# Assessment of uncertainty in computer experiments, from Universal to Bayesian Kriging.

C. Helbert[*], D. Dupuy[*] and L. Carraro[*]

July 30, 2008

### Abstract

Kriging was first introduced in the field of geostatistics. Nowadays, it is widely used to model computer experiments. Since the results of deterministic computer experiments have no experimental variability Kriging is appropriate in that it interpolates observations at data points. Moreover Kriging quantifies prediction uncertainty which plays a major role in many applications. Among practitioners we can distinguish those who use Universal Kriging where the parameters of the model are estimated and those who use Bayesian Kriging where model parameters are random variables. The aim of this paper is to show that the prediction uncertainty has a correct interpretation only in the case of Bayesian Kriging. Different cases of prior distributions have been studied and it is shown that in one specific case, Bayesian Kriging supplies an interpretation as a conditional variance for the prediction variance provided by Universal Kriging. Finally, a simple petroleum engineering case study presents the importance of prior information in the Bayesian approach.

**Keywords**: Universal Kriging, Bayesian Kriging, Informative Prior, Gaussian Random Field, Computer Experiments, Markov chain Monte-Carlo.

# 1    Introduction

Kriging is widely used to model computer experiments ([1], [2]). Since the results of deterministic computer experiments have no experimental variability Kriging is appropriate in that it interpolates observations at data points. But the major advantage of Kriging is its ability to quantify the prediction uncertainty which plays

---

[*]Ecole Nationale Supérieure des Mines de Saint-Etienne, Département 3MI, 158 Cours Fauriel - 42023 Saint-Etienne (France) e-mail: name@emse.fr

a major role in many applications from uncertainty propagation [3] to global optimisation [4]. These works are often based on the traditional "plug-in Kriging", also called Universal Kriging in reference of the initial work from geostatistics. However, several publications ([5], [6], [7] and [8]) illustrate that the traditional "plug-in Kriging" underestimate the true variance. Indeed it does not take into account the uncertainty that comes from the covariance parameters estimation. The authors of [5] show that the correct approach is to use the Bayesian framework. Several supplemental discussions deal with the effects of the priors ([9], [10]).

From these earlier perspectives, the aim of the article is to emphasize that Bayesian Kriging supplies an interpretation as a conditional variance for the prediction variance provided by Universal Kriging (prior information on the mean only). This article is also a way to study the impact on prediction of different standard prior distributions (conjugate and non informative) based on both the location parameter as well as covariance parameters and to validate theoretically the simulations which are used in the last part of this article. Our main contribution is to illustrate the interest of considering informative priors derived from physics to conduct Bayesian analyses instead of considering traditional non informative or conjugate priors. The article is divided in 3 parts. Section 2 recalls the equations of Universal Kriging (predictor and prediction variance) together with the expressions of the parameters estimators obtained by maximum likelihood. In section 3 a link is made between Universal Kriging and Bayesian Kriging by studying different traditional prior distributions and using simulation results. A case study is developed in section 4 where Bayesian Kriging with different informative prior distributions is computed, and the accuracy of predictions are compared.

## 2   Prediction uncertainty in Universal Kriging

Let $D$ be included in $\mathbb{R}^k$, $k \geq 1$. We suppose that the output $y$ is a function of $x \in D$. We assume that $y$ is the realization of a Gaussian random field $(Y(x))_{x \in D}$ such that:

$$E\left(Y(x)\right) = f(x)^T \beta \tag{1}$$

$$\text{and } Cov\left(Y(x), Y(x+h)\right) = \sigma^2 R(h|\theta) \tag{2}$$

where $f(x) = \left(f_0(x) \ldots f_p(x)\right)^T$ is a known trend vector, $\beta = \left(\beta_0 \ldots \beta_p\right)^T$ is the vector of trend parameters, $R(.|\theta)$ is a correlation function and $\theta = \left(\theta_1 \ldots \theta_k\right)^T$ is the vector of correlation parameters, often called the range parameter since its unit is homogeneous to a distance.

Note the Gaussian spatial correlation function is used for the examples of sections 3 and 4 which is defined by

$$\forall h \in \mathbb{R}^k, \ R(h|\theta) = \exp\left(-\sum_{i=1}^{k}\left(\frac{h_i}{\theta_i}\right)^2\right).$$

This choice supposes that the output is an infinitely differentiable function. Depending on the characteristics of the studied response, other correlation functions such as spherical or exponential could be used.

Furthermore, let $y = (y_1 \ldots y_n)^T$ be the output observed at locations $X = (x_1 \ldots x_n)^T$.

In the case where all the parameters of the model (trend, range and variance) are known, the kriging predictor[1], also called Simple Kriging $Y_{SK}(x_0)$, and the prediction variance $\sigma_{SK}^2(x_0)$ at a new location $x_0$ are given by ([11], [2]):

$$Y_{SK}(x_0) = f(x_0)^T \beta + r_\theta^T R_\theta^{-1}(y - F\beta) \tag{3}$$

$$\text{and } \sigma_{SK}^2(x_0) = \sigma^2\left(1 - r_\theta^T R_\theta^{-1} r_\theta\right) \tag{4}$$

where

$$R_\theta = \left(R\left(x_i - x_j|\theta\right)\right)_{1 \le i,j \le n}$$

$$r_\theta = \left(R_\theta\left(x_0, x_1\right) \ldots R_\theta\left(x_0, x_n\right)\right)^T$$

$$F = \left(f\left(x_1\right) \ldots f\left(x_n\right)\right)^T$$

Here, from the theoretical point of view, the predictor and its variance can be interpreted as the expectation and the variance of $Y(x_0)$ conditional on the observations.

