No. 2007–68

KRIGING MODELS THAT ARE ROBUST WITH RESPECT TO SIMULATION ERRORS

By A.Y.D. Siem, D. den Hertog

August 2007

ISSN 0924-7815

Kriging models that are robust with respect to simulation errors

A.Y.D. Siem* D. den Hertog‡

August 29, 2007

Abstract

In the field of the Design and Analysis of Computer Experiments (DACE) meta-models are used to approximate time-consuming simulations. These simulations often contain simulation-model errors in the output variables. In the construction of meta-models, these errors are often ignored. Simulation-model errors may be magnified by the meta-model. Therefore, in this paper, we study the construction of Kriging models that are robust with respect to simulation-model errors. We introduce a robustness criterion, to quantify the robustness of a Kriging model. Based on this robustness criterion, two new methods to find robust Kriging models are introduced. We illustrate these methods with the approximation of the Six-hump camel back function and a real life example. Furthermore, we validate the two methods by simulating artificial perturbations. Finally, we consider the influence of the Design of Computer Experiments (DoCE) on the robustness of Kriging models.

Keywords: Kriging, robustness, simulation-model error

JEL Classification: C60.

1 Introduction

Kriging models were originally proposed and used in Geostatistics; see e.g. Cressie (1991). Later on, Kriging models were more and more used in the design and analysis of computer experiments (DACE). This trend was started by Sacks et al. (1989). Since then, many others followed; see Van Beers and Kleijnen (2004), Jones et al. (1998), Jones (2001), Koehler and Owen (1996), Santner et al. (2003), and Stehouwer and Den Hertog (1999). Nowadays, computer simulations are frequently used to simulate products or processes; see e.g. Fu (2002). One can think of technical products like televisions, mobile phones, laptops, etc. However, these computer simulations can be very time-consuming, i.e., one simulation run may cost hours to run; see e.g. Rommel and Shoemaker (2007). In DACE, Kriging models are used as meta-models to

*Department of Econometrics and Operations Research/CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, Phone:+31 13 4663254, Fax:+31 13 4663280, E-mail: a.y.d.siem@uvt.nl.

‡Department of Econometrics and Operations Research/CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, Phone:+31 13 4662122, Fax:+31 13 4663280, E-mail: d.denhertog@uvt.nl.

approximate time-consuming computer simulations. These meta-models are used for two main purposes. First, they are used to gain more insight into the relationship between the output variables and the input variables of computer simulations. Second, they are used for optimization of the simulated system. Since many (nonlinear) optimization techniques need a lot of function evaluations, the computer simulation model is not appropriate for this purpose. Therefore, instead of the computer simulation model, the meta-model is optimized. Meta-models are also called response surface models, compact models, surrogates, or emulators.

Computer simulations represent a mathematical model of a real-world phenomenon. Therefore they may contain errors. These errors may be model errors, but also numerical errors. In Stinstra and Den Hertog (2007) these kinds of errors in the computer simulation are referred to as 'simulation-model errors', and are defined as the difference between reality and the computer simulation model. Even though simulation-model errors can become rather big, they are often neglected. In Oden et al. (2006) it is stated that quantification of these simulation-model errors will have 'a profound impact on the reliability and utility of simulation methods in the future'.

In constructing meta-models, we have to consider these simulation-model errors, since meta-models can be sensitive to these errors. The meta-model, based on the incorrect data may deviate a lot from the meta-model based on the correct data; e.g., an error of 5% in the data may be 'magnified' to 20% by the meta-model.

In this paper, we show through some examples that Kriging models may be non-robust with respect to simulation-model errors. We introduce a robustness criterion to quantify the amount of robustness of a Kriging model. We present two robust Kriging methods that are more robust against simulation-model errors. These robust Kriging models are less sensitive to errors in the output data, i.e., the difference between the Kriging models based on the correct and the perturbed output data is kept as small as possible. If there is little uncertainty about the values of the output data, we keep the deviation of the Kriging model with respect to the Kriging model based on the correct data as small as possible. Another advantage of the new methods is that the Kriging models become numerically more stable. We validate the methods by artificially incorporating simulation-model errors in the output data. Finally, we also study the influence of the Design of Computer Experiments (DoCE) on the robustness of Kriging models. It turns out that space-filling designs are good designs with respect to the robustness of a Kriging model.

Until now, not so much research has been done yet on robust Kriging models. In Stinstra and Den Hertog (2007), robust optimization of different kinds of meta-models is studied. They consider robust optimization of meta-models with respect to simulation-model errors, meta-model errors, and implementation errors. Note that in Stinstra and Den Hertog (2007) the goal is to find a robust optimum using a classical and possibly non robust Kriging model, whereas in this thesis, the goal is to find a robust meta-model. In Hawkins and Cressie (1984), a Kriging method is proposed that is robust against outliers. In that paper, the data is edited before the actual Kriging is carried out. In Matías and González-Manteiga (2003) a regularized Kriging method is introduced, which performs well if the data contain outliers. In that paper, the MSE of the predictor is decomposed into a variance part and a bias part and instead of the minimization of the variance under an unbiasedness constraint, a linear combination between the variance part

and the bias part is minimized. In Salazar Celis et al. (2007), rational models are constructed that account for simulation-model errors by requiring the model to intersect vertical segments through the output data. It is shown how this problem is reduced to a quadratic programming problem with a convex objective function.

This paper is organized as follows. In Section 2 we summarize the theory on Kriging models. In Section 3, we give examples of non-robust Kriging models, and introduce a robustness criterion. In Section 4, we introduce two new methods to obtain robust Kriging models. In Section 5, we study the influence of different Designs of Computer Experiments on the robustness of Kriging models. Finally, in Section 6 we present our conclusions and propose directions for further research.

2 Kriging models

In this section we summarize some Kriging theory according to Sacks et al. (1989). The function $y : \mathcal{U} \mapsto \mathbb{R}$, with $\mathcal{U} \subseteq \mathbb{R}^q$ that we want to approximate, is treated as a realization of a stochastic process $Y(x)$, where x denotes a q-dimensional input variable. This stochastic process is assumed to consist of a regression part and a stochastic part:

$$Y(x) = \sum_{j=0}^{k} \beta_j f_j(x) + Z(x), \tag{1}$$

where $k+1$ is the number of regression functions including $f_0(x) \equiv 1$. Often, the regression functions f_j are left out except for $f_0(x)$, because they do not yield better Kriging models. The stochastic part $Z(x)$ is assumed to have zero mean and constant process variance (say) σ^2. The covariance between $Z(w)$ and $Z(x)$, with $w, x \in \mathcal{U}$, is given by

$$V(w, x) = \sigma^2 R(w, x),$$

where $R(w, x)$ denotes the correlation between $Z(w)$ and $Z(x)$. Given is a vector of computer simulation input data $[x^1, \ldots, x^n]^T$ and a vector of corresponding output data $y_s = [y(x^1), \ldots, y(x^n)]^T$. We assume y_s is a realization of the stochastic vector $Y_s = [Y(x^1), \ldots, Y(x^n)]^T$, defined by (1). Further, we assume a scalar output, as most of the Kriging literature does.

The Kriging model is given by the so-called Best Linear Unbiased Predictor (BLUP):

$$\hat{y}(x) = c^T(x) y_s.$$

It is called BLUP, because the weights $c(x)$ are chosen such that they solve the following minimization problem

$$\begin{array}{ll} \min\limits_{c(x)} & \mathrm{MSE}[\hat{y}(x)] = E[c^T(x) Y_s - Y(x)]^2 \\ \text{s.t.} & E[c^T(x) Y_s] = E[Y(x)]. \end{array} \tag{2}$$

In other words, the MSE is minimized subject to the unbiasedness constraint.

Before we proceed, we introduce some further notation. We write

$$f(x) = [f_0(x), \ldots, f_k(x)]^T$$

for the $k+1$ regression functions in (1), and

$$F = \begin{bmatrix} f^T(x^1) \\ \vdots \\ f^T(x^n) \end{bmatrix}$$

for the values of these regression functions in the n design points. Furthermore, let R be the correlation matrix with elements

$$R_{ij} = R(x^i, x^j), \text{ for } i = 1, \ldots, n \text{ and } j = 1, \ldots, n,$$

i.e., R_{ij} is the correlation between $Z(x^i)$ and $Z(x^j)$. Let

$$r(x) = \big[R(x^1, x), \ldots, R(x^n, x)\big]^T$$

be the vector with correlations between $Z(x^i)$ and $Z(x)$.

Classical Kriging assumes that $c(x)$ is independent of the output data. Then we can rewrite the optimization problem in (2) as (see Santner et al. (2003))

$$\begin{array}{ll} \min\limits_{c(x)} & \text{MSE}[\hat{y}(x)] = \sigma^2[1 + c^T(x)Rc(x) - 2c^T(x)r(x)] \\ \text{s.t.} & F^T c(x) = f(x). \end{array} \tag{3}$$

To solve (3), Lagrange multipliers $\lambda(x)$ are used. This gives the following system of equations:

$$\begin{bmatrix} 0 & F^T \\ F & R \end{bmatrix} \begin{bmatrix} \lambda(x) \\ c(x) \end{bmatrix} = \begin{bmatrix} f(x) \\ r(x) \end{bmatrix}. \tag{4}$$

Solving this system of equations for $c(x)$ and $\lambda(x)$ gives

$$\begin{array}{rcl} \lambda(x) & = & (F^T R^{-1} F)^{-1}(F^T R^{-1} r(x) - f(x)) \\ c(x) & = & R^{-1}(r(x) - F\lambda(x)), \end{array} \tag{5}$$

which yields the Kriging predictor:

$$\begin{array}{rcl} \hat{y}(x) & = & c^T(x) y_s \\ & = & f^T(x)\hat{\beta} + r^T(x)R^{-1}(y_s - F\hat{\beta}), \end{array} \tag{6}$$

where

$$\hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} y_s \tag{7}$$

is the generalized least-squares (GLS) estimate of β in (1).

The MSE of the predictor – also known as the Kriging variance – becomes (see also Lophaven et al. (2002)):

$$\mathrm{MSE}[\hat{y}(x)] = \sigma^2(1 + u^T(x)(F^T R^{-1} F)^{-1} u(x) - r^T(x) R^{-1} r(x)),$$

where $u(x) = F^T R^{-1} r(x) - f(x)$.

Until now, we have not discussed the form of the correlation function $R(w, x)$. Most publications assume that the correlation structure is stationary; i.e. $R(w, x) = R(w - x)$. Usually a parametric family of correlation functions is chosen. A popular choice is the exponential family

$$R^{\theta,p}(w, x) = \prod_{j=1}^{q} \exp(-\theta_j |w_j - x_j|^{p_j}), \tag{8}$$

where, as noted earlier, q is the dimension of the input variable. In this thesis, we will mostly use (8) with $p_j = 2$, as done in Sacks et al. (1989); then (8) is called the Gaussian correlation function.

