ACCOUNTING FOR PARAMETER UNCERTAINTY
IN SIMULATION INPUT MODELING

Faker Zouaoui

Sabre, Inc.
Research Group
Southlake, TX 76092, U.S.A.

James R. Wilson

Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695-7906, U.S.A.

ABSTRACT

We formulate and evaluate a Bayesian approach to probabilistic input modeling. Taking into account the parameter and stochastic uncertainties inherent in most simulations, this approach yields valid predictive inferences about the output quantities of interest. We use prior information to construct prior distributions on the input-model parameters. Combining this prior information with the likelihood function of sample data observed on the input processes, we compute the posterior parameter distributions using Bayes' rule. This leads to a Bayesian Simulation Replication Algorithm in which: (a) we estimate the parameter uncertainty by sampling from the posterior distribution of the input model's parameters on selected simulation runs; and (b) we estimate the stochastic uncertainty by multiple independent replications of those selected runs. We also formulate some performance evaluation criteria that are reasonable within both the Bayesian and frequentist paradigms. An experimental performance evaluation demonstrates the advantages of the Bayesian approach versus conventional frequentist techniques.

1 INTRODUCTION

Discrete-event simulations, especially those modeling complex systems, are almost all driven by random input processes. A simulation experiment therefore typically requires a number of streams of random variates drawn from specified distributions or input models. The inherent variation in the output of a simulation experiment arising from its dependence on these random inputs is often called *stochastic uncertainty* (Helton 1998). We generally assume that the input models driving the simulation belong to known parametric families. However, uncertainty typically occurs when choosing between different input models. We refer to this second source of variation as *model uncertainty* (Raftery et al. 1995). The parameters on which these models depend are usually assumed to be fixed. In practice,

these parameters are estimated from subjective information (expert opinion) or from real data observed on the input processes. The estimation of unknown parameters gives rise to another source of variation often referred to in the literature as *parameter uncertainty* (Raftery et al. 1995).

Cheng and Holland (1997) consider two methods of assessing how the variation in the simulation output depends on parameter uncertainty and stochastic uncertainty. The first method is based on classical differential analysis, or the δ-method (Stuart and Ord 1994). The main result of Cheng and Holland (1997) shows that under general conditions the total variation in the simulation output is composed of two distinct terms, depending respectively on the parameter uncertainty and stochastic uncertainty. One problem with this method is that certain sensitivity coefficients have to be estimated, and the effort needed to do this increases linearly with the number of unknown input parameters. Moreover, when the number of parameters is large, a problem can occur with spurious variation overinflating the variance estimate. The second method that Cheng and Holland (1997) consider for assessing the variation in the simulation output is the parametric bootstrap (Efron and Tibshirani 1993). Although computationally more expensive, this method does not suffer from the difficulties of the δ-method. It can also be more competitive on the grounds of computational efficiency if the number of unknown parameters is large.

Both of the above methods rely on the assumption that the parameters of the input models are unknown but deterministic quantities. Moreover, the output inferences are implicitly conditional on the selected single input model. The objective of the simulation experiment is therefore to estimate the mean output response as a function of the "true" but unknown parameter values. The most fundamental problem with such approaches to input model selection is that conditional on a single input model and on given values of the parameters for that input model, the output inference underestimates the overall (unconditional) uncertainty in the output quantities of interest, sometimes to a dramatic extent (Kass and Raftery 1995). Moreover, the usual approach to

model selection in the simulation community is commonly guided by a series of goodness-of-fit tests (Law and Kelton 2000). These tests can be highly misleading and very difficult to interpret in a classical statistical framework (Berger and Delampady 1987).

All these difficulties can be avoided, if one adopts a Bayesian approach that incorporates prior information on competing models and their parameters in a rigorous manner. We can compute the posterior probabilities using Bayes' rule for all competing models and their parameters; and then we can make a composite inference that takes account of model and parameter uncertainty in a formally justifiable way (Zouaoui 2001). Even if prior information is not readily available, there are methods to perform a full Bayesian analysis that rely on noninformative priors and thus will give more weight to the observed data, but will still incorporate the model and parameter uncertainties that are due to our lack of knowledge about the input processes driving our simulation experiment. These methods generalize the classical inferences conditional on the choice of a single input model and its parameters, and they work for both small and large sample sizes.

In this paper, we use a Bayesian approach to account for both stochastic uncertainty and parameter uncertainty in simulation; and we estimate the effects of these sources of uncertainty on the output quantity of interest. Hence, we assume that the functional forms of all input models are known, perhaps based on our prior knowledge of the processes driving the simulation model—a situation that sometimes occurs in simulation applications. In Zouaoui and Wilson (2001b), we extend our approach to account for model uncertainty as well as parameter uncertainty and stochastic uncertainty.

The rest of this paper is organized as follows. In Section 2 we define the symbolism used to describe our layout for probabilistic simulation experiments. In Section 3 we survey some recent developments concerning the use of the bootstrap method for classical frequentist simulation input modeling. In Section 4 we detail our Bayesian framework for handling parameter and stochastic uncertainties in discrete-event simulation; and this development leads to our "Bayesian Simulation Replication Algorithm" for designing simulation experiments. To evaluate the performance of both Bayesian and frequentist input modeling techniques on a fair and consistent basis, in Section 5 we formulate appropriate evaluation criteria and illustrate the application of these criteria to a simulation of a single-server queue. For more complete details on the development presented in this paper, see Zouaoui and Wilson (2001a).

2 THE SIMULATION EXPERIMENT

A simulation experiment, in its basic form, consists of making m independent runs of the simulation model and observing a single output performance measure of interest, y, from each run. For simplicity, we assume that the simulation model is driven by a single sequence $\{X_1, X_2, \ldots\}$ of independent and identically distributed input random variables, from which we observe the random sample $\boldsymbol{x} = (x_1, \ldots, x_n)$. Multiple independent random input sequences will be treated in a similar fashion.

