EFFICIENT UNCERTAINTY ANALYSIS IN A THREE DIMENSIONAL HYDROGEOLOGICAL MODEL OF FLOW AND TRANSPORT

LIODAKIS S.C.[1], KYRIAKIDIS P.C.[1,2] and GAGANIS P.[3]

[1]Department of Geography, University of the Aegean (University Hill, Mytilene, Lesvos, 81100, Greece), [2]Department of Geography, University of California, Santa Barbara (CA 93106-4060, USA), [3]Department of Environmental Studies, University of the Aegean (University Hill, Mytilene, Lesvos, 81100, Greece)
E-mail: stelioslio@geo.aegean.gr

ABSTRACT

In three dimensional hydrogeological investigations involving heterogeneous porous media, three dimensional realizations of saturated hydraulic conductivity are often input to physically-based simulators of flow and transport in a Monte Carlo framework to evaluate the uncertainty in the spatial distribution of solute concentration due to the uncertainty in the spatial distribution of hydraulic conductivity (Gutjahr and Bras, 1993) Realistic uncertainty analysis requires a large number of simulated conductivity realizations, often comprising millions of values. When such conductivity grids are generated via simple random (SR) sampling, uncertainty analysis quickly becomes extremely expensive in terms of both time and computer resources.

A more efficient alternative to SR sampling is Latin hypercube (LH) sampling, a form of stratified random sampling, which often yields a more representative distribution of simulated parameter values (in terms of smaller sampling variability of their statistics) with fewer realizations (Helton and Davis, 2003). The most widely used methods for generating LH samples from a multivariate distribution are those of Iman and Conover, 1982 and Stein, 1987, which do not rely on any Gaussian assumption and can be used for simulation with or without conditioning data; relevant representative applications include the work of Pebesma and Heuvelink, 1999 and Zhang and Pinder, 2003, respectively.

This work compares the performance of LH sampling to SR sampling in the context of a 3D hydrogeological model involving flow and transport. More specifically, 3D lognormal fields of hydraulic conductivity are generated via SR and LH sampling, and are then input to a hydrogeological model to compute the corresponding 3D fields of solute concentration. The sampling methods adopted are evaluated in terms of the reproduction of ensemble statistics of hydraulic conductivity and solute concentration computed from a very large ensemble set generated via SR sampling. The results show that LH sampling is more efficient than SR sampling, in that it can overall reproduce to a similar extent statistics of the conductivity and concentration fields, yet with smaller sampling variability than the latter.

Keywords: Spatial variability, Uncertainty analysis, Monte Carlo, Latin hypercube sampling, 3D stochastic hydrogeology,

1. Introduction

Three dimensional Monte Carlo simulation, or simple random (SR) sampling from a univariate or multivariate probability distribution, is routinely used for uncertainty and sensitivity analysis of model predictions in a wide spectrum of scientific disciplines, such as engineering science, hydrology, and more generally earth sciences, to name but a few. Monte Carlo simulation consists of generating alternative samples (realizations) from the input parameters, evaluating the model response for each of these realizations, and constructing the corresponding distribution of model predictions. In a spatial context, the spatial distribution of 3D geo-referenced variables is typically modelled within a geostatistical framework via a random field; that is, a set of spatially correlated random variables, one per location (Chiles and Delfiner, 1999). In hydrogeological investigations involving flow and transport in heterogeneous porous media, for example, the spatial distribution of saturated hydraulic conductivity is often parameterized in terms of a 3D lognormal random field model. Realizations of such a random field are then used along with physically-based simulators of flow and transport in a Monte Carlo framework for evaluating, for example, the uncertainty in the spatial distribution of 3D solute concentration realizations due to the uncertainty in the spatial distribution of hydraulic conductivity and possibly other relevant variables (Gutjahr and Bras, 1993).

Any realistic uncertainty analysis, however, calls for the availability of a representative distribution of model outputs, and can quickly become expensive in terms of both time and computer resources in the case of complex models (Helton and Davis, 2002). This problem is far more pronounced in earth and environmental science applications, where, in hydrogeology for example, three dimensional grids of hydraulic conductivity values are used along with other parameters to simulate flow and transport in porous media (Gutjahr and Bras, 1993). An efficient alternative to classical Monte Carlo simulation is Latin hypercube (LH) sampling, a form of stratified random sampling, aiming at generating representative samples or realizations from a set of random variables with a given multivariate probability distribution. Here, the term representative implies realizations spanning efficiently the range of possible attribute realizations corresponding to that probability distribution. LH sampling has been shown to lead to model outputs with smaller sampling variability in their statistics than SR sampling for the same number of input simulated realizations; that efficiency, however, decreases the more non-linear that model becomes in the parameters (McKay, Beckman and Conover, 1979; Helton and Davis, 2003; Helton, Johnson, Salaberry and Storlie, 2006).