Furthermore, we assume that the stochastic process $Z(x)$ is Gaussian. Then, its log likelihood is a function of the process variance σ^2, the regression parameters β, and the correlation parameters θ. The MLE $\hat{\beta}$ of β equals the GLS estimator, and is given by (7); the MLE $\hat{\sigma}^2$ of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n}(y_s - F\hat{\beta})^T R^{-1}(y_s - F\hat{\beta}).$$

To find the MLE $\hat{\theta}$ of θ, we should solve (see Sacks et al. (1989))

$$\min_{\theta} |R|^{1/n} \hat{\sigma}^2. \tag{9}$$

Solving (9) is achieved by using some numerical optimization procedure; we use the Matlab toolbox DACE provided by Lophaven et al. (2002).

3 A robustness criterion

We suppose that the output data are subject to simulation-model errors, i.e., instead of the exact data $y(x^1), \ldots, y(x^n)$, we have perturbed data $\tilde{y}^i = y(x^i) + \varepsilon_y^i$, where ε_y^i is the perturbation of the i-th data point $y(x^i)$. As a consequence, the Kriging model may deviate from the Kriging model based on the correct data.

In Section 3.1, we give an example of a non-robust Kriging model. In Section 3.2, we introduce the so-called rc-value which is a measure for the robustness of a Kriging model.

3.1 An example of a non-robust Kriging model

In this subsection we give an example which shows that Kriging models may be non-robust.

Example 3.1

We consider the approximation of the so-called Six-hump camel back function; see Dixon and Szego (1978). This function is defined as

$$f_4(x_1, x_2) = x_1^2(4 - 2.1x_1^2 + x_1^4/3) + x_1x_2 + x_2^2(-4 + x_2^2);$$

see also Figure 1. In Figure 2, a Kriging model based on the correct data, and a Kriging model

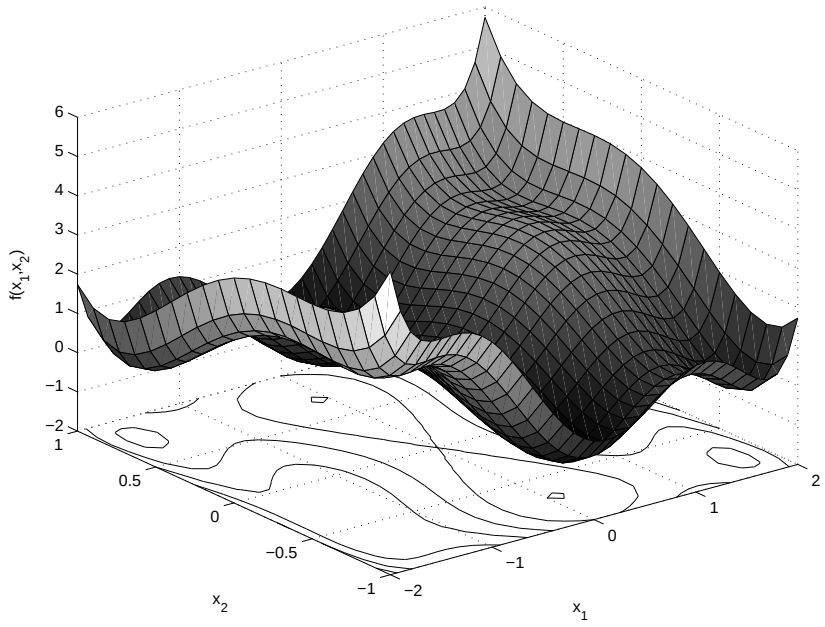


Figure 1: Graph of $f(x_1, x_2) = x_1^2(4 - 2.1x_1^2 + x_1^4/3) + x_1x_2 + x_2^2(-4 + x_2^2)$.

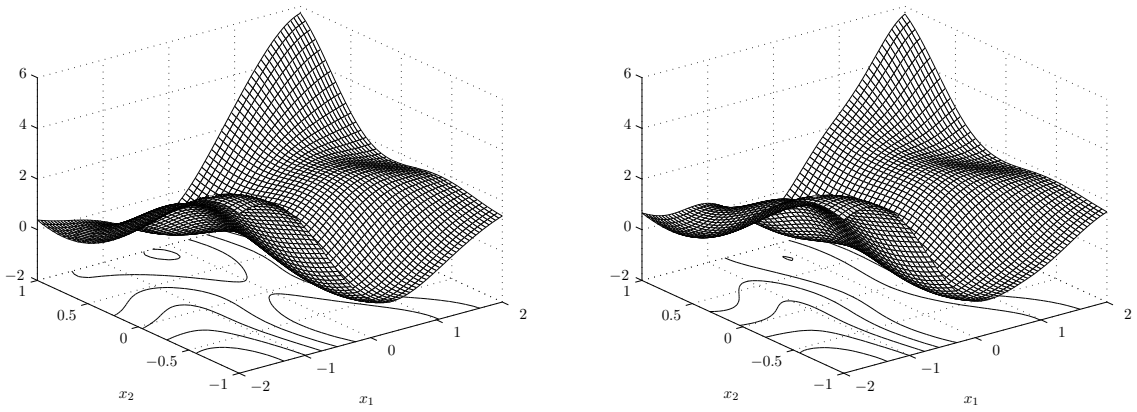
based on perturbed data are shown, for $n = 16$ data points. We used Kriging models without a regression part, i.e., in (1) we took $f_0(x) \equiv 1$, and $f_j(x) = 0$, for $j \geq 1$. The locations of the input data points are shown in Figure 3. The vector of perturbations is given by:

$$\begin{aligned}\varepsilon_y =&[0.017, -0.018, 0.052, -0.024, 0.0053, -0.052, 0.021, 0.022,\\ &-0.084, 0.022, -0.0037, 0.013, -0.065, -0.024, -0.031, 0.12]^T.\end{aligned}$$

This vector is sampled from the uniform distribution on a hypersphere around 0 with radius 0.2. For this vector it holds that $\|\varepsilon_y\|_2 = 0.1893$. The nominal output data values are given by

$$\begin{aligned}y =&[3.9966, 2.0788, 0.5542, 3.5276, 1.7050, 0.9322, -0.8909, -0.7797,\\ &-0.0688, 0.9572, 1.8925, 0.5688, 1.9661, 2.2138, 5.7333, 1.2014]^T.\end{aligned}$$

In Figure 3, a contour plot of the absolute difference of both Kriging models in Figure 2 is shown. It can be seen from this figure that deviations of about 0.7 are reached, which is a relatively large number compared to the perturbations of the data ε_y. The maximal deviation is almost



(a) nominal data (b) perturbed data

Figure 2: A nominal and perturbed Kriging model with $\|\varepsilon_y\|_2 = 0.1893$ and $n = 16$ in Example 3.1.

magnified 6 times with respect to the maximal value in the vector ε_y, and almost 4 times with respect to $\|\varepsilon_y\|_2$.

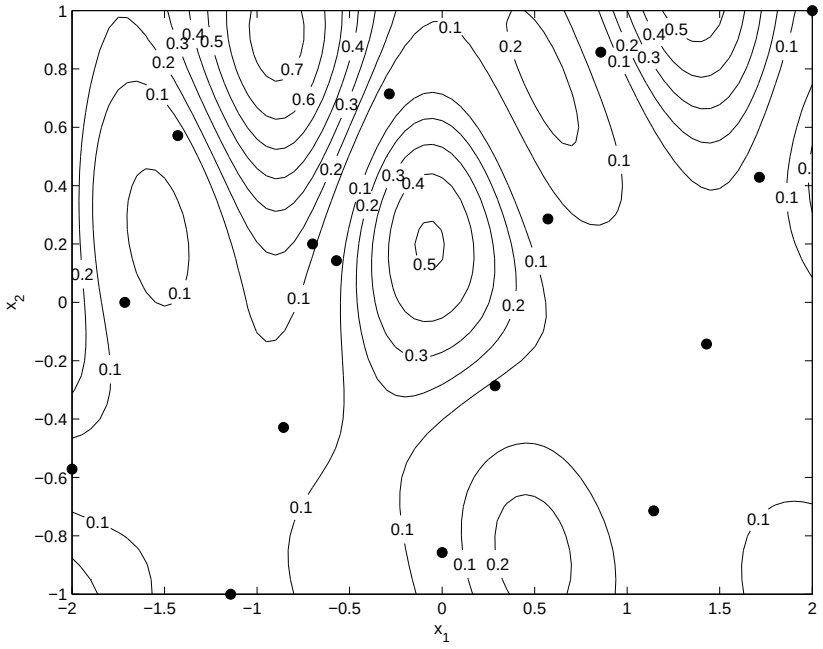


Figure 3: Absolute difference between the nominal and a perturbed Kriging model in Example 3.1.

Next, we consider the Kriging prediction of the point $(-0.15, 0)$ based on these 16 data points. In particular, we consider the Kriging weights $c(x)$ of the predictor in this point. The Kriging weights associated with the input data points are shown in Figure 4. The point $(-0.15, 0)$ is indicated by a '+'. Note that there are two points that have relatively large weights, namely -2 and 2.6. If the data points associated with these weights contain errors, these errors also get relatively large weights. This means that an error in the output data in these points is magnified by these large Kriging weights. Furthermore, it seems that it is not really necessary

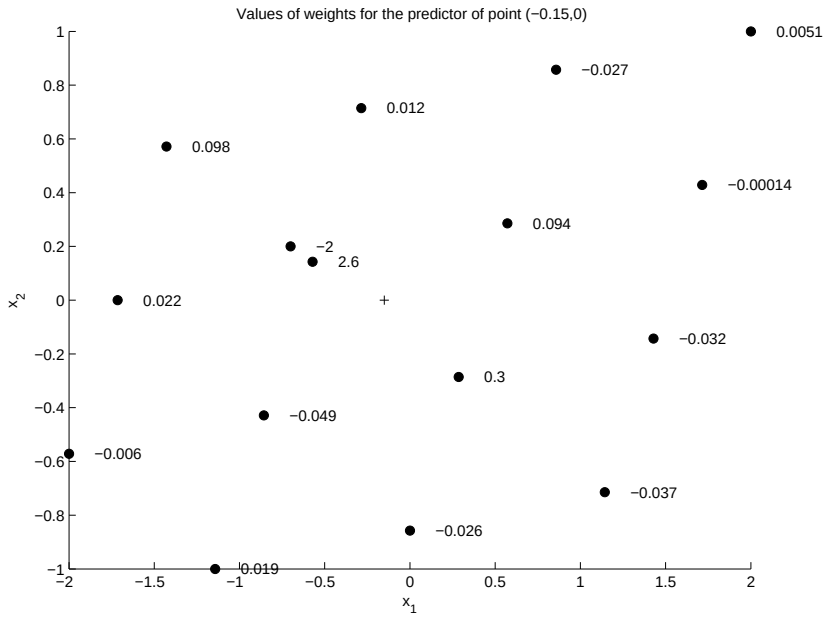


Figure 4: Kriging weights for the prediction of $y(-0.15, 0)$ in Example 3.1.

