weight in the predictor. When predicting the output for a location that has already been observed, the prediction equals the observed value. (In deterministic simulation this property is certainly attractive, as mentioned above.)

In Kriging, a crucial role is played by the *variogram*, that is, a diagram of the variance of the difference between the measurements at two input locations (also see Figure 1), which has symbols explained in the next section. The assumption of a second-order stationary covariance process implies that the variogram is a function of the distance (say) h between two locations. Moreover, the further apart the two inputs, the smaller this dependence—until the effect is negligible.

Formal model for Kriging

A *random process* $Z(\cdot)$ can be described by $\{Z(\boldsymbol{s})\colon \boldsymbol{s} \in D\}$ where D is a fixed subset of R^d and $Z(\boldsymbol{s})$ is a random function at location $\boldsymbol{s} \in D$; see Cressie[6] (p 52).

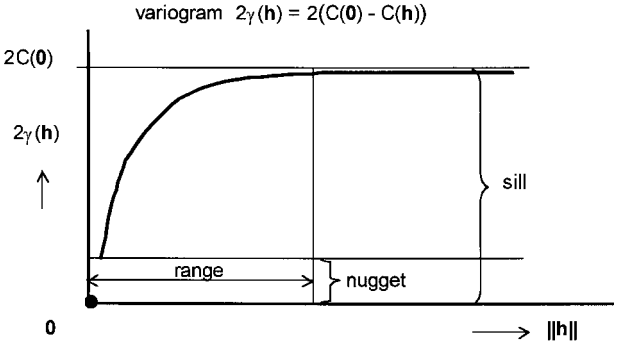

Figure 1 An example variogram.

There are several types of Kriging, but we limit this section to *Ordinary Kriging*, which gives the following two assumptions (already mentioned above, but not yet formalized):

(i) The *model assumption* is that the random process consists of a constant μ and an error term $\delta(\boldsymbol{s})$:

$$\boldsymbol{Z}(\boldsymbol{s}) = \mu + \delta(\boldsymbol{s}) \text{ with } \boldsymbol{s} \in D, \ \mu \in R \tag{1}$$

(ii) The *predictor assumption* is that the predictor for the point $\boldsymbol{s}_0$—denoted by $p(Z(\boldsymbol{s}_0))$—is a weighted linear function of all the observed output data:

$$p(Z(\boldsymbol{s}_0)) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{s}_i) \text{ with } \sum_{i=1}^{n} \lambda_i = 1 \tag{2}$$

To select the weights λ_i in (2), the *criterion* is minimal mean-squared prediction error (say) σ_e^2 defined as

$$\sigma_e^2 \equiv E[\{Z(\boldsymbol{s}_0) - p(Z(\boldsymbol{s}_0))\}^2] \tag{3}$$

To minimize (3) given (2), let m be the Lagrangian multiplier ensuring $\sum_{i=1}^{n} \lambda_i = 1$. Then we can write the prediction error as

$$E[Z(\boldsymbol{s}_0) - \sum_{i=1}^{n} \lambda_i \cdot Z(\boldsymbol{s}_i)]^2 - 2m\left[\sum_{i=1}^{n} \lambda_i - 1\right] \tag{4}$$

To minimize (4), we utilize the *variogram*; also see Figure 1. By definition, the variogram is $2\gamma(\boldsymbol{h}) = var[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})]$, where $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$ as explained by the stationary covariance process assumption with $\boldsymbol{h} \in R^d$ and $i,j = 1,\ldots,n$. Obviously, we have $var[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})] = 2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = 2\gamma(\boldsymbol{h})$. The spacing $\boldsymbol{h}$ is also called the lag.

