# Groundwater pollution risk using a modified Latin hypercube sampling

Husam Baalousha

## ABSTRACT

Characterisation of groundwater modelling involves significant uncertainty because of estimation errors of these models and other different sources of uncertainty. Deterministic models do not account for uncertainties in model parameters, and thus lead to doubtful output. The main alternatives for deterministic models are the probabilistic models and perturbation methods such as Monte Carlo Simulation (MCS). Unfortunately, these methods have many drawbacks when applied in risk analysis of groundwater pollution. In this paper, a modified Latin Hypercube Sampling method is presented and used for risk, uncertainty, and sensitivity analysis of groundwater pollution. The obtained results were compared with other sampling methods.

Results of the proposed method have shown that it can predict the groundwater contamination risk for all values of probability better than other methods, maintaining the accuracy of mean estimation. Sensitivity analysis results reveal that the contaminant concentration is more sensitive to longitudinal dispersivity than to velocity.

**Key words** | contaminant transport modelling, Latin hypercube sampling, Monte Carlo simulation, probabilistic modelling, risk analysis

**Husam Baalousha**
Institute of Hydraulic Engineering and Water Resources Management,
Aachen University of Technology (RWTH),
Mies-van-der-Rohe-Strasse 1, 52056, Aachen,
Germany
Tel: +49 241 802 7343
Fax: +49 241 802 2348
E-mail: *Baalousha@web.de*

## INTRODUCTION

Mathematical models are usually used to simulate the groundwater flow and to predict the fate of contaminants in groundwater aquifers. Despite the power of the mathematical models, there is a large doubt about their accuracy. It is usually the case that these models are used in a deterministic way. However, the hydrogeological properties of the aquifer may vary significantly in time and space, and thus cannot be treated in a deterministic way. Many stochastic models have been developed and used in groundwater and contaminant transport modelling to account for the uncertainty in hydrogeological properties. These methods can be classified into two main categories: sampling techniques and moment approaches. Among the sampling methods, the most famous and widely used one is the Monte Carlo Simulation (MCS). The advantages of MCS are that it is a straightforward approach to generate a large number of equally likely random realisations, and it is independent of the dimension of the problem of concern. In addition, it can be used in different applications and can be applied on all groundwater models without any restriction. MCS, however, has many drawbacks such as the large number of runs needed to achieve reliable results. In groundwater models, for which large numbers of discritisation points usually exist, MCS becomes very crude. For instance, to apply MCS with a contaminant transport model, thousands of model runs will be required to achieve reliable results.

There are some other alternatives to MCS like moment methods. The First Order Second Moment method (FOSM) uses only the first two moments of the Taylor series expansion around the mean. This approach has been used with small groundwater and contaminant transport models to analyse risk and uncertainty (Lioua & Der Yeh 1997). The FOSM method is good for mean prediction, but it is poor for risk analysis since the tail probability is not considered. Moreover,

doi: 10.2166/hydro.2006.018

the covariance matrix in FOSM becomes very large when dealing with correlated variables. Thus, the computation efforts become as inefficient and as huge as for the case of MCS.

Some stratification methods were developed to decrease the number of realisations of MCS (quasi-Monte Carlo sampling). Latin Hypercube Sampling (LHS) is one of these approaches, which was developed from MCS.

It is based on a highly controlled selection of the input values and their random permutations (McKay *et al.* 1979). As with the Monte Carlo method, the mean and standard deviation for each parameter should be specified. The probability distribution for each input random parameter is divided into a number of segments with equal probability.

LHS was further developed to reduce the variance and to make it more deterministic. This development is called Mean-Value Lattice Sampling (MVLS), and it depends on the same idea as LHS, but it uses the mean value of each segment instead of a random value. MVLS has been widely used in different applications and mathematical models. Although the results obtained by MVLS are never worse than LHS, this method has some drawbacks. The main disadvantage of MVLS is that it does not reproduce the tail probability very well (Zhang & Pinder 2003).

Despite the huge research to improve the MCS and quasi-MC, the application of quasi-Monte Carlo methods in groundwater modelling and risk assessment is very rare (Dimov *et al.* 1998). To the knowledge of the author, all the work done has concentrated on classical MCS. In this paper, a new sampling approach is proposed and presented with illustrative examples for risk and sensitivity analysis in groundwater modelling. This proposed approach takes the advantages of LHS and MVLS for better stratification. It depends on stratification of LHS, and thus it considers the tail probability, which MVLS does not consider. The results obtained by the proposed methodology are compared with other methods to check the accuracy and the convergence rate of the proposed method. MCS was used as a reference for comparison with other different methods.

## MONTE CARLO SIMULATION

Monte Carlo Simulation (MCS) has been used to solve integration with multiple dimensions. Let $f(X)$ denote a vector of random variables with a joint probability density function PDF $= f(x)$. In groundwater models, random variables may be hydraulic conductivity, recharge, etc. Thus, the model output $h = g(X)$ is the groundwater head or the contamination level, for example, at a certain time and location. The general form of the multiple-dimensional integral to be solved can be written as

$$\mu_h = E[h] = E[g(X)] = \int_R g(X)f(X)dx \tag{1}$$

where $R$ is the probability space $R \in [0,1]$, $E$ is the statistics expectation, and $\mu_h$ is the mean value. The variance equals

$$\sigma^2 = E([g(X) - \mu_h]^2) = \int_R [g(X) - \mu_h]^2 f(X)dx. \tag{2}$$

Since it is very difficult to evaluate the integral in Equation (1) using analytical methods, sampling methods were employed to solve it. Thus, the mean in Equation (1) can be estimated using MCS. In crude MCS, we generate $n$ independent samples $x_1, x_2, \ldots, x_n$ from the density function $f(X)$ over a uniform space [0,1]. Thus, MCS estimation of the integral in Equation (1) is

$$E[g(X)]_{MC} = \mu_{MC} = \frac{1}{n}\sum_{i=1}^{n} g(x_i). \tag{3}$$

According to the strong law of large number, it is known that the sum in Equation (3) will converge to the exact value of Equation (1). Based on the Limit State Theorem, it is known that

$$\Pr\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) = \int_R g(X)f(X)dx. \tag{4}$$

From Equations (1), (3) and (4), it is evident that the mean estimated by MCS is unbiased.

