Practical considerations for Monte Carlo simulations: From characterization of complex spatial uncertainty to prediction of groundwater flows

Z. LAKDAWALA[1*] and F. CORNATON[2]

[1,2]FEFLOW Development - DHI-WASY GmbH, Germany zla@dhigroup.com
[*] presenting author

Abstract

Stochastic subsurface hydrology has been an area of intensive research over the past few decades, but there are limited software tools to analyze and model heterogeneous groundwater systems. The focus of this work is to provide an overview of some stochastic approaches to quantify uncertainty using the Monte Carlo approach. The aim is to work towards a module for uncertainty quantification in the existing FEFLOW software, a development of DHI-WASY GmbH. For the scope of this work, the Monte Carlo (MC) technology is discussed which includes generating the stochastic input (using two-point/multiple-point statistics in FEFLOW), transforming the input into the output (using FEFLOW deterministic solvers), and analyzing the output. Each of these will be briefly discussed with practical relevance to solve for flows related to groundwater applications. A few numerical results will be presented for a complex spatially heterogeneous model for the Rhone alluvial aquifer in Switzerland.

1 Introduction

The management of a groundwater system includes decision making related to groundwater quality, the volume of water that may be withdrawn from an aquifer, the placement/location of pumping and recharge wells and/or the rates of pumping/recharge wells. The groundwater quality is compromised with the presence of contamination sites such as hazardous industrial wastes, road salts, leachate from landfills, excessive use of fertilizers for agricultural activities and/or other chemicals. It is desirable to know how the contaminants would spread in the groundwater in such cases not only to maintain groundwater quality but also to incorporate measures to prevent such an occurrence. Predicting groundwater flow or solute transport, like any natural process, is a very difficult task due to their complexity in terms of lacking subsurface data, heterogeneity and associated uncertainty.

Deterministic approaches based on classical theories still dominate groundwater simulations, but with certain limitations. Their main drawback is that they generally do not account for the heterogeneity in subsurface geology. However, the deterministic tools can still help to model the dynamics of the aquifers by incorporating complex sources/sinks, complex and trends in hydraulic properties, irregular boundaries, and transient effects. Advanced groundwater modeling tools, such as FEFLOW, can be used to simulate a multitude of processes involving fluid flow, groundwater age and contaminant under fully or variably saturated conditions.

It only makes sense to think in terms of stochastic groundwater modeling and simulations owing to the uncertainty arising from heterogeneity, incomplete characterization, and sparse and uncertain measurements at a few locations of aquifers. Former work by (Li and McLaughlin, 1991) speak about the incorrectness of defining effective hydraulic conductivity as a property of the aquifer maintaining that large scale non uniformity/heterogeneity is a part of most field applications in reality, suggesting the importance of stochastic approach in modeling groundwater flow and solute transport. Dagan (2002) proposed to treat aquifer properties such as hydraulic conductivity as stochastic variable. In fact, there are many ways in which randomness could potentially enter the problem in the form of initial values, random boundary value, forces and/or input parameters. This mainly extends to treating the aquifer properties as stochastic(random) parameters that influence flow and transport. The randomness reflects the uncertainty of their values, for example, the hydraulic conductivity varies in space by orders of magnitude. The idea is to have several realizations of the parameter field, conditioned to available data and estimate uncertainty via the variability of the responses from the different realizations. In theory, the task of estimating or quantifying uncertainty covers distinct components of fundamental techniques, such as stochastic estimation and simulations for modeling data uncertainty (geostatistics, inverse modeling, reduced order modeling, Bayesian statistics, etc.), and numerical solution of stochastic differential equations and sensitivity analysis. There are many numerical approaches that can be used to study non-stationary and heterogeneous flow problems in complex field conditions. The most commonly used is the Monte Carlo approach, which typically entails generating input parameter fields, using deterministic solvers and performing sensitivity analysis. This approach is practically used due to its simplicity, however they come with high computation costs.