After some tedious manipulations, (4) gives

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\sum_{i=1}^{n} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - 2m\left(\sum_{i=1}^{n} \lambda_i - 1\right) \tag{5}$$

Differentiating (5) with respect to $\lambda_1,\ldots,\lambda_n$ and m gives the optimal $\lambda_1,\ldots,\lambda_n$:

$$\boldsymbol{\lambda}' = \left(\boldsymbol{\gamma} + \mathbf{1}\frac{1 - \mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma}}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}\right)' \Gamma^{-1} \text{ and}$$

$$m = -(1 - \mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma})/(\mathbf{1}'\Gamma^{-1}\mathbf{1}) \tag{6}$$

where $\boldsymbol{\gamma}$ denotes the vector $(\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_1),\ldots,\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_n))'$, $\boldsymbol{\Gamma}$ denotes the $n \times n$ matrix whose (i, j)th element is $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)$, and $\mathbf{1}$ denotes a vector of ones; also see Cressie[6] (p 122).

We emphasize that these optimal Kriging weights λ_i depend on the specific point $\boldsymbol{s}_0$ that is to be predicted, whereas linear-regression metamodels use fixed estimated parameters (say) $\hat{\boldsymbol{\beta}}$.

The optimal weights (6) give the minimal mean-squared prediction error: (3) becomes (also see Cressie[6] (p 122))

$$\begin{aligned}\sigma_e^2 &= \sum_{i=1}^{n} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) + m \\ &= \boldsymbol{\gamma}'\Gamma^{-1}\boldsymbol{\gamma} - \frac{(\mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma} - 1)^2}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}\end{aligned} \tag{7}$$

However, in (6) and (7) $\gamma(\boldsymbol{h})$ is *unknown.* The usual *estimator* is

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2 \tag{8}$$

where $|N(\boldsymbol{h})|$ denotes the number of distinct pairs in $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}; i, j = 1,\ldots, n\}$; see Matheron.[10] The estimator in (8) is unbiased if the process $Z(\cdot)$ is indeed second-order stationary; see Cressie[6] (p 71).

Given (8) for different $||\boldsymbol{h}||$ values, the variogram is estimated by *fitting* a curve through the estimated values $2\hat{\gamma}(\boldsymbol{h})$. This curve displays the following important *characteristics* (see Figure 1):

(i) For large values of $||\boldsymbol{h}||$, the variogram $2\hat{\gamma}(\boldsymbol{h})$ approaches a constant $C(\mathbf{0})$, called the *sill*: for these large $||\boldsymbol{h}||$ values, all variances of the differences $Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})$ are invariant with respect to $\boldsymbol{h}$.

To prove this property, we define the *covariogram* $C(\boldsymbol{h}) = Cov(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$. Obviously, $Cov(Z(\boldsymbol{s}), Z(\boldsymbol{s})) = Var(Z(\boldsymbol{s}))$. Then it is easy to derive

$$2\gamma(\boldsymbol{h}) = 2(C(\mathbf{0}) - C(\boldsymbol{h})) \tag{9}$$

Since $C(\boldsymbol{h}) \downarrow 0$ as $||\boldsymbol{h}|| \uparrow \infty$, the variogram has the upper limit $2C(\mathbf{0})$.

(ii) The interval of $||\boldsymbol{h}||$ on which the curve does increase (to the sill) is called the *range* (say) r; that is, $C(\boldsymbol{h}) < \varepsilon$ for $||\boldsymbol{h}|| > r + r_\varepsilon$. We shall give a specific model in (10).

(iii) Although (9) implies $\gamma(\mathbf{0}) = 0$, the fitted curve does not always pass through zero: it may have a positive intercept called the *nugget* variance. This variance estimates noise. For example, in geostatistics this nugget effect means that when going back to the 'same' spot, a completely different output (namely, a gold nugget) is observed.

We add that in *random* simulation, the same input (say, the same traffic rate in queueing simulation) gives different outputs because different pseudorandom numbers are used. Below we shall return to this issue.

