

Applied Latin Hypercube stochastic method to quantify the uncertainty in groundwater equation model simulations

Lhoussaine El Mezouary¹, Bouabid El Mansouri²

¹Laboratories of mathematical and analysis, noncommutative geometry and applications, Faculty of Science, Ibn Tofail University, Campus Maamora, BP.133, 1400 Kénitra, Maroc.

²Laboratory of hydroinformatics, Faculty of Sciences, University Ibn Tofail, Campus Maamora, BP. 133 1400 Kenitra, Morocco.

Abstract. It is accepted that digital models simplify the physical reality that is the object of the modeling. Hydrodynamic modeling is an approach with high uncertainties in this context. Indeed, the deterministic modeling approach assumes the existence of a functional relationship between the observed variables. The variables are observed by a series of measurements riddled with errors. Because of this, there is always a significant amount of uncertainty associated with a hydrogeological model. This uncertainty can be associated with the conceptual model or with the data and parameters associated with the different components of the model. Some model parameters such as hydraulic conductivity and recharge are particularly susceptible to uncertainty. Stochastic modeling of the hydrodynamics of a groundwater reservoir is an adequate response to allow us to take a step back on the significance of the results. The study is based on the development of a direct problem-solving model which represents the best estimate of the real hydrodynamic system. This model is used to make predictions. With a stochastic approach, a set of models is constructed where each model, as a whole, is considered to be equally likely. Each model is then used to make the prediction or simulate a given scenario. The MODFLOW-STOCHASTIC-GMS code allows us to do randomization simulations (Latin Hypercube method) and with parameter indicators.

1 Introduction

Latin Hypercube Sampling (LHS) is a frequently used sampling method in the areas of simulation experiment design, computational physics [1, 2], computational chemistry [3, 4], computational biology [5, 6], and other similar disciplines uncertainty analysis, adaptive metamodeling, reliability analysis, and probabilistic load flow analysis [2, 3, 7-9].

The LHS method is preferred by many researchers when it comes to sensitivity analysis, for example, Helton et al [10] used this method to study the propagation of uncertainty in analyses of complex systems, while Manache et al used the LHS method to sensitivity analysis of a water-quality model [11], This method is still preferred in many new researches in various fields [12-14]. LHS has a superior space-filling impact, resilience, and convergence character when compared to other random or stratified sampling methods.

The goal of the LHS extension is to create a larger LHS while keeping the current LHS (or the original LHS). In at least two situations, sampling should be increased, especially for time-consuming simulation systems. Sequential sampling, adaptive metamodeling, and other methods are only a few examples. If the original LHS is subsequently shown to be too small, the second approach is to investigate LHS enlargement, and creating a new LHS with bigger sampling points without the original sample points would be time-intensive. However, due to the LHS structure, expanding the size

of an original LHS while retaining its stratification features is challenging.

One example is the integral multiple extension, in which the new LHS is integral times the size of the original sample. For stratified sampling methods, such as LHS, integral-multiple extension algorithms are used, were suggested by Tong [15]. Sallaberry et al. [16] proposed an LHS with correlated variables two-multiple extension method. Later publications titled "nested Latin hypercube design" [17] and "nested orthogonal array-based Latin hypercube design" [18] described two similar methods. A two-layer nested Latin hypercube design is a Latin hypercube design that contains a subset of a smaller Latin hypercube design. The K-extended LHS technique [19] is a particular integral-multiple extension method in which the new LHS includes K smaller LHSs. Vorechovsky [20-22] published several similar studies in which The new sample size is much larger than the old one. Using integral multiple extension methods allows you to get a more rigorous LHS while retaining all of your original sample points. The objective of this work is the implementation of the method of a method of stochastic resolution of the diffusivity equation based on LHS, we are interested in this work to simulate random fields of the parameter of the hydraulic conductivity and to solve for each given realization, the flow equation by classical methods of numerical resolutions (finite elements, finite differences, ...).

2 Stochastic description

Consequently, as in many fields of Physics, we are led to define macroscopic quantities called local quantities in hydrogeology. Two approaches coexist for the definition of the local properties of a porous medium: the deterministic approach based on the theory of R.E.V. "Representative Elementary Volume" and the stochastic approach. Here we describe the first approach. This simply consists of defining a local (macroscopic) quantity using an appropriate spatial mean on a volume element; this local quantity being associated with the centroid of this elementary volume.

The geological medium can therefore be seen as a continuous medium. It is implicit in this approach that the characteristic length of this elementary volume (v) satisfies the inequalities: $l_p \ll l_v \ll l_d$ where l_d represent the length scale of the geological domain studied.

