Advances in Water Resources 16 (1993) 305–312

On the optimal selection of interpolation points for use in Monte Carlo simulations

Rao S. Govindaraju

Department of Civil Engineering, Seaton Hall, Kansas State University, Manhattan, Kansas 66506, USA

Ahlfeld and Pinder recently proposed a method for reducing the enormous computational burden associated with Monte Carlo simulations. It was suggested that the solution be evaluated at only a few select points (a knot sequence) over the parameter space of the random quantity and the solution elsewhere be obtained through interpolation. This paper addresses the issue of the choice of points in the knot sequence. The main thesis of the present work is that these points need to be chosen appropriately to maximize their utility and yield the fewest interpolation errors for the Monte Carlo procedure. An optimal choice of interpolation points should be based on the underlying probability distribution of the random variable(s). This is illustrated through an example for which analytical solutions are available. The statistical quantities of interest considered in this paper are the mean, variance, and probability densities of the state variable.

Key words: Monte Carlo, simulations, interpolation modeling.

INTRODUCTION

The modeling of many physical phenomena requires the use of stochastic approaches to account for uncertainty in the external and internal influences on the system. These could result from randomness in the parameters, in the initial and boundary conditions, and in the forcing functions. For example, the partial differential equations governing water movement and solute transport in porous media are rendered stochastic largely because of spatial variability exhibited by many field soils (Nielsen *et al.*,[19] Biggar and Nielsen,[4] Vieira *et al.*,[25] Dagan[10]). Analytical solutions to determine water flow and solute transport in the subsurface are available in special instances of regular boundaries and homogeneous media. When the equations are treated as stochastic, such analytical solutions become even more scarce. Some of the common stochastic techniques applied to subsurface environments are: the conditional-and-unconditional-probabilities approach (Dagan[8,9]); the spectral approach in conjunction with the perturbation approximation (Bakr *et al.*,[3] Gutjahr *et al.*,[14] Mizell *et al.*, 18): the cumulant-expansion method (Chu and Sposito,[6] Sposito and Barry,[24] Kavvas *et al.*[16]): and the semigroup approach (Serrano[20,21]). However, most of these techniques (except perhaps the cumulant-expansion approach) have restricted application in terms of needing special boundary conditions and analytical solutions for the deterministic problem.

A brute-force technique to solve any stochastic groundwater-flow and transport problem is through Monte Carlo simulations (Freeze,[11] Smith and Freeze,[22,23] Hopmans *et al.*[15]). The major criticism of this method is the enormous computational burden that is frequently associated with it. Realizations are generated from the known probability distributions of the random quantities. A deterministic solution of the problem based on realized values of the uncertain quantities results in a corresponding realization of the dependent variable. A large number of such trials form a synthetic ensemble for the dependent variable. The number of Monte Carlo trials that need to be performed depends on factors such as the nature of stochasticity, the convergence criteria chosen, and the characteristics of the physical problem under consideration. Despite the computational effort involved, the Monte Carlo technique is often used to test theoretical results (Govindaraju and Kavvas[13]) and is the only viable method for complicated problems when simpler solutions do not exist.

Ahlfeld and Pinder[1] recently proposed a method for reducing the computational requirement during Monte Carlo simulations. They used the problem of stochastic

Advances in Water Resources 0309-1708/94/$07.00

solute transport in a homogeneous saturated aquifer (whose hydraulic conductivity is a random quantity) as an example to illustrate their method. The basic idea in their work consists in selection of a limited number of points (called a knot sequence in this study) in the parameter space of the random hydraulic conductivity. By using the finite-element method, the concentration at particular space-time points is evaluated for each of the hydraulic conductivites in the knot sequence. During the Monte Carlo procedure, the concentrations for different values of hydraulic conductivity (i.e. those not in the knot sequence) are then evaluated through interpolation, rather than by solving the physical equations of groundwater flow and solute transport. Since interpolation usually requires negligible effort compared with the numerical solution of the governing partial differential equations, the procedure presented by Ahlfeld and Pinder[1] may be as much as two orders of magnitude faster than the regular Monte Carlo simulations. The study of Ahlfeld and Pinder[1] showed that ten points in the knot sequence were sufficient to obtain results with acceptable accuracy when compared with results from 1000 Monte Carlo simulations. Since 1000–10000 Monte Carlo trial are not uncommon in dealing with spatial variability of hydraulic conductivity, interpolation of state variables over the parameter space results in significant savings in computer effort.

Consequently, the choice of the points in the knot sequence becomes an important issue. Because the number of interpolation points is far fewer than the total number of Monte Carlo simulations, an optimal placement of these points needs to be designed to maximize their utility or minimize the error in predictions. The purpose of this paper is to examine the influence of different knot sequences based on various strategies on the prediction of the statistical properties of the state variables (mean, variance, and probability densities). To define appropriate criteria for comparison of different choices of knot sequences, a hypothetical example is chosen for which analytical solutions are available. Two different strategies for selecting the knot sequence are presented. One of them places the interpolation points at equal distances, whereas the placement of interpolation points is based on the underlying probability distribution of the random parameter in the other strategy. Results from 9000 Monte Carlo simulations are also included. Whether a particular knot sequence is optimal (in terms of the smallest error) depends on the statistical property of interest and the dependence of the state variable on the random quantity. The results clearly indicate that the choice of knot sequence may have serious consequences in terms of the quality of predicted results.