During the jth simulation run, a stream of random numbers $\boldsymbol{u}_j$ is generated internally within the simulation model. This stream is used to generate the input random variates $\tilde{\boldsymbol{x}}_j$ by some transformation method, from which the output y_j is computed. It is more convenient to think of both streams as finite for a fixed simulation run length.

One possible method for generating the input random variates $\tilde{\boldsymbol{x}}_j$, which is commonly used by the simulation software packages, is the inverse transform method. If we let u_{ji} denote the ith random number sampled on the jth simulation run, then $\tilde{x}_{ji}$ can be generated using the inverse transform method as

$$\tilde{x}_{ji} = G_M^{-1}(u_{ji}, \theta_M), \tag{1}$$

where $G_M^{-1}(\cdot)$ is the inverse of the distribution function $G_M(\cdot, \theta_M)$ of the simulation input model M, having θ_M as its d_M-dimensional vector of parameters. Given θ_M, we shall assume in this paper that the conditional distributions of $\tilde{X}_{ji}$ and X_i are the same. Hence, the real data observations are assumed to have been drawn from the distribution $G_M(x, \theta_M)$.

The model and parameter uncertainties are represented by the random variables M and θ_M, respectively, both of which are assumed to depend only on the subjective information or data observed on the target input processes; and the stochastic uncertainty depends only on the randomness of $\boldsymbol{u}$. Thus the output of interest from the simulation run, y, can be regarded as an unknown complicated function of $\boldsymbol{u}$, M, and θ_M,

$$y = y(\boldsymbol{u}, M, \theta_M). \tag{2}$$

In this paper, we limit our attention to the effects of parameter uncertainty and stochastic uncertainty on the distribution of y. For simplicity we drop M from our subsequent expressions, recognizing that they are implicitly dependent on the input model M. (In Zouaoui and Wilson (2001b), we will relax the simplifying assumption that the input model M is known.) Thus equation (2) becomes

$$y = y(\boldsymbol{u}, \theta), \tag{3}$$

and we let

$$\eta(\theta) = \int y(\boldsymbol{u}, \theta) \, d\boldsymbol{u} \tag{4}$$

denote the expected value of y given θ.

The objective of a classical simulation experiment is generally to estimate $\eta(\theta_0)$, where θ_0 is the true but unknown parameter value, estimated separately from the simulation experiment using real data. It is also of interest to compute a measure of the variability of the simulation output, from which a confidence interval for $\eta(\theta_0)$ can be constructed.

3 CONVENTIONAL APPROACHES TO ESTIMATING PARAMETER AND STOCHASTIC UNCERTAINTY

From the structure of the simulation model described in the previous section, we see that the responses or outputs of the simulation runs for a fixed parameter vector $\theta \in \mathbb{R}^d$ can be written as

$$y_j = y(\boldsymbol{u}_j, \theta) = \eta(\theta) + e_j(\boldsymbol{u}_j, \theta), \quad j = 1, \ldots, m.$$

The error variable e_j is the random difference between the output of the jth simulation run and $\eta(\theta)$. We generally assume that $\mathrm{E}(e_j|\theta) = 0$ and $\mathrm{Var}(e_j|\theta) = \tau^2(\theta)$ for $j = 1, \ldots, m$, so that $\mathrm{E}(y_j|\theta) = \eta(\theta)$. Hence the mean of the simulation outputs,

$$\overline{y} = \frac{1}{m} \sum_{j=1}^{m} y_j,\qquad (5)$$

is an unbiased estimator of $\eta(\theta)$.

If the maximum likelihood estimator $\widehat{\theta}$ of θ_0 is used in (3), then the output is $y_j = y(\boldsymbol{u}_j, \widehat{\theta}) = \eta(\widehat{\theta}) + e_j(\boldsymbol{u}_j, \widehat{\theta})$ for $j = 1, \ldots, m$, where $\widehat{\theta}$ and $\boldsymbol{u}_j$ are independent random vectors. We have the variance decomposition

$$
\begin{aligned}
\mathrm{Var}(y_j) &= \mathrm{Var}_{\widehat{\theta}}\{\mathrm{E}[\, y_j(\boldsymbol{u}_j, \widehat{\theta}) \,|\, \widehat{\theta}\,]\} \\
&\quad + \mathrm{E}_{\widehat{\theta}}\{\mathrm{Var}[\, y_j(\boldsymbol{u}_j, \widehat{\theta}) \,|\, \widehat{\theta}\,]\} \\
&= V_{\mathrm{par}} + V_{\mathrm{sto}}, \qquad (6)
\end{aligned}
$$

in which V_{par} denotes the parameter variance and V_{sto} denotes the stochastic or simulation variance.

It is generally easy to estimate the stochastic variance V_{sto} in (6). Based on m independent replications, the most commonly used estimator is

$$\widehat{V}_{\mathrm{sto}} = \frac{1}{m-1} \sum_{j=1}^{m} (y_j - \overline{y})^2, \qquad (7)$$

where $y_j = y(\boldsymbol{u}_j, \widehat{\theta})$ is the output of the jth replication, and $\overline{y}$ is the average of all y_j's and the estimate of the mean response. The parameter variance V_{par} in (6) is more

difficult to estimate. Cheng and Holland (1997) developed a finite-difference estimator of V_{par}.

In the rest of this section, we review a bootstrap method (Efron 1979) for estimating V_{par}. Following the notation of Section 2, we substitute the estimator $\widehat{\theta}$ into $G(x, \theta)$ to obtain the fitted distribution $G(x, \widehat{\theta})$. We then draw a sample $\boldsymbol{x}_1^* = (x_{11}^*, x_{12}^*, \ldots, x_{1n}^*)$ from $G(x, \widehat{\theta})$, perhaps via the inverse transform (1). Corresponding to this bootstrap sample is the bootstrap estimate $\widehat{\theta}^*(1)$ of θ_0, computed from $\boldsymbol{x}_1^*$ in exactly the same way that $\widehat{\theta}$ was computed from $\boldsymbol{x}$. Repeating this sampling-and-estimation operation independently B times yields the estimates $\{\widehat{\theta}^*(i) : i = 1, \ldots, B\}$. We can then carry out B bootstrap simulation experiments, one for each $\widehat{\theta}^*(i)$, where each run using the input parameter vector $\widehat{\theta}^*(i)$ has the same length as in the original experiment; and altogether m' simulation runs are performed using each $\widehat{\theta}^*(i)$. For $i = 1, \ldots, B$, this yields the ith set of simulation-generated output responses $\{y_{ij}^* : j = 1, \ldots, m'\}$ from which we calculate the sample mean $\overline{y}_i^* = \sum_{j=1}^{m'} y_{ij}^*/m'$. Let $\overline{\overline{y}}^* = \sum_{i=1}^{B} \overline{y}_i^*/B$ denote the grand mean of the bootstrap sample means.

