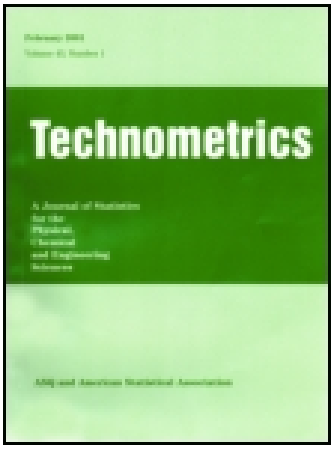


Technometrics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utch20

# An Efficient Sampling Technique for Bayesian Inference With Computationally Demanding Models

Peter Reichert[a], Mark Schervish[b] & Mitchell J Small[c]

[a] Swiss Federal Institute for Environmental Science and Technology (EAWAG), 8600 Dübendorf, Switzerland

[b] Department of Statistics Carnegie Mellon University, Pittsburgh, PA 15213

[c] Departments of Engineering and Public Policy and Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

Published online: 01 Jan 2012.

To cite this article: Peter Reichert, Mark Schervish & Mitchell J Small (2002) An Efficient Sampling Technique for Bayesian Inference With Computationally Demanding Models, Technometrics, 44:4, 318-327, DOI: 10.1198/004017002188618518

To link to this article: http://dx.doi.org/10.1198/004017002188618518

# An Efficient Sampling Technique for Bayesian Inference With Computationally Demanding Models

**Peter REICHERT**

Swiss Federal Institute for Environmental Science and Technology (EAWAG)
8600 Dübendorf, Switzerland
(*reichert@eawag.ch*)

**Mark SCHERVISH**

Department of Statistics Carnegie Mellon University
Pittsburgh, PA 15213
(*mark@stat.cmu.edu*)

**Mitchell J. SMALL**

Departments of Engineering and Public Policy and
Civil and Environmental Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
(*ms35@andrew.cmu.edu*)



In the environmental sciences, a large knowledge base is typically available on an investigated system or at least on similar systems. This makes the application of Bayesian inference techniques in environmental modeling very promising. However, environmental systems are often described by complex, computationally demanding simulation models. This strongly limits the application of Bayesian inference techniques, because numerical implementation of these techniques requires a very large number of simulation runs. The development of efficient sampling techniques that attempt to approximate the posterior distribution with a relatively small parameter sample can extend the range of applicability of Bayesian inference techniques to such models. In this article a sampling technique is presented that tries to achieve this goal. The proposed technique combines numerical techniques typically applied in Bayesian inference, including posterior maximization, local normal approximation, and importance sampling, with copula techniques for the construction of a multivariate distribution with given marginals and correlation structure and with low-discrepancy sampling. This combination improves the approximation of the posterior distribution by the sampling distribution and improves the accuracy of results for small sample sizes. The usefulness of the proposed technique is demonstrated for a simple model that contains the major elements of models used in the environmental sciences. The results indicate that the proposed technique outperforms conventional techniques (random sampling from simpler distributions or Markov chain Monte Carlo techniques) in cases in which the analysis can be limited to a relatively small number of parameters.

KEY WORDS: Bayesian inference; Computationally demanding model; Copula; Low-discrepancy sequence; Markov chain Monte Carlo; Numerical technique.

## 1. INTRODUCTION

Due to the accumulation of much knowledge over the past centuries, much information on environmental systems is available. For this reason, Bayesian techniques provide a natural framework for statistical inference for decision-oriented problems. The explicit use of prior knowledge and its updating with specific data leads to an optimal use of all available sources of information. However, because prior knowledge is often vague and cannot be unambiguously formulated in a prior probability distribution, a subjective element is introduced into the analysis.

The usefulness of Bayesian methods in the environmental sciences has been pointed out by various authors (Freeze, Massmann, Smith, Sperling, and James 1990; Brand and Small 1995; Ellison 1996; Reichert and Omlin 1997; Steinberg, Reckhow, and Wolpert 1996; Omlin and Reichert 1999). Other authors have questioned the Bayesian approach, mainly because of the problem of uniquely specifying prior distributions (e.g., Dennis 1996). Early applications of Bayesian techniques and similar approaches in the environmental sciences focused on sensitivity and uncertainty analyses of models for aquatic systems (Hornberger and Spear 1981; Dilks, Canale, and Meier 1992). The difficulty in accessing groundwater systems, which makes measurements very expensive, led to the recognition of the usefulness of formal incorporation of prior knowledge by Bayesian techniques in hydrogeology (Freeze et al. 1990; Abbaspour, Schulin, Schläppi, and Flühler 1996; Sohn, Small, and Pantazidou 2000). Other environmental applications of Bayesian methods include modelling of air pollution (Smith and French 1993; Ranyard and Smith 1997), environmental exposure and risk assessment (Finkel and Evans 1987; Taylor 1993), and global climate change impacts (Patwardhan and Small 1992). However, even today, the application of Bayesian techniques is largely limited to the research literature and is only very rarely done in "field" applications.


TECHNOMETRICS, NOVEMBER 2002, VOL. 44, NO. 4
DOI 10.1198/004017002188618518

The application of Bayesian techniques usually requires a large number of simulation runs to numerically approximate the posterior distribution of model parameters. This can seriously limit the applicability of these techniques in environmental modeling, because many environmental models require significant computation time to perform a single simulation. Typical causes for this long computation time include the consideration of many chemical constituents, a high degree of spatial resolution, and long simulation periods. In many practical applications performed in the past, this problem was further increased by the use of inefficient sampling schemes such as random sampling from the prior [Bayesian Monte Carlo (BMC)] (Dilks et al. 1992; Brand and Small 1995; Dakins, Toll, Small, and Brand 1996).

It is the goal of this article to describe a procedure that can lead to an approximation of a low-dimensional posterior parameter distribution and of the corresponding distribution of model results, with significantly fewer simulations compared to often-used sampling techniques. (This is what is meant by "efficient" in the title.) The article is structured as follows. In section 2 a brief survey of techniques for the numerical approximation of the posterior distribution is given, with special emphasis on importance sampling. In Section 3 the proposed extensions to importance sampling are discussed. These involve the use of a larger class of sampling distributions (arbitrary marginals coupled by a copula with given correlation coefficients) and the use of low-discrepancy samples instead of random samples. Both of these techniques have been used for many years, but not in the current context. In Section 4 the proposed technique is applied to a simple example of practical relevance. Finally, in Section 5 the results are summarized and their implications for other applications discussed.