MATHEMATICAL STATEMENT OF THE PROBLEM

Let us say that the outcome of a particular state variable Y depends on some parameter X, which behaves in a random fashion. (Here Y may be identified with solute concentration at some location of interest, while X represents the random hydraulic conductivity of the homogeneous aquifer.) The dependence of Y on X may be expressed in the general form:

$$Y = \mathrm{g}(X; \mathbf{p}) \tag{1}$$

where $\mathbf{p}$ is the vector of other non-random parameters that influence the solution. An analytical solution to the stochastic problem requires that the functional form of $\mathrm{g}(\cdot)$ be explicitly known as a first step. In most practical cases of interest, this is not available, and numerical solutions have to be used for this evaluation. Let $\mathrm{f}_x(x)$ denote the known probability density of X. Then the kth moment of Y may be defined as:

$$\langle Y^k \rangle = \int \mathrm{g}(x, \mathbf{p})^k \mathrm{f}_x(x)\mathrm{d}x \tag{2}$$

where $\langle \cdot \rangle$ denotes the expectation operation and the integration on the right-hand side is over all possible values of X (i.e. the entire parameter space). If X is a discrete random variable, then the integral in eqn (2) would be replaced by the appropriate sum. For the purposes of this study, X is assumed to be a continuous random variable. In a Monte Carlo approach, an approximate evaluation of eqn (2) is conducted by random generation of sample values of $X(= x_i)$, evaluating Y numerically for each x_i, and summing over all simulations, i.e.:

$$\langle Y^k \rangle = \left\{ \sum_{i=1}^{NM} \mathrm{g}(x_i; \mathbf{p})^k \right\} \Big/ NM \tag{3}$$

where NM is the total number of simulations. This procedure is computer-intensive because each $\mathrm{g}(x_i, \boldsymbol{p})$ needs to be computed independently and usually entails the numerical solution of a large problem. The larger the number of Monte Carlo simulations NM, the more accurate are the predictions from eqn (3), but there is also a corresponding increase in cost.

This brute-force Monte Carlo procedure does not take any advantage of the nature of the relationship between Y and X through the functional form $\mathrm{g}(\cdot)$. In most cases of practical interest, the problem is well posed, so that a small variation in X results in a correspondingly small change in the solution $Y = \mathrm{g}(X)$. The 'smoothness' or continuity requirements of $\mathrm{g}(X)$ depend on what derivative of Y will be used in the interpolation procedure. We consider evaluating $Y_j(= \mathrm{g}(x_j; \boldsymbol{p}))$ for a limited number of x_j values

($j = 1, 2, \ldots, ninpol$) called interpolation points (knot sequence), where *ninpol* is the number of interpolation points. Ahlfeld and Pinder[1] used Hermite-cubic polynomials for interpolation because they were interested in the state variable and its first derivative. For simplicity, linear-basis functions are chosen for interpolation in this study. Thus the interpolated value of Y for any x not belonging to the knot sequence is:

$$\hat{Y}(x) = \hat{g}(x:p) = \sum_{j=1}^{ninpol} Y_j o_j(x) \tag{4}$$

where $o_j(x)$ are linear-basis functions with the usual compact support. These functions occur frequently in the numerical solution of partial differential equations (Lapidus and Pinder[17]). It can be shown that, as the number of interpolation points increases, the interpolation error reduces to zero. More specifically (Conte and de Boor,[7] Atkinson[2]), we have:

$$|Y(x) - \hat{Y}(x)| \leq \frac{\max |x_j - x_{j-1}|^2}{8} * \max \left| \frac{d^2 Y}{dx^2} \right| \tag{5}$$

where $x_1 \leq x \leq x_{ninpol}$ in the above equation. It is clear that $\hat{Y}(x) = Y(x)$ when x coincides with an interpolation point. A decrease in the error $|\hat{Y}(x) - Y(x)|$ may be achieved by reducing the maximum separation between nodal points. The error term depends on the nature of $g(X)$ through the term d^2Y/dx^2 in eqn (5). Thus, to maximize the use of the interpolation formula, eqn (4), in Monte Carlo simulations, the location of interpolation points in the knot sequence becomes important. Apart from Monte Carlo simulations without interpolations, two different strategies for choosing interpolation points are studied here to show the disparity in results that may occur on the basis of the choice of interpolation points in the knot sequence.

Strategy 1: Knot sequence based on probability density of X

The basic idea in this strategy is as follows. The number of interpolation points in a given interval in the parameter space is proportional to the probability mass within the interval. Suppose that we decide on *ninpol* interpolation points. Then the number of interpolation points within the interval ($X_1 < X < X_2$) is given by:

$$N_{1-2} = ninpol * \left(\int_{x1}^{x2} f_x(x)\, dx \right) \tag{6}$$

This means that *ninpol*-1 intervals are used to partition the parameter space in such a way that the probability mass within each interval is the same ($= 1.0/(ninpol - 1)$). This method allows the points in the knot sequence to be distributed according to the probability density of X. More interpolation points are therefore concentrated in regions of higher probability of occurrence of X. It is likely that more realizations will have a smaller interpolation error. The upper (X_{max}) and lower (X_{min}) limits of interpolation are obtained from the maximum and minimum generated values of X.