2. Latin hypercube simulation of random fields

The procedure of Latin hypercube sampling from a multivariate distribution is introduced here for the case of a random field, a collection of geo-referenced random variables, modelling the spatial distribution an attribute, such as hydraulic conductivity.

2.1. LH sampling from random fields

Sampling from a random field amounts to generating S alternative realizations (images in 2D) of the spatial distribution of a geo-referenced variable Y over the study area A. LH sampling from a random field amounts to (a) generating a spatially correlated SR sample from the random field, and (b) transforming that SR sample into a spatially correlated LH sample with marginally (per location) stratified entries.

Stein's method for LH sampling (Stein, 1987) from a random field amounts to transforming a $(S \times M)$ SR sample matrix $\boldsymbol{Y}_R$ into a LH sample matrix $\mathbf{Y}_L = [y_s^L(\mathbf{c}_m), s=1,...,S, m=1,...,M]$ as

$$y_s^L(\mathbf{c}_m) = G^{-1}\left(\frac{r_s(\mathbf{c}_m) - u_s(\mathbf{c}_m)}{S}\right), \quad s=1,\dots,S \tag{1}$$

where $r_s(\boldsymbol{c}_m)$ = rank($y_s(\boldsymbol{c}_m)$) is the rank of the s-th value $y_s(\boldsymbol{c}_m)$ of the SR sample $y(\boldsymbol{c}_m)$ at location with coordinate vector $\boldsymbol{c}_m$, and $G(\,)^{-1}$ is the inverse Gaussian cumulative distribution function (CDF).

In essence, Eq. 2 involves a set of S spatially correlated probability values (the argument of the inverse CDF), which are also marginally (location-wise) stratified. The rank value $r_s(\boldsymbol{c}_m)$ identifies the probability stratum associated with an original value $y_s(\boldsymbol{c}_m)$ of the SR sample at location $\boldsymbol{c}_m$. The addition of a random number $u_s(\boldsymbol{c}_m)$ uniformly distributed in $[0,1]$ furnishes a random probability perturbation within than stratum. Those stratified probability values are then used to derive the corresponding stratified Gaussian quantiles via the inverse CDF; see, Fig. 1.