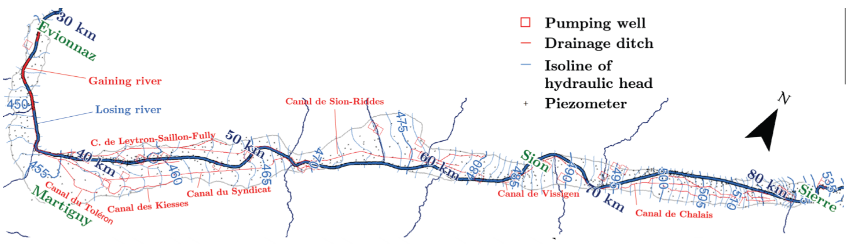


Figure 1: Contour map of groundwater heads during High Waters including major drainage ditches, pumping wells and piezometer locations (Figure taken from Glenz (2013))

As rightly pointed out, by (Renard, 2007), there is lacking software, data sets and computing infrastructure to solve applications related to stochastic subsurface hydrogeology on a real scale despite progress in theories and methodologies. In an attempt to bridge this gap, this is a first attempt towards including Monte Carlo engine for uncertainty quantification in the FEFLOW software. The objective of this paper is to illustrate the ongoing work towards solving groundwater flow through stochastic modelling and numerical simulations using FEFLOW. We use uncertainty quantification, for example, to investigate how the variances of input parameter, such as hydraulic conductivity, contribute to the variances of the observed heads and velocity. This is illustrated with the example of the Rhone alluvial aquifer (as shown in Figure 1). Figure 1 illustrates the contour map of groundwater heads including the major drainage ditches, pumping wells and piezometer locations. The incomplete characterization and sparsely available data is partly the reason why Glenz (2013) did a quantitative evaluation of the impact of the Third Rhone correction on groundwater within the Rhone Alluvial aquifer. Here, we use the same model to account for uncertainty in the steady state modeling of the aquifer flow system using the Monte Carlo approach with FEFLOW. Random generators based on two-point and multi-point statistics for stochastic spatial simulations are typically used to generate the 'space of uncertainty' of an underlying phenomenon. Collectively, the multiple realizations represent the uncertainty of the simulated variable. For the scope of this paper, the conditioned stochastic realizations of hydraulic conductivity are generated using the Turning Bands method (TBM). The random parameter fields are conditioned to observed data via residual based kriging methods ((Emery and Lantuejoul, 2006)). The random generator relying on the TBM is included as a routine in FEFLOW for unstructured computational grids as one of the methods to account for the two-point variogram-based spatial variability in input parameter towards predicting groundwater flow and solute transport. There is also ongoing work on generating the space of uncertainty using Direct Sampling Technique ((Mariethoz et al., 2010)) and using the Distance Kernel methods (discussed in (Scheidt and Caers, 2009)) to reduce the dimensionality of the space of input parameters thereby reducing the computational costs significantly. The deterministic solvers of FEFLOW are then used as transfer functions to generate a response based on the space of uncertainty. Lastly, a sensitivity analysis is done by illustrating the variance plots of the hydraulic conductivity and hydraulic heads respectively, to quantify the variability of the ensemble of solutions based on the initial space of uncertainty.

2 Problem Statement

Here, we consider the steady state flow within a confined/unconfined aquifer. The flow in the aquifer is governed by the Darcy equations coupled with water balance equation:

$$\mathbf{u}(\omega, \mathbf{x}) = -K(\omega, \mathbf{x})\nabla h(\omega, \mathbf{x}) \tag{1}$$

$$\nabla \cdot \mathbf{u}(\omega, \mathbf{x}) = f(\mathbf{x}) \tag{2}$$

Here $\mathbf{u}$, h and K denotes the Darcy velocity, hydraulic head and hydraulic conductivity(permeability divided by water viscosity) respectively.

The system of equations is equipped with appropriate boundary conditions, depending on the problem under consideration. The boundary conditions usually take the form of a constant head value/head dependent flux/a known flux (to model recharge, evaporation, pumping wells, stream discharge etc.), and/or no-flow conditions.