## 2. BAYESIAN INFERENCE AND NUMERICAL APPROXIMATIONS

In the Bayesian approach to statistical inference, probability distributions describe the state of knowledge of the investigator or the research team on model structure, parameters, and/or results, rather than a limit of observed frequency distributions of a stochastic system for a large number of repetitions of an experiment. To conduct a Bayesian analysis for a given model structure with continuous parameters, prior knowledge of model parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^{\mathrm{T}}$, must be summarized in the form of a prior probability density, $f_{\mathrm{pri}}(\boldsymbol{\theta})$. In addition, the model, $M$, must be specified by the probability density of model results (for a given experimental layout), $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, conditional on the values of the model parameters, $f_M(\mathbf{y} \mid \boldsymbol{\theta})$. If the observations, $\mathbf{y}_{\mathrm{meas}} = (y_{\mathrm{meas},1}, \ldots, y_{\mathrm{meas},n})^{\mathrm{T}}$, are substituted in the probability density, $f_M$, and this function is viewed primarily as a function of the model parameters, then it is called the likelihood function of the model, $L_M(\boldsymbol{\theta}, \mathbf{y}_{\mathrm{meas}}) = f_M(\mathbf{y}_{\mathrm{meas}} \mid \boldsymbol{\theta})$. Bayesian inference then involves updating prior knowledge about parameter values to posterior knowledge according to

$$f_{\mathrm{post}}(\boldsymbol{\theta} \mid \mathbf{y}_{\mathrm{meas}}) = \frac{L_M(\boldsymbol{\theta}, \mathbf{y}_{\mathrm{meas}}) f_{\mathrm{pri}}(\boldsymbol{\theta})}{\int L_M(\boldsymbol{\theta}', \mathbf{y}_{\mathrm{meas}}) f_{\mathrm{pri}}(\boldsymbol{\theta}') \mathrm{d}\boldsymbol{\theta}'} \tag{1}$$

(e.g. Gelman, Carlin, Stern, and Rubin 1995). In this equation, $f_{\mathrm{post}}$ is the posterior probability density of the model parameters, which considers prior knowledge and data, $\mathbf{y}_{\mathrm{meas}}$. This posterior probability density of the parameters can then be used to calculate the posterior distribution of model results. Although proper formulation of the prior distribution and the likelihood function are difficult tasks (Wolpert 1989; Chaloner 1996; Kadane and Wolfson 1998), and the sensitivity of model results to these assumptions should be checked (Berger 1984; Wolfson, Kadane, and Small 1996; Small and Fischbeck 1999), in this article only the computational evaluation of (1) is addressed.

Except for very simple examples, the posterior distribution given by (1) cannot be computed analytically. The two most important numerical techniques applied instead are importance sampling and Markov chain Monte Carlo simulation (Gelman et al. 1995; Gamerman 1997). We focus here on a brief review of importance sampling, because we propose an extension of this technique.

Importance sampling is based on a sample from a distribution that is different from the posterior and is corrected by weights to approximate the posterior (e.g. Gelman et al. 1995). Typically this technique starts with a numerical determination of the maximum of the posterior distribution. This maximum can be calculated using the numerator of (1) only,

$$\boldsymbol{\theta}_0(\mathbf{y}_{\mathrm{meas}}) = \arg\max_{\boldsymbol{\theta}} \left( \log\left( L_M(\boldsymbol{\theta}, \mathbf{y}_{\mathrm{meas}}) f_{\mathrm{pri}}(\boldsymbol{\theta}) \right) \right). \tag{2}$$

The logarithm is introduced for numerical convenience. Next, the curvature of the log posterior density at its maximum is estimated, and the covariance matrix of a multivariate distribution approximating the posterior in the vicinity of its maximum is calculated according to

$$\boldsymbol{\Sigma} \approx -\left( \frac{\partial^2 \log(L_M(\boldsymbol{\theta}, \mathbf{y}_{\mathrm{meas}}) f_{\mathrm{pri}}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^{\mathrm{T}} \partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^{-1}. \tag{3}$$

This information is used to construct a sampling distribution, $f_{\mathrm{samp}}$, that approximates the posterior. Often a multivariate normal distribution with mean $\boldsymbol{\theta}_0$ and covariance matrix $\boldsymbol{\Sigma}$ is used as a sampling distribution. However, other distributions, including a multivariate $t$ distribution and a mixture of different analytical distributions, may be appropriate (Gelman et al. 1995). (Mixtures are of special importance in the case of the existence of several local maxima.) Then a sample is drawn from the sampling distribution. For each point of the sample, $\{\boldsymbol{\theta}_k\}_{k=1}^N$, the posterior weight

$$w_k = \frac{\frac{f_{\mathrm{pri}}(\boldsymbol{\theta}_k) \cdot L_M(\boldsymbol{\theta}_k, \mathbf{y}_{\mathrm{meas}})}{f_{\mathrm{samp}}(\boldsymbol{\theta}_k)}}{\sum_{l=1}^N \frac{f_{\mathrm{pri}}(\boldsymbol{\theta}_l) \cdot L_M(\boldsymbol{\theta}_l, \mathbf{y}_{\mathrm{meas}})}{f_{\mathrm{samp}}(\boldsymbol{\theta}_l)}} \tag{4}$$

is calculated. These weights correct the sample to approximate the posterior distribution. Expectations of arbitrary functions, $g$, with respect to the posterior distribution (1) can then be approximated by

$$\mathrm{E}_{f_{\mathrm{post}}}(g) \approx \sum_{k=1}^N w_k \, g(\boldsymbol{\theta}_k). \tag{5}$$

Note that the prior sampling technique (BMC) mentioned in Section 1 is a special case of importance sampling with $f_{\mathrm{samp}} = f_{\mathrm{pri}}$. This technique is obviously very inefficient if the posterior is significantly different from the prior.

## 3. PROPOSED EXTENSIONS TO IMPORTANCE SAMPLING

The major disadvantages of importance sampling are

- the limited flexibility of the parameterized sampling distribution to approximate the posterior distribution
- the slow convergence of random samples to the sampling distribution.

It is not possible to completely eliminate these problems. However, extensions of this technique that decrease the severity of these problems, at least for low-dimensional parameter spaces, are discussed in the next two sections. Both of the techniques used in these extensions have been available for many years and are frequently used in other fields. However, because they are not usually applied in the current context, their usefulness in improving the convergence of numerical approximations to posterior distributions in Bayesian inference has not yet been appreciated.

### 3.1 Extension of the Class of Sampling Distributions

Instead of using any of the analytical distributions commonly used in Bayesian inference, we propose constructing a sampling distribution from given marginals and given Kendall correlation coefficients using the copula approach. This significantly increases the analyst's flexibility in approximating the posterior distribution. In contrast to the distribution-free approach of constructing parameter samples with given rank correlation coefficients (Iman and Conover 1982), our approach uses a multivariate distribution, the density of which can be easily calculated. This is necessary to calculate the weights for importance sampling (4).

The copula approach starts from the observation that any multivariate cumulative distribution function can be written in the form

$$F(\boldsymbol{\theta}) = C\big(F_1(\theta_1), \ldots, F_m(\theta_m)\big) \tag{6}$$

(Schweizer 1991; Nelson 1995; Clemen and Reilly 1999), where the $F_i$'s are the marginal cumulative distribution functions and $C$ is the copula. The density can be written as

$$f(\boldsymbol{\theta}) = f_1(\theta_1) \cdot \cdots \cdot f_m(\theta_m) c\big(F_1(\theta_1), \ldots, F_m(\theta_m)\big), \tag{7}$$

where $f_i$ is the marginal density corresponding to $F_i$ and

$$c = \frac{\partial^m C}{\partial F_1 \cdots \partial F_m} \tag{8}$$

is the copula density or dependence function. The representations (6) and (7) of a probability distribution have the advantage of separating the marginal densities and the correlation structure of a given multivariate distribution. On the other hand, these equations can be used to construct a multivariate density given the marginal distributions and the correlation structure. This is what we do, technically in the same way as has been done for combining expert opinion (Clemen and Reilly 1999), but with another application in mind. The decomposition of the multivariate sampling distribution into one-dimensional marginal distributions and a copula may be preferred to the decomposition into a marginal distribution for one parameter and a series of conditional distributions for the other parameters if the model structure does not naturally lead to such a decomposition. For this alternative case, more specific numerical techniques for estimating marginal posterior densities are available (Johnson 1995).