To be more precise, let us define $V(r)$ as the volume of a spatial domain included in D (the studied geological domain) having for the center, the position vector r and $u(r, t)$ a tensor quantity, vector or scalar in a porous medium. The spatial mean value of u over V at time t is defined by:

$$u(r, t, V(r)) = \frac{1}{V(r)} \int_{V(r)} u(r, t) dr$$

This usually depends on the size, shape, and orientation of $V(r)$ at time t . So that u depends only on r and t , we have to define a volume V_0 between two spheres of volume, This generally depends on the size, shape, and orientation of $V(r)$ at time t . So that u depends only on r and t , we must define a volume V between two spheres of volume $V_{min} = \frac{\pi}{6} l_{min}^3$ and $V_{max} = \frac{\pi}{6} l_{max}^3$.

$$\frac{\partial u(r, t, V(r))}{\partial V_0} = 0, \text{ and } \frac{\pi}{6} l_{min}^3 < V_0 < \frac{\pi}{6} l_{max}^3$$

If we can find for V_0 common limits l_{min} and l_{max} at any point r of the domain D , then can define a field $u(r, t)$ through D and treat D as a continuous medium for the quantity u . The volume V is the representative elementary volume (R.E.V.). In the stochastic approach, any property of the geological medium (such as geometry or physical quantities) is treated as a random function of space. The probabilistic tool works as a great simplifier in the sense that the diversity of the geometry of a physical quantity of a porous medium is reduced to the knowledge of average and higher-order moments. In addition, the stochastic approach was able to justify the differences observed (for example) between the dispersion coefficients measured in the laboratory and those measured during experiments.

De Marsily [23] uses the simple example of porosity at a midpoint, at the laboratory scale. This variable, noted n , takes the value $\varphi = 0$ when the point is located in a grain and $\varphi = 1$ when it is in a pore.

FIG. 1 thus represents an embodiment i of a random process (the distribution of the grains) characteristic of the porous medium. We suppose then that we have a number M of realizations of this medium. At any point P , we have a population of values of the variable porosity φ for which we can calculate the statistical

properties: mathematical expectation, variance, moments of higher order. In the real case, however, we only have a porous medium to study, that is to say, that a single realization of the random processes that characterize it. For example, we obtain the expectation of the porosity at a point P defined by:

$$\varphi_P = \frac{1}{M} \sum_i^M \varphi_i^P$$

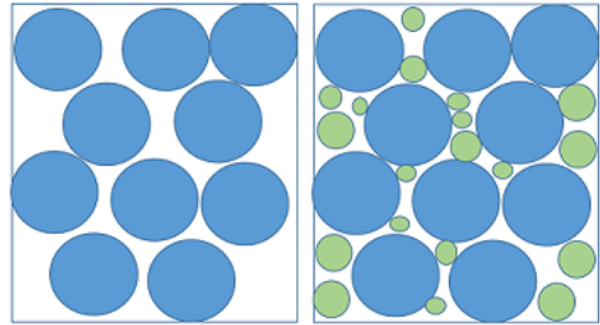


Fig. 1. Example of an implementation of the "grain distribution" variable (high porosity with large spaces and small spaces).

3 Groundwater flow equation, stochastic approach

By considering the properties of the natural formation studied as random spatial functions associated with a statistical structure characterized by its moments (mathematical expectation, variance, and spatial covariance law), the stochastic approach makes it possible to integrate this heterogeneity into the very formalism of considered equations.

The resolution of the equations thus obtained, however, requires the use of approximations and simplifying assumptions leading to analytical solutions which may differ from one author to another (see for example [24-26]). Therefore, these analytical solutions apply in principle only to specific cases (uniform flow, infinite or semi-infinite medium, low heterogeneity in particular). Their validation, therefore, requires recourse to in situ experiments (for example [27]) or to numerical simulations [6, 28-33]. The results of these different works show that these solutions can be applied to cases where the spatial variability of Kh is low. Pursuing the goal of extending the potential field of application of the results from stochastic theory, some authors have focused on solving the flow and transport equations in non-uniform flow cases [32, 34].

We take the scale of the natural formation, which in most cases consists of different horizontal strata of depth varying from a few meters to several hundred meters. On this scale, the properties of the medium, as well as the flow variables, are magnitudes calculated or measured on volumes large enough to ensure that they are macroscopic concerning the scale of the pores. These properties can be considered deterministic and attached to each point in space. The medium and the fluid are assimilated to a continuous medium where one will be able to represent the properties and the macroscopic variables like vectors or continuous spatial tensors.