Cheng and Holland (1997) show how the variance of $\overline{y}_i^*$ depends on both the stochastic variance and the parameter variance. When each bootstrap is an exact replica of the original experiment (i.e. $m' = m$), Cheng and Holland suggest using the sample variance of the $\{\overline{y}_i^*\}$,

$$S_B^2 = \frac{1}{B-1} \sum_{i=1}^{B} \left(\overline{y}_i^* - \overline{\overline{y}}^*\right)^2, \qquad (8)$$

as an estimate for the variance of the original sample mean $\overline{y}$ specified in (5). Using (7) as an estimate for V_{sto} and (8), we estimate the parameter variance by

$$\widehat{V}_{\mathrm{par}} = S_B^2 - \widehat{V}_{\mathrm{sto}}. \qquad (9)$$

An alternative bootstrap-based approach to estimating V_{par} is to compute $\widehat{V}_{\mathrm{sto}}^*(i) = \sum_{j=1}^{m'} \left(y_{ij}^* - \overline{y}_i^*\right)^2/(m'-1)$ as an additional estimate of $\tau^2(\theta_0)$ for $i = 1, \ldots, B$; and then we can use $\widehat{V}_{\mathrm{sto}}^* = \left[\widehat{V}_{\mathrm{sto}} + \sum_{i=1}^{B} \widehat{V}_{\mathrm{sto}}^*(i)\right]/(B+1)$ in place of $\widehat{V}_{\mathrm{sto}}$ in (9).

Finally, we discuss how we can construct a bootstrap confidence interval for $\eta(\theta_0)$. One possibility is to construct a classical t-type confidence interval using the bootstrap variance estimates given in this section. However, this interval generally performs poorly in practice. Efron and Tibshirani (1993) explain this phenomenon and suggest methods to overcome such a drawback. A confidence interval that generally behaves better for such cases is the bootstrap

percentile-type confidence interval. From the order statistics $\overline{y}_{(1)}^* \leq \overline{y}_{(2)}^* \leq \cdots \leq \overline{y}_{(B)}^*$ of the $\{\overline{y}_i^* : i = 1, \ldots, B\}$, we obtain the $100(1 - \alpha)\%$ percentile-type confidence interval

$$[\beta_L^*, \beta_U^*] \approx [\overline{y}_{(\lceil B\alpha/2 \rceil)}^*, \overline{y}_{(\lceil B(1-\alpha/2) \rceil)}^*]. \qquad (10)$$

4 BAYESIAN APPROACH

We describe in this section the applicability of the Bayesian approach to our basic structure (3) of the discrete-event simulation model. We provide methods to estimate the posterior mean response and to construct a confidence interval for that quantity.

4.1 Estimating Mean Response

We observe a random sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ from our selected input model. Let θ be its d-dimensional vector of parameters with prior distribution $p(\theta)$. We assume that the hyperparameters of the prior distribution are either known or estimated using moment matching or some other empirical Bayes method (Carlin and Louis 1996). We are not being fully Bayesian here by stopping the hierarchy at the second stage. However, we argue here that in most simulation applications in operations research and industrial engineering (as opposed, for example to applications in econometrics), prior information will be generally vague if it exists at more than one level down the hierarchy.

For our basic simulation model structure, we derive the posterior mean response given $\boldsymbol{x}$.

Theorem 1 *If the simulation response y has the form (3), then*

$$\mathrm{E}(y|\boldsymbol{x}) = \int \eta(\theta)\, p(\theta|\boldsymbol{x})\, d\theta.$$

Zouaoui (2001) and Zouaoui and Wilson (2001a) contain proofs of all results presented in this paper.

Our first objective is to develop an approach that accounts fully for parameter and stochastic uncertainty and that can be extended easily to account for model uncertainty. This approach should also have a theoretical appeal, and most importantly it should perform well in practice. Figure 1 summarizes the Bayesian Simulation Replication Algorithm that we have developed to implement a Bayesian approach to simulation input modeling. The net effect of the algorithm is to account fully for the uncertainty in the parameters of the input model as well as to account fully for the usual stochastic uncertainty. The algorithm can be seen as an uncertainty decomposition algorithm. The inner loop will be used to generate estimates of the stochastic uncertainty, whereas the outer loop will estimate the parameter uncertainty. The next subsection gives a detailed explanation of how to estimate the stochastic and param-

```
for r = 1, . . . , R
    generate the parameter vector θʳ from p(θ|x)
    set the parameter vector θ ← θʳ
        for j = 1, . . . , m
            set the random-number input u ← uⱼ
            perform the simulation run using u and θ
            calculate the output response y_rj = y(u, θ)
        end loop
    compute ȳ_r = ∑_{j=1}^{m} y_rj / m
end loop
compute the grand mean ȳ̄ = ∑_{r=1}^{R} ȳ_r / R
    as an estimate for E(y|x)
```

Figure 1: Bayesian Simulation Replication Algorithm

eter variances and construct a confidence interval on the posterior mean response.

4.2 Assessing Output Variability

We try to assess the variability of the simulation output based on simple response surface models, given the objective of estimating the mean response in our simulation study. The following analysis can be extended to more complicated models, but such an extension is beyond the scope of this paper. We develop two methods of estimating the parameter and stochastic variances in the simulation output.

