ROBUST DESIGN WITH NONPARAMETRIC MODELS: PREDICTION OF SECOND-ORDER CHARACTERISTICS OF PROCESS VARIABILITY BY KRIGING [1]

Luc Pronzato * **Éric Thierry** *

* *Laboratoire I3S, CNRS/Université de Nice–Sophia Antipolis, Les Algorithmes, 2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France*

Abstract: We use kriging to predict the mean and variance of a response $y(\mathbf{x})$ when the input factors $\mathbf{x}$ are subject to random variability. Uncertainty on these predictions is obtained by considering fluctuations along one trajectory y of the process due to fluctuations of $\mathbf{x}$, and then averaging over the possible trajectories, conditionally on input-output data. Possible applications include robust design engineering, where the data that are obtained from prototypes in laboratory experiments, or from simulation codes, are used to construct models for the responses of interest to the designer, but mass-production involves variability of input factors around the specifications the designer will indicate.

Keywords: Kriging, Robust Design, Prediction, Computer Experiments, Process Variability, Nonparametric Modelling, Propagation of errors, Taguchi Method

1. INTRODUCTION

Assume that, based on controlled, say laboratory, or prototype, experiments, we construct a model of the response $y(\mathbf{x})$ of a system to input factors $\mathbf{x} \in \mathbb{R}^d$. Several objectives may be considered, such as y should be maximized, y should be set equal to some target T, a constraint $y \leq c$ should be satisfied, etc., that is, y may correspond either to an objective or a constraint in a multi-objective optimization problem, see for instance (Bates *et al.* 1999).

We consider the situation where the prediction $\eta(\mathbf{x})$ of $y(\mathbf{x})$ at some unobserved $\mathbf{x}$ is obtained by kriging (Krige 1951, Matheron 1963). The approach is particularly attractive for computer experiments, see, *e.g.*, (Sacks *et al.* 1989*a*, Sacks *et al.* 1989*b*, Welch *et al.* 1992), due to its flexibility and the possibility to predict model accuracy (or rather, inaccuracy) from deterministic responses: roughly, the unknown response $y(\mathbf{x})$ is described as a mean value plus the trajectory of a zero-mean stationary process, and statistical inference is made from observations of the response at given (design) points $\mathbf{x}_1, \ldots, \mathbf{x}_N$. The second-order characteristics of the process are estimated from the data $\mathbf{Y} = [y(\mathbf{x}_1), \ldots, y(\mathbf{x}_N)]$, and maximum-likelihood can be used when the process is Gaussian and its covariance is suitably parameterized. The best linear unbiased predictor for the value of $y(\mathbf{x})$ at new inputs $\mathbf{x}$ is then constructed from $\mathbf{Y}$, together with the variance of $y(\mathbf{x})$ due to model uncertainty. Classical extensions concern the case where observation errors are present (which corresponds to physical, as opposed to computer, experiments) and the mean value of the response is parameterized, with possibly a Bayesian prior on these parameters. We refer to the papers mentioned above and to (Stein 1999) for a more accurate introduction. The choice of appropriate design points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ has received little attention, see (Sacks and Schiller 1988, Shewry and Wynn 1988), but Latin Hypercube designs are generally adopted for their suitable space-filling property (it is a most attractive feature of kriging to be able to generate fairly accurate models from very

[1] This work is part of the European project 'TITOSIM' (TIme TO market reduction via Statistical Information Management), GRD1-2000-25724

few data, see (Costa *et al.* 2000) for an application in signal processing). In robust design problems, see below, this is a definite advantage over the so-called Taguchi method, see (Vuchkov and Boyadjieva 2001). When the design points are generated sequentially, the method to be used generally depends on the final objective of the model, which may for instance correspond to the optimization of the response, see (Schonlau *et al.* 1998, Bates and Pronzato 2001).

In this paper, we focuss our attention on the following robust design problem: in mass production, $\mathbf{x}$ cannot be chosen accurately and must be considered as a random variable [2]. We shall assume that its second-order characteristics are known (for instance, it may be normal with known mean and variance). This induces variability on $y(\mathbf{x})$, which we want to predict, again in terms of mean and variance. Such predictions can then be taken into account, for instance by choosing the mean value for $\mathbf{x}$ such that the mean value of $y(\mathbf{x})$ is maximized and the variability of $y(\mathbf{x})$ is minimized. More complex situations can of course be considered, leading to various multi-objective optimization problems. The crucial point here is that, starting from a single response (which may be an objective or a constraint in the original problem), we get two responses due to the variability of $\mathbf{x}$ in mass production: the mean and the variance of $y(\mathbf{x})$. Also, since the model is constructed from a finite data sample $\mathbf{Y}$, both responses are uncertain, and their variances are of interest.