The solution depends strongly on the hydraulic conductivity fields that are highly variable and never perfectly known. Uncertainty crops into these parameters by describing them as as random fields, i.e. the hydraulic conductivity $K(\omega_i, x_i)>0$ in each point x_i is a random variable ω_i, where $i = 1, ..., n$. n is taken to be the total number of points/elements in the computational grid and $x_1, ..., x_n$, the random variables $K(\omega, x_1), ..., K(\omega, x_n)$ are in general spatially (and temporally) correlated. The hydraulic conductivity is generated by a random generator using TBM (more details in section 3. The entire set of realizations for hydraulic conductivity contribute to the input space of uncertainty. Large number of realizations are typically required to account for uncertainty and for real-scale problems, computational overheads are essential to deal with.

3 Methodology

Although the deterministic modeling approach is traditionally used, it is unable to incorporate the heterogeneity and quantify uncertainty. For the stochastic modeling approach, a random field generator using TBM is employed to get a set of realizations of the spatially correlated conductivity field with a stochastic set of underlying variograms (equivalent to sets of variances) are simulated. Using the FEFLOW deterministic solvers, we investigate how the variances in hydraulic conductivity contribute to the variances of the observed heads and velocities.

3.1 Generating input parameter fields for Monte Carlo simulations

For generating stochastic realizations, the field of geostatistics covers various ways by focusing on spatial or spatio-temporal data sets. It is usually desired to have an initial estimate of model parameter uncertainties, typically consisting of probability density functions, and preferably consisting of multivariate normal or log-normal probability density functions definable by mean vectors and covariance matrices. Most often, the parameter fields are represented by multi-Gaussian random fields. Such a representation allows finding important relations and understanding the impact of uncertainty, which is essential for a theoretical perspective. Mostly, the efforts are devoted to finding the most accurate and most general relations between the various parameters appearing in the stochastic partial differential equations. For the Gaussian fields, there are various two point statistical methods, both exact and approximate. The methods include a long list of convolution methods, LU decomposition methods, spectral approaches based on discrete, continuous and fast Fourier techniques, variants of sequential Gaussian algorithms, and the turning bands method. However, such techniques use the oversimplified two-point statistics (in the form of variograms) to represent the geological phenomena which, in reality, may have complex geometrical configurations.

In order to locate certain connectivity/structures, such as curvilinear channels, in the complex geometrical configurations, there is little interest in the Gaussian family of fields. The limitation of two point statistical approach to capture structural patterns despite all its advantages turns one towards considering alternative approaches, such as the multi-point and the truncated pluri-Gaussian method. These are facies modeling technique based on multiple point statistics (MPS) instead of the conventional variograms models built on two point statistics. The methods commonly include simulating the facies and then simulating the property values within the facies. MPS framework appears to be a popular choice in the recent years. It relies on a training image (e.g. a conceptual geological model or a given pattern). From this image, multi-point statistics are calculated and used for simulation. The advantage of the MPS as compared to traditional variograms or transition probabilities is that they integrate the possibility of modeling complex spatial relations between the facies. It can be said that two point statistical approaches are unable to account for connectivity (but controls the flow and transport reasonably well). The MPS approach is one of the techniques that further equips in reproducing a global connectivity. One of the main challenges here is to acquire the correct representative training image including the 3D complexity and heterogeneity at the right scale. These can, however, be generated by expert knowledge (drawing them by hand, or a conceptual model), digitizing existing maps, using model output etc. In a 3D configuration, fitting a pattern to a multidimensional training image becomes more challenging.

Towards including MPS-based methods into FEFLOW, it is worthwhile to mention that there is ongoing work in interfacing the direct sampling technique by (Mariethoz et al., 2010)with FEFLOW. This, however, will be the subject of another upcoming publication.