However, in practice, the parameters of the external trend and/or those of the covariance function are not known. They are usually estimated through the optimization of a criterion, such that maximum likelihood (ML) or cross-validation. The kriging predictor $Y_{UK}(x_0)$, called Universal Kriging, and its prediction variance $\sigma_{UK}^2(x_0)$ are then modified to take into account parameter estimation. Using ML estimation, the expressions of kriging predictor and prediction variance are ([11], [2]):

$$Y_{UK}(x_0) = f(x_0)^T \hat{\beta}_{ML} + r_{\hat{\theta}_{ML}}^T R_{\hat{\theta}_{ML}}^{-1}\left(y - F\hat{\beta}_{ML}\right) \tag{5}$$

---

[1]The kriging predictor and the studied output, as functions of space variables, will be sometimes called *surface* hereafter.

and

$$\sigma^2_{UK}(x_0) = \hat{\sigma}^2 \left[ 1 - r^T_{\hat{\theta}_{ML}} R^{-1}_{\hat{\theta}_{ML}} r_{\hat{\theta}_{ML}} \right.$$
$$+ \left( f(x_0)^T - r^T_{\hat{\theta}_{ML}} R^{-1}_{\hat{\theta}_{ML}} F \right) \left( F^T R^{-1}_{\hat{\theta}_{ML}} F \right)^{-1} \left. \left( f(x_0)^T - r^T_{\hat{\theta}_{ML}} R^{-1}_{\hat{\theta}_{ML}} F \right)^T \right] \quad (6)$$

where the parameters are obtained by solving recursively the following simultaneous equations:

$$\hat{\beta}_{ML} = \left( F^T R^{-1}_{\theta} F \right)^{-1} F^T R^{-1}_{\theta} y$$

$$\hat{\sigma}^2_{ML} = \frac{\left( y - F\hat{\beta}_{ML} \right)^T R^{-1}_{\theta} \left( y - F\hat{\beta}_{ML} \right)}{n - (p+1)}$$

$$\hat{\theta}_{ML} = \text{argmin} \left[ \frac{n}{2} + \frac{n}{2} \log \left( 2\pi\hat{\sigma}^2_{ML} \right) + \frac{1}{2} \log \left( \det R_{\theta} \right) \right]$$

It can be noted that $Y_{UK}(x_0)$ of expression (5) is obtained by substituting $\beta$ by its estimate in (3). Besides, the variance of Universal Kriging (6) is larger than Simple Kriging (4) since uncertainty on $\beta$ is included. Unfortunately, expressions (5) and (6) can not be interpreted as conditional expectation and variance. Indeed, the probability law of $\hat{\theta}_{ML}$ is not known, this is therefore the same for $\hat{\beta}_{ML}$ and $\hat{\sigma}^2_{ML}$ whose expressions depend on $\hat{\theta}_{ML}$. Besides, expression (6) considers neither uncertainty due to covariance parameters estimation nor to variance estimation.

The following part emphasises the already known results that only the Bayesian context allows interpreting $\sigma^2_{UK}(x_0)$ as a conditional variance ([1]). Moreover this section recalls the impact on prediction of different standard prior distributions (conjugate and non informative) based on the location parameter as well as on the covariance parameters. Finally it validates the simulations used in the last part of this article.

# 3 The Bayesian approach to interpreting Universal Kriging's prediction variance

This section will be illustrated with the set of data of [12], where the output is the temperature of a chemical reaction (Figure 1). The mass ratio of oxidant to fuel being burned (the input) is increased from no oxidant to a surfeit of oxidant. In this process, the reaction increases in temperature to a maximum and then decreases as excess oxidant is added. The output is observed on 11 regularly
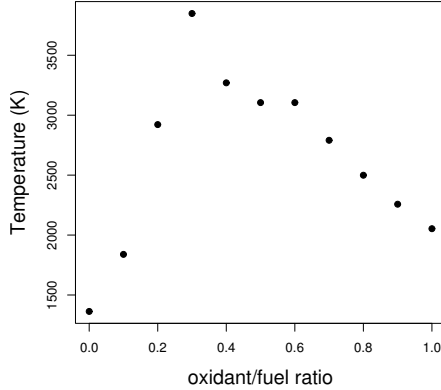
4

Figure 1: Data set

distributed values on the interval $[0, 1]$.

From here, we will assume that $(Y(x)|\beta, \sigma^2, \theta)_{x \in D}$ is a Gaussian random field such that expectation and spatial covariance function are equal to $f(x)^T \beta$ and $\sigma^2 R(.|\theta)$ respectively. Moreover, model parameters are considered random with prior joint density denoted by $\pi$.

In this Bayesian context, the predicted value at any point $x$ of domain $D$ and the prediction variance are simply given by the expectation and the variance of the output conditional on the observations, *i.e.* $E(Y(x)|y)$ and $\text{Var}(Y(x)|y)$.

The Bayesian rules give the following general expression for any function $g$:

$$E(g(Y(x))|y) = \int_\beta \int_{\sigma^2} \int_\theta E(g(Y(x))|y, \beta, \sigma^2, \theta) \pi(\beta, \sigma^2, \theta|y) \, d\beta d\sigma^2 d\theta \quad (7)$$

The conditional variance, the conditional density etc. derive from this formula.

In the right term of expression (7) one can recognize the Simple Kriging: $Y(x)|y, \beta, \sigma^2, \theta$ is indeed a Gaussian random variable with mean $Y_{SK}(x)$ and variance $\sigma_{SK}^2(x)$.

## 3.1 Known variance, known correlation parameters and conjugate prior for trend parameters.

This case is interesting because analytical calculations can be conducted when the prior law of trend parameters is assumed to be Gaussian.
Let $\beta$ be a Gaussian random vector with mean $\mu$ and variance $\lambda \Sigma$, where $\lambda$ is a positive scalar and $\Sigma$ is a symmetric definite positive matrix with its maximum

eigenvalue equal to 1. The posterior distribution of $\beta$ is also Gaussian with the following parameters:

$$E\left(\beta|y\right) = \mu + \lambda\Sigma F^T\left(\lambda F\Sigma F^T + \sigma^2 R_\theta\right)^{-1}\left(y - F\mu\right)$$

$$\mathrm{Var}\left(\beta|y\right) = \lambda\Sigma - \lambda^2\Sigma F^T\left(\lambda F\Sigma F^T + \sigma^2 R_\theta\right)^{-1} F\Sigma$$

As mentioned before, $Y\left(x_0\right)|y,\beta$ is Gaussian with the same parameters as the Simple Kriging:

$$E\left(Y\left(x_0\right)|y,\beta\right) = \left(f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F\right)\beta + r_\theta^T R_\theta^{-1} y$$

$$Var\left(Y\left(x_0\right)|y,\beta\right) = \sigma^2\left(1 - r_\theta^T R_\theta^{-1} r_\theta\right)$$

The posterior distribution for the output is also Gaussian:

$$E\left(Y\left(x_0\right)|y\right) = \left(f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F\right)\left[\mu + \Sigma F^T\left(\lambda F\Sigma F^T + \sigma^2 R_\theta\right)^{-1}\left(y - F\mu\right)\right] + r_\theta^T R_\theta^{-1} y$$

$$Var\left(Y\left(x_0\right)|y\right) = \left(f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F\right)\left[\lambda\Sigma - \lambda^2\Sigma F^T\left(\lambda F\Sigma F^T + \sigma^2 R_\theta\right)^{-1} F\Sigma\right]\left(f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F\right)^T$$

Two particular cases can be noticed.
Firstly, when $\lambda$ tends to zero, we obtain the equations of Simple Kriging given by (3) and (4).