Consider a permanent flow in the absence of a sink/source term. The equation which governs such a flow is written:

$$\nabla q(x) = -\nabla[K(x)\nabla H(x)] = 0$$

with, $H(x)$: the hydraulic head, $K(x)$: is hydraulic conductivity, and $q(x)$ is the well/source term. We can relate the flow velocity U to the Darcy flow q by:

$$U = \frac{q}{\varphi_e}$$

φ_e is the kinematic porosity. We consider that the flow problem is of direct and unconditional type, that is to say, that we propose to try to identify H and U (through their statistical properties). The stochastic methods of solving this problem are based in particular on the properties of the spatial distribution of hydraulic conductivity. The experimental data compiled in [35] allowed us to deduce that the hydraulic conductivity $K(r)$ is highly variable and should be characterized by a logarithmic distribution density [36]. Freeze's analysis concludes that the distribution that best fitted the data was a lognormal distribution for K ; which means.

$$Y = \ln(K)$$

The flow problem was stated as follows by Dagan [37]: "given the domain W and the properties of the medium in the form of random functions (eg: Y), it is necessary to determine the random spatial function H which satisfies the flow equation and the boundary conditions".

3.1 Latin Hypercube simulation

A stratified Monte Carlo sampling technique is known as Latin Hypercube Sampling (LHS) (MC). The sample region is partitioned in a certain manner by dividing the range of each component of x . Only the case where the components of x are independent or can be transformed to an independent basis will be examined. Sample generation for linked components with a Gaussian distribution is also straightforward. To generate a sample size N from the x variables $x_1, x_2, x_3, \dots, x_4$, LHS operates in the following manner. The range of each variable is partitioned into N non-overlapping intervals based on equal probability size $1/N$. In terms of the probability density in the interval, one value is selected at random from each interval. The N values for x_1 are then randomly matched with the N values for x_2 . These N pairings are combined at random with the N values of x_3 to produce N n -triplets, and so on, until a set of N n -tuples is created. N n -tuples make up the Latin hypercube sample. As a consequence, for given values of N and n , there are $(N!)n_1$ different interval options for an LHS. For three normalized variables, a 10-run LHS using the uniform probability density function (p.d.f.) is shown below (range [0,1]). In this case, the equal probability spaced values are $0, \dots, 0.8, 1$. Performance of the LHSMC considers the case when x is an n -vector random variable for which p.d.f. $f_x(x)$ exists. Let h denote an objective function that is defined like this:

$$h = q(x)$$

Consider now the following class of estimators:

$$T = \frac{1}{N} \sum_{i=1}^N g(H_i)$$

$H_i = q(x_i)$, where $g(\cdot)$ is any known function and is an arbitrary known function T represents an estimate of $[h]$ if $g(h) = h$, i.e. if h is a fixed point for g . The r th sample moment is obtained when $g(h) = H_i$. The empirical distribution function of h at position c may be obtained by selecting $g(h) = u(ch)u(\cdot)$ as a step function.

The approach of the Latin Hypercube then is that the parameter space (all possible combinations of random variable values) should be sampled as completely as possible with a limited number of model scenarios. Once the segments are defined, each parameter is then randomly sampled until the true value is found within each probability segment. The random numbers for each parameter are combined with the random numbers for the other parameters, such that all possible combinations of the segments are sampled. The total number of runs of the model is the product of the number of segments for each parameter.

We consider a heterogeneous aquifer characterized by two zones of distinct hydraulic conductivity HK . If we decompose each zone into three segments. The total number of combinations would be 2^3 equals 8 possible solutions. From which we have framed the total area in the probability curve taking into account the specificity of the maximum and the minimum to have the maximum range of parameters while maintaining the best chance for the stability of the model.

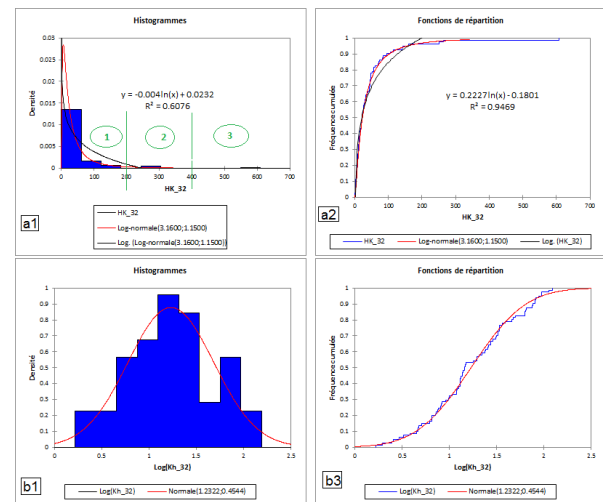


Fig. 2. Latin Hypercube segmentation method of kh parameter according to the **Log – Normal** law with three segments. (Figure a1 and Figure a2): respectively represent the histogram of the probability density function (PDF) and the cumulative distribution function (CDF). (Figure b1 and Figure b2): represent respectively PDF and CDF of **log (kh)**.