Propagation of errors is standard for classical models, such as polynomials, see (Vuchkov and Boyadjieva 2001). It is the aim of the paper to show that *propagation of errors is also feasible when the model is obtained by kriging*. The predictions of the mean and variance of $y(\mathbf{x})$ when $\mathbf{x}$ varies are constructed in Section 2. Uncertainty due to estimation from a finite data sample is considered in Section 3. Illustrative examples are presented in Section 4. Throughout the paper we restrict our attention to the case of computer experiments, the extension to physical experiments where observation errors are present does not raise particular difficulties (only the covariance structure must be modified, see, *e.g.*, Costa *et al.* (2000)).

2. BAYESIAN KRIGING FOR PREDICTING MEAN AND VARIANCE

We remind the construction of the predictor $\eta(\mathbf{x})$, using a Bayesian approach (Bayesian kriging), and derive the joint posterior distribution of two responses $y(\mathbf{x}_a), y(\mathbf{x}_b)$. This is used later on to construct predictions that take the variability of $\mathbf{x}$ into account.

[2] In practise, it happens that some of the factors can still be controlled during mass production, however, this does not modify the methodology presented below.

2.1 *Bayesian kriging*

We model the observations by $Y_i = y(\mathbf{x}_i) = \beta^\top \mathbf{r}(\mathbf{x}_i) + Z(\mathbf{x}_i)$ with $Z_i = Z(\mathbf{x}_i)$ a zero mean second order stationary stochastic process and $\beta^\top \mathbf{r}(\mathbf{x})$ the deterministic part (which corresponds to *universal kriging*). The realizations Z_i, Z_j are correlated, and we define

$$V(Z_i, Z_j) = \mathtt{E}_Z\{Z(\mathbf{x}_i)Z(\mathbf{x}_j)\}\,. \tag{1}$$

Since the process is assumed to be stationary, we write

$$V(Z_i, Z_j) = \sigma_z^2 C(\mathbf{x}_i - \mathbf{x}_j)\,.$$

An usual model for the covariance $C(\cdot)$ is

$$C(\mathbf{z}) = C(\theta, \mathbf{z}) = \exp\left(-\sum_{i=1}^{d} \theta_i z_i^2\right)\,, \tag{2}$$

which gives a process $Z(\mathbf{x})$ infinitely mean square differentiable, see, *e.g.*, (Stein 1999). Other models for the covariance (Matérn class) yield exactly one, two or m times mean square differentiability [3].

We assume that the parameters β have a normal prior $\mathcal{N}(\mu, \sigma^2\Omega)$, and that the process Z is Gaussian and independent of β. We denote by $\mathbf{Y}$ the N observations $Y_1, Y_2, \ldots, Y_N$ and $\mathbf{V}$, $\mathbf{C}$ the matrices defined by $[\mathbf{V}]_{i,j} = V(Z_i, Z_j)$, $[\mathbf{C}]_{i,j} = C(\mathbf{x}_i - \mathbf{x}_j)$, $i, j = 1, \ldots, N$.

The posterior distribution of β (conditional to $\mathbf{Y}$) is normal $\mathcal{N}(\bar{\beta}, \mathbf{W})$, with

$$\begin{aligned}\mathbf{W} &= [\mathbf{R}^\top \mathbf{V}^{-1}\mathbf{R} + (\sigma^2\Omega)^{-1}]^{-1}\,,\\ \bar{\beta} &= \mathbf{W}[\mathbf{R}^\top \mathbf{V}^{-1}\mathbf{Y} + (\sigma^2\Omega)^{-1})\mu]\,,\end{aligned}$$

where the i-th row of the matrix $\mathbf{R}$ equals $\mathbf{r}^\top(\mathbf{x}_i)$.

We want to predict two responses $y_a = y(\mathbf{x}_a), y_b = y(\mathbf{x}_b)$. The joint distribution $\pi(\mathbf{Y}, y_a, y_b|\beta)$ of $\mathbf{Y}, y_a$, and y_b conditional to β is normal $\mathcal{N}(\tilde{\beta}, \tilde{\mathbf{V}})$, with

$$\tilde{\beta} = \begin{pmatrix}\mathbf{R}\\ R_{ab}\end{pmatrix}\beta\,,\ \tilde{\mathbf{V}} = \sigma^2\begin{pmatrix}\mathbf{C} & \Gamma\\ \Gamma^\top & \Sigma\end{pmatrix}$$

where

$$R_{ab} = \begin{pmatrix}\mathbf{r}^\top(\mathbf{x}_a)\\ \mathbf{r}^\top(\mathbf{x}_a)\end{pmatrix}$$

$\Sigma_{1,1} = \Sigma_{2,2} = 1, \Sigma_{1,2} = C(\mathbf{x}_a - \mathbf{x}_b), \Gamma_{i,1} = C(\mathbf{x}_i - \mathbf{x}_a)$, and $\Gamma_{i,2} = C(\mathbf{x}_i - \mathbf{x}_b)$, $i = 1, \ldots, N$.