4.2.1 Classical Output Analysis

Following our basic simulation structure of Section 2, we see that the output responses from the simulation runs performed by the Bayesian Simulation Replication Algorithm of Figure 1 are given by

$$y_{rj} = y(\boldsymbol{u}_j, \theta^r) = \eta(\theta^r) + e_j(\boldsymbol{u}_j, \theta^r), \qquad (11)$$

for $r = 1, \ldots, R$ and $j = 1, \ldots, m$. We generally assume that

$$\mathrm{E}(e_j|\theta^r) = 0 \quad \text{and} \quad \mathrm{Var}(e_j|\theta^r) = \tau^2, \qquad (12)$$

where τ^2 does not depend on θ^r. Given that our main objective is to estimate the overall mean response, we further assume that

$$\eta(\theta^r) = \beta + \delta_r(\theta^r), \qquad (13)$$

where $\beta = \beta(\boldsymbol{x}) = E_{\theta^r}[\eta(\theta^r)] = \int \eta(\theta) \, p(\theta|\boldsymbol{x}) \, d\theta = E(y|\boldsymbol{x})$ from Theorem 1; and

$$E_{\theta^r}(\delta_r) = 0 \quad \text{and} \quad \text{Var}_{\theta^r}(\delta_r) = \sigma^2. \qquad (14)$$

Based on these assumptions, the posterior variance can be written as the sum of two variances measuring the stochastic and parameter uncertainty, respectively.

Theorem 2 *If (11)–(14) hold, then*

$$\text{Var}(y|\boldsymbol{x}) = \tau^2 + \sigma^2.$$

The response surface model given by (11)–(14) is known in the statistical literature as the classical random-effects model (Rao 1997), where one estimates β, τ^2, and σ^2 using the simulation-generated statistics specified in Figure 1 as follows: the posterior mean response is estimated by

$$\widehat{\beta} = \overline{\overline{y}}; \qquad (15)$$

the variance due to stochastic uncertainty is estimated by

$$\widehat{\tau}^2 = \frac{\displaystyle\sum_{r=1}^{R} \sum_{j=1}^{m} (y_{rj} - \overline{y}_r)^2}{R(m-1)}; \qquad (16)$$

and finally the variance due to parameter uncertainty is estimated by

$$\widehat{\sigma}^2 = \frac{\displaystyle\sum_{r=1}^{R} (\overline{y}_r - \overline{\overline{y}})^2}{(R-1)} - \frac{\widehat{\tau}^2}{m}. \qquad (17)$$

In many applications, especially the ones where we are totally ignorant about the output performance measure, the above formulation delivers reasonable point estimates. However, there are two important drawbacks of using such an approach. The first drawback concerns the estimate $\widehat{\sigma}^2$, which can be negative. The problem of a negative estimate for a variance component can be avoided by setting it equal to zero, but this creates new issues (Rao 1997). The main drawback comes from the fact that substituting point estimates for the above parameters ignores our real uncertainty about them. We suggest in §4.2.2 a full Bayesian treatment of the same model with noninformative priors based on normally distributed simulation output observations.

In addition to point estimates, we can also construct a confidence interval for β from the output of the Bayesian Simulation Replication Algorithm given in Figure 1. Similar to the bootstrap approach, we can construct an approximate $100(1-\alpha)\%$ Bayesian percentile confidence interval for β

as

$$[\beta_{\mathrm{L}}, \beta_{\mathrm{U}}] \approx [\overline{y}_{(\lceil R\alpha/2\rceil)}, \overline{y}_{(\lceil R(1-\alpha/2)\rceil)}], \qquad (18)$$

where the quantities $\overline{y}_{(1)} \le \overline{y}_{(2)} \le \cdots \le \overline{y}_{(R)}$ denote the order statistics of the $\{\overline{y}_r : r = 1, \ldots, R\}$ specified in Figure 1. These intervals generally perform better in practice than t-type confidence intervals (Efron and Tibshirani 1993).

4.2.2 Bayesian Output Analysis

Under the hierarchical normal model, we assume that the output data $\{y_{rj} : r = 1, \ldots, R \text{ and } j = 1, \ldots, m\}$ specified in Figure 1 are independently normally distributed, that is

$$y_{rj} \mid \mu_r, \tau \sim \mathrm{N}(\mu_r, \tau^2). \qquad (19)$$

For $r = 1, \ldots, R$, we also assume that the parameter μ_r, which corresponds to $\eta(\theta^r)$ in equation (11), is also normally distributed:

$$\mu_r \mid \beta, \sigma \sim \mathrm{N}(\beta, \sigma^2), \ r = 1, \ldots, R. \qquad (20)$$

Note that although assumption (19) generally holds in practice because of the simulation output usually being an average of a large number of output random variables, assumption (20) may not be consistent with the form of the posterior distribution $p(\theta|\boldsymbol{x})$.

To complete the Bayesian formulation, we assume a noninformative prior distribution for (β, τ, σ), with $\sigma > 0$ and $\tau > 0$; specifically we take $p(\beta, \tau, \sigma) \propto \tau^{-1}$. It follows from this assumption that the joint posterior density of all parameters satisfies the relation

$$p(\boldsymbol{\mu}, \beta, \tau, \sigma \mid \boldsymbol{y})$$
$$\propto \ \frac{1}{\tau} \prod_{r=1}^{R} p(\mu_r \mid \beta, \sigma) \prod_{r=1}^{R} \prod_{j=1}^{m} p(y_{rj} \mid \mu_r, \tau), \quad (21)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_R)$ and $\boldsymbol{y} = \{y_{rj} : r = 1, \ldots, R; \ j = 1, \ldots, m\}$.

Many numerical methods such as conditional maximization can be used to obtain posterior point estimates for the parameters of interest. However, we are also interested to compute posterior confidence intervals for these parameters; and Markov Chain Monte Carlo (MCMC) methods are appropriate for such inferences. In fact, we can even obtain a large sample from the posterior distribution of each parameter from which we can estimate its density.