Secondly, when $\lambda$ tends to infinity, the first moments of the posterior distribution of $\beta$ tend to the expectation and the variance of the ML estimator (see Figure 2):

$$E\left(\beta|y\right)_{\lambda\to\infty} = \left(F^T R_\theta^{-1} F\right)^{-1} F^T R_\theta^{-1} y = \hat{\beta}_{ML}$$

$$\mathrm{Var}\left(\beta|y\right)_{\lambda\to\infty} = \sigma^2\left(F^T R_\theta^{-1} F\right)^{-1} = \mathrm{Var}\left(\hat{\beta}_{ML}\right)$$

Figure 2 presents the change over $\lambda$ of $E\left(\beta|y\right)$ and $\mathrm{Var}\left(\beta|y\right)$ with $\lambda$ varying from 1 to $10^9$, the prior distribution is Gaussian with mean equal to 2500. When the variance tends to infinity (non informative prior), the posterior distribution tends to the ML's one. In such a case, the moments of the posterior distribution of $Y\left(x_0\right)$ tend to:

$$E\left(Y\left(x_0\right)|y\right) = \left(f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F\right)\hat{\beta}_{ML} + r_\theta^T R_\theta^{-1} y \tag{8}$$
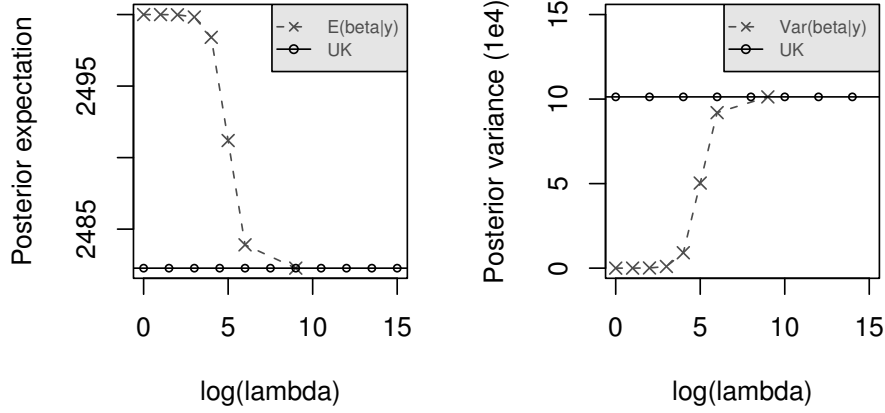
Figure 2: Expectation and variance of the posterior distribution of $\beta$.

$$Var\left(Y\left(x_0\right)|y\right) = \sigma^2 \left[ \left( f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F \right) \left( F^T R_\theta^{-1} F \right)^{-1} \left( f\left(x_0\right)^T - r_\theta^T R_\theta^{-1} F \right)^T \right.$$
$$\left. + \left( 1 - r_\theta^T R_\theta^{-1} r_\theta \right) \right]$$

$$(9)$$

This result can be observed in Figure 3 which presents the evolution of $Var\left(Y\left(x_0\right)|y\right)$ with respect to $\lambda$ for different values of $x_0$, the prior distribution is Gaussian with mean equal to 2500. In this example, the trend function $f\left(x\right)$ is equal to 1. In expressions (8) and (9) one recognizes the predicted value and the prediction variance supplied by Universal Kriging. Hence, Universal Kriging is confounded with Bayesian Kriging in the particular case of a uniform prior distribution for $\beta$, and when $\sigma^2$ and $\theta$ are constants. This is not yet the case for other prior distributions as will be shown in section 3.2.
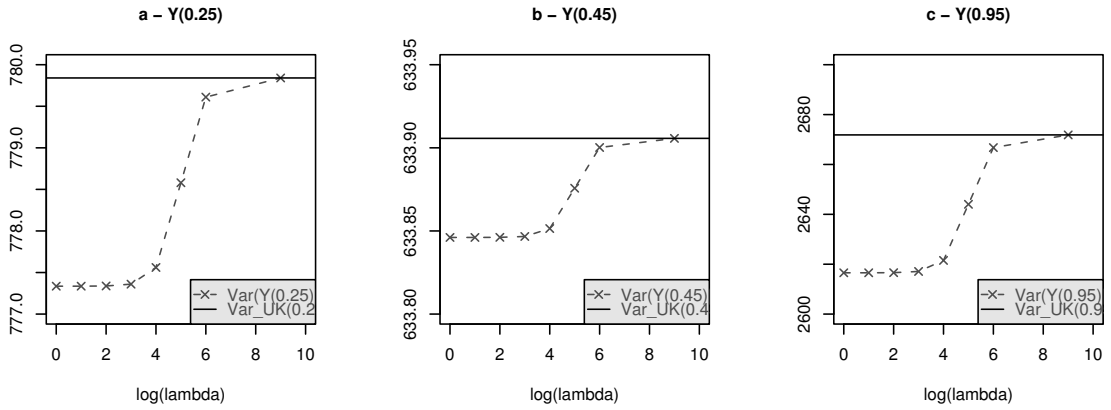


Figure 3: Variance of the posterior distribution of $Y\left(x_0\right)$ for $x_0 = 0.25$, 0.45 and 0.95 with a Gaussian prior distribution.

## 3.2 Known correlation parameters and non informative priors for trend and variance parameters.

This second case is interesting because the posterior distributions are centred on the maximum likelihood estimators. Nevertheless, the prediction variance of Universal Kriging does not correspond to a conditional variance.

Let us define the joint prior density as $\pi(\beta, \alpha) = \dfrac{1}{\alpha}$, where $\alpha = \dfrac{1}{\sigma^2}$.