The variance of the estimator is

$$\mathrm{var}[g(X)]_{MC} = \sigma^2_{MC} = \frac{1}{n}\sum_{i=1}^{n} (g(x_i) - \mu_h)^2. \tag{5}$$

Similarly, it was shown that the variance of MCS equals (Owen 1998)

$$\sigma^2_{MC} = \frac{\sigma^2}{n}. \tag{6}$$

According to Equation (6), the standard error of the estimated integral in Equation (1) using MCS is a function of $1/\sqrt{n}$. That means, the error of MCS is inversely proportional to the square root of the number of runs. For instance, to decrease the error by a factor of two, the number of runs should be increased by a factor of four. In other words, to decrease the standard error of MCS, the sampling size should be increased. However, increasing the sampling size can be computationally expensive. A better solution is to employ some technique of variance reduction. These techniques incorporate more information about the analysis directly into the estimator. The methods of variance reduction make MCS more deterministic, and thus, decrease error.

An example of $10 \times 10$ sampling realisations using MCS is shown in Figure 1(a). From this figure, one can see how the sampling points are randomly scattered. There are indeed some areas in the probability domain which are not sampled at all. This explains why too many sampling points are needed to get reliable output by MCS.

## Variance reduction of Monte Carlo

According to Equation (6), one should increase the sample size to reduce the error of MCS. In groundwater and contaminant transport modelling, however, it will be very time-consuming and computationally inefficient to increase the sampling size, which means increasing the number of runs. Therefore, different techniques have been developed to reduce the variance of MCS, and thus reducing the sampling size while preserving the accuracy. An outline of some of these techniques is presented in the following subsections.

## Importance sampling

Importance Sampling (IS) is aimed at reducing the variance of the classical MCS while maintaining its accuracy and efficiency. The premise behind IS is to generate samples where all the points contribute equally in Equation (3). Thus, it concentrates on drawing samples that have more impact on the system output ($g(X)$) than others. This can be

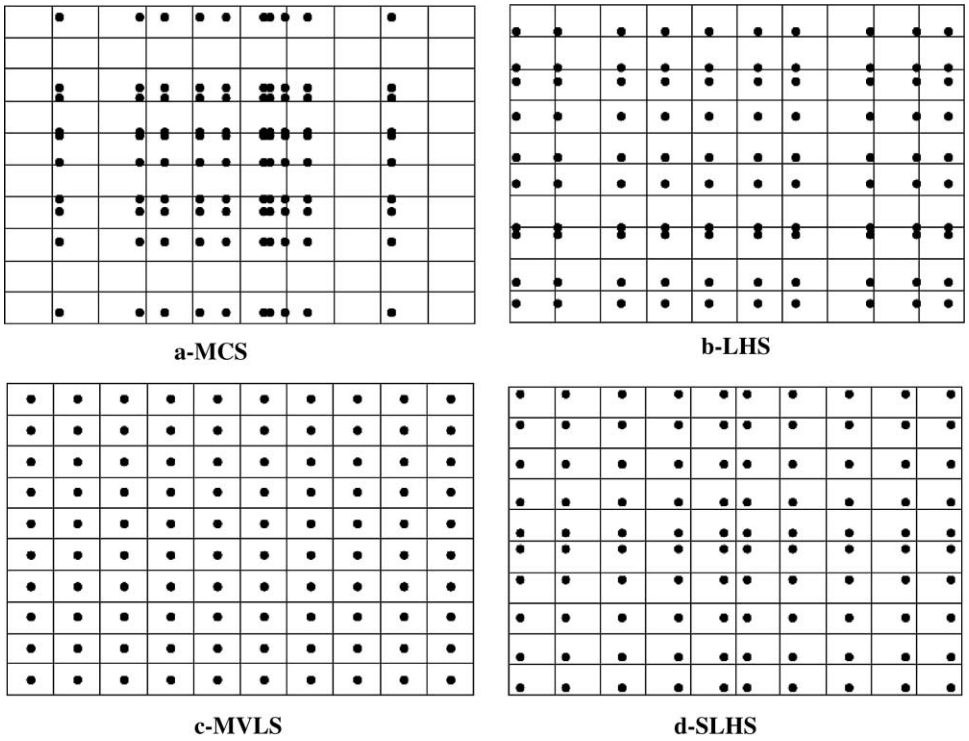


**Figure 1** | Example of $10 \times 10$ sampling points in MCS, LHS, MVLS and SLHS.

approached by sampling from another PDF and not from that of input random variables. Therefore, there will be a bias in mean estimation based on the new PDF. To eliminate the bias, a new function $g_1(X)$ can be defined as follows:

$$g_1(X) = g(X) \bullet p(X) \tag{7}$$

where $p(X) = f(X)/m(X)$ is called a weight function on $R$ such that $p(X) > 0$ whenever $p(X)|f(X)| > 0$. The efficiency of IS is highly dependent on the new selected PDF $m(X)$. The integration in Equation (1) can be rewritten in the following form:

$$\mu_1 = E[g_1(X)] = \int_R g_1(X)m(X)dx = \int_R g(X)p(X)m(X)dx$$

$$= \int_R \frac{g_1(X)f(X)}{m(X)}m(X)dx = \int_R g(X)f(X)dx. \tag{8}$$

From Equations (1) and (8), it is obvious that the mean remains the same though the samples are taken from different density functions. Similar to Equation (3), estimation of the mean based on importance sampling can be written as follows:

$$\mu_1 = \frac{1}{n}\sum^{n} g_1(x_i) = \frac{1}{n}\sum^{n} g(x_i)m(x_i). \tag{9}$$