Figure 2a and figure 2b show an example of a PDF with parameter kh according to the Log-normal law as well as the corresponding distribution $Log(kh)$. We will also define in the parameterization of LHS a variation interval for kh to keep the values in an acceptable range. The sampling of values of the parameter kh is based on the results of Table 1, specifying the mean, the standard

deviation, and the coefficient of variation. We have chosen to subdivide each distinct area of hydraulic conductivity by three segments as mentioned in figure 2. We then have 8 possible realizations and consequently 8 possible stochastic solutions for the problem some of which have shown a great agreement at the reference hydraulic heads (Figure 3).

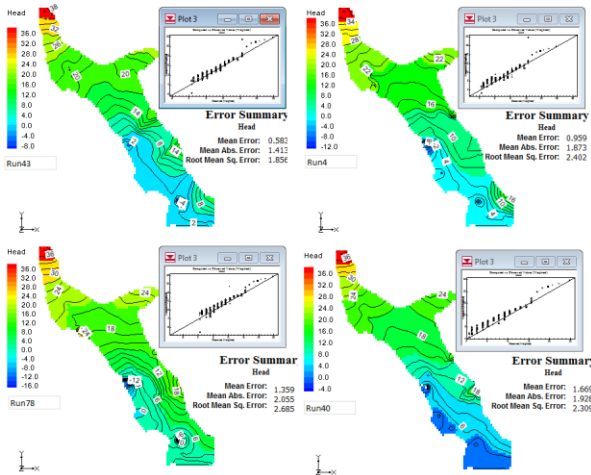


Fig. 3. Illustration of four best simulations based on kh given by LHS

4 Conclusion

This work implements the stochastic model simulating the hydrodynamic behavior of the ground flow equation based on the probabilistic distributions of the permeability parameter. This is made possible through the use of the LHS method and criteria on the choice of the type parameter in the first place (hydraulic conductivity). The use of the LHS distribution model allowed, in the beginning, to focus on the type of model likely to better fit the hydrodynamic parameters calibrated during the direct modeling phase and most vulnerable to estimation uncertainties, and for guided towards the choice of the stochastic solution.

References

1. Fedon, C., et al., Application of deterministic sampling methods for uncertainty quantification in manufacturing tolerances in neutron physics. 2021. 373: p. 111023.
2. Yu, H., et al., A probabilistic load flow calculation method with Latin hypercube sampling. 2009. 33(21): p. 32-35.
3. Helton, J.C., et al., A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. 2005. 89(3): p. 305-330.
4. Scott, S.N., et al., Validation of heat transfer, thermal decomposition, and container pressurization of polyurethane foam using mean value and latin hypercube sampling approaches. 2016. 52(1): p. 121-147.

