A Monte Carlo Simulation of the Kriging Model

Jay D. Martin*
Applied Research Laboratory, State College, PA 16804

and

Timothy W. Simpson†
The Pennsylvania State University, University Park, PA 16802

This paper investigates the resulting probability distribution of a kriging model, a Gaussian process model, when the model parameters must be estimated from observations of the process being modeled. The output of a kriging model is defined as being Gaussian given the model parameters. In practice, these model parameters must be estimated from observations of the process being modeled (typically a computationally expensive computer mode). It was found that when the model parameters are treated as random variables instead of known values, the resulting probability distribution of the kriging model is a Student-t distribution, a distribution with fatter tails than the Gaussian or normal distribution. The Markov Chain Monte Carlo (MCMC) method was used to determine the probability distributions of the model parameters and the output of the kriging model given the observations. The resulting model parameters were validated against the results of a Bayesian analysis of the simple one-dimensional test problem. The results were also compared to the standard method of Maximum Likelihood Estimation as an alternative method to estimate model parameters.

I. Introduction

A Monte Carlo simulation of a kriging model is presented in this work to characterize the probabilistic nature of the kriging model. A kriging model can surrogate the mapping process of a computationally expensive computer simulation with a more computationally efficient approximation that interpolates a set of observations of the computationally expensive model. A kriging model treats this mapping process as a spatial stochastic process, approximating the computationally expensive model as a spatial Gaussian process, providing a mean and variance for the output of the surrogate. One of the most difficult tasks in using a kriging model as a metamodel is estimating the values of the model parameters associated with the form of the kriging model based upon the set of known observations. When kriging models are used as deterministic approximations of computer simulations, the parameter estimation task is typically accomplished by utilizing Maximum Likelihood Estimation (MLE)[1]. One deficiency with this technique is that it maximizes the conditional probability of the given observations assuming the model parameters are known. In reality these parameters are not known but are random variables, potentially resulting in the underestimation of the stochastic process variability[2]. The underestimation of the process variability is meaningless when the output of the kriging model is used as a deterministic approximation of the computer simulation since only the expected value is used. A second deficiency of MLE is it is characterized as an optimization process and frequently cost function for the model parameters is not convex, having multiple maxima points. This requires the use of stochastic optimization algorithms[1], a computationally expensive process, to guarantee the selection of the global maximum likelihood.

The proper estimation of the variability of the stochastic process is important when the kriging model is used as a probabilistic model, quantifying the uncertainty that exists in the model. In Reliability-based Design Optimization (RBDO), it is important to include model uncertainty while assessing the sensitivity of the system design to input uncertainty[3-5]. RBDO typically requires finding an optimal design point that satisfies the system constraints with a

* Research Assistant, Manufacturing Systems, P.O. Box 30, Member AIAA. Email: jdm111@psu.edu.
† Associate Professor, Departments of Mechanical & Nuclear Engineering and Industrial & Manufacturing Engineering, 329 Leonhard Bldg., Member AIAA. Email: tws8@psu.edu.

specified percentage of certainty. Most optimal designs occur at a constraint boundary, such as a size, weight, or cost. By understanding the impact of input variability on the actual system performance, a design can be selected that satisfies the system constraints with a prescribed probability. Evaluation of probabilities at potential design points requires a large number of computer simulation evaluations to properly characterize the variability of the system performance as a function of input variation. Accurate uncertainty characterization can become infeasible for computationally expensive computer simulations. As a result, a metamodel, a model of the model, can be used to assess the system uncertainty, reducing the computational expense of RBDO[6-8]. The use of this approximation to the computer simulation introduces another source uncertainty, model uncertainty, that must be included in the system uncertainty assessment[3]. This work is motivated by the need to correctly quantify this model uncertainty in system uncertainty assessment.

In this work, a Markov Chain Monte Carlo (MCMC)[9] method is used as a fully Bayesian analysis method to estimate the kriging model parameters and the output of the kriging model. The MCMC method accounts for the uncertainty in the model parameters that is ignored by the MLE method. The MCMC parameter selection method is characterized as an integration process and as a result is much less sensitive to the multimodality of the likelihood function. The results of this parameter estimation technique are compared to those found using MLE for a one-dimensional test problem. Section 2 provides some motivation for the need of computationally efficient models for both Multidisciplinary Design Optimization (MDO) and Reliability-based Design Optimization (RBDO). Section 3 gives a brief description of the test problem and the procedure used to compare MCMC and MLE parameter estimation methods. A brief review of the kriging model and the MLE and Bayesian analysis parameter estimation methods are given in Section 4. Section 5 provides some background and comparison of results for the MCMC method used in this work for parameter estimation. Section 6 offers some conclusions and future.

II. Motivation

This work is concerned with quantifying model uncertainty[3] and including it in RBDO. It is impossible for a mathematical model to exactly reproduce reality; as a result all mathematical models have uncertainty associated with them. High fidelity models, such as finite element analyses (FEA) and computational fluid dynamics (CFD) models attempt to minimize this uncertainty by modeling as much of the small-scale structure of the system with physics-based models. In general, it is assumed the uncertainty in a high fidelity model is small compared to the variability of the inputs to the model. This level of detail in a high fidelity model comes at the cost added computational expense.

Multidisciplinary Design Optimization (MDO) of a system results in an optimal design that typically resides on a system constraint such as maximum system mass, size, cost or durability. In a perfect world the design can be specified exactly and the system constraints can be satisfied exactly. In a real world there exists variation in the materials used to make the system, variations in the manufacturing process to create the system, and variations in the system's lifecycle that make it impossible to exactly satisfy the system constraints. Inclusion of these variations during design optimization is termed Reliability-based Design Optimization (RBDO). RBDO is a method that finds the optimal design point that has a specified probability of satisfying the system constraints. RBDO takes MDO from a deterministic method to a probabilistic method, adding significantly to the complexity of the calculations by at least doubling the number of variables and constraints considered in the optimization[10].

A reliable design should not be confused with a robust design. A robust design is a design that is minimally sensitive to input variations from the nominal design. An optimally robust design is one that trades off optimality of design performance and sensitivity to design variation about that optimal design point[11, 12]. As identified above, a reliable design is an optimal design that satisfies design constraints with a specified probability.

RBDO requires the definition of the performance metric as a function of the design parameters. Given this function, the variability of the design parameters (inputs to the function) can be propagated through the function to provide a resulting performance metric and its variability. This performance function is actually characterized with a probability distribution function (pdf) in which the expected value occurs at the expected value of the design parameters. The pdf of the performance function is defined as a function of the design parameters. The system constraints are also functions of the design parameters and may be either linear constraints, such as a maximum or minimum value of a parameter or they may nonlinear such as maximum or minimum performance value like cost. The reliability of a design point is then the integral of the performance function's pdf that lies within the feasible region. This is a very difficult integral to evaluate directly and is almost approximated. The different assumptions used to approximate this integral differentiate many of the RBDO methods published in the literature[13-16].

The Monte Carlo Simulation (MCS) method is frequently considered the standard method used to evaluate the integral upon which all other methods are compared in terms of accuracy[15, 17, 18]. It is also the most computationally

expensive method, requiring tens to hundreds of thousands of performance function evaluations for even simple problems. This has prompted the desire to use metamodels to act as surrogates of the actual performance function[6, 7]. A metamodel is a computationally efficient approximation to the performance function. Metamodels are also incorporated into non-MCS RBDO methods[8].

A limitation of the RBDO methods previously mentioned is they do not account for uncertainty in the mathematical model used to evaluate the performance function. This is an issue made worse when using a metamodel since it is an approximation to the mathematical model and will introduce additional uncertainty into the performance estimate. The inclusion of model uncertainty in RBDO calculations will tend to broaden the performance function's pdf (increase its variability). This will result in an optimal reliable design that is more conservative than when model uncertainty is ignored[19].

III. Example Problem

A one-dimensional problem is used to demonstrate some of the many aspects of the kriging model. The computer simulation calculates the temperature of a chemical reaction. The mass ratio of oxidant to fuel being burned is increased from no oxidant to an excess of oxidant. In this process, the reaction increases in temperature to a maximum and then decreases as excess oxidant is added as shown in Fig. 1. There are flat portions in the output temperature as the reaction moves through regions of phase changes with the products of combustion. The sample points used to fit the kriging models are indicate with the dots. They are evenly spaced at 0.1 increments from 0 to 1 for a total of 11 points.