To fit a variogram curve through the estimates resulting from (8), analysts usually apply the *exponential model*

$$\gamma(\boldsymbol{h}) = \begin{cases} c_0 + c_1(1 - e^{-||\boldsymbol{h}||/a}) & \text{if } \boldsymbol{h} \neq \mathbf{0} \\ 0 & \text{if } \boldsymbol{h} = \mathbf{0} \end{cases} \tag{10}$$

where obviously c_0 is the nugget, $c_0 + c_1$ the sill, and a the range. However, other models are also fitted; for example,

the *linear model*

$$\gamma(\boldsymbol{h}) = \begin{cases} c_0 + b||\boldsymbol{h}|| & \text{if } \boldsymbol{h} \neq \boldsymbol{0} \\ 0 & \text{if } \boldsymbol{h} = \boldsymbol{0} \end{cases} \quad (11)$$

where again c_0 is the nugget; see Cressie[6] (p 61). Actually, we shall apply (11) in our experiments, because it is the simplest model and yet gives acceptable results (eg it estimates the nugget effect very well).

In *deterministic simulation*, analysts use more general distance formulas than (8). For example, Sacks *et al*[3] (p 413) and Jones *et al*,[9] (p 5) use the weighted distance formula

$$h(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{g=1}^{k} \theta_g |x_{i(g)} - x_{j(g)}|^{p_g} \quad (12)$$

where θ_g (with $\theta_g \geq 0$) measures the importance of the input x_g, and p_g controls the smoothness of the distance function. To estimate θ_g, maximum likelihood estimation (MLE) is used. The p_g are fixed such that $0 < p_g \leq 2$. (We shall briefly return to (12) in the section Conclusions and future research.)

Detrended Kriging

Ordinary Kriging was defined by (1), where $\mu \in R$ was the *constant* mean of the random process $Z(\cdot)$. This assumption, however, limits the application of Ordinary Kriging to rather simple models of the process $Z(\cdot)$. A more general assumption is that μ is not a constant, but an unknown linear combination of known functions $\{f_0(\boldsymbol{s}), \ldots, f_n(\boldsymbol{s})\}, \boldsymbol{s} \in D$. This is called *Universal Kriging*; see Huijbregts and Matheron[11] (p 160) and also Cressie[6] (p 151). Cressie[6] discusses real (nonsimulated) coal-ash data, and Régnière and Sharov[12] discuss simulated spatial and temporal output data of a random simulation model for ecological processes.

Now we introduce a novel type of Kriging that we call *Detrended Kriging*. Detrended Kriging preprocesses the original data, and then applies Ordinary Kriging to the resulting data so we can apply software for Ordinary Kriging. For Universal Kriging, however, software is available only for spatial and temporal data, not for simulation with an arbitrary number of inputs—to the best of our knowledge.

We assume that the process mean $\mu(\boldsymbol{s})$ satisfies the decomposition

$$\mu(\boldsymbol{s}) = S(\boldsymbol{s}) + \eta(\boldsymbol{s}) \quad (13)$$

where $S(\boldsymbol{s})$ is a known signal function (see, however, the text below (14)) and $\eta(\boldsymbol{s})$is a *white-noise* process that models the measurement error; that is, $\eta(\boldsymbol{s})$ is normally identically and independently distributed with zero mean (NIID). So, we replace (1) by

$$Z(\boldsymbol{s}) = S(\boldsymbol{s}) + \eta(\boldsymbol{s}) + \delta(\boldsymbol{s}) \quad (14)$$

In practice, the signal function $S(\boldsymbol{s})$ in (14) is unknown. Therefore, we estimate $S(\boldsymbol{s})$ through $\hat{S}(\boldsymbol{s})$, from the set of observed (noisy) I/O data $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_j)): i = 1, \ldots, n\}$. Owing to the assumed white noise, we use ordinary least squares (OLS) to obtain the estimator $\hat{S}(\boldsymbol{s})$.

Next, we apply Ordinary Kriging to the detrended set $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}) - \hat{S}(\boldsymbol{s}_i)): i = 1, \ldots, n\}$. Our predictor for the output of location $\boldsymbol{s}_0$ is the sum of this Ordinary Kriging prediction and the estimator $\hat{S}(\boldsymbol{s}_0)$.