Strategy 2: Uniform distribution of interpolation points

In this strategy, the interpolation points are spaced equidistantly over the parameter space. No effort is made to relate the spacing of the interpolation points to the underlying distribution of the random parameter X.

Strategy 3: Monte Carlo simulations without interpolation

This strategy does not use any interpolation and is based on the conventional simulation strategy whereby Y is evaluated for each realized value of X by solving the physical problem. These results are also included here for comparison purposes.

It is clear that the results of strategies 1 and 2 should approach those of strategy 3 asymptotically as the number of interpolation points increases indefinitely. Ahlfeld and Pinder[1] chose a selection procedure based on bisection of the parameter space to choose new interpolation points. Their study does not make any comparisons of the performance of different strategies for choice of interpolation points.

EXAMPLE AND RESULTS

To illustrate the relevance of the proper placement of interpolation points, an example is considered that yields analytical solutions. This example has been particularly designed to bring out the differences in solutions between strategies 1 and 2. In more practical problems, the differences may not be as pronounced as the results shown in this paper. Let X have an exponential distribution given by:

$$f_x(x) = \lambda \exp(-\lambda x). \quad x > 0. \quad \lambda > 0$$
$$= 0. \quad x < 0 \tag{7}$$

where λ (spelt as lambda in all the figures) is the parameter of the exponential distribution. Suppose the function $Y = g(X)$ has the following form:

$$Y(X) = g(X) = a \exp(bX) \tag{8}$$

where a and b are constants. Depending on the values of the constants a and b in eqn (8), the functional dependance of Y on X can vary quite dramatically.

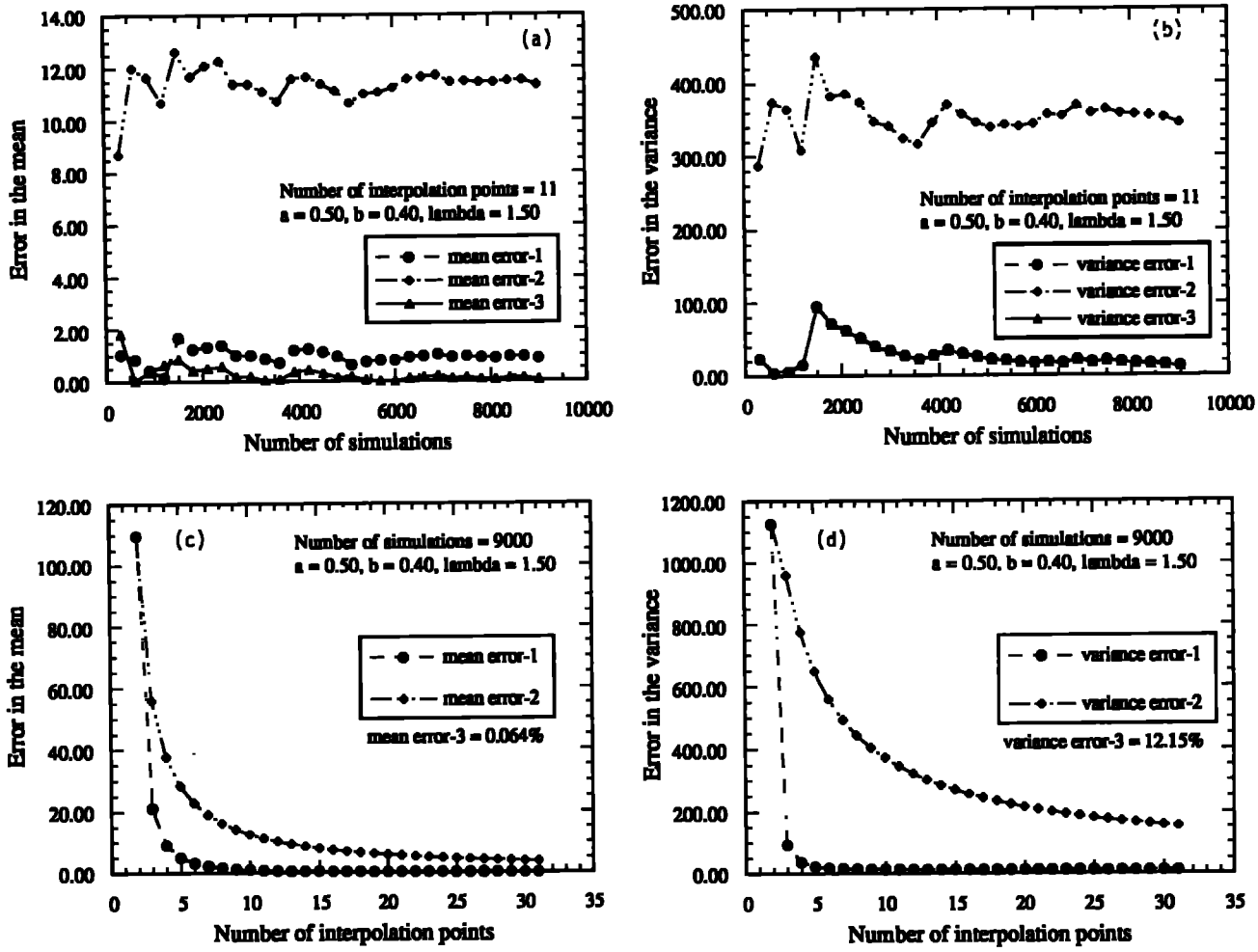

Fig. 1. Comparison of errors (%) for the mean and variance by using strategies 1, 2, and 3 as a function of the number of Monte Carlo simulations (a) and (b) and number of interpolation points (c) and (d). The numbers in the legends refer to the strategy.