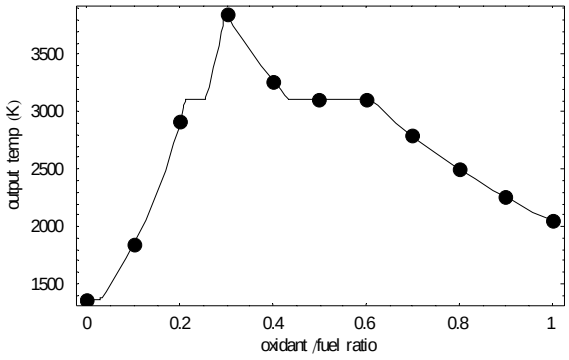


Figure 1. Plot of Output Temperature vs. Oxidant.

IV. Background on the Kriging Model

The kriging model[20] is a popular metamodel for deterministic computer models that have been used in a variety of applications including MDO[21-26] and RBDO[7] because it interpolates all of observations used to create it. In these applications it is used as a deterministic approximation of a computationally expensive model. A kriging model approximates the output of a computer simulation, $y(x)$, by quantifying the correlation that exists between n observations, $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}$, of the computer simulation at n input sites, $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$, that lie within the domain of $y(x)$. The kriging model is capable of exactly reproducing (interpolating) a set of n known observations, **Y**, an important feature for metamodels of deterministic computer models[27-29]. The kriging model is actually a probabilistic model, capable of producing a probability density function (pdf) for the output of the computer model given the set of observations. A probabilistic model is "a model in which the results are determined by using one or more random variables to represent uncertainty about a process or in which the given input will produce an output according to some statistical distribution."[30] Only the expected value of a kriging model is used when it is used as a deterministic metamodel.

A kriging model summarizes the probability distribution of its output as a Gaussian process and provides equations of the expected value or mean and the variance, sufficient statistics to completely specify pdf of the output. More details on kriging models can be found in many sources[28, 29, 31, 32]. The kriging model provides the expected value of the computer simulation, $y(x)$, with the following equation:

$$\mathrm{E}\left[y(x)\right] = \mathbf{f}^{\mathrm{T}}(x)\boldsymbol{\beta} + \mathbf{r}^{\mathrm{T}}(x)\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}) \tag{1}$$

and the variance is:

$$\mathrm{var}\left[y(x)\right] = \sigma^2\left(1 - \begin{bmatrix}\mathbf{f}^{\mathrm{T}}(x) & \mathbf{r}^{\mathrm{T}}(x)\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{F}^{\mathrm{T}}\\ \mathbf{F} & \mathbf{R}\end{bmatrix}\begin{bmatrix}\mathbf{f}(x)\\ \mathbf{r}(x)\end{bmatrix}\right) \tag{2}$$

In the case of ordinary kriging, the regressors for the linear trend function are defined as: $\mathbf{f}^{\mathrm{T}}(x) = \{1\}$. The matrix **F** is the values of the regressors at all of the observation sites. The coefficients of the regressors, $\boldsymbol{\beta}$, are selected to remove any bias that may exist in the observations, leaving a Gaussian process with zero mean and variance of σ^2. An example of an ordinary kriging model for the example problem is shown in Fig. 2. This plot indicates the expected value with a heavy line and the 95% confidence interval of that expected value with a light dashed line. The confidence interval (variance) of the interpolating kriging model approaches zero as the expected value approaches an observation site.

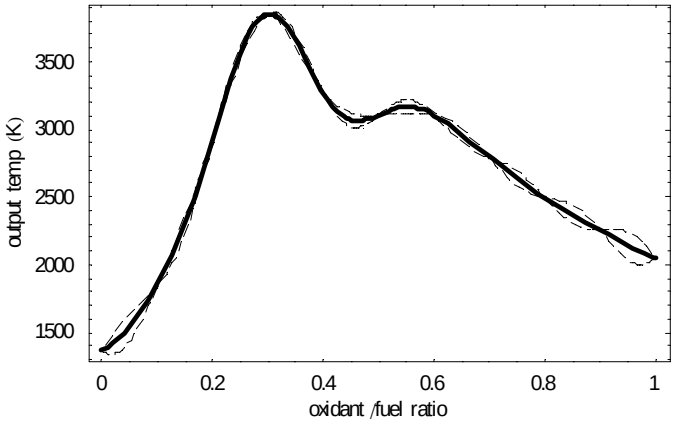


Figure 2. Plot of Ordinary Kriging Model of the 11 Observations of the Example Problem.

The correlation between two observations is quantified by a spatial correlation function. The Gaussian function is one of the most commonly used spatial correlation functions for kriging models used in engineering design, and it is the spatial correlation function used in this work. There are many other possible alternatives[33]. The Gaussian spatial correlation function provides a very smooth and infinitely differentiable surface and is defined with only one parameter, θ, which controls the range of influence of nearby points as follows:

$$R(x_1, x_2) = e^{-\left(\frac{|x_2 - x_1|}{\theta}\right)^2}, \text{ where } \theta > 0 \tag{3}$$

The correlation matrix, **R** used in Eqs. (1) and (2), quantifies the correlation between all of the observations of the system, **Y**, by using the spatial correlation function, $R(x_1, x_2)$.

The Gaussian or *normal* probability density function specified by the expected value and variance of the kriging model (Eqs. (1) and (2)) is conditional on the *n* observations, **Y**, of the computer simulation and the kriging model parameters, $\boldsymbol{\gamma} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}\}$. In reality, only the *n* observations are known; the kriging model parameters are unknown and must be estimated from the observations. Estimating the best model parameters based upon the observations is the most difficult task to complete when using kriging models. The use of cross-validation, a method that does not assume any probability distributions, and Maximum Likelihood Estimation (MLE), a method that assumes the observations follow a known pdf, to select the best values of the model parameters have been investigated by many researchers[1, 29, 34, 35]. They concluded that MLE typically selects the optimal model parameter values better than cross-validation. As a result, cross-validation is not included in the comparison of MLE and Bayesian analysis to determine kriging model parameters.

A. Maximum Likelihood Estimation

The MLE method of parameter estimation maximizes the likelihood of the model parameters given the observations. If the output distribution of the computer model comes from a Gaussian distribution, then the likelihood of the parameters, $\boldsymbol{\gamma}$, is defined as the multivariate normal distribution for the *n* observations of **Y** given the model parameters, $\boldsymbol{\gamma}$, and is given as:

$$L(\boldsymbol{\gamma}|\mathbf{Y}) = p(\mathbf{Y}|\boldsymbol{\gamma}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n |\mathbf{R}(\theta)|}} e^{\frac{(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta})^T \mathbf{R}(\theta)^{-1}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta})}{2\sigma^2}} \quad (4)$$

The MLE method find the values of the model parameters, $\boldsymbol{\gamma}$, that maximize this likelihood. Unfortunately, the multivariate probability distribution is defined conditionally on the model parameters and does not account for the uncertainty that exists in these parameters. This has prompted the use of Bayesian analysis methods to determine the model parameters since they do account for the uncertainty in the model parameters[36-38].

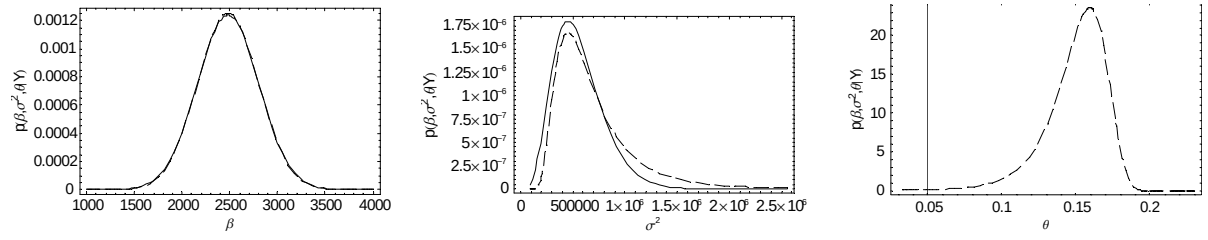
The MLE method is characterized as an optimization process. For a well-behaved likelihood function, a smooth concave function, this is a very efficient method to determine the best model parameters. More often, especially in the multi-dimensional input setting, there can be regions of the model parameter space where the likelihood function becomes unstable due to the required inversion of the correlation matrix, **R**. It is also prone to multiple maxima and/or long ridges of near maximum values, causing extreme difficulty for most gradient based optimization algorithms[1, 39, 40]. As a result, a stochastic optimization method, such as simulated annealing, should be used to determine the optimal kriging model parameters. The use of a stochastic optimization method significantly increases the computational expense of using MLE as a parameter estimation method.

