Use of design sensitivity information in response surface and kriging metamodels

Citation for published version (APA):

Rijpkema, J. J. M., Etman, L. F. P., & Schoofs, A. J. G. (2001). Use of design sensitivity information in response surface and kriging metamodels. *Optimization and Engineering*, 2(4), 469-484.
<https://doi.org/10.1023/A:1016098623669>

DOI:

[10.1023/A:1016098623669](https://doi.org/10.1023/A:1016098623669)

Document status and date:

Published: 01/01/2001

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

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.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Use of Design Sensitivity Information in Response Surface and Kriging Metamodels

J. J. M. RIJPKEMA, L. F. P. ETMAN, A. J. G. SCHOOFS

Eindhoven University of Technology, Faculty of Mechanical Engineering, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

email: j.j.m.rijpkema@tue.nl

email: l.f.p.etman@tue.nl

email: a.j.g.schoofs@tue.nl

Received March 27, 2001; Revised January 4, 2002

Abstract. Metamodels based on responses from designed (numerical) experiments may form efficient approximations to functions in structural analysis. They can improve the efficiency of Engineering Optimization substantially by uncoupling computationally expensive analysis models and (iterative) optimization procedures. In this paper we focus on two strategies for building metamodels, namely Response Surface Methods (RSM) and kriging. We discuss key-concepts for both approaches, present strategies for model training and indicate ways to enhance these metamodeling approaches by including design sensitivity data. The latter may be advantageous in situations where information on design sensitivities is readily available, as is the case with e.g. Finite Element Models. Furthermore, we illustrate the use of RSM and kriging in a numerical model study and conclude with some remarks on their practical value.

Keywords: metamodels, response surface models, kriging, design sensitivities, computer experiments

1. Introduction

In engineering optimization a direct coupling between analysis models and optimization routines may be very inefficient, as during optimization a large number of iterative calls to possibly time-consuming analysis models may be necessary. In those situations it is preferred to uncouple analysis and optimization through the use of so called metamodels or surrogate models: fast to evaluate approximations for objective and constraint functions. These metamodels may also be helpful in large-scale multidisciplinary optimization (MDO), to exchange relevant information between the disciplines involved and to study the effect of design changes on coupled subproblems.

For the construction of metamodels several strategies exist, such as Response Surface Methods (RSM) (Myers et al., 1995; Schoofs et al., 1992), Kriging (Sacks et al., 1989; Koehler et al., 1996), Radial Basis Functions (RBF) (Powell, 1987), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), Smoothing Splines (Lancaster, 1986) and Neural Networks (Zhang et al., 2000). They all estimate the response for a specific design on the basis of information from the full analysis in a limited number of so called training designs. However, they differ with respect to their underlying conceptual ideas, the calculation effort needed for training and the applicability to specific situations, such as large-scale

optimization problems or analysis models based on numerical simulations. Through numerical model studies one tries to find guidelines for their proper use and selection (e.g., Giunta et al., 1998; Jin et al., 2001; Schoofs et al., 1997).

In this paper we will focus on two metamodeling strategies, namely RSM and kriging. RSM was originally developed for the fitting of data from physical experiments through regression models (Myers et al., 1995). Based on statistical assumptions, such as uncorrelated approximation errors, it can profit from concepts of statistical Design of Experiments (DOE) for an efficient planning of training designs (Montgomery, 1997). However, when used with data from deterministic computer simulation models implicit statistical assumptions used in RSM may be disputed. For those situations kriging interpolation models may form an alternative (Sacks et al., 1989). Being developed originally in the field of geostatistics for the prediction of properties of ore layers they can take spatial correlation information into account and predict analysis results in the training designs exactly. Furthermore, kriging is more flexible with respect to the functional behavior of the responses, as postulating a model function beforehand is not strictly necessary.

Both for RSM and kriging we review key-concepts and present strategies for model training. Furthermore, we discuss ways to enhance these approaches by the incorporation of design sensitivity data. This may lead to a reduction of the actual number of full model analyses that is necessary for model training and estimation. It may be very effective, especially in those situations where design sensitivities are easily available, as is the case for analysis models based on the Finite Element Method. Finally, we illustrate the use of RSM with and without design sensitivity information and the use of kriging in a numerical model study, which throws some light on strengths and weaknesses of these approaches for practical applications.

2. Response surface models

2.1. Key concepts

Construction of Response Surface Models through regression techniques is well known from their use in fitting data from physical experiments (Draper et al., 1998). The approach can also be applied successfully to data from numerical experiments, though implicit statistical assumptions may be subject to discussion in situations where data are generated from deterministic numerical models (Sacks et al., 1989).

