

## OPTIMIZATION BY SIMULATION METAMODELLING METHODS

Russell C.H. Cheng  
Christine S.M. Currie

School of Mathematics  
University of Southampton  
Southampton, SO17 1BJ, U.K.

### ABSTRACT

We consider how simulation metamodels can be used to optimize the performance of a system that depends on a number of factors. We focus on the situation where the number of simulation runs that can be made is limited, and where a large number of factors must be included in the metamodel. Bayesian methods are particularly useful in this situation and can handle problems for which classical stochastic optimization can fail. We describe the basic Bayesian methodology, and then an extension to this that fits a quadratic response surface which, for function minimization, is guaranteed to be positive definite. An example is presented to illustrate the methods proposed in this paper.

### 1 INTRODUCTION

In simulation output analysis, it is often more convenient to use a simple statistical (usually regression) model of the simulation output rather than to work directly with the simulation model. These simplified models are termed *metamodels* (see Barton 1998). Let the output of one run of our simulation model be  $y$ . We assume that this depends in a continuous way on  $p$  factors  $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ , and can therefore be regarded as a function of these factors. We treat these factors as *decision variables* that are set by the controller of the system under investigation. We thus assume the following simulation metamodel

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

where  $q(\mathbf{x}, \boldsymbol{\theta})$  is a regression function describing the underlying dependence of the response  $y$  on the factor variables  $\mathbf{x}$ , and  $\varepsilon$  is a random noise variable, to allow for the inherent random variability in the observed output value. The vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)^T$  is a set of coefficients on which the regression function depends. Some, possibly all, of these coefficients will be unknown and have to be estimated by fitting the regression function to observed simulation output.

Geometrically the regression function can be thought of as defining a surface in  $p$ -dimensional space that describes

our estimate of the dependence of the simulation response on the  $p$  factors  $(x_1, x_2, \dots, x_p)$ . This interpretation gives rise to the name *response surface methodology*. Good basic references are Barton (1998) and Hood and Welch (1993).

Response surface methodology is useful if we wish to find the factor setting  $\mathbf{x}^*$  for which the system response is optimal. For example, if the metamodel describes the performance of the system, so that  $q(\mathbf{x}, \boldsymbol{\theta})$  is simply a *performance index*, then  $\mathbf{x}^*$  will be the factor setting that optimizes  $q(\mathbf{x}, \boldsymbol{\theta})$ .

For a discussion of simulation optimization and the use of response surface methodology for this, see Azadivar (1999), Kelton (1999), Glover, Kelly and Laguna (1999), Fu et al. (2000), Gonda et al. (2000), Law and McComas (2000) and Swisher et al. (2000).

One serious and common difficulty remains. If the number of factors is large but the simulation model is complex so that only few runs can be made, then it may not be possible to estimate all the coefficients of the metamodel using for example classical methods of maximum likelihood. Here a Bayesian approach has a great advantage.

In the Bayesian approach we have a prior probability distribution  $\pi(\boldsymbol{\theta})$  that expresses our prior belief about the likely value of the parameters, which gives us a viable starting point for the simulation runs. As we make our simulation runs we can progressively modify and make more precise our belief about the distribution of  $\boldsymbol{\theta}$ . We can stop the procedure at any point, with the knowledge about the behavior of the system output improving with the number of simulation runs made.

One further potential problem that can occur with this methodology is that the response surface fitted is not positive (negative) definite when searching for a minimum (maximum). For situations where we are fairly certain that there is an optimum (which will be assumed to be a minimum in what follows) we present here a simple methodology for a quadratic  $q(\mathbf{x}, \boldsymbol{\theta})$ , based on a restriction of the parameter values, such that the response surface is always positive definite.

Thus in summary we examine the Bayesian approach for fitting  $q(\mathbf{x}, \boldsymbol{\theta})$  and then obtaining the factor setting  $\mathbf{x}^*$  for which  $q(\mathbf{x}, \boldsymbol{\theta})$  is optimized. The basic methodology was given by Cheng (2004). For completeness we outline the methodology again here for the case where we fit a quadratic  $q(\mathbf{x}, \boldsymbol{\theta})$ . We then go on to discuss the issue of ensuring how to ensure that the fitted regression is positive definite, when it is known at the outset that the problem is well-posed so that  $q(\mathbf{x}, \boldsymbol{\theta})$  can be expected to satisfy this condition.