The expectation of Y is:

$$\langle Y\rangle = a\langle \exp(bX)\rangle = a\frac{\lambda}{\lambda - b} \tag{9}$$

where is is assumed that $\lambda > b$[1]. Equation (9) results from the definition of the moment generating function of X given by:

$$F_x(t) = \langle e^{tX}\rangle = \frac{\lambda}{\lambda - t} \tag{10}$$

Similarly, the second moment of Y may be obtained as:

$$\langle Y^2\rangle = \langle a^2 e^{2bX}\rangle = a^2\frac{\lambda}{\lambda - 2b} \tag{11}$$

For eqn (11) to hold, $\lambda > 2b$. The analytical expression for the variance of Y is then given as:

$$\text{var}(Y) = a^2\left\{\frac{\lambda}{\lambda - 2b} - \frac{\lambda^2}{(\lambda - b)^2}\right\} \tag{12}$$

and finally, by using probabilistic transforms, the probability density of Y may be expressed as:

$$f_Y(y) = \frac{\lambda}{|by|}\exp\left[-\frac{\lambda}{b}\ln(y/a)\right] \tag{13}$$

Equations (9), (12), and (13) are the theoretical expressions for the mean, variance, and probability density of Y. Comparisons of the performances of various strategies can be made by studying the percentage errors with respect to the theoretical expressions for each strategy. The error definitions in this study are as follows:

$$\text{error in the mean} = \text{EM} = \frac{\text{theoretical mean} - \text{approximate mean}^1}{\text{theoretical mean}} * 100 \tag{14a}$$

$$\text{error in the variance} = \text{EV} = \frac{\text{theoretical variance} - \text{approximate variance}}{\text{theoretical variance}} * 100 \tag{14b}$$

The theoretical means and variances are calculated from eqns (9) and (12), respectively. The approximate values are obtained by using Monte Carlo simulations under strategies 1, 2, and 3.

This discussion first analyzes the percentage errors in

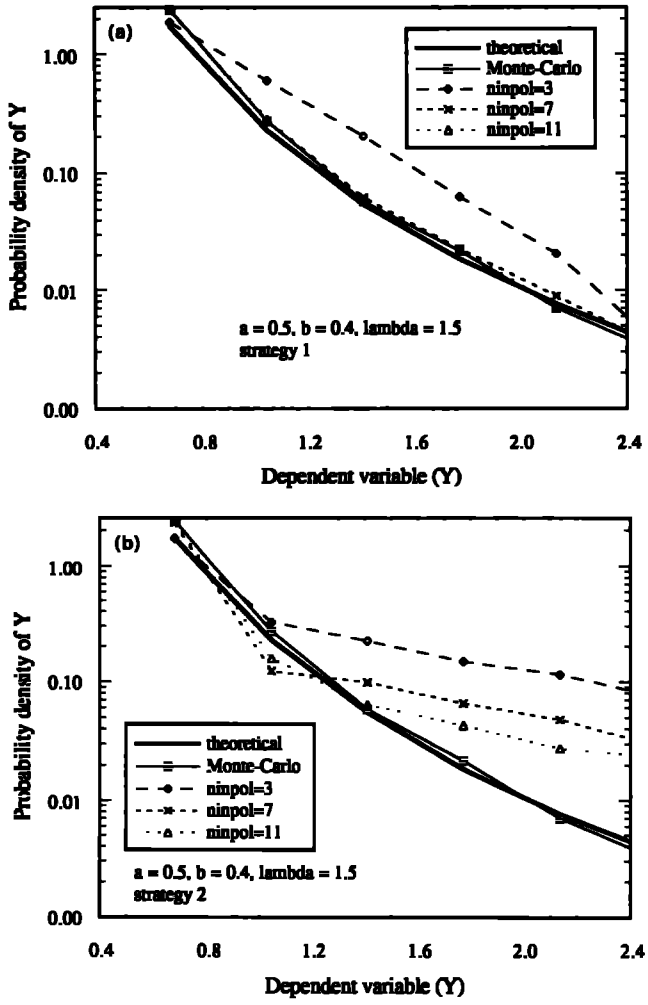

Fig. 2. Comparisons of the theoretical and simulated probability densities by using strategies 1(a) and 2(b). The parameters are the same as in Fig. 1.

the mean and variance. In Figs. 1(a)–(d), the values of the parameters in eqns (7) and (8) are: $a = 0{\cdot}5$, $b = 0{\cdot}40$, and $\lambda = 1{\cdot}50$. In Fig. 1(a), EM is shown as a function of the total number of Monte Carlo simulations. The numbers in the figure legend refer to the particular strategy. The total number of interpolation points (*ninpol*) in Fig. 1(a) is eleven. The EM value from strategy 1 (approximately 1%) is considerably less than that for strategy 2 (about 12%). The errors incurred with strategy 3 (Monte Carlo simulations without interpolation) are the least, as would be expected. With an increase in the number of Monte Carlo trials, the EM appears to settle at 1% for strategy 1 and 12% for strategy 2. Meanwhile, the regular Monte-Carlo-simulation results labelled as strategy 3 show that the EM reduced to zero with an increasing number of simulations. This is because, once the number of interpolation points is fixed for strategies 1 and 2 (11 in this instance), increasing the number of Monte Carlo simulations beyond a certain amount will not reduce the interpolation error.