5. Chaudhary, N., et al., Characterizing the optimal flux space of genome-scale metabolic reconstructions through modified latin-hypercube sampling. 2016. 12(3): p. 994-1005.
6. Lu, Y., et al., Risk analysis of African swine fever in Poland based on spatio-temporal pattern and Latin hypercube sampling, 2014–2017. 2019. 15(1): p. 1-12.
7. Husslage, B.G., et al., Space-filling Latin hypercube designs for computer experiments. 2011. 12(4): p. 611-630.
8. Wang, G.G. and S. Shan, Review of metamodeling techniques in support of engineering design optimization. 2007.
9. Olsson, A., G. Sandberg, and O.J.S.s. Dahlblom, On Latin hypercube sampling for structural reliability analysis. 2003. 25(1): p. 47-68.
10. Helton, J.C., F.J.J.R.E. Davis, and S. Safety, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. 2003. 81(1): p. 23-69.
11. Manache, G., C.S.J.J.o.W.R.P. Melching, and Management, Sensitivity analysis of a water-quality model using Latin hypercube sampling. 2004. 130(3): p. 232-242.
12. Pang, M., et al., Uncertainty and Sensitivity Analysis of Input Conditions in a Large Shallow Lake Based on the Latin Hypercube Sampling and Morris Methods. 2021. 13(13): p. 1861.
13. Saleem, M., A.J.C. Hosoda, and B. Materials, Latin Hypercube Sensitivity Analysis and Non-destructive Test to Evaluate the Pull-out Strength of Steel Anchor Bolts Embedded in Concrete. 2021. 290: p. 123256.
14. Ye, X., et al. Approximate Modeling of Gear Torque of Permanent Magnet Synchronous Motor Based on Improved Latin Hypercube Sampling. in The Proceedings of the 9th Frontier Academic Forum of Electrical Engineering. 2021. Springer.
15. Tong, C.J.R.E. and S. Safety, Refinement strategies for stratified sampling methods. 2006. 91(10-11): p. 1257-1265.
16. Sallaberry, C.J., et al., Extension of Latin hypercube samples with correlated variables. 2008. 93(7): p. 1047-1059.
17. Qian, P.Z.J.J.o.t.A.S.A., Sliced Latin hypercube designs. 2012. 107(497): p. 393-399.
18. He, X. and P.Z.J.B. Qian, Nested orthogonal array-based Latin hypercube designs. 2011. 98(3): p. 721-731.
19. Williamson, D.J.E., Exploratory ensemble designs for environmental models using k-extended Latin Hypercubes. 2015. 26(4): p. 268-283.
20. Vorechovský, M.J.P., Hierarchical subset latin hypercube sampling. 2006. 6: p. 285-98.
21. Vorechovský, M. Extension of sample size in Latin hypercube sampling—methodology and software. in Proceedings of the 3rd Symposium on Life-Cycle and Sustainability of Civil

- Infrastructure Systems, Vienna, Austria, IALCCE, Vienna, BOKU University, Vienna: CRC Press/Balkema. 2012.
22. Vořechovský, M.J.C.A.C. and I. Engineering, Hierarchical refinement of latin hypercube samples. 2015. 30(5): p. 394-411.
 23. De Marsily, G., Quantitative hydrogeology. 1986, Paris School of Mines, Fontainebleau.
 24. Atangana, A. and D.J.J.o.E.M. Baleanu, Caputo-Fabrizio derivative applied to groundwater flow within confined aquifer. 2017. 143(5): p. D4016005.
 25. Tolikas, P.K., E.G. Sidiropoulos, and C.D.J.W.r.r. Tzimopoulos, A simple analytical solution for the Boussinesq one-dimensional groundwater flow equation. 1984. 20(1): p. 24-28.
 26. Neuman, S.P. and Y.K.J.W.R.R. Zhang, A quasi-linear theory of non-Fickian and Fickian subsurface dispersion: 1. Theoretical analysis with application to isotropic media. 1990. 26(5): p. 887-902.
 27. Kitanidis, P.K., Introduction to geostatistics: applications in hydrogeology. 1997: Cambridge university press.
 28. Draoui, Y., et al., Quantification of Surface Water-Groundwater Exchanges by GIS Coupled with Experimental Gauging in an Alluvial Environment.
 29. El Mansouri, B. and L.J.P.o.t.I.A.o.H.S. El Mezouary, Enhancement of groundwater potential by aquifer artificial recharge techniques: an adaptation to climate change. 2015. 366: p. 155-156.
 30. El Mezouary, L., Modélisation mathématique et numérique des phénomènes d'écoulements et de transport de pollutions dans les milieux poreux: application a l'aquifère alluvial de la rivière de Magra, Italie. 2016.
 31. El Mezouary, L., B. El Mansouri, and M. El Bouhaddioui. Groundwater forecasting using a numerical flow model coupled with machine learning model for synthetic time series. in Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, Geo-IT and Water Resources 2020. 2020.
 32. El Mezouary, L., et al., Modélisation numérique de la variation saisonnière de la qualité des eaux souterraines de l'aquifère de Magra, Italie. 2015(2): p. 25-31.
 33. El Mezouary, L., et al. Coupling of numerical flow model with the Susceptibility Index method (SI) to assess the groundwater vulnerability to pollution. in Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, Geo-IT and Water Resources 2020. 2020.
 34. Craig, J.R. and T.J.A.i.w.R. Heidlauf, Coordinate mapping of analytical contaminant transport solutions to non-uniform flow fields. 2009. 32(3): p. 353-360.
 35. Freeze, R.A.J.W.r.r., A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. 1975. 11(5): p. 725-741.
 36. BIR, S., Ecoulements à travers les milieux poreux. Approche stochastique. 2012, Université de Tizi Ouzou-Mouloud Mammeri.
 37. Dagan, G.J.T.i.P.M., Higher-order correction of effective permeability of heterogeneous isotropic formations of lognormal conductivity distribution. 1993. 12(3): p. 279-290.