TILBURG UNIVERSITY

Tilburg University

Kriging for Interpolation in Random Simulation

van Beers, W.C.M.; Kleijnen, J.P.C.

Publication date:
2001

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
van Beers, W. C. M., & Kleijnen, J. P. C. (2001). *Kriging for Interpolation in Random Simulation*. (CentER Discussion Paper; Vol. 2001-74). Operations research.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Aug. 2022

No. 2001-74

KRIGING FOR INTERPOLATION IN RANDOM SIMULATION

By Wim C.M. van Beers and Jack P.C. Kleijnen

October 2001

ISSN 0924-7815

D:\Data\WP\PAPERS\Beers\KrigDef7.WPD

Written: 27 September 2001

Printed: 8 October 2001 (4:19)

Kriging for Interpolation in Random Simulation

Wim C.M. van Beers [1] and Jack P.C. Kleijnen [2]

[1] Department of Information Systems, Tilburg University (KUB),

Postbox 90153, 5000 LE Tilburg, The Netherlands

Phone: +3113-4662016; Fax: +3113-4663377; E-mail: wvbeers@kub.nl

[2] Department of Information Systems/Center for Economic Research (CentER),

Tilburg University (KUB), Postbox 90153, 5000 LE Tilburg, The Netherlands

Phone: +3113-4662029; Fax: +3113-4663377; E-mail: kleijnen@kub.nl

Web: http://www.tilburguniversity.nl/faculties/few/im/staff/kleijnen/

Abstract

Whenever simulation requires much computer time, interpolation is needed. There are several interpolation techniques in use (for example, linear regression), but this paper focuses on Kriging. This technique was originally developed in geostatistics by D. G. 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: A perfectly specified detrending function gives the best predictions, but Ordinary Kriging gives quite acceptable results; traditional linear regression gives the worst predictions.

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.

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.

Metamodeling 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, Kleijnen, and Andres[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 never been applied to random simulation* - as far as we know. 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 [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 non-linear 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 conclusions and mention possible future research topics.

Kriging

History of Kriging

Kriging is an interpolation technique originally developed by D. G. 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.[3] 's pioneering article. Many more publications followed; for example, Meckesheim et al.[7] give 35 references. Also see Koehler and Owen[8], and Jones, Schonlau, and Welch[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 are, the more positively correlated the prediction errors are*. Mathematically, this assumption is modeled 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, then the prediction equals the observed value. (In deterministic simulation this property is certainly attractive, as we said above.)

In Kriging, a crucial role is played by the *variogram*: 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 subsection. 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 two inputs are, the smaller this dependence is - until the effect is negligible.

Formal Model for Kriging

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

There are several types of Kriging, but we limit this subsection to *Ordinary Kriging*, which makes 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 $\delta(\boldsymbol{s})$:

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

(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 $\Sigma_{i=1}^{n} \lambda_i = 1$. Then we can write the prediction error as

$$E[Z(\mathbf{s}_0) - \Sigma_{i=1}^{n} \lambda_i \cdot Z(\mathbf{s}_i)]^2 - 2m[\Sigma_{i=1}^{n} \lambda - 1] \; . \qquad (4)$$

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 $i, j = 1, ..., n$. 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

$$-\Sigma_{i=1}^{n} \Sigma_{j=1}^{n} \lambda_i \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\Sigma_{i=1}^{n} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - 2m(\Sigma_{i=1}^{n} \lambda_i - 1) \; . \qquad (5)$$

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

$$\boldsymbol{\lambda}' = (\boldsymbol{\gamma} + \mathbf{1}\frac{1 - \mathbf{1}' \Gamma^{-1} \boldsymbol{\gamma}}{\mathbf{1}' \Gamma^{-1} \mathbf{1}})' \Gamma^{-1} \text{ and } m = -(1 - \mathbf{1}' \Gamma^{-1} \boldsymbol{\gamma}) / (\mathbf{1}' \Gamma^{-1} \mathbf{1}) \qquad (6)$$

where $\boldsymbol{\gamma}$ denotes the vector $(\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_1), ..., \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_n))'$ and Γ denotes the $n \times n$ matrix whose $(i, j)^{th}$ element is $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)$; 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) $\hat{\beta}$.