From this we can compute the conditional

$$\pi(y_a, y_b|\beta, \mathbf{Y}) = \frac{\pi(\mathbf{Y}, y_a, y_b|\beta)}{\pi(\mathbf{Y}|\beta)}$$

which is normal $\mathcal{N}(\tilde{\mathbf{y}}, \sigma_z^2\mathbf{H})$ with

[3] Note, however, that the analytic properties of the sample function, that is of individual trajectories, are not necessarily related to mean square properties. This is considered in (Cramér and Leadbetter 1967), Chapter 5.

$$\tilde{\mathbf{y}}(\beta) = \Gamma^\top \mathbf{C}^{-1}\mathbf{Y} + \mathbf{U}^\top \beta\,,$$
$$\mathbf{H} = \Sigma - \Gamma^\top \mathbf{C}^{-1}\Gamma\,,$$

where

$$\mathbf{U}^\top = [R_{ab} - \Gamma^\top \mathbf{C}^{-1}\mathbf{R}]\,.$$

The joint $\pi(y_a, y_b|\mathbf{Y})$ is finally obtained by

$$\pi(y_a, y_b|\mathbf{Y}) = \int \pi(y_a, y_b|\beta, \mathbf{Y})\pi(\beta|\mathbf{Y})d\beta\,,$$

which gives after some calculation the normal $\mathcal{N}(\eta_{ab}, \sigma_z^2\mathbf{H} + \mathbf{U}^\top\mathbf{W}\mathbf{U})$, with $\eta_{ab} = \tilde{\mathbf{y}}(\bar{\beta})$.

We return to non-Bayesian kriging by letting σ^2 tend to infinity (that is, using a non informative prior for β). This gives

$$(y_a, y_b) \sim \mathcal{N}(\eta_{ab}, \sigma_z^2\mathbf{P})$$

with

$$\eta_{ab} = \{\Gamma^\top \mathbf{C}^{-1} + \mathbf{U}^\top[\mathbf{R}^\top \mathbf{C}^{-1}\mathbf{R}]^{-1}[\mathbf{R}^\top \mathbf{C}^{-1}]\}\mathbf{Y}\,,$$
$$\mathbf{P} = \Sigma - \Gamma^\top \mathbf{C}^{-1}\Gamma + \mathbf{U}^\top[\mathbf{R}^\top \mathbf{C}^{-1}\mathbf{R}]^{-1}\mathbf{U}\,.$$

Note that the values of σ_z^2 and the parameters θ that appear in $C(\mathbf{z})$, see *e.g.* (2), can be estimated from the data, for instance by maximum likelihood when the process is assumed to be Gaussian.

In what follows we shall only consider the case of simple kriging, where $\beta = \beta_0 \in I\!R$, $\mathbf{R} = \mathbf{1}_N$ (the N dimensional vector with all components equal to 1) and $R_{ab} = \mathbf{1}_2$. Other situations could be treated similarly. This gives

$$\mathbf{P}_{1,1} = 1 - \mathbf{c}_a^\top \mathbf{C}^{-1}\mathbf{c}_a + \frac{(1 - \mathbf{c}_a^\top\gamma)^2}{S}\,,$$
$$\mathbf{P}_{2,2} = 1 - \mathbf{c}_b^\top \mathbf{C}^{-1}\mathbf{c}_b + \frac{(1 - \mathbf{c}_b^\top\gamma)^2}{S}\,,$$
$$\mathbf{P}_{1,2} = C(\mathbf{x}_a - \mathbf{x}_b) - \mathbf{c}_a^\top \mathbf{C}^{-1}\mathbf{c}_b + \frac{(1 - \mathbf{c}_a^\top\gamma)(1 - \mathbf{c}_b^\top\gamma)}{S}\,,$$

where $\mathbf{c}_a = (C(\mathbf{x}_a - \mathbf{x}_1), \ldots, C(\mathbf{x}_a - \mathbf{x}_N))^\top$, $\mathbf{c}_b = (C(\mathbf{x}_b - \mathbf{x}_1), \ldots, C(\mathbf{x}_b - \mathbf{x}_N))^\top$, $\gamma = \mathbf{C}^{-1}\mathbf{1}_N$ and $S = \sum_{i,j=1}^N [\mathbf{C}^{-1}]_{i,j}$. Note that $P_{1,1} = P_{1,2} = 0$ if $\mathbf{x}_a = \mathbf{x}_i$, the inputs used for the observation Y_i, $i = 1, \ldots, N$.