The basic approach to RSM-construction (Myers et al., 1995) is to model the response $y(\mathbf{x})$ as the realization of a stochastic variable:

$$y = g(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon \quad (1)$$

In fact it is the combination of a ‘global’ model $g(\mathbf{x}, \boldsymbol{\beta})$, which depends on design variables $\mathbf{x}$ and model parameters $\boldsymbol{\beta}$, and a random error ε . The random error ε is assumed to have zero mean and a variance σ^2 , whereas different realizations are independent. For the ‘global’ model one often chooses a functional form which is linear in its parameters $\boldsymbol{\beta}$:

$$g(\mathbf{x}, \boldsymbol{\beta}) = \beta_1 * f_1(\mathbf{x}) + \dots + \beta_k * f_k(\mathbf{x}) = \mathbf{f}^T(\mathbf{x}) * \boldsymbol{\beta} \quad (2)$$

Training the model boils down to estimating the model parameters β . To achieve this an experimental design $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ has to be selected, containing N points in the design space for which (numerical) experiments have to be carried out. For the purpose of efficiency the number of training points is limited, as gathering of the experimental data may be time-consuming e.g. in large numerical FEM- or BEM-models. Concepts from classical or optimal statistical Designs of Experiments (Montgomery et al., 1997; Atkinson et al., 1992), such as fractional factorial designs, central composite designs or D-optimal designs, may be helpful in this. However, if in deterministic simulation models the assumption of independent random errors ε is not valid other experimental designs such as space-filling designs or lattice designs (Koehler et al., 1996) may be preferred.

Once all information at the training designs has been collected, one can compare actual responses with model predictions for the N experimental points used:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \leftrightarrow \begin{pmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \cdots & f_k(\mathbf{x}_1) \\ \vdots & \vdots & & \vdots \\ f_1(\mathbf{x}_N) & f_2(\mathbf{x}_N) & \cdots & f_k(\mathbf{x}_N) \end{pmatrix} * \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix} \quad (3)$$

More formally, this can be summarized in matrix notation by:

$$\mathbf{y} = \mathbf{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

where $\mathbf{X}$ represents the so called design matrix and $\boldsymbol{\varepsilon}$ is a column with model errors. Using a standard Least Squares (LS) approach estimates, $\mathbf{b}$, for the model parameters, $\boldsymbol{\beta}$, can be calculated from the available data through:

$$\mathbf{b} = (\mathbf{X}^T * \mathbf{X})^{-1} * \mathbf{X}^T * \mathbf{y} \quad (5)$$

Procedures from statistics, such as residual analysis, stepwise regression or subset regression (Rawlings et al., 1998) may be used to check or improve model specification, though with data from deterministic numerical models one has to be careful with respect to the implicit statistical assumptions regarding the model error ε .

From the estimated parameter values, $\mathbf{b}$, one can easily generate predictions for other designs, $\mathbf{x}_0$:

$$\hat{y}(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0) * \mathbf{b} \quad (6)$$

2.2. Incorporation of design sensitivities

With information on design sensitivities easily available, as is the case with e.g. Finite Element Models, efficiency of the model training can improve by including this (additional) information in the training process, simultaneously reducing the number of design points for which training experiments have to be carried out. However, if the response is non-smooth

or contains discontinuities, sensitivity data can disturb the global behavior of the response and should not be used.

To incorporate design sensitivity information into the modeling procedure one augments the postulated relation, Eq. (2), with corresponding relations for all n partial derivatives:

$$\begin{aligned} y &= \mathbf{f}^T(\mathbf{x}) * \boldsymbol{\beta} + \varepsilon_1 \\ \frac{\partial y}{\partial x_1} &= \frac{\partial \mathbf{f}^T(\mathbf{x})}{\partial x_1} * \boldsymbol{\beta} + \varepsilon_2 \\ &\dots \\ \frac{\partial y}{\partial x_n} &= \frac{\partial \mathbf{f}^T(\mathbf{x})}{\partial x_n} * \boldsymbol{\beta} + \varepsilon_{n+1} \end{aligned} \quad (7)$$

These relations can be summarized by a matrix-equation:

$$\mathbf{u}(\mathbf{x}) = \mathbf{F}^T(\mathbf{x}) * \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (8)$$

with:

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) & \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ f_k(\mathbf{x}) & \frac{\partial f_k(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_k(\mathbf{x})}{\partial x_n} \end{pmatrix} \quad (9)$$

Analogous to the standard RSM-approach, one now selects an appropriate experimental design and collects experimental data both for responses and for design sensitivities to train the model. All data found in this way for the N training points can be compared with their respective model predictions. A matrix equation results, which is similar to Eq. (4) for standard RSM:

$$\mathbf{U} = \mathbf{G} * \boldsymbol{\beta} + \mathbf{E} \quad (10)$$

However, the vectors of responses and errors, respectively $\mathbf{U}$ and $\mathbf{E}$, as well as the design matrix $\mathbf{G}$ now are $(n + 1)$ -times as large, as they contain additional information on the n partial derivatives. On first sight it seems attractive to estimate the model parameters, $\boldsymbol{\beta}$, by the standard Least Squares approach described earlier. However, as response quantities and their sensitivities are two entities with different (physical) dimensions, weighting factors must be used to express the importance of sensitivity data with respect to response data. This can be achieved through a Weighted Least Squares approach (WLS) (Myers, 1990). With an appropriate weighting matrix, $\mathbf{B}$, estimates for the model parameters follow from:

$$\mathbf{b} = (\mathbf{G}^T * \mathbf{B} * \mathbf{G})^{-1} * \mathbf{G}^T * \mathbf{B} * \mathbf{y} \quad (11)$$