The idea behind MCMC (Gilks et al. 1996) is to simulate a random walk in the space of $(\boldsymbol{\mu}, \beta, \tau, \sigma)$ which converges to a stationary distribution that is the joint posterior distribution $p(\boldsymbol{\mu}, \beta, \tau, \sigma \mid \boldsymbol{y})$. The most widely used MCMC method is the *Gibbs Sampler* algorithm (Casella and George 1992). Figure 2 summarizes briefly the steps of the algorithm. For

our problem this algorithm can be easily implemented in any statistical software package, given its simplicity and the fact that we can generate variates easily from the following required conditional distributions (Gelman et al. 1995):

$$\mu_r \mid \beta, \tau, \sigma, \boldsymbol{y} \sim \mathrm{N}\left(\frac{\frac{\beta}{\sigma^2} + \frac{m\overline{y}_r}{\tau^2}}{\frac{1}{\sigma^2} + \frac{m}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2} + \frac{m}{\tau^2}}\right), \qquad (22)$$

for $r = 1, \ldots, R$;

$$\beta \mid \boldsymbol{\mu}, \tau, \sigma, \boldsymbol{y} \sim \mathrm{N}\left(\frac{1}{R}\sum_{r=1}^{R}\mu_r, \frac{\sigma^2}{R}\right); \qquad (23)$$

$$\tau^2 \mid \boldsymbol{\mu}, \beta, \sigma, \boldsymbol{y} \sim \mathrm{IG}\left(\frac{mR}{2}, \frac{1}{2}\sum_{r=1}^{R}\sum_{j=1}^{m}(y_{rj} - \mu_r)^2\right); \quad (24)$$

and

$$\sigma^2 \mid \boldsymbol{\mu}, \beta, \tau, \boldsymbol{y} \sim \mathrm{IG}\left(\frac{R-1}{2}, \frac{1}{2}\sum_{r=1}^{R}(\mu_r - \beta)^2\right), \qquad (25)$$

where the density function of a random variable Z having an Inverse Gamma distribution $\mathrm{IG}(\nu, \phi)$ with shape parameter ν and scale parameter ϕ is defined as

$$p(z|\nu, \phi) = \frac{\phi^\nu e^{-\phi/z}}{\Gamma(\nu)z^{\nu+1}}, \ z > 0, \ \nu > 0, \ \phi > 0,$$

and where $\Gamma(\cdot)$ denotes the gamma function. The classical estimates of β, τ^2, and σ^2 given by equations (15), (16), and (17), respectively, are good initial estimates with which to start the Gibbs sampler algorithm. If the initial estimate $\widehat{\sigma}^2$ is negative, then we can set it to a small value, say 0.001. The output inference from the Gibbs sampler algorithm is based on $T = 50000$ iterations, with a warm-up period of $T^* = 5000$ iterations. These relatively large numbers of iterations are generally required for convergence because of the high autocorrelation between successively sampled values of some parameters (Spiegelhalter et al. 1996). The point estimates of β, τ^2, and σ^2 are given in Figure 2. A $100(1 - \alpha)\%$ credible interval for β can be constructed from the output of the Gibbs sampler algorithm as

$$[\beta_\mathrm{L}, \beta_\mathrm{U}] \approx [\beta_{(\lceil(T-T^*)\alpha/2\rceil)}, \beta_{(\lceil(T-T^*)(1-\alpha/2)\rceil)}], \qquad (26)$$

where the quantities $\beta_{(1)} \leq \beta_{(2)} \leq \cdots \leq \beta_{(T-T^*)}$ denote the order statistics of the $\{\beta_t : t = T^*+1, \ldots, T\}$ generated by the Gibbs sampler algorithm of Figure 2. Credible intervals similar to (26) can also be constructed for τ^2 and σ^2.

Set β_0, τ_0^2, and σ_0^2 to their classical estimates
   (15), (16), and (17), respectively.
For $t = 1, \ldots, T^*, T^*+1, \ldots, T$, generate:
   $\mu_{rt} \sim p\left(\mu_r \mid \beta_{t-1}, \tau_{t-1}, \sigma_{t-1}, \boldsymbol{y}\right)$ as in (22),
      for $r = 1, \ldots, R$, and
      take $\boldsymbol{\mu}_t = (\mu_{1t}, \ldots, \mu_{Rt})$;
   $\beta_t \sim p\left(\beta \mid \boldsymbol{\mu}_t, \tau_{t-1}, \sigma_{t-1}, \boldsymbol{y}\right)$ as in (23);
   $\tau_t^2 \sim p\left(\tau^2 \mid \boldsymbol{\mu}_t, \beta_t, \sigma_{t-1}, \boldsymbol{y}\right)$ as in (24); and
   $\sigma_t^2 \sim p\left(\sigma^2 \mid \boldsymbol{\mu}_t, \beta_t, \tau_t, \boldsymbol{y}\right)$ as in (25).
end loop
Compute estimates for the parameters of interest:
   $\widehat{\beta} = \sum_{t=T^*+1}^{T} \beta_t / (T - T^*)$
   $\widehat{\tau}^2 = \sum_{t=T^*+1}^{T} \tau_t^2 / (T - T^*)$
   $\widehat{\sigma}^2 = \sum_{t=T^*+1}^{T} \sigma_t^2 / (T - T^*)$

Figure 2: Gibbs Sampler Algorithm

5 PERFORMANCE EVALUATION

In this section we propose some criteria for comparing Bayesian and frequentist approaches within the discrete-event simulation framework. We will test our ideas empirically by applying them to a queueing system. In the following discussion, we take the view that a frequentist is an analyst who seeks to estimate an unknown parameter based only on the model that has been adopted for the observable data, whereas a Bayesian is one who seeks to estimate the parameter by appropriately combining his or her prior intuition with the information content in the data.

5.1 Evaluation Criteria

5.1.1 Point Estimation

Suppose that a sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ of independent random variables is to be drawn; and given the value of a scalar parameter θ_0, $\boldsymbol{x}$ has the conditional probability density $g_{\theta_0}(\boldsymbol{x})$. We assume that the unknown parameter θ_0 is a random variable and refer to the distribution π_0 of θ_0 as the "true prior" distribution. Here it is important to distinguish between (a) the "true" parameter vector θ_0 of the inputs $\boldsymbol{x}$ to the real system with response y_0 versus (b) the parameter θ of the input model for the simulation with response $y = y(\boldsymbol{u}, \theta)$, even when θ is sampled from the posterior distribution $p(\theta|\boldsymbol{x})$ as in the Simulation Replication Algorithm of Figure 1.