When one is only interested into prediction at some particular point $\mathbf{x}$, only the first component of η_{ab} has to be considered, which corresponds to $\eta(\mathbf{x})$, and the $(1,1)$ component of $\mathbf{P}$ gives the uncertainty about this prediction.

2.2 *Models for mean and variance*

We consider now $\mathbf{x}$ as a random variable, with $\mathtt{E}_x\{\cdot\}$ the expectation with respect to $\mathbf{x}$, and denote

$$\bar{\mathbf{x}} = \mathtt{E}_x\{\mathbf{x}\}\,,\quad \Sigma_x = \mathtt{E}_x\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^\top\}\,.$$

In a robust design problem, both $\bar{\mathbf{x}}$ and Σ_x may depend on factors that have to be settled by the designer.

From the results above, $\eta(\mathbf{x})$ can be considered as the mean of $y(\mathbf{x})$ conditional on $\mathbf{Y}$, under a noninformative prior for β, that is,

$$\mathtt{E}_y\{y(\mathbf{x})|\mathbf{Y}\} = \eta(\mathbf{x})\,.$$

A naive approach for taking variability of $\mathbf{x}$ into account is to use a Taylor series development of $\eta(\mathbf{x})$ at $\bar{\mathbf{x}}$:

$$\eta(\mathbf{x}) = \eta(\bar{\mathbf{x}}) + (\mathbf{x} - \bar{\mathbf{x}})^\top \frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}}_{|\bar{\mathbf{x}}} + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^\top \frac{\partial^2\eta(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}(\mathbf{x} - \bar{\mathbf{x}}) + HOT$$

which gives

$$\mathtt{E}_x\{\eta(\mathbf{x})\} = \eta(\bar{\mathbf{x}}) + \frac{1}{2}\mathrm{trace}\left[\frac{\partial^2\eta(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\right] + HOT$$

and, denoting Var_x the variance with respect to $\mathbf{x}$,

$$\mathrm{Var}_x\{\eta(\mathbf{x})\} = \frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}}_{|\bar{\mathbf{x}}} + HOT$$

However, a more careful analysis shows that this approach is inexact. Moreover, it does not permit to take uncertainty on the prediction into account, that is, to derive variance models for $\mathtt{E}_x\{\eta(\mathbf{x})\}$ and $\mathrm{Var}_x\{\eta(\mathbf{x})\}$, as it will be done in Section 3.

In fact, when $\mathbf{x}$ fluctuates, the true response, that is, the trajectory, remains the same for different values of $\mathbf{x}$. We assume that the fourth derivative of $C(\cdot)$ at zero exists and is finite and consider a process

$$y_2(\mathbf{x}) = y(\bar{\mathbf{x}}) + (\mathbf{x} - \bar{\mathbf{x}})^\top G(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^\top H(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})\,,$$

with suitable means and covariances for the processes $G(\cdot)$ and $H(\cdot)$, such that

$$y_{h,h',h'',h'''}(\mathbf{x}) = y(\bar{\mathbf{x}}) + \sum_i (\mathbf{x} - \bar{\mathbf{x}})_i \frac{y(\bar{\mathbf{x}} + h_i\mathbf{e}_i) - y(\bar{\mathbf{x}})}{h_i} + \frac{1}{2}\sum_{i,j}(\mathbf{x} - \bar{\mathbf{x}})_i(\mathbf{x} - \bar{\mathbf{x}})_j \frac{1}{h_j'''}\left[\frac{y(\bar{\mathbf{x}} + h_i'\mathbf{e}_i) - y(\bar{\mathbf{x}})}{h_i'} - \frac{y(\bar{\mathbf{x}} + h_i''\mathbf{e}_i - h_j'''\mathbf{e}_j) - y(\bar{\mathbf{x}} - h_j'''\mathbf{e}_j)}{h_i''}\right]$$

tends to $y_2(\mathbf{x})$ in quadratic mean when $h, h', h'', h''' \to 0$:

$$\lim_{h,h',h'',h'''\to 0} E_y\{[y_2(\mathbf{x}) - y_{h,h',h'',h'''}(\mathbf{x})]^2\} = 0\,,$$