for the weights to have such a size, because these weights cancel each other more or less. This behavior typically occurs when two or more data points are relatively close to each other. □

3.2 The rc-value

The example illustrated in Figure 4 suggests that the values of the Kriging weights $c(x)$ play an important role in the robustness of a Kriging model. The Kriging model based on the perturbed data is given by:

$$\hat{y}_\varepsilon(x) = c^T(x)(y_s + \varepsilon_y) = c^T(x)y_s + c^T(x)\varepsilon_y. \tag{10}$$

The perturbed Kriging predictor $\hat{y}_\varepsilon(x)$ will be least sensitive to simulation-model errors if the values of $c(x)$ are small. Note that in (10), we assume that $c(x)$ is independent of the output data y_s, which is not the case in general. I.e., in standard Kriging analysis, output data y_s would result in a different $c(x)$ than output data $y_s + \varepsilon_y$. Note that this assumption is also made when rewriting (2) into (3).

In practice, an estimate of the maximal simulation-model error is often available. Let ε_y be the vector containing the perturbations ε_y^i. The errors could be inside a box: $-\nu \leq \varepsilon_y^i \leq \nu, \forall i = 1, \ldots, n$. The errors could also be inside a hyperellipse: $\|A\varepsilon_y\|_2$, where A is an $n \times n$-matrix, and $\|\cdot\|_2$ the Euclidean norm. In particular, the errors could be inside a hypersphere: $\|\varepsilon_y\|_2 \leq \nu$.

Lemma 1. *Suppose that the Euclidean norm of the simulation-model error vector ε_y of output data is bounded by ν, i.e., $\|\varepsilon_y\|_2 \leq \nu$. Then*

$$c^T(x)\varepsilon_y \leq \nu\|c(x)\|_2. \tag{11}$$

Proof. To determine the maximal value that $c^T(x)\varepsilon_y$ can attain, we solve the maximization problem:

$$\begin{array}{ll} \max\limits_{\varepsilon_y} & c^T(x)\varepsilon_y \\ \text{s.t.} & \varepsilon_y^T\varepsilon_y \leq \nu^2. \end{array} \tag{12}$$

It is easy to verify that (12) has optimal value $\nu\|c(x)\|_2$, which implies that $c^T(x)\varepsilon_y \leq \nu\|c(x)\|_2$. □

We may regard the righthand side of (11) as an approximation of the maximal deviation of the nominal Kriging model. Therefore, we regard $\|c(x)\|_2$ as a robustness-criterion value (rc-value) of a Kriging model. The larger the value of $\|c(x)\|_2$, the less robust the Kriging model is. Note that in the data points we have that $\|c(x)\|_2 = 1$. Furthermore, if $\sum_{i=1}^n c_i(x) = 1$ and $c_i \geq 0$, then $\|c(x)\| \leq 1$.

Next, we are interested in the deviation of the nominal Kriging model relative to the maximal perturbation, $\|\varepsilon_y\|_\infty$, i.e.,

$$\frac{c^T(x)\varepsilon_y}{\|\varepsilon_y\|_\infty}. \tag{13}$$

Note that the maximum of maximization problem (12) is attained by the vector ε_y, with $\|\varepsilon_y\|_2 = \nu$, with the same direction of the vector $c(x)$, i.e., the maximum is attained for

$$\varepsilon_y = \nu\frac{c(x)}{\|c(x)\|_2}.$$

Substituting this into (13), we obtain for the maximal deviation of the nominal Kriging model relative to the maximal perturbation, which we call the $\overline{\text{rc}}$-value:

$$\overline{\text{rc}}\text{-value} = \nu\frac{c^T(x)c(x)}{\|c(x)\|_2}\cdot\frac{1}{\nu}\frac{\|c(x)\|_2}{\|c(x)\|_\infty} = \frac{\|c(x)\|_2^2}{\|c(x)\|_\infty}.$$

Note that the $\overline{\text{rc}}$-value is a measure that quantifies the magnification factor by which the Kriging model may magnify the maximal perturbation $\|\varepsilon_y\|_\infty$. Furthermore, note that the relation between the rc-value and the $\overline{\text{rc}}$-value is given by:

$$\overline{\text{rc}}\text{-value} = \frac{\|c(x)\|_2}{\|c(x)\|_\infty}\text{rc-value}.$$

Example 3.2

In this example, we again consider the approximation of the Six-hump camel back function as we also did in Example 3.1. We take the same $n = 16$ input data points. The location of the input data points can be seen in Figure 5. We calculate the corresponding correct output values, from which we fit a Kriging model. Then, we calculate the rc-value $\|c(x)\|_2$. The rc-value of this

Kriging model is shown in Figure 5. Rc-values of more than $\|c(x)\|_2 = 3$ are reached. In Figure 6, the $\overline{\text{rc}}$-value is shown. From this figure we can see that $\overline{\text{rc}}$-values of 6 are reached. This means that an error of (say) 5% may be magnified to 30%! □

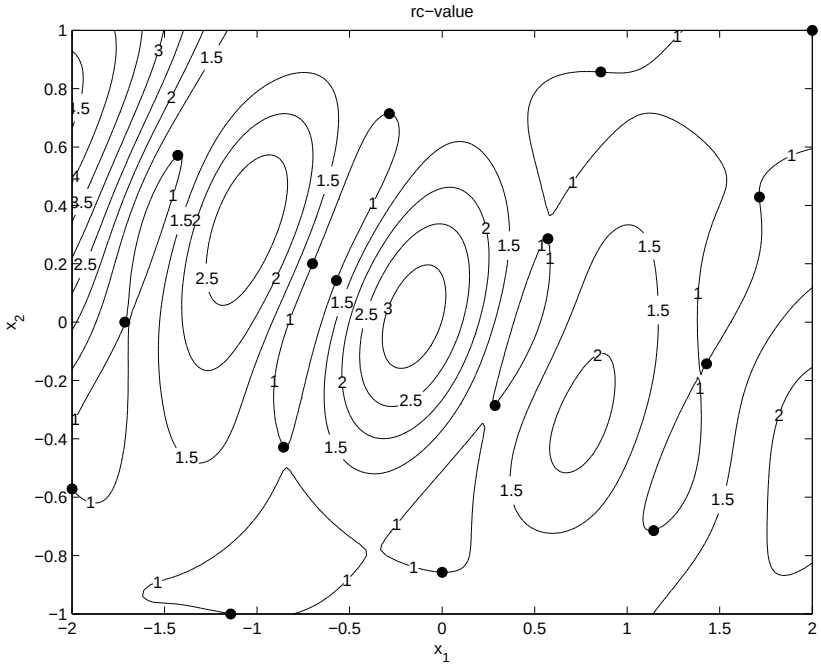


Figure 5: Rc-value of the Kriging prediction of the Six-hump camel back function based on 16 data points in Example 3.2.

The example given in this section motivates the need for Kriging models that are more robust. In Section 4, we shall describe two different methods to construct Robust Kriging models.

4 Robust Kriging models

The classical Kriging method minimizes the MSE of the predictor subject to an unbiasedness constraint; see (3). In Section 3.2, we introduced the rc-value. We also want to make this rc-value as small as possible. Hence, besides minimizing the MSE, we want to minimize the rc-value. More mathematically, we are interested in solving the following bi-objective optimization problem:

$$\begin{array}{ll} \min\limits_{c(x)} & \left\{E[c^T(x)Y_s - Y(x)]^2, \|c(x)\|_2^2\right\} \\ \text{s.t.} & E[c^T(x)Y_s] = E[Y(x)]. \end{array} \tag{14}$$

In this section, we use two well-known methods to solve this bi-objective optimization problem:

1. weighted sum method
2. ε-constraint method.

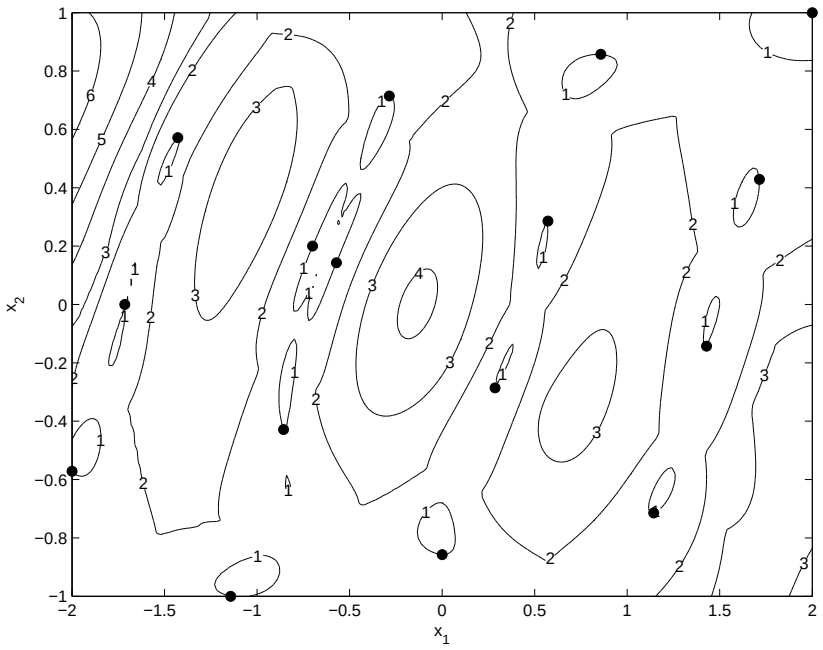


Figure 6: $\overline{\text{rc}}$-value of the Kriging prediction of the Six-hump camel back function based on 16 data points in Example 3.2.

We refer to Miettinen (1999) for an elaborate treatment of these methods.

In Section 4.1, we use the weighted sum method to solve (14). In Section 4.2, we solve (14) by using the ε-constraint method. Both methods yield robust Kriging models. In Section 4.4, we discuss the trade-off between the MSE and the rc-value. In Section 4.5, we validate the new Kriging methods by artificially generating perturbations in the output data, and checking the performance of the robust Kriging methods.