For technical convenience, and because a normal distribution is a natural first approximation to the posterior close to its maximum, we use the copula of the multivariate normal distribution to specify the correlation structure for given pairwise Kendall correlation coefficients and marginal distributions (Clemen and Reilly 1999). The definition of the Kendall correlation coefficient starts from a random sample of two points, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, from the parameter distribution. The Kendall correlation coefficient, of the components $i$ and $j$ of the parameter vector $R^*_{i,j}$, is then equal to the probability that differences in the components $i$ and $j$ of the two points have the same sign (concordant components) minus the probability that these differences have different signs (disordant components),

$$\begin{aligned} R^*_{i,j} &= \mathrm{P}\big(([\boldsymbol{\theta}_1]_i - [\boldsymbol{\theta}_2]_i)([\boldsymbol{\theta}_1]_j - [\boldsymbol{\theta}_2]_j) > 0\big) \\ &\quad - \mathrm{P}\big(([\boldsymbol{\theta}_1]_i - [\boldsymbol{\theta}_2]_i)([\boldsymbol{\theta}_1]_j - [\boldsymbol{\theta}_2]_j) < 0\big) \\ &= 2\mathrm{P}\big(([\boldsymbol{\theta}_1]_i - [\boldsymbol{\theta}_2]_i)([\boldsymbol{\theta}_1]_j - [\boldsymbol{\theta}_2]_j) > 0\big) - 1, \end{aligned} \tag{9}$$

where $[\cdot]_i$ refers to the $i$th component of the argument. Kendall correlation coefficients are used because it is easier to construct a multivariate distribution with given Kendall correlation coefficients compared to given moment-based correlation coefficients using the copula approach. This is because Kendall correlation coefficients are not affected by monotone transformations of individual parameters necessary to adapt the marginal distributions. If we denote the multivariate normal distribution with mean values $\boldsymbol{\mu}$, standard deviations $\boldsymbol{\sigma}$, moment-based correlation coefficients $\mathbf{R}$ by $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{R})$, and the univariate normal distribution by $\mathrm{N}(\mu, \sigma)$ and $(x_1, \ldots, x_n)^{\mathrm{T}}$ by $\mathbf{x}$, then, using the definitions (6) and (7), we can express the copula of the multivariate normal distribution as

$$C_{\mathrm{N}(\mathbf{R})}(\mathbf{x}) = F_{\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{R})}\big(F^{-1}_{\mathrm{N}(\mu_1, \sigma_1)}(x_1), \ldots, F^{-1}_{\mathrm{N}(\mu_m, \sigma_m)}(x_m)\big) \tag{10}$$

and the density as

$$\begin{aligned} &c_{\mathrm{N}(\mathbf{R})}(\mathbf{x}) \\ &= \frac{f_{\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{R})}\big(F^{-1}_{\mathrm{N}(\mu_1, \sigma_1)}(x_1), \ldots, F^{-1}_{\mathrm{N}(\mu_m, \sigma_m)}(x_m)\big)}{f_{\mathrm{N}(\mu_1, \sigma_1)}\big(F^{-1}_{\mathrm{N}(\mu_1, \sigma_1)}(x_1)\big) \cdot \cdots \cdot f_{\mathrm{N}(\mu_m, \sigma_m)}\big(F^{-1}_{\mathrm{N}(\mu_m, \sigma_m)}(x_m)\big)}. \end{aligned} \tag{11}$$

It will become evident in eq. (15) that these copulas do not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. For the normal distribution, the moment-based correlation coefficients can be calculated from Kendall correlation coefficients according to

$$\mathbf{R}: \quad R_{i,j} = \sin\left(\frac{\pi}{2} R^*_{i,j}\right) \tag{12}$$

(Kruskal 1958; Clemen and Reilly 1999). If these correlation coefficients are used for the copula, then the resulting distribution will have the desired Kendall correlation coefficients, $\mathbf{R}^*$, whereas the moment-based correlation coefficients will no longer be $\mathbf{R}$ because of the transformation of the marginals.

After substituting the density of the multivariate normal distribution

$$f_{\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\sigma},\mathbf{R})}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \quad (13)$$

with the covariance matrix given by

$$\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\sigma})\mathbf{R}\operatorname{diag}(\boldsymbol{\sigma}) \quad (14)$$

into (11), algebraic manipulation leads to

$$c_{\mathrm{N}(\mathbf{R})}(\mathbf{x}) = \frac{1}{\mid \mathbf{R} \mid^{1/2}} \exp\left(-\frac{1}{2}\begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}(x_1) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}(x_m) \end{pmatrix}^{\mathrm{T}} \times (\mathbf{R}^{-1}-\mathbf{I}_m)\begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}(x_1) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}(x_m) \end{pmatrix}\right), \quad (15)$$

where $\mathbf{I}_m$ is the $m$-dimensional identity matrix (Clemen and Reilly 1999). This is a joint copula density of $m$ uniformly distributed random variables having the same Kendall correlation matrix as the normal distribution with correlation matrix (12).

Combining (7) and (15) leads to the density of the sampling distribution

$$f_{\mathrm{samp}}(\boldsymbol{\theta}) = f_1(\theta_1)\cdot\cdots\cdot f_m(\theta_m)\frac{1}{\mid \mathbf{R} \mid^{1/2}} \times \exp\left(-\frac{1}{2}\begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}\big(F_1(\theta_1)\big) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}\big(F_m(\theta_m)\big) \end{pmatrix}^{\mathrm{T}} \times (\mathbf{R}^{-1}-\mathbf{I}_m)\begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}\big(F_1(\theta_1)\big) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}\big(F_m(\theta_m)\big) \end{pmatrix}\right), \quad (16)$$

where $f_1, \ldots, f_m$ and $F_1, \ldots, F_m$ are the given densities and cumulative distribution functions of the marginals and $\mathbf{R}$ is the correlation matrix calculated with the aid of (12) from the given Kendall correlation matrix $\mathbf{R}^*$.

Substituting (10) in (6) and considering the fact [observed in (15)] that this distribution does not actually depend on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ leads to the following expression for the cumulative distribution function of the sampling distribution:

$$F_{\mathrm{samp}}(\boldsymbol{\theta}) = F_{\mathrm{N}(\mathbf{0},\mathbf{1},\mathbf{R})}\big(F^{-1}_{\mathrm{N}(0,1)}\big(F_1(\theta_1)\big), \ldots, F^{-1}_{N(0,1)}\big(F_m(\theta_m)\big)\big). \quad (17)$$

To apply (16), marginal distributions and Kendall correlation coefficients must be specified. In the first iteration step of the procedure, where a local normal approximation has been estimated by using (2) and (3), the Kendall correlation coefficients can be estimated by an inversion of (12) which is valid for normal distributions (Kruskal 1958):

$$\mathbf{R}^*: \quad R^*_{i,j} = \frac{2}{\pi}\arcsin(R_{i,j}). \quad (18)$$

Normal marginal distributions can be used, or marginal distributions can be selected according to prior knowledge. In any case, the maximum of the sampling distribution constructed in this way should be at least approximately equal to the maximum, $\boldsymbol{\theta}_0$, determined in the maximization procedure. In later iteration steps of approximating the posterior distribution, marginal distributions can be estimated from the weighted sample (4) of the previous iteration step. Similarly, Kendall correlation coefficients of the posterior distribution can be estimated from the weighted sample of the previous iteration step according to

$$\mathbf{R}^*: \; R^*_{i,j} \approx \frac{4\cdot \sum_{\substack{k>l \\ ([\boldsymbol{\theta}_1]_i-[\boldsymbol{\theta}_2]_i)([\boldsymbol{\theta}_1]_j-[\boldsymbol{\theta}_2]_j)>0}} w_{\mathrm{post},k}\, w_{\mathrm{post},l}}{1-\sum_{k=1}^{N} w^2_{\mathrm{post},k}} - 1. \quad (19)$$

Projections of the values of the weights (4) can help diagnose the need for a shift of the sampling distribution. See Section 4 for an illustration.