## 2 BAYES METHOD

Without loss of generality, we shall assume that the objective is to estimate the value of  $\mathbf{x}$  which *minimizes*  $q(\mathbf{x}, \boldsymbol{\theta})$ . We further assume that we can only perform a limited number of simulated trials but that there is some prior expert understanding of how the performance index  $q$  might vary as a function of  $\mathbf{x}$ . We use the specific example first given by Cheng (2004) to show how the approach works. This allows for a clearer explanation of the methods. For further discussion of Bayesian methodology in simulation see Cheng (1999) and Chick (2000).

Our example is based on a real engineering application in which the aim of the modeling was to optimize the set-up of a piece of complex equipment so that the time that it took to perform a given task was minimized. The time taken to perform the task was therefore taken to be the performance measure. There were 6 factor variable settings to be set: A, B, C, D, E and F, and we denote their values by  $\mathbf{x} = (x_1, x_2, \dots, x_6)$ .

The simulation model was an accurate model of the real system and took some time to run. For our analysis only 7 simulation runs were made, each at a different value of the factor settings. The output from the simulation runs can be described by

$$y_i = q(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, 2, \dots, 7 \quad (2)$$

where  $\varepsilon$  is initially taken to have mean zero with a constant variance.

As we wish to carry out an unrestricted optimization, we only require  $q$  to be a good approximation to the unknown true system performance in the region of its minimum. We therefore use a quadratic expression for  $q$  and put

$$q(\mathbf{x}, \boldsymbol{\theta}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \dots + \theta_7 x_6 + \theta_8 x_1^2 + \theta_9 x_1 x_2 + \dots + \theta_{28} x_6^2 \quad (3)$$

The Bayesian approach treats  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{28})$  as being a random variable, with a *prior distribution*. We set the  $E[\theta_j]$  equal to the best estimates of the factor values prior to the simulation experiments, and the standard deviation  $\text{StdDev}[\theta_j]$  is set to a value indicative of the confidence that we have in this estimate.

Bayes' Theorem allows updating of the distribution of  $\boldsymbol{\theta}$ , as we get new data points  $y_i$ , using the formula

$$p(\boldsymbol{\theta}|\mathbf{y}) = K^{-1} p(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \quad (4)$$

where, assuming a continuous distribution for simplicity,

$$K = \int p(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (5)$$

is a normalizing constant to ensure that  $p(\boldsymbol{\theta}|\mathbf{y})$  is a proper probability density.

Here  $p(\boldsymbol{\theta}|\mathbf{y})$  is the *posterior distribution* of the parameters  $\boldsymbol{\theta}$  given the observed simulation values  $\mathbf{y}$ , and  $p(\mathbf{y}|\boldsymbol{\theta})$  is the *likelihood* of observing  $\mathbf{y}$  given that the parameters have value  $\boldsymbol{\theta}$ .

The main advantage of the Bayesian approach over the classical methods of maximum likelihood is that it allows the incorporation of prior knowledge. A further attraction is that it allows incremental updating of  $\boldsymbol{\theta}$ , and hence  $q(\mathbf{x}, \boldsymbol{\theta})$ , as additional values of  $\mathbf{x}$  are obtained.

The choice of *design points*  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  at which to perform simulations is a classical statistical experimental design problem. The aim is to minimize the number of design points needed to estimate  $q$ . In our case, we need to estimate first and second order terms of  $q$ .

A classical non-central composite design (see Cochran & Cox (1957) for details) would use  $2^{(6-1)} + 2*6 + 1 = 45$  design points, where there are 32 points to estimate the linear and mixed terms of  $q$ , and 13 points to estimate the squared terms. Thus the upper bound on the number of trial runs required is 45. The number of design points can be reduced if some of the mixed terms are assumed to be zero. We do not consider design aspects further here.

