Uncertain Predictions of Flow and Transport in Random Porous Media: The Implications for Process Planning and Control

Shlomo Orr, MRDS, Inc.

Traditional predictions of flow and transport in porous media are based on mass balance equations in the form of partial differential equations (PDEs), where the flux at every point is defined by Darcy’s law, $\mathbf{q} = -\mathbf{K} \nabla h$, i.e., the flux is proportional to hydraulic head gradient, where $\mathbf{K}$ is the hydraulic conductivity of the medium (a tensor or a scalar; essentially, a material property); it is further assumed that Darcy’s law applies to transient multiphase flow in three dimensions [14,26,36,55]. The solutions of these PDEs constitute groundwater models, oil reservoir simulators, geothermal models, and models of flow and transport in soils/vadose-zone. Due to the similarity between the linear Darcy’s law and Ohm’s law in electricity, Fourier law in heat conduction, and Hooke’s law in elasticity, such models (or PDE solutions) are similar and commonly interchangeable between these fields.

Natural porous formations are heterogeneous, and display spatial variability of their geometric and hydraulic properties. This variability is of irregular and complex nature. It generally defies a precise quantitative description because of insufficient information on all relevant scales [9,18,26,29,30,32,33,86,91]. In practice, only sparse measurements are available (limited by cost of drilling and monitoring). Under lack of exhaustive information, the higher the variability, the higher the uncertainty. Geostatistics is commonly used to analyze and interpolate between measurements in mining and oil explorations, as well as hydrology and soil sciences, using methods such as “kriging”, where the uncertainties in “krigged” values are also quantified [33,35,36,42,56-58,76,77,81,83,84,104,105]. Frequently, these data are collected on different scales that may differ from the required scale of predictions. The task of quantitatively relating measurements and properties on different scales is difficult and intriguing [4,5,7,13,27,29,30,38-40,46,59,78,86,91,101,108,109]. Lack of information in both observed results (output) and measured material properties (parameters) causes uncertain predictions. Spatial variability and uncertainty have lead engineers and geologists to use probabilistic theories that translate the uncertainty to a random space function (RSF) or a random field, consisting of an ensemble of (infinite number of) equally probable “realizations” of parameter values, all having the same spatial statistics, particularly correlation structure [107,76,77,23,33,35,36,42,56,58,83,84,85,87,91]. Imbedded in this approach is a geostatistical model of an assumed joint pdf. In practice, only the first two moments are considered, with an underlying assumption of multivariate normal distribution; in particular, the theoretical semi-variogram (or simply, “variogram”) - the reciprocal of the covariance function, and the mean and variance of the pdf. Since these joint moments are inferred from spatial data, the assumption of ergodicity (i.e., assuming that the ensemble and spatial statistics are identical – a theorem that cannot be proven on real data) must be invoked, which, in turn, implies some kind of stationarity (or statistical homogeneity) [33,35,36,49]. Further, in order to determine the variogram model from available spatial data, an inverse method has to be used to estimate the parameters of this variogram; sensitivity to data errors on one hand, and identifiability problems (of model parameters) on the other hand [81,87,104,105] lead to uncertainty in the geostatistical model itself, which is usually ignored (in fact, the common practice is to fit the variogram model to the experimental variogram by eye and by subjective judgement of model type, degree of stationarity (drift), and statistical anisotropy). Another ignored uncertainty is in the “measured” hydraulic conductivity value that are actually inferred from hydraulic tests interpreted by simplistic models that assume local homogeneity, which is somewhat inconsistent with the RSF approach.

The use of RSF to predict behavior of uncertain systems is not limited to flow in porous media; great efforts have been devoted in all science and engineering fields to (a) estimate or predict mean behavior of the system under different stresses, and (b) compute the uncertainty associated with these predictions (expressed by the variance-covariance of the solution), i.e., the first two statistical moments of system output [4,5,7,10-12,15,16-22,25,28,31-33-35,44,47-51,57,62,63,68-71,73-78,82-93,97,99-101,103,106,107,110,111,112,114]. The resulted approximate solutions are usually limited to simple geometry...
Uncertain Predictions of Flow and Transport in Random Porous Media: The Implications for Process Planning and Control

MRDS, Inc., 5116 Bissonnet, Bellaire, TX, 77401

Proceedings of the 2002 Performance Metrics for Intelligent Systems Workshop (PerMIS -02), Gaithersburg, MD on August 13-15, 2002

Approved for public release; distribution unlimited

14. ABSTRACT

Same as Report (SAR)
and boundary conditions, and to moderate to low variability, as they mostly rely on variations of small-perturbation methods. Recently developed approximations for higher variability in material properties using integro-differential representations have been limited to simple 2D geometry and simple boundary conditions [91,114], with little use, yet.

