Analysis of hydrogeological structure uncertainty by estimation of hydrogeological acceptance probability of geostatistical models

Dylan R. Harp*, Velimir V. Vesselinov

Los Alamos National Laboratory
MS T003, Los Alamos, NM 87544, USA

Abstract

The following describes a proposed approach to account for the equifinality of solutions that result from comparing observations to flow simulations when using realizations of geostatistical models. We introduce hydrogeological acceptance probability to rank the propensity of a geostatistical model to produce acceptable realizations with respect to the consistency of their simulations with observations. The estimation of hydrogeological acceptance probability is equivalent to the calculation of the sample mean of a binomial distribution. This allows the sampling of realizations to be preemptively terminated based on the current estimate and subject to the desired confidence level and interval length. We propose a uncertainty analysis of the hydrogeological heterogeneity utilizing acceptable realizations from a selected set of geostatistical models. In the case of a crisp definition of realization acceptance, this produces a facies probability map. If the definition of realization

*Correponding author
Email addresses: dharp@lanl.gov (Dylan R. Harp), vvv@lanl.gov (Velimir V. Vesselinov)

acceptance is imprecise, the analysis yields upper and lower bounds on the facies probability map in the form of facies plausibility and belief maps, respectively.

Keywords: preemptive sampling, stochastic hydrology, Dempster-Shafer Theory, model ranking

---

1. Introduction

Collecting a sufficient number of direct observations to adequately characterize hydrogeological heterogeneity is not feasible in most, if not all cases. This has led to approaches to infer hydrogeological heterogeneity on the basis of comparisons between output of flow and transport simulators and available observations. Typically, the only available information regarding the hydrogeological heterogeneity of a site are sparsely located direct point estimates and indirect (e.g. water level and/or concentration) observations. Considering reducible (epistemic) and irreducible (aleatoric) uncertainty [1], multiple models of heterogeneity will produce acceptable simulations of the system when compared to the available observations [2]. Well known approaches to account for the equifinality of solutions (i.e. multiple equally acceptable simulations of a system [3]) include maximum likelihood Bayesian model averaging (MLBMA) [4], generalized likelihood uncertainty estimation (GLUE) [2], and the Bayesian approach of [5]. MLBMA provides a relatively efficient approach to rank calibrated models; however, acceptable sub-optimal solutions may be ignored. GLUE provides a general framework for evaluating model uncertainty given a Monte Carlo sampling and parameters of a deterministic model (e.g. hydraulic conductivities of a distributed model). The

Bayesian approach of [5], also requiring a Monte Carlo sampling, provides a formal Bayesian approach to estimate model uncertainty. An approach by [6] combines GLUE and MLBMA, utilizing the statistical approach of MLBMA without ignoring suboptimal solutions. This paper presents a comprehensive approach to evaluate model uncertainty, including imprecision in the definition of model acceptance, given parameters defining geostatistical (stochastic) models, utilizing a preemptive sampling scheme based on the estimation of hydrogeological acceptance probability.

For clarity, we will refer to a geostatistical functional form as a geostatistical framework, while a particular instance of a geostatistical framework, defined by the specification of its parameters, as a geostatistical model. We will refer to a particular instance of a geostatistical model, usually specified by the designation of a random seed, as a realization of the geostatistical model. A geostatistical model defines an infinite set of equally-probable realizations with respect to the statistical characteristics of the spatial variation (geostatistically-equally-probable). Realizations can be conditioned to honor values at particular locations while maintaining the global statistics of the field by using conditional simulation [7]. Various geostatistical frameworks have been developed to characterize and evaluate the structure of aquifer heterogeneity, including variogram [7], multiple-point [8], and Markov-chain geostatistics [9]. While alternatives to geostatistical modeling of heterogeneity exist as well [10, 11, 12, 13], we explore a geostatistical approach here.

Many of the current geostatistical inverse approaches utilized in hydrogeological studies explore uncertainty by evaluating residuals between simulated and observed values (e.g. water levels, concentrations) produced us-

ing geostatistical realizations from a single geostatistical model (i.e. a single conceptual model of the statistical characteristics of the aquifer heterogeneity) which is inferred from the observed or assumed hydrogeology of a site (e.g. [14, 15, 16, 17, 18]). These approaches provide unrealistic estimates of uncertainty in cases where an appropriate statistical model of the heterogeneity is uncertain, a common scenario in hydrogeological studies with sparse data. [19] quantify uncertainty in estimating semivariogram parameters from empirical semivariograms using a maximum likelihood estimation, and demonstrate how this uncertainty can be propagated to predictions of head. [20] developed an approach for geostatistical design that they refer to as continuous Bayesian model averaging considering uncertainty in parameters of the Matèrn family of covariance functions. Other approaches evaluate the consistency of geostatistical models (conceptual model of the statistical properties of the heterogeneity) with observations in a statistical sense [21, 22, 23]. These approaches assume that the statistical characteristics of the 'true' heterogeneity can be inferred by the observed hydrogeology (hydrogeological observations from a single realization of the 'true' geostatistical model).