Thus, theoretical results ([13]) give

$$\pi(\beta, \alpha|y) = \phi\left(\beta|\sigma^2\right)\gamma(\alpha) \tag{10}$$

In expression (10), $\phi$ is the density of the normal distribution centred on $\hat{\beta}_{ML}$ and with variance $\sigma^2\left(F^T R^{-1} F\right)^{-1}$, and $\gamma$ is the density of the gamma distribution with a shape of $\dfrac{n-(p+1)}{2}$ and a scale of $\dfrac{2}{(n-(p+1))\hat{\sigma}_{ML}^2}$. Hence, the mean of this distribution is exactly $\dfrac{1}{\hat{\sigma}_{ML}^2}$. Thus, in Bayesian context with non informative prior (defined as above), the joint posterior distribution of the trend parameters and the variance is centred on the ML estimators, *i.e.* on the same parameters as those used in Universal Kriging. Therefore, it is interesting to compare the two approaches into details: Figure 4 represents a comparison between $Y_{UK}(x)$ and $Y_{BK}(x) = E\left(Y(x)|y\right)$ (figure 4a) and between $\sigma_{UK}(x)$ and $\sigma_{BK}(x) = \sqrt{\mathrm{Var}\left(Y(x)|y\right)}$ (figure 4b) when the prior is non informative.
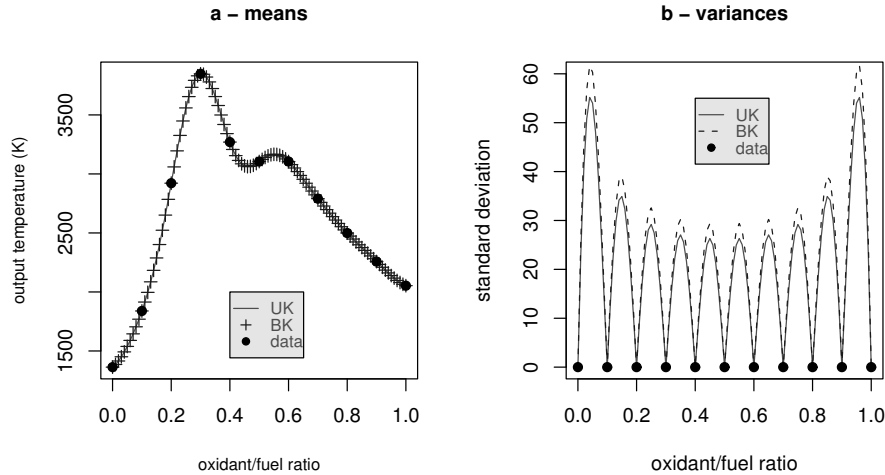


Figure 4: Comparison between Universal and Bayesian Kriging when the prior is non informative.

One can observe on this figure that $Y_{UK}(x)$ and $Y_{BK}(x)$ give the same results. In this particular case, the Universal Kriging estimator can be interpreted as a conditional expectation. Nevertheless, it is not the same for the prediction variance of Universal Kriging $\sigma_{UK}^2(x)$ which is inferior to Bayesian variance $\sigma_{BK}^2(x)$. This difference is mainly explained by the fact that $\sigma_{UK}^2$ takes only into account the uncertainty due to the estimation of $\beta$ and not the uncertainty due to the one of $\sigma^2$.

Thus, this short example shows that $\sigma_{UK}^2(x)$ can not be interpreted as a conditional variance and that Universal Kriging underestimates uncertainty when compared to uncertainty of non informative Bayesian Kriging.

## 3.3 Known correlation parameters and conjugate priors for trend and variance parameters.

The aim of this case is to validate the Markov chain Monte-Carlo simulations (MCMC), useful in practice to get samples from posterior distributions which are not explicitly known (for example, when there is a prior distribution on $\theta$).

Let the prior distribution be the conjugate prior (Gaussian for $\beta$ and Gamma for $\alpha = \sigma^{-2}$):

$$\pi(\beta, \alpha) = \phi\left(\mu, \frac{\Sigma}{\alpha}\right)\gamma\left(a_1, \frac{1}{a_2}\right)$$

The posterior distribution is then well known (Gaussian for $\beta$ and Gamma for $\alpha = \sigma^{-2}$):

$$\pi(\beta, \alpha|y) = \phi\left(\mu_Y, \frac{\Sigma_Y}{\alpha}\right)\gamma\left(a_{11}, \frac{1}{a_{22}}\right)$$

where

$$\mu_Y = \Sigma_Y\left(F^T R^{-1}y + \Sigma^{-1}\mu\right) \text{ and } \Sigma_Y = \left(\Sigma^{-1} + F^T R^{-1}F\right)^{-1}$$

$$a_{11} = a_1 + \frac{n}{2} \text{ and } a_{22} = a_2 + \frac{\mu^T\Sigma^{-1}\mu + y^T R^{-1}y - \mu_Y^T\Sigma_Y^{-1}\mu_Y}{2}$$

The Metropolis Hastings algorithm (see [14]) is used to compute MCMC simulations with a Gaussian random walk.

Figure 5 presents the law adequacy between theoretical posterior distributions and data simulated using a MCMC method in the case of the conjugate prior. Figure 5a presents the posterior distribution of $\beta|\sigma^2$ and Figure 5b the posterior distribution of $\sigma^{-2}$.
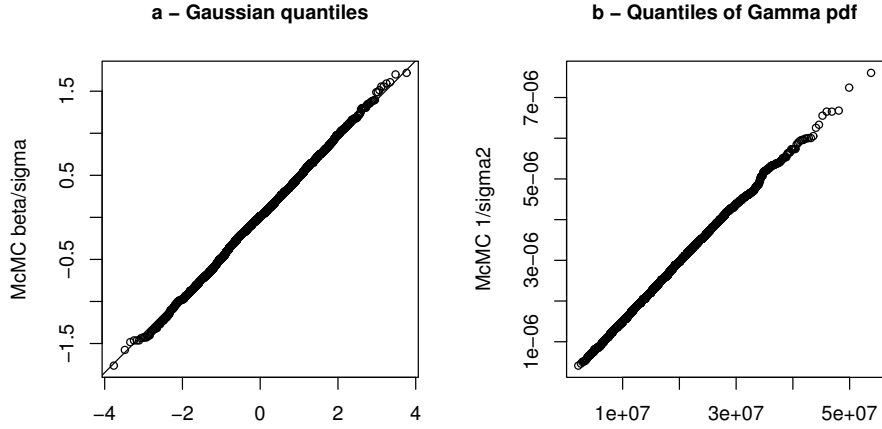
Figure 5: Law adequacy between theoretical posterior distribution (left) and data simulated using a MCMC method in case of the conjugate prior (right).