With respect to mean behavior, it is especially desired to define effective properties of heterogeneous media [1-3,7,8,25,33,37-41,43,45, 52,54,60,61,64-66,71-73,75,77,78,91,93,97,98, 102,108,109,114]. With the two statistical moments of system output, one hopes to optimize and control systems such as oil production, groundwater remediation, irrigation, leaching, etc. However, approximating the statistical behavior of a complex system of flow in random porous media based on the statistics of the hydraulic parameters is a formidable task, at best [46,114, 91-93], because this implies solving the stochastic flow and transport equations (analogous to the Heat/Diffusion Equation) or other stochastic PDEs in other fields (ibid). Since a direct, explicit (closed-formed) solution to the problem of random parameters (or coefficients) is practically impossible\(^1\) [74,93,112,114], only approximate solutions have been reported in the literature for relatively simple cases (ibid,28,31-33,49-50,83- 93,100-101]. Interest in this class of stochastic differential equations has its origins in quantum mechanics, wave propagation, turbulence theory, random eigenvalues, and functional integration [8,20,51,48,62,63,68-71,74,78,99,103,111]. Due to the limited types of problems that can be tackled by stochastic theories (closure approximations), in practice, numerical approximations in the form of high-resolution approximations in the form of high-resolution approximations (e.g., perturbation methods, Neumann series) or by numerical approximations, i.e., Monte Carlo simulations (MCS). Monte Carlo simulations (MCS) are used; however, MCS require ample computer power and CPU time [46,114,88-89] Orr [114] describes and analyzes other difficulties and non-quantifiable uncertainty associated with MCS, particularly the generation of correlated random fields that are faithful to the geostatistical model, and simulations of flow in highly heterogeneous/erratic media.

The geostatistical model provides the statistics of the parameters, particularly the permeability; in MCS, it provides the spatial distribution of parameter values for each realization. Based on these values and assumed model structure (i.e., the conceptual model of flow and transport, including large geologic features, boundary- and initial-conditions, sources and sinks), stochastic solutions are approximated (analytically or numerically). Since the model structure itself is frequently uncertain due to (a) unknown boundary and initial conditions, (b) extent of large-scale geologic features, and (c) information pertaining to geochemical reactions and phase transition, such sets of 500-1000 MCS need to be repeated for several, if not many alternative conceptual models or equally probable model structures [116]. Subsequent optimization requires many repetitions of each Monte Carlo simulation in order to build the search space (i.e., generate sufficient number of scenarios or trajectories); hence, rigorous optimization under uncertainty is prohibitive in terms of computer power and time for most practical applications. In an attempt to optimize best new well placement in an oil reservoir, Guyaguler and Horne [53] were forced to perform optimization on only 23 randomly selected, “history matched” realizations in order to overcome the obstacle of prohibitive computer power and time, while continuously verifying their results against a “truth” model (apparently based on extensive calibration and some “effective” properties). Indeed, the authors concluded that “a decision based on a single realization (though with perfect history match) may differ substantially from the true optimum”[53, p.4]. As was shown theoretically by Neuman and Orr [91], unique effective properties (of random media) that are data independent do not generally exist except for a few special cases. This explains why parameter estimates obtained by traditional inverse methods tend to vary as one modifies the database and/or the imposed stress [113,115]; consequently, calibration of deterministic models may be meaningless in term of predictive power.

Thus, on the way to optimal solutions using stochastic predictions we already encounter

---

\(^1\) A solution to a stochastic PDE consists of specifying the (joint) probability density function (pdf) of the response, \(h(x)\), given those of \(K(x)\) (and forcing functions and boundary conditions). Unfortunately, one cannot obtain the joint cumulative distribution function (CDF) of the random response at all (infinite number of) points. Even for a finite set of points, one cannot obtain closed-form equations for a finite number of moments. This problem can be circumvented by either approximations (e.g., perturbation methods, Neumann series) or by numerical approximations, i.e., Monte Carlo simulations (MCS).
significant theoretical and practical barriers, particularly, uncertain interpretations, model limitations, prohibitive computer power, and non-existing effective parameters – all of which render these predictions highly uncertain, while only part of this uncertainty is being quantified. Our presentation will discuss these barriers, and will bring two examples where traditional stochastic approximations (i.e., approximate solutions of the governing stochastic PDEs) are unable to provide reliable practical solutions. One example deals with the difficult problem of finding the best location of the next well in an oil reservoir [53]). The following simplified block diagram describes the work flow as described by Guyaguler and Horne [53]. Note that “data” are inferred values from field tests. Note also that at each step, there are inherent errors, inconsistencies, and unaccounted (as well as counted) uncertainty. Another example involves optimal control of heap leaching in the mining industry [94,95].