Figure 1(b) makes comparisons of percentage errors in the variance EV as a function of number of Monte Carlo trials. The parameter values and the number of interpolation points are the same as in Fig. 1(a). The errors resulting from strategies 1 and 3 are almost identical. Strategy 1 leads to a smaller EV than strategy 2. This behavior is similar to that of Fig. 1(a). Increasing the number of Monte Carlo trials leads to a constant value of 10% for strategy 1 and 350% for strategy 2.

Figure 1(c) examines the influence of increasing the number of interpolation points on the EM. Clearly, as the number of interpolation points increases, the interpolation error decreases, and consequently both strategy 1 and strategy 2 show a decrease in mean percentage error. However, it may be noted that strategy 1 shows a smaller error than strategy 2 when the number of interpolation points ranges from three to ten. The difference in error is particularly significant when the number of interpolation points is less than ten. The choice of interpolation points becomes particularly crucial when we decide upon a small number of interpolation points. This choice becomes less significant as the number of interpolation points is increased, so that the interpolation error is practically negligible irrespective of which interpolation strategy is used. As indicated in Fig. 1(c), the mean error obtained from 9000 Monte Carlo simulations is 0·064%. This is the asymptotic value that both strategy 1 and strategy 2 would achieve if the number of interpolation points were increased indefinitely and the interpolation error made negligible in the process.

Figure 1(d) shows the behavior of the EV with an increasing number of interpolation points. The reduction in error with increasing interpolation points is quite dramatic for strategy 1, whereas strategy 2 shows a more gradual reduction. The EV for strategy 3 after 9000 simulations is given as 12.15%. This is the asymptotic value that seems to have been achieved by strategy 1 with ten interpolation points. Figure 1(b) confirms this conclusion. For this example, it may therefore be concluded that strategy 1 would be preferable to strategy 2 because of smaller EV and EM values.

Figures 2(a) and 2(b) compare the predictions for the probability densities for the same parameter values as in Fig. 1(a) (i.e. $a = 0{\cdot}5$, $b = 0{\cdot}4$, and $\lambda = 1{\cdot}5$). Figure 2(a) shows the theoretical density and the simulated densities obtained from 9000 simulations and by choosing three, seven and eleven interpolation points and using strategy 1. Figure 2(b) shows similar results when strategy 2 is used for interpolation. Figures 2(a) and 2(b) are plotted on a semi-log graph for comparison purposes. These figures indicate that strategy 1 has a faster rate of convergence to the theoretical density for a given number of interpolation points. The results for strategy 1 with $ninpol = 7$ in Fig. 2(a) are better than those with $ninpol = 11$ for strategy 2 in Fig. 2(b). These figures demonstrate that the choice of interpolation