The asymptotic properties of Maximum Likelihood Estimators state that the inverse of the observed Fisher information matrix provide the covariance matrix of the kriging model parameters[2, 41]. The observed Fisher information matrix is defined as the Hessian of the logarithm of the likelihood function (Eq. (4)) and can be used to measure the "information" in the observations of the computer model. The observed Fisher information matrix for the example problem at the model parameter values of maximum likelihood is given in Table 1. The diagonal terms are the variances of the model parameters and the off-diagonal terms are the covariances between the model parameters. If the observations were independent, then the correlation matrix would be the identity matrix and the distribution of the trend function coefficients would have a Gaussian distribution and the variance would have a Gamma distribution[42]. The distributions of the trend function coefficients are Gaussian with the mean and variance as given in Table 1, the process variance, σ^2, is approximated with a Gamma distribution, and the distribution family of the correlation parameter is not known.

Table 1. Covariance Matrix and Values of the Model Parameters for the Example Problem.

	β	σ^2	θ
β	104,000	-7.191×10^{6}	-0.9787
σ^2	-7.191×10^{6}	5.826×10^{10}	2692
θ	-0.9787	2692	0.000370
mode	2480.7	458,000	0.159
$\sqrt{\text{variance}}$	322.3	241,000	0.019

The plots of the likelihood function as a function of each of the three model parameters for the example problem are shown in Fig. 3. The likelihoods (shown as dashed lines) are calculated by holding varying each model parameter over its range of feasibility while holding the other model parameters constant at their most likely value. They are normalized to result in a proper probability distribution. The distributions predicted from the Fisher information matrix shown with solid lines, using the mode and variance as given in Table 1, appear to match the likelihoods fairly well with an underestimation of the variance of the process variance. The authors know of no distribution family for the predicted distribution for the correlation range parameter.



a) likelihood over β b) likelihood over σ^2 c) likelihood over θ

Figure 3. Likelihood function plots of model parameters.

One desire of this study is to determine the marginal probability distributions of the model parameters and compare them to the likelihoods given in Fig. 3. The marginal probability distribution of the model parameters is a measurement of the probability distribution of each model parameter, independent of the of model parameters given the observations of the system being modeled. It is relatively easy to compute the observed information matrix at selected values of the model parameters but it is computationally expensive to calculate the marginal probability distributions. The existence of non-zero off-diagonal terms in Table 1 may indicate that the model parameters are correlated and as a result dependent upon each other. If the model parameters were independent then little difference between the shape of the likelihood functions and the marginal distribution would be expected.

B. Bayesian Analysis

A Bayesian analysis[43] of the model parameter estimation process begins with the definition of the posterior distribution of the model parameters given the observations, **Y**:

$$\pi\left(\beta,\sigma^2,\theta|\mathbf{Y}\right)=\frac{p\left(\mathbf{Y}|\beta,\sigma^2,\theta\right)\pi\left(\beta,\sigma^2,\theta\right)}{\int_{\gamma} p\left(\mathbf{Y}|\beta,\sigma^2,\theta\right)\pi\left(\beta,\sigma^2,\theta\right)d\boldsymbol{\gamma}} \tag{5}$$

Eq. (5), also known as Bayes rule for continuous probability distributions, is used to "turn around" the likelihood equation of Eq. (4), providing the probability distribution of the model parameters given the observations. From this result the model parameter uncertainty is determined by calculating the marginal probability distributions of the model parameters given the observations as:

$$p\left(\beta|\mathbf{Y}\right)=\iint_{\sigma^2,\theta}\pi\left(\beta,\sigma^2,\theta|\mathbf{Y}\right)d\sigma^2 d\theta \tag{6}$$

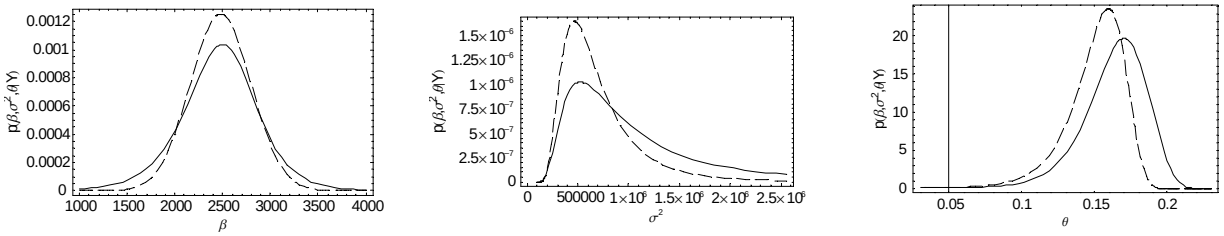
The marginal distribution of each model parameter is determined by "integrating out" the other model parameters.

The Bayesian analysis process first requires the selection of a *prior* distribution for the model parameters, $\pi\left(\beta,\sigma^2,\theta\right)$ used in Eq. (5). Selection of the prior is based upon previous knowledge of the model parameters. Often, there is little prior information available on the model parameters which prompt the use of a *noninformative prior*, a prior that makes no assumptions on the distributions of the model parameters. Most priors chosen are *improper*, *i.e.* their integral is not finite, and have the general form:

$$\pi\left(\beta,\sigma^2,\theta\right)\propto\frac{\pi\left(\theta\right)}{\left(\sigma^2\right)^a},\quad a\in\mathbb{R}, \tag{7}$$

for various choices of $\pi\left(\theta\right)$ and a[36, 44]. One of the simplest forms is the Laplace prior where $\pi\left(\theta\right)=1$ and $a=0$ resulting in $\pi\left(\beta,\sigma^2,\theta\right)=1$. A second alternative is the *independence Jeffreys* prior where $\pi\left(\theta\right)=1$ and $a=1$[36]. In most cases of model parameter selection (statistical inference), both of these priors tend to result in proper posterior distributions. The likelihood function dominates Eq. (5), making the posterior distribution insensitive to the prior model used. Berger *et al.*[44], found this to not be the case when selecting the model parameters for spatial stochastic processes such as kriging. In their study to find a general form for a prior distribution of the model parameters that yields a proper posterior distribution for the spatial stochastic process model, they determined a proper posterior distribution would result by using the reference prior[45] or the through the use of $\pi\left(\theta\right)=1$ for a large enough (for the case of the Gaussian correlation function $a>\frac{3}{2}$) in Eq. (7).

A Bayesian analysis was completed using the data from the test problem in a manner similar to Kennedy and O'Hagan[46]. The plots in Fig. 4 overlay the marginal probability density function plots (the solid curves) over the likelihood functions (the dashed curves) from Fig. 3. The likelihood function is normalized so it represents a proper probability density function (the area under the probability curve is 1). The solid lines in the plots show the marginal probability density function for each of the model parameters as a comparison. The pdfs for β are both symmetric with their maximum value (the mode) occurring at nearly the same location. The actual values are presented in



a) Marginal probability of β b) Marginal probability of σ^2 c) Marginal probability of θ

Figure 4. Plots of Likelihoods and Marginal Probabilities for $\pi\left(\beta,\sigma^2,\theta\right)=1$

Table 2. The mode of the likelihood function represents the model parameter values found using the MLE method. The broader distributions of the marginal pdfs appear to suggest the estimates are less accurate than what the likelihood functions imply. The plot of the marginal pdf of the process variance, σ^2, appears to confirm the conclusion from Berger *et al.*[44]: the improper prior, $\pi\left(\beta,\sigma^2,\theta\right)=1$ results in an improper posterior for σ^2 since the probability does not become infinitely small as σ^2 becomes large. For this reason the improper prior of Eq. (7) with $\pi\left(\theta\right)=1$ and $a=\frac{3}{2}$ is also compared in Table 2 and Fig. 5.

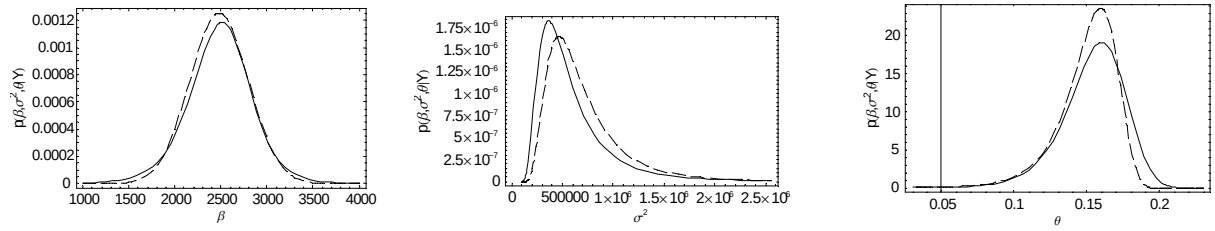
Table 2. Model Parameter Estimates for Different Prior Assumptions.