The variance of $g(X)$ based on the importance function is

$$\sigma_{IS}^2 = E[(g_1(X) - \mu_1)^2] = \left(\int g_1(X)m(X)dx - \mu_{hl}\right)^2. \tag{10}$$

The variance of $g(X)$ based on IS can be estimated as

$$\sigma_1^2 = \frac{1}{n}\sum_{i=1}^{n} (g(x_i)m(x_i) - \mu_1)^2. \tag{11}$$

The variance of the estimator $\mu_1$ equals

$$\sigma_{IS}^2 = \frac{\sigma_1^2}{n}. \tag{12}$$

Comparing Equations (6) and (12), the variance of IS is not equal to the variance of MCS. That means, the root mean square error (RMSE) can be affected by IS for better or worse. The value of the RMSE depends on the importance density function $m(X)$. Therefore, if $m(X)$ was chosen properly, the variance of MCS can be reduced, otherwise it can be worse. As a result, selecting a good importance density function can reduce the variance achieved with classical MCS. The reader is referred to Owen (1998) for more details about importance sampling.

## Latin hypercube sampling

Latin Hypercube Sampling (LHS) is one type of stratified MCS. It was first suggested by McKay *et al.* (1979), and then further developed to improve its efficiency. LHS was used in different computer models for sensitivity and uncertainty analysis (Iman & Conover 1982; Iman & Helton 1988). This method seeks to make the samples more regular than random as in MCS. The idea of the LHS depends on subdivision of the sampling space into $N_s$ numbers of segment with equal probability and drawing one sample from each segment.

Once the segments are defined, each parameter is then randomised until a value that lies within each probability segment is found. That is, the samples are chosen randomly in such a way that each interval contains one sample. Then, the random numbers for each parameter are combined with the random numbers from the other parameters such that all possible combinations of segments are sampled. The LHS realisations can be generated based on the following formula:

$$x_{ij} = F^{-1}\left(\frac{\pi_{j(i)} - U_{ij}}{N_s}\right) \tag{13}$$

where:

$\pi_{ij}$: the random permutation of $1, \ldots, n$;
$n$: total number of realisations;
$F^{-1}$: inverse cumulative probability density function;
$U_{ij}$: a $U[0,1]$ random variable;
$N_s$: number of segments.
$j$: $1,2, \ldots, k$, where $k$ is the dimension of input vector $X$.

Doing so, instead of performing a large number of simulations, a smaller number of runs is needed with LHS. In LHS, the sampling points are scattered randomly in

each segment as shown in Figure 1(b). The LHS estimator of the function $g(X)$ is

$$\mu_{LHS} = \frac{1}{N_s}\sum_{i=1}^{N_s} g(x_i) \tag{14}$$

where
$x_i$: the random value selected in interval $i$;
$N_s$: number of segments.

The relation between variance of LHS and variance of MCS was shown to be equal to (Kollig & Keller 2002)

$$\mathrm{var}[g(X)_{LHS}] \leq \frac{N_s}{N_s - 1}\sigma^2_{MCS} \tag{15}$$

for all $N_s \geq 2$. According to Owen (1997), LHS with $N_s > 1$ segments is never worse than Monte Carlo with $n - 1$ numbers of samples. Stein (1987) has shown that LHS does not reduce the variance relative to simple random sampling, but the reduction depends on the simulated function itself. As a conclusion, LHS can be used with asmaller number of runs than MCS, and it preserves the accuracy of the output.

### Mean-value lattice sampling (MVLS)

The Mean-Value Lattice Sampling (MVLS) was developed from LHS and it has been used in different stochastic models. The idea of the MVLS depends on subdivision of the probability domain into segments of equal probability in the same way as LHS. Unlike LHS, MVLS has less randomness since realisations in MVLS are chosen at the mean value of each segment instead of at random values. Thus, the MVLS is more deterministic than LHS. As with LHS, the distribution of all the input random variables should be transformed to standard normal distribution. Then, subdivision of probability space into segments and determination of the mean value can be done easily. After identification of the mean value of each segment, these points of mean values can be used in sampling. To ensure all possible combinations of sampling are considered, the sampling points of each random variable are combined with sampling points of other variables. For instance, given two random variables $X$ and $Y$, with $n$ and $m$ the number of realisations, respectively, the number of runs required are $n \times m$. Figure 1(c) shows an example of $10 \times 10$ sampling points in MVLS. Obviously, MVLS is more deterministic than LHS, and thus has less variance in the mean estimation, on the one hand. On the other hand, sampling at the mean values makes MVLS omit some other areas in the probability domain.

The estimator of MVLS is equal to

$$\mu_{MVLS} = \frac{1}{N_s}\sum_{i=1}^{N_s} g(\mu_{Si}) \tag{16}$$

where $\mu_{Si}$ is the mean value of segment $i$. It was shown that the variance of MVLS is not always less than that of LHS (Owen 1992). Besides, MVLS suffers from different drawbacks. MVLS does not reproduce the tail probability very well, and it is not better than LHS if the dimension of the problem is high (Zhang & Pinder 2003). In risk analysis, the probability is usually located at the tail, and thus MVLS does not work well in this case. Therefore, a new methodology that overcomes the limitations of MVLS is proposed and used in groundwater pollution risk assessment and uncertainty analysis.