4.1 Weighted sum method

In the weighted sum method, a linear combination of the two objectives in (14) is optimized, i.e., instead of (3), we solve

$$\begin{aligned} \min_{c(x)} \quad & \sigma^2[1 + c^T(x)Rc(x) - 2c^T(x)r(x)] + \rho c^T(x)c(x) \\ \text{s.t.} \quad & F^T c(x) = f(x), \end{aligned} \tag{15}$$

where ρ is the weight we give to minimizing the rc-value. We can solve (15) by using Lagrange multipliers $\lambda(x)$. This gives the following linear system of equations (cf. (4))

$$\begin{bmatrix} 0 & F^T \\ F & R + \frac{\rho}{\sigma^2} I \end{bmatrix} \begin{bmatrix} \lambda(x) \\ c(x) \end{bmatrix} = \begin{bmatrix} f(x) \\ r(x) \end{bmatrix}. \tag{16}$$

Solving this system of equations for $c(x)$ and $\lambda(x)$ gives (cf. (5))

$$\begin{aligned} \lambda(x) &= (F^T(R+\tfrac{\rho}{\sigma^2}I)^{-1}F)^{-1}(F^T(R+\tfrac{\rho}{\sigma^2}I)^{-1}r(x)-f(x)) \\ c(x) &= (R+\tfrac{\rho}{\sigma^2}I)^{-1}(r(x)-F\lambda(x)), \end{aligned}$$

which yields the robust Kriging predictor (cf. (6)):

$$\begin{aligned} \tilde{y}(x) &= c^T(x)y_s \\ &= f^T(x)\hat{\beta}+r^T(x)(R+\tfrac{\rho}{\sigma^2}I)^{-1}(y_s-F\hat{\beta}), \end{aligned} \tag{17}$$

where

$$\hat{\beta} = (F^T(R+\frac{\rho}{\sigma^2}I)^{-1}F)^{-1}F^T(R+\frac{\rho}{\sigma^2}I)^{-1}y_s. \tag{18}$$

Note that the difference between the predictor in (17) and the predictor in (6) is that in (17) instead of R, we have $R+\frac{\rho}{\sigma^2}I$, i.e., to the diagonal elements of R, $\frac{\rho}{\sigma^2}$ is added. Note that this implies that the resulting robust Kriging models are not interpolating. The choice of a suitable value for ρ can be made by trial and error. After one choice for ρ the Kriging model (17) can be build and the corresponding Kriging variance and rc-values can be calculated. Based on these values one may decide to try a new value for ρ.

Example 4.1
In this example, we apply the weighted sum method to the approximation of the Six-hump camel back function. We use the same dataset as in Examples 3.1 and 3.2. We select $\rho = 0.05$, $\rho = 0.15$, and $\rho = 0.3$. The resulting rc-values can be seen in Figures 7, 8, and 9. These examples show that, compared to the rc-value using classical Kriging, shown in Figure 5, the rc-value can be reduced a lot. The rc-value before applying the weighted sum method reaches values of more than 4.5. After applying the weighted sum method, it reaches values of at most 1.4 for $\rho = 0.05$, 1.05 for $\rho = 0.15$, and 1 for $\rho = 0.3$. □

The Kriging predictor in (17) is similar to the predictor that is obtained in the literature, when measurement errors are taken into account. Kriging models that consider measurement errors can be found e.g. in Sacks et al. (1989), Santner et al. (2003), and Sasena et al. (2002). In these papers, an extra term $\varepsilon(x)$ is added to the model (1):

$$Y(x) = \sum_{j=0}^{k}\beta_j f_j(x) + Z(x) + \varepsilon(x),$$

where $\varepsilon(x)$ is a zero mean white noise process, and is independent of $Z(x)$. The variance of $\varepsilon(x)$ is denoted by σ_ε^2. For the MSE$[\hat{y}(x)]$ (cf. the objective in (3)) the following is obtained:

$$\text{MSE}[\hat{y}(x)] = \sigma^2[1+\frac{\sigma_\varepsilon^2}{\sigma^2}+c^T(x)(R+\frac{\sigma_\varepsilon^2}{\sigma^2}I)c(x)-2c^T(x)r(x)].$$

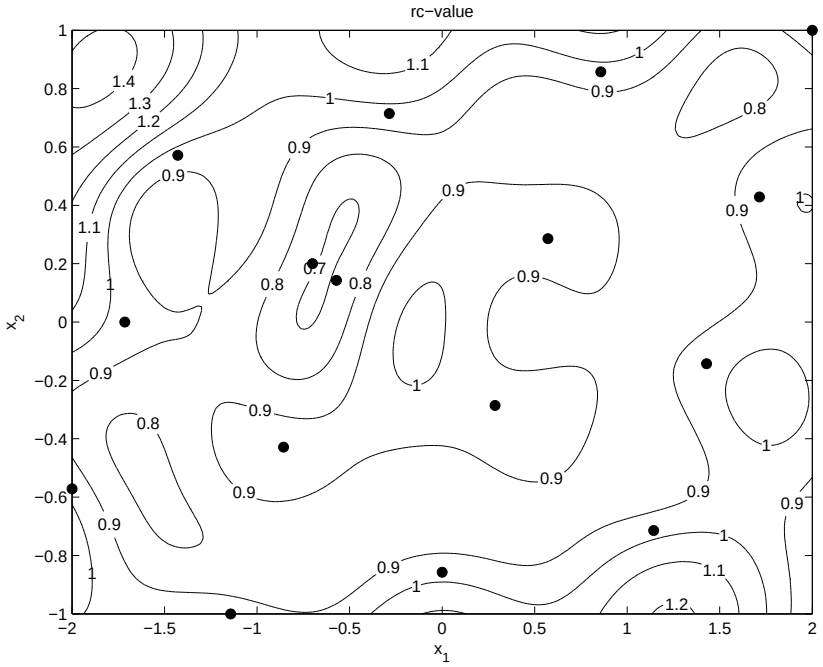


Figure 7: rc-value of the robust Kriging model, using the weighted sum method in Example 4.1 with $\rho = 0.05$.

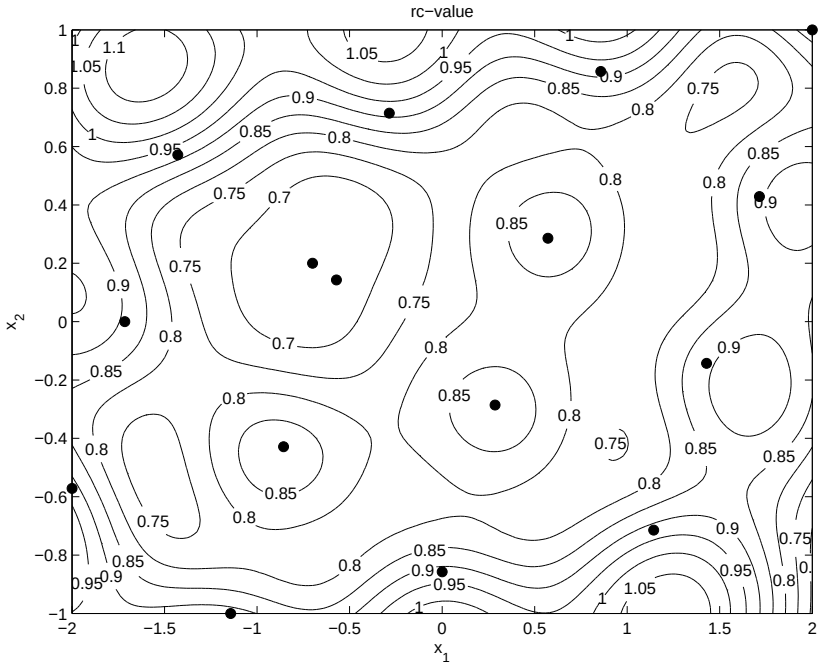


Figure 8: rc-value of the robust Kriging model, using the weighted sum method in Example 4.1 with $\rho = 0.15$.

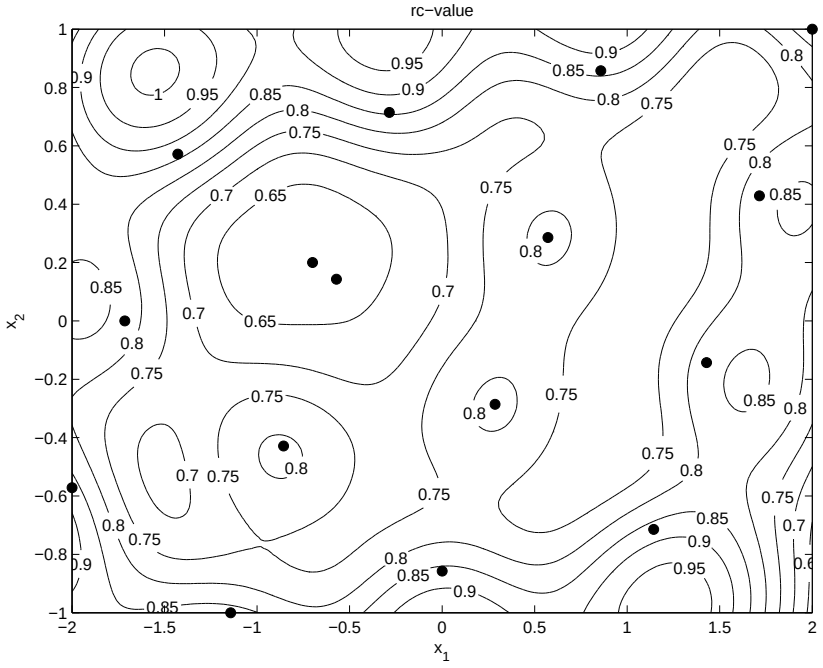


Figure 9: rc-value of the robust Kriging model, using the weighted sum method in Example 4.1 with $\rho = 0.3$.

Now the optimization problem

$$\begin{aligned} \min_{c(x)} \quad & \sigma^2[1 + \tfrac{\sigma_\varepsilon^2}{\sigma^2} + c^T(x)(R + \tfrac{\sigma_\varepsilon^2}{\sigma^2} I)c(x) - 2c^T(x)r(x)] \\ \text{s.t.} \quad & F^T c(x) = f(x), \end{aligned}$$

has to be solved. Using Lagrange multipliers, exactly the same linear system of equations as in (16) is obtained, if we use $\rho = \sigma_\varepsilon^2$. In Sasena et al. (2002) the value of σ_ε^2 is estimated by using the Maximum likelihood estimator. If we treat the simulation-model errors of black-box functions as realizations from a white noise process, we can also estimate $\rho = \sigma_\varepsilon^2$ by using the Maximum likelihood estimator.