A common choice for $\mathbf{B}$ is the inverse of the covariance matrix of the experimental data. Assuming independent results between the N training sites and equal covariance matrices,

V , at individual training sites it reduces to:

$$\mathbf{B} = [\text{cov}(\mathbf{U})]^{-1} = (\mathbf{I}_{N \times N} \otimes \mathbf{V}_{(n+1) \times (n+1)})^{-1} = \mathbf{I}_{N \times N} \otimes \mathbf{V}_{(n+1) \times (n+1)}^{-1} \quad (12)$$

An estimate of the covariance matrix V can be determined from experimental data (Johnston et al., 1997):

$$\hat{V} = \frac{1}{N - k} * \sum_{j=1}^N (\mathbf{u}_j - \mathbf{F}^T(\mathbf{x}_j) * \boldsymbol{\beta}) * (\mathbf{u}_j - \mathbf{F}^T(\mathbf{x}_j) * \boldsymbol{\beta})^T \quad (13)$$

In this so called *covariance oriented* approach correlations between responses and sensitivities are taken into account. An alternative is the so called *variance oriented* approach, where V is assumed to be diagonal. Elements follow from corresponding diagonal elements in Eq. (13) and represent variances of the data considered.

For both choices an iterative approach should be employed as estimates for the model parameters both depend on and influence the weighting matrix through V . Effectively, the model parameters estimated in stage k , $\mathbf{b}_k$, are used to calculate an appropriate weighting matrix through V , from which new model parameters, $\mathbf{b}_{k+1}$ are estimated. Once convergence has been reached the estimated model parameters can be used to predict both responses and sensitivities for other designs, $\mathbf{x}_0$.

2.3. Numerical model study

To illustrate the practical performance of the RSM-methods described, namely a direct LS-approach on response data as well as a variance and a covariance oriented WLS-approach on response and sensitivity data, they are applied to a simple two-variable analytical test function:

$$y(\mathbf{x}) = 2 + 4x_1 + 4x_2 - x_1^2 - x_2^2 + a * \sin(bx_1) * \sin(bx_2) \quad 0.5 \leq x_1, x_2 \leq 3.5 \quad (14)$$

Parameter choices are restricted to three representative cases, namely $a = 0.5, b = 2$ (smooth behavior), $a = 2, b = 2$ (smoothly fluctuating behavior) and $a = 0.5, b = 10$ (strongly fluctuating behavior), as illustrated in figure 1.

For all cases training data from a 3^2 full factorial design with factor levels $\{0.5, 2, 3.5\}$ are generated and used to train a full quadratic model:

$$g(\mathbf{x}, \boldsymbol{\beta}) = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_1 x_2 \quad (15)$$

Once model parameters are estimated, they are used to generate model predictions for both responses and sensitivities in the design space. Results for response predictions obtained through the direct LS-approach on (only) response training data are illustrated in figure 2.

The general trend, compared to the true functional behavior as illustrated in figure 1, is that for the smooth case response predictions look rather adequate, whereas in the strongly

Figure 1. Test function values for parameter choices indicated.

Figure 2. RSM response predictions from a direct LS-approach, trained on a factorial 3^2 -design.

fluctuating case there is correspondence for the global behavior while local details are filtered out. In the smoothly fluctuating case a second order polynomial model seems inadequate to capture the real functional behavior. Augmenting the number or the position of the training sites does not substantially alter this behavior, as this is predominantly described by the form of the model, Eq. (15), presumed.

To compare model predictions and exact values from the analytic test function quantitatively the so-called Empirical Integrated Squared Error criterion can be used, defined as:

$$EISE = \frac{1}{N_{grid}} * \sum_{grid} (f(x) - g(x, b))^2 \quad (16)$$

where both predictions and true function values are calculated on a regular grid in the design space.

Results for response, f , and for sensitivities, $\partial f / \partial x$, from a 101×101 regular grid are summarized in Table 1, with training data obtained from a 3^2 full factorial design.

They indicate that inclusion of sensitivity information in the ‘smooth’ case leads to an improvement of the RSM-models both for response, f , and for sensitivities, $\partial f / \partial x$, with the covariance-oriented WLS-approach being slightly superior. In the ‘smoothly fluctuating’ case, though WLS-results show some improvement over the LS-results, a second order polynomial function seems inadequate to capture either response, f , or sensitivities, $\partial f / \partial x$,

Table 1. Summary of *EISE*-results for the RSM-approaches described.