Table 1: Comparison of the parameters of the sample simulated by MCMC and theoretical parameters in the case of the conjugate prior.

|  | $\beta$ | | $\alpha$ | |
|---|---|---|---|---|
| Prior distribution | $\mu$ | $\sigma^2$ | $a_1/a_2$ | $a_1/a_2^2$ |
| | 2500 | $1000^2$ | $10^{-5}$ | $10^{-10}$ |
| Posterior distribution | $\mu_Y$ | $\Sigma_Y$ | $a_{11}/a_{22}$ | $a_{11}/a_{22}^2$ |
| Theoretical results | 2477 | 0.23 | 2.50 $10^{-6}$ | 9.17 $10^{-13}$ |
| Simulation results | 2482 | 0.22 | 2.50 $10^{-6}$ | 9.61 $10^{-13}$ |

Table 1 and Figure 5 show that the resulting posterior distributions are close to the theoretical ones. The short difference comes from imprecision of the sampling method. Thus, simulations will be used to compute the distribution of the output at any point of the domain conditional on the observations and for every kind of prior distributions, proper or improper. Note that the surfaces are generally compared through the first two moments of the distribution: posterior expectation and posterior variance. In this Bayesian context, the posterior variance includes all sources of uncertainty that comes from the trend, the variance and the correlation function.

## 3.4   Prior distribution on $\beta$, $\sigma^2$ and $\theta$

Let us consider the same experimental case as Martin and Simpson [12] where the prior is:

$$\pi\left(\beta, \frac{1}{\sigma^2}, \theta\right) = \frac{1}{(\sigma^2)^{3/2}}$$

Here, the posterior distributions are sampled for all model parameters (trend, variance and also correlation) using MCMC techniques. The results are validated by a comparison to the posterior distributions of Martin and Simpson's paper.

In this example, one can compare Universal Kriging where the model parameters are estimated by maximum likelihood and Bayesian Kriging which is a mixture of Kriging models where parameters follow the posterior distribution. Figure 6 presents a comparison between Universal Kriging and Bayesian Kriging. Bellow, the graph on the left presents $Y_{BK}(x) - Y_{UK}(x)$, and shows that estimators are different, especially near the origin of the domain. Regarding uncertainty (graph on the right), standard deviation of Universal Kriging is always smaller than that of Bayesian Kriging. Thus, expectation and variance provided by Universal Kriging and Bayesian Kriging are different. In particular, Universal Kriging underestimates uncertainty, a result already observed previously.
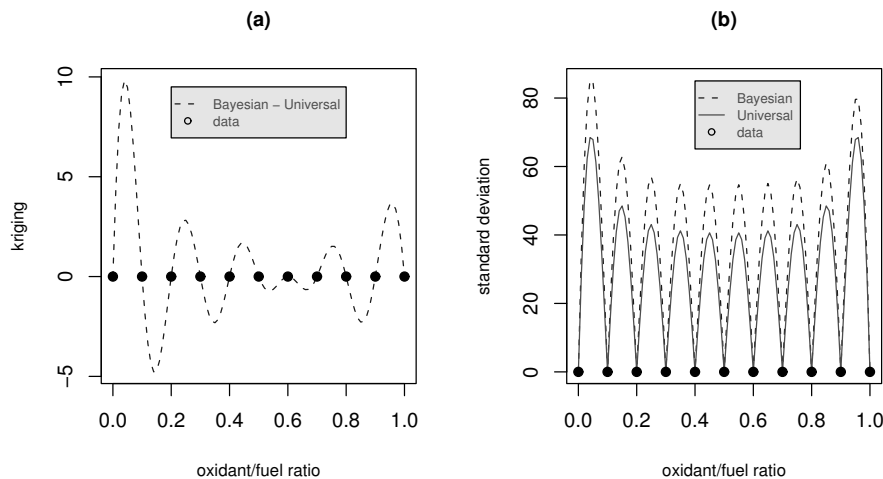


Figure 6: Comparison between Universal Kriging and Bayesian Kriging in the non informative case where the prior is $\pi\left(\beta, \sigma^{-2}, \theta\right) = \sigma^{-3}$.

Another advantage of the Bayesian approach is the assessment of the whole distribution of the predicted values. For example, a very asymmetric posterior

distribution will not be detected by universal approach.

At the same time, Bayesian Kriging avoids the optimization of the likelihood function which is often badly conditioned, especially in high dimension, when little information is available. Besides, the difference between approaches increases with dimension. Several publications ([3], [13]) have dealt with this topic. However, these authors only use generic prior distributions: conjugate or non informative. The section 4 proposes a novel way to get prior information.

# 4    Case study: Impact of the choice of the prior

A petroleum case study is considered to illustrate the general interest of Bayesian Kriging. The goal is to propagate uncertainty from geophysical parameters to the field oil production. The geophysical parameters (permeability, porosity etc.) which are used to describe the oil reservoir are not precisely known yet have a profound influence on the oil production: indeed, a high permeability and a high porosity generally yield high production levels. The resulting oil production probability distribution will be carefully analyzed to decide if the field should be exploited.

The oil production of a given field can be simulated using computationally intensive computer experiments. In this case study we consider a 3D streamlines oil production simulator (namely 3DSL® [15]). The simulator output is the field oil production (FOPT) after 7000 days. We deal with 3 inputs which have a major impact on the output: two permeability factors (LMULTKZ and KRWMAX) and the well's bottom hole pressure (LBHP). They were transformed to belong to the range [-1, 1]. Then the simulator is used to propagate uncertainty from geophysics parameters (inputs) to the oil production (output). A single run of 3DSL® takes a long time (2.5 CPU time[2]), so very few simulations are available. A metamodel, obtained by Kriging for example, is built and uncertainty is propagating through the surrogate, instead of through the simulator itself. In this section we show how Bayesian Kriging can more accurately estimate the production probability law than Universal Kriging. How the choice of the prior influences the results will be demonstrated: by comparing Bayesian Kriging with a non informative and informative priors.

The idea presented here consists in using simplified simulations which go much faster, to derive informative prior for Bayesian Kriging. Indeed we can easily tune

---

[2]Since CPU time has been higly dependant on the hardware, we have provided relative time values for the simulators' responses

the 3DSL® simulation parameters to speed up simulations. However, these simulations become less accurate and can not be directly used as a data set to fit Kriging models. Nevertheless they can be used to provide prior information for Bayesian Kriging. Two different approches are considered to speed up simulations: in both ways the number of nodes on each stream line is smaller by a factor of ten. Some constraints such as the actualization of the field pressure are relaxed. Nonetheless, the main difference comes from the time step used to do the calculus which is more accurate in the first approach. Simulations obtained by the first approach (resp. second approach) will be called Degraded 1 (resp. Degraded 2) simulations.

In order to compare the simulations, a full factorial design at 11 levels in each direction has been done. The whole surface contains 1331 points. Table 2 shows that 3DSL® simulations and Degraded 1 simulations are highly correlated with a correlation coefficient of 0.95. As expected, Degraded 1 simulations are closer to 3DSL® than Degraded 2. It should be noted that the correlation is very good in the directions of KRWMAX and LBHP (coefficient higher than 0.97). Figure 7, which presents a comparison between the 3DSL®, Degraded 1 and Degraded 2 simulations in directions LMULTKZ, KRWMAX and LBHP, also displays this accuracy. It is not the case in the direction of LMULTKZ with a correlation coefficient of 0.61. Degraded 2 simulations are less correlated with 3DSL® than Degraded 1. This can be observed by lower figures in Table 2 and also on Figure 7, which shows the mean of the response in each direction.