	Likelihood			$\pi\left(\beta,\sigma^2,\theta\right)=1$			$\pi\left(\beta,\sigma^2,\theta\right)=\frac{1}{\left(\sigma^2\right)^{\frac{3}{2}}}$		
model parameter	β	σ^2	θ	β	σ^2	θ	β	σ^2	θ
mode	2481	458,000	0.159	2494	526,000	0.170	2511	362,000	0.160
median	2481	601,000	0.154	2471	813,000	0.166	2495	499,000	0.156
mean	2481	698,000	0.150	2460	952,000	0.163	2485	600,000	0.154
$\sqrt{\text{variance}}$	319	371,000	0.021	456	533,000	0.023	362	361,000	0.023

The *median* (the 50% probability point) and the *mean* (the expected value or the first moment) are two commonly used estimates of the model parameters and are frequently better estimates of the model parameters than the mode (Berger[43] pg. 134). All three of the estimates of the model parameters and the square root of the variance (standard error) are provided in Table 2 for the likelihood function and the two different priors. The square root of the variance provides a measurement for the accuracy of the estimate. The variances, the second moment of the pdf, for the likelihood functions compare well to the variances determined by the observed Fisher information matrix in Table 1.

The model parameter estimates from Table 2 indicate that the selection of the prior makes little difference to the estimate of β for this test example. Since the resulting posterior distributions of β are nearly symmetric, there is little difference between the mode, median and mean. The standard error for β is larger for either choice of prior than for the likelihood function. The three different estimates, mode, median, and mean, for the two other model parameters, σ^2 and θ, are significantly different since their corresponding pdfs are asymmetric. The selection of the prior also appears to have a significant impact on the resulting posterior distributions of the model parameters. The second prior choice results in model estimates that have smaller standard errors than the first prior choice. This is to be expected since the second prior choice indicates there is a preference for model parameters that result in a smaller process variance.

The resulting marginal probability distribution for the model parameters using the second prior choice, $\pi(\beta,\sigma^2,\theta)=1/(\sigma^2)^{\frac{3}{2}}$, are shown in Fig. 5. The resulting posterior distributions in this case are more similar to the likelihood functions than in the previous case. The major difference between the posterior distributions and the likelihood functions occurs in the pdf of σ^2, where the selected prior directly influences the resulting posterior distribution by shifting it to the left (smaller values) and compressing the resulting standard error of the resulting model parameter estimate. The posterior distribution of the process variance, σ^2, does appear to become infinitely small as σ^2 gets large, resulting a proper posterior distribution for this choice of prior distribution. The second difference is the increased standard error in the estimate of the correlation range, θ.



a) Marginal probability of β b) Marginal probability of σ^2 c) Marginal probability of θ

Figure 5: Plots of Likelihood and Marginal Probabilities for $\pi(\beta,\sigma^2,\theta)=1/(\sigma^2)^{\frac{3}{2}}$

The results of this Bayesian analysis of the simple example problem indicate that including the model parameter uncertainty in the estimation of the model parameters is important. It also shows the accuracy of the model parameter estimates, the standard errors, is typically underestimated by using the MLE method and the resulting information matrix at the selected model parameters. The Bayesian analysis method is predominated by integration as compared to MLE which is mostly an optimization problem. The result is the Bayesian method does not have the computational difficulties experienced by MLE in situations where the likelihood function is multimodal or has ridges or near optimal values. This improvement comes at the expense of the increased computational expense of evaluating all of the multidimensional integrals and is usually infeasible for most problems other than the simplest ones such as the example problem in the work. For this reason, the more computationally efficient Bayesian analysis method of Markov Chain Monte Carlo simulation was developed.

V. Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a general method to simulate complex, nonstandard multivariate distributions. It can be used to simulate stochastic processes having probability densities known up to a constant of proportionality[9]. This permits the use of Eq. (5) to determine the posterior distributions of the model parameters without requiring the computationally expensive evaluation of integral in the denominator. For this work, MCMC is used to explore the posterior distributions of the kriging model parameters, $\boldsymbol{\gamma}$ and the output of the kriging model at unobserved locations in the domain in a manner similar to that described by Diggle *et al.*[38].

The MCMC method for the simulation of a distribution *f* is any method producing an ergodic Markov chain whose stationary distribution is *f* [47]. A Markov Chain is a process that generates a sequence of values that converge to a single value independent of the starting value of the sequence. In this work, the Metropolis-Hastings algorithm is used due to its universality of approximating unknown probability distributions[47-50]. The chain is created by generating a random candidate, x', according to a *candidate-generating density*, $q(x,x')$. This density specifies the probability of generating a new point, x', given the current point, *x*. If $q(x,x')$ satisfies the reversibility condition:

$$\pi(x)q(x,x')=\pi(x')q(x',x) \tag{8}$$

then the density $q(x,x')$ is the desired distribution. Unfortunately, it is seldom possible to know this density exactly; the equality in Eq. (8) is replaced by the left-hand being greater than the right-hand side, implying the

transition from x to x' occurs too often and from x' to x too rarely. This condition can be remedied by introducing a probability $\alpha(x,x') < 1$ that the move is made. The reversibility condition of Eq. (8) is then satisfied as follows:

$$\pi(x)q(x,x')\alpha(x,x') = \pi(x')q(x',x)\alpha(x',x) \tag{9}$$

The probability of move, $\alpha(x',x)$, should be maximized to satisfy the reversibility condition; its upper limit is 1. The probability of moving from the current state, x, to candidate state x' is:

$$\begin{aligned} \alpha(x,x') &= \min\left[\frac{\pi(x')q(x',x)}{\pi(x)q(x,x')},1\right], \text{ if } \pi(x)q(x,x') > 1 \\ &= 1, \quad \text{otherwise} \end{aligned} \tag{10}$$

The Metropolis-Hastings algorithm is specified by:

1. candidate-generating density, $q(x,x')$,
2. if the candidate-generating density is symmetric, $q(x,x') = q(x',x)$, then the probability of move (Eq. (10)) reduces to $\pi(x')/\pi(x)$,
3. the probability of move, $\alpha(x,x')$, does not require the knowledge of the normalizing constant of posterior distribution, $\pi(x)$, the denominator of Eq. (5), and
4. if a candidate value is rejected, the current value is taken as the next value in the sequence.

In this work, a normal distribution is used as a symmetric candidate-generating density. The candidate value is selected such that the current value is the mean and variance is some constant selected prior to the simulation of the chain. The resulting chain is a random walk where the candidate value is accepted unconditionally if $\pi(x') > \pi(x)$ and with probability given by $\pi(x')/\pi(x)$[51]. This algorithm is also the basis for the simulated annealing optimization algorithm commonly used when maximizing the likelihood equation for model parameter estimation.

The desire of generating these chains or sequences of values is to approximate the probability density functions of the model parameters and simultaneously simulate the probabilistic nature of the kriging model, taking into account the uncertainty of the model parameters. The mean (expected value) and variance of the sequence of length k can be estimated if the elements of the chain are independent by:

$$E(y) = \text{Mean}(y) = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{11}$$

$$\text{Variance}(y) = \frac{1}{k-1}\sum_{i=1}^{k}\left(y_k - E(y)\right)^2 \tag{12}$$

The exact algorithm used to generate the Markov chains in this work is:

1. Start with an initial value x_1 and variance s^2 for the candidate-generating function
2. Repeat for $j = 1,2,\ldots,k$
3. Generate x' from $q(x_j,x') \sim N(x_j,s^2)$ and u from $U(0,1)$
4. If $u \le \alpha(x_j,x')$ accept $x_{j+1} = x'$ else maintain $x_{j+1} = x_j$.

a. $\alpha\left(x_j, x'\right)$ is calculated using the likelihood equation from Eq. (4) and the prior probability, $\pi\left(\beta, \sigma^2, \theta\right) = 1 \Big/ \left(\sigma^2\right)^{\frac{3}{2}}$

b. $\alpha\left(x_j, x'\right)$ is additionally defined to reject candidates outside of the bounds of the model parameters for this specific problem:

i. $\beta \rightarrow [1000, 4000]$

ii. $\sigma^2 \rightarrow [90000, 2560000]$

iii. $\theta \rightarrow [0.03, 0.23]$

5. Return the values of the chain $\{x_1, x_2, \ldots, x_k\}$

The value of the candidate-generating function variance or spread must be selected to ensure the Markov chain covers the probability space of the model parameters. This ability to cover the space is related to the acceptance rate, the rate at which a candidate is accepted. For the random walk, the optimal acceptance rate is dependent upon the number of model parameters being fit: 0.45 for one parameter and 0.23 as the number of parameter approach infinity[52]. In accordance with these theoretical results, the variance of the candidate-generating functions was selected to result in an acceptance rate of 0.3. Table 3 shows that little differences in the mean and variance of the chains generated with acceptance rates from 0.3 to 0.7. The resulting chains were sampled for every 3rd element to reduce serial correlation and were run to a length of 10,000 sampled values after an initial burn-in of 1000 steps. In Table 4, the impact of sampling the chains at different rates is provided for an acceptance rate of 0.3. Little difference was seen for sampling at any rate of 2 or greater.