	Smooth $a = 0.5, b = 2$		Smoothly fluctuating $a = 2, b = 2$		Strongly fluctuating $a = 0.5, b = 10$	
	$EISE_f$	$EISE_{\partial f/\partial x}$	$EISE_f$	$EISE_{\partial f/\partial x}$	$EISE_f$	$EISE_{\partial f/\partial x}$
LS	7.82×10^{-2}	2.75×10^{-1}	1.25	4.39	6.84×10^{-2}	6.26
WLS (variance)	6.49×10^{-2}	2.56×10^{-1}	1.04	4.10	7.11×10^{-2}	6.26
WLS (covariance)	6.20×10^{-2}	2.50×10^{-1}	9.92×10^{-1}	4.01	6.56×10^{-2}	6.26

Table 2. Summary of *EISE*-results for RSM-approach specific training designs.

	Smooth $a = 0.5, b = 2$		Smoothly fluctuating $a = 2, b = 2$		Strongly fluctuating $a = 0.5, b = 10$	
	$EISE_f$	$EISE_{\partial f/\partial x}$	$EISE_f$	$EISE_{\partial f/\partial x}$	$EISE_f$	$EISE_{\partial f/\partial x}$
LS (3^2 -design)	7.82×10^{-2}	2.75×10^{-1}	1.25	4.39	6.84×10^{-2}	6.26
WLS (variance; 2^2 -design)	1.06×10^{-1}	2.54×10^{-1}	1.67	4.06	3.74	7.73

as was already suggested from figure 2. In the ‘strongly fluctuating’ case gains in response prediction, f , are small or even absent, whereas models for sensitivities, $\partial f/\partial x$, perform poorly. In fact, the second order polynomial function is not able to capture the real behavior of the sensitivities properly, as is clear from the true functional behavior, sketched in figure 2.

Including sensitivity information in the model training makes it possible to reduce the necessary number of training sites or to estimate a higher order RSM-model. This is illustrated by training a full quadratic model, Eq. (15), on the basis of a 2^2 full factorial design with factor levels $\{0.5, 3.5\}$ using a variance oriented WLS-approach and comparing *EISE*-results with those obtained from the LS-approach on a 3^2 -training design. Results are presented in Table 2.

Results show that a (substantial) reduction in training effort by including sensitivity information seems to pay off in the ‘smooth’ case, where the proposed RSM model is flexible enough to capture the behavior of both responses and sensitivities properly. However, if the response is non-smooth or even contains discontinuities sensitivity information can disturb the global behavior of the response, as is illustrated in the ‘strongly fluctuating’ case. If sensitivity information can only be obtained numerically, through additional model analyses, there is no clear advantage to spend additional effort on sensitivity information instead of using it directly for gathering response information from a larger training design.

3. Kriging

3.1. Key concepts

Kriging originates from the field of geostatistics as a method to predict responses for spatially correlated data, such as the thickness of ore layers, from a limited number of experimental

data (Cressie, 1993). It is, in contrast to RSM, an interpolation method, predicting responses for training designs exactly. As both spatial correlation and the need for exact representation of the training data can also be important with responses from (deterministic) simulation models the kriging approach was introduced for the design and analysis of computer experiments, DACE (Welch et al., 1989).

The basic approach to the construction of kriging models (Koehler et al., 1996) is to model the response $y(\mathbf{x})$ as the realization of a (Gaussian) stochastic process:

$$y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x}) * \boldsymbol{\beta} + Z(\mathbf{x}) \quad (17)$$

In fact it is the combination of a ‘global’ linear regression model $\mathbf{f}^T(\mathbf{x}) * \boldsymbol{\beta}$ and a random process $Z(\mathbf{x})$ which allows for ‘local’ corrections to the ‘global’ model. In this way an exact representation of responses at all training designs can be achieved. The random process $Z(\mathbf{x})$ is assumed to have zero mean as well as a spatial covariance for design sites $\mathbf{x}$ and $\mathbf{w}$ which is the product of a process variance σ^2 and a correlation function $R(\mathbf{x}, \mathbf{w})$:

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{w})) = \sigma^2 * R(\mathbf{x}, \mathbf{w}) \equiv \sigma^2 * R(\mathbf{d}) \quad (18)$$

with $\mathbf{d} = \mathbf{x} - \mathbf{w}$. The correlation function $R(\mathbf{d})$ is user specified and in general restricted to be of the product correlation type:

$$R(\mathbf{x}, \mathbf{w}) = \prod_{j=1}^n R_j(d_j) \quad (19)$$

Furthermore it may depend on some parameters, $\boldsymbol{\theta}$, which have to be estimated from the training data. Several functional forms are available such as cubic, (generalized) exponential or Matérn-type correlation functions. However, in this paper we will restrict to a Gaussian correlation function:

$$R(\mathbf{d}) = \prod_{j=1}^n \exp(-\theta_j * d_j^2), \quad (20)$$

which results in a stochastic process that is infinitely mean square differentiable.