see (Stein 1999), Chapter 2. This gives

$$\mathtt{E}_x\{y_2(\mathbf{x})\} = y(\bar{\mathbf{x}}) + \frac{1}{2}\,\mathrm{trace}\,[H(\bar{\mathbf{x}})\Sigma_x]$$

and

$$\text{Var}_x\{y_2(\mathbf{x})\} = G^\top(\bar{\mathbf{x}})\Sigma_x G(\bar{\mathbf{x}}) + HOT$$

where the higher order terms HOT will be neglected. Next step is to consider expectation with respect to the processes, that is, with respect to possible trajectories, conditional on the observations. We get

$$\begin{aligned}
&\mathtt{E}_y\{\mathtt{E}_x\{y_2(\mathbf{x})\}|\mathbf{Y}\} \\
&= \eta(\bar{\mathbf{x}}) + \frac{1}{2}\,\text{trace}\,[\mathtt{E}_y\{H(\bar{\mathbf{x}})|\mathbf{Y}\}\Sigma_x] \\
&= \eta(\bar{\mathbf{x}}) + \frac{1}{2}\,\text{trace}\left[\frac{\partial^2\eta(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\right] \qquad (3)
\end{aligned}$$

for the prediction of the mean response over $\mathbf{x}$. Note that it is larger or smaller than the prediction at the mean value $\bar{\mathbf{x}}$ depending on the sign of the second order derivative of the predictor $\eta(\cdot)$ at this point.

For the expected variance, neglecting terms of order in $\mathbf{x}$ higher than two, we consider

$$\begin{aligned}
&\mathtt{E}_y\{\text{Var}_x\{y_{h,h'}(\mathbf{x})\}|\mathbf{Y}\} = \\
&\sum_{i,j}[\Sigma_x]_{i,j}\,\mathtt{E}_y\left\{\frac{[y(\bar{\mathbf{x}}+h_i\mathbf{e}_i)-y(\bar{\mathbf{x}})]}{h_i}\right. \\
&\left.\times\,\frac{[y(\bar{\mathbf{x}}+h'_j\mathbf{e}_j)-y(\bar{\mathbf{x}})]}{h'_j}\right\}.
\end{aligned}$$

Using the results of Section 2.1 on the joint posterior distribution of $[y(\bar{\mathbf{x}}+h_i\mathbf{e}_i), y(\bar{\mathbf{x}})]$, $[y(\bar{\mathbf{x}}+h'_j\mathbf{e}_j), y(\bar{\mathbf{x}})]$ and $[y(\bar{\mathbf{x}}+h_i\mathbf{e}_i), y(\bar{\mathbf{x}}+h'_j\mathbf{e}_j)]$, we get

$$\begin{aligned}
&\lim_{h\to 0,h'\to 0}\mathtt{E}_y\left\{\frac{[y(\bar{\mathbf{x}}+h_i\mathbf{e}_i)-y(\bar{\mathbf{x}})]}{h_i}\right. \\
&\left.\times\,\frac{[y(\bar{\mathbf{x}}+h'_j\mathbf{e}_j)-y(\bar{\mathbf{x}})]}{h'_j}\right\} = \\
&\frac{\partial\eta(\mathbf{x})}{\partial x_i}_{|\bar{\mathbf{x}}}\frac{\partial\eta(\mathbf{x})}{\partial x_j}_{|\bar{\mathbf{x}}} - \sigma_z^2\frac{\partial C(\mathbf{z})}{\partial z_i\partial z_j}_{|\mathbf{0}} \\
&-\sigma_z^2\frac{\partial\mathbf{c}^\top(\mathbf{z})}{\partial z_i}_{|\bar{\mathbf{x}}}\left(\mathbf{C}^{-1}-\frac{\gamma\gamma^\top}{S}\right)\frac{\partial\mathbf{c}(\mathbf{z})}{\partial z_j}_{|\bar{\mathbf{x}}},
\end{aligned}$$

where $\mathbf{c}(\mathbf{x}) = (C(\mathbf{x}-\mathbf{x}_1),\ldots,C(\mathbf{x}-\mathbf{x}_N))^\top$.