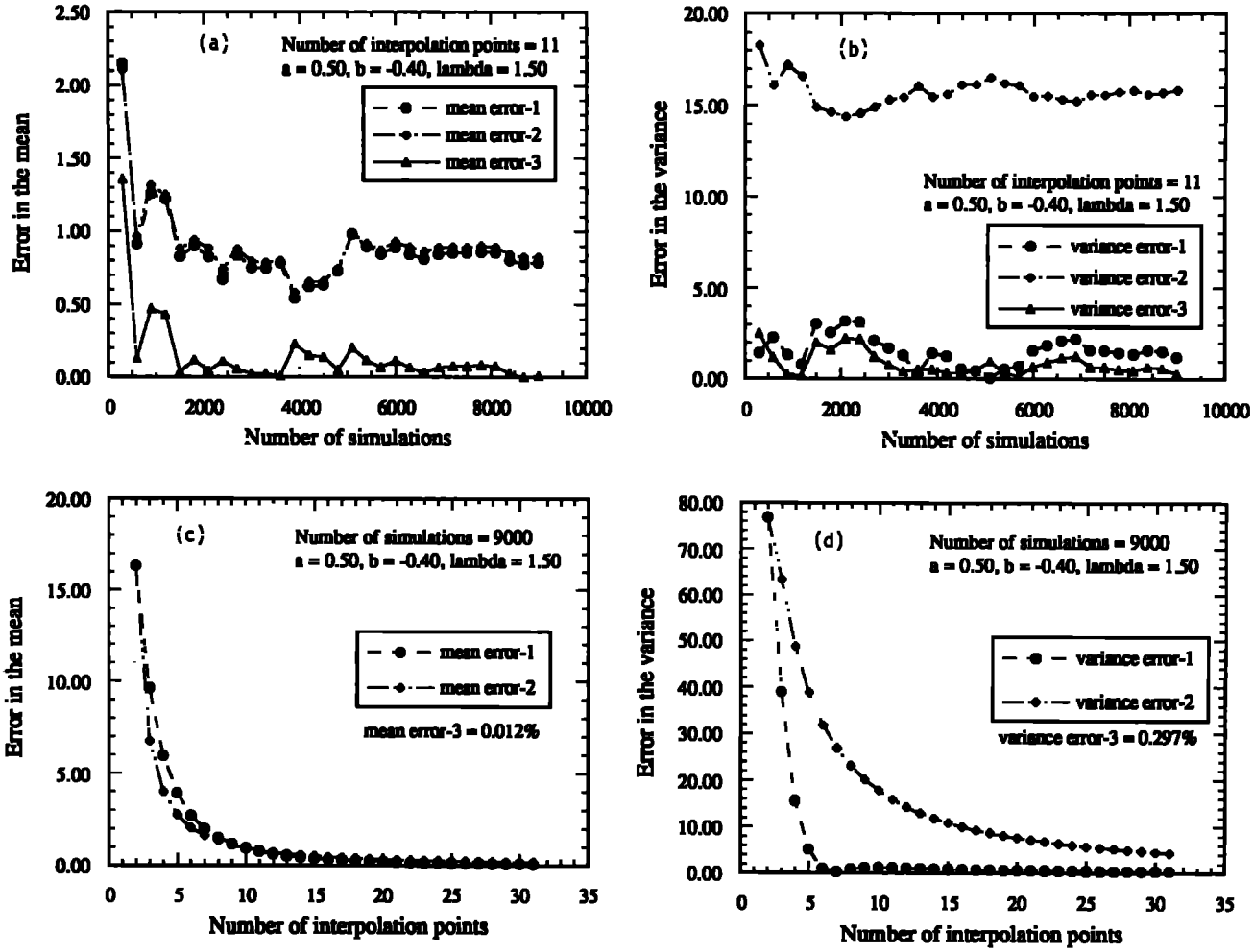

Fig. 3. Comparisons of errors (%) for the mean and variance by using strategies 1, 2, and 3 as a function of the number of Monte Carlo simulations (a) and (b) and number of interpolation points (c) and (d). The numbers in the legends refer to the strategy.

points has a significant influence on statistical predictions under a Monte Carlo framework. Similar conclusions were obtained when the parameter values were changed to $a = 2{\cdot}0$. $b = 0{\cdot}1$, and $\lambda = 0{\cdot}5$ (results not shown).

Figures 3(a)–(d) also compare the mean and variance percentage errors for parameter values of $a = 0{\cdot}50$. $b = -0{\cdot}40$, and $\lambda = 1{\cdot}5$ in eqns (7) and (8). Unlike Figs 1(a)–(d), Figs 3(a)–(d) show differences in behavior for the errors. In Fig. 3(a), the EM values for strategies 1 and 2 are practically identical and eventually settle to about 1%. The error from strategy 3 is much smaller and reduces to zero with an increasing number of simulations (as would be expected). This is in contrast to Fig. 1(a), in which strategy 1 was shown to be distinctly superior to strategy 2. Figure 3(b) shows similar behavior to Fig. 1(b), and strategy 1 leads to a smaller percentage error for the variance than strategy 2. Strategy 1 produces an error pattern that is quite similar to the one obtained from Monte Carlo simulations without any interpolation (i.e. strategy 3). Figure 3(c) shows a different performance when compared with corresponding results for different parameters in Fig. 1(c). The percentage error in the mean decreases with increasing interpolation points. But strategy 2 seems to perform marginally better when the number of interpolation points is less than seven. Both of the strategies approach the asymptotic value of 0·012% with increasing interpolation points. This asymptotic value is indicated by 9000 Monte Carlo simulations carried out without any interpolation (strategy 3). Figure 3(d) compares the EV for an increasing number of interpolation points. Whereas both strategy 1 and strategy 2 approach the asymptotic value of 0·297%, it is clear that strategy 1 is better and reaches this value with a smaller number or interpolation points.

Figures 4(a) and 4(b) compare the simulated and theoretical probability densities for strategies 1 and 2, respectively, for the parameter values of Fig. 3. Figure 4(a) shows that, for a smaller number of interpolation points (*ninpol* = 3), the simulated probability density does not predict the theoretical density well under strategy 1. Some small oscillations are also observed. However, with an increasing number of interpolation point (*ninpol* = 7. 11), the comparisons are quite