To test our new Detrended Kriging, we apply classic *cross-validation*; see Kleijnen and van Groenendaal[13] (p 156). Cross-validation eliminates one I/O combination, say $(\boldsymbol{s}_k, Z(\boldsymbol{s}_k))$, from to the original data set $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i) - \hat{S}(\boldsymbol{s}_i)): i = 1, \ldots, n\}$, so the remaining $n-1$ data combinations are $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i): i = 1, \ldots, k-1, k+1, \ldots, n\}$. This new set gives a prediction $p(Z(\boldsymbol{s}_k))$. This process of elimination and prediction is repeated for (say) c different combinations ($c \leq n$). Obviously, if we sort the original set such that the first c observations are deleted one at a time, then we get $k = 1, 2, \ldots, c$.

To summarize the resulting prediction accuracy, we use the L_2 norm of the difference vector $||P(Z(\boldsymbol{s}_k)) - Z(\boldsymbol{s}_k)||$ (the L_2 norm $||\boldsymbol{x}||$ is defined as $\left(\sum_{k=1}^{c} x_k^2\right)^{1/2}$). In our experiments we find that the L_1 and L_∞ norms give similar conclusions.

Note that in Kriging, all prediction errors may be zero at the I/O points that are actually used to estimate the Kriging model. Therefore we use cross-validation.

Two examples and five metamodels

We are interested in the application of Kriging to discrete-event simulation models, such as simulated queueing systems. As Law and Kelton[14]—the best-selling textbook on simulation—state (on p. 12), a *single server* queueing system is quite representative of more complex, dynamic, stochastic simulation models. For further simplification, we suppose that the output of interest is the mean waiting time in the steady state, $E(W)$. This output can be estimated through a simulation that uses the following nonlinear stochastic difference equation:

$$W(i) = \max(W(i-1) + S(i-1) - A(i), 0) \quad \text{with} \quad i = 1, 2, \ldots \quad (15)$$

where $W(i)$ denotes the waiting time of customer i, $S(i)$ the service time of customer i, and $A(i)$ the interarrival time between customers i and $i-1$. It is standard to start the simulation run in the empty (idle) state: $W(0) = 0$. For additional simplification, we assume that the arrival times form a Poisson process, and so do the service times. This gives the well-known *M/M/1* (which can actually be solved analytically; see Equation (18) below). By definition, M/M/1 implies that both $S(i)$ and $A(i)$ are identically and independently distributed (IID), so simulation is straightforward. The output $E(W)$ is usually estimated through the

simulation run's *average*

$$\overline{W} = \sum_{i=b}^{n} \frac{W(i)}{n-b+1} \qquad \text{with } 0 \leq b < n \tag{16}$$

where b denotes the length of the initialization (start-up, transient) phase (which may be zero) and n is the run length. (In M/M/1 analysis and simulation through renewal analysis, this initialization is no issue. In practical simulations, however, it is a major problem; see Law and Kelton[14] (pp 496–552).) In other words, the dynamic simulation model generates the time series (15), but this series is characterized through the single statistic (16).

Actually, simulation is done for sensitivity analysis (possibly followed by optimization). Such an analysis aims at estimating the *I/O function* (say)

$$E(Z(\boldsymbol{s})) = S(\boldsymbol{s}) \tag{17}$$

where, following (14), Z denotes the (multiple) output and $\boldsymbol{s}$ the (multiple) input. In the M/M/1 example, we have $Z = \overline{W}$ and $s = \lambda/\mu$ with arrival rate λ and service rate μ.