Table 2: Correlation among the 3 sets of simulations.

| Correlation Coefficient | ALL | LMULTKZ | KRWMAX | LBHP |
|---|---|---|---|---|
| 3DSL®/Degraded 1 | 0.95 | 0.61 | 0.97 | 0.98 |
| 3DSL®/Degraded 2 | 0.80 | -0.19 | 0.89 | 0.91 |
| Degraded 1/Degraded 2 | 0.89 | 0.56 | 0.96 | 0.94 |

Let us focus on two points of Figure 7:

- curve's level: Degraded 2 simulations seem to be closer to 3DSL® than to Degraded 1,

- curve's variations: Degraded 1 simulations seem to be closer to 3DSL® than to Degraded 2, a result which corresponds to the correlation coefficients of Table 2.
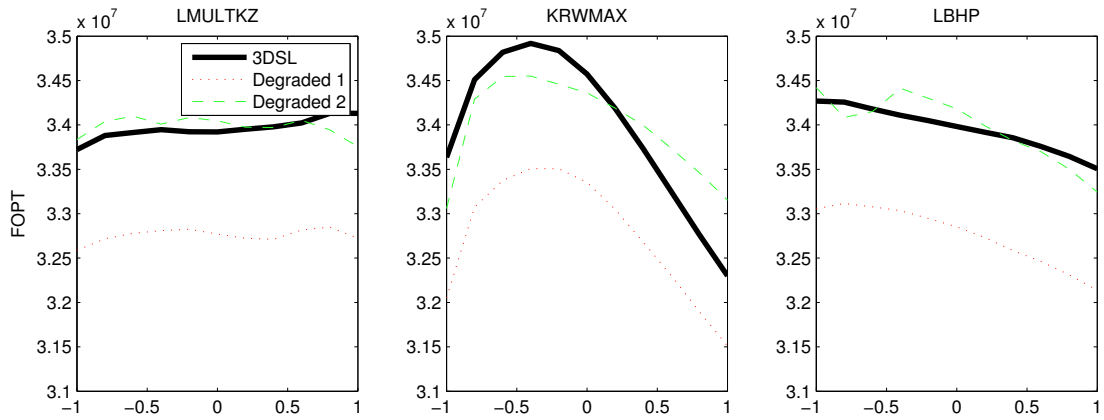
Figure 7: Comparison among 3DSL®, Degraded 1 and Degraded 2.

Regarding CPU time, Figure 8 presents a comparison of CPU times which illustrates that one 3DSL® simulation takes around 2.5 CPU time as opposed to 0.25 for a Degraded 1 run and only 0.18 for a Degraded 2. Thus, Degraded 1 and Degraded 2 simulations are quite close to 3DSL® but are less time consuming. These simple-to-compute metamodels will quickly provide useful information to compute Bayesian Kriging.
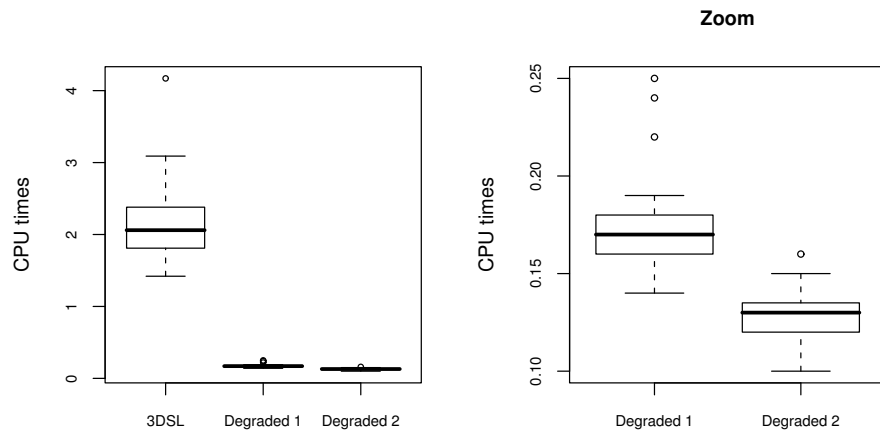


Figure 8: Comparison of CPU times for the full factorial design $3^3$.

As afore mentioned, we want to build a predictive metamodel for 3DSL® in order to propagate uncertainty. Our study is limited to surrogates obtained by Kriging (Universal or Bayesian) estimated using very few runs. The trend of the

14

Kriging model will supposed to be linear with respect to the factors[3] . This choice can be discussed in the direction of KRWMAX, where the trend is more quadratic than linear (cf. Figure 7). However, it is not a problem to underestimate the curvature, the residual term for Kriging is then more substantial, which simplifies identifying the correlation parameters. A comparison of different trends could have been interesting but has not yet been done.

We will compare four different strategies equivalent with respect to CPU time consumption:

- **"UK"** strandard approach using Universal Kriging on 20 3DSL® simulations,

- **"no info BK"** a Bayesian approach using a non informative Kriging on the same 20 runs,

- **"info 1 BK"** a Bayesian approach on 17 3DSL® simulations using a prior distribution obtained from 24 Degraded 1 simulations,

- **"info 2 BK"** a Bayesian approach on 18 3DSL® simulations using a prior distribution obtained from 24 Degraded 2 simulations.

Indeed, the four strategies almost take the same amount of CPU computation time: 41.55 units for the first two strategies, 39.20 for the third and 40.36 for the last one.

The embedded space filling designs used to fit kriging in the different strategies are plotted in Figure 9. The 17 runs design is plotted in black circles, the run which is added to compose the 18 runs design is represented by a triangle. Finally the two last runs to compose the 20 runs design are represented by squares.

Note that the non informative law (see section 3.3) used to compute strategy **"no info BK"** is the following:

$$\pi\left(\beta, \sigma, \theta\right) = \frac{\pi\left(\theta\right)}{\sigma} \qquad (11)$$

To the best of our knowledge, little information is known about $\theta$. We will use a uniform distribution on [0,10]. A longer range would not be coherent with the size of the domain, where the maximal estrangement between two points is equal to 2.