### 3.2 Quasi-Random Sequences

In different fields of numerical analysis, quasi-random sequences have been used instead of random numbers to increase the efficiency of Monte Carlo techniques (Hammersley and Handscomb 1964; Niederreiter 1978, 1992; Hua and Wang 1981). A quasi-random, or low-discrepancy, sequence is a sequence of points, $\{\mathbf{z}_k\}_{k=1}^N$, that is uniformly distributed in the unit cube, $[0,1]^m$. This is quantified by its discrepancy, $D$, the supremum of the difference between the empirical distribution function of the sample and the distribution function to be approximated (in this case, of the uniform distribution). For low-discrepancy sequences, the discrepancy is significantly smaller than the expected discrepancy of a random sample. The major field of application of quasi-random sequences is numerical integration (Niederreiter 1978) however, quasi-random sequences have also been applied in statistics (Fang, Wang, and Bentler 1994), for global optimization (Kalagnanam and Diwekar 1997), and for Bayesian inference (Shaw 1988).

Two simple low-discrepancy sequences due to Hammersley and Halton (Hammersley 1960; Halton 1960; Hammersley and Handscomb 1964) can be constructed using the following expansion of a nonnegative integer, $k$:

$$k = a_0 + a_1 p + a_2 p^2 + \cdots + a_r p^r, \quad (20)$$

where $p$ is a prime number and $a_i$ are nonnegative integer coefficients smaller than $p$. These coefficients can be used to construct a number between 0 and unity according to

$$\Phi_p(k) = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} + \cdots + \frac{a_r}{p^{r+1}}. \quad (21)$$

Any sequence of different prime numbers, $p_1, \ldots, p_{m-1}$, leads to a sequence of $N$ Hammersley points defined by

$$\mathbf{z}_k = \left( \frac{k-1/2}{N}, \Phi_{p_1}(k), \Phi_{p_2}(k), \ldots, \Phi_{p_{m-1}}(k) \right), \quad k = 1, \ldots, N. \quad (22)$$

Here the first component has been shifted by $1/(2N)$ compared to other definitions to avoid a value of 0 or unity that would be transformed to infinity by the transformation given by (27). Any sequence of different prime numbers, $p_1, \ldots, p_m$, leads to a sequence of $N$ Halton points defined by

$$\mathbf{z}_k = \left( \Phi_{p_1}(k), \Phi_{p_2}(k), \ldots, \Phi_{p_m}(k) \right), \quad k = 1, \ldots, N. \quad (23)$$

Compared with the Hammersley sequence, the Halton sequence has the advantage that a sample can easily be extended to a larger sample without modifying the old sample points. It has been shown that the discrepancy of these sequences with respect to the uniform distribution in $[0, 1]^m$ is much smaller than that of a random sample (Hammersley 1960; Halton 1960; Hammersley and Handscomb 1964). This is illustrated by the comparison of sampling techniques in the unit square in Figure 1 (Kalagnanam and Diwekar 1997). The sampling techniques used are grid sampling, random sampling, latin hypercube sampling (McKay, Beckman, and Conover 1979), and Hammersley sampling according to (22). It is evident that the Hammersley points are much more uniformly distributed than the points from random or latin hypercube sampling. Note that the two quasi-random sequences due to Hammersley and Handscomb described in this section are very useful examples for demonstrating the basic ideas, because their construction is very simple [see (20)–(23)]. Other quasi-random sequences have been proposed to further improve the efficiency of numerical integration procedures (Sobol 1988; Bratley and Fox 1988).

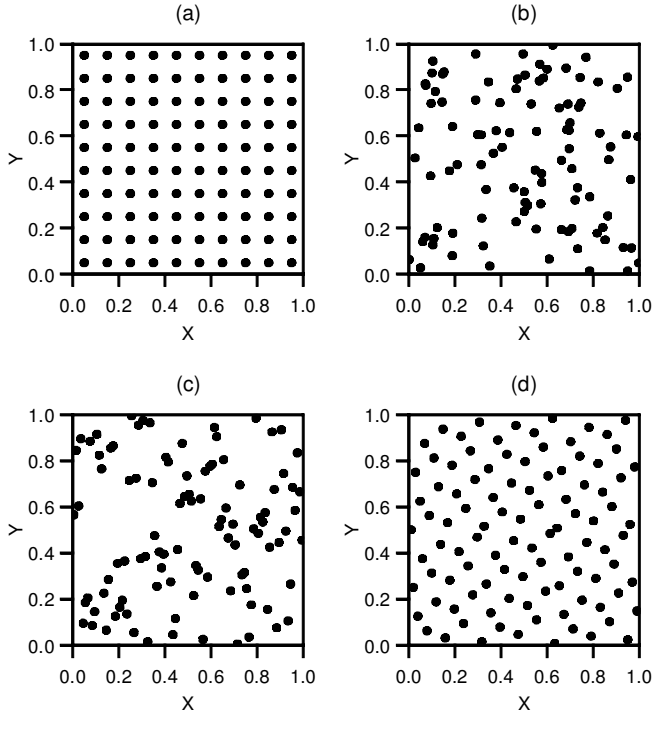


*Figure 1. Bivariate Uniform Distribution Without Correlation. Samples of 100 points are shown for the following sampling techniques: (a) grid (D = .090), (b) random (D = .087), (c) latin hypercube (D = .067), (d) Hammersley (D = .029).*

### 3.3 Combination of the Extensions

The form (17) of the cumulative distribution function of the sampling distribution shows that the values of

$$\begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}\left(F_1(\theta_1)\right) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}\left(F_m(\theta_m)\right) \end{pmatrix} \quad (24)$$

are distributed according to a multivariate normal distribution with means equal to 0, standard deviations equal to unity, and correlation coefficients equal to $\mathbf{R}$ if a sample of values of $\boldsymbol{\theta}$ is distributed according to the sampling distribution. If the matrix $\mathbf{A}$ is a Cholesky factor of the correlation matrix $\mathbf{R}$, that is,

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{R}, \quad (25)$$

then its inverse can be used to transform a normally distributed sample with correlation matrix $\mathbf{R}$ to an uncorrelated normally distributed sample (Gelman et al. 1995; Gamerman 1997). Application of the standard normal distribution to the marginals then leads to a uniform distribution of the points

$$\mathbf{z} = \begin{pmatrix} F_{\mathrm{N}(0,1)}\left( \left[ \mathbf{A}^{-1} \begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}\left(F_1(\theta_1)\right) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}\left(F_m(\theta_m)\right) \end{pmatrix} \right]_1 \right) \\ \vdots \\ F_{\mathrm{N}(0,1)}\left( \left[ \mathbf{A}^{-1} \begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}\left(F_1(\theta_1)\right) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}\left(F_m(\theta_m)\right) \end{pmatrix} \right]_m \right) \end{pmatrix} \quad (26)$$

in $[0, 1]^m$. Inversion of this formula shows that sample points $\boldsymbol{\theta}$ of the sampling distribution can be constructed from sample points $\mathbf{z}$ of a uniform distribution in $[0, 1]^m$ according to

$$\boldsymbol{\theta} = \begin{pmatrix} F_1^{-1}\left( F_{\mathrm{N}(0,1)}\left( \left[ \mathbf{A} \begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}(z_1) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}(z_m) \end{pmatrix} \right]_1 \right) \right) \\ \vdots \\ F_m^{-1}\left( F_{\mathrm{N}(0,1)}\left( \left[ \mathbf{A} \begin{pmatrix} F^{-1}_{\mathrm{N}(0,1)}(z_1) \\ \vdots \\ F^{-1}_{\mathrm{N}(0,1)}(z_m) \end{pmatrix} \right]_m \right) \right) \end{pmatrix}. \quad (27)$$