In general, $S(\boldsymbol{s})$ in (17) has unknown shape and parameters. However, when studying the performance of a specific simulation methodology, researchers often use the M/M/1 simulation model because the I/O function $S(\boldsymbol{s})$ is then known — assuming that the methodology has selected an appropriate initialization length b in (16) (obviously, knowledge of $S(\boldsymbol{s})$ may not be used by the methodology itself):

$$E(W) = \frac{\lambda}{\mu(\mu-\lambda)} \qquad \text{with } \frac{\lambda}{\mu} < 1 \tag{18}$$

Unfortunately, the latter assumption is very questionable: it is well known that selecting an appropriate transient-phase length b and run length n in (16) is difficult.

Moreover, Kriging assumes that, in general, the simulation observations Z have additive *white noise*; see (14). In the M/M/1 example, (14) gives the following: (i) $Z = \overline{W}$; (ii) $S(\boldsymbol{s}) = \lambda/(\mu(1-\lambda))$; (iii) normality holds if $\overline{W}$ in (16) uses a sample size $(n-b+1)$ such that an asymptotic central limit theorem holds; (iv) constant variances result if different simulated traffic rates use different and appropriate sample sizes—see Kleijnen and van Groenendaal;[15] and (v) independence results if no common pseudorandom numbers are used. Altogether, Kriging's (1) or (14) applies to the M/M/1 example only if a slew of assumptions hold!

Hence, it is much more efficient and effective to generate Kriging test data through sampling from (13) with $S(\boldsymbol{s}) = \lambda/(\mu(1-\lambda))$ instead of (15) and (16). Indeed, our approach requires less computer time, and guarantees that the white noise assumption holds, including the desired value for the variance of the white noise. The alternative using (15) and (16) would require very long runs; especially for high traffic rates $\lambda/\mu \uparrow 1$ this alternative requires $n \uparrow \infty$.

In conclusion, to test the Kriging methodology we generate data through a static, random Monte Carlo model like (13) instead of a dynamic stochastic simulation model such as (15) combined with (16). Therefore, the Monte Carlo technique is both efficient and effective.

Besides Example I, we study Example II representing simulations with *multiple local maxima*, which are interesting when optimizing simulation outputs. Example I represents queueing simulations that show 'explosive' mean waiting times as the traffic rate approaches the value one. Example II has no specific interpretation.

We sample the white-noise term $\eta(\boldsymbol{s})$ in (14) through the Matlab function called 'randn', which gives standard NIID variates; that is, $\eta(\boldsymbol{s})$ has zero mean and unit variance. We also experiment with a larger variance, namely 25; this results in larger error terms, but not in other conclusions.

To estimate possible values of the L_2 norm (defined above), we use 100 *macroreplications*. From these macroreplications we estimate L_2's median, 0.10 quantile $Q_{0.1}$, and 0.90 quantile $Q_{0.9}$.

In both examples we take $n = 21$ equally spaced input values: s_i with $i = 1,\ldots,21$. For cross-validation, we select (rather arbitrarily) $c = 5$ input values: We eliminate $i = 2$, 8, 9, 15, 16, respectively. We compared the following five metamodels:

(i) Ordinary Kriging;
(ii) second-degree detrending: $\hat{S}(\boldsymbol{s})$ is a second-degree polynomial;
(iii) perfectly specified detrending function: $\hat{S}(\boldsymbol{s})$ is a hyperbola in Example I, and a fourth-degree polynomial in Example II;
(iv) fifth-degree detrending: $\hat{S}(\boldsymbol{s})$ is a fifth-degree polynomial (overfitting);
(v) linear regression model, that is a second-degree polynomial estimated through OLS.

Note that we also study a perfectly specified detrending function (see model iii), because this function provides a utopian situation. This function gives the mimimum predicion error; in practice, it is unknown (otherwise simulation would not be used). Furthermore, it helps us verify the correctness of our computer program.

Example I: M/M/1 hyperbola

We take $S(s) = s/(1-s)$ on $D = [0.01, 0.99] \subset R^1$; this hyperbolic function may represent the mean steady-state waiting time for a traffic rate s in an M/M/1 queueing system. This function gives Figure 2, which also displays an example of the noisy output $Z(\boldsymbol{s})$. The input locations are $s_i \in \{0.01, 0.05, 0.1, \ldots, 0.95, 0.99\}$. The cross-validation is carried out at $s = 0.05$, 0.35, 0.40, 0.70, 0.75.