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

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

However, in (6) and (7) $\gamma(\mathbf{h})$ is *unknown*. Usually it is *estimated* by

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

where $|N(\boldsymbol{h})|$ denotes the number of distinct pairs in $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j): \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}; i, j = 1, ..., n\}$; 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 $Z(\boldsymbol{s} + h) - Z(\boldsymbol{s})$ 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

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

Because $C(\boldsymbol{h}) \downarrow 0$ 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, $C(\boldsymbol{h}) <$ for $\| \boldsymbol{h} \| > r + r$. 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 effects 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 pseudo-random 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*

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

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*

$$\begin{aligned} \gamma(\boldsymbol{h}) &= c_0 + b\|\boldsymbol{h}\| \quad \text{it} \quad \boldsymbol{h} \neq \mathbf{0} \\ &= 0 \qquad\quad \text{if} \quad \boldsymbol{h} = \mathbf{0} \end{aligned} \tag{11}$$

where again c_0 is the nugget; see Cressie[6] (p. 61). Actually, we shall apply (11) in our experiments.

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 our 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 $\{f_0(\boldsymbol{s}), ..., f_n(\boldsymbol{s})\}$, $\boldsymbol{s} \in D$. This is called *Universal Kriging*; see Huijbregts and Matheron[11] (p. 160) and also Cressie[6] (p. 151).

Now we introduce a novel type of Kriging that we call *Detrended Kriging*. We assume that the process mean $\mu(\boldsymbol{s})$ satisfies the decomposition

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

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

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

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 $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i)): i = 1, ..., n\}$. Because of 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 $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i) - \hat{S}(\boldsymbol{s}_i)): i = 1, ..., n\}$. 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[12] (p. 156). Cross-validation eliminates one I/O combination, say $(\boldsymbol{s}_k, Z(\boldsymbol{s}_k))$, from to the original data set $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i): i = 1, ..., n\}$, so the remaining $n-1$ data combinations are $\{(\boldsymbol{s}_i, Z(\boldsymbol{s}_i): i = 1, ..., k-1, k+1, ..., n\}$. 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 $k = 1, 2, ..., c$.

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 $(\sum_{k=1}^{c} x_k^2)^{1/2}$). In our experiments we find that the L_1 and L_∞ norms give simular 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 Monte Carlo Examples and Five Metamodels

We are interested in the application of Kriging to discrete-event simulation models, such as simulated queueing systems. Unfortunately, such systems have unknown and uncontrolled I/O

behavior. For example, even when simulating the simple M/M/1 system, it is unknown when exactly the steady state is reached. Therefore, we now create a laboratory; that is, we represent the simulation's I/O behavior - $S(\boldsymbol{s})$ in (14) - by a fourth-order polynomial (example I) or by a hyperbole (example II), both augmented with white noise. In this way, we control the approximation error (bias) of our Kriging metamodel. Moreover, we control the so-called intrinsic simulation noise; that is, we control the variance of the white noise. Finally, we assure that the other white-noise assumptions hold: Normality, statistical independence, and constant variance. Such perfect control is impossible in experiments with the M/M/1 system! (In future research - see last section - we shall apply Kriging to discrete-event systems such as queueing or inventory systems. Such applications, however, only illustrate the behavior of Kriging; they do not permit truly controlled experimentation.)

Example I represents simulations with multiple local maxima, which are interesting when optimizing simulation outputs. Example II represents queueing simulations that show 'explosive' mean waiting times as the traffic rate approaches the value one.

We sample the white noise-term $(\boldsymbol{s})$ in (14) through the Matlab function called 'randn', which gives standard NIID variates; that is, $(\boldsymbol{s})$ 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 *macro-replications*. From these macro-replications 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 $i = 1, ..., 21$. For cross-validation we select (rather arbitrarily) $c = 5$ inputs 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 fourth degree polynomial in Example I, and a hyperbolic function 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.

Example I: 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 2. We obtain output for the following 21 input locations $s_i \in \{0, 0.5, 1, ..., 10\}$; see again Figure 2, which also displays an example of the noisy output $Z(\boldsymbol{s})$. We cross-validate at $s = 0.5, 3.5, 4, 7,$ and 7.5.

The estimated distribution of L_2 is summarized in Table 1. 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.

Example II: Hyperbole

Now we take $S(s) = s/(1-s)$ on $D = [0.01, 0.99] \subset R^1$. This hyperbolic function can represent the mean steady-state waiting time for a traffic rate s in an M/M/1 queueing system;

see Kleijnen and Van Groenendaal[12] (pp. 100-116). This function gives Figure 3. The input locations are $s_i \in \{0.01, 0.05, 0.1, ..., 0.95, 0.99\}$. The cross-validation is carried out at s = 0.05, 0.35, 0.40, 0.70, 0.75.