This gives the following prediction for the variance:

$$\begin{aligned}
&\mathtt{E}_y\{\text{Var}_x\{y_2(\mathbf{x})\}|\mathbf{Y}\} = \frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}}_{|\bar{\mathbf{x}}} \\
&-\sigma_z^2\,\text{trace}\left[\frac{\partial^2 C(\mathbf{z})}{\partial\mathbf{z}\partial\mathbf{z}^\top}_{|\mathbf{0}}\Sigma_x\right] \qquad (4) \\
&-\sigma_z^2\,\text{trace}\left[\frac{\partial\mathbf{c}^\top(\mathbf{z})}{\partial\mathbf{z}}_{|\bar{\mathbf{x}}}\left(\mathbf{C}^{-1}-\frac{\gamma\gamma^\top}{S}\right)\frac{\partial\mathbf{c}(\mathbf{z})}{\partial\mathbf{z}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\right].
\end{aligned}$$

It tends to the value $\text{Var}_x\{\eta(\mathbf{x})\} \simeq \partial\eta(\mathbf{x})/\partial\mathbf{x}^\top_{|\bar{\mathbf{x}}}\Sigma_x\,\partial\eta(\mathbf{x})/\partial\mathbf{x}_{|\bar{\mathbf{x}}}$ of the prediction of the variance for the deterministic model when σ_z tends to zero, and, of course, it tends to zero when Σ_x tends to zero. Note that the second term is usually positive, whereas the third one is negative, so that situations may exist where $\mathtt{E}_y\{\text{Var}_x\{y_2(\mathbf{x})\}|\mathbf{Y}\}$ is smaller than $\text{Var}_x\{\eta(\mathbf{x})\}$.

3. MODEL UNCERTAINTY

Uncertainty on the true response due to the fact that the data sample is finite induces uncertainty in the predictions (3) and (4). Using the same approach as in Section 2, we obtain after some (lengthy) calculations

$$\begin{aligned}
&\text{Var}_y\{\mathtt{E}_x\{y_2(\mathbf{x})\}|\mathbf{Y}\} = \\
&\sigma_z^2\left[1-\mathbf{c}^\top(\bar{\mathbf{x}})\mathbf{C}^{-1}\mathbf{c}(\bar{\mathbf{x}}) + \frac{(1-\mathbf{c}^\top(\bar{\mathbf{x}})\gamma)^2}{S}\right] \\
&+\frac{\sigma_z^2}{4}\sum_{i,j,k,l=1}^{d}[\Sigma_x]_{i,j}[\Sigma_x]_{k,l}\left[\frac{\partial^4 C(\mathbf{z})}{\partial z_i\partial z_j\partial z_k\partial z_l}_{|\mathbf{0}}\right. \\
&\left.-\frac{\partial^2\mathbf{c}^\top(\mathbf{z})}{\partial z_i\partial z_j}_{|\bar{\mathbf{x}}}\left(\mathbf{C}^{-1}-\frac{\gamma\gamma^\top}{S}\right)\frac{\partial^2\mathbf{c}(\mathbf{z})}{\partial z_k\partial z_k}_{|\bar{\mathbf{x}}}\right] \\
&+\sigma_z^2\sum_{i,j=1}^{d}[\Sigma_x]_{i,j}\left[\frac{\partial^2 C(\mathbf{z})}{\partial z_i\partial z_j}_{|\mathbf{0}}\right. \\
&\left.-\left(\mathbf{c}^\top(\bar{\mathbf{x}})\mathbf{C}^{-1}+\frac{1-\mathbf{c}^\top(\bar{\mathbf{x}})\gamma}{S}\gamma^\top\right)\frac{\partial^2\mathbf{c}(\mathbf{z})}{\partial z_i\partial z_j}_{|\bar{\mathbf{x}}}\right]
\end{aligned}$$

for the variance of the mean response. Note that it tends to the variance of $y(\bar{\mathbf{x}})$ when Σ_x tends to zero. Also, the second correcting term is negligible with respect to the third when Σ_x is small.