Bayesian updating of  $\boldsymbol{\theta}$ , and hence  $q(\mathbf{x}, \boldsymbol{\theta})$ , can be performed with any number of data points, making it particularly attractive in this example, where only 7 data points are available.

## 3 NUMERICAL EXAMPLE

We intend to investigate the properties of the method more fully including tests on problems where the solution is known. However we limit ourselves in this paper to the example introduced in the previous section.

A prior model for  $q(\mathbf{x})$  was specified as follows. We take, as our initial guess of the minimum point, the values  $\mathbf{x}_0 = (7.586, 2.794, 5.282, 2.313, 1.601, 1.273)$ . These were obtained from discussion with engineers familiar with the system.

The mixed terms  $x_i x_j$  of  $q(\mathbf{x}, \boldsymbol{\theta})$  were all taken to be zero and the remaining terms of  $q(\mathbf{x}, \boldsymbol{\theta})$  were selected so

that  $q(\mathbf{x}, \boldsymbol{\theta})$  increases by 0.1 at a distance +1 or -1 from  $\mathbf{x}_0$  along any axis. This gave

$$q(\mathbf{x}, \boldsymbol{\theta}) = 82.9230 - 1.52x_1 - 0.56x_2 - 1.06x_3 - 0.46x_4 - 0.32x_5 - 0.26x_6 + 0.1x_1^2 + 0.1x_2^2 + 0.1x_3^2 + 0.1x_4^2 + 0.1x_5^2 + 0.1x_6^2 \quad (6)$$

This selection was made to give a starting shape for the regression surface  $q(\mathbf{x}, \boldsymbol{\theta})$  with a definite, but not sharply defined, minimum at the point  $\mathbf{x}_0$  indicated by the engineers

The uncertainty about the coefficients is not so dependent on expert opinion. We therefore assume that each coefficient has a normal distribution, with standard deviation 1. This builds in a fair amount of uncertainty about parameter values.

The 7 seven simulation output results ( $y$ ) are displayed in Table 1, with the corresponding factor settings ( $\mathbf{x}$ ). The selection of the design points was made in a rather simplistic way. In the first run the settings were selected to be  $\mathbf{x}_0$ , the engineers' best estimates of the optimal settings. Subsequent factor settings were chosen by varying one parameter at a time and in a direction that was thought might give an improved performance.

Table 1: Results of 7 Simulation Runs

Run	1	2	3	4	5	6	7
A	7.586	7.586	7.586	6.597	7.586	7.586	7.586
B	2.794	2.794	2.794	2.794	4.348	2.794	2.794
C	5.282	4.124	5.494	5.282	5.282	5.282	5.282
D	2.313	2.649	4.319	2.313	2.313	2.313	2.313
E	1.601	1.601	1.601	1.601	1.601	2.402	1.601
F	1.273	1.273	1.273	1.273	1.273	1.273	0.909
Y	72.55	72.6	72.67	72.57	72.56	72.55	72.56

Only factors C and D could not be varied independently, and these were instead varied simultaneously in runs 2 and 3, in a way that was physically possible.

Updating the model, using this simulation data, was carried out by calculating the posterior distribution as given by Bayes' formula (5). There are many ways of doing the calculation and many packages exist for doing this. We used the WinBUGS package available on the Internet. The updated values of the parameters are given in Table 2. The updated function  $q$  has a minimum of 72.43 at the point  $\mathbf{x}^* = (7.394, 3.502, 5.091, 3.019, 1.814, 1.362)$ .

There are two points to note about this procedure.

Firstly, the updating moved the minimum from  $\mathbf{x}_0$  to  $\mathbf{x}^*$ . This shows the influence of the observations in changing our view of the minimum point. Note also that the value of the updated minimum is  $q(\mathbf{x}^*) = 72.43$ . This is slight improvement compared with the value at the initial guess of  $q(\mathbf{x}_0) = 72.55$ .