In both cases:
1. Rigorous MCS is already prohibitive in terms of computing resources
2. Models cannot capture the full complexity; hence, predictions are unreliable
3. Rigorous optimization is impossible due to time and computer limitations

In each of these cases, decisions have been made at every step of the solution. Many of these decisions are made subjectively, based on experience, knowledge, thoroughness, understanding (or conceptual models of the process), computer resources, and time limitation, possessed by the modeler. These factors affect and are being affected by the degree of belief (by the modelers) in each decision made.

Initially, sampling (network) design decisions have to be made re sampling locations (an optimization procedure on its own that depends on the end results as well, i.e., a feedback mechanism with the goal of minimizing prediction uncertainty [119]). Then, the following decisions (and sub-decisions) have to be made: (a) type of model and/or curve fitting to use to determine the hydraulic conductivity or permeability (a mini-inverse model that could be ill-posed); in the case of two-phase flow (e.g., oil-water in a reservoir, water-air in a heap), several other decisions (or assumptions) have to be made, and an ill-posed inverse procedure must take place [117-121, and author’s personal experience]; (b) determining and eliminating outliers; (c) determining optimal lag distance for the experimental variogram; (d) determining the pdf, variogram model, and model parameters; particularly, choosing between Gaussian and Indicator models, judging between drift and/or anisotropy, and determining the drift (requires a complex inverse procedure, with typical ill-posed cases; see [36,81,104,105]); (e) determining dimensionality, domain size, and mesh resolution with respect to correlation scales, measurement scales, and property/parameter representation scales; (f) determining conceptual model (or model structure) of flow and transport, including the parameters to be treated as random,
the governing equations (PDEs, including reactive transport and multiphase flow), and uncertain boundary- and initial-conditions, boundary locations, zones of specific character/features (based on geologic and hydraulic information) – i.e., the simulator (these decisions should be made first, and re-evaluated as more information is being analyzed); (g) for MCS: random number generator (RNG; portable or not; RNG type; seed, etc.); (h) random field generator (RFG), including type and RFG parameters (in addition to the variogram model and pdf); (i) number of simulations (should be based on the behavior of point output variance as a function of the number of simulations, i.e., a trial & error procedure which usually is not being done, including in the above two cases), given the limited computer power; (j) number of realizations required for optimization; ideally, all (hundreds of) realizations are used; one compromise may be reducing the number of realizations [53]; another compromise would consider mean system behavior (using the resulting ensemble mean results, and optimize that mean behavior (a major uncertain decision); (k) type of optimization/search algorithms (l) number of simulations (per realization) for constructing a sufficiently dense search state-space; (m) objective function and cost variables. The last three decisions have to be made by all optimization procedures.

In the well placement problem [53], due to limited computer power, a decision was made to use only 23 realizations (while 500-1000 simulations are typically needed to provide meaningful ensemble statistics [114,46]), and a decision was made to calibrate randomly selected random realizations, individually, which contradicts the concept of random fields (or RSF), but may have served a practical purpose (i.e., to find the maximum NPV). Similar prohibitive computer power problem prevails in the heap leaching simulations. While unstable oil-water fronts in the well placement case cannot be captured by the simulator, unstable wetting fronts during heap leaching cannot be captured even by high-resolution two-phase models (like the one used by Orr and Vesselinov [95]). In the latter case, lack of information on essential reactive transport properties, and unmonitored dynamic changes in heap structure (particularly sealing of pore space by clays and erosion products) cannot be determined and modeled. Consequently, the simulators are weak, missing on mean system behavior (or predictions) with uncounted uncertainty, resulting in weak optimization and control.

We see that along the track of approximate solutions and partial optimization of oil reservoirs and heap leaching operations based on predetermined stochastic PDEs, the confidence of modelers and decision-makers is being eroded with each decision being made, depending on their knowledge and degree of belief at each decision point. By the end of this process, decision-makers find themselves with very little confidence and very little decision power. Commonly, in this stochastic approach, the degree of belief is not quantified, though it contributes to the total uncertainty. Alternative fuzzy logic techniques do quantify the uncertainty associated with the degree of belief.