Training the model boils down to estimating the model parameters $\boldsymbol{\beta}$, the process variance σ^2 and the correlation function parameters $\boldsymbol{\theta}$. To achieve this an experimental design $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ has to be selected, containing N points in the design space for which (numerical) experiments have to be carried out. For the purpose of efficiency the number of training points is limited, as gathering of the experimental data may be time-consuming. However, as the model, Eq. (17), assumes no independent random errors, concepts different from classical or optimal statistical Design of Experiments are preferred, such as entropy designs, space filling designs or Latin hypercube designs (Koehler et al., 1996).

Once all training information is collected, one can compare actual responses with model predictions for the N experimental points used. A matrix equation results:

$$\mathbf{y} = \mathbf{X} * \boldsymbol{\beta} + \mathbf{Z} \quad (21)$$

which is similar to Eq. (4) for standard RSM, though now residuals, $\mathbf{Z}$, are correlated according to the user specified correlation function $R(\mathbf{x}, \mathbf{w})$:

$$\text{cov}(\mathbf{Z}) = \sigma^2 * \begin{pmatrix} R(\mathbf{x}_1, \mathbf{x}_1) & \cdots & R(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ R(\mathbf{x}_N, \mathbf{x}_1) & \cdots & R(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix} \equiv \sigma^2 * \mathbf{R}_D \quad (22)$$

As the residuals in the training design points are correlated, model parameters β are best estimated on the basis of Weighted Least Squares (WLS), minimizing:

$$L(\beta) = \mathbf{Z}^T * \mathbf{B} * \mathbf{Z} = (\mathbf{y} - \mathbf{X} * \beta)^T * \mathbf{B} * (\mathbf{y} - \mathbf{X} * \beta) \quad (23)$$

with weighting matrix:

$$\mathbf{B} = (\text{cov}(\mathbf{Z}))^{-1} = \sigma^{-2} * \mathbf{R}_D^{-1} \quad (24)$$

In this way an estimate, $\mathbf{b}$, for the model parameters, β , can be derived:

$$\mathbf{b} = (\mathbf{X}^T * \mathbf{R}_D^{-1} * \mathbf{X})^{-1} * \mathbf{X}^T * \mathbf{R}_D^{-1} * \mathbf{y} \quad (25)$$

as well as an estimate, s^2 , for the process variance σ^2 :

$$s^2 = \frac{1}{N} * (\mathbf{y} - \mathbf{X} * \mathbf{b})^T * \mathbf{R}_D^{-1} * (\mathbf{y} - \mathbf{X} * \mathbf{b}) \quad (26)$$

However, results for $\mathbf{b}$ and s^2 depend on the correlation function parameters θ through $\mathbf{R}_D$. Therefore, these parameters should be estimated first from the experimental data. Using a Maximum Likelihood approach they result from maximizing the log-likelihood function:

$$L(\theta) = -(N * \ln(s^2) + \ln(\det(\mathbf{R}_D))) \quad (27)$$

Notice that practical maximization of $L(\theta)$ may become computationally expensive, as for every evaluation of $L(\theta)$ one has to calculate explicit estimates for s^2 and consequently, Eq. (26), for $\mathbf{b}$ as well as $\det(\mathbf{R}_D)$. Therefore one strives for a small number of correlation function parameters.

Once model parameters and correlation function parameters are estimated the best linear unbiased prediction of the response at a design $\mathbf{x}_0$ can be generated from:

$$y(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0) * \mathbf{b} + \mathbf{r}^T(\mathbf{x}_0) * \mathbf{R}_D^{-1} * \mathbf{Z}_D \quad (28)$$

In this expression $\mathbf{Z}_D$ represents the residuals for the training designs:

$$\mathbf{Z}_D = \mathbf{y} - \mathbf{X} * \mathbf{b} \quad (29)$$

whereas $\mathbf{r}(\mathbf{x}_0)$ represents the spatial correlations between design $\mathbf{x}_0$ and training designs $\mathbf{x}_1, \dots, \mathbf{x}_N$:

$$\mathbf{r}(\mathbf{x}_0) = \begin{pmatrix} R(\mathbf{x}_0, \mathbf{x}_1) \\ \vdots \\ R(\mathbf{x}_0, \mathbf{x}_N) \end{pmatrix} \quad (30)$$

The second term in the expression for the predicted response:

$$\mathbf{r}^T(\mathbf{x}_0) * \mathbf{R}_D^{-1} * \mathbf{Z}_D \quad (31)$$

is in fact an interpolation of the residuals of the regression model $\mathbf{f}^T(\mathbf{x}_0) * \mathbf{b}$, resulting in exact predictions at the training sites used.