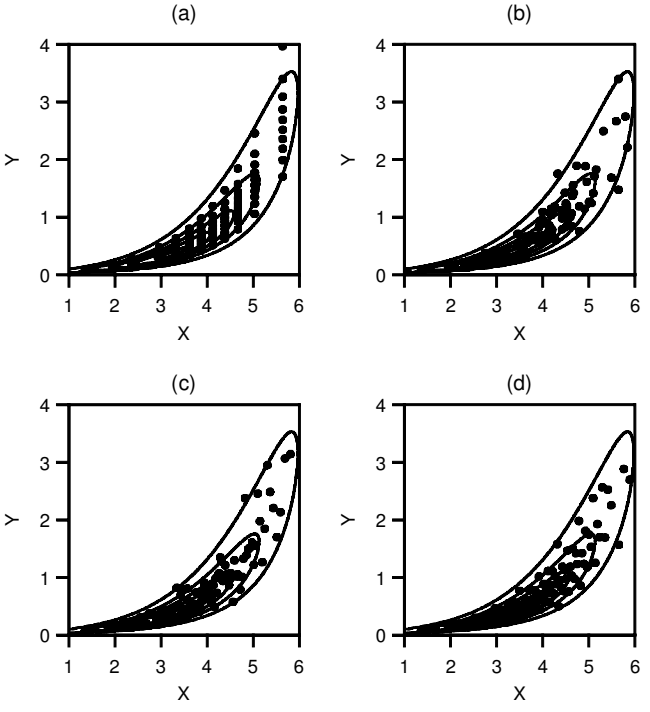


*Figure 2. Bivariate Distribution With N(4,1) and LN(1,1) Marginals and a Kendall Correlation Coefficient of .8. Highest probability density contour lines with probability contents of .05, .25, .5, .75, and .95 and samples of 100 points for the following sampling techniques: (a) grid (D = .052), (b) random (D = .078), (c) latin hypercube (D = .060), (d) Hammersley (D = .029).*

This equation makes it very simple to construct a sample of the sampling distribution based on an arbitrary sample from the uniform distribution in $[0, 1]^m$. The sample, $\{\boldsymbol{\theta}_k\}_{k=1}^N$, is based on the cumulative distribution functions of the marginals, $F_1, \ldots, F_m$, on the Kendall correlation coefficients, $\mathbf{R}^*$, used to calculate $\mathbf{R}$ according to (12) and then $\mathbf{A}$ according to (25), and on a sample, $\{\mathbf{z}_k\}_{k=1}^N$, of the uniform distribution in $[0, 1]^m$. Obviously, a low-discrepancy sequence, as described in Section 3.2, can be used for this purpose.

Figure 2 shows that the extended class of distributions introduced in Section 3.1 leads to an extension of typically used distribution shapes. The "banana-shaped" distribution is a difficult case for importance sampling (Gelman et al. 1995). The graphic evidence in Figures 1 and 2 that the low-discrepancy sample is a good approximation to the distribution is supported by the values of the discrepancies for these distributions (see the figure captions). The discrepancies for the Hammersley sampling technique are significantly smaller than for the other techniques for both the uniform and the transformed distributions.

## 4. ILLUSTRATIVE APPLICATION

### 4.1 Model and Data

The model chosen to illustrate and compare the methods is a simple biogeochemical model commonly used as a submodel for reaction kinetics in wastewater treatment plants (Henze, Grady, Gujer, Marais, and Matsuo 1986; Gujer et al. 1995; Gujer, Henze, Mino, and van Loosdrecht 1999) rivers

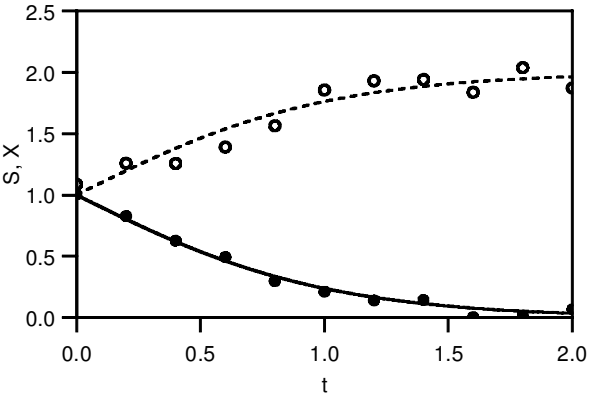


*Figure 3. Model Results and Synthetic Data (Dimensionless Units). The solid circles and solid line represent synthetic measurements and simulation for dissolved substrate (S); the open circles and dashed line, synthetic measurements and simulation for bacterial population (X).*

(Reichert et al. 2001), lakes (Omlin, Reichert, and Forster 2001), and other systems in which microbially mediated biogeochemical processes occur. The model describes degradation of a dissolved organic substrate in a completely stirred batch reactor caused by the growth of a bacterial population in the reactor. The process equations are

$$\frac{\mathrm{d}X}{\mathrm{d}t} = kSX \qquad \text{and} \qquad \frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{1}{Y} \cdot kSX. \tag{28}$$

The first equation describes the growth of the bacterial population, which is assumed to be proportional to the product of the concentration of substrate, $S$, and the concentration of bacteria, $X$, with a growth rate constant, $k$. The second equation describes degradation of the substrate, which is assumed to be proportional to growth with a yield coefficient, $Y$. To keep the example as simple as possible, it is assumed that all parameters are nondimensionalized and that the values of the model parameters $k$ and $Y$ as well as the initial conditions of $S$ and $X$, $S_{\text{ini}}$ and $X_{\text{ini}}$, are equal to unity. For this example, we produced simulation results and synthetic measurements that were assumed to be normally distributed around the simulation results with a standard deviation of .025 for the substrate and .1 for the bacteria. This difference in standard deviation accounts for the fact that it is usually much simpler to measure dissolved substrate than bacterial populations. Figure 3 shows the results of a model simulation of $X$ and $S$.

### 4.2 Implementation of Model and Solution Techniques

The model equations were implemented in the program AQUASIM (Reichert 1994, 1995), which is designed for identification and simulation of models for aquatic systems. The batch version of this program has been extended by interfaces that allow the input of parameter sets and by options for posterior maximization and Markov chain Monte Carlo techniques. The creation and evaluation of samples from sampling distributions as described in Section 3 were done using the UNCSIM package (Reichert 2002). This package provides simple programs for producing and evaluating sample files. It is designed to perform this type of analysis with any

simulation program that can read sample files and write corresponding result files. Finally, the plots used for illustrating the samples and for analyzing the marginal distributions were produced with the program R, designed for statistical computing (Ihaka and Gentleman 1996).