3.1.1 Turning bands method to generate Gaussian, spatially correlated grids on unstructured grids

The turning bands method, as described in (Emery and Lantuejoul, 2006), is implemented in the FEFLOW software for structured and unstructured grids. It is a multidimensional random number generator for the simulation of spatially correlated random fields. If measurement data is available, the fields can be conditioned using simple or ordinary kriging techniques, making sure that the input parameter honors the data from field measurements.

3.2 Boosting the efficiency of Monte Carlo simulations

Due to the heterogeneity and incomplete characterizations, a large number of realizations (from TPS or MPS) are needed to incorporate uncertainty into the model, which result in high computational costs. Parallel computing is particularly helpful for cost reduction of Monte Carlo based simulations, but even with the modern day computers, Monte Carlo simulations are limited in its usage. Different strategies have been discussed to overcome this issue, mainly to reduce the number of realizations and/or reduce the computational efforts for deterministic solvers. In the last decade, the works by Dagan (2002) (Scheidt and Caers, 2009) have addressed the issue of reducing the dimensionality of the stochastic space and finding and optimal subset of realizations to estimate uncertainty. Ranking methods based on measures of connectivity by (McLennan and Deutsch, 2005), Karhuenen Loewe Expansions, (Kernel-) Principal Component Analysis by Sarma et al. (2008), distance kernel method (DKM) by (Scheidt and Caers, 2009) are a some commonly used ways of reducing computational efforts by smartly selecting fewer realizations in a multidimensional space, such that the reduced space of uncertainty also spans the original space of uncertainty. DKM has the added advantage that the selection is based on the flow responses obtained by an approximate model that are not only computationally cheaper but also result in an unbiased estimate of uncertainty.

For real-scale simulations, it only makes sense to do a cluster analysis/dimensionality reduction for improving efficiency. Towards this end, there is ongoing work to employ DKM by (Scheidt and Caers, 2009) as means of model reduction into FEFLOW.

Another strategy to overcome the computational costs is to try and reduce the cost of computing the responses/solutions. Many suggestions include numerical upscaling techniques where hydraulic conductivity is up-scaled on a coarser grid Durlofsky (2005), (Renard and de Marsily, 1997) and the fine scale model is replaced by approximate models that are less computationally expensive. Multi-scale approaches allow better representation of the hydraulic conductivity on the fine scale((Hou and Wu, 1997), (Jenny and Lunati, 2009)). These methods have been studied in a deterministic context where iterative schemes are developed to reduced the error of the solution between fine and coarse scales. In the stochastic context, the works of (Chen and Durlofsky, 2008), Chen et al. (2011) discuss the representation of variability of the large number of solutions, thereby suggesting that multi-scale approaches are well suited as inexpensive deterministic solvers (response functions) in the stochastic context. In a recent publication, (Josset and Lunati, 2013) employs a subset of realizations using DKM and construct an error model to correct the potential bias of multi-scale finite volume estimate. The work indicates that the use of ranking methods such as DKM can be used together with multi-scale approaches to significantly reduce costs of estimating uncertainty using MC methods.

3.3 FEFLOW deterministic solvers as transfer functions

Each realization from TBM (or any random generator) generates an output sample space using a transfer function. Many realizations are required to get an accurate representation of the input sample space and accurate statistics of output sample space. Variants of Monte Carlo techniques have overcome many limitations relating to multiple scales and the large number of simulations to obtain reliable statistics. These come in the form of multi-scale and multi-level Monte Carlo methods. Parallel computing is particularly helpful for Monte Carlo based simulations. As already emphasized in Renard (2007), an all-encompassing software tool for predicting uncertainty in modeling ground water flow and transport from a practitioner's perspective could be very useful. Towards this end, a Monte Carlo engine is developed for FEFLOW. The deterministic solvers in FEFLOW for solving flow (and contaminant transport) can be used to solve the groundwater problem using the parameter fields generated (as described in 3.1). The FEFLOW solvers play the role of transfer functions translating the input space of uncertainty to output response maps (in the form of hydraulic head, flow, and/or contaminant).