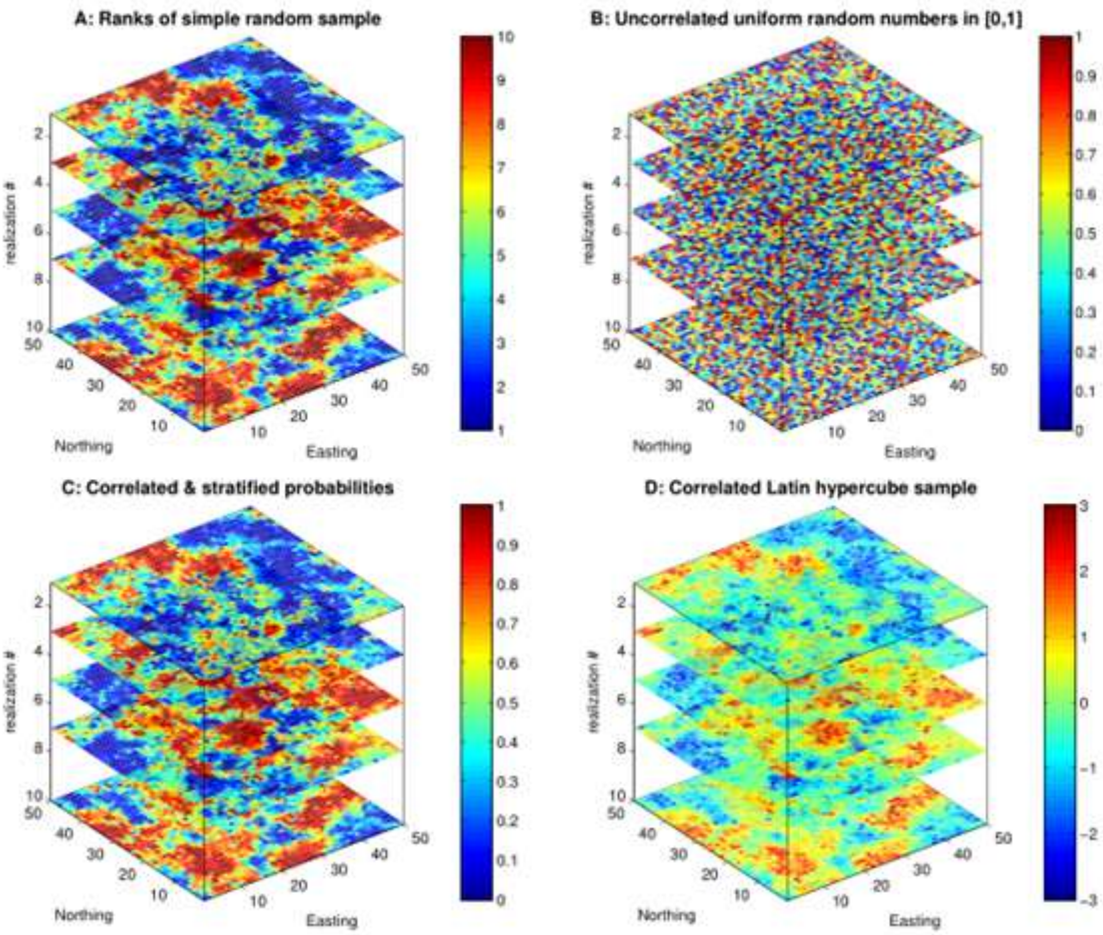

Figure 1: A: Rank ordered version of SR sample, **B:** Uniform random numbers in $[0,1]$, simulated independently at each grid node, **C:** Correlated probability values, stratified at each grid node, derived from **A** and **B**, **D**: Final LH sample of size $S = 10$, whose values are derived as quantiles of a standard Gaussian RV for the stratified probabilities in **C**

3. Hydrogeological case study

This Section presents a synthetic case study comparing simple random and Latin hypercube sampling from a lognormal random field modelling the spatial distribution of saturated hydraulic conductivity in a hydrogeological flow and transport problem.

A two-dimensional synthetic groundwater flow system is considered, similar to that used in Zhang and Pinder (2003). The dimensions of the flow system are $1005m$ by $1005m$ by $125m$ discretized into a $201 \times 201 \times 25$ grid with uniform rectangular cells of size $5m$ by $5m$. Porosity was assumed constant throughout the domain and equal to 0.25. No flow conditions $(\partial h/\partial n = 0)$ were assigned to the rest of the domain boundaries.

In terms of sample size or number of realizations per method, three such sizes are considered; namely, $S = 20,50,$ and 80. Once a sample, say of size $S = 20,$ is generated, the discrepancy between the statistics of the simulated ensemble and the reference statistics* is quantified using the root mean squared error (RMSE). The computation of such error statistic is repeated over a set of $I = 100$ batches of realizations, with each batch containing the same sample size, $S = 20$ for example, thus estimating the sampling distributions of RMSE values for each sample size and for each method; these distributions are presented in terms of their means and medians, as well as their 75% and 95% probability intervals.

For the solute transport problem, an initial concentration equal to 0 is assumed throughout the model domain. At time $t = 0$, a contaminant is introduced at the central cell (middle of 3D domain), along the upstream constant head boundary, with constant concentration $C_0 = 100\, mg/l$.

* Reference ensemble statistics are derived from a set of 10000 realizations of hydraulic conductivity generated via simple random (SR) sampling and the corresponding 10000 solutions of the transport problem – concentration fields. Fig. 2 illustrates a pictorial representation of a 3D simulated conductivity field (left) and the respective 3D concentration realization (right).

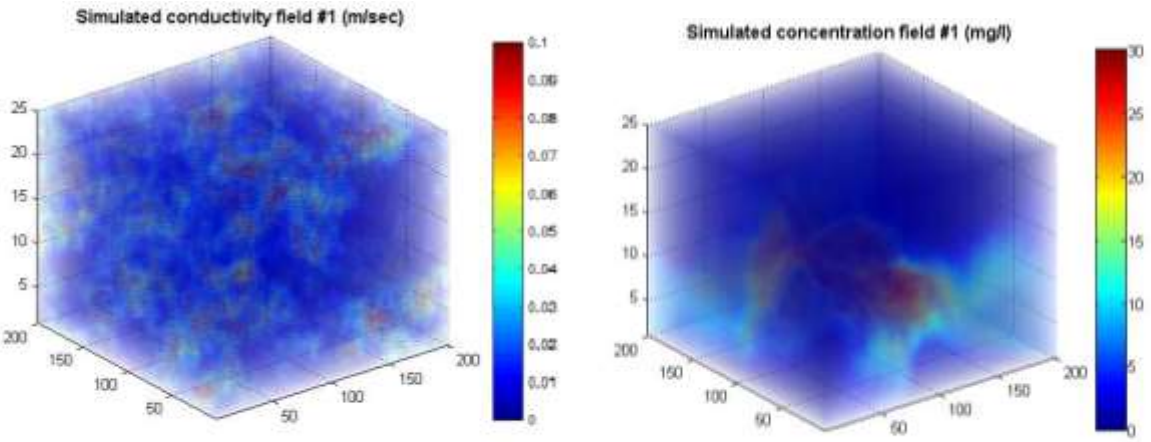

Figure 2: A: Realization of a 3D lognormal random field, **B:** 3D solute concentration realization corresponding to the hydraulic conductivity realization shown in **A**.