Numerical properties

Another attractive aspect of our method is that the condition number of R improves. In Kriging, this matrix R may become very badly conditioned. Since the inverse R^{-1} appears in the Kriging predictor (6), this property may give numerical problems; see Sacks et al. (1989).

If λ is an eigenvalue of R, then $\mu = \lambda + \rho$ is an eigenvalue of $R + \rho I$. Therefore, since R is positive definite, $R + \rho I$ is also positive definite. Since both matrices are positive definite, its singular values are equal to its eigenvalues. Therefore, the 2-norm condition number of both matrices are $\kappa_2(R) = \lambda_{\max}/\lambda_{\min}$ and $\kappa_2(R+\rho I) = \mu_{\max}/\mu_{\min}$, respectively. It is straightforward

to derive the following relationship between the condition numbers of R and $R + \rho I$:

$$\kappa_2(R + \rho I) = \frac{\lambda_{\min}\kappa_2(R) + \rho}{\lambda_{\min} + \rho}.$$

Therefore, if ρ is large enough, it will substantially decrease $\kappa_2(R + \rho I)$. Since $\lambda_{\min}$ is usually small, ρ does not need to be very large to improve the condition number. Therefore, by adding a small number ρ on the diagonal of the correlation matrix R, numerical problems with R^{-1} may be avoided. Ababou et al. (1994) study the condition numbers of covariance matrices based on different correlation functions.

4.2 ε-constraint method

In the ε-constraint method, one of the objectives in (14) is optimized, whereas the other objective is used as an extra constraint. Minimizing the MSE of the predictor and using the rc-value as extra constraint, we obtain:

$$\begin{array}{ll} \min\limits_{c(x)} & \sigma^2[1 + c^T(x)Rc(x) - 2c^T(x)r(x)] \\ \text{s.t.} & F^T c(x) = f(x) \\ & c^T(x)c(x) \leq \gamma^2, \end{array} \tag{19}$$

where γ is the upper bound for the rc-value. The choice of γ^2 can be made by the user, depending on the amount of robustness he/she wants to reach since the value of γ is is an upper bound for the rc-value. Clearly, (19) is an optimization problem with a quadratic objective. We cannot solve (19) explicitly. We solve (19) by using the Karush-Kuhn-Tucker conditions.

Let $\mu(x)$ be the Lagrange multiplier associated with the inequality in (19). We distinguish between two cases: $\mu(x) = 0$ and $\mu(x) > 0$. If $\mu(x) = 0$, we get the same equations as in (4), and the Kriging function becomes the same as in (6). If $\mu(x) > 0$, we obtain the following system of equations:

$$Rc(x) + \mu(x)c(x) + F\lambda(x) = r(x) \tag{20}$$

$$F^T c(x) = f(x) \tag{21}$$

$$c^T(x)c(x) = \gamma^2. \tag{22}$$

We can write $c(x)$ and $\lambda(x)$ as a function of $\mu(x)$ using (20) and (21):

$$\lambda(\mu(x), x) = (F^T(R + \mu(x)I)^{-1}F)^{-1}(F^T(R + \mu(x)I)^{-1}r(x) - f(x)) \tag{23}$$

$$c(\mu(x), x) = (R + \mu(x)I)^{-1}(r(x) - F\lambda(\mu(x), x)), \tag{24}$$

where (24) can be obtained by solving (20) for $c(x)$ and (23) by substituting (24) into (21) and solving this for $\lambda(x)$. Substituting $\lambda(x)$ into $c(x)$ and $c(x)$ into (22), we obtain an equation with only $\mu(x)$ as unknown variable. This equation can be solved numerically. The following lemma helps solving (22).

Lemma 2. *The function $G : (0, \infty) \to \mathbb{R}$ defined by*

$$G_x(\mu) = c^T(\mu, x)c(\mu, x), \tag{25}$$

i.e., the left hand side of (22), is decreasing in μ.

Proof. For the sake of simplicity in notation, we do not explicitly show the dependence of x in this proof. Since the associate Jacobian is nonsingular, it follows from the Implicit function theorem that $\lambda(\mu)$ and $c(\mu)$ are differentiable with respect to μ. Differentiating (20) and (21) to μ gives

$$Rc'(\mu) + \mu c'(\mu) + c(\mu) + F\lambda'(\mu) = 0 \tag{26}$$

$$F^T c'(\mu) = 0. \tag{27}$$

Multiplying (26) by $(c')^T(\mu)$, we obtain

$$(c')^T(\mu)Rc'(\mu) + \mu(c')^T(\mu)c'(\mu) + (c')^T(\mu)c(\mu) + (c')^T(\mu)F\lambda'(\mu) = 0.$$

Using (27) and the fact that $R + \mu I$ is positive definite, we obtain:

$$(c')^T(\mu)c(\mu) = -(c')^T(\mu)(R + \mu I)c'(\mu) \leq 0.$$

This implies that (25) is decreasing in μ. □

The result in Lemma 2 makes it easier to solve (22) numerically. We can solve (22) e.g. by using bisection.

To solve (22) numerically, it is necessary to repeatedly calculate $(R + \mu I)^{-1}$, for various values of μ. The following lemma shows that this computation can be done efficiently.

Lemma 3. *Let $R \succ 0$ be a correlation matrix, and let $\mu > 0$. Then we can write $(R+\mu I)^{-1}$ as:*

$$(R + \mu I)^{-1} = Q\Lambda_\mu^{-1}Q^T,$$

where $\Lambda_\mu^{-1} = diag(1/(\lambda_1^R + \mu), \ldots, 1/(\lambda_n^R + \mu))$ with $\lambda_1^R, \ldots, \lambda_n^R$ the eigenvalues of R, and Q is an orthogonal matrix containing the eigenvectors of R.

Proof. Note that since R is positive definite and $\mu > 0$, $R + \mu I$ is also positive definite. This means that we can decompose $R + \mu I$ as

$$R + \mu I = Q\Lambda_\mu Q^T,$$

where Q is an orthogonal matrix containing the eigenvectors of $R + \mu I$, and Λ contains the eigenvalues of $R + \mu I$. Note that $R + \mu I$ and R have the same set of eigenvectors. Furthermore, the eigenvalues of $R + \mu I$ are given by $\lambda_1^R + \mu, \ldots, \lambda_n^R + \mu$. Using this property, $(R + \mu I)^{-1}$ is

given by

$$(R + \mu I)^{-1} = Q\Lambda_\mu^{-1}Q^T,$$

where $\Lambda_\mu^{-1} = \text{diag}(1/(\lambda_1^R + \mu), \ldots, 1/(\lambda_n^R + \mu))$. □

Using the result of Lemma 3, the calculation of $(R + \mu I)^{-1}$ becomes a lot easier. Now, the matrix Q has to be calculated only one time, and the matrix Λ_μ^{-1} is very easy to calculate. Note that in (23) and (24) the matrix $(R+\mu I)^{-1}$ is multiplied on the righthand side by a vector, say v. Then $(R + \mu I)^{-1}v = Q\Lambda_\mu^{-1}Q^T v$. The matrix vector multiplication $Q^T v$ has to be calculated only once. Since $\Lambda_\mu^{-1}Q^T v$ is a matrix vector multiplication, and $Q\Lambda_\mu^{-1}Q^T v$ is also a matrix vector multiplication, the calculations are much easier than calculating $(R + \mu I)^{-1}v$ by solving a new linear system of equations, for every μ.

Note that also in the ε-constraint method, adding the number μ on the diagonal of R resulting in $R + \mu I$ gives a better condition number for $R + \mu I$.

Example 4.2

In this example we apply the ε-constraint method to the approximation of the Six-hump camel back function, based on the same data as in Examples 3.1, 3.2, and 4.1. We use the ε-constraint method with $\gamma^2 = 2$. This choice is also made subjectively. The resulting rc-value can be seen in Figure 10. The robustness measured by the rc-value is indeed smaller than $\sqrt{2}$ everywhere. □

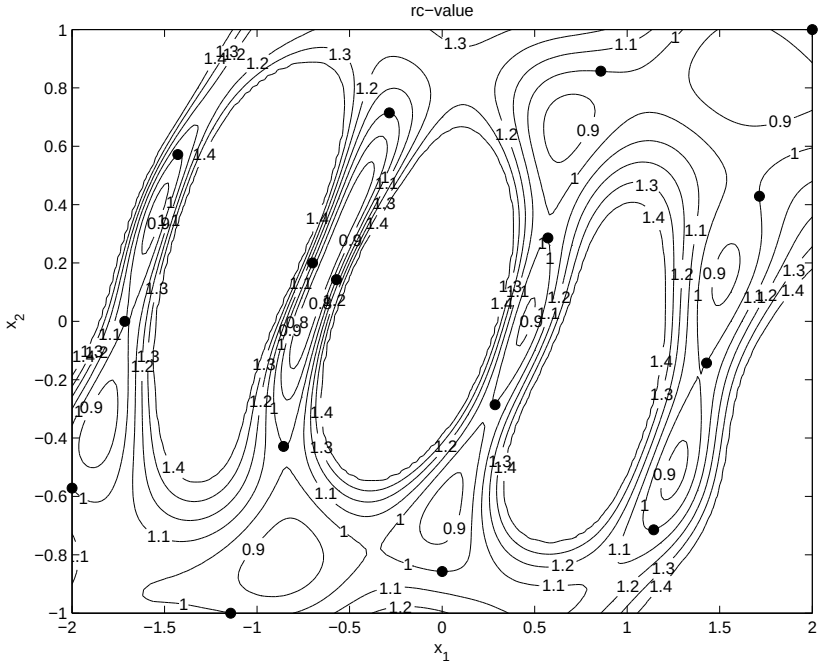


Figure 10: rc-value of the robust Kriging model, using the ε-constraint method in Example 4.2.

Optimization problem (19) is not feasible for all values of γ^2. The minimal value for γ^2 in a

certain input data point x, can be determined by solving the following problem:

$$\begin{aligned} \gamma^2_{\min} = \min_{c(x)} \quad & c^T(x)c(x) \\ \text{s.t.} \quad & F^T c(x) = f(x). \end{aligned}$$

For the often used case that $f(x) = 1$ and $F = e_n$ we have $\gamma^2_{\min} = 1/n$.

In using the ε-constraint method we can also reverse the roles of the MSE of the predictor and the rc-value and minimize the rc-value and impose an upper bound on the MSE, i.e.,

$$\begin{aligned} \min_{c(x)} \quad & c^T(x)c(x) \\ \text{s.t.} \quad & F^T c(x) = f(x) \\ & \sigma^2[1 + c^T(x)Rc(x) - 2c^T(x)r(x)] \leq \tau^2. \end{aligned} \tag{28}$$

It is easy to verify that if the constraint on the MSE is not binding, the optimum of (28) will be attained for $c(x) = F(F^T F)^{-1} f(x)$. Note that the predictor is then exactly the same as the regression model.