The estimated distribution of L_2 is summarized in Table 2. Like example I, 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 Monte Carlo Example and Nugget Effect

We also wish to better understand the relationship between the nugget effect in (11) and the variance of the noise $\varepsilon(s)$ in (13). Therefore we perform a simple Monte Carlo example: We take $Z(s) = 10 + \varepsilon(s)$ where $\varepsilon(s)$ is NIID with $\mu = 0$ and σ^2 = 1, 4, 9, 16, and 25 respectively. We sample two macro-replicates, 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 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; 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 found that Kriging applied to detrended data gives good predictions.

Further, we found that the nugget effect equals the noise variance.

We restricted our examples to a single input. Therefore we gave each weight $_g$ 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.[13] (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 non-constant 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 $_g$, which assumes normality). Finally, we shall apply Kriging to practical queueing and inventory simulations.

References

1. Kleijnen, J.P.C. (1998), Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: *Handbook of Simulation*, edited by J. Banks, Wiley, New York, pp. 173-223
2. Campolongo, F., J.P.C. Kleijnen, and T. Andres (2000), Screening methods. In: *Sensitivity Analysis*, edited by A. Saltelli, K. Chan, and E.M. Scott, Wiley, Chichester (England), pp. 65-89
3. Sacks, J. , Welch, W.J., Mitchell, T.J. & Wynn, H.P. (1989), Design and analysis of

computer experiments. *Statistical Science*, 4, no.4, pp. 409-435

4. Groenendaal, W.J.H. Van and J.P.C. Kleijnen (1997), On the assessment of economic risk: factorial design versus Monte Carlo methods. *Reliability Engineering and Systems Safety,* 57, pp. 91-102
5. Kleijnen, J.P.C. and J.Helton (1999), Statistical analyses of scatter plots to identify important factors in large-scale simulations. *Reliability Engineering and Systems Safety,* 65, no. 2, pp. 147-197
6. Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Wiley, New York
7. Meckesheimer, M., R.R. Barton, T.W. Simpson, and A.J. Booker (2001), Computationally inexpensive metamodel assessment strategies. *American Institute of Aeronautics and Astronautics Journal* (submitted)
8. Koehler, J.R. and A.B. Owen (1996), Computer experiments. In: *Handbook of Statistics*, Volume 13, edited by S. Ghosh and C.R.Rao, Elsevier, Amsterdam, pp. 261-308
9. Jones, D.R., M. Schonlau, and W.J. Welch (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 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, C.J. and Matheron, G. (1971). Universal Kriging (An optimal method for estimating and contouring in trend surface analysis). In *Proceedings of Ninth International Symposium on Techniques for Decision-making in the Mineral Industry*
12. Kleijnen, J.P.C. and W. van Groenendaal (1992). *Simulation, a Statistical Perspective*, Wiley, Chichester (Engeland)
13. Simpson, T.W. , Mauery, T.M. , Korte, J.J., and Mistree, F. (2001). *Kriging Metamodels*

for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *American Institute of Aeronautics and Astronautics Journal* (submitted)