## THE PROPOSED METHODOLOGY: STRATIFIED LHS

In the previously discussed methods, the accuracy is always measured by investigation of the mean estimator of each method. In risk analysis, however, not only the mean is of concern, but also the probability. Therefore, the probability of exceeding a pre-defined threshold value should be considered when investigating the performance of sampling methods. The new methodology (Stratified Latin Hypercube Sampling, (SLHS)) is proposed to obtain better results in risk analysis than the previously mentioned methods. SLHS depends on the same principle of stratification, but with different realisations points to consider all areas of probability. In MVLS, selection of realisations is made in the centre of each segment ($U_{ij} = 0.5$ in Equation (13)). In probabilistic modelling and risk analysis, the probability of failure is very often located at the tail (low or high probability). Therefore, since the tail probability is not well presented in MVLS (Zhang & Pinder 2003), it fails to predict the small or high probability.

As discussed before, LHS and MVLS have less variance (i.e. less error) than crude MCS. Owen (1992) has shown that the difference between estimates using LHS and MVLS is a function of $(N_s^{1/2})$, where $n$ is the number of realisations. Therefore, MVLS is superior in the case where the bias of LHS is not a function of $(N_s^{1/2})$. As a result, MVLS is better than LHS only for small $n$ (Owen 1992). It has been shown the proper selection of sampling points can reduce the variance (Lemieux & L'Ecuyer 2000). The proposed methodology is a combination of LHS and MVLS. That is, in this method, the same procedure of LHS is followed but with four realisations in each hypercube, in addition to the sampling point at the mean. Thus, the total number of realisations in each hypercube is five.

The five realisations in SLHS are chosen to cope with the tail probability and to cover the whole hypercube. These points are 5th quantile, 25th quantile, 50th quantile, 75th quantile and 95th quantile in each segment (see Figure 1(d)). Therefore, SLHS covers more areas of PDF than any other method does, and promotes the accuracy as will be shown in the illustrative example. To avoid any bias in estimation, and to get all the possible results in the outcome domain, the maximum and minimum values of each random variable are considered in this SLHS.

### Illustrative example

In the following subsections, the power of the proposed methodology will be demonstrated using a one-dimensional contaminant transport model. The velocity vector and longitudinal dispersion in this example are assumed to be random variables. No correlation will be considered since there is no specific method for coupling correlation techniques with this sampling methodology. Also, the aim of this study is to demonstrate the new methodology and to compare it with other methods. Investigation of correlation structure in sampling is left to further research.

### One-dimensional contaminant transport model

In this example, a one-dimensional contaminant transport model is considered. The one-dimensional transport equation (without chemical reaction) is

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - V\frac{\partial C}{\partial x} \tag{17}$$

where $C$ is the concentration of pollutant, $V$ is the groundwater velocity and $D$ is the dispersion coefficient. The initial and boundary conditions of Equation (17) are

$$C(x,0) = 0$$

$$C(0,t) = c_0$$

$$C(\infty,t) = 0.$$

The solution of Equation (17) based on the given boundary and initial condition is (Freeze & Cherry 1979)

$$C = \frac{c_0}{2}\left[erfc\left(\frac{x - Vt}{\sqrt{4\alpha_L Vt}}\right) + \exp\left(\frac{x}{\alpha_L}\right)erfc\left(\frac{x + Vt}{\sqrt{4\alpha_L Vt}}\right)\right] \tag{18}$$

where $c_0$ is the initial concentration, erfc is the complementary error function, $V$ is the velocity, $\alpha_L$ is the longitudinal dispersivity and $x$ is the distance from the source of contamination. In Equation (18), two random variables are considered: the velocity and the longitudinal dispersion. Both random variables were assumed to be constant in the physical space and variable in the probability space. The probability distribution of the velocity and the longitudinal dispersion are assumed to be log-normal.

## METHODOLOGY AND APPLICATION

A one-dimensional contaminated field is given here as an example to demonstrate the new methodology and compare it with other sampling methods discussed before. Therefore, a random velocity field with log-normal distribution is assumed. The mean and standard deviation of velocity are assumed to be 2.5 m/d and 0.3 m/d, respectively (in real space).

Similarly, the mean and standard deviation of longitudinal dispersion are 15.0 m and 5.0 m, respectively. Other parameters are given as: $c_0 = 100$ mg/l, $x = 50.0$ m and $t = 10$ d. The maximum and minimum values of velocity are 5.0 m/d and 0.0 m/d, respectively, and the maximum and minimum values of longitudinal dispersion are 30 and 1.0, respectively. Both parameters were assumed to have log-normal distribution. The same data were used with the

different sampling methods and the results are compared with the proposed method.

The first step is to transfer the vector of random variables from log-normal to normal distribution. This transformation was done based on the Rackwitz & Fiessler (1978) approach. All computations were carried out based on the transformed values of the random parameter.

## RESULTS AND DISCUSSION

In the following subsections, results of the SLHS as well as other methods are presented. First, the mean value of contaminant concentration was obtained using different methods. Second, risk and sensitivity analysis were carried out.

### Mean-value estimation

The four different methods (SLHS, MCS, MVLS and LHS) were used to estimate the mean value of contaminant concentration given the previously mentioned initial and boundary conditions. The result of mean estimation is shown in Figure 2. The true mean value of $c/c_0$ equals 0.948746 based on substituting mean values of random parameters in Equation (18). To get a better view of different methods' performance, Figure 2 was separated into two diagrams to illustrate the performance of all methods. The small diagram shows the performance of all methods and the big one shows the performance of SLHS, MVLS and LHS with the true mean.

From Figure 2, one can see that MCS has the worst output among the other methods.