Weighted sum method versus ε-constraint method

The weighted sum method improves improves the robustness throughout the whole space and also improves the numerical stability of the Kriging model. In the ε-constraint method the rc-value is improved only in those places where the rc-value of the classical Kriging model exceeds the value of γ. Also, the ε-constraint method improves the numerical stability only in the input data points, where the robustness is improved. Furthermore, the ε-constraint method yields interpolating Kriging models. A drawback of the weighted-sum method is that a good value for ρ cannot be determined a priori. An advantage of the ε-constraint method is that the maximal rc-value of the Kriging model can be chosen a priori. A drawback of the ε-constraint method is that no explicit formula for the Kriging predictor is available.

4.3 Multiplicative perturbations

Until now we assumed that the perturbations in the output data y_s are additive. In this section we assume that the perturbations in the output data are multiplicative, i.e., instead of the correct data $y(x^1), \ldots, y(x^n)$, we have perturbed data $(1 + \varepsilon_y^i)y(x^i)$. Then, instead of (10), we obtain:

$$\hat{y}_\varepsilon(x) = c^T(x)(1 + \varepsilon_y)y_s = c^T(x)y_s + c^T(x)\bar{Y}_s\varepsilon_y, \tag{29}$$

where $\bar{Y}_s$ is a diagonal matrix with the elements of y_s on its diagonal. Note that (29) is similar to (10), but instead of $c^T(x)$, in (29), we have $c^T(x)\bar{Y}_s$. All results for additive perturbations also hold for multiplicative perturbations with $\bar{Y}_s c(x)$ instead of $c(x)$. The additive case is a special case of the multiplicative case with $\bar{Y}_s = I$. We summarize the results briefly. First, we can prove a similar result for the maximal deviation of the nominal Kriging model as in Lemma

1.

Lemma 4. *Suppose that the Euclidean norm of the vector of the multiplicative simulation-model errors is bounded by ν, i.e., $\|\varepsilon_y\|_2 \leq \nu$. Then*

$$c^T \bar{Y}_s \varepsilon_y \leq \nu \|\bar{Y}_s c(x)\|_2$$

Proof. The proof is similar to the proof of Lemma 1. □

We now apply the weighted sum method and the ε-constraint method with $\|c^T(x)\bar{Y}_s\|_2$ as rc-value. Applying the weighted sum method, we obtain as Kriging predictor:

$$\begin{aligned} \tilde{y}(x) &= c^T(x) y_s \\ &= f^T(x)\hat{\beta} + r^T(x)(R + \tfrac{\rho}{\sigma^2}\bar{Y}_s^2)^{-1}(y_s - F\hat{\beta}), \end{aligned}$$

where

$$\hat{\beta} = (F^T(R + \frac{\rho}{\sigma^2}\bar{Y}_s^2)^{-1}F)^{-1}F^T(R + \frac{\rho}{\sigma^2}\bar{Y}_s^2)^{-1}y_s.$$

Note that this is the same as (17) and (18), where I is replaced by $\bar{Y}_s^2$.

If we apply the ε-constraint method, we also obtain similar results as for the additive case. Instead of (20), (21), and (22), we obtain:

$$\begin{aligned} Rc(x) + \mu(x)\bar{Y}_s^2 c(x) + F\lambda(x) &= r(x) \\ F^T c(x) &= f(x) \\ c^T(x)\bar{Y}_s^2 c(x) &= \gamma^2. \end{aligned} \tag{30}$$

Again, similarly to (23) and (24), $c(x)$ and $\lambda(x)$ can be written as a function of $\mu(x)$:

$$\begin{aligned} \lambda(\mu(x), x) &= (F^T(R + \mu(x)\bar{Y}_s^2)^{-1}F)^{-1}(F^T(R + \mu(x)\bar{Y}_s^2)^{-1}r(x) - f(x)) \\ c(\mu(x), x) &= (R + \mu(x)\bar{Y}_s^2)^{-1}(r(x) - F\lambda(\mu(x), x)). \end{aligned}$$

We also have a similar result to Lemma 2:

Lemma 5. *The function $G : (0, \infty) \to \mathbb{R}$ defined by*

$$G_x(\mu) = c^T(\mu, x)\bar{Y}_s^2 c(\mu, x),$$

i.e., the left hand side of (30), is decreasing in μ.

Proof. The proof is similar to the proof of Lemma 2. □

4.4 Trade-off between MSE and rc-value

In Figure 11 a plot of the so-called Pareto curve for the prediction of the point $(0.15, 0)$ is shown. This Pareto curve is constructed by repeatedly solving (19) for different values of γ. Obviously,

there is a trade-off between the MSE and the rc-value. There is much room for improving the rc-value, without giving up a lot in the MSE, so it is really worthwhile, to improve the robustness.

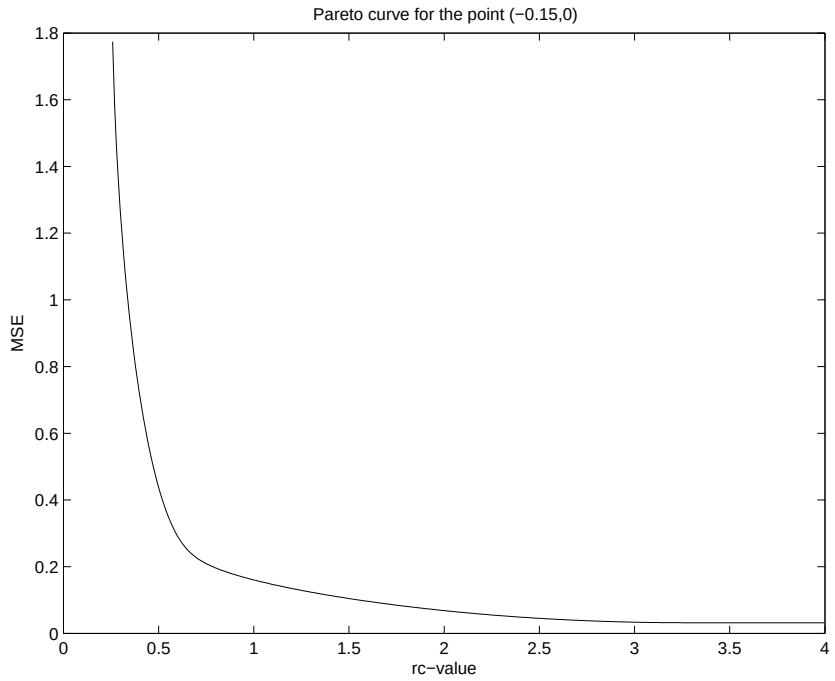


Figure 11: Pareto curve of the MSE against the rc-value for the point $(-0.15, 0)$.

4.5 Assessing the robustness of a Kriging model by Monte Carlo simulation

When applying the weighted sum method or the ε-constraint method (to obtain robust Kriging models), the Kriging weights $c(x)$ are changed. Let us drop the assumption that the Kriging weights $c(x)$ are independent of the perturbations in the output ε_y. Let $\varepsilon_c(x)$ be the perturbation of the Kriging weights $c(x)$ due to the perturbation in the output data ε_y. Furthermore, let $r_c(x)$ be the change that is made in these weights by either the weighted sum method or the ε-constraint method. Then the robust Kriging model $\hat{y}_r(x)$ becomes (cf. (10)):

$$\begin{aligned}\hat{y}_r(x) &= (c(x) + \varepsilon_c(x) + r_c(x))^T (y_s + \varepsilon_y) \\ &= c^T(x) y_s + (c(x) + \varepsilon_c(x) + r_c(x))^T \varepsilon_y + (\varepsilon_c(x) + r_c(x))^T y_s. \end{aligned} \tag{31}$$

In the weighted sum method and the ε-constraint method the second term in (31), $(c(x)+\varepsilon_c(x)+ r_c(x))^T\varepsilon_y$, is made small by minimizing the rc-value or putting an upper bound to the rc-value respectively. The third term is also small, since in both the weighted sum method and the ε-constraint method also the MSE is minimized, which keeps the robust Kriging models accurate to some extent. However, the more weight we put on minimizing the rc-value, the bigger $r_c(x)$ can become, which shows the trade-off between accuracy and robustness mathematically.

The two robust Kriging methods introduced in the previous sections, do not take the third

term of (31) into account. To check that our methods still make the Kriging models more robust, we assess the robustness of a Kriging model in this section by simulating the perturbations ε_y, using Monte Carlo simulation.

We simulate the perturbations ε_y, by sampling $\varepsilon_{y;j}$, for $j = 1, \ldots, N$ from a uniform distribution on a hypersphere around 0 with radius ν. Based on these N perturbed datasets, we fit N Kriging models and calculate the relative deviation

$$rd(x) = \frac{1}{\nu} \max_{j=1,\ldots,N} |\hat{y}(x) - y_j^{\varepsilon}(x)|,$$

where $\hat{y}(x)$ is the nominal Kriging model, which is based on the correct data, and $\hat{y}_j^{\varepsilon}(x)$ is the Kriging model based on the j-th perturbed dataset. Dividing by ν makes the value of the relative deviation $rd(x)$ comparable to the rc-value. This procedure is summarized in Algorithm 1.

Algorithm 1 Algorithm to calculate the relative deviation $rd(x)$

```
Fit a Kriging model ŷ(x) based on the correct data y_s
FOR j = 1 to N
     Sample ε_{y;j} from the uniform distribution on hypersphere around 0
     with radius ν.
     Fit Kriging model ŷ_j^ε(x) based on perturbed data ỹ_s = y_s + ε_{y;j}.
     Calculate the absolute difference |ŷ(x) − ŷ_j^ε(x)|.
ENDFOR
Calculate rd(x) = (1/ν) max_{j=1,...,N} |ŷ(x) − y_j^ε(x)|.
```

Example 4.3

We again consider the Six-hump camel back function, again based on the same dataset as in the Examples 3.1, 3.2, 4.1, and 4.2. We calculate the relative deviation $rd(x)$ based on the classical Kriging method, based on the weighted sum method with $\rho = 0.15$, and based on the ε-constraint method with $\gamma^2 = 2$. We perturb the data by randomly sampling perturbations $\varepsilon_{y;j}$ for which $\|\varepsilon_{y;j}\|_2 \leq 0.2$ with $j = 1, \ldots, 500$. Applying Algorithm 1, we obtain the $rd(x)$ shown in Figures 12 and 13. In Figure 12 it can be seen that in certain regions, the relative deviation $rd(x)$ is greater than the rc-value of Figure 5. This is a consequence of ignoring $\varepsilon_c(x)$ in the calculation of the rc-value. By using the classical Kriging method, the maximal relative deviation is 3.5. By applying the weighted sum method, the maximal relative deviation is reduced to 2. By applying the ε-constraint method, the relative deviation is especially reduced where the rc-values of the classical Kriging model exceeds the value of $\sqrt{2}$. □

Example 4.4 (Dialectric breakdown strength)

In this example, we consider the analysis of performance degradation data in accelerated tests. The dataset is given in Table 1. The output variable y is the dialectric breakdown strength in kilo-Volts, and the input variables x_1 and x_2 are the time in weeks and the temperature in degrees Celcius. See Nelson (1981) for more about information on this data.