3.2. Incorporation of design sensitivities

With information on design sensitivities easily available the efficiency of the model training may be improved by including this (additional) information in the training process (Morris et al., 1993). Analogous to the procedure, as described in Section 2.2, the postulated relation, Eq. (17), is augmented with corresponding relations for all n partial derivatives. A matrix equation results:

$$\mathbf{u}(\mathbf{x}) = \mathbf{F}^T(\mathbf{x}) * \boldsymbol{\beta} + \tilde{\mathbf{Z}}(\mathbf{x}) \quad (32)$$

which is similar to the result from Eq. (8), except for the vector $\tilde{\mathbf{Z}}(\mathbf{x})$, which contains the random process $Z(\mathbf{x})$ and its partial derivatives $\partial Z(\mathbf{x})/\partial x_i$. Analogous to the standard kriging approach, one now selects an appropriate experimental design $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and collects experimental data both for responses and for design sensitivities to train the model. All data found in this way for the N training designs can be compared with their respective model predictions. From this a matrix equation results, which is similar to Eq. (21) for standard kriging:

$$\mathbf{U} = \mathbf{G} * \boldsymbol{\beta} + \tilde{\mathbf{Z}}^* \quad (33)$$

However, the vectors of responses and residuals, respectively $\mathbf{U}$ and $\tilde{\mathbf{Z}}^*$, as well as the design matrix $\mathbf{G}$ now are $(n + 1)$ -times as large, as they contain additional information on the n partial derivatives. Incorporation of these changes in the training approach from standard kriging is straightforward and leads, in principle, to an estimation procedure for model parameters $\boldsymbol{\beta}$, the process variance σ^2 and the correlation function parameters $\boldsymbol{\theta}$. However, the covariance matrix to be used in model training:

$$\text{cov}(\tilde{\mathbf{Z}}^*) \equiv \sigma^2 * \tilde{\mathbf{R}}_D \quad (34)$$

now not only should account for covariances between responses at different training sites $\mathbf{x}$ and $\mathbf{w}$ from $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$:

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{w})) = \sigma^2 * R(\mathbf{x} - \mathbf{w}) \equiv \sigma^2 * R(\mathbf{d}), \quad (35)$$

as is the case with standard kriging, but also for covariances between responses and sensitivities:

$$\begin{aligned} \text{cov}\left(Z(\mathbf{x}), \frac{\partial Z}{\partial x_i} \Big|_{\mathbf{w}}\right) &= \frac{\partial R}{\partial d_i} \Big|_{\mathbf{x}-\mathbf{w}}; & \text{cov}\left(\frac{\partial Z}{\partial x_i} \Big|_{\mathbf{x}}, Z(\mathbf{w})\right) &= -\frac{\partial R}{\partial d_i} \Big|_{\mathbf{x}-\mathbf{w}}; \\ \text{cov}\left(\frac{\partial Z}{\partial x_i} \Big|_{\mathbf{x}}, \frac{\partial Z}{\partial x_j} \Big|_{\mathbf{w}}\right) &= -\frac{\partial^2 R}{\partial d_i * \partial d_j} \Big|_{\mathbf{x}-\mathbf{w}} \end{aligned} \quad (36)$$

Notice that the correlation function $R(\mathbf{d})$ has to be twice differentiable.

Model predictions for both response and design sensitivities at a design $\mathbf{x}_0$ can be generated from:

$$\mathbf{u}(\mathbf{x}_0) = \mathbf{F}^T(\mathbf{x}_0) * \mathbf{b} + \tilde{\mathbf{r}}^T(\mathbf{x}_0) * \tilde{\mathbf{R}}_D^{-1} * \tilde{\mathbf{Z}}_D^* \quad (37)$$

where $\tilde{\mathbf{Z}}_D^*$ represent the residuals for responses and sensitivities at the training designs:

$$\tilde{\mathbf{Z}}_D^* = \mathbf{U} - \mathbf{G} * \mathbf{b} \quad (38)$$

and $\tilde{\mathbf{r}}^T(\mathbf{x}_0)$ represent the spatial correlation between responses and sensitivities at design $\mathbf{x}_0$ and respective training designs $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$.

3.3. Numerical model study

To illustrate the practical performance of kriging it is applied to the two-variable analytical test function, used previously to illustrate RSM:

$$y(\mathbf{x}) = 2 + 4x_1 + 4x_2 - x_1^2 - x_2^2 + a * \sin(bx_1) * \sin(bx_2) \quad 0.5 \leq x_1, x_2 \leq 3.5 \quad (39)$$

Parameter choices are restricted to $a = 0.5, b = 2$ (smooth behavior), $a = 2, b = 2$ (smoothly fluctuating behavior) and $a = 0.5, b = 10$ (strongly fluctuating behavior). For the kriging model a simple form is chosen:

$$y(\mathbf{x}) = \beta_0 + Z(\mathbf{x}) \quad (40)$$

with a constant term, β_0 , as ‘global’ linear regression model and a Gaussian random process $Z(\mathbf{x})$ with:

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{w})) = \exp\left(-\theta * \sum_j (\mathbf{x}_j - \mathbf{w}_j)^2\right) \quad (41)$$

Figure 3. Response predictions from a kriging model, trained on a 3^2 factorial design.

which depends only on a single parameter, θ . This choice was motivated by the practical experience that results appear to be rather insensitive to the specific choice of the ‘global’ linear regression model, whereas a simple, twice differentiable correlation function is preferred, so that it can easily be extended to include design sensitivity data in the model training.