In general, water pressure residuals will vary substantially between realizations from the same geostatistical model, indicating that geostatistically-equally-probable realizations are not necessarily hydraulically-equally probable. Conversely, it is also possible that realizations from various alternative geostatistical models will produce similar hydraulic responses. Therefore, evaluating modeling residuals in the unavoidable presence of epistemic and aleatoric uncertainty means that multiple geostatistical models will produce

realizations resulting in acceptable simulations of the system given the available observations. As a result, the identification of the geostatistical model that is the most consistent with observations in a statistical sense does not necessarily ensure that the true geostatistical model has been identified. This is an example of asking a question that cannot be answered given the typical information provided. The question that can be asked (and answered) is what is the propensity of a geostatistical model to produce realizations resulting in simulated values consistent with observations given the epistemic and aleatoric uncertainty.

One can rank stochastic realizations by how consistent their simulated values are to observed values. Researchers have proposed many functional forms to accomplish this [2], with perhaps the most common being the sum of squared residuals (SSR). Using a ranking suited to a particular application, criteria for the acceptibility of a realization can be established to incorporate epistemic and aleatoric uncertainty. We propose that geostatistical models can be ranked by estimating their hydrogeological acceptance probability in terms of their propensity to produce "acceptable realizations". A higher acceptance probability of a geostatistical model indicates its propensity to produce realizations resulting in simulated water levels (heads) consistent with observations accounting for measurement (aleatoric) and conceptual (epistemic) uncertainty. The hydrogeological acceptance probability is estimated in a relative sense conditional to the selected flow model.

The fact that multiple geostatistical models can produce realizations with indistinguishable acceptability indicates that an uncertainty analysis must consider more than just the geostatistical model with the highest acceptance

probability. A comprehensive uncertainty analysis would require that all possible geostatistical models be evaluated. Of course, in practice this is typically not feasible. Therefore, such uncertainty analyses are relative, conditional on the considered geostatistical models. We propose that the relative uncertainty associated with hydrogeological heterogeneity can be evaluated by computing the acceptance probability of a sampling of geostatistical models from one or several geostatistical frameworks that incorporate all acceptable realizations from a sampling of geostatistical models. Such an analysis can be utilized to produce facies probability maps of hydrogeological properties.

As the distinction between an acceptable and unacceptable realization with respect to epistemic uncertainty will undoubtedly be imprecise, the use of fuzzy sets [24, 25] is attractive. Fuzzy sets allow an expert to assign grades of membership to the acceptable and unacceptable sets of realizations. Fuzzy sets can also be constructed by integrating information from multiple experts [24]. Previous approaches using fuzzy sets in a hydrological context include mapping flood propagation model outputs to likelihoods in order to evaluate roughness coefficient uncertainty within the GLUE framework [26]; mapping imprecise remotely sensed observations of energy partition variability into model functional types of evapotranspiration [27]; approximation of the unsaturated Darcy law using a fuzzy rule-based model [28]; and a fuzzy least-squares regression approach to the Cooper-Jacob method [29].

Considering a dual relationship between the membership functions for the acceptable and unacceptable sets ($\mu_A = 1 - \mu_U,\ \mu \in [0, 1]$, where μ_A and μ_U are the membership functions for the acceptable and unacceptable sets, respectively), implies that there is a lack of conflict in the information. As a

result, it is possible to estimate the acceptance probability of a geostatistical model while using an imprecise definition of acceptance.

An uncertainty analysis of the hydrogeological heterogeneity in the case of an imprecise definition of acceptibility will produce upper and lower bounds on the facies probability map. These upper and lower probability bounds are consistent with Dempster-Shafer theory (DST) [30], and are referred to as plausibility and belief measures, respectively. Recently, [31] demonstrated the use of DST to account for epistemic and aleatoric uncertainty of permeability measurements. In the limiting case where the fuzzy membership function becomes a crisp indicator function, the gap between the belief and plausibility measures, defining the span of possible probability distributions, collapses to a single probability distribution.

This paper proposes the concept of hydrogeological acceptance probability of a geostatistical model to allow for an imprecise definition of realization acceptance, and demonstrates its ability to characterize the hydraulic conductivity heterogeneity and associated uncertainty. The following sections demonstrate the estimation of acceptance probability, mapping of acceptance probabilities over a geostatistical parameter space, and the mapping of facies probabilities, beliefs and plausibilities based on discrete samplings of geostatistical parameters (models).

2. Estimating hydrogeologic acceptance probability

The concept of the hydrogeological acceptance probability of a geostatistical model incorporates aleatoric and epistemic uncertainty. In the present case, the source of aleatoric uncertainty (i.e. uncertainty dependent on a ran-

dom event and irreducible without the improvement of measuring techniques and/or collection of new measurements [1]) is the measurement of the water levels. The source of epistemic uncertainty (i.e. uncertainty related to the lack of information about the system) is in the definition of the conceptual model (e.g. boundary and initial conditions of the flow simulator and details of the geostatistical model, such as the number of facies). In the following, we define criteria for determining the membership of a realization in the set of acceptable realizations. It should be realized that other criteria could be defined for other applications to fit a particular scenario.