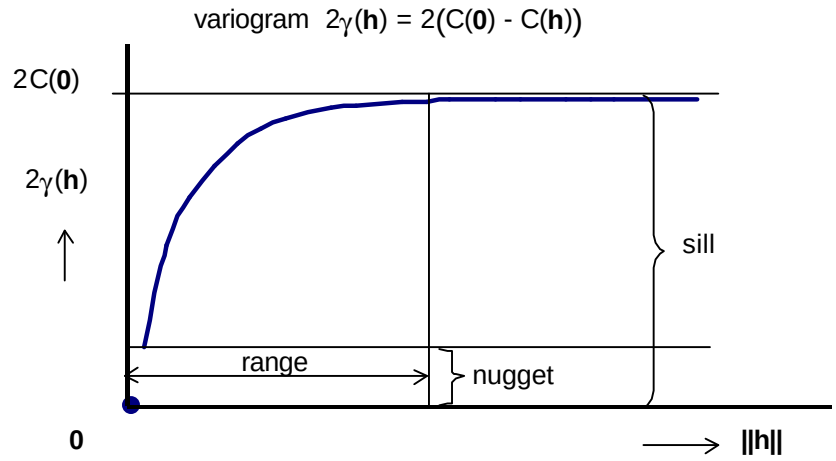

Figure 1: An example variogram

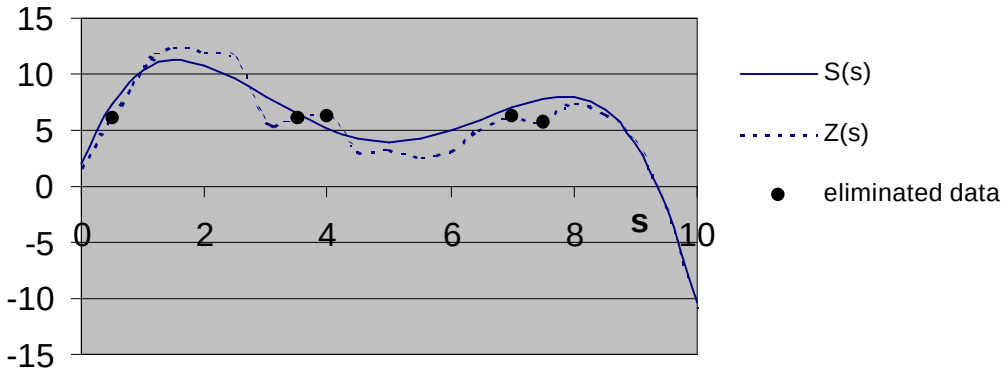

Figure 2: $S(s) = -0.0579s^4 + 1.11s^3 - 6.845s^2 + 14.1071s + 2$ and example sample output $Z(\boldsymbol{s})$

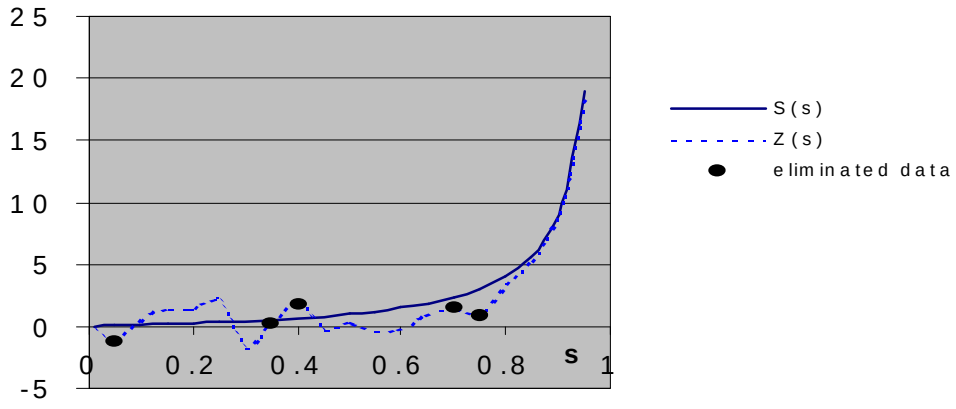

Figure 3: $S(s) = s/(1-s)$ and example sample output $Z(\boldsymbol{s})$

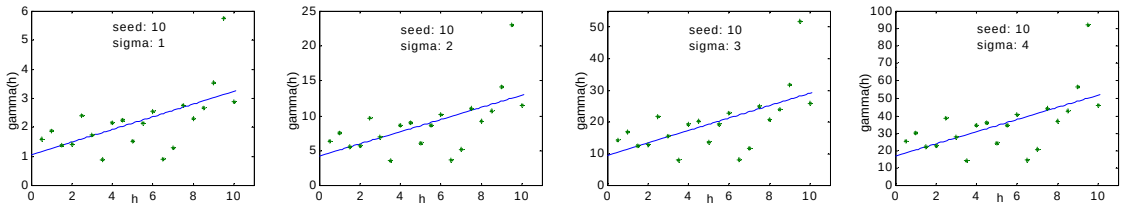

Figure 4: Variogram estimates for different variances

	Metamodel				
L2 norm	**i**	**ii**	**iii**	**iv**	**v**
$Q_{0.1}$	1.5976	1.515	1.2094	1.2173	5.5965
Median	2.4713	2.41	1.8748	1.9117	6.0363
$Q_{0.9}$	3.3226	3.246	2.6424	2.6959	6.5048

Table 1: Estimated quantiles of L_2 distribution for example I

	Metamodel				
L2 norm	i	ii	iii	iv	v
$Q_{0.1}$	1.2429	1.3622	1.1972	2.7411	17.593
Median	1.8832	2.1522	1.8419	3.6678	18.17
$Q_{0.9}$	2.5698	2.925	2.5829	4.4677	18.652

Table 2: Estimated quantiles of L_2 distribution for example II

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

Table 3: Estimated nugget effects for different white noise variances σ^2

Figure captions and table headings

Figure 1: An example variogram

Figure 2: $S(s) = -0.0579s^4 + 1.11s^3 - 6.845s^2 + 14.1071s + 2$ and example sample output $Z(\boldsymbol{s})$

Figure 3: $S(s) = s/(1-s)$ and example sample output $Z(\boldsymbol{s})$

Figure 4: Variogram estimates for different variances

Table 1: Estimated quantiles of L_2 distribution for example I

Table 2: Estimated quantiles of L_2 distribution for example II

Table 3: Estimated nugget effects for different white noise variances