First the model is trained from a 3^2 full factorial design with factor levels $\{0.5, 2, 3.5\}$ through the standard kriging procedure described earlier. At the start one has to estimate the correlation function parameter θ , maximizing the log-likelihood function $L(\theta)$, or equivalently minimizing its negative, $-L(\theta)$. Both in the smooth and in the strongly fluctuating case an optimal value, θ_{opt} , can be located, which is used in further training of the model. However, in the smoothly fluctuating case an interior minimum was not found on the interval considered. In the latter case, after experimenting with different values of θ , a boundary value $\theta = 2.0$ was taken for further training of the model. Results from the kriging models obtained in this way are illustrated in figure 3.

General trend is that for the smooth case response predictions look rather adequate, whereas in the strongly fluctuating case there is correspondence for the global behavior while local details are filtered out. However, notice that in this case the experimental design chosen somehow matches the symmetry of the true functional behavior, figure 1. In the smoothly fluctuating case, as with RSM, the kriging model seems inadequate to capture the real functional behavior. Results from the Empirical Integrated Squared Error criterion $EISE$, Eq. (16), for the response function, f , calculated on a regular 101×101 grid in design space, are reported in Table 3. They confirm these findings and show that for the 3^2 full factorial design RSM outperforms kriging slightly.

Table 3. Summary of $EISE$ -results for kriging and RSM models, trained on equal designs.

	Smooth $a = 0.5, b = 2$		Smoothly fluctuating $a = 2, b = 2$		Strongly fluctuating $a = 0.5, b = 10$	
	$EISE_f$	θ_{opt}	$EISE_f$	θ_{opt}	$EISE_f$	θ_{opt}
Kriging (3^2 full factorial)	8.77×10^{-2}	0.22	1.65	2 ^a	1.61×10^{-1}	0.28
RSM (3^2 full factorial)	7.82×10^{-2}	–	1.25	–	6.84×10^{-2}	–

^aValue used in model prediction as no optimal value was found.

Figure 4. Response predictions from a kriging model, trained on a 5^2 factorial design.

However, in contrast to RSM, differences between predictions and true functional behavior are not predominantly caused by the kriging model assumed but by the experimental design used for training. This can be illustrated by training the same kriging model on a 5^2 full factorial design with factor levels $\{0.5, 1.25, 2, 2.75, 3.5\}$. For all cases an optimal value, θ_{opt} , could be determined. Results are presented in figure 4, and show that now the smoothly fluctuating case is captured far more adequately than with the 3^2 -design or with RSM. This is in line with results from other numerical studies (e.g.: Giunta et al., 1998; Jin et al., 2001).

With kriging models other types of experimental designs may be appropriate. A common class of designs is the Latin hypercube design. One of their specific properties is that projecting the design points to a lower dimensional space by deleting a design variable does not lead to a replication of design points. This may be advantageous with responses from deterministic numerical models.

Results for the kriging model described earlier, when trained on a $N = 9$ and a $N = 25$ Latin hypercube design, are illustrated in figures 5 and 6. They show that even with only $N = 9$ training designs the trend of the smoothly fluctuating function is captured in some detail. However, for the strongly fluctuating case, the behavior of the kriging model does not represent the true functional behavior very well, as it is influenced too much by local

Figure 5. Response predictions from a kriging model, trained on a $N = 9$ Latin hypercube design.

Figure 6. Response predictions from a kriging model, trained on a $N = 25$ Latin hypercube design.

values from the incidental training points used. In this case use of the kriging model as metamodel in optimization may cause some problems.

Values for the Empirical Integrated Squared Error criterion $EISE$, Eq. (16), calculated on a regular 31×31 grid in design space, are reported in Table 4.

According to these results the full factorial designs, which can be considered as space filling designs, outperform the Latin hypercube designs in most cases. However, for higher dimensional problems the number of experimental designs needed may grow fast and other designs, such as Latin hypercube designs, may be the more practical choice. Furthermore, as for use of metamodels in optimization representation of the true functional behavior with respect to the type and location of the extrema may be far more important than the functional value, measures other than $EISE$ may be more representative.

Though the influence of including sensitivity information into the training of kriging models has not yet been studied, it is expected to show the same trend as with RSM, where it pays off in the ‘smooth case’ whereas non-smooth or even discontinuous behavior may disturb the global behavior of the response. However, the process of finding the optimal correlation function parameters through a maximum likelihood approach may be numerically more time consuming as the correlation matrix now becomes $N(n+1)$ by $N(n+1)$ instead of N by N with regular kriging.

Table 4. Summary of $EISE$ -results for kriging models, trained on different designs.