As appears in the figure, SLHS and MVLS methods have a better convergence rate than LHS. The results of SLHS and MVLS are close to each other but in all cases better than the result of LHS. SLHS and MVLS have converged to the true mean after 1000 runs, while LHS needed about 2000 runs to converge to the true mean. That is also clear in the variance of the mean estimator of each method as shown in the variance diagram in Figure 3. Similar to the mean estimation, SLHS and MVLS have less variance in comparison with LHS and MCS methods. Although the variance of LHS is less than that of MCS, but it is not as good as SLHS and MVLS. In general, it can be concluded that all the methods (except MCS) have a good performance for mean estimation.

### Risk analysis

Any hydrogeological system is usually subjected to different stresses in the form of pollution or pumping. The probability of survival of a hydrogeological system depends on its ability to resist the stresses without being polluted or overexploited. The failure occurs when the pollution, for example, exceeds the maximum permissible level. The factors that lead to such a failure of the hydrogeological

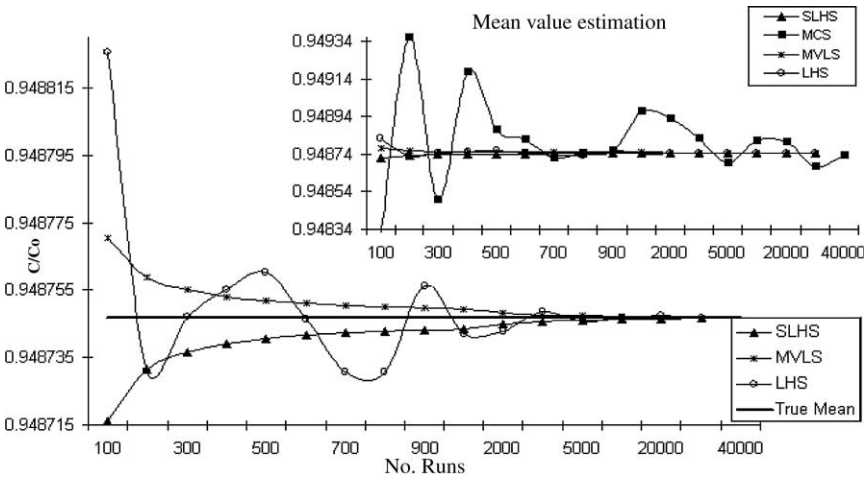


**Figure 2** | Estimation of mean pollutant concentration using different sampling methods (the small figure shows the results of all methods and the large one shows results without MCS).

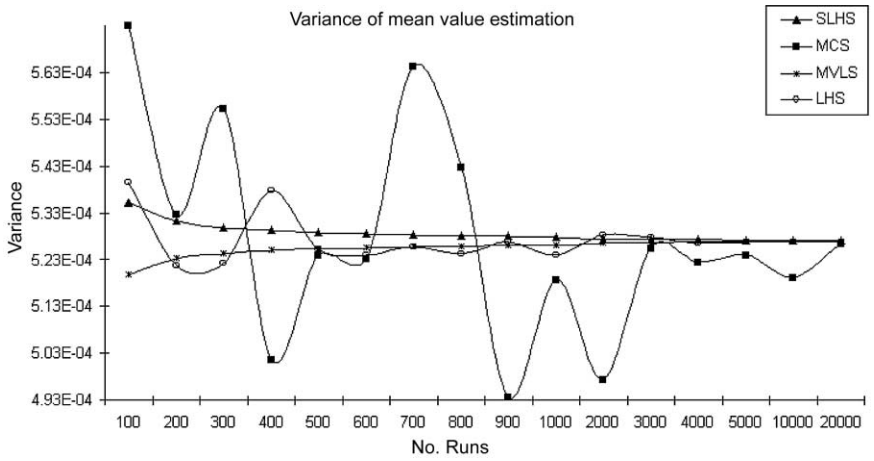


**Figure 3** | Variance of mean estimators.

system and the imposed stresses on that system are very often uncertain. Therefore, building and analysis of any hydrogeological model is subjected always to uncertainty.

The reliability of the system ($P_s$) is defined as the probability of sustaining a pollution load in which the resistance $R$ of the system exceeds the stresses ($L$); that is:

$$P_s = P(L \leq R). \tag{19}$$

On the other hand, the risk $P_f$ is the complement of the reliability, which can be expressed as

$$Pf = P(L \geq R) = 1 - P_s. \tag{20}$$

In hydrological engineering, the loads arise usually from the natural events such as contamination load. Since natural events occur randomly in time and space, they should be considered as random variables. The probability of failure is given by

$$P_f = P[g(X) \leq C_{\max}] = \int_{g(X) \leq C_{\max}} f_x(X)dx \tag{21}$$

where $f_x(X)$ is the joint probability density function of the random variables $X = (x_1, x_2, \ldots, x_n)$ and $C_{\max}$ is the maximum permissible value. The problem now is to evaluate the integral in Equation (21). This integral is very difficult to evaluate using classical analytical methods. Many other methods, such as the reliability method and Monte Carlo method, were used to evaluate this integral. In this study, the risk will be estimated using the proposed method of SLHS and the results are compared with results of MCS, MVLS and LHS. The input samples are drawn randomly from the basic variables according to their corresponding probabilistic properties (Press *et al.* 1992), and then substituted in Equation (21). The general procedure for sampling simulation can be described as follows:

- Generation of a single value for each random variable based on the PDF of each one.
- Assessment of the performance function if $g(X) < 0 \Rightarrow$ system failure. Therefore, the following function can be identified:

$$I(x_i) = \begin{cases} 1 \rightarrow g(x_i) \leq 0 \\ 0 \rightarrow g(x_i) > 0 \end{cases} \tag{22}$$

- Repeat steps 1 and 2 $n$ times.
- Estimate the probability of failure as follows:

$$P_f = \int_{G(x) \leq 0} f_x(x)dx = \int_{x \subset R} I(x) f_x(x)dx = E[I(x)]. \tag{23}$$

Based on Equation (23), the probability of failure is

$$E[P_f] = \frac{1}{n}\sum_{i=1}^{n} p_i = \frac{n_f}{n} \tag{24}$$

where $n_f$ is the number of failures and $n$ is the total number of simulation. It was found that the output values of

the model vary between 83 and 99 mg/l based on the predefined upper and lower limits of the input values. Therefore, three values of maximum permissible concentrations were identified to cope with low, medium and high probabilities. These three probability values are

- Low risk ⇒ probability of $C \leq 84.0$ mg/l.
- Medium risk ⇒ probability of $C \leq 95.0$ mg/l.
- High risk ⇒ probability of $C \leq 99.0$ mg/l.