Table 3: Mean and Variances of the Model Parameters for Many Acceptance Rates

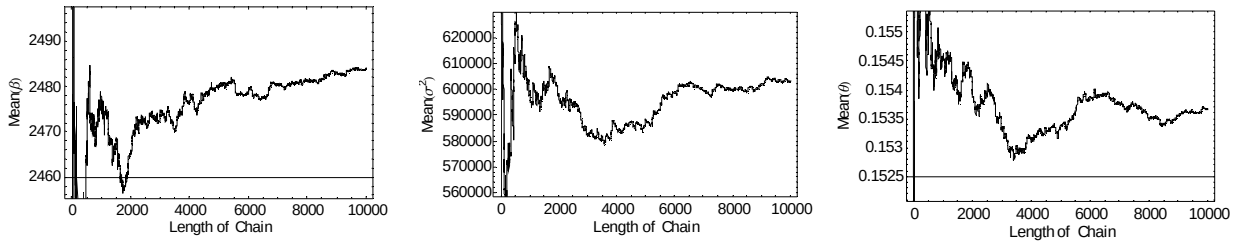
parameter		0.3	0.4	0.5	0.6	0.7
β	mean	2484	2485	2478	2486	2483
	$\sqrt{\text{variance}}$	372	385	370	366	368
σ^2	mean	603,000	616,000	606,000	600,000	603,000
	$\sqrt{\text{variance}}$	364,000	377,000	370,000	356,000	366,000
θ	mean	0.154	0.154	0.155	0.154	0.154
	$\sqrt{\text{variance}}$	0.023	0.023	0.022	0.022	0.023

Table 4: Mean and Variances of the Model Parameters for Many Acceptance Rates

parameter		1	2	3	4	5
β	mean	2481	2486	2484	2490	2490
	$\sqrt{\text{variance}}$	377	373	372	372	374
σ^2	mean	642,000	599,000	603,000	597,000	602,000
	$\sqrt{\text{variance}}$	395,000	363,000	364,000	358,000	368,000
θ	mean	0.156	0.154	0.154	0.154	0.154
	$\sqrt{\text{variance}}$	0.023	0.022	0.023	0.022	0.023

The MCMC method is independent of the starting values used. To minimize the effect of the initial conditions, the chain is permitted "burn-in" of 1,000 values before values are saved. The chains are built till

10,000 values are saved. The mean of the model parameters as each element of the chain is added is shown in Fig. 6. The mean appears to be relatively stable at 10,000 iterations, "however, there can never be any (positive) empirical guarantee that a sufficiently long run has been taken[53]."

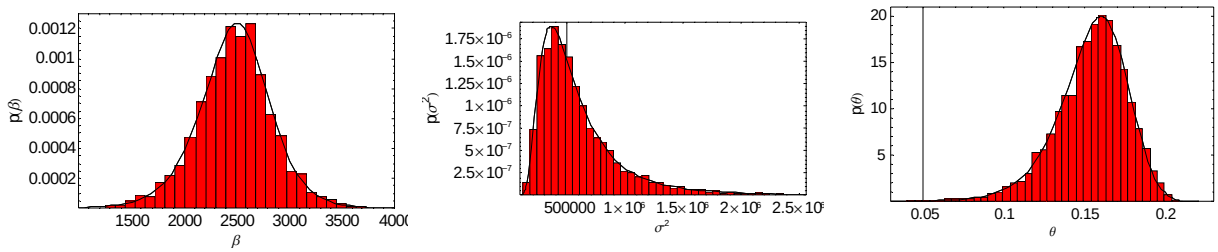


a) Mean of β b) Mean of σ^2 c) Mean of θ

Figure 6: Plots of the Mean of the Model Parameters vs. Length of the Chain

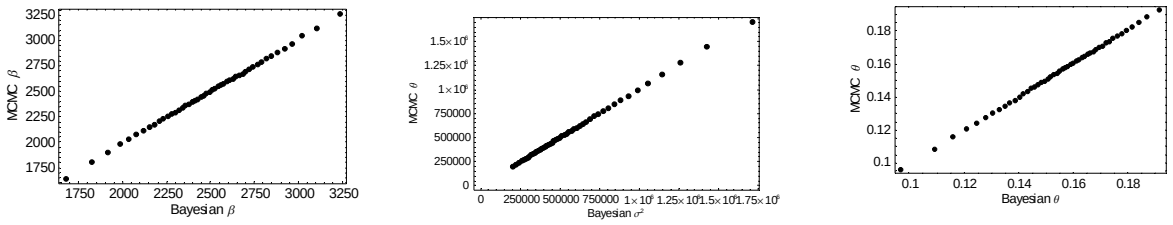
The resulting histograms of the model parameters are shown in Fig. 7. The actual pdf calculated from the Bayesian analysis is shown as the overlaying line. The probability density functions appear to be very good approximations to the actual pdfs. Experience has shown that comparing the pdfs shapes in this manner can be misleading. A much more accurate comparison can be made by observing a Quantile-Quantile plot of the density functions. Fig. 8 shows the quantiles of the densities at increments of 0.02. The R^2 measurement of all three plots is greater than 0.9995, indicating a very good probability the densities are the same.

The output of the kriging model was also simulated at each saved value of the Markov chain. The value of the



a) Marginal probability of β b) Marginal probability of σ^2 c) Marginal probability of θ

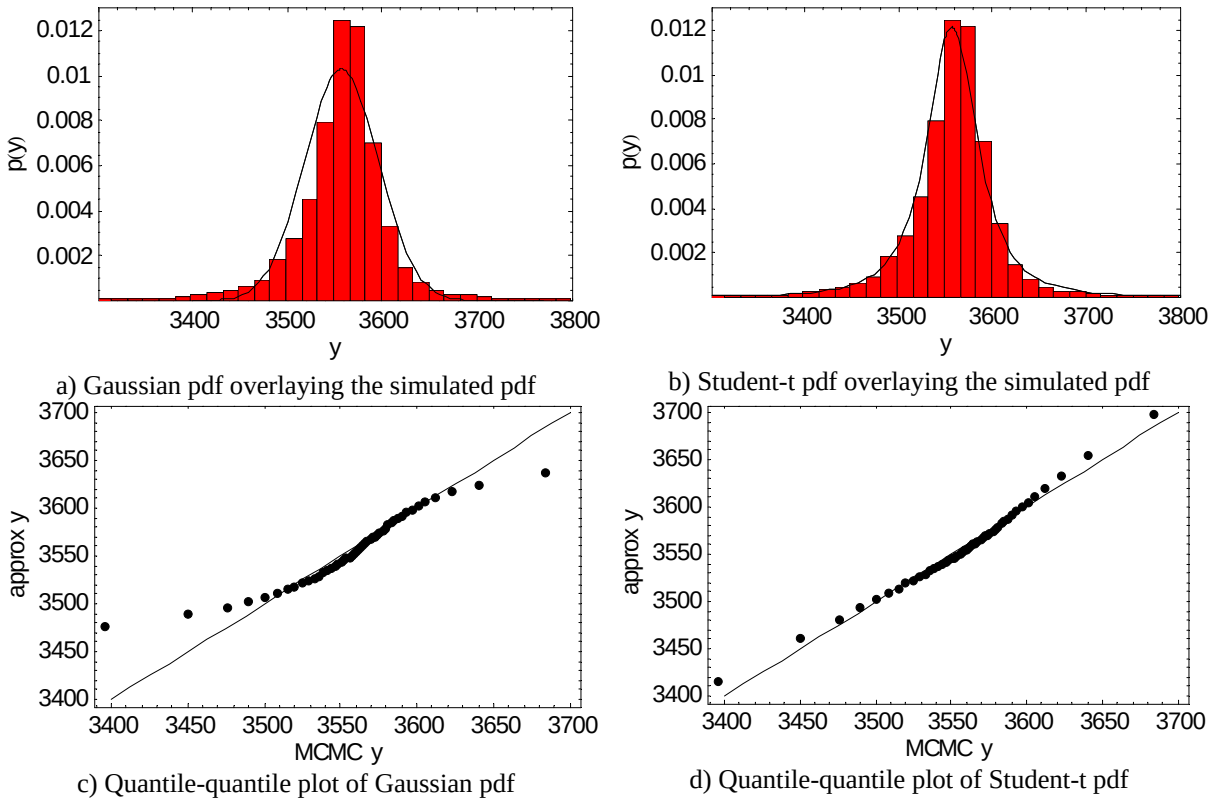
Figure 8: Plots of the Simulated pdfs of the Model Parameters



a) Q-Q plot of β b) Q-Q plot of σ^2 c) Q-Q plot of θ

Figure 7: Quantile-Quantile plots of the Model Parameters

output at a given input location with given model parameters was generated from $N\left(E\left(y\left(x\right)\right), \text{var}\left(y\left(x\right)\right)\right)$, where the expected value and variance are defined in Eqs. (1) and (2) respectively. This simulation technique is very similar to the Gibbs Sampler[47, 53]. The resulting probability density of the output at an input location of 0.25 (halfway between the known observations at 0.2 and 0.3) is shown in Fig. 9.