	Smooth $a = 0.5, b = 2$		Smoothly fluctuating $a = 2, b = 2$		Strongly fluctuating $a = 0.5, b = 10$	
	$EISE_f$	θ_{opt}	$EISE_f$	θ_{opt}	$EISE_f$	θ_{opt}
Kriging (3^2 full factorial)	6.85×10^{-2}	0.22	1.34	2 ^a	1.11×10^{-1}	0.28
Kriging (4^2 full factorial)	1.85×10^{-2}	0.42	3.54×10^{-1}	1.28	1.85×10^{-2}	0.42
Kriging (5^2 full factorial)	1.68×10^{-3}	0.38	1.97×10^{-2}	0.74	2.94×10^{-4}	0.22
Kriging ($N = 9$ Latin hypercube)	1.82×10^{-1}	0.38	1.13	1.36	3.05×10^{-1}	2 ^a
Kriging ($N = 16$ Latin hypercube)	2.15×10^{-2}	0.40	2.88×10^{-1}	0.86	2.79×10^{-1}	1.68
Kriging ($N = 25$ Latin hypercube)	2.27×10^{-3}	0.40	2.99×10^{-2}	0.52	1.02×10^{-1}	1.52

^aValue used in model prediction as no optimal value was found.

4. Conclusion

Both Response Surface Models and kriging models are useful as metamodels in Engineering Optimization. RSM's are very easy to train, though an appropriate model function has to be chosen in advance, restricting flexibility of the model especially in non-smooth situations. They may be preferred when working on a global scale, e.g. for identifying promising regions in design space for optimal designs. Furthermore, polynomial functions, often used for RSM's, tend to average out non-smooth response behavior, preventing from premature convergence of optimization algorithms to local extremes. However, on a local- or mid-range scale RSM's may be too inflexible to capture detailed local functional behavior, as was illustrated for the smoothly fluctuating case. In these situations kriging may be more adequate, as all experimental training data are exactly fitted then. In fact, the model relies on the response values and therefore is far more flexible. However, as was illustrated for the strongly fluctuating case, if responses exhibit large variation on a local scale or are subject to substantial noise, interpolating the specific training data obtained may not capture the true functional behavior with respect to the type and location of the extrema. Furthermore, the procedure for training the model is less straightforward for kriging than for RSM, as correlation function parameters have to be estimated from a maximum likelihood approach. For a large number of design variables and/or large training designs this may become computationally expensive and may hinder a successful application.

The efficiency of model training can be improved by including information on design sensitivities in the model training. This holds especially for those situations where design sensitivities are easily available, as is the case with e.g. Finite Element Models. However, if the response is non-smooth or contains discontinuities sensitivity data can disturb the global behavior of the response and should not be used. Furthermore, the proposed metamodel should be flexible enough to capture the behavior of both responses and sensitivities properly. If sensitivity information can only be obtained numerically, through additional model analyses, there is no clear advantage to spend additional effort on sensitivity information instead of using it directly for gathering response information for a larger training design.

References

- A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*, Clarendon Press: Oxford, 1992.
- N. A. C. Cressie, *Statistics for Spatial Data*, J. Wiley: New York, 1993.
- N. R. Draper and H. Smith, *Applied Regression Analysis*, J. Wiley: New York, 1998.
- J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics* vol. 19, no. 1, pp. 1–141, 1991.
- A. Giunta, L. T. Watson, and J. Koehler, "A comparison of approximation modeling techniques: Polynomial versus interpolating models," in *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Paper AIAA-98-4758, 1998.
- R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodeling techniques under multiple modeling criteria," *Struct. Multidisc. Optim.* vol. 13, pp. 1–13, 2001.
- J. Johnston and J. DiNardo, *Econometric Methods*, McGraw-Hill: New York, 1997.
- J. R. Koehler and A. B. Owen, "Computer experiments," in *Handbook of Statistics* vol. 13, S. Ghosh and C. R. Rao, eds., Elsevier Science: Amsterdam, pp. 261–308, 1996.
- P. Lancaster and K. Salkauskas, *Curve and Surface Fitting: An Introduction*, Academic Press: London, 1986.
- D. C. Montgomery, *Design and Analysis of Experiments*, J. Wiley: New York, 1997.

- M. D. Morris, T. J. Mitchell, and D. Ylvisaker, "Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction," *Technometrics* vol. 35, pp. 243–255, 1993.
- R. H. Myers, *Classical and Modern Regression with Applications*, PWS-Kent: Boston, 1990.
- R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, J. Wiley: New York, 1995.
- M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review," in *Algorithms for Approximation*, J. C. Mason and M. G. Cox, eds., Oxford University Press: London, pp. 143–167, 1987.
- J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied Regression Analysis: A Research Tool*, Springer Texts in Statistics, Springer: Berlin, 1998.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science* vol. 4, pp. 409–435, 1989.
- A. J. G. Schoofs, M. B. M. Klink, and D. H. van Campen, "Approximation of structural optimization problems by means of designed numerical experiments," *Structural Optimization* vol. 4, pp. 206–212, 1992.
- A. J. G. Schoofs, M. H. van Houten, L. F. P. Etman, and D. H. van Campen, "Global and mid-range function approximation for engineering optimization," *Mathematical Methods of Operations Research* vol. 46, pp. 335–359, 1997.
- Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House Publishers, Norwood, MA, 2000.