3.4 Quantifying uncertainty in the response

The ultimate goal of uncertainty analysis is to predict and to understand the factors that affect, for example, the residual error of the estimation. Obviously, both input parameter variations and errors in head measurements are mostly responsible for a non-zero estimated residual head. Hence, the variances of the head that are obtained from the stochastic conductivity fields could be used to determine the intrinsic errors of the head measurements. A simple form of sensitivity analysis in the form of computing and visualizing first and second moments can be done using FEFLOW.

4 Numerical results

4.1 Realizing the Rhone Alluvial Aquifer model in FEFLOW

The methodology described above is applied to the Rhone Alluvial Aquifer (as shown in Figure 1) to solve a steady state 2D groundwater flow in heterogeneous isotropic unconfined aquifer. The virtual representation of the aquifer is realized using the FEFLOW software for groundwater flow and contaminant transport modeling. Figure 2 illustrates a part of the aquifer with an adapted mesh respecting the surface waters and pumping wells. The entire 2D model consists of 54974 mesh elements. The flux across the boundary (as shown in Figure 3 (Left)) between the

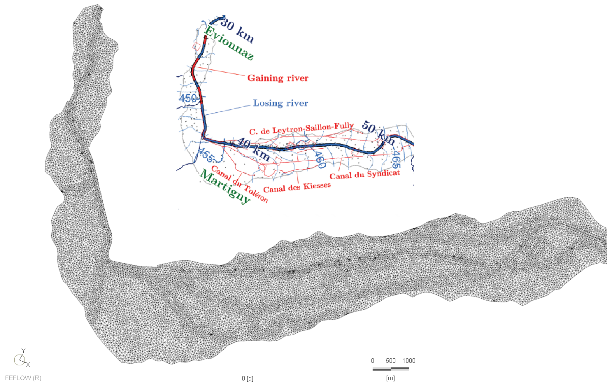


Figure 2: Illustration of the finite element mesh for the Rhone aquifer near Evionnaz and Martigny region.

streambed and the aquifer is calculated from the head loss between the specified surface water level and the model-calculated head in the aquifer. Storage is negligible and the transfers are asuumed to be occurring instantaneously. The Rhone river to the underlying alluvial aquifer is considered as a surface water body to consider Cauchy/mixed type boundary condition in the model. This type of boundary condition is defined by the water level of the surface water and a transfer coefficient related to the hydraulic conductivity divided by its thickness as shown in Figure 3 (middle). Additionally, well boundary conditions are specified at the pumping wells (marked in red) , along with constant hydraulic head value imposed at one location (marked in purple), as seen in Figure 3(right). The remaining unmarked boundaries assume no-flow condition. For more details on the exact boundary condition values, see Glenz (2013).

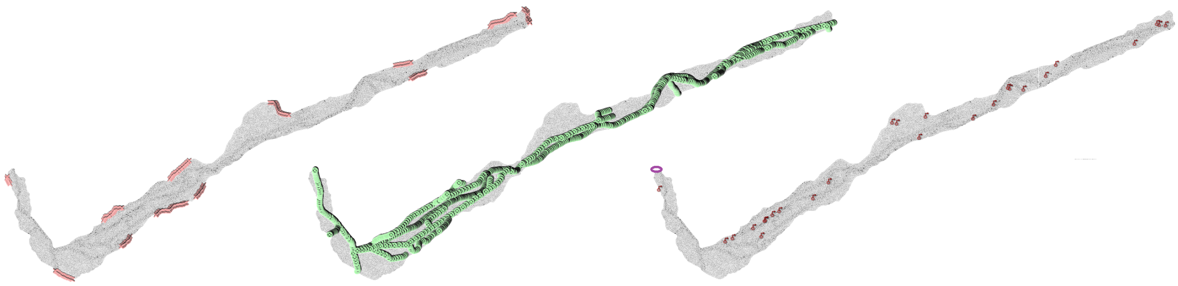
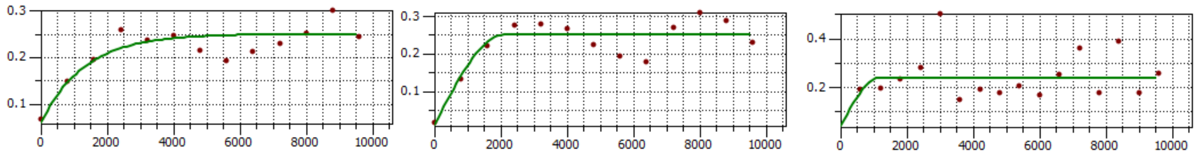