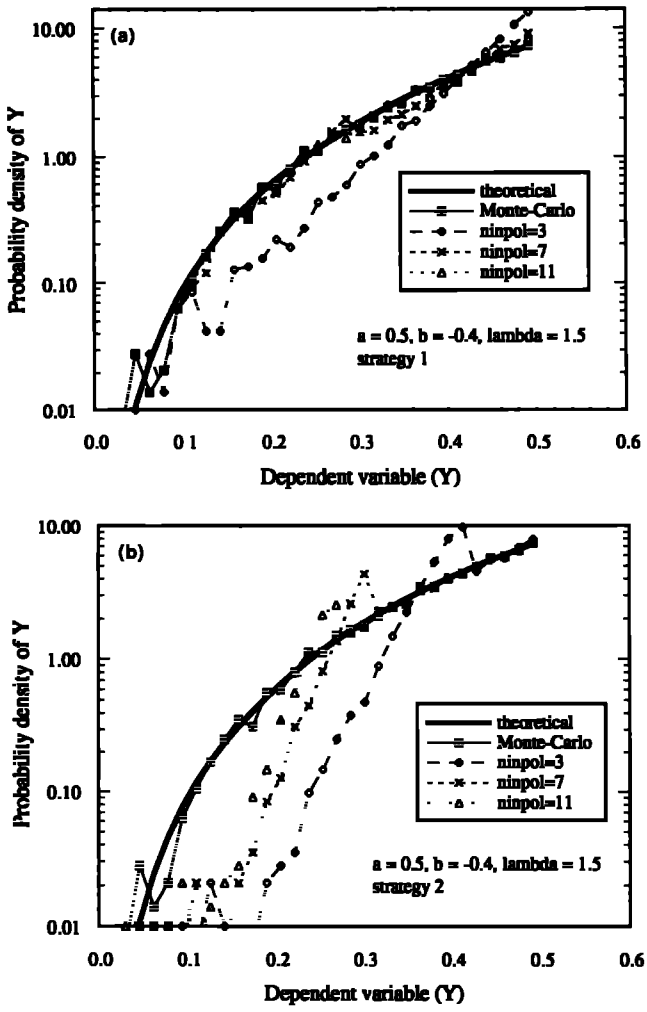

Fig. 4. Comparison of the theoretical and simulated probability densities by using strategies 1(a) and 2(b). The parameters are the same as in Fig. 3.

good. Figure 4(b) shows that the simulated probability-density predictions also improve with increasing *ninpol* for strategy 2. However, the magnitude of the oscillations is much larger in this case and persists even for $ninpol = 11$. For a small number of interpolation points ($ninpol = 3$), the magnitude of these oscillations (measured as deviations from the theoretical solution) is quite large but they are fewer. As the number of interpolation points is increased ($ninpol = 7.11$), the magnitude of these oscillations decreased, but their number increases. Figures 4(a) and 4(b) also include the simulated probability-density function obtained from 9000 Monte Carlo trials without interpolation for comparison purposes. Under strategy 2, the knot sequence is placed equidistantly, and regions of the largest interpolation error are assigned equal weight. Fewer interpolation points therefore lead to fewer oscillations. As the number of interpolation points increases, the relative percentage of realizations falling in the high-error zones increases, which thereby increases the number of oscillations. Simultaneously, there are more realizations with smaller interpolation error, so the magnitude of the oscillations decreases with an increasing number of points in the knot sequence under strategy 2. The conclusions of Figs. 3 and 4 remained unaltered for parameter values of $a = 2{\cdot}0$, $2 = -1{\cdot}0$, and $\lambda = 0{\cdot}5$ (results not shown). Figures 1 and 3 show the performance of the two strategies when the modeling quantities of interest are the mean and the variance, while Figs 2 and 4 show the predictive ability for the two strategies when the probability density is the quantity of modeling interest.

SUMMARY AND CONCLUSIONS

The aim of this paper was to demonstrate that the choice of interpolation points is important in dealing with Monte Carlo simulations. This was done by using an example for which analytical solutions are available. This allowed for a definition of percentage errors for the mean and variance and comparisons with the theoretical probability-density function. In most practical problems of groundwater and solute transport, such analytical solutions would not be available. Two strategies for the selection of interpolation points were considered. Strategy 1 chose the interpolation points based on the underlying probability-density function of the random parameter. Strategy 2 distributes the interpolation points uniformly over the parameter space. For comparison purposed, results obtained from Monte Carlo simulations without interpolation are labelled as strategy 3 in this paper. For the examples considered in this paper, strategy 1 seems to be superior to strategy 2. This was indicated by smaller percentage errors in the mean and variance. Strategy 2 resulted in smaller oscillatory behavior during prediction of the probability density in the second example. The exception to this general trend was encountered in predicting the mean behavior with a small number of interpolation points. Typically, for most problems of practical interest, the number of interpolation points would be ten or more for each random variable in the physical problem. Strategy 1 tends to concentrate the nodal points in regions of higher probability of occurrence. It would not be suitable when one is interested in extreme events that are governed by the behavior of the tails of a distribution.

The examples used in this study were designed to bring out the differences between strategy 1 and strategy 2. Thus the discrepancies in the performance of these two strategies are probably more pronounced than those that would be encountered in most practical groundwater problems. It should also be noted that the strength of the Monte Carlo procedure is not in integration of smooth functions in one dimension. In fact, previous research studies have used numerical-integration

methods in estimating moments (Bresler and Dagan,[5] Govindaraju *et al.*[13]). Nevertheless, this paper demonstrates that the choice of interpolation points is important and warrants careful consideration before undertaking a Monte Carlo procedure with interpolation.

A natural extension of this approach is to cases in which the random property X (say, hydraulic conductivity) is non-homogeneous and even has spatial correlation. The definitions of the moments would change and the form of the probability-density function of X would change. This would lend to a change in the generation procedure for realizations of X. However, the evaluation of the unknown variable $Y = g(X)$ remains unchanged, and interpolation strategy would still be applicable. The number of interpolation points is typically from one to two orders of magnitude smaller than the number of Monte Carlo simulations (see Ahlfeld and Pinder[1] for actual estimates). This would be true even if more than one parameter were random in the physical problem under consideration. However, we should then have to extend this interpolation over multiple dimensions, but the underlying concept remains the same. In any instance, this scheme would reduce the enormous computational burden that is frequently associated with Monte Carlo solutions for problems in stochastic framework.