### 4.3 Problem to be Solved

Our simple model has four parameters: $Y$, $k$, $S_{\text{ini}}$, and $X_{\text{ini}}$. There is typically much less uncertainty in the stoichiometric parameter $Y$ than in the kinetic parameter $k$. Similarly, $S_{\text{ini}}$ is usually known to much higher accuracy than $X_{\text{ini}}$. Therefore, it seems reasonable to reduce the dimension of the problem for this didactical application, by assuming the model parameters $Y = 1$ and $S_{\text{ini}} = 1$ to be perfectly known. For the other two model parameters, we assume independent lognormal prior distributions with means and standard deviations equal to 1.2 and .3 for $k$, and 1.0 and 1.0 for $X_{\text{ini}}$. The likelihood function of the model is a product of normal distributions with standard deviations of .025 for results in $S$ and .1 for results in $X$ around the result of the deterministic model (28) at those points in time where data are assumed to be available (Fig. 3). We distinguish two cases. For the first case, we assume that all data shown in Figure 3 are available; for the second case, we assume that only data for substrate concentrations are available (the solid circles in Fig. 3). The latter case is not unrealistic, because due to the greater difficulty in measuring bacterial concentrations, the attempt is often made to estimate changes in bacterial concentrations indirectly through changes in observed conversion rates of dissolved substances.

Note that due to the short simulation time required by the simple model given by (28), it is possible to calculate a very accurate approximation to the true solution of this problem on a very fine grid in the parameter space. This allows us to compare the samples and marginal distributions with the "true" solution of this problem. In the subsequent section, this is done for the proposed efficient importance sampling technique and for a simple Markov chain Monte Carlo technique.

### 4.4 Results

To take advantage of the distributions and sampling techniques described in Section 3, we apply the following procedure:

1. Estimate the maximum of the posterior distribution and the standard deviations and correlation coefficients of a local normal approximation to the maximum.
2. Construct a low-discrepancy sample for a multivariate normal distribution at the maximum of the posterior with the estimated correlation coefficients but enlarged standard deviations. Calculate the posterior weights, posterior means, standard deviations, and Kendall correlation coefficients.
3. Construct a low-discrepancy sample based on more realistic, but still widened marginals (using prior information and results from the previous steps), and calculate posterior weights and posterior means, standard deviations, and Kendall correlation coefficients. Repeat this step until approximate convergence is achieved (i.e., posterior means, standard deviations, Kendall correlation coefficients, and the shape of the marginals no longer change significantly).
4. Construct a low-discrepancy sample based on realistic marginals (without extension of the range of values). Check marginals and weights and use them for inference if the approximation of the posterior seems adequate.

This procedure has led to good results in the present example, as well as in the example used during the development phase of the UNCSIM package (Reichert 2002). However, this is not a universal procedure, but rather a procedure that requires changes depending on the dimensionality of the problem and on the results of intermediate steps. In steps 2 and 3, the standard deviations of the sampling distribution are enlarged to have an extended range of sample points that can support detection of deviations of the sampling distribution from the posterior. In the present example, the standard deviations were enlarged by a factor of 1.5; this factor would have to be reduced in higher dimensions.

The top row of Figure 4 shows the samples according to step 2 of the aforementioned procedure for the two cases of a complete dataset (left column) and substrate data only (right column). This sample is a low-discrepancy (Hammersley) sample of a normal distribution of size 50 with mean and correlation coefficient as determined locally from the maximum of the posterior distribution, but with standard deviations 50% larger than those determined locally. The weights of the sample points, represented by the size of their circles, obviously decrease with increasing distance from the center of the distribution. This indicates that no significant shift of the sampling distribution is necessary to fit the posterior and that

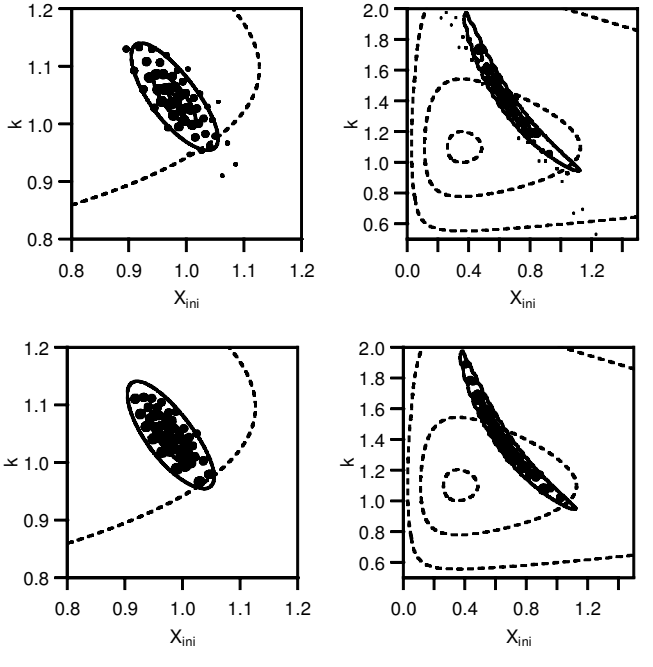


*Figure 4. The First and Third Samples Produced by the Efficient Importance Sampling Technique (Rows) for the Two Cases of a Complete Dataset (Left Column) or Data for Substrate Only (Right Column). The surface area of the circles characterizing each point of the sample is proportional to its posterior weight. The dashed and solid lines indicate 5%, 50%, and 95% highest probability density region boundaries for the prior and the posterior distribution. Note the differences in scales between the left and right columns.*

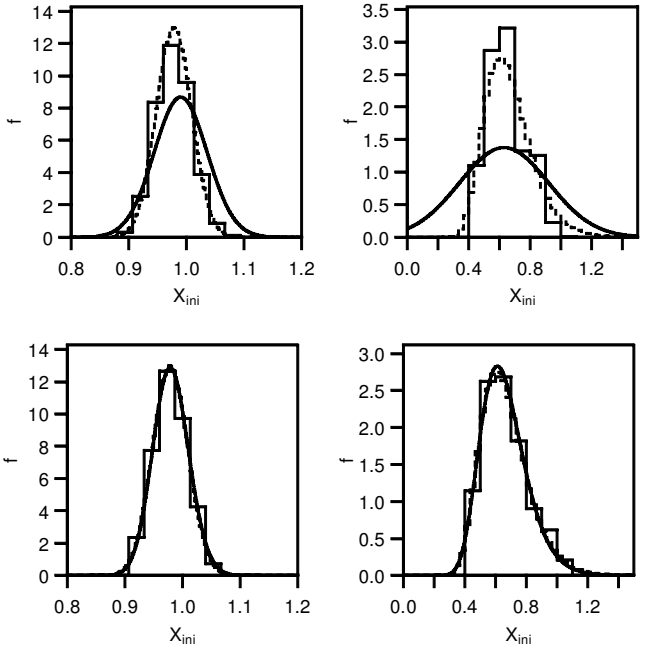


*Figure 5. Marginal Distributions for $X_{ini}$ for the First and Third Samples Produced by the Efficient Importance Sampling Technique (rows) for the Two Cases of a Complete Dataset (Left Column) or Data for Substrate Only (Right Column). The solid line represents the continuous curve: sampling distribution. The solid line represents a step function: histogram characterizing the distribution of the sample. The dashed line indicates the "exact" marginal of the posterior.*

the sampling distribution can be narrowed to some degree. For the case with the complete dataset (left), a minor shift to the top left seems adequate, because there are more sample points with small weight in the bottom right. The very small dots at both ends of the sample for the second case indicate that the posterior is narrower or bended away from the sample. (Note that the solid lines characterizing the posterior are usually not available.) These findings are supported by the analysis of the marginal distributions for $X_{\text{ini}}$ shown in the top row of Figure 5. It is evident that the histogram calculated from the weighted sample (solid line, step function), which represents the marginal of the posterior distribution, is narrower than the marginal of the sampling distribution (solid line, continuous function) for both cases. (Note that the exact marginal shown by the dashed line is usually not available.)