---

[3] $Y\left(x|\beta, \sigma^2, \theta\right) = \beta_0 + \beta_1 MULTKZ + \beta_2 KRWMAX + \beta_3 LBHP + Z\left(x\right)$ where $Cov\left(Z(x), Z(x+h)\right) = \sigma^2 R\left(h|\theta\right)$
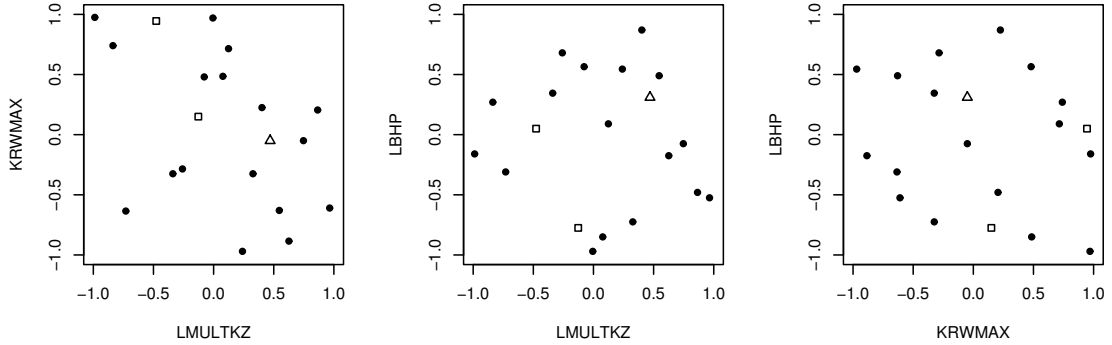
Figure 9: Projection of the embedded designs used to fit kriging.

Note also that Bayesian Kriging with the same non informative prior has been computed on the 24 degraded simulations in order to extract the prior distribution needed by **"info 1 BK"** and **"info 2 BK"**.

Expectations and variances of posterior distributions obtained on Degraded 1 & 2 simulations are summed up on Table 3[4]. There are several differences between these two sets of data, especially on $\beta_1$, $\theta$ and $\sigma^2$. In the case Degraded 2 simulations, $\beta_1$ is equal to zero (LMULKZ has no influence on FOPT). Moreover, the range parameters are smaller and the total variance appears also smaller.

Table 3: Parameters posterior distributions on the 24 runs.

|  |  | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma^2$ | $\theta_1$ | $\theta_2$ | $\theta_3$ |
|---|---|---|---|---|---|---|---|---|---|
| 24 runs of | $E\left(.|Y\right)$ | -0.40 | -0.23 | -0.52 | -0.64 | 2.12 | 3.57 | 0.76 | 2.71 |
| Degraded 1 | $Std\left(.|Y\right)$ | 0.78 | 0.57 | 0.88 | 0.53 | 0.88 | 1.93 | 0.31 | 1.30 |
| 24 runs of | $E\left(.|Y\right)$ | -0.56 | 0.00 | -0.63 | -0.58 | 1.69 | 2.25 | 0.48 | 0.97 |
| Degraded 2 | $Std\left(.|Y\right)$ | 0.61 | 0.42 | 0.68 | 0.70 | 0.68 | 0.78 | 0.19 | 0.35 |

The prior used for **"info 1 BK"** (resp. **"info 2 BK"**) is centred on parameters shown in the first two lines (resp. the last two lines) of Table 3. For example, mean and variance of $\sigma^2$ are *a priori* equal to 2.12 and 0.88 in strategy **"info 1 BK"**, whereas they are equal to 1.69 and 0.68 in **"info 2 BK"**.
Concerning distributions, the Normal law is chosen for $\beta$ and $\theta$ and Lognormal for $\sigma^2$.

---

[4]The parameter values are not expressed in the same scale as the output's (see Figure 7). Actually, data were centred and reduced before modelling.

Table 4 presents the comparison of the four strategies which are evaluated through 4 indicators computed on the whole surface of N=1331 points:

- the **R**oot **M**ean **S**quare **E**rror: $RMSE = \sqrt{\dfrac{1}{N} \sum\limits_{i=1}^{N} \left( FOPT\left(x_i\right) - \hat{Y}\left(x_i\right) \right)^2}$

- the **M**ean **A**bsolute **E**rror: $MAE = \dfrac{1}{N} \sum\limits_{i=1}^{N} \left| FOPT\left(x_i\right) - \hat{Y}\left(x_i\right) \right|$

- the **A**verage **S**tandard **D**eviation: $AStD = \dfrac{1}{N} \sum\limits_{i=1}^{N} \sqrt{\sigma_{\hat{Y}}^2\left(x_i\right)}$

- and the **Pr**obability of exceeding: $PR = \dfrac{1}{N} \sum\limits_{i=1}^{N} 1_{\left| FOPT(x_i) - \hat{Y}(x_i) \right| > 2\sigma_{\hat{Y}}(x_i)}.$

The $RMSE$ and the $MAE$ are average distances between the real surface composed of 1331 runs of 3DSL$^{\circledR}$ and the surface which has been estimated by one of the four strategies. The $AStD$ represents the average uncertainty provided by each strategy. The $Pr$ is the proportion of the points which are outside the interval: $\left[ \hat{Y}\left(x\right) - 2\sigma_{\hat{Y}}\left(x\right); \hat{Y}\left(x\right) - 2\sigma_{\hat{Y}}\left(x\right) \right].$

Table 4: Comparison of the four strategies.

|  | ”UK” | ”no info BK” | ”info 1 BK” | ”info 2 BK” |
|---|---|---|---|---|
| $RMSE$ | 205 042 | 204 042 | 166 770 | 223 845 |
| $MAE$ | 159 402 | 143 190 | 125 748 | 171 786 |
| $AStD$ | 99 457 | 141 573 | 188 411 | 258 394 |
| $Pr$ | 40 % | 19 % | 8 % | 4 % |

**"Info 1 BK"** appears the best strategy. Indeed, its $RMSE$ and its $MAE$ are smaller than those of other strategies. The average uncertainty provided by this strategy is well estimated: the interval $\left[ \hat{Y}\left(x\right) - 2\sigma_{\hat{Y}}\left(x\right); \hat{Y}\left(x\right) - 2\sigma_{\hat{Y}}\left(x\right) \right]$ contains 92% of the output of 3DSL$^{\circledR}$. Thus, the information that is obtained from Degraded 1 simulations and introduced through Bayesian Kriging is useful: the prediction surface and its uncertainty are accurate.