Similarly, we get for the variance of the variance

$$\begin{aligned}
&\text{Var}_y\{\text{Var}_x\{y(\mathbf{x})\}|\mathbf{Y}\} = 2\sigma_z^4\,\text{trace}\left(\left\{\Sigma_x\left[\frac{\partial^2 C(\mathbf{z})}{\partial\mathbf{z}\partial\mathbf{z}^\top}_{|\mathbf{0}}\right.\right.\right. \\
&\left.\left.\left.+\,\frac{\partial\mathbf{c}^\top(\mathbf{z})}{\partial\mathbf{z}}_{|\bar{\mathbf{x}}}\left(\mathbf{C}^{-1}-\frac{\gamma\gamma^\top}{S}\right)\frac{\partial\mathbf{c}(\mathbf{z})}{\partial\mathbf{z}^\top}_{|\bar{\mathbf{x}}}\right]\right\}^2\right) \\
&-\,4\sigma_z^2\frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}^\top}_{|\bar{\mathbf{x}}}\Sigma_x\left[\frac{\partial^2 C(\mathbf{z})}{\partial\mathbf{z}\partial\mathbf{z}^\top}_{|\mathbf{0}}+\frac{\partial\mathbf{c}^\top(\mathbf{z})}{\partial\mathbf{z}}_{|\bar{\mathbf{x}}}\right. \\
&\left.\times\left(\mathbf{C}^{-1}-\frac{\gamma\gamma^\top}{S}\right)\frac{\partial\mathbf{c}(\mathbf{z})}{\partial\mathbf{z}^\top}_{|\bar{\mathbf{x}}}\right]\Sigma_x\frac{\partial\eta(\mathbf{x})}{\partial\mathbf{x}}_{|\bar{\mathbf{x}}}.
\end{aligned}$$

Notice that we ignored the fact that in practise the parameters θ and σ_z^2 in the covariance model, (2) for instance, are estimated values. Taking uncertainty on these estimates into account in the evaluation of predictions is a challenging but difficult task.

4. EXAMPLES

Consider a one-dimensional process ($x \in I\!R$), with zero mean and covariance $\exp(-0.2z^2)$ ($\sigma_z = 1$). We observe the responses at the design points $0, 2, 3, 4, 7$ and 10. Figure 1 presents a typical realization of the process (full line), with observed values $\mathbf{Y}$ indicated by stars, and the associated predicted response (dashed

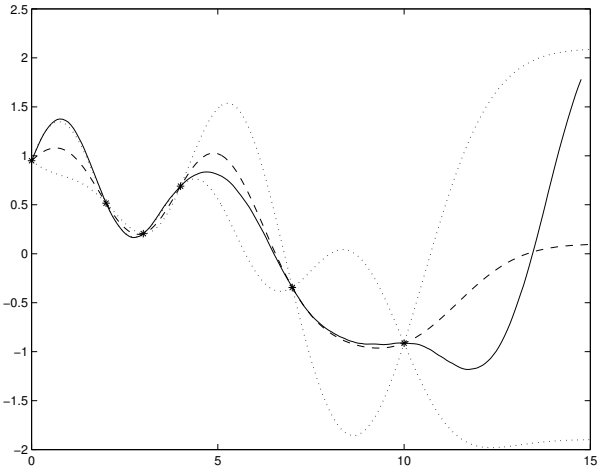

Fig. 1. Typical realization of the process (full line), observations (stars), prediction (dashed lines) and 2σ confidence bounds (dotted lines)

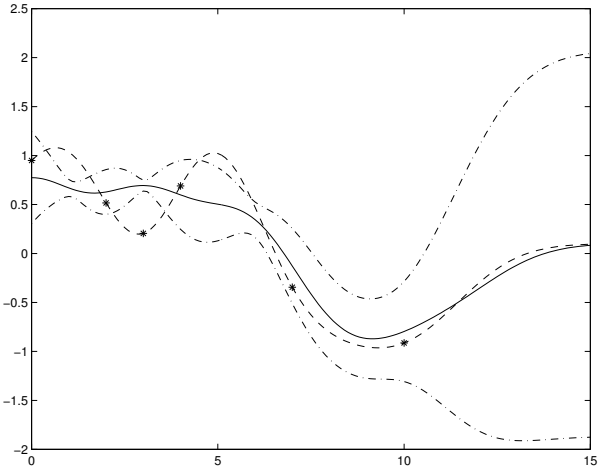

Fig. 2. Prediction $\eta(\bar{x})$ (dashed line), observations (stars), prediction $\mathtt{E}_y\{\mathtt{E}_x\{y_2(x)\}|\mathbf{Y}\}$ (full line) and 2σ confidence bounds (dash-dotted lines)

line) together with 2σ confidence bounds around this prediction (dotted lines). Notice how the uncertainty increases with the distance to observation points.

Assume now that x is normally distributed $\mathcal{N}(\bar{x}, 0.01)$.

Figure 2 presents the evolution of the prediction $\eta(\bar{x})$ at $\bar{x}$ (dashed line, as in Figure 1) and the prediction of the mean response $\mathtt{E}_y\{\mathtt{E}_x\{y_2(x)\}|\mathbf{Y}\}$ (full line), together with 2σ confidence bounds obtained from $\mathrm{Var}_y\{\mathtt{E}_x\{y_2(x)\}|\mathbf{Y}\}$ (dash-dotted lines), when $\bar{x}$ varies from 0 to 15.