The results of the above-mentioned probabilities are shown in Figures 4–6. As seen in the figures, SLHS converges faster than other methods in cases of high and low probabilities. This is indeed because the SLHS pays more attention to the tail (0.05% and 0.95% quantiles). In the medium probability, however, SLHS does not seem to have better performance than MVLS or LHS.

All the methods, except MCS, converge almost with the same rate to the true value with medium probability (Figure 5).

## Sensitivity analysis

Sensitivity analysis is very important in groundwater modelling since it is used to allocate and design the sampling site. This analysis is very helpful in designing site investigation wells since the sampling points should be located at points of high sensitivity. Moreover, sensitivity analysis can help in the modelling process to pay more attention on the important input variables, and thus improve the calibration process. As a result, uncertainty of model output can be reduced if sensitivity analysis is being carried out. Relative sensitivity is the ratio of the change of one variable with respect to the change in the other. Given a dependent variable $y$ and independent variable $x_i$, the relative sensitivity can be defined as (Tung 1999)

$$S_{\%} = \left(\frac{\partial y}{\partial x}\right)\left(\frac{x_i}{y}\right) \tag{25}$$

where $S_{\%}$ is a dimensionless quantity measuring the percentage change in the dependent variable $y$ due to one percentage change in the independent variable $x_i$. In this study, the sensitivity of model output (Equation (18)) to each input random variable was evaluated. That is, the sensitivity of pollutant concentration to velocity vector and longitudinal dispersion were calculated using the different methods mentioned earlier. The sensitivity measure can be obtained by evaluating the partial derivative of Equation (18) with respect to velocity and longitudinal dispersion coefficient. The resulted derivative of pollution concentration with respect to velocity and longitudinal dispersion, respectively,are as follows:

$$\frac{\partial C}{\partial V} = \frac{c_0 x}{2\sqrt{\pi \alpha_L V t}} \exp\left[-\frac{(x - Vt)^2}{4\alpha_L V t}\right] \tag{26}$$