We define an acceptance metric (Φ) as

$$\Phi(\theta) = \sum_{i=1}^{N_o} \begin{cases} \text{if } \hat{h}_i(\theta) > \overline{h}_{i,max} & [\overline{h}_{i,max} - \hat{h}_i(\theta)]^2 \\ \text{if } \hat{h}_i(\theta) < \overline{h}_{i,min} & [\overline{h}_{i,min} - \hat{h}_i(\theta)]^2 \\ \text{if } \overline{h}_{i,min} \leq \hat{h}_i(\theta) \leq \overline{h}_{i,max} & 0 \end{cases} , \qquad (1)$$

where θ is a vector of parameters defining the realization (geostatistical model parameters and random seed), N_o is the number of observations, $\hat{h}_i(\theta)$ is the ith simulated value given θ, and $\overline{h}_{i,max}$ and $\overline{h}_{i,min}$ define an interval of values for the ith simulated value consistent with measurement (aleatoric) uncertainty determined by the resolution and precision of the measuring instruments (e.g. pressure transducers).

Accounting for epistemic uncertainty requires expert opinion of values considered to be consistent with observations given uncertainty in the conceptual model (i.e. boundary and initial conditions, geostatistical framework, etc.). In general, the fuzzy membership function defining membership in the set of acceptable realizations μ_A can be expressed as

$$\mu_A(\theta) = \begin{cases} 1 & \text{if } \Phi(\theta) \leq \Phi_1 \\ 0 & \text{if } \Phi(\theta) \geq \Phi_2 \\ f(\theta) & \text{if } \Phi_1 < \Phi(\theta) < \Phi_2 \text{ where } f(\theta) : \Phi(\theta) \rightarrow [0,1] \end{cases}, \quad (2)$$

where Φ_1 and Φ_2 define the limits of maximum and minimum membership, respectively, in the acceptable set of realizations. $f(\theta)$ defines the grades of membership between the maximum and minimum and is required to decrease monotonically. For demonstration purposes, $f(\theta)$ is assumed to be a linear function here, defined as

$$f(\theta) = \frac{\Phi_2 - \Phi(\theta)}{\Phi_2 - \Phi_1}, \ \Phi_1 < \Phi(\theta) < \Phi_2 \quad (3)$$

Other possible functional forms for $f(\theta)$ include Gaussian, exponential, and step functions [24]. In the special case where $\Phi_1 = \Phi_2$, equation (2) reduces to an indicator function $\chi_A(\theta)$ as

$$\chi_A(\theta) = \begin{cases} 1 & \text{if } \Phi(\theta) \leq \Phi_c \\ 0 & \text{if } \Phi(\theta) > \Phi_c \end{cases}, \quad (4)$$

producing crisp sets consistent with classical set theory [25], where $\Phi_c = \Phi_1 = \Phi_2$ is used to denote the crisp cutoff of acceptibility.

Considering a crisp definition of an acceptable realization (equation (4)), the acceptance probability p_A of a geostatistical model can be estimated as the sample mean of a binomial distribution as

$$\hat{p}_A = \frac{1}{N_r} \sum_{i=1}^{N_r} \chi_A(\theta_i), \quad (5)$$

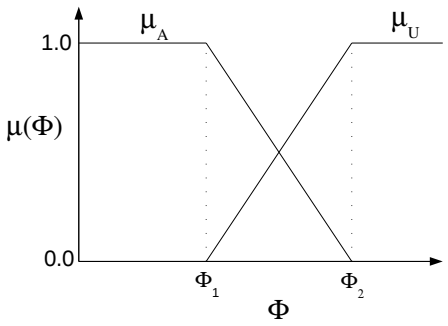

Figure 1: Membership functions for the set of acceptable μ_A and unacceptable μ_U realizations.

where N_r is the total number of realizations. Assuming that membership in the set of unacceptable realizations U, denoted by μ_U, has a dual relationship to μ_A (i.e. $\mu_A = 1 - \mu_U$), indicating that the information is not conflicting [25]) (see Figure 1), a similar equation can be used in the fuzzy case as

$$\hat{p}_A = \frac{1}{N_r} \sum_{i=1}^{N_r} \mu_A(\theta_i), \tag{6}$$

Considering equation (6), it can be demonstrated that the information is not conflicting by verifying that $\hat{p}_A + \hat{p}_U = 1$ ($\hat{p}_U = 1/n \sum_{i=1}^{n} \mu_U(\theta_i)$), which is consistent with the axioms of probability theory.

The two-sided confidence interval length L for $\hat{p}_A$ can be calculated as

$$L = 2z_{\alpha/2} \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n}}, \tag{7}$$

where $z_{\alpha/2}$ is the z-score at $\alpha/2$ where $((1-\alpha) \times 100)\%$ is the confidence level [32, page 485]. By inspecting equation (7), it is apparent that L depends on the current value of $\hat{p}_A$. An equation defining the required sample size to

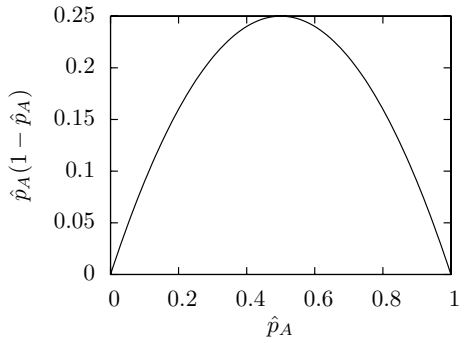

Figure 2: Values of $\hat{p}_A(1-\hat{p}_A)$ as a function of $\hat{p}_A$ demonstrating the influence of $\hat{p}_A$ on the calculation of the sample size n (equation (8)).

estimate $\hat{p}_A$ within L at the $(1-\alpha)*100\%$ confidence level can be determined by solving equation (7) for n as

$$n = \frac{4z^2_{\alpha/2}\hat{p}_A(1-\hat{p}_A)}{L^2}. \tag{8}$$

By considering equation (8) and inspecting Figure 2, it is apparent that the largest number of samples will be required when $\hat{p}_A = 0.5$. Therefore, estimation of small or large values of $\hat{p}_A$ will be less computationally intensive and/or time consuming. This property allows preemptive termination of a sampling at a specified confidence level $((1-\alpha)\times 100)\%$ and interval length (L) based on the current value of $\hat{p}_A$. Values of $\hat{p}_A$ provide a metric to rank the propensity of geostatistical models to produce acceptable realizations.