We calculate the relative deviation $rd(x)$ based on the classical Kriging method, the weighted sum method with $\rho = 0.69$, and the ε-constraint method with $\gamma^2 = 2$. We perturb the data

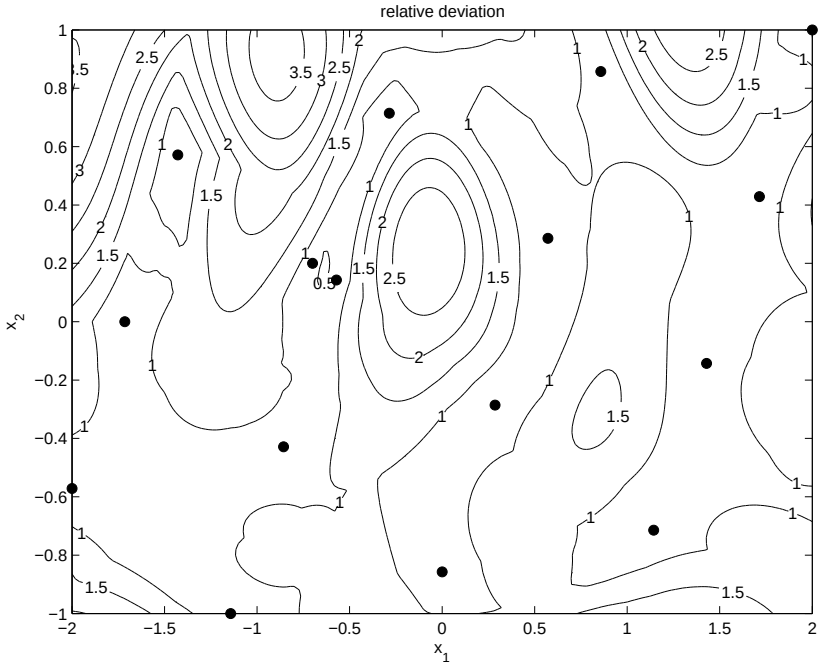
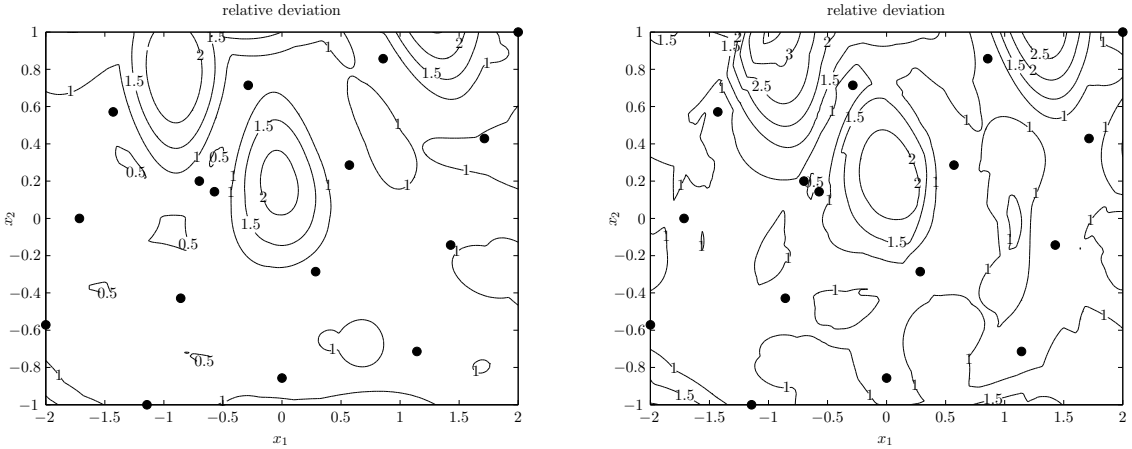


Figure 12: Relative deviation of classical Kriging model of the Six-hump camel back function in Example 4.3.



(a) weighted sum method (b) ε-constraint method

Figure 13: Relative deviation of the robust Kriging models of the Six-hump camel back function in Example 4.3.

x_1	x_2	y
1	225	15.0
1	250	12.5
1	180	15.5
2	225	13.0
2	250	12.0
2	180	14.0
4	225	12.5
4	250	13.0
4	180	17.5
16	225	12.5
16	250	12.0
16	180	17.0
32	180	13.0
32	225	11.0
32	250	10.5

Table 1: Performance degradation data.

by randomly sampling perturbations $\varepsilon_{y;j}$ for which $\|\varepsilon_{y;j}\|_2 \leq 2$ for $j = 1, \ldots, 500$. Applying Algorithm 1, we obtain the $rd(x)$ shown in Figures 14 and 15. These plots show that the relative deviation $rd(x)$ decreased after applying both robust Kriging methods. The classical Kriging method gives a maximal relative deviation of 3.5. The weighted sum method reduces the maximal relative deviation to 1.4. The ε-constraint method reduces the maximal relative deviation to 1.8. □

Examples 4.3 and 4.4 show that the robust Kriging methods introduced in Section 4 indeed improve the robustness of Kriging models.

5 Influence of DoCE on robustness

The Design of Computer Experiments (DoCE) also influences the robustness of a Kriging model. In Kleijnen et al. (2005), an overview of designing simulation experiments is given. The following example compares the robustness of a Latin Hypercube Sample (LHS) (see McKay et al. (1979)) with a maximin Latin Hypercube design (LHD) (see Van Dam et al. (2007)), and a maximin design (see Szabó et al. (2007)). All three designs have 30 input data points.

By definition, in an LHS, the interval of every dimension is divided into subintervals of equal length, and for every subinterval, a design point is drawn from a uniform distribution on that subinterval. The points for each dimension are combined randomly. In a maximin LHD, every dimension is divided into an equal number of levels. The data points are placed on the levels such that in every dimension, there is a data point on every level, and such that the minimal distance between the data points is maximized. In a maximin design the data points are placed such that the minimal distance between the data points is maximized.

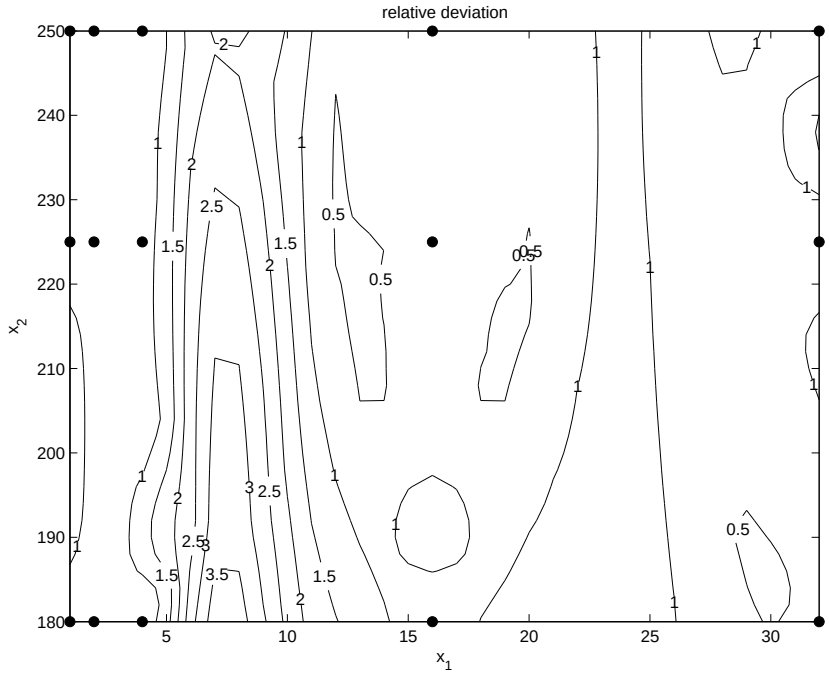
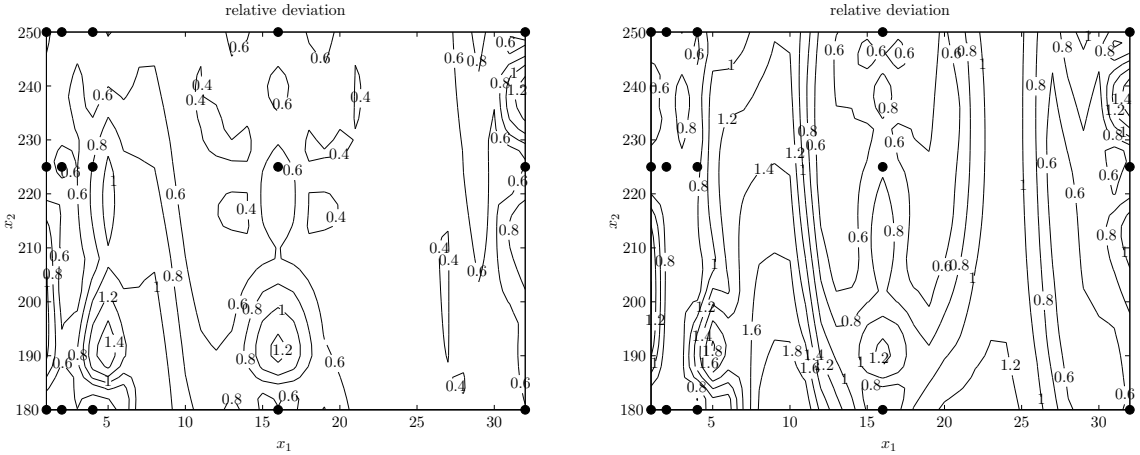


Figure 14: Relative deviation of classical Kriging model of the performance degradation data in Example 4.4.



(a) weighted sum method (b) ε-constraint method

Figure 15: Relative deviation of the robust Kriging models of the performance degradation data in Example 4.4.

Example 5.1

In this example, we compare the rc-values of Kriging models of the six-hump camel back function based on an LHS, a maximin LHD, and a maximin design. We fix the number of design points at 30. The designs as well as the resulting rc-values are shown in Figure 16.