Figure 3: Rhone Aquifer model with marked boundary realized with FEFLOW.

The Turning Bands Method is used to generate hydraulic conductivity fields for the Monte Carlo simulations. The generated realizations not only honor the prescribed statistics(in the form of a variogram) but also honor the data (that is usually inferred or measured by the engineers) at the prescribed locations. For the Rhone aquifer case, Glenz (2013) estimated the values for hydraulic conductivity of the aquifer on 375 pilot points using PEST software. The pilot points are uniformly distributed along the model (this is onsidered as an idealized scenario), and are taken to be the conditioning data set. This workd as a good benchmark to test the routines implemented in FEFLOW. Different spatial correlations are described via different variogram models that are constructed using the conductivity values for the 375 points, 188 points and 94 points (as shown in Figure 4). Various models are fitted, such that each



Conditioning data	Type	Range	Sill	Nugget
375	Exponential	1333	0.25	0.06
188	Spherical	700	0.25	0.01
94	Exponential	400	0.075	0.04

Figure 4: Three sets of variograms (using 3 different conditioning sets with 375, 188 and 94 points) are used to define the uncertainty in spatial correlations.

variogram is identified by type, sill, range and nugget coefficients as shown in Figure 4.

The TBM random generator is used to generated 100 realizations for each variogram. In Figure 5, a single realization for each spatial correlation is shown. It can be seen that the hydraulic conductivity has a more defining spatial correlation for 375 points as opposed to the field generated using 94 points. The ensemble of 300 realizations define our space of uncertainty

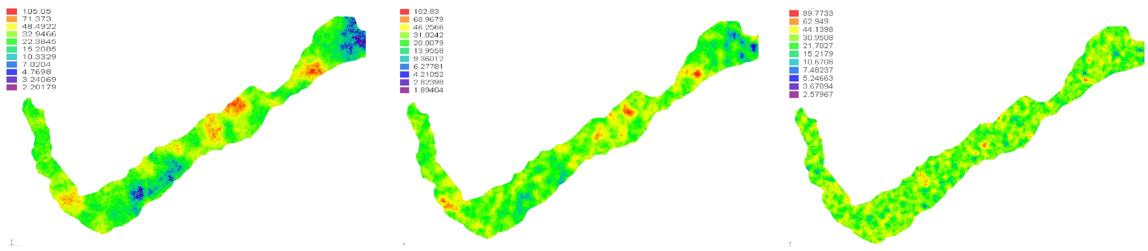


Figure 5: Random hydraulic conductivity for different spatial correlations (i.e conditioning sets).

for the hydraulic conductivity and are provided as input to the deterministic solvers of FEFLOW to generate the hydraulic head values. In Figure 6, one sees the variance plots for a single realizatzion for each of the spatial correlations. It can be seen that the variance is lowest at the conditioning points, and highest in the absence of conditioning points (more room for uncertainty where the values are not restricted to prescribed data). Also, the variance is higher when less points are considered in the conditioning set.

The deterministic solver in FEFLOW is used for obtaining the flow and hydraulic head profiles in the model. Each solution is tied to a realization generated in the previous step. Figure 7 show the mean(over 300 realizations) of hydraulic conductivity, head and flow distributions in the aquifer.