3. Mapping hydrogeological uncertainty using a set of acceptable realizations

The estimation of the acceptance probability of a geostatistical model produces a set of realizations deemed acceptably consistent with observa-

tions. The hydrogeological uncertainty within the context of this geostatistical model can be evaluated from this set. This analysis of uncertainty can be extended indefinitely by considering the union of acceptable sets from multiple geostatistical models as

$$\bigcup_{i=1}^{N_m} A_i \tag{9}$$

where N_m is the number of geostatistical models and A_i is the acceptable set of realizations from the ith geostatistical model. The geostatistical models are not limited to a particular geostatistical framework. For instance, it would be possible to combine acceptable sets from Markov-chain [9] and indicator kriging [7] geostatistical frameworks in a single analysis.

In the following, we consider the case of geologic units with distinct hydraulic conductivities. A similar analysis can be performed for continuous hydraulic conductivities using discretized intervals. If A is considered crisp (equation (4)), it is possible to map the probability of a particular geologic unit (facies) by approximating the one-location marginal probabilities at each location as

$$\hat{p}(l;\mathbf{x}) = \frac{\sum_{i=1}^{N_r} I_{l,i}(\mathbf{x})\chi_A(\theta_i)}{\sum_{i=1}^{N_r} \chi_A(\theta_i)}, \tag{10}$$

where $\hat{p}(l;\mathbf{x})$ denotes the probability of the lth facies at location $\mathbf{x}$ and $I_{l,i}(\mathbf{x})$ is the lth facies indicator function at location $\mathbf{x}$ associated with the ith realization. By restricting our analysis to the union of sets of acceptable realizations, equation (10) can be simplified as

$$\hat{p}(l;\mathbf{x}) = \frac{1}{N_A}\sum_{i=1}^{N_A} I_{l,i}(\mathbf{x}), \tag{11}$$

where i is now an index for the for the collection of acceptable realizations and N_A is the number of realizations in the union of acceptable sets ($N_A = \sum_{i=1}^{N_r} \chi_A(\theta_i)$). The facies indicator functions are defined as

$$I_{l,i}(\mathbf{x}) = \begin{cases} 1 & \text{if unit } l \text{ occurs at location } \mathbf{x} \text{ in realization } i \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

A facies probability map for the lth facies can be generated by mapping $p(l;\mathbf{x})$ for all $\mathbf{x}$.

If A is considered a fuzzy set (equation (2)), the calculation of the one-location marginal probability is not valid. In this case, equation (11) must be generalized by transferring the information about the acceptance of a realization to a measure that a facies exists at a location as

$$m(l;\mathbf{x}) = \frac{1}{N_A}\sum_{i=1}^{N_A} I_{l,i}(\mathbf{x})\mu_A(\theta_i) \tag{13}$$

where $m(l;\mathbf{x})$ is a basic probability assignment (bpa) consistent with Dempster-Shafer theory (DST) [30]. DST provides a generalization of probability theory, relaxing certain axiomatic constraints to allow for the representation of conflicting information.

In the current context, the properties of $m(l;\mathbf{x})$ can be expressed as

$$m(\varnothing;\mathbf{x}) = 0, \tag{14}$$

$$m(l;\mathbf{x}) \geq 0, \tag{15}$$

$$\sum_{B \in P(\mathbf{X})} m(B) = 1. \tag{16}$$

where $\mathbf{X}$ is the universe of discourse (commonly referred to as the frame of discernment in DST) and $P(\mathbf{X})$ is the set of all subsets of $\mathbf{X}$, where subsets (B) consist of singletons and collections of singletons (the power set of $\mathbf{X}$) [25]. Therefore, m does not adhere to the axioms of probability, as it is defined on $P(\mathbf{X})$, where probabilities are defined on $\mathbf{X}$ (i.e. singletons). By inspecting equation (13), it is apparent that the properties defined by equations (14) and (15) are satisfied. The property defined by equation (16) is also satisfied due to the assumption that the facies are mutually-exclusive and exhaustively defined implicit in a Markov chain or indicator geostatistical framework [9]. This implies that, prior to analysis, the bpa m that at least one of the facies exists at a location ($m(\mathbf{X}; \mathbf{x})$) is equal to one. This is the case of total ignorance with respect to the frame of discernment (i.e. $Bel(\mathbf{X}; \mathbf{x}) = 1$ and $Bel(l; \mathbf{x}) = 0$ for $l = 1, \ldots, N_f$, where N_f is the number of facies). During the analysis, a portion of this evidence is shifted onto the singletons of the power set, reducing the initial uncertainty as information is obtained through the sampling. Therefore, the evidence that is not assigned to a singleton during the analysis remains on $\mathbf{X}$ as

$$m(\mathbf{X}; \mathbf{x}) = 1 - \sum_{l=1}^{N_f} m(l; \mathbf{x}), \tag{17}$$

Rearranging equation (17) demonstrates that the property defined in equation (16) is satisfied.