For a realization θ_0 obtained from π_0 (which will be unknown to the Bayesian and frequentist analysts), we as-

sume that we can compute the true output performance measure of the real system $\beta_0(\theta_0) = \text{E}(y_0|\theta_0)$ (which will also be unknown to both analysts). The analysts will then attempt to estimate $\beta = \beta(\boldsymbol{x}) = \text{E}(y|\boldsymbol{x})$ using their respective simulation procedures based solely on observing the experimental outcomes $\{x_1, x_2, \ldots, x_n\}$ drawn from g_{θ_0}.

Our criterion for assessing the performance of a given estimator $\widehat{\beta}$ of $\beta_0(\theta_0)$ will be the average risk $\mathcal{R}$ of $\widehat{\beta}$ relative to the true prior distribution π_0 of θ_0,

$$\mathcal{R}(\pi_0, \widehat{\beta}) = \text{E}_{\pi_0}\Big(\text{E}_{g_{\theta_0}}\Big\{\big[\widehat{\beta}(\boldsymbol{x}) - \beta_0(\theta_0)\big]^2\Big\}\Big) \qquad (27)$$

$$= \int \int \big[\widehat{\beta}(\boldsymbol{x}) - \beta_0(\theta_0)\big]^2 \, g_{\theta_0}(\boldsymbol{x})\pi_0(\theta_0) \, d\boldsymbol{x} \, d\theta_0,$$

where we take the simulation-based estimator $\widehat{\beta}(\boldsymbol{x}) = \overline{\overline{y}}$ when we adopt the Bayesian Simulation Replication Algorithm of Figure 1. When we adopt the frequentist approach, we compute the maximum likelihood estimator $\widehat{\theta}^{\text{mle}} = \widehat{\theta}^{\text{mle}}(\boldsymbol{x})$ based on the original input data, and then in (27) we take

$$\widehat{\beta}(\boldsymbol{x}) = \frac{1}{m}\sum_{j=1}^{m} y\big(\boldsymbol{u}_j, \widehat{\theta}^{\text{mle}}(\boldsymbol{x})\big) \qquad (28)$$

based on a set of m independent simulation runs.

In the Bayesian approach we will use the "operational" prior π, chosen by the Bayesian analyst, to compute the posterior distribution $p(\theta|\boldsymbol{x})$ of θ. The operational prior will generally be noninformative or minimally informative. It is also important to stress that the true prior π_0 is entirely unknown to the Bayesian and frequentist analysts. The real contest will be to find the estimator that minimizes $\mathcal{R}(\pi_0, \widehat{\beta})$.

The criterion we propose for assessing performance makes sense and is reasonable within the Bayesian and frequentist paradigms. Although the notion of a "true prior" seemingly conflicts with the "degree of belief" interpretation of prior distributions espoused by many Bayesians, it should be noted that the Bayesians' degree of belief about θ is reflected in π, not π_0, and that the idea of a true prior distribution makes perfect sense in a computer-assisted experiment where θ_0 is generated at random from π_0, whose form is unknown to both analysts. If the true θ_0 is constant and π_0 is taken as degenerate at that constant, then the average risk reduces to the mean squared error (MSE), which is widely used by frequentists to judge estimators. In our framework, the average risk is precisely the squared error in estimating $\beta_0(\theta_0)$ by $\widehat{\beta}$, averaged over all the randomness in the problem.

5.1.2 Interval Estimation

The Bayesian approach can be also used to develop interval estimation procedures (§4.2.2) to assess its performance compared to the classical and bootstrap methods. As in the previous subsection, our strategy will be to use noninformative priors. The main properties used to assess the performance of a Bayesian confidence interval for $\beta_0(\theta_0)$ are the average interval length and variance of the interval length as well as its coverage probability. We now examine the coverage of a Bayesian confidence interval. Adopting the notation

$$\mathcal{B}(\boldsymbol{x}) \equiv \big[\widehat{\beta}_{\text{L}}(\boldsymbol{x}), \widehat{\beta}_{\text{U}}(\boldsymbol{x})\big], \qquad (29)$$

we define the coverage probability of the Bayesian confidence interval (29) as

$$\mathcal{C}(\beta_0(\theta_0)) = \text{E}_{\pi_0}\Big\{\text{E}_{g_{\theta_0}}\big[I_{\{\beta_0(\theta_0)\in\mathcal{B}(\boldsymbol{x})\}}\big]\Big\} \qquad (30)$$

$$= \int \int I_{\{\beta_0(\theta_0)\in\mathcal{B}(\boldsymbol{x})\}} \, g_{\theta_0}(\boldsymbol{x})\pi_0(\theta_0) \, d\boldsymbol{x} \, d\theta_0,$$

where $I_{\{\beta_0(\theta_0)\in\mathcal{B}(\boldsymbol{x})\}}$ denotes the indicator function of the event $\{\beta_0(\theta_0) \in \mathcal{B}(\boldsymbol{x})\}$ that the Bayesian confidence interval (29) contains $\beta_0(\theta_0)$.