We therefore propose to replace these formidable stochastic approaches by a simpler yet intelligent stochastic control, particularly, the multiresolution decision support system (MRDS, which includes fuzzy logic as one of its components) in order to reach more reliable and efficient optimal solutions, with reduced, accountable uncertainty, in real time, with minimal computer resources. Moreover, unlike the rigid stochastic PDEs, MRDS can be naturally extended to optimal control of linked processes. In the oil field case, this includes exploration, all surface installations and operations, delivery system, and distribution. In the heap leaching case, this includes subsequent solvent extraction and electrowinning, as well as all antecedent processes – from exploration and blasting to transportation, crushing, agglomeration, conveying, placement, and design of the irrigation systems.

REFERENCES

49. Gelhar, L.W., Stochastic subsurface hydrology: From theory to applications, Water Resour. Res., 22(9),


Roberts, P.H., Analytical theory of turbulent diffusion,
Poley, A.D., Effective permeability and dispersion in
Paleologos, E., Effective Hydraulic Conductivity of
Papoulis, A.,
Orr, S. and S.P. Neuman, "Operator and Integro-
Neuman, S. P., Tartakovsky, D. M., Wallstrom, T. C. &
Neuman, S. P., & Orr S., Prediction of steady state flow in
Tartar, L., Remarks on homogenization,
Schetzen, 
Samper, F.J. and S.P. Neuman, Estimation of spatial
Samper, F.J. and S.P. Neuman, Estimation of spatial
covariance structures by adjoint state maximum
Samper, F.J. and S.P. Neuman, Estimation of spatial
covariance structures by adjoint state maximum
Shvidler, M.I., Filtration Flows in Heterogeneous Media
Tartar, L., Remarks on homogenization, Homogenization
and Effective Modules in Materials and Media , edited by J.L. Erickson. D. Kinderlehrer, R. Kohn, and J.L.
Tartar, L., Nonlocal effects induced by homogenization,
in Partial differential equations and the calculus of variations, Vol. 2, Essays in honor of Ennio De Giorgi,
edited by F. Colombini, A. Marino, I. Modica, and S.
Spagnolo, Progress in nonlinear differential equations
Tsokos C.P., and W.J. Padgett, Random Integral
Equations with Applications to Life Sciences and
Engineering, Mathematics in Science and Engineering,
Van Kampen, N. G., Stochastic Processes in Physics and
Vannarcke, E.H., Random Fields: Analysis and
Carrera, J., J. Heredia, S. Vomvoris, and P. Hufschmied,
Fracture flow modelling: Application of automatic
calibration techniques to a small fractured monzonitic
gneiss block, in Hydrogeology of Low Permeability
Environments, edited by S.P. Neuman and I.
Neutriks, Hydrogeology, Selected Papers, Intern.
Assoc. Hydrogeol., Verlag Heinz Heise, Hannover,
Germany, 2, 115-167, 1990.
Orr, S., Stochastic Approach to Steady State Flow in
Neuman, S.P., and A. Guadagnini, A new look at
traditional deterministic flow models and their
 calibration in the context of randomly heterogeneous
media. MODELCARE – 1999.
James, A.L., and C.M. Oldenburg, Linear and Monte
Carlo uncertainty analysis for subsurface contaminant
Carrera, J., Neuman, S.P, Estimation of aquifer
parameter under transient and steady state conditions :
1. Maximum likelihood method incorporating prior
Carrera, J., Neuman, S.P., Estimation of aquifer
parameter under transient and steady state conditions :
2. Uniqueness, stability, and solution algorithms.
Water Resources Research, v. 22, 2, p. 211-227,
1986b.
Carrera, J., Neuman, S.P., Estimation of aquifer
parameter under transient and steady state conditions :
3. Application to synthetic and field data. Water
Carrera, J., State of the art of inverse problem applied to
the flow and solute transport equations. Groundwater
Flow and Quality Modeling, Custodio, E. et al., eds.,
NATA ASI Series, Serie C : Mathematical and
Gh. de Marsily, J.P. Delhomme, A. Coudrain-Ribstein, A.M. Lavenue, Four Decades of Inverse Problems in
Hydrogeology, in celebration of S.P. Neuman’s 60th
Birthday, 30 pp., 1998.