According to DST, lower and upper bounds on a cumulative distribution function (CDF) can be defined as belief (Bel) and plausibility (Pl) measures,

respectively. The belief of an arbitrary event D can be defined a

$$Bel(D) = \sum_{E|E\subseteq D} m(E) \tag{18}$$

where it is required that the following properties are adhered to:

$$Bel(\varnothing) = 0, \tag{19}$$

$$Bel(\mathbf{X}) = 1, \tag{20}$$

and

$$Bel(D) + Bel(\overline{D}) \leq 1. \tag{21}$$

where $\overline{D}$ indicates not D.

In the current context, the belief that facies l exists at location $\mathbf{x}$ is

$$Bel(l;\mathbf{x}) = m(l;\mathbf{x}) = \frac{1}{N_A}\sum_{i=1}^{N_A} I_{l,i}(\mathbf{x})\mu_A(\theta_i). \tag{22}$$

It is apparent that the properties defined by equation (19) and (20) are satisfied considering the arguments above related to the properties of a bpa (equations (13), (14), (16), and (17)). To demonstrate that equation (22) satisfies equation (21), consider that

$$Bel(\bar{l};\mathbf{x}) = \sum_{m=1|m\neq l}^{N_f} Bel(m;\mathbf{x}) \tag{23}$$

and

$$\sum_{l=1}^{N_f} Bel(l;\mathbf{x}) = \sum_{l=1}^{N_f}\left[\frac{1}{N_A}\sum_{i=1}^{N_A} I_{l,i}(\mathbf{x})\mu_A(\theta_i)\right] \leq \sum_{l=1}^{N_f}\left[\frac{1}{N_A}\sum_{i=1}^{N_A} I_{l,i}(\mathbf{x})\right] = 1. \tag{24}$$

Therefore,

$$Bel(l;\mathbf{x}) + Bel(\bar{l};\mathbf{x}) = Bel(l;\mathbf{x}) + \sum_{m=1|m\neq l}^{N_f} Bel(m;\mathbf{x}) = \sum_{l=1}^{N_f} Bel(l;\mathbf{x}) \leq 1. \tag{25}$$

in accordance with equation (21).

The plausibility of an arbitrary event D can be defined as

$$Pl(D) = \sum_{E|D\cap E\neq\varnothing} m(E), \tag{26}$$

requiring that the following properties are adhered to

$$Pl(\varnothing) = 0, \tag{27}$$

$$Pl(\mathbf{X}) = 1, \tag{28}$$

and

$$Pl(D) + Pl(\overline{D}) \geq 1. \tag{29}$$

In the current context, according to the duality of belief and plausibility measures [25], the plausibility that facies l exists at $\mathbf{x}$ can be defined as

$$Pl(l;\mathbf{x}) = 1 - Bel(\bar{l};\mathbf{x}). \tag{30}$$

Equation (30) satisfies equation (27) and (28) as

$$Pl(\varnothing;\mathbf{x}) = 1 - Bel(\overline{\varnothing};\mathbf{x}) = 1 - Bel(\mathbf{X}) = 0 \tag{31}$$

and

$$Pl(\mathbf{X};\mathbf{x}) = 1 - Bel(\overline{X};\mathbf{x}) = 1 - Bel(\varnothing) = 1, \tag{32}$$

respectively. Adherence of equation (30) to equation (29) can be demonstrated considering equation (21) as

$$Pl(l;\mathbf{x}) + Pl(\bar{l};\mathbf{x}) = 2 - (Bel(l;\mathbf{x}) + Bel(\bar{l};\mathbf{x})) \geq 1. \tag{33}$$

Mapping equations (22) and (30) for all $\mathbf{x}$ provides upper and lower bounds for the facies probability map in the form of plausibility and belief maps, respectively. The gap between the plausibility and belief measures indicates the range of possible values for the facies probability, indicating the epistemic uncertainty in defining the acceptability of a realization. In the case where $\Phi_1 = \Phi_2$, the gap between the plausibility and belief collapses to the facies probability map.

As the acceptance information (degree of membership) is applied to the entire realization, and not to individual cells (locations) within the realization, the gap between the belief and plausibility will be constant over the model domain, with its magnitude dependent on the imprecision in the definition of realization acceptance. In other words, the conflict in the information due to imprecision in the definition of the acceptance metric is constant over the model domain ($Pl(l;\mathbf{x}) - Bel(l;\mathbf{x}) = c,\ c \geq 0$).