Secondly, the standard deviations of all the  $\boldsymbol{\theta}$  coefficients are reduced. This is reassuring, and suggests that our

Table 2: Prior and Posterior Parameter Distributions

Parameter	Prior		Posterior	
	Mean	Std Dev	Mean	Std Dev
$\theta[1]$	82.92	1	82.93	0.9866
$\theta[2]$	-1.52	1	-1.525	0.8789
$\theta[3]$	-0.56	1	-0.5436	0.9783
$\theta[4]$	-1.06	1	-1.066	0.9322
$\theta[5]$	-0.46	1	-0.4723	0.9744
$\theta[6]$	-0.32	1	-0.3268	0.9915
$\theta[7]$	-0.26	1	-0.2691	0.9991
$\theta[8]$	0.1	1	0.1005	0.3642
$\theta[9]$	0	1	2.84E-04	0.7717
$\theta[10]$	0	1	0.01587	0.7022
$\theta[11]$	0	1	-0.01009	0.7412
$\theta[12]$	0	1	-0.00298	0.7123
$\theta[13]$	0	1	-0.00525	0.6889
$\theta[14]$	0.1	1	0.08672	0.7811
$\theta[15]$	0	1	-0.00476	0.8783
$\theta[16]$	0	1	-0.0107	0.9405
$\theta[17]$	0	1	-0.00605	0.9434
$\theta[18]$	0	1	0.001177	0.9564
$\theta[19]$	0.1	1	0.09316	0.6225
$\theta[20]$	0	1	-0.00251	0.8703
$\theta[21]$	0	1	0.00933	0.8667
$\theta[22]$	0	1	0.005517	0.8551
$\theta[23]$	0.1	1	0.09786	0.7854
$\theta[24]$	0	1	-0.00139	0.956
$\theta[25]$	0	1	0.00652	0.9567
$\theta[26]$	0.1	1	0.09282	0.9263
$\theta[27]$	0	1	-0.00733	0.9843
$\theta[28]$	0.1	1	0.09889	0.9826

knowledge about the coefficients has been improved by the information obtained from the simulation data points.

The posterior estimate  $\mathbf{x}^*$  depends on  $\mathbf{x}_0$ . This dependence will decrease as we get more data. However, the better our initial guess, the faster we converge to the true optimum.

#### 4 POSITIVE DEFINITENESS

A weakness of the above methodology is that there is no guarantee that the posterior distribution of the regression metamodel will possess a minimum, even if we are confident that a minimum exists in the true model. We therefore consider a modification of the above procedure where we use a model for which the quadratic regression function is guaranteed to have a minimum. We write  $\boldsymbol{\theta}$  as 3 separate parameters:  $\mu$ ,  $\mathbf{a}$  and  $\boldsymbol{\Sigma}$ . The regression function  $q(\mathbf{x}, \boldsymbol{\theta})$  can then be written as

$$q(\mathbf{x}, \boldsymbol{\theta}) = \mu + (\mathbf{x} - \mathbf{a})^T \boldsymbol{\Sigma} (\mathbf{x} - \mathbf{a}) \quad (7)$$

where  $\mathbf{\Sigma}$  is symmetric. This is a slightly different definition from that used in (3) which used an essentially upper triangular form. We now ensure that  $\mathbf{\Sigma}$  is positive definite by defining the elements of the matrix to be

$$\Sigma_{ij} = \sum_{k=1}^m \delta_{ik} \delta_{jk} \quad 1 \leq i, j \leq p \quad (8)$$

where  $\sum_{k=1}^m \delta_{ik} = 0$  for  $1 \leq i \leq p$ .