The estimated distribution of the prediction accuracy L_2 is summarized in Table 1. This example suggests that metamodel iii (perfectly specified detrending function) gives

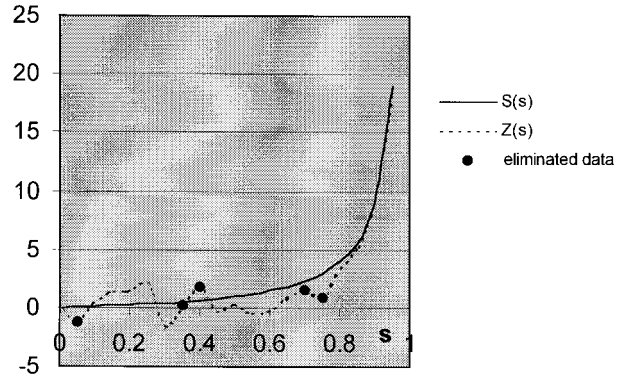

Figure 2 $S(s) = s/(1–s)$ and example sample output $Z(s)$.

Table 1 Prediction accuracy: estimated quantiles of L_2 distribution for Example I

	Metamodel				
L_2 *norm*	*i*	*ii*	*iii*	*iv*	*v*
$Q_{0.1}$	1.2429	1.3622	1.1972	2.7411	17.5931
Median	1.8832	2.1522	1.8419	3.6678	18.1702
$Q_{0.9}$	2.5698	2.9250	2.5829	4.4677	18.6524

the best results. Model i (Ordinary Kriging) is not too bad. Model v (OLS) is simply bad.

Example II: fourth-degree polynomial

We take the following specific polynomial: $S(s) = -0.0579s^4 + 1.11s^3 - 6.845s^2 + 14.1071s + 2$ on $D = [0, 10] \subset R^1$. This polynomial has two maxima: a local one and a global one (see Figure 3). We obtain output for the following 21 input locations $s_i \in \{0, 0.5, 1, \ldots, 10\}$. We cross-validate at $s = 0.5$, 3.5, 4, 7, and 7.5.

The estimated distribution of L_2 is summarized in Table 2. This example suggests that metamodel iii (perfectly specified detrending function) gives the best results. Model i (Ordinary Kriging) is not too bad. Model v (OLS) is simply bad.

Third example and nugget effect

We also wish to better understand the relationship between the nugget effect in (11) and the variance of the noise $\eta(\mathbf{s})$ in (13). Therefore, we perform a simple Monte Carlo experiment. We take $Z(s) = 10 + \eta(s)$, where $\eta(s)$ is NIID with $\mu = 0$ and $\sigma^2 = 1$, 4, 9, 16, and 25, respectively. We sample two macroreplicates, setting the seed of Matlab's 'randn'—rather arbitrary—to the values 10 and 20. In the various Kriging metamodels, we fit the linear variogram of (11); see

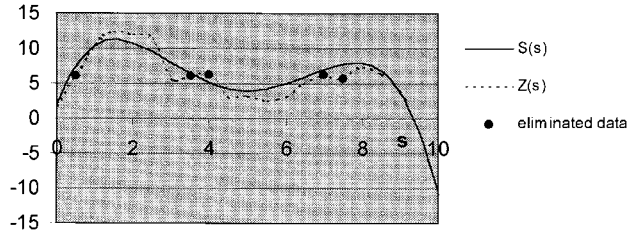

Figure 3 $S(s) = -0.0579s^4 + 1.11s^3 - 6.845s^2 + 14.1071s + 2$ and example sample output $Z(s)$.