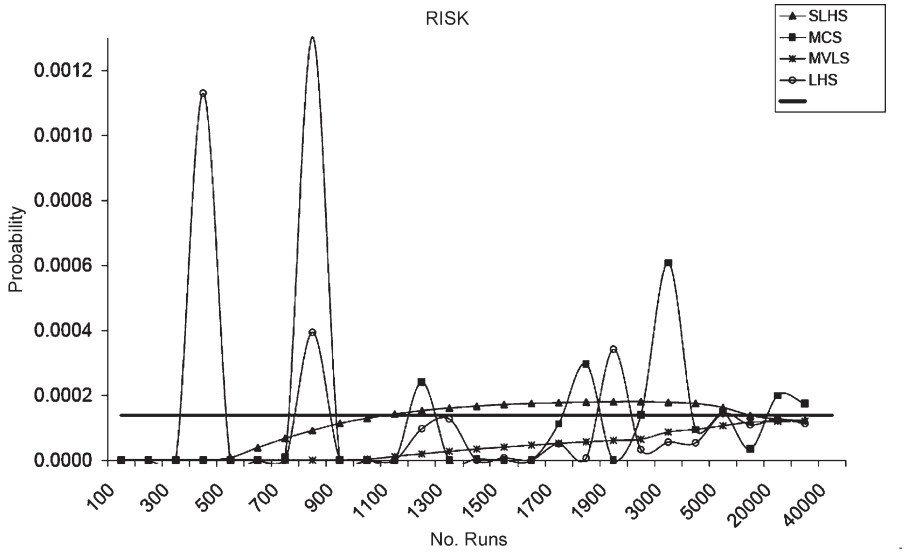


**Figure 4** | Low risk: probability of $C \leq 84.0$ mg/l.

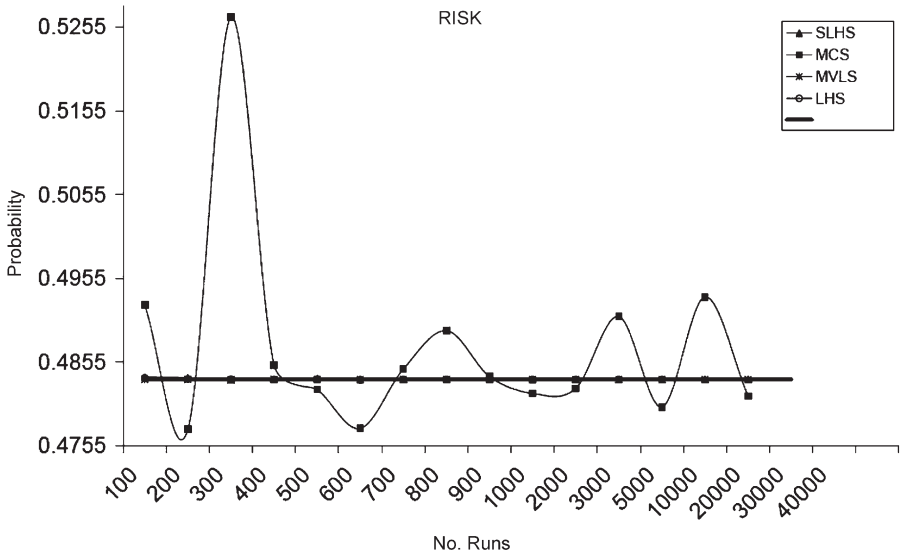


**Figure 5** | Medium risk: probability of $C \leq 95.0$ mg/l.

and

$$\frac{\partial C}{\partial \alpha_L} = \frac{c_0 x}{2}\left[\frac{\exp\left[\frac{(x-Vt)^2}{4\alpha_L Vt}\right]}{\sqrt{\pi \alpha_L Vt}} - \frac{\exp\frac{x}{\alpha_L}\, erfc\left[\frac{(x+Vt)}{\sqrt{4\alpha_L Vt}}\right]}{\alpha_L}\right]. \quad (27)$$

Equations (26) and (27) were evaluated using four different methods: MCS, MVLS, LHS and SLHS. The results of sensitivity analysis are shown in Figures 7 and 8.

From sensitivity figures, the contaminant concentration is directly proportional to the velocity vector and inversely proportional to longitudinal dispersion. MVLS and SLHS have almost the same convergence rate while LHS has a little slower convergence rate, and MCS is very slow in the sensitivity calculation. It is also found the contaminant concentration is more sensitive to likely changes in longitudinal dispersion than to velocity.