ACKNOWLEDGEMENT

This paper benefited from the comments of an anonymous reviewer.

REFERENCES

1. Ahlfeld, D.A. & Pinder, G.F., A fast and accurate method for solving subsurface contaminant transport with a single uncertain parameter. *Adv. Water Resour.*, **15** (1992) 143–50.
2. Atkinson, K.E., *An Introduction to Numerical Analysis*. John Wiley, New York, NY, USA, 1978.
3. Bakr, A.A., Gelhar, L.W., Gutjahr, A.L. & MacMillan J.R., Stochastic analysis of spatial variability in subsurface flow. 1: Comparison of one- and three-dimensional flows. *Water Resour. Res.*, **14** (1978) 263–71.
4. Biggar, J.W. & Nielsen D.R., Spatial variability of the leaching characteristics of field soil. *Water Resour. Res.*, **12** (1976) 78–84.
5. Bresler, E. & Dagan, G., Unsaturated flow in spatially variable field soils. 2: Application of water flow models to various fields. *Water Resour. Res.*, **19** (1983) 421–8.
6. Chu, S.-Y. & Sposito, G., (1980), A derivation of the macroscopic solute transport equation for homogeneous saturated porous media. *Water Resour. Res.*, **16** (1980) 542–6.
7. Conte, S.D. & de Boor, C., *Elementary Numerical Analysis: An Algorithmic Approach*. McGraw Hill, New York, NY, USA, 1965.
8. Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities. 1: Conditional simulation and the direct problem. *Water Resour. Res.*, **18** (1982). 813–33.
9. Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities. 2: The solute transport, *Water Resour. Res.*, **18** (1982) 835–48.
10. Dagan, G., Statistical theory of groundwater flow and transport: pore to laboratory, laboratory to formation, and formation to regional scale. *Water Resour. Res.*, **22** (1986) 120s–34s.
11. Freeze, R.A., A stochastic conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resour. Res.*, **11** (1975) 725–41.
12. Govindaraju, R.S. & Kavvas, M.L., Stochastic overland flows. II: Numerical solutions for the evolutionary probability density functions. *Stochastic Hydrol. Hydraul.*, **5** (1991) 105–24.
13. Govindaraju, R.S., Or, D., Kavvas, M.L., Rolston, D.E. & Biggar, J., Error analysis of simplified unsaturated flow models under large uncertainty in hydraulic properties. *Water Resour. Res.*, **28** (1992) 2913–24.
14. Gutjahr, A.L., Gelhar, L.W., Bakr, A.A. & MacMillan, J.R., Stochastic analysis of spatial variability in subsurface flows. 2: Evaluation and application, *Water Resour. Res.*, **14** (1978) 953–9.
15. Hopmans, J.W., Schukking, H. & Torfs, P.J.J., Two-dimensional steady state unsaturated flow in heterogeneous soils with autocorrelated soil hydraulic properties, *Water Resour. Res.*, **24** (1988) 2005–17.
16. Kavvas, M.L., Govindaraju, R.S., Rolston, D.E., Or, D. & Biggar, J., On the stochastic pollution transport equations. In *Heat and Mass Transfer in Porous Media*, ed. M. Quintard & M. Todorovic. Elsevier, Amsterdam, The Netherlands, 1992, pp. 137–42.
17. Lapidus, L. & Pinder, G.F., *Numerical Solution of Partial Differential Equations in Science and Engineering*, John Wiley, New York, NY, USA, 1982.
18. Mizell, S.A., Gutjahr, A.L. and Gelhar, L.W., Stochastic analysis of spatial variability in two-dimensional steady groundwater flow assuming stationary and nonstationary heads. *Water Resour. Res.*, **18** (1982) 1053–67.
19. Nielsen, D.R., Biggar, J.W. & Erh, K.T., Spatial variability of field measured soil-water properties. *Hilgardia*, **42** (1973) 215–60.
20. Serrano, S.E., General solution to random advective-dispersive equations in porous media. I: Stochasticity in sources and in the boundaries. *Stochastic Hydrol. Hydraul.*, **2** (1988) 79–98.
21. Serrano, S.E., General solution to random advective-dispersive equations in porous media. II: Stochasticity in the parameters. *Stochastic Hydrol. Hydraul.*, **2** (1988) 99–112.
22. Smith, L. & Freeze, R.A., Stochastic analysis of steady state ground water flow in a bounded domain. 1: One-dimensional simulations. *Water Resour. Res.*, **15** (1979) 521–8.
23. Smith, L. and Freeze, R.A., Stochastic analysis of steady state groundwater flow in a bounded domain. 2: Two-dimensional simulations. *Water Resour. Res.*, **15** (1979) 1543–59.
24. Sposito, G. & Barry, D.A., On the Dagan model of solute transport in groundwater: foundational aspects. *Water Resour. Res.*, **23** (1987) 1867–75.
25. Viera, S.R., Nielsen, D.R. & Biggar, J.W., Spatial variability of field measured infiltration rate. *Soil Sci. Soc. Am. J.*, **45** (1981) 1040–8.