Table 2 Estimated quantiles of L_2 distribution for Example II

	Metamodel				
L_2 *norm*	*i*	*ii*	*iii*	*iv*	*v*
$Q_{0.1}$	1.5976	1.5153	1.2094	1.2173	5.5965
Median	2.4713	2.4097	1.8748	1.9117	6.0363
$Q_{0.9}$	3.3226	3.2462	2.6424	2.6959	6.5048

Figure 4 (we display results for the seed value of 10 only; note the different scales for the y-axis in the four plots).

The intercept in (11) estimates the nugget effect, and this intercept is presented for different σ^2 values in Table 3. Obviously, these results confirm our conjecture: the nugget effect is the variance of the noise.

Conclusions and future research

We assume that, in practice, the mean μ of the Kriging metamodel (1) is not a constant, but is a composition of a signal function and white noise. We presented results for two examples of I/O behaviour of the underlying random simulation model: Ordinary Kriging and Detrended Kriging give quite acceptable predictions, whereas traditional linear regression gives the worst results.

OLS predicts so poorly because OLS assumes that the fitting errors are white noise, whereas Kriging allows errors that are correlated. More specifically, the closer the inputs, the more positive the correlation. Moreover, OLS uses a single estimated function for all input values, whereas Kriging adapts its predictor as the input changes. Note that OLS may be attractive when looking for an explanation—not a prediction—of the simulation's I/O behaviour; for example, which inputs are most important; does the simulation show a behaviour that the experts find acceptable (also see Kleijnen[1])?

OLS is also compared with (universal) Kriging by Regniere and Sharov[12]. Their OLS model, however, is a rather complicated metamodel (involving terms of order six). The resulting prediction accuracies are similar for OLS and Kriging. Further, we found that the nugget effect equals the noise variance.

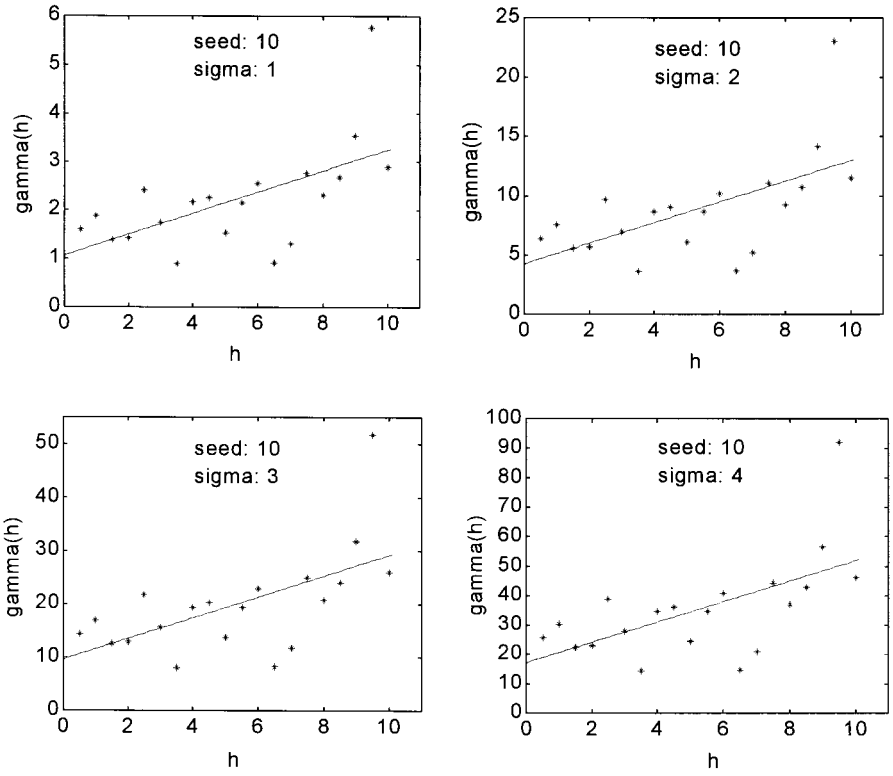

Figure 4 Variogram estimates for different variances.