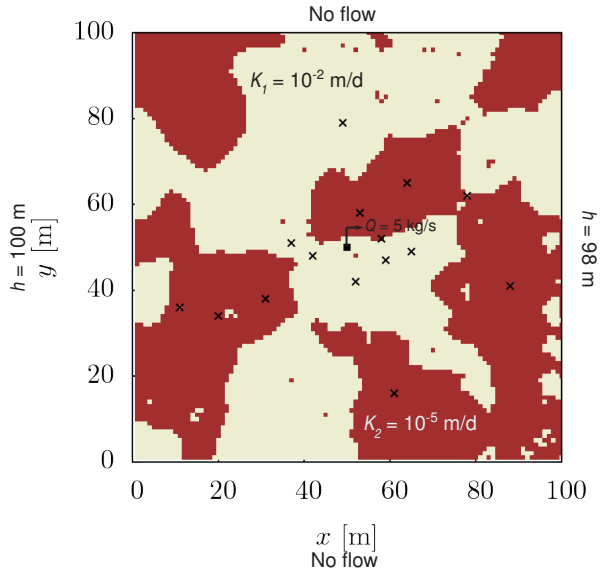

Figure 3: Map of horizontal plane of synthetic model. Colored regions indicate geologic units with distinct hydraulic conductivities. Black x's indicate locations of water-level monitoring wells (15). The pumping well is indicated by a black square with its associated pumping rate at the approximate center of the model domain. Boundary conditions are noted on the four sides of the map.

4. Synthetic case study

The use of hydrologic acceptance probability is demonstrated on a numerically-generated synthetic pumping test where the 'true' hydrogeological property structure, boundary conditions, and initial conditions are specified. In this way, the feasibility of using hydrological acceptance probability can be evaluated. The model (presented in Figure 3) is a 2-dimensional representation of the horizontal plane of an aquifer composed of 2 distinct geologic units (facies) with uniform hydraulic conductivities (i.e. $K_1 = 10^{-2}$ m/s, $K_2 = 10^{-5}$ m/s). The 'true' mapping of the facies is generated as a real-

ization of a Markov-chain geostatistical model [9, 33] with facies volumetric proportions of $p_1 = p_2 = 0.5$ and mean facies lengths of $\bar{l}_{1,x} = \bar{l}_{1,y} = 20$ m where the numerical subscripts indicate the facies and letter subscripts indicate direction (refer to [9, 33, 23] for detailed discussions of Markov-chain geostatistics).

The flow simulations are performed on an orthogonal grid with 1 m spacing (10,000 nodes) using FEHM [34]. The geostatistical grid cells are concurrent with the flow simulation grid cells. An ambient gradient of 0.02 m/m is induced in the positive x-direction by imposing constant head boundaries of 100 m and 98 m along lines x=0 m and 100 m, respectively. A pumping well is located in the approximate center of the model. The pumping well discharges at a constant rate of 5 kg/s beginning at t=0 s. Fifteen observation wells are located throughout the model domain (black x's in Figure 3). For demonstration purposes, it is assumed that the hydraulic conductivity is perfectly known at the pumping and observation wells. Head measurements are collected at the pumping and observation wells at 11 times ($t = 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12$ and 10.0 days), resulting in 176 (16 wells $\times$ 11 times) observations. Gaussian noise with a mean of zero and standard deviation of 2.5×10^{-5} m was added to the simulated values for the 'true' case to produce observed values with measurement error.

5. Results and discussion

This paper introduces the application of acceptance probability to evaluate structural hydrogeological uncertainty due to limited pressure and hydro-

geologic observations. In order to investigate the acceptance probability of various geostatistical models within a Markov-chain geostatistical framework, discrete geostatistical parameter sets are evaluated, where each parameter set defines a geostatistical model. The parameters are the mean facies lengths in the x and y directions, $\bar{l}_x$ and $\bar{l}_y$ [9], respectively. A matrix of parameter combinations are considered, varying both parameters from 10 to 50 m at 5 m increments (81 parameter sets), where $\bar{l}_x = \bar{l}_y = 20$ m are the parameters used to generate the 'true' realization (Gaussian noise with zero mean and standard deviation equal to 0.005 m is added to the simulated values from the 'true' realization to produce observed values with measurement error). It is important to note that these stochastic geostatistical parameters do not define the 'true' spatial distribution of hydraulic conductivity, but do define the geostatistical model used to generate the 'true' realization of hydraulic conductivity. Other sampling schemes could have been chosen (e.g. Monte Carlo, Markov chain Monte Carlo, Latin Hypercube Sampling) in order to search the parameter space more thoroughly.

This demonstration uses discrete parameter sets in order to decrease the computational requirements. It includes (1) example calculations of hydrogeological acceptance probability (Figures 4 and 5); (2) mapping of acceptance probability on the parameter space and mappings of facies probability, using crisp definitions of acceptance demonstrating the effects of modifying acceptance probability estimation parameters (Figures 6, 7, 8, and 9); and (3) facies plausibility and belief maps for imprecise definitions of acceptance (Figure 10). In all cases, a value of 0.015 m is added and subtracted from the observed value to obtain values for $\overline{h}_{max}$ and $\overline{h}_{min}$, respectively. These bounds

for measurement error are significantly larger than the Gaussian noise added to the 'true' simulated values (zero mean, standard deviation 2.5×10^{-5} m), providing a sufficiently large range of values considered consistent with respect to measurement error.