A first observation is that averaging with respect to x smoothes the prediction, compare $\mathtt{E}_y\{\mathtt{E}_x\{y_2(x)\}|\mathbf{Y}\}$ to $\eta(\bar{x})$. Next, there is no clear relation between the distance to observation points and uncertainty of the prediction of the mean response: for instance, the uncertainty is close to a local maximum at the observed point $x = 2$, but close to local minimum at $x = 3$.

Figure 3 gives the evolution of the prediction of the variance $\mathtt{E}_y\{\mathrm{Var}_x\{y_2(x)\}|\mathbf{Y}\}$ (full line) together

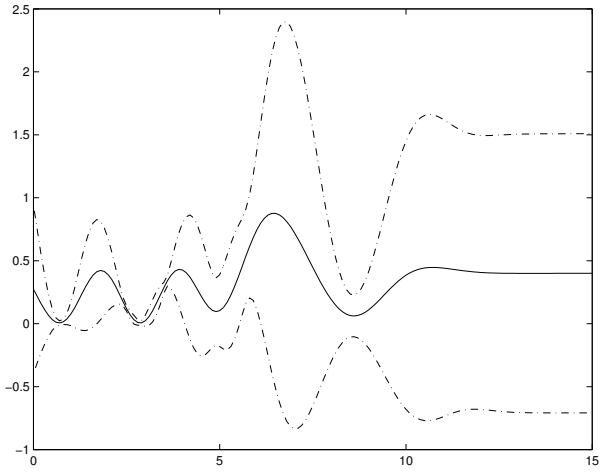

Fig. 3. Prediction $\mathtt{E}_y\{\mathrm{Var}_x\{y_2(x)\}|\mathbf{Y}\}$ (full line) and 2σ confidence bounds (dash-dotted lines)

with 2σ confidence bounds obtained from $\mathrm{Var}_y\{\mathrm{Var}_x\{y_2(x)\}|\mathbf{Y}\}$ (dash-dotted lines) [4] .

Observe that the prediction of the variance tends to be small when $\eta(x)$ is close to a stationary point, and to be large when the slope of $\eta(x)$ is large. The uncertainty on the variance tends to be small when the variance itself is small. If, for instance, the purpose were to maximize the expected response and minimize the variance, the point $x = 1$ would be a good compromise.

5. REFERENCES

Bates, R. and L. Pronzato (2001). Emulator-based global optimisation using lattices and Delaunay tesselation. In: *Proc. 3rd Int. Symp. on Sensitivity Analysis of Model Output* (P. Prado and R. Bolado, Eds.). Madrid. pp. 189–192.

Bates, R., R. Fontana, L. Pronzato and H.P. Wynn (1999). Computer experiments for concurrent engineering. In: *Proc. 6th Int. Conf. ATA on the New Role of Experimentation in the Modern Automotive Product Development Process*. Firenze. pp. 1585–1588.

Costa, J.-P., L. Pronzato and E. Thierry (2000). Nonlinear prediction by kriging, with application to noise cancellation. *Signal Processing* **80**, 553–566.

Cramér, H. and M.R. Leadbetter (1967). *Stationary and Related Stochastic Processes*. Wiley. New York.

Krige, D.G. (1951). A statistical approach to some mine valuation and allied problems on the Witwatersrand. Master Thesis, University of Witwatersrand.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246–1266.

[4] Since the variance is a positive quantity, only the region above 0 gives an indication of the uncertainty on the prediction.

Sacks, J. and S. Schiller (1988). Spatial designs. In: *Statistical Decision Theory and Related Topics IV* (S.S. Gupta and J.O. Berger, Eds.). Vol. 2. pp. 385–399. Springer. Heidelberg.

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

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

Schonlau, M., W.J. Welch and D.R. Jones (1998). Global versus local search in constrained optimization of computer models. In: *New Developments and Applications in Experimental Design, Lecture Notes — Monograph Series, vol. 34*. pp. 11–25. IMS. Hayward.

Shewry, M.C. and H.P. Wynn (1988). Maximum entropy sampling with application to simulation codes. In: *Prep. 12th IMACS World Congress on Scientific Computation*. Paris. pp. 517–519.

Stein, M.L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. Springer. Heidelberg.

Vuchkov, I.N. and L.N. Boyadjieva (2001). *Quality Improvement with Design of Experiments. A Response Surface Approach*. Kluwer. Dordrecht.

Welch, W.J., R.J. Buck, J. Sacks, H.P. Wynn, T.J. Mitchell and M.D. Morris (1992). Screening, predicting and computer experiments. *Technometrics* **34**(1), 15–25.