The average length of the confidence interval (29) is

$$\mathcal{L} = \text{E}_{\pi_0}\Big\{\text{E}_{g_{\theta_0}}\big[\widehat{\beta}_{\text{U}}(\boldsymbol{x}) - \widehat{\beta}_{\text{L}}(\boldsymbol{x})\big]\Big\} \qquad (31)$$

$$= \int \int \big[\widehat{\beta}_{\text{U}}(\boldsymbol{x}) - \widehat{\beta}_{\text{L}}(\boldsymbol{x})\big] \, g_{\theta_0}(\boldsymbol{x})\pi_0(\theta_0) \, d\boldsymbol{x} \, d\theta_0,$$

and the variance of the confidence interval length is given by

$$\mathcal{V} = \text{E}_{\pi_0}\Big(\text{E}_{g_{\theta_0}}\Big\{\big[\widehat{\beta}_{\text{U}}(\boldsymbol{x}) - \widehat{\beta}_{\text{L}}(\boldsymbol{x}) - \mathcal{L}\big]^2\Big\}\Big) \qquad (32)$$

$$= \int \int \big[\widehat{\beta}_{\text{U}}(\boldsymbol{x}) - \widehat{\beta}_{\text{L}}(\boldsymbol{x}) - \mathcal{L}\big]^2 \, g_{\theta_0}(\boldsymbol{x})\pi_0(\theta_0) \, d\boldsymbol{x} \, d\theta_0.$$

5.2 Experimental Design

We developed in the previous subsections point and interval estimation criteria to assess the performance of the frequentist and Bayesian approaches in a discrete-event simulation framework. However, criteria (27), (30), (31), and (32) will be generally hard to compute analytically in most of the applications and should be estimated from the output realizations. The performance evaluation of the Bayesian and frequentist approaches is generally conducted by Monte Carlo computer-assisted simulation experiments. The output of these experiments can be used to estimate our performance criteria. Figure 3 summarizes the protocol we used to

conduct N Monte Carlo experiments and deliver estimates for the average risk of each estimator and its standard error, the mean and the variance of the interval lengths, and their coverage probabilities.

5.3 Application to a Single Server Queue

To provide a practical illustration of the above performance evaluation framework, we consider a simple single server queueing system. For this example, we will just compare the performance of the Bayesian and bootstrap approaches since the computational effort is approximately the same. Moreover, the average risk for the classical and bootstrap method is the same since they both use the same estimator for the mean response. In Zouaoui and Wilson (2001a),

Choose the true prior distribution π_0

For $i = 1, \ldots, N$

Sample $\theta_{0,i}$ from π_0

Compute the true output measure $\beta_{0,i} = \beta_0(\theta_{0,i})$

Sample x_i from $g_{\theta_{0,i}}(x)$

Compute the Bayes estimate $\widehat{\beta}_i^{\text{Bayes}} = \overline{\overline{y}}$
   using an operational prior π

Compute the frequentist estimate $\widehat{\beta}_i^{\text{Freq}}$
   as specified in (28) using $\widehat{\theta}^{\text{mle}}$

Compute a CI for $\beta_{0,i}$ using each approach

end loop

For each simulation-based approach, compute:

   the estimated average risk as

$$\widehat{\mathcal{R}} = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{\beta}_i - \beta_{0,i}\right)^2$$

   with Monte Carlo standard error estimate

$$\widehat{\text{SE}}(\widehat{\mathcal{R}}) = \left\{ \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[\left(\widehat{\beta}_i - \beta_{0,i}\right)^2 - \widehat{\mathcal{R}}\right]^2 \right\}^{\frac{1}{2}}$$

   the estimated coverage probability as

$$\widehat{\mathcal{C}} = \frac{1}{N} \sum_{i=1}^{N} I_{\{\beta_{0,i} \in \mathcal{B}_i(x_i)\}}$$

   the estimated mean of the CI length as

$$\widehat{\mathcal{L}} = \frac{1}{N} \sum_{i=1}^{N} \left[\widehat{\beta}_U(x_i) - \widehat{\beta}_L(x_i)\right]$$

   the estimated variance of the CI length as

$$\widehat{\mathcal{V}} = \frac{1}{N-1} \sum_{i=1}^{N} \left[\widehat{\beta}_U(x_i) - \widehat{\beta}_L(x_i) - \widehat{\mathcal{L}}\right]^2$$

Figure 3: Monte Carlo Experimental Design

we consider a larger scale example with a degenerate true prior distribution, where the main objective is to compare the interval estimation performance of the Bayesian method with the classical and bootstrap results of Cheng and Holland (1997).

The single server system has customers enter the queue according to a Poisson process with rate λ, having a "true" Gamma prior distribution π_0 with shape parameter $\nu_1 = 50$ and scale parameter ϕ_1 that will take a range of values corresponding to increasing levels of prior traffic intensity. The customers receive an exponential service time with rate μ, having a "true" Gamma prior distribution with shape parameter $\nu_2 = 50$ and scale parameter $\phi_2 = 100$. The objective is to estimate the average time in the system. Monte Carlo experiments are based on data samples of size $n = 1000$, and are repeated $N = 500$ times.

For the Bayesian approach, we chose the operational prior distribution for the arrival rate λ to be noninformative having a density proportional to λ^{-1}. After observing a sample $x = \{x_1, x_2, \ldots, x_n\}$ of interarrival times, we compute the posterior density of λ using Bayes' rule to obtain a Gamma density with shape parameter n and scale parameter $1/(\sum_{j=1}^{n} x_j)$. Similarly, choosing the same functional form for the prior distribution of the service rate μ produces a proper posterior Gamma density having a shape parameter n and a scale parameter $1/(\sum_{j=1}^{n} z_j)$, where $z = \{z_1, z_2, \ldots, z_n\}$ is the observed sample of service times.

Table 1 shows the results for the average risk in estimating the posterior mean response. It appears that the Bayesian point estimator averages a slightly smaller risk in all cases. As shown in Table 2, the main conclusion concerns the excellent performance of the Bayesian method in terms of interval estimation. The coverage probability is very close to the nominal coverage with shorter confidence intervals than their bootstrap counterparts. Although the system considered here is simple and these results may not be generalized for all systems, there is strong empirical evidence in favor of the Bayesian approach. The other examples detailed in Zouaoui and Wilson (2001a, b) further support this conclusion.

6 CONCLUSIONS

In this paper we have proposed a Bayesian approach that uses prior information and data observations for inferences on the distribution of a simulation-generated output response. We have derived an expression for the posterior mean, and we have developed a Bayesian Simulation Replication Algorithm to compute point and confidence interval estimators for the posterior mean response. We have also proposed two response surface models to decompose the posterior variance into two components measuring the parameter and stochastic uncertainty, respectively.