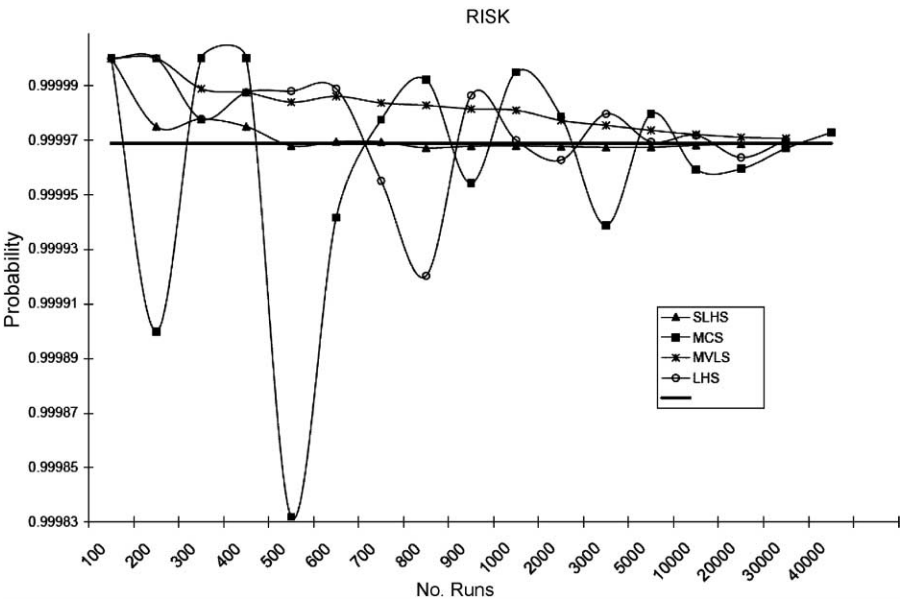


**Figure 6** | High risk: probability of $C \leq 99.0$ mg/l.

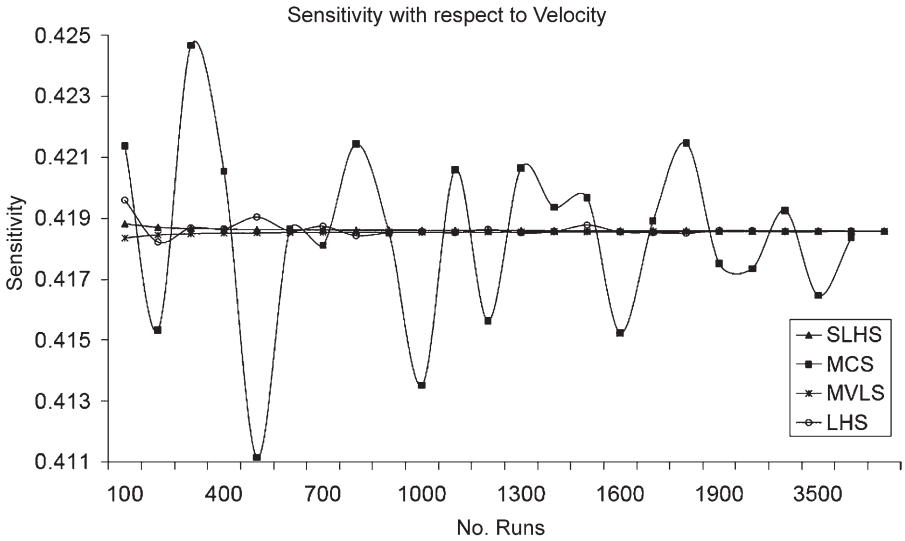


**Figure 7** | Sensitivity of contaminant concentration with respect to velocity.

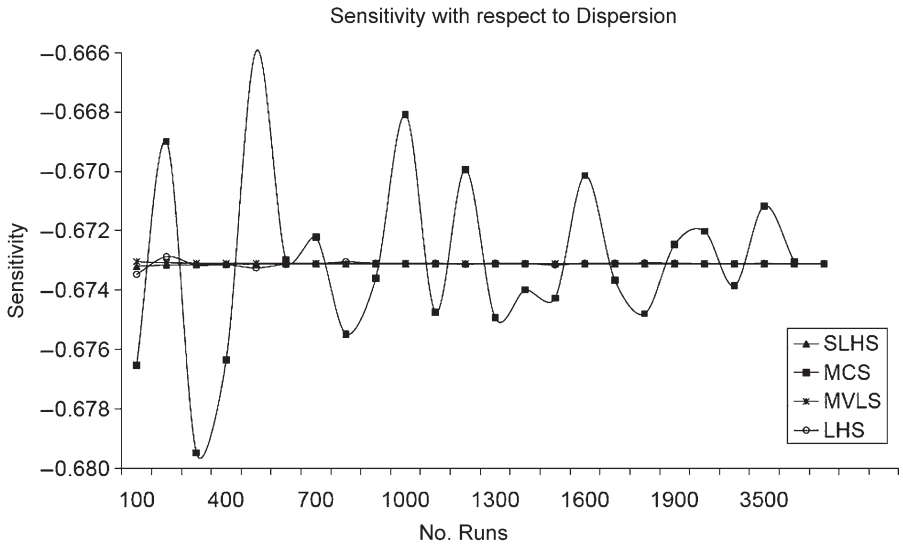


**Figure 8** | Sensitivity of contaminant concentration with respect to longitudinal dispersion.

## SUMMARY AND CONCLUSIONS

In this study, a new sampling approach was introduced and demonstrated using a one-dimensional contaminant transport model. The new approach stratifies the LHS considering all the probability space. Also, the maximum and minimum values of each input random variable were considered to avoid any bias and to avoid obtaining results outside the outcome domain. Because different random realisations might have different outputs, combinations of all possible input parameters were made to avoid any bias in computations.

In mean estimation, it is clear that all the methods (except MCS) converge quickly when used for mean value estimation. Comparing SLHS, MVLS and LHS, it is found that SLHS and MVLS converge to the true mean almost with the same rate, whereas LHS has a slower rate. MCS has always a slow convergence rate when compared with

the other methods. This is also cogent in the variance of mean estimate, which is shown in Figure 3.

SHLS was superior in risk analysis compared to the other methods, especially at high and low risk. This is reasonable because SLHS reproduces the tail probability more than any other method does. In high risk case, SLHS converges to the true value after about 500 runs, while MVLS required more than 30,000 runs to converge. LHS has oscillated and has shown the worse performance. The worst performance is that of MCS as shown in Figure 6.

Moving the risk from high to medium, none of the methods has been shown to be advantageous (except MCS). SHLS, MVLS and LHS have almost similar results.

In low risk cases, the performance of the methods is almost similar to their performance in high risk cases.

From sensitivity analysis, SLHS and MVLS have shown no difference in their results, while the result of LHS was a little worse. Similar to risk and mean analysis, MCS has the worst results for the previously mentioned reasons. From sensitivity analysis, it is found that the contaminant concentration is more sensitive to likely changes in longitudinal dispersion than to likely changes in groundwater velocity.

In summary, the SLHS has shown to be advantageous in risk analysis. It has a faster convergence rate than MVLS or LHS when used in problems with a high or low probability of occurrence. In addition, SLHS maintains the same accuracy as MVLS and LHS when used with events that have a medium probability of occurrence. To apply the proposed methodology of SLHS to real hydrogeological problems, it should be extended to consider correlated variables.

## REFERENCES

Dimov, I., Jaekel, U., Vereecken, H. & Wendt, D. 1998 A numerical approach for determination of sources in reactive transport. In *Monte Carlo and Quasi-Monte Carlo Methods. Lecture Notes in Statistics*, vol 127 (eds Niederreiter H., Hellekalek P., Larcher G. & Zinterhof P.), pp. 189–205. Springer-Verlag, New York.

Freeze, R. & Cherry, J. 1979 *Groundwater*. Prentice-Hall, Englewood Cliffs, NJ.

Iman, R. & Conover, W. 1982 A distribution-free approach to inducing rank correlation among input variables. *Statistics-Simulation and Computation* **11**, 311–334.

Iman, R. & Helton, J. 1988 An investigation of uncertainty and sensitivity techniques for computer models. *Risk Analysis* **8** (1), 71–90.

Kollig, T. & Keller, A. 2002 Efficient multidimensional sampling. *Computer Graphics Forum* **21** (3), 557 doi:10.1111/1467-8659.00706.

Lemieux, L. & L'Ecuyer, P. 2000 Using lattice rules for variance reduction in simulation. In *Proceedings of the 2000 Winter Simulation Conference*, pp. 509–516. IEEE Press. Piscataway, NJ.

Lioua, T. & Der Yeh, H. 1997 Conditional expectation for evaluation of risk groundwater flow and solute transport: one dimensional analysis. *Journal of Hydrology* **199** (3–4), 378–402.

McKay, M., Conover, W. & Beckman, R. 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.

Owen, A. 1992 Orthogonal array for computer experiments, integration and visualization. *Statistica Sinica* **2**, 439–452.

Owen, A. 1997 Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis* **34** (5), 1884–1910.

Owen, A. 1998 Latin supercube sampling for very high dimensional simulations. *ACM Transactions on Modeling and Computer Simulation* **8** (2), 71–102.

Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. 1992 *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edn. Cambridge University Press, Cambridge.

Rackwitz, R. & Fiessler, B. 1978 Structural reliability under combined random load sequences. *Computers and Structures* **9**, 489–494.

Stein, M. 1987 Large sample properties of simulations using latin hypercube sampling. *Technometrics* **29** (2), 143–151.

Tung, Y. 1999 Risk/reliability-based hydraulic engineering design. In *Hydraulic Design Handbook* (ed. L. Mays), pp. 7.1–7.55. McGraw-Hill, New York.

Zhang, Y. & Pinder, G. 2003 Latin hypercube lattice sample selection strategy for correlated random hydraulic conductivity fields. *Water Resources Research* **39** (8) doi:11-1/11-3.