a) Gaussian pdf overlaying the simulated pdf

b) Student-t pdf overlaying the simulated pdf

c) Quantile-quantile plot of Gaussian pdf

d) Quantile-quantile plot of Student-t pdf

Figure 9: Simulation Results of Kriging Model Output at x=0.25

The main purpose of this study is to assess the approximation of the output of the kriging model as a Gaussian process with mean and variance given by Eqs. (1) and (2). Fig. 9a shows a Gaussian pdf with mean and variance as estimated from Eqs. (1) and (2) using the expected values of the model parameters from the simulation. The data is slightly asymmetric; it is skewed to the left with a skewness of -3.2. The results shown in Fig. 9b reveals the Gaussian pdf underestimates the tails of the distribution. It appears that the output is poorly approximated by the Gaussian pdf when the model parameters must be estimated from the observations.

Given that the output of the kriging model is not well approximated by a Gaussian pdf when the model parameters are estimated from the data, it was desired to find a pdf family that does approximate the output of the kriging model. A Cauchy distribution was investigated next as a pdf that is symmetric with fatter tails than the Gaussian pdf. Unfortunately, it overestimated the thickness of the tails. The Student-t distribution provides a pdf family that permits control of the tail thickness by using a shape factor, ν, from that of the Cauchy ($\nu = 1$) to the Gaussian ($\nu = \infty$). The Student-t distribution is parameterized by only the shape factor (or degrees of freedom) resulting in a distribution that has a zero mean and a unit scale. The standard Student-t distribution is defined as:

$$t(\nu) = \frac{\Gamma\left((\nu+1)/2\right)}{(\pi\nu)^{1/2}\,\Gamma(\nu/2)\left(1+\left(x^2/\nu\right)\right)^{(\nu+1)/2}},\ \nu \text{ a positive integer} \tag{13}$$

where Γ is the Gamma function. For this application, the distribution must be moved (non-zero mean) and scaled appropriately by transforming x with a location and a scale parameter (a and b) as follows:

$$x = \frac{y-a}{b} \tag{14}$$

and by scaling the resulting probability by dividing by b to result in a proper pdf. The location parameter, a, is the expected value from Eq. (1). The scale parameter, b, is equal to the square root of the variance (standard deviation)

from Eq. (2), but is must be adjusted with parameter, λ, by matching the quantile values from one standard deviation of the normal distribution (.84135).

$$b = \frac{\sqrt{\text{variance}}}{\lambda} \tag{15}$$

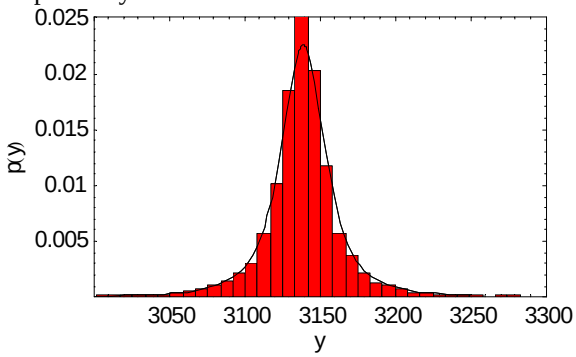
The value of λ for a given shape factor, ν, is given in Table 5. As the shape factor approaches infinity, λ approaches 1 as expected since such a distribution is the normal distribution.

Table 5: Scale Correction Factor, λ, as a Function of Student-t Shape Factor, ν

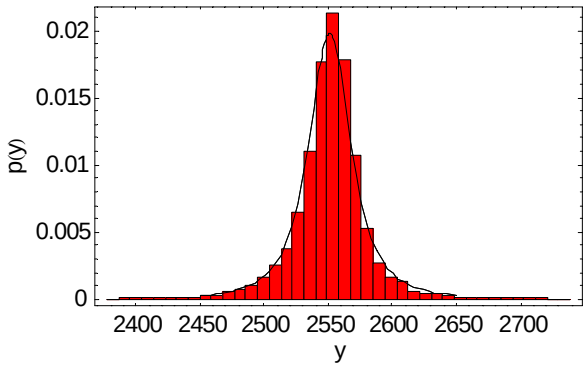
ν	1	2	3	4	5	6
λ	1.84	1.32	1.20	1.14	1.11	1.09

The results shown in Figs. 9b and 9d compare the pdf from the MCMC simulation to that approximated by a Student-t distribution using a shape factor of 2 and the transformation of Eq. (14). The shape factor of 2 was selected to maximize the R^2 (0.9843) of the line fit to plot of Fig. 9d. The standard use of the Student-t distribution arises in the problem of estimating the mean of a normally distributed random sample where the shape factor is equal to the degrees of freedom or the number of observations minus one and is always a positive integer. The definition of the Student-t pdf in Eq. (13) is valid for any positive, non-zero real value for the shape factor, ν. By dropping the restriction that ν is a positive integer, the maximum R^2 of 0.9851 occurs at $\nu = 1.83$.

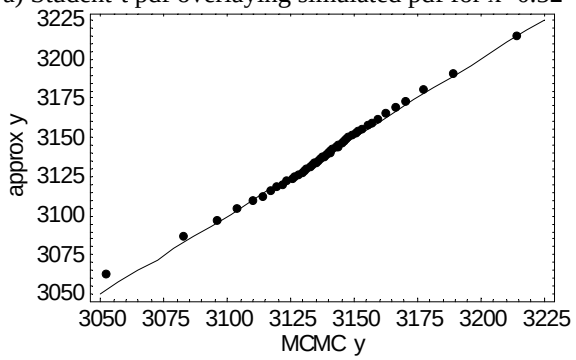
The output distribution of the test problem was simulated at two other locations, $x = 0.52$ and 0.78 to determine if the output can be approximated by the same Student-t distribution. The two locations to be closer to observed values and to closer to the expected mean of the entire process (an output close to β). The results of simulation at these two locations are shown in Fig. 10. The same shape factor, $\nu = 2$, as was used in Fig. 9 was used, provided very good estimations of the pdf of the output. The R^2 of the straight line in Figs. 9c and 9d were 0.993 and 0.998, respectively.



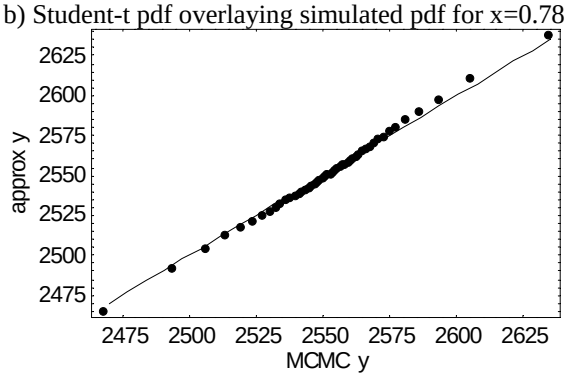
a) Student-t pdf overlaying simulated pdf for x=0.52



b) Student-t pdf overlaying simulated pdf for x=0.78



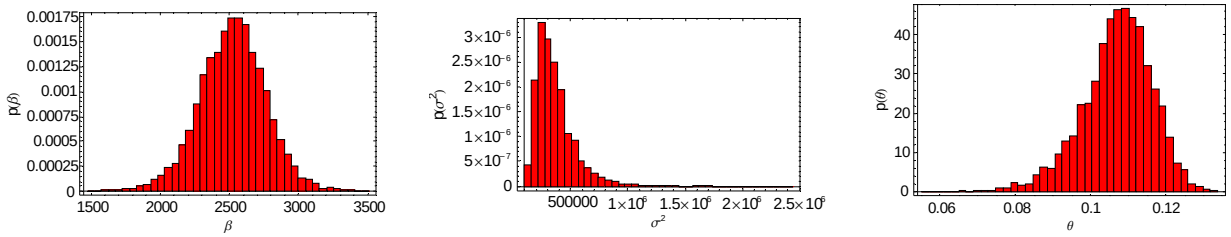
c) Q-Q plot of Student pdf for x=0.52



d) Q-Q plot of Student-t pdf for x=0.78

Figure 10: Simulation Results of Kriging Model Output at x=0.52 and 0.78

A second example was used to verify the results of the first problem; the output of a kriging model can be approximated by a Student-t distribution when the model parameters must be estimated from the observations. In this example, 16 evenly spaced samples were taken from the same computer model as in the first example and the MCMC simulation was completed. The Bayesian analysis was completed since the MCMC results of the first example matched Bayesian analysis very well. The resulting marginal distributions of the model parameters are shown in Fig. 11. The mean and variance of the marginal distributions are compared with those found using 11 observations in Table 6.