Figure 4 presents plots of simulated versus observed heads for a range of acceptance metric Φ values. The simulated values are from different realizations generated by the geostatistical model used to generate the 'true' heterogeneity ($\bar{l}_x = \bar{l}_y = 20$ m). From Figure 4, it is apparent that as the consistency of the realizations to the observed hydraulic response decreases (Φ increases), initially there is increased scatter both above and below the 1:1 line. However, in the plot in the lower left with $\Phi = 1010$, there is a bias for increased simulated drawdown for locations with higher observed drawdown (lower observed head). This is due to changes in heterogeneity patterns at locations near the pumping well (i.e. simulating low permeability near monitoring wells with high permeability nearby in the 'true' heterogeneity. While all monitoring wells and the pumping well are conditioned to the 'true' facies, it is possible to produce significantly different heterogeneity near these locations. In the extreme case ($\Phi = 3.27 \times 10^5$), the heterogeneity is so inconsistent with the observed water levels the majority of simulated values are less than the observed values (greater simulated drawdown than observed).

Figure 5 presents acceptance metric Φ values (equation (1)) for realizations from three geostatistical models: (a) the geostatistical model used to generate the 'true' realization ($\bar{l}_x = \bar{l}_y = 20$ m), (b) the geostatistical model with the highest acceptance probability ($\bar{l}_x = 10$m, $\bar{l}_y = 10$m), and (c) the

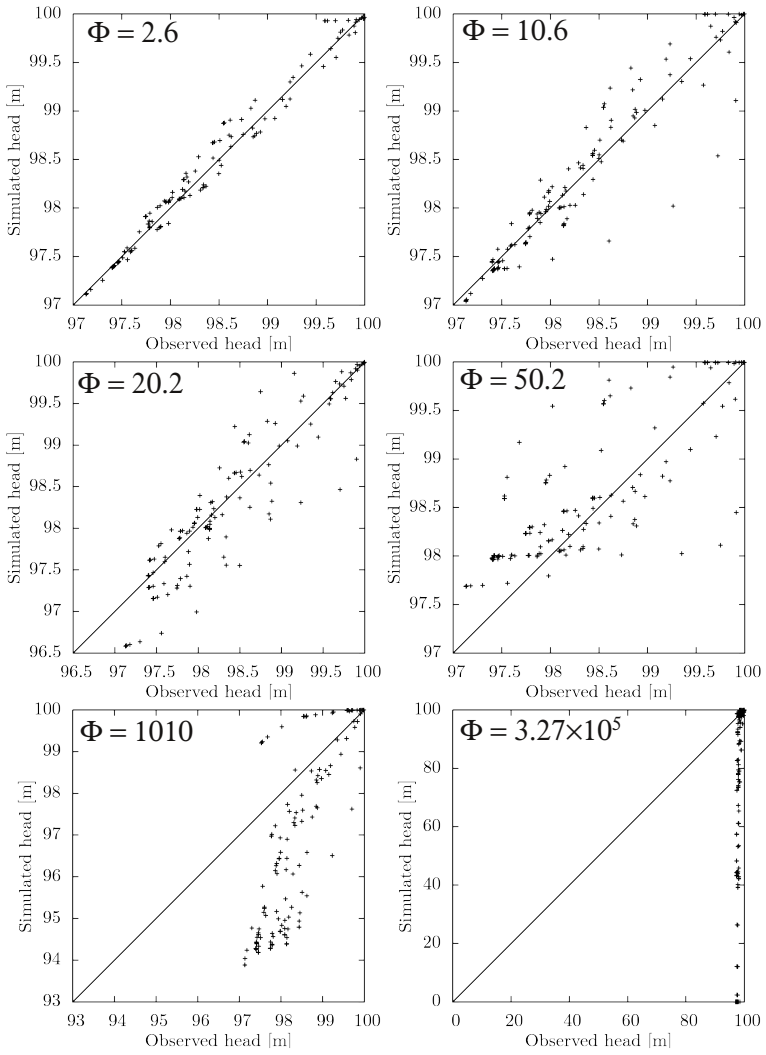

Figure 4: Plots of simulated heads for realizations from the geostatistical model used to generate the 'true' heterogeneity versus observed heads from the 'true' heterogeneity. Cases with a range of acceptance metric Φ values are presented and noted on each plot. The diagonal line has a 1:1 slope. Note differences in scale between plots.

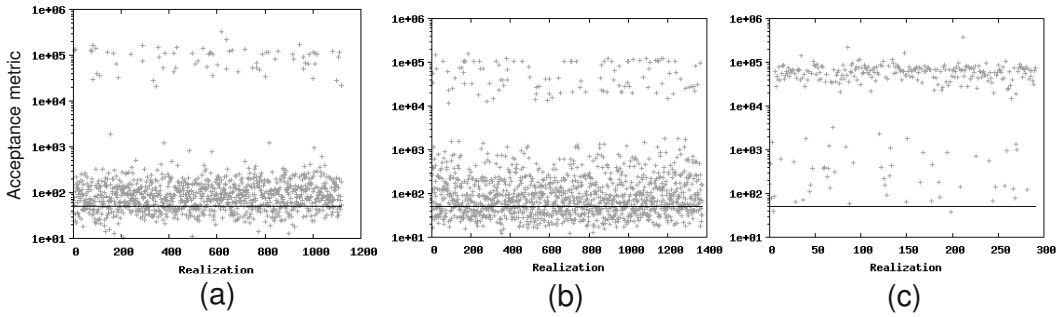

Figure 5: Acceptance metric values for realizations from three geostatistical models. The horizontal line at $\Phi = 50$ indicates the values of Φ_c used to produce the samples of realizations ($\alpha = 0.05$, $L = 0.05$). The geostatistical models and resulting acceptance probabilities (rounded to their specified precision) are the geostatistical model (a) used to generate the 'true' realization $\hat{p}_A = 0.25$, (b) with highest acceptance probability $\hat{p}_A = 0.35$, and (c) with lowest acceptance probability $\hat{p}_A = 0.00$.

lowest acceptance probability ($\bar{l}_x = 10$ m, $\overline{y} = 50$ m). Each cross in Figure 5 signifies the results of one plot from Figure 4.