Table 3 Estimated nugget effects for different white-noise variances

σ^2	*Seed 10*	*Seed 20*
1	1.1	0.9
4	4	4
9	9.6	8.5
16	17.1	15.5
25	26.5	24.1

We restricted our examples to a single input. Therefore, we gave each weight θ_g in the more general distance formula (12) the fixed value of one. In design optimization, however, these parameters are used to control the importance of the input variable x_g; see for example Simpson *et al*[16] (p 8) and Jones *et al* [9] (p 5). In future work we shall investigate multiple inputs.

Further, we shall relax the assumption of white noise: we shall investigate the effects of nonconstant variances (which occur in queueing simulations), common random number usage (which creates correlations among the simulation outputs), and non-normality (Kriging uses maximum likelihood estimators of the weights θ_g, which assumes normality). Finally, we shall apply Kriging to practical queueing and inventory simulations.

Acknowledgements—We thank the two anonymous referees for their most useful comments on an earlier version.

References

1 Kleijnen JPC (1998). Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: Banks J (ed) *Handbook of Simulation*. Wiley: New York, pp 173–223.
2 Campolongo F, Kleijnen JPC and Andres T (2000). Screening methods. In: Saltelli A, Chan K and Scott EM (eds) *Sensitivity Analysis*. Wiley: Chichester, England, pp 65–89.
3 Sacks J, Welch WJ, Mitchell TJ and Wynn HP (1989). Design and analysis of computer experiments. *Stati Sci* **4**: 409–435.
4 Groenendaal WJH van and Kleijnen JPC (1997). On the assessment of economic risk: factorial design versus Monte Carlo methods. *Reliability Eng Systems Saf* **57**: 91–102.
5 Kleijnen JPC and Helton J (1999). Statistical analyses of scatter plots to identify important factors in large-scale simulations. *Reliability Eng Systems Saf* **65**: 147–197.
6 Cressie NAC (1993). *Statistics for Spatial Data*. Wiley: New York.
7 Meckesheimer M, Barton RR, Simpson TW and Booker AJ (2001). Computationally inexpensive metamodel assessment strategies. *Am Inst Aeronaut Astronaut J* **40**: 2053–2060.
8 Koehler JR and Owen AB (1996). Computer experiments. In: Ghosh S and Rao CR (eds) *Handbook of Statistics*, Vol **13**. Elsevier; Amsterdam, pp 261–308.
9 Jones DR, Schonlau M and Welch WJ (1998). Efficient global optimization of expensive black-box functions. *J Global Optim* **13**: 455–492.

10 Matheron G (1962). *Traite de Geostatistique Appliquee*, Memoires du Bureau de Recherches Geologiques et Minieres, No. 14, Editions Technip, Paris, pp 57–59.
11 Huijbregts CJ and Matheron G (1971). Universal Kriging (an optimal method for estimating and contouring in trend surface analysis). In: *Proceedings of the Ninth International Symposium on Techniques for Decision-Making in the Mineral Industry*. Toronto, Canada.
12 Régnière JandSharov A (1999). Simulating temperature-dependent ecological processes at the sub-continental scale: male gypsy moth flight phenology. *Int J Biometereol* **42**: 146–152.
13 Kleijnen JPC and van Groenendaal W (1992). *Simulation, a Statistical Perspective*. Wiley: Chichester, England.
14 Law AM and Kelton WD (2000). *Simulation Modeling and Analysis*, 3rd edn McGraw-Hill: Boston.
15 Kleijnen JPC and van Groenendaal W (1995). Two-stage versus sequential sample-size determination in regression analysis of simulation experiments. *Am J Math Manage Sci* **15**: 83–114.
16 Simpson TW, Mauery TM, Korte JJ and Mistree F (2001). *Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization*. Am Inst Aeronaut Astronaut J **39**: 2233–2241.

Received October 2001;
accepted September 2002 after one revision