We observed that "**info 2 BK**" gives a $RMSE$ and a $MAE$ higher than other strategies. The information introduced using Degraded 2 simulations degrades the estimation of the surface. The difference between the surfaces obtained by the three Bayesian strategies can be observed through the difference of posterior distributions (see Table 5). The posterior distribution obtained after using the fastest and less accurate Degraded 2 simulations appears remarkable: mean of $\sigma^2$, $\theta_2$ and $\theta_3$ are smaller in this distribution than in the others. Note that this distribution is less dispersed.

"**No info BK**" and "**UK**" give similar results according to $RMSE$ and $MAE$ (see Table 4). Indeed, when the prior is non informative, information used with Bayesian Kriging is only given by data. Thus, this case is close to Universal Kriging. However the average standard deviation ($AStD$) appears much smaller with UK than with BK. Thus, the interval $\left[\hat{Y}(x) - 2\sigma_{\hat{Y}}(x); \hat{Y}(x) - 2\sigma_{\hat{Y}}(x)\right]$ contains only 60% of the output with UK against 81% with BK. The uncertainty announced by UK is widely underestimated, especially because it does not take into account the uncertainty on correlation parameters.

Table 5: Parameters posterior distributions for the three Bayesian strategies.

| | | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma^2$ | $\theta_1$ | $\theta_2$ | $\theta_3$ |
|---|---|---|---|---|---|---|---|---|---|
| "info 1 | $E(.|Y)$ | -0.41 | 0.19 | -0.41 | -0.47 | 1.88 | 3.74 | 0.87 | 3.32 |
| BK" | $Std(.|Y)$ | 0.55 | 0.30 | 0.64 | 0.32 | 0.40 | 1.16 | 0.19 | 0.88 |
| "info 2 | $E(.|Y)$ | -0.41 | 0.17 | -0.60 | -0.21 | 1.00 | 2.98 | 0.70 | 1.48 |
| BK" | $Std(.|Y)$ | 0.42 | 0.22 | 0.51 | 0.37 | 0.14 | 1.21 | 0.18 | 0.96 |
| "no info | $E(.|Y)$ | -0.50 | 0.54 | 0.32 | -0.18 | 2.64 | 3.54 | 1.05 | 4.75 |
| BK" | $Std(.|Y)$ | 0.99 | 0.46 | 1.07 | 0.41 | 0.75 | 1.37 | 0.25 | 2.00 |

One last remark must be added relating to the impact of the design. The good results obtained with strategy "**info 1 BK**" are not only due to appropriate prior information but also from the impact of the design. For example (not presented here in detail) it is puzzling that the 20 runs design gives poorer results than that of 17 runs (see Table 6) with all kind of Kriging methods, since the former includes the latter. A deeper study is needed in order to provide a better understanding of the phenomenon.
In addition, the results mentioned above are still the same, even when the source of variability coming from the design is removed.

Table 6: Comparison between Universal Kriging and Bayesian Kriging on the 17 runs design.

|  | ”UK” | ”no info BK” | ”info 1 BK” | ”info 2 BK” |
|---|---|---|---|---|
| $RMSE$ | 175 595 | 186 976 | 166 770 | 200 200 |
| $MAE$ | 132 935 | 135 422 | 125 748 | 151 905 |
| $AStD$ | 136 832 | 199 505 | 188 411 | 281 317 |
| $Pr$ | 22 % | 7 % | 8 % | 2 % |

# 5    Conclusions

The first objective of this paper was to show that in most cases, prediction variance of Universal Kriging can not be interpreted as the variance of the response conditional on the observations. Indeed, it only takes into account the uncertainty induced by the estimation of trend parameters and not those created by approximating variance and correlation parameters. Therefore, it underestimates the resulting uncertainty on the response. This result has also been observed in the 3D case study.

The second aim of this paper is to propose different strategies to get informative prior information. The performance of Bayesian kriging then depends on the choice of the prior:

- If one uses a non informative prior there is no risk: the results will be close to those obtained by universal kriging: close predictors but a better prediction variance. It is then very interesting to use Bayesian Kriging especially in contexts where techniques used are based on the prediction variance [4]. However, before applying Bayesian Kriging on a data set, one must compute the posterior distribution using MCMC techniques which can be hard task.

- If one uses an informative prior, two different things can occur. First if the prior is well adapted Bayesian Kriging gives excellent results: accurate predictor and precise prediction variance. On the contrary, if the prior is not well adapted Bayesian Kriging furnishes poor results. Hence, if the confidence in the prior is weak, one should use a non informative prior.

# Aknowledgements

# References

[1] Sacks J., Welch W. J., Mitchell T. J., and Wynn H.P. Design and Analysis of Computer Experiments. *Statistical Science* **4**: 409-435 (1989).

[2] Santner T., Williams B., and Notz W. The Design and Analysis of Computer Experiments. Springer (2003).

[3] Oakley J., and O'Hagan A. Probabilistic Sensitivity Analysis of Complex Models: a Bayesian Approach. *Journal of the Royal Statistical Society* **66**: 751-769 (2004).

[4] Jones D., Schonlau M., and Welch W. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* **13**: 455-492 (1998).

[5] Handcock M.S., and Stein M.L. A Bayesian Analysis of Kriging. *Technometrics* **35**: 403-410 (1993).

[6] Zimmerman D.L., and Cressie N. Mean Squared Prediction Error in the Spatial Linear Model with Estimated Covariance Parameters. *Annals of Institude of Statistical Mathematics* **44**: 27-43 (1992).

[7] Prasad N.G.N., and Rao J.N.K. The Estimation of Mean Squared Error of Small Area Estimators. *Journal of American Statistical Association* **85**: 163-171 (1990).

[8] Fuller W.A., and Hasza D.P. Properties of Predictors for Autoregressive Time Series. *Journal of the American Statistical Association* **76**: 155-161 (1981).

[9] Oakley J. Eliciting Gaussian Process Priors for Complex Computer Codes. *The statistician* **51**: 81-97 (2002).

[10] Berger J.O., De Oliveira V., and Sanso B. Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association.* **96**: 1361-1374 (2001).

[11] Ripley B.D. Spatial Statistics. New York, Wiley (1981).

[12] Martin J., and Simpson T. A Monte Carlo Simulation of the Kriging Model. *American Institute of Aeronautics and Astronautics* Paper 4483 (2004).

[13] Goria S., Armstrong M., and Galli A. Using a Bayesian Approach to Incorporate New Information when Estimating Resources. *8th annual conference of the International Association for Mathematical Geology* (IAMG), 15-20 September 2002, Berlin: 75-79 (2002).

[14] Robert C. Méthodes de Monte Carlo par chaînes de Markov. Economica (1996).

[15] 3DSL® `http://www.streamsim.com` (June 2008).