To account for the (small) asymmetry in the histogram approximating the marginal posterior distribution, the second samples (step 3 of the procedure) are low-discrepancy (Hammersley) samples of size 50 based on lognormal marginals with means and Kendall correlation coefficients estimated from the first step, but with standard deviations still enlarged by a factor of 1.5. The graphical analyses of the marginals (not shown) now lead to more symmetric deviations between the sampling distribution and the posterior. These findings allow us to proceed to step 4 of the procedure and to construct a sample based on the means, Kendall correlation coefficients, and unchanged standard deviations resulting from the second sample. Such a sample of size 50 is shown in the bottom row of Figure 4. The smaller difference in observed posterior weights indicates the closer approximation of the sampling distribution to the posterior. This is again confirmed by the

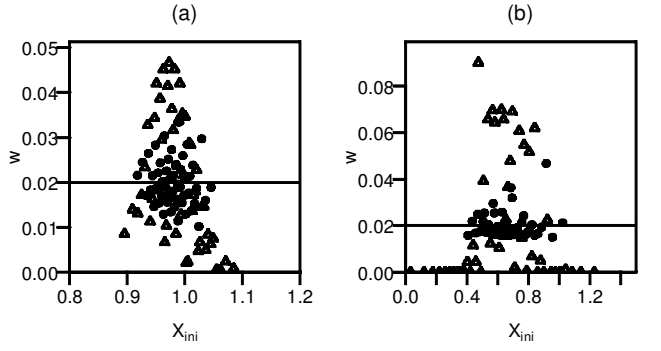


*Figure 6. Posterior Weights of the First (Triangles) and Third (Solid Circles) Samples Produced by the Efficient Importance Sampling Technique for the two Cases of the Complete Data set (a) or Data for Substrate only (b), Plotted as a Function of $X_{ini}$.*

analysis of the marginals shown in the bottom row of Figure 5. The histograms corresponding to the weighted sample do not significantly deviate from the marginals of the sampling distribution.

Figure 6 shows another useful diagnostic plot, in which the weights of the sample points are plotted as a function of $X_{\text{ini}}$. These "marginal weights" demonstrate that the range of weights decreases significantly from sample 1 to sample 3, indicating that the deviation between the sampling distribution and the posterior decreases. Because for the final samples there is no systematic trend in weights and the range of weights is not large, this sample can be accepted as a reasonable sample for Bayesian inference. If possible, the sample size could now be increased to improve the accuracy of the posterior distributions of the parameters and the corresponding model results.

To quantitatively support our graphical results regarding the efficiency of the proposed technique, we calculated the discrepancies of the final samples for six techniques: the efficient importance sampling technique using both proposed extensions, three simplified importance sampling techniques using only one or none of the extensions, a short Markov chain (with the same total number of simulations as in the three importance samples of size 50 each), and a long Markov chain. The results, given in Table 1, show the superiority of the efficient importance sampling technique over all other techniques based on 150 sample points (in addition to the

*Table 1. Discrepancies of Weighted Posterior Samples for Different Techniques*

| | Case | |
|---|---|---|
| *Technique* | *Data for S and X* | *Data for S only* |
| Efficient importance sampling | .06 | .06 |
| Copula distribution, random sampling | .15 | .11 |
| Normal distribution, Hammersley sequence | .08 | .08 |
| Normal distribution, random sampling | .17 | .15 |
| Markov chain 1–150 | .27 | .15 |
| Markov chain 1,000–2,000 | .06 | .07 |

function evaluations needed for posterior maximization common to all techniques) for the present example. A comparison of rows 2 and 3 indicates that in the current example, using the low-discrepancy sequence instead of random samples contributed more to the improvement than the flexible shape of the distribution allowed by the copula representation. This conclusion is not generalizable however, the opposite result can be expected for posteriors with shapes that deviate more significantly from the normal. The Markov chain technique is not competitive at such a small sample size, but does lead to good results if more simulations can be performed.

## 5. CONCLUSIONS

In importance sampling for Bayesian inference, the posterior distribution of model parameters is approximated by a sample from a multivariate normal distribution or from another analytically tractable distribution, and the sample is then corrected with weights to improve the approximation. The procedure proposed in this article extends this approach by using multivariate distributions constructed from arbitrary marginals and given pairwise correlation coefficients with the aid of a copula and by using low-discrepancy sequences for sampling. The first of these extensions makes the approach more flexible in improving the approximation of the shape of the posterior by the sampling distribution, and the second leads to a better approximation of the distribution by the sample. The resulting estimates of the marginals and correlation coefficients can then be used to improve the sampling distribution for the next iteration of the procedure, and the weights can be used to diagnose the quality of the approximation. This combination of techniques can be expected to lead to good estimates of properties of the posterior distribution with a minimum of simulation runs. The advantages of the proposed procedure are most relevant for problems concerning inference of a small number of parameters (less than about 10), because analysis of a large number of marginal distributions can become cumbersome, the chosen form of the copula may limit the degree of approximation of high-dimensional distributions, and the low-discrepancy sequences used in this article need many points to cover a high-dimensional distribution. Still, the proposed procedure can lead to significant efficiency gains for computationally demanding models with a relatively small number of influential parameters. If computing time is not a limiting factor, however, then Markov chain Monte Carlo techniques may be preferable because of their greater flexibility.



## ACKNOWLEDGMENTS

The authors thank Urmila Diwekar for stimulating discussions and Mark Borsuk, two anonymous reviewers, and the editor for suggestions for improving the manuscript. The first author also thanks the Swiss Federal Institute for Environmental Science and Technology (EAWAG) and the Department of Engineering and Public Policy (EPP) at Carnegie Mellon University for the opportunity to work temporarily at EPP.

*[Received May 2001. Revised March 2002.]*

## REFERENCES

Abbaspour, K. C., Schulin, R., Schläppi, E., and Flühler, H. (1996), "A Bayesian Approach for Incorporating Uncertainty and Data Worth in Environmental Projects," *Environmental Modeling and Assessment*, 1, 151–158.

Berger, J. O. (1984), "The Robust Bayesian Viewpoint," in *Robustness of Bayesian Analyses*, ed. J. Kadane., Amsterdam: North-Holland, pp. 63–124.

Brand, K. P., and Small, M. J. (1995), "Updating Uncertainty in an Integrated Risk Assessment: Conceptual Framework and Methods," *Risk Analysis*, 15, 719–731.

Bratley, P., and Fox, B. L. (1988), "ALGORITHM 659: Implementing Sobol's Quasi-Random Sequence Generator," *ACM Transactions on Mathematical Software*, 14, 88–100.

Chaloner, K. M. (1996), "The Elicitation of Prior Distributions," in *Bayesian Biostatistics*, eds. D. A. Barry and D. K. Stangl, New York, Marcel Dekker, pp. 141–156.

Clemen, R. T., and Reilly, T. (1999), "Correlations and Copulas for Decision and Risk Analysis," *Management Science*, 45, 208–224.