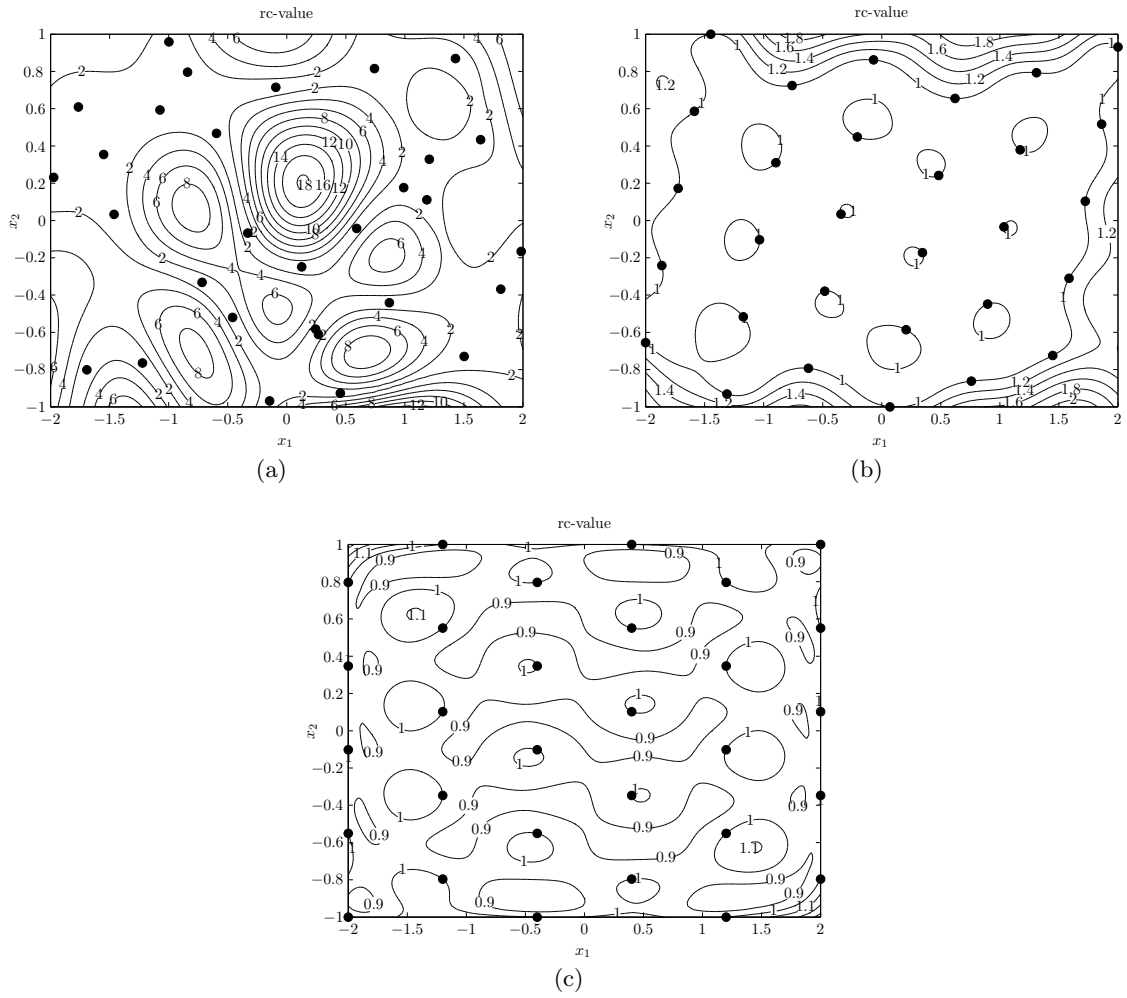


Figure 16: rc-value for LHS (a), maximin LHD (b), and maximin design (c) for Six-hump camel back function in Example 5.1.

These figures show that the rc-values associated with the LHS are much larger than the rc-values associated with the maximin LHD and the maximin design. For the LHS the rc-value reaches 18, however, for the maximin LHD and the maximin design the rc-value reaches only 2 and 1.3 respectively. We obtained similar results for other sizes of datasets. Note that for the LHS, the rc-values are very large in a region where there are relatively few data points. For the maximin LHD, only at the boundary of the region the rc-values are relatively large. The rc-values are the smallest for the maximin design. □

Example 5.1 suggests that to obtain a robust Kriging model, the input data points should be spread as regularly as possible over the input design space.

In sequential optimization, input data points are added iteratively. This strategy is expected

to converge to the optimum of the computer simulation function; see e.g. Jones (2001). Hence, data points will be clustered in a region around the optimum. Therefore we cannot expect the data to be spread regularly over the input design space. Especially if the Kriging models are fitted globally, i.e., based on all data points, this may result in non robust Kriging models.

If sequential designs are used for sensitivity analysis, as done in Kleijnen and Van Beers (2004), the data points may also cluster. However, it is expected that the amount of clustering is less than in case of sequential optimization.

6 Conclusions and further research

Kriging models may be not so robust against simulation-model errors. An error in the output data may be magnified a lot by the Kriging model. We introduced a measure of robustness for Kriging models, the so-called rc-value. This value can also be seen as a quality measure. Robust Kriging models can be obtained by using the rc-value as a criterion. Such robust Kriging models can be obtained by using the weighted-sum method and the ε-constraint method. The choice of Design of Computer Experiments (DoCE) may have a big impact on the robustness of Kriging models. The more space-filling the DoCE is, the smaller the rc-values are.

For further research we propose to extend our methods to robust Polynomials, Rational functions, and robust Radial Basis Functions. For polynomial interpolation, the so-called Lebesgue constant is comparable to the rc-value. Furthermore, it would be interesting to do more research on the the influence of the DoCE on the robustness of Kriging models. From Section 5, it follows that the more space-filling the design is, the smaller the rc-values are, and the more robust the Kriging model will be. Data balancing could be a solution to obtain more robust Kriging models. Data balancing involves leaving out data points, to obtain better models.

Acknowledgements

The authors would like to thank prof. dr. J.P.C. Kleijnen for proof-reading this paper and for his valuable comments.

References

Ababou, R., A.C. Bagtzoglou, and E.F. Wood (1994). On the condition number of covariance matrices in Kriging, estimation and simulation of random fields. *Mathematical Geology 26*(1), 99–132.

Beers, W.C.M. van and J.P.C. Kleijnen (2004). Kriging interpolation in simulation: a survey. In R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters (Eds.), *Proceedings of the 2004 Winter Simulation Conference*, Washington D.C., pp. 113–121.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: Wiley-Interscience.

Dam, E.R. van, B.G.M. Husslage, D. den Hertog, and J.B.M. Melissen (2007). Maximin Latin hypercube designs in two dimensions. *Operations Research 55*(1), 158–169.

Dixon, L.C.W. and G.P. Szego (1978). The optimization problem: An introduction. In L.C.W. Dixon and G.P. Szego (Eds.), *Towards Global Optimization II*. New York: North Holland.

Fu, M.C. (2002). Optimization for simulation: Theory vs. practice. *Journal on Computing 14*(3), 192–215.

Hawkins, D.M. and N. Cressie (1984). Robust Kriging — a proposal. *Mathematical Geology 16*(1), 3–18.

Jones, D., M. Schonlau, and W. Welch (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**, 455–492.

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, **21**, 345–383.

Kleijnen, J.P.C. and W.C.M. van Beers (2004). Application-driven Sequential Designs for Simulation Experiments: Kriging Metamodeling. *Journal of the Operational Research Society*, **55**, 876–883.

Kleijnen, J.P.C., S.M. Sanchez, T.W. Lucas, and T.M. Cioppa (2005). A user's guide to the brave new world of designing simulation experiments. *Journal on Computing 17*(3), 263–289.

Koehler, J.R. and A.B. Owen (1996). Computer experiments. In S. Gosh and C.R. Rao (Eds.), *Handbook of Statistics*, Volume 13, pp. 261–308. New York: Elsevier Science B.V.

Lophaven, S.N., H.B. Nielsen, and J. Sondergaard (2002). DACE: A Matlab Kriging toolbox version 2.0. Technical Report IMM-TR-2002-12, Technical University of Denmark, Copenhagen.

Matías, J.M. and W. González-Manteiga (2003). Regularized Kriging: The support vectors method applied to Kriging. In O. Kaynak, E. Alpaydin, E. Oja, and L. Xu (Eds.), *Artificial Neural Networks and Neural Information Processing - ICANN/ICONIP 2003, Joint International Conference ICANN/ICONIP 2003, Istanbul, Turkey, June 26-29, 2003, Proceedings*, Volume 2714 of *LNCS*, Berlin, Heidelberg, pp. 209–216.

McKay, M.D., R.J. Beckman, and W.J. Conover (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*(2), 239–245.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers.

Nelson, W. (1981). Analysis of performance-degradation data. *IEEE Transactions on Reliability 2*(2), 149–155.

Oden, J.T., T. Belytschko, J. Fish, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip (2006). Simulation-based engineering science. Technical report, National Science Foundation.

Rommel, G.R. and C.A. Shoemaker (2007). A stochastic radial basis function method for the global optimization of expensive functions. *Journal on Computing*. To appear.

Sacks, J., S.B. Schiller, and W.J. Welch (1989). Designs for computer experiments. *Technometrics*, **31**, 41–47.

Sacks, J., W.J. Welch, T.J. Mitchell, and H.P. Wynn (1989). Design and analysis of computer experiments. *Statistical Science 4*(4), 409–435.

Salazar Celis, O, A. Cuyt, and B. Verdonk (2007). Rational approximation of vertical segments. *Numerical Algorithms*. To appear.

Santner, T.J., B.J. Williams, and W.I. Notz (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

Sasena, M., M. Parkinson, P. Goovaerts, P. Papalambros, and M. Reed (2002). Adaptive experimental design applied to an ergonomics testing procedure. In *Proceedings of DETC'02 ASME 2002 Design Engineering Technical Conferences And Computers and Information in Engineering Conference*, Montreal, Canada.

Stehouwer, P. and D. den Hertog (1999). Simulation-based design optimization: Methodology and applications. In *Proceedings of the First ASMO UK/ISSMO Conference on Engineering Design Optimization*, Ilkly, UK.

Stinstra, E.D. and D. den Hertog (2007). Robust optimization using computer experiments. *European Journal of Operational Research*. To appear.

Szabó, P.G., M.Cs. Markót, T. Scendes, E. Specht, L.G. Casado, and I.García (2007). *New approaches to circle packing in a square*. New York: Springer.