Lastly, the variances for head and flow profiles are obtained from the ensemble of 300 realizations of hydraulic conductivity. Figure 8 illustrates how the input uncertainty in hydraulic conductivity affects the simulation results of hydraulic head and flow velocity throughout the model domain.

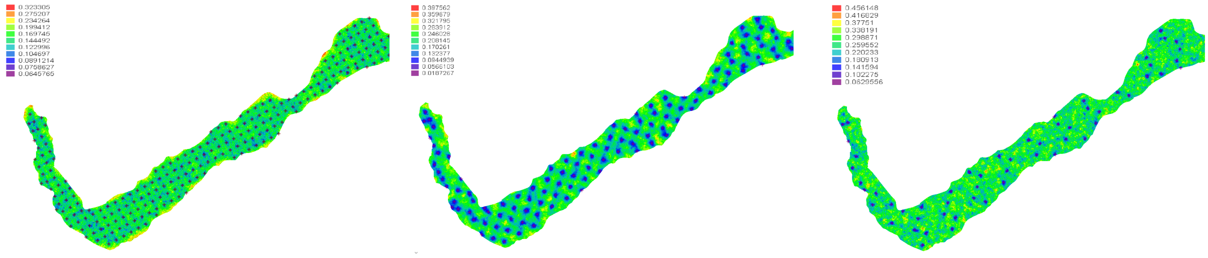


Figure 6: Second order moments for different spatial correlations (i.e conditioning sets).

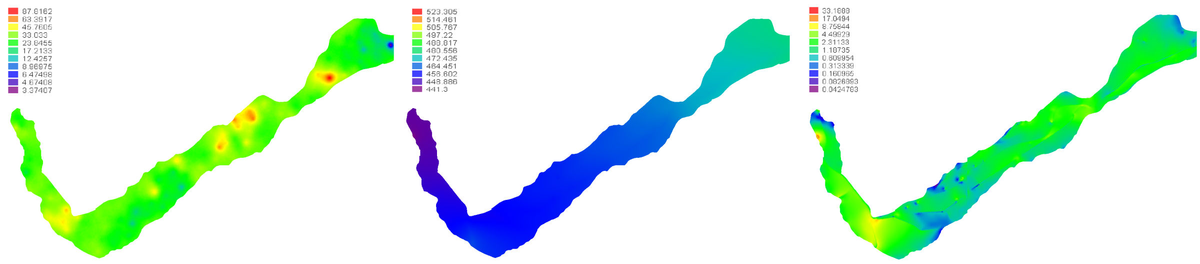


Figure 7: First order moments for 300 realizations of hydraulic conductivity, head and flow velocity.

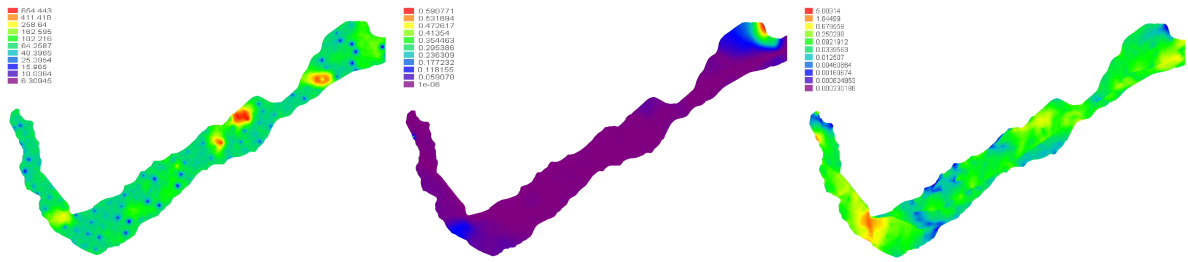


Figure 8: Second order moments for 300 realizations of hydraulic conductivity, head and flow velocity.