If we assume normal errors, the likelihood of the data, given the parameters  $\mu$ ,  $\mathbf{a}$  and  $\mathbf{\Sigma}$  is given by

$$L(\mathbf{y}|\boldsymbol{\theta}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_{ij} - \mu - (\mathbf{x}_i - \mathbf{a})^T \mathbf{\Sigma}(\mathbf{x}_i - \mathbf{a})]^2\right) \quad (9)$$

We now have to consider the prior for  $\mathbf{a}$  and  $\mathbf{\Sigma}$ . If we consider them as being the mean and variance of a  $p$ -multivariate normal distribution, then in this case, convenient priors for  $\mathbf{a}$  and  $\mathbf{\Sigma}^{-1}$  would be a normal and a Wishart distribution:

$$\mathbf{a}|\mathbf{\Sigma} \sim N_p(\boldsymbol{\xi}, n_0^{-1}\mathbf{\Sigma}); \quad (10)$$

$$\mathbf{\Sigma}^{-1} \sim \text{Wishart}_p(\chi, \mathbf{W}), \quad (11)$$

The prior distribution for  $\mathbf{a}$  is then

$$\pi(\mathbf{a}|\boldsymbol{\xi}, \mathbf{\Sigma}, n_0) \propto \{\det[(n_0^{-1}\mathbf{\Sigma})]\}^{1/2} \exp\left(-\frac{n_0}{2}(\mathbf{a} - \boldsymbol{\xi})^T \mathbf{\Sigma}^{-1}(\mathbf{a} - \boldsymbol{\xi})\right) \quad (12)$$

whilst the prior for  $\mathbf{\Sigma}$  is

$$\pi(\mathbf{\Sigma}|\chi, \mathbf{W}) \propto \{\det \mathbf{\Sigma}\}^{-(\chi + p + 1)/2} \times \exp\left(-\frac{1}{2} \text{trace}\{\mathbf{W}^{-1}\mathbf{\Sigma}^{-1}\}\right) \quad (13)$$

We also need a prior for the other two parameters  $\mu$  and  $\sigma$ . For  $\mu$  we have used a normal prior so that

$$\pi(\mu|\mu_0, \tau) \propto \exp\left(-\frac{1}{2\tau^2}[\mu - \mu_0]^2\right) \quad (14)$$

whilst for  $\sigma$  we have used the Jeffreys reference prior

$$\pi(\sigma) \propto \sigma^{-2}. \quad (15)$$

The posterior distribution is proportional to the product of (9), (12), (13), (14) and (15), i.e.

$$\pi(\mathbf{a}, \mathbf{\Sigma}, \mu, \sigma | \mathbf{y}, \mathbf{x}) \propto L(\mathbf{y} | \boldsymbol{\theta}) \times \pi(\mathbf{a} | \boldsymbol{\xi}, \mathbf{\Sigma}, n_0) \pi(\mathbf{\Sigma} | \chi, \mathbf{W}) \pi(\mu | \mu_0, \tau) \pi(\sigma).$$

We also have that

$$E(\mathbf{\Sigma}) = \frac{\mathbf{W}^{-1}}{\chi - p - 1} \quad (16)$$

and

$$\text{Var}(\Sigma_{ij}) = \frac{2(\omega^{ij})}{(\chi - p - 3)(\chi - p - 1)^2} \quad (17)$$

where

$$\omega^{ij} = (\mathbf{W}^{-1})_{ij}.$$

The hyperparameters,  $n_0$ ,  $\boldsymbol{\xi}$ ,  $\chi$  and  $\mathbf{W}$ ,  $\mu_0$  and  $\tau$  are set by the user, given the available prior information (see Robert (2001) for more detail). In our case they are chosen as follows in order to correspond reasonably closely with the values used in the original formulation of the problem given in Section 2.

We let  $E(\Sigma_{ii}) = 0.1$ . Moreover we set  $\text{Var}(\Sigma_{ii}) = 1$  and because we use a symmetric matrix for  $\mathbf{\Sigma}$ , rather than an upper triangular form we set  $\text{Var}(\Sigma_{ij}) = 1/2$  when  $i \neq j$ . Then use of (16) and (17) where  $p = 6$  in our example yields  $\omega^{ii} = 0.202$   $\omega^{ij} = 0.202/\sqrt{2} \cong 0.1428$  when  $i \neq j$  and  $E(\Sigma_{ij}) = 0.1/\sqrt{2}$ , again when  $i \neq j$ , and  $\chi = 2.02$ .