For reference, a horizontal line is drawn through $\Phi = 50$, representing an example value for Φ_c. This horizontal line denotes the expert defined cutoff between acceptable and unacceptable realizations, separating the realizations into these two sets. A fuzzy definition of acceptance would require a range of values of Φ where the membership would transition monotonically from one set to the other. The acceptance probability of the geostatistical model is calculated by dividing the number of realizations in the acceptable set (points below Φ_c in the crisp case presented in Figure 5) by the total number of realizations. The acceptance probabilities are (a) 0.25, (b) 0.35, and (c) 0.0 rounded to the specified resolution of the estimation ($L = 0.05$) at 95% confidence.

As illustrated in Figure 2, probability estimates closer to 0.5 will require more samples n (equation (8)) for the same confidence level and interval length, as apparent in the number of realizations presented in the plots in Figure 5, where (b), with probability closest to 0.5, requires the most realizations at 1377, and (c), with probability farthest from 0.5, requires the least at 292. Figure 5 (a) requires 1120 realizations.

Figure 6 presents acceptance probability maps on the geostatistical parameter space estimated at 95% confidence with $\Phi_c = 50$. Confidence interval lengths (L) of (a) 0.05, (b) 0.025, and (c) 0.01 are presented. Acceptance probability estimates are rounded to increments spaced at the confidence interval length to avoid presenting greater resolution than the estimates warrant. This is evident as an increase in the number of colors in moving from Figure 6(a) to 6(c). The geostatistical model (geostatistical parameter set) with the highest acceptance probability is $\bar{l}_x = \bar{l}_y = 10$ m (Figure 5 (b)), not the model used to generate the 'true' realization, $\bar{l}_x = \bar{l}_y = 20$ m (Figure 5 (a)). This is not a surprising result, as the head observations come from a single realization of this geostatistical model, and should not be expected to uniquely characterize the hydraulic response of the geostatistical model. In practice, this is the case as well, where the head observations are produced by a single spatial distribution of heterogeneity, where not only will the geostatistical parameters be unknown, but the appropriate geostatistical framework may be uncertain as well. It is therefore reasonable to expect that the observed hydraulic response may be more consistent with geostatistical models that are not the 'true' model. It is for this reason that the identification of an optimal geostatistical model from hydraulic observations is an

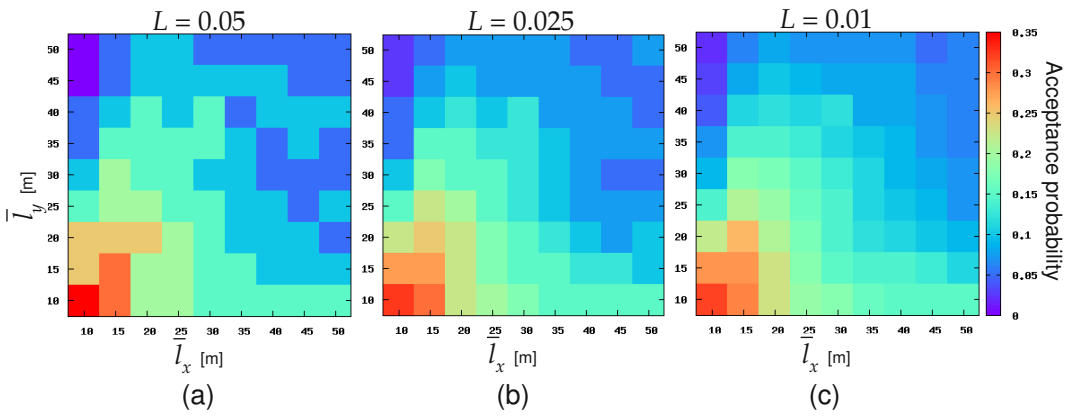

Figure 6: Maps of hydrologic acceptance probability on the parameter space of geostatistical model parameters. Probabilities are estimated at the 95% confidence level for interval lengths L of (a) 0.05, (b) 0.025, and (c) 0.01. Φ_c is 50 for all three cases.

ill-conceived strategy.

A similar presentation to Figure 6 is made in Figure 7, in this case demonstrating the effect of the perceived conceptual model uncertainty by modifying values of Φ_c ((a) 40, (b) 60, and (c) 80) holding $L = 0.05$. Once again, the confidence level of the estimation is 95%. By inspecting Figure 7, it is apparent that increasing the perceived conceptual model uncertainty (increasing Φ_c) increases the acceptance probability values of the matrix of geostatistical models as a greater proportion of realizations become acceptable.

Facies probability maps are presented in Figure 8, presenting the probability of the high conductivity unit ($K_1 = 10^{-2}$m/s). Probabilities near 1.0 or 0.0 can be interpreted as locations with high certainty of belonging to the high (K_1) or low (K_2) conductivity geologic unit, respectively. Locations with probabilities approaching 0.5 indicate that the geologic unit at that location is uncertain. Therefore, equivalent information could be presented