Given the difference in frequentist and Bayesian views, we have developed some criteria for comparing the performance of both approaches in a simulation framework. The basic criteria are: the Bayes risk for point estimation, and the expected length of the confidence interval and its coverage probability for interval estimation. We have conducted Monte Carlo experiments on a queueing application that show clear empirical evidence in favor of the Bayesian approach mainly in terms of coverage probability.

In Zouaoui and Wilson (2001b) we will extend the Bayesian framework presented in this paper to account for model uncertainty as well as parameter and stochastic uncertainty. It is not clear how the δ and bootstrap methods could be extended to account for model uncertainty.

Table 1: Average Risk of Classical Bootstrap and Proposed Bayesian Estimators for Average Sojourn Time in a Single Server Queue

ϕ_1	Method	$\widehat{\mathcal{R}}$	$\widehat{\text{SE}}(\widehat{\mathcal{R}})$
10	Bootstrap	9.3×10^{-11}	6.0×10^{-12}
	Bayesian	9.1×10^{-11}	5.9×10^{-12}
30	Bootstrap	1.4×10^{-10}	1.0×10^{-11}
	Bayesian	1.3×10^{-10}	9.8×10^{-12}
50	Bootstrap	2.7×10^{-9}	2.0×10^{-10}
	Bayesian	2.4×10^{-9}	1.8×10^{-10}
70	Bootstrap	8.0×10^{-9}	1.1×10^{-9}
	Bayesian	5.9×10^{-9}	6.8×10^{-10}
90	Bootstrap	2.5×10^{-4}	1.2×10^{-4}
	Bayesian	1.1×10^{-5}	4.5×10^{-6}

Table 2: Performance of Nominal 90% Confidence Intervals for Average Sojourn Time in a Single Server Queue

ϕ_1	Method	$\widehat{\mathcal{L}}$	$\widehat{\mathcal{V}}^{1/2}/\widehat{\mathcal{L}}$	$\widehat{\mathcal{C}}$
10	Bootstrap	4.9×10^{-5}	9.5×10^{-2}	98.2
	Bayesian	3.0×10^{-5}	9.6×10^{-2}	88.8
30	Bootstrap	6.3×10^{-5}	1.0×10^{-1}	99.0
	Bayesian	3.9×10^{-5}	1.1×10^{-1}	89.6
50	Bootstrap	2.6×10^{-4}	1.7×10^{-1}	99.4
	Bayesian	1.5×10^{-4}	1.5×10^{-1}	90.0
70	Bootstrap	4.0×10^{-4}	1.8×10^{-1}	98.8
	Bayesian	2.3×10^{-4}	1.8×10^{-1}	90.4
90	Bootstrap	7.5×10^{-3}	9.3×10^{-1}	99.4
	Bayesian	4.3×10^{-3}	6.0×10^{-1}	91.6

REFERENCES

Berger, J. O., and M. Delampady. 1987. Testing precise hypotheses. *Statistical Science* 2 (3): 317–352.

Carlin, B. P., and T. A. Louis. 1996. *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall.

Casella, G., and E. I. George. 1992. Explaining the Gibbs sampler. *American Statistician* 46:167–174.

Cheng, R. C. H., and W. Holland. 1997. Sensitivity of computer simulation experiments to errors in input data. *Journal of Statistical Computation and Simulation* 57:219–241.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7:1–26.

Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap*. New York: Chapman & Hall.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. London: Chapman & Hall.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. *Markov chain Monte Carlo in practice*. London: Chapman & Hall.

Helton, J. C. 1998. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation* 57:3–76.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90 (430): 773–795.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis*. 3rd ed. New York: McGraw-Hill.

Raftery, A. E., D. Madigan, and C. T. Volinsky. 1995. Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). In *Bayesian statistics 5: Proceedings of the fifth Valencia international meeting, June 5–9, 1994*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 323–349. Oxford: Oxford University Press.

Rao, P. 1997. *Variance components estimation: Mixed models, methodologies and applications*. London: Chapman & Hall.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks. 1996. *BUGS 0.5: Bayesian inference using Gibbs sampling manual (version ii)*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, UK.

Stuart, A., and J. K. Ord. 1994. *Kendall's advanced theory of statistics, volume 1: Distribution theory*. 6th ed. London: Edward Arnold.

Zouaoui, F. 2001. Accounting for input uncertainty in discrete-event simulation. Doctoral dissertation, Graduate Program in Operations Research, North Carolina State University, Raleigh, NC. Available on-line via `http://www.lib.ncsu.edu/etd/` `public/etd-3949188410121271/etd.pdf` [accessed June 16, 2001].

Zouaoui, F., and J. R. Wilson. 2001a. Accounting for parameter uncertainty in simulation input modeling. Technical Report, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available on-line via `ftp://ftp.ncsu.edu/pub/` `eos/pub/jwilson/msc1.pdf` [accessed June 16, 2001].

Zouaoui, F., and J. R. Wilson. 2001b. Accounting for input model and parameter uncertainty in simulation. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, M. W. Rohrer, and D. J. Medeiros, to appear. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available on-line with a more detailed development via `ftp://ftp.ncsu.edu/pub/eos/pub/` `jwilson/msc2.pdf` [accessed June 16, 2001].

AUTHOR BIOGRAPHIES

FAKER ZOUAOUI is an operations research consultant in the Research Group at Sabre, Inc. He received his bachelor's and master's degrees in industrial engineering from Bilkent University, and he received his Ph.D. degree in operations research from North Carolina State University. He is a member of INFORMS. His e-mail address is `<faker.zouaoui@sabre.com>`.

JAMES R. WILSON is Professor and Head of the Department of Industrial Engineering at North Carolina State University. He currently serves as the corepresentative of INFORMS–College on Simulation to the Board of Directors of the Winter Simulation Conference. He is a member of ASA, ACM, IIE, and INFORMS. His e-mail address is `<jwilson@eos.ncsu.edu>`, and his web address is `<www.ie.ncsu.edu/jwilson>`.