Also we set  $\mu_0 = 72.5$  so that the mean corresponded approximately to the prior estimate of the optimum as given by the engineers. Similarly the value of  $\xi$  was taken to be the prior estimates of the optimum factor values as given by the engineers. To allow for our uncertainty about the prior values we set  $n_0 = 10^{-10}$  and  $\tau = 10^5$ .

We did not carry out a full Bayesian analysis for this model but instead carried out a numerical optimization to identify the mode of the posterior distribution  $\pi(\mathbf{a}, \mathbf{\Sigma}, \mu, \sigma | \mathbf{y}, \mathbf{x})$ .

Minimization of  $-\log \pi(\mathbf{a}, \Sigma, \mu, \sigma | \mathbf{y}, \mathbf{x})$  was carried out using the Nelder-Mead numerical method. We did not attempt to alter  $\Sigma$  directly but instead, together with  $\mathbf{a}$ ,  $\mu$ , and  $\sigma$ , treated the  $\delta_{ij}$  in (8) as being variables in the Nelder-Mead procedure, subject to  $\sum_{k=1}^m \delta_{ik} = 0$  for  $1 \leq i \leq p$ .

The choice of  $m$  was based on the following consideration. We need to have sufficient flexibility to allow a free choice of the elements of  $\Sigma$ . Now, as it is symmetric, there are  $0.5p(p+1)$  distinct elements in  $\Sigma$ . Now we have  $p(m-1)$  distinct choices for the  $pm$   $\delta_{ij}$ , as they are subject

the  $m$  restrictions  $\sum_{k=1}^m \delta_{ik} = 0$ ,  $1 \leq i \leq p$ . To retain the same degree of freedom of choice we therefore need to have  $p(m-1) \geq 0.5p(p+1)$ , i.e.  $m \geq (3+p)/2$ . With  $p = 6$  we therefore must have  $m \geq 5$ . In the example we took  $m = 7$ .

This gave a minimum at  $\mathbf{x}^* = \hat{\mathbf{a}} = (8.13, 2.86, 5.13, 2.53, 1.76, 1.29)$ ; with

$$\hat{\Sigma} = \begin{pmatrix} 0.0082 & 0.0065 & 0.013 & 0.012 & 0.0024 & 0.011 \\ . & 0.011 & 0.016 & 0.012 & 0.0032 & 0.013 \\ . & . & 0.043 & 0.031 & -0.001 & 0.029 \\ . & . & . & 0.035 & 0.0029 & 0.025 \\ . & . & . & . & 0.0098 & 0.0029 \\ . & . & . & . & . & 0.028 \end{pmatrix}$$

with minimum value 72.54. The value of  $\sigma$  was  $\sigma = 5.45\text{E-}8$ , indicating a good fit to the data points which is to be expected.

The minimized value in this case was actually still close to the initial estimate provided by the engineers.

An interesting aspect of the model is that if we had been able to obtain a set of normal deviates  $\mathbf{Z}$ , with sufficient statistics  $\bar{\mathbf{z}}$  and  $\mathbf{S}$ , directly then the posterior distribution for  $\mathbf{a}$  and  $\Sigma$  is given directly by

$$\mathbf{a} | \xi, \Sigma, \bar{\mathbf{z}}, \mathbf{S} \sim N_p \left( \frac{n_0 \xi + n \bar{\mathbf{z}}}{n + n_0}, \frac{\Sigma}{n + n_0} \right) \quad (18)$$

$$\Sigma^{-1} | \bar{\mathbf{z}}, \mathbf{S} \sim W_p(\chi + n, \mathbf{W}_1(\bar{\mathbf{z}}, \mathbf{S}))$$

where

$$\mathbf{W}_1(\bar{\mathbf{z}}, \mathbf{S})^{-1} = \mathbf{W}^{-1} + \mathbf{S} + \frac{nm_0}{n + n_0} (\bar{\mathbf{z}} - \xi)(\bar{\mathbf{z}} - \xi)^T$$

which is written in terms of the sufficient statistics for  $\mathbf{Z}$

$$\bar{\mathbf{z}} = \frac{1}{m} \sum_{i=1}^m \mathbf{z}_i \quad \mathbf{S} = \sum_{i=1}^m (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \quad (19)$$

## 5 SUMMARY AND CONCLUSIONS

In summary, we have shown how simulation metamodels can be used to find the optimal factor settings for a system.