4. Results and conclusions

The performance of simple random (SR) and Latin hypercube (LH) sampling, was investigated in a three dimensional synthetic case study involving flow and transport in a porous medium, assuming a lognormal random field with for conductivity with known parameters. Performance comparison for the two methods included reproduction of (a) the reference ensemble average conductivity field, and (b) the reference ensemble average concentration field from concentration realizations obtained by solving a flow and transport boundary value problem for each hydraulic conductivity realization.

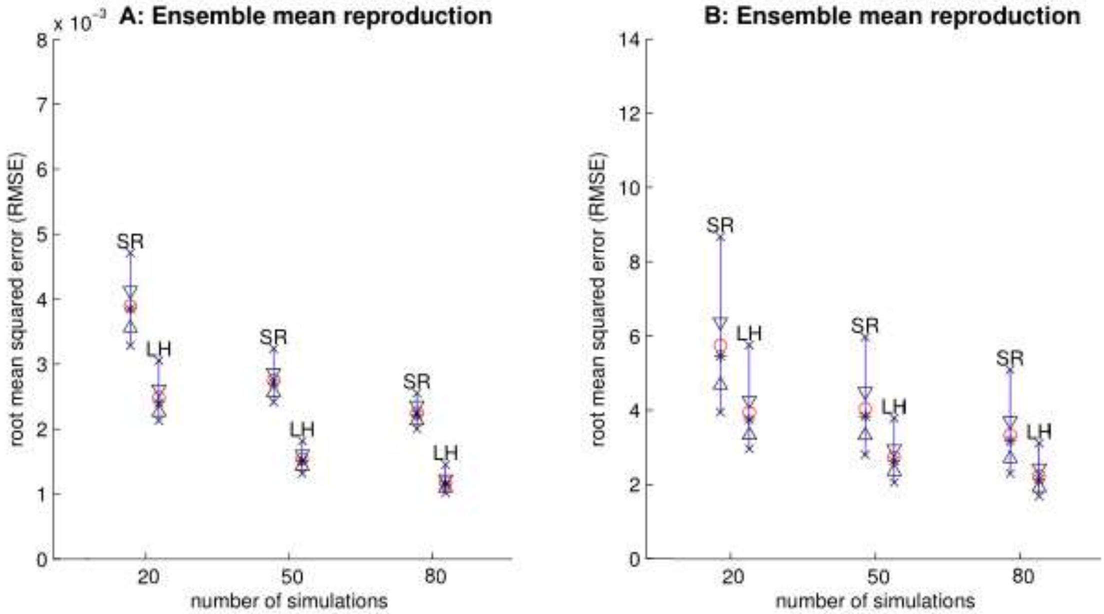

Figure 3: A: Reproduction of hydraulic conductivity reference ensemble mean, **B:** Reproduction of concentration reference ensemble mean.

Figure 3 depicts the reproduction of the ensemble average hydraulic conductivity (left) and concentration (right) fields for the two sampling methods and the three sample sizes considered. It is easily appreciated that LH yields a better reproduction of these ensemble average fields than SR sampling, for both input and output of the model. Concluding, it could be argued that LH sampling illustrates the smaller sampling variability for the same number of simulated realizations than SR sampling. LH sampling could thus lead to efficient uncertainty propagation with fewer model runs due to more representative inputs, thus reducing the time and computer resources for such an endeavor.

REFERENCES

1. Chilès, J. P. and Delfiner, P. (1999), Geostatistics: Modeling Spatial Uncertainty, John Wiley & Sons.
2. Gutjahr A.L. and Bras R.L., (1993), 'Spatial variability in subsurface flow and transport: A review'. Reliability Engineering & System Safety, 42, 293–316.
3. Helton J.C. and Davis F.J. (2003), 'Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems'. Reliability Engineering & System Safety, 81, 23–69.
4. Helton, J. C., Johnson, J. D. Salaberry, C. J. and Storlie, C. B. (2006), 'Survey of sampling based methods for uncertainty and sensitivity analysis', Reliability Engineering and System Safety 91.
5. Iman R.L. and Conover W.J. (1982), 'A distribution-free approach to inducing rank correlation among input variables'. Communications in Statistics, Part B Simulation and Computation, 11, 311—334.
6. 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(2), 239–245.
7. Pebesma E.J. and Heuvelink G.B.M. (1999), 'Latin hypercube sampling of Gaussian random fields'. Technometrics, 41, 303—312.
8. Stein M. (1987), 'Large sample properties of simulations using Latin hypercube sampling'. Technometrics, 29, 143—151.
9. Zhang Y. and Pinder G.F., (2003), 'Latin hypercube lattice sample selection strategy for correlated random hydraulic conductivity fields'. Water Resources Research, 39(8), 1226. DOI: 10.1029/2002WR001822.