a) Marginal probability of β b) Marginal probability of σ^2 c) Marginal probability of θ

Figure 11: Plots of the Marginal pdfs of the Model Parameters Using 16 Observations

Table 6: Comparison of Model Parameters for 11 and 16 Observations of Test Problem

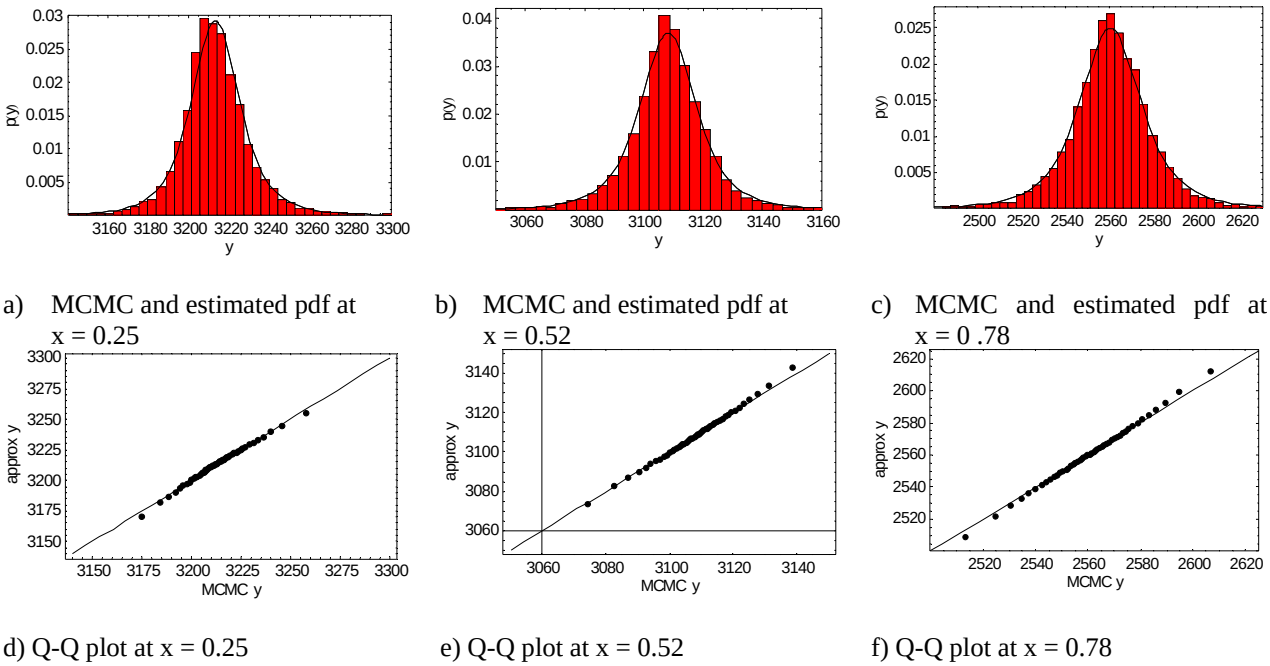
Model parameter	# of observations	mean	$\sqrt{\text{variance}}$
β	11	2484	372
	16	2524	365
σ^2	11	603,000	364,000
	16	365,000	183,000
θ	11	0.154	0.023
	16	0.107	0.009

Table 6 indicates that adding five observations did little change in the marginal distribution of the β parameter. The mean of the other two model parameters did change as a result of the added observations and the variance in their estimate was significantly reduced. It is impossible to directly compare the simulated output pdfs for the three locations since they now fall at different locations relative to the observations used to fit the models. The three simulated output pdfs can be compared to the Student-t distribution using the definition presented above. The results of the simulated outputs at $x = 0.25, 0.52, 0.78$ are shown in Fig. 12. The shape factor of $\nu = 3$ for the Student-t distribution was selected to maximize the R^2 of the line fit to the quantile-quantile plots of Fig. 12. The simulated mean (as calculated per Eq. (11) for the observations is compared to the expected value from Eq. (1) using the mean of the model parameters and the corresponding R^2 value of the line fit to the quantile-quantile plots are shown in Table 7. The simulated means and expected values were found to match almost exactly in the three locations.

Table 7: Summary of MCMC pdfs and Student-t approximated pdfs for 16-Observation Problem

	x = 0.25	x = 0.52	x = 0.78
Simulated Mean	3213	3108	2561
Expected Value	3213	3108	2561
R^2	0.996	0.998	0.999

VI. Conclusions



a) MCMC and estimated pdf at x = 0.25

b) MCMC and estimated pdf at x = 0.52

c) MCMC and estimated pdf at x = 0 .78

d) Q-Q plot at x = 0.25

e) Q-Q plot at x = 0.52

f) Q-Q plot at x = 0.78

Figure 12: Simulation Results of Output at x = 0.25, 0.52, and 0.78 for 16-Observation Problem

A Markov Chain Monte Carlo (MCMC) simulation of a simple kriging model was found to match the results of a Bayesian analysis for estimating the marginal probability distributions of the model parameters. The MCMC method is more computationally efficient than the Bayesian analysis since it permits the simultaneous simulation of all of the model parameters rather than the evaluation of the *n*-dimensional integral of Eq. (5) to determine the joint probability of the model parameters given the observations and the *n* different (*n*-1)-dimensional integrals of Eq. (6) for each model parameter in order to determine the marginal probability distributions of the model parameters.

The MCMC method also permits the simulation of the output of the kriging model that incorporates the uncertainty in the model parameters. The kriging model is defined as a Gaussian process model given the model parameters, but in practice the model parameters must be estimated from a set of observations, altering the actual probability distribution of the kriging model. It was found that this actual probability distribution has thicker tails than the normal distribution predicted by the kriging model with known parameters. As a result a Student-t distribution was used to approximate the output distribution of a kriging model when the model parameters are not known. The Student-t distribution was moved and scaled using Eqs. (14) and (15). The shape parameter, ν, was selected to maximize the R^2 value of a line fit to a quantile-quantile plot of the approximated and simulated pdfs. This value was found to remain constant across all input values to the kriging model. Therefore, best shape factor can then be selected by simulating the output of the kriging model at one location during the MCMC simulation and used for estimating the pdfs at all other locations.

Future work should investigate alternative methods for determining the shape factor for the Student-t distribution approximation of the output of the kriging model when the model parameters are estimated from the set of observations. It should also investigate the impact of using additional model parameters to estimate the trend of the model rather than just a single constant value (universal kriging vs. ordinary kriging). The investigation should finally include examples of more than one input dimension, adding to the number of range parameters to be fit for the correlation functions.

Acknowledgments

The authors gratefully acknowledge support from Dr. Kam Ng of ONR 333, Contract No. N00014-00-G-0058.

References

[1]Martin, J. D. and T. W. Simpson, "On the Use of Kriging Models to Approximate Deterministic Computer Models," *ASME 2004 Design Engineering Technical Conference*, Salt Lake City, UT, 2004.

[2]Stein, M. L., *Interpolation of Spatial Data : Some Theory for Kriging*. Springer Series in Statistics, Springer-Verlag, New York, 1999.

[3]Martin, J. D. and T. W. Simpson, "On Using Kriging Models as Probabilistic Models in Design," *2004 SAE World Congress*, 2004-01-0430, Detroit, MI, 2004.

[4]Wojtkiewicz, S. F., M. S. Eldred, R. V. J. Field, A. Urbina and J. R. Red-Horse, "A Toolkit for Uncertainty Quantification in Large Computational Engineering Models," *42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA 2001-1455, Seattle, WA, 2001.

[5]Draper, D., "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society, Series B*, Vol. 57, No. 1, 1995, pp. 45-97.

[6]Qu, X., S. Venkataraman and R. T. Haftka, "Reliability, Weight and Cost Tradeoffs in the Design of Composite Laminates for Cryogenic Environments," *42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA 2001-1327, Seattle, WA, 2001.