The example considered illustrates how Bayesian methods can be used in practice when time permits only a few runs of a simulation model to be made, but the number of factors to be considered is large.

When the response surface is constrained to be positive definite, the optimum obtained is similar to that obtained without this constraint. Having a positive definite response surface also ensures that a minimum does exist.

Although we have not constructed a full posterior distribution for the data, we have carried out an estimate of optimal coefficient values (using modal estimators) which incorporates prior knowledge to better inform the optimization routine used to find the modal estimators. In the future, we hope to refine this method to find the true posterior distribution of the parameters.

## REFERENCES

- Azadivar, F. 1999. "Simulation Optimization Methodologies." In *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 93-100.
- Barton, R. R. 1998. "Simulation Metamodels." In *Proceedings of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 167-174.
- Cheng, R. C. H. 1999. "Regression Metamodelling using Bayesian Methods." In *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 330-335.
- Cheng, R. C. H. 2004. Optimization of Systems by Simulation Metamodelling Methods. In *Proceedings of OR Society Simulation Workshop 2004*, S. Robinson and S. Taylor, eds., Birmingham: OR Society. 39-44.
- Chick, S. E. 2000. "Bayesian Methods for Simulation." In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., IEEE, Piscataway, NJ. 109-118.
- Cochran, W. G. and Cox, G. M. 1957. "Experimental Designs, 2<sup>nd</sup> Ed.", New York: John Wiley.
- Fu, M. C., S. Andradottir, J.S. Carson, F. Glover, C.R. Harrel, Y.-C. Ho, J.P. Kelly and S.M. Robinson. 2000. "Integrating Optimization and Simulation: Research and Practice." In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 610-616.

- Glover, F., J.P. Kelly and M. Laguna. 1999. "New advances for wedding optimization and simulation." In *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 255-260.
- Gonda, H., H.G. Neddermeijer, J. Gerrit, G.J. van Oortmarssen, N. Piersma and R.A. Dekker. 2000. "Framework for Response Surface Methodology for Simulation Optimization." In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 129-136.
- Hood, S. J. and P.D. Welch. 1993. "Response surface methodology and its application in simulation." In *Proceedings of the 1993 Winter Simulation Conference*, G.W. Evans, M. Mollaghasemi, E.C. Russell, and W.E. Biles, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 115-122.
- Kelton, W. D. 1999. "Designing Simulation Experiments." In *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 33-38.
- Law, A. M. and M.G. McComas. 2000. "Simulation-Based Optimization." In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 46-49.
- Robert, C. 2001. "The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation." New York: Springer-Verlag.
- Swisher, J. R., S.H. Jacobson, P.D. Hyden and L.W. Schruben. 2000. "A Survey of simulation optimization techniques and procedures." In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., Piscataway, NJ: Institute of Electronics and Electrical Engineers. 119-128.

## AUTHOR BIOGRAPHIES

**RUSSELL C.H. CHENG** is Professor, Head of Operational Research, and Deputy Dean of the Faculty of Mathematical Studies at the University of Southampton. He has an M.A. and the Diploma in Mathematical Statistics from Cambridge University, England. He obtained his Ph.D. from Bath University. He is a former Chairman of the U.K. Simulation Society, a Fellow of the Royal Statistical Society, Member of the Operational Research Society. His research interests include: variance reduction methods and parametric estimation methods. He was a Joint Editor of the *IMA Journal of Management Mathematics*. His email and web addresses are R.C.H.Cheng@maths.soton.ac.uk and <www.maths.soton.ac.uk/staff/Cheng>.

**CHRISTINE CURRIE** is a PhD student at the University of Southampton. She has an MPhys from Oxford University and an MSc in Operational Research from the University of Southampton. Her research interests include mathematical modeling of epidemics, Bayesian statistics and variance reduction methods. Her email address is: <C.S.M.Currie@maths.soton.ac.uk>.