5 Conclusion

In order to quantify uncertainty, the following aspects have been explored. The random generator for generating the input space of uncertainty uses the Turning Band method. The spatial correlation is defined by the fitted variograms on three different conditioning data sets. The stochastic results from TBM are considered as input for uncertainty quantification of hydraulic heads using the deterministic solver in FEFLOW. For more practical relevance and to avoid upscaling to different discretization grids, the algorithm is modified to handle arbitrary shaped domains. The method is validated against spherical and exponential covariance models, however the results have not been presented here. Using the realizations from the TBM random generator, the Monte Carlo engine in FEFLOW is used to demonstrate how to generate mean and variance plots of solution. The results can be further used to indicate additional uncertainties in the fitted model by comparing the predicted variances against the variances of the computed head/velocity values.

Moreover, there is ongoing work towards providing an interface for using the Direct sampling technique by Mariethoz et al. (2010), reducing the dimension of the input space of uncertainty using the Distance Kernel methods by Scheidt and Caers (2009) and further parallelization of the Monte-Carlo engine in FEFLOW.

Acknowledgments

The authors would like to thank Europaeische Fonds for regionale Entwicklung(EFRE) and Investitions Bank Berlin (IBB) for their support.

References

Chen, Y., K. Park, and L.J. Durlofsky (2011). Statistical assignment of upscaled flow functions for an ensemble of geological models. *Journal Computers and Geosciences 15*(1), 35–51.

Chen, Y and Durlofsky, L.J. (2008). Ensemble-level upscaling for efficient estimation of fine-scale production statistics, *SPE 13*(4), 400–411.

Dagan, G. (2002). An Overview of stochastic modeling of groundwater flow and transport: From theory to applications. *Eos, Transactions American Geophysical Union 83(53)*, 621–625.

Durlofsky, L. (2005). Upscaling and gridding of fine scale geological models for flow simulation. *8th international forum on reservoir simulation*, Iles Borromees, Stresa, Italy, 20–24.

Emery, X. and Lantuejoul, C. (2006). TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via turning bands method, *Journal Computers and Geosciences* 32(10), 1615–1628.

Glenz, D. (2013). Inverse Modeling of groundwater flow in the Rhone alluvial aquifer- impact of the Third Rhone correction. PhD Thesis, University of Neuchatel.

Hou, T.Y. and Wu, X.H.(1997). A multiscale finite element method for elliptic problems in composite materials and porous media, *Journal of Computational Physics 134*(1), 169–189.

Jenny, P. and Lunati, I.(2009). Modeling complex wells with the multi-scale finite volume method, *Journal of Computational Physics 228*(3), 687–702.

Josset, L. and Lunati, I. (2013). Local and Global Error Models to Improve Uncertainty, *Mathematical Geosciences 45(5)*, 601–620.

Li, S.G. and McLaughlin, D. (1991). A Nonstationary Spectral Method for Solving Stochastic Groundwater Problems: Unconditional Analysis, *Water Resources Research 27*(7), 1589–1605.

Mariethoz,G. P. Renard, J. Straubhaar (2010). The Direct Sampling method to perform multiple point geostatistical simulations. *Water Resources Research* 46(11), W11536.

McLennan, J. and Deutsch, C.V. (2005). Ranking geostatistical realizations by measures of connectivity, *SPE: international thermal operations and heavy oil symposium,* Alberta, Canada, p 98168.

Renard, P. and de Marsily, G. (1997). Calculating equivalent permeability: a review, *Advances in Water Resources 20*(5), 253–278.

Renard, P. (2007). Stochastic hydrogeology: what professionals really need? *Ground Water 45*(5), 531–541.

Sarma, P., L.J. Durlofsky, and K. Aziz (2008). Kernel Principal Component Analysis for Efficient, Differentiable Parameterization of Multipoint Geostatistics. *Mathematical Geosciences 40*(1), 3–32.

Scheidt, C. and Caers, J. (2009). Representing Spatial Uncertainty Using Distances and Kernels, *Mathematical Geosciences* Springer publications,*41 (4)*, 397–419.