[7]Jin, R., X. Du and W. Chen, "The Use of Metamodeling Techniques for Optimization Under Uncertainty," *Structural and Multidisciplinary Optimization*, Vol. 25, 2003, pp. 99-116.

[8]Youn, B. D. and K. K. Choi, "A New Response Surface Methodology for Reliability-Based Design Optimization," *Computers & Structures*, Vol. 82, 2004, pp. 241-256.

[9]Geyer, C. J., "Practical Markov Chain Monte Carlo," *Statistical Science*, Vol. 7, No. 4, 1992, pp. 473-511.

[10]DeLaurentis, D. A. and D. N. Mavris, "Uncertainty Modeling and Management in Multidisciplinary Analysis and Synthesis," *The 38th AIAA Aerospace Sciences Meeting*, AIAA 2000-0422, Reno, NV, 2000.

[11]Sundaresan, S., K. Ishii and D. R. Houser, "A Robust Optimization Procedure with Variation of Design Variables and Constraints," *Advances in Design Automation*, Vol. 65-1, 1993, pp. 379-386.

[12]Chen, W., J. K. Allen, K.-L. Tsui and F. Mistree, "A Procedure for Robust Design: Minimizing Variations Caused by Noise Factors and Control Factors," *ASME Journal of Mechanical Design*, Vol. 118, No. 4, 1996, pp. 478-485.

[13]Youn, B. D. and K. K. Choi, "Selecting Probabilistic Approaches for Reliability-Based Design Optimization," *AIAA Journal*, Vol. 42, No. 1, 2004, pp. 124-131.

[14]Choi, K. K. and B. D. Youn, "On Probabilistic Approaches for Reliability-Based Design Optimization (RBDO)," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, AIAA 2002-5472, Atlanta, GA, 2002.

[15]Du, X. and W. Chen, "A Most Probable Point Based Method for Uncertainty Analysis," *ASME 2000 Design Engineering Technical Conferences and Computers and Information in Engineering Conferences*, DETC2000/DAC-14263, Baltimore, MD, 2000.

[16]Tu, J., K. K. Choi and Y. H. Park, "Design Potential Method for Robust System Parameter Design," *AIAA Journal*, Vol. 39, No. 4, 2001, pp. 667-677.

[17]Du, X. and W. Chen, "Towards a Better Understanding of Modeling Feasibility Robustness in Engineering Design," *ASME Journal of Mechanical Design*, Vol. 122, No. 4, 2000, pp. 385-394.

[18]Emch, G. and A. R. Parkinson, "Using Engineering Models to Control Variability: Feasibility Robustness for Worst-Case Tolerances," *Advances in Design Automation*, Vol. 65-1, 1993, pp. 411-418.

[19]Padmanabhan, D. and S. M. Batill, "An Iterative Concurrent Subspace Robust Design Framework," *8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, AIAA 2000-4841, Long Beach , CA, 2000.

[20]Cressie, N. A. C., *Statistics for Spatial Data*, John Wiley & Sons, Inc., New York, 1991.

[21]Booker, A. J., A. R. Conn, J. E. Dennis, Jr., P. D. Frank, M. Trosset and V. Torczon, "Global Modeling for Optimization: Boeing/IBM/Rice Collaborative Project 1995 Final Report," The Boeing Company ISSTECH-95-032, Seattle, WA, Dec. 1995.

[22]Simpson, T. W., J. J. Korte, T. M. Maurey and F. Mistree, "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, AIAA-98-4755, St Louis, MO, 1998.

[23]Simpson, T. W., "Comparison of Response Surface and Kriging Models in the Multidisciplinary Design of an Aerospike Nozzle," Langley Research Center, NASA NASA/CR-1998-206935, Hampton, VA, Feb. 1998.

[24]Mitchell, T. J. and M. D. Morris, "Bayesian Design and Analysis of Computer Experiments: Two Examples," *Statistica Sinica*, Vol. 2, 1992, pp. 359-379.

[25]Jones, D. R., M. Schonlau and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, pp. 455-492.

[26]Koch, P. N., B. A. Wujek, O. Golovidov and T. W. Simpson, "Facilitating Probabilistic Multidisciplinary Design Optimization Using Kriging Approximation Models," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, AIAA 2002-5415, Atlanta, GA, 2002.
[27]Simpson, T. W., J. Peplinski, P. N. Koch and J. K. Allen, "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129-150.
[28]Simpson, T. W., T. M. Maurey, J. J. Korte and F. Mistree, "Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233-2241.
[29]Sacks, J., W. J. Welch, T. J. Mitchell and H. P. Wynn, "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409-435.
[30]5000.59, D. D., "DoD Modeling and Simulation (M&S) Management," Jan. 4, 1994.
[31]Martin, J. D. and T. W. Simpson, "A Study on the Use of Kriging Models to Approximate Deterministic Computer Models," *2003 ASME Design Engineering Technical Conference*, DETC2003/DAC48762, Chicago, IL, 2003.
[32]Ripley, B. D., *Spatial Statistics*. Wiley Series in Probability and Mathematical Statistics: Applied Probability & Math Section, John Wiley & Sons, Inc., New York, 1981, 252.
[33]Koehler, J. R. and A. B. Owen, *Computer Experiments*, in *Handbook of Statistics*, Ghosh, S. and C. R. Rao, Editors. 1996, Elsevier Science: New York. p. 261-308.
[34]Currin, C., T. J. Mitchell, M. D. Morris and D. Ylvisaker, "A Bayesian Approach to the Design and Analysis of Computer Experiments," Oak Ridge National Laboratory ORNL-6498, Oak Ridge, TN.
[35]Stein, M. L., "A Comparison of Generalized Cross Validation and Modified Maximum Likelihood for Estimating the Parameters of a Stochastic Process," *The Annals of Statistics*, Vol. 18, No. 3, 1990, pp. 1139-1157.
[36]Handcock, M. S. and M. L. Stein, "A Bayesian Analysis of Kriging," *Technometrics*, Vol. 35, No. 4, 1993, pp. 403-410.
[37]Le, N. D. and J. V. Zidek, "Interpolation with Uncertain Spatial Covariances: A Bayesian Alternative to Kriging," *Journal of Multivariate Analysis*, Vol. 43, 1992, pp. 351-374.
[38]Diggle, P. J., R. A. Moyeed and J. A. Tawn, "Model-based Geostatistics," *Applied Statistics*, Vol. 47, No. 3, 1997, pp. 299-350.
[39]Ripley, B. D., *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge, 1988.
[40]Warnes, J. J. and B. D. Ripley, "Problems with Likelihood Estimation of Covariance Function of Spatial Gaussian Processes," *Biometrika*, Vol. 74, No. 3, 1987, pp. 640-2.
[41]Efron, B. and D. V. Hinkley, "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika*, Vol. 65, No. 3, 1978, pp. 457-482.
[42]Arnold, S. F., *Mathematical Statistics*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1990.
[43]Berger, J. O., *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer Series in Statistics, ed. Fienberg, S., J. Gani, K. Krickeberg, J. O. Berger and B. Singer, Springer, New York, 1985.
[44]Berger, J. O., V. De Oliveira and B. Sanso, "Objective Bayesian Analysis of Spatially Correlated Data," *Journal of the American Statistical Association*, Vol. 96, 2001, pp. 1361-1374.
[45]Berger, J. O. and J. M. Bernardo, *On the Development of the Reference Prior Method*, in *Bayesian Statistics 4*, Bernardo, J. M., J. O. Berger, A. P. Dawid and A. F. M. Smith, Editors. 1992, Oxford University Press: London. p. 35-60.
[46]Kennedy, M. C. and A. O'Hagan, "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society, Series B*, Vol. 63, No. 3, 2001, pp. 425-464.
[47]Robert, C. P. and G. Casella, *Monte Carlo Statistical Methods*. Springer Texts in Statistics, Springer-Verlag, New York, NY, 1999.
[48]Hastings, W. K., "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, Vol. 57, No. 1, 1970, pp. 97-109.
[49]Chib, S. and E. Greenburg, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, No. 4, 1995, pp. 327-335.
[50]Chen, M.-H., Q.-M. Shao and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computing*. Springer Series in Statistics, Springer, New York, 2000.
[51]Metropolis, A., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087-1092.
[52]Roberts, G. O., A. Gelman and W. R. Gilks, "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *The Annals of Applied Probability*, Vol. 7, No. 1, 1997, pp. 110-120.
[53]Smith, A. F. M. and G. O. Roberts, "Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, Vol. 55, No. 1, 1993, pp. 3-23.