Dakins, M. E., Toll, J. E., Small, M. J., and Brand, K. P. (1996), "Risk-Based Environmental Remediation: Bayesian Monte Carlo Analysis and the Expected Value of Sample Information," *Risk Analysis*, 16, 67–79.

Dennis, B. (1996), "Discussion: Should Ecologists Become Bayesians?," *Ecological Applications*, 6, 1095–1103.

Dilks, D. W., Canale, R. P., and Meier, P. G. (1992), "Development of Bayesian Monte Carlo Techniques for Water Quality Model Uncertainty," *Ecological Modelling*, 62, 149–162.

Ellison, A. M. (1996), "An introduction to Bayesion Inference for Ecological Research and Environmental Decision-Making," *Ecological Applications*, 6, 1036–1046.

Fang, K.-T., Wang, Y., and Bentler, P. M. (1994), "Some Applications of Number-Theoretic Methods in Statistics," *Statistical Science*, 9, 416–428.

Finkel, A. M., and Evans, J. S. (1987), "Evaluating the Benefits of Uncertainty Reduction in Environmental Health Risk Management," *Journal Air Pollution Control Association*, 37, 1164–1172.

Freeze, R. A., Massmann, J., Smith, L., Sperling, T., and James, B. (1990), "Hydrogeological Decision Analysis: 1. A Framework," *Ground Water*, 28, 738–766.

Gamerman, D. (1997), *Markov Chain Monte Carlo—Statistical Simulation for Bayesian Inference*, London: Chapman and Hall.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.

Gujer, W., Henze, M., Mino, T., Matsuo, T., Wentzel, M. C., and Marais, G. v. R. (1995), "The Activated Sludge Model No. 2: Biological Phosphorus Removal," *Water Science and Technology*, 31, 1–11.

Gujer, W., Henze, M., Mino, T., and van Loosdrecht, M. (1999), "Activated Sludge Model No. 3," *Water Science and Technology*, 39, 183–193.

Halton, J. H. (1960), "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals," *Numerische Mathematik*, 2, 84–90.

Hammersley, J. M., (1960), "Monte Carlo Methods for Solving Multivariate Problems," *Annals of the New York Academy of Science*, 86, 844–874.

Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods*, London: Spottiswoode, Ballantyne & Co.

Henze, M., Grady, C. P. L., Gujer, W., Marais, G. v. R., and Matsuo, T. (1986), "Activated Sludge Model No. 1," Scientific and Technical Report 1, IAWPRC Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment Processes.

Hornberger, G. M., and Spear, R. C. (1981), "An Approach to the Preliminary Analysis of Environmental Systems," *Journal of Environmental Management*, 12, 7–18.

Hua, L. K., and Wang, Y. (1981), *Applications of Number Theory to Numerical Analysis*, Berlin: Springer.

Ihaka, R., and Gentleman, R. (1996), "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299–314.

Iman, R. L., and Conover, W. J. (1982), "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communication in Statistics—Simulation and Computation*, 11, 311–334.

Johnson, V. E. (1995), "A Technique for Estimating Marginal Posterior Densities in Hierarchical Models Using Mixtures of Conditional Densities," *Journal of the American Statistical Association*, 87, 852–860.

Kadane, J. B., and Wolfson, L. J. (1998), "Experiences in Elicitation" (with discussion), *Journal of the Royal Statistical Society*, Ser. D 47, 3–19.

Kalagnanam, J. R., and Diwekar, U. M. (1997), "An Efficient Sampling Technique for Off-Line quality control," *Technometrics*, 39, 308–319.

Kruskal, W. H. (1958), "Ordinal Measures of Association," *Journal of the American Statistical Association*, 53, 814–861.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, 21, 239–245.

Nelson, R. (1995), "Copulas, Characterization, Correlation, and Counterexamples," *Mathematics Magazine*, 68, 193–198.

Niederreiter, H. (1978), "Quasi-Monte Carlo Methods and Pseudo-Random Numbers," *Bulletin of the American Mathematical Society*, 84, 957–1041.

——— (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Philadelphia: Society for Industrial and Applied Mathematics.

Omlin, M., and Reichert, P. (1999), "A Comparison of Model Extrapolation Uncertainty Estimation Techniques," *Ecological Modelling*, 115, 45–59.

Omlin, M., Reichert, P., and Forster, R. (2001), "Biogeochemical Model of Lake Zürich: Model Equations and Results," *Ecological Modelling*, 141, 77–103.

Patwardhan, A., and Small, M. J. (1992), "Bayesian Methods for Model Uncertainty With Application to Future Sea Level Rise," *Risk Analysis*, 12, 513–523.

Ranyard, D., and Smith, J. Q. (1997), "Building a Bayesian Model in a Scientific Environment: Managing Uncertainty After an Accident," in *The Practice of Bayesian Analysis*, eds. S. French and J. Q. Smith, London: Arnold, pp. 245–258.

Reichert, P. (1994), "AQUASIM—A Tool for Simulation and Data Analysis of Aquatic Systems," *Water Science and Technology*, 30, 21–30.

——— (1995), "Design Techniques of a Computer Program for the Identification of Processes and the Simulation of Water Quality in Aquatic Systems," *Environmental Software*, 10, 199–210.

——— (2002), "UNCSIM—A Program Package for Uncertainty Analysis and Bayesian Inference," In preparation.

Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., and Vanrolleghem, P. (2001), "River Water Quality Model No. 1 (RWQM1): II. Biochemical Process Equations," *Water Science and Technology*, 43, 11–30.

Reichert, P., and Omlin, M. (1997), "On the Usefulness of Overparameterized Ecological models," *Ecological Modelling*, 95, 289–299.

Schweizer, B. (1991), "Thirty Years of Copulas," in *Advances of Probability Distributions With Given Marginals*, eds. G. Dall'Aglio, S. Kotz, and G. Salinetti, Dordrecht: Kluwer Academic Publishers.

Shaw, J. E. H. (1988), "A Quasi-random Approach to Integration in Bayesian Statistics," *The Annals of Statistics*, 16, 895–914.

Small, M. J., and Fischbeck, P. F. (1999), "False Precision in Bayesian Updating With Incomplete Models," *Human and Ecological Risk Assessment*, 5, 291–304.

Smith, J. Q., and French, S. (1993), "Bayesian Updating of Atmospheric Dispersion Models for Use After an Accidental Release of Radiation," *The Statistician*, 42, 501–511.

Sobol, I. M. (1988), "On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals," *USSR Computers and Mathematical Physics*, 7, 51–70.

Sohn, M. D., Small, M. J., and Pantazidou, M. (2000), "Reducing Uncertainty in Site Characterization Using Bayes Monte Carlo Methods," *Journal of Environmental Engineering*, 126, 893–902.

Steinberg, L. J., Reckhow, K. H., and Wolpert, R. L. (1996), "Bayesian Model for Fate and Transport of PCB in Upper Hudson River," *Journal of Environmental Engineering*, 122, 341–349.

Taylor, A. C. (1993), "Using Objective and Subjective Information to Generate Distributions for Probabilistic Exposure Assessment," *Journal of Exposure Analysis and Environmental Epidemiology*, 3, 285–298.

Wolfson, L. J., Kadane, J. B., and Small, M. J. (1996), "Bayesian Environmental Policy Decisions: Two Case Studies," *Ecological Applications*, 6, 1056–1066.

Wolpert, R. L. (1989), "Eliciting and Combining Subjective Judgments About Uncertainty," *International Journal of Technology Asessment in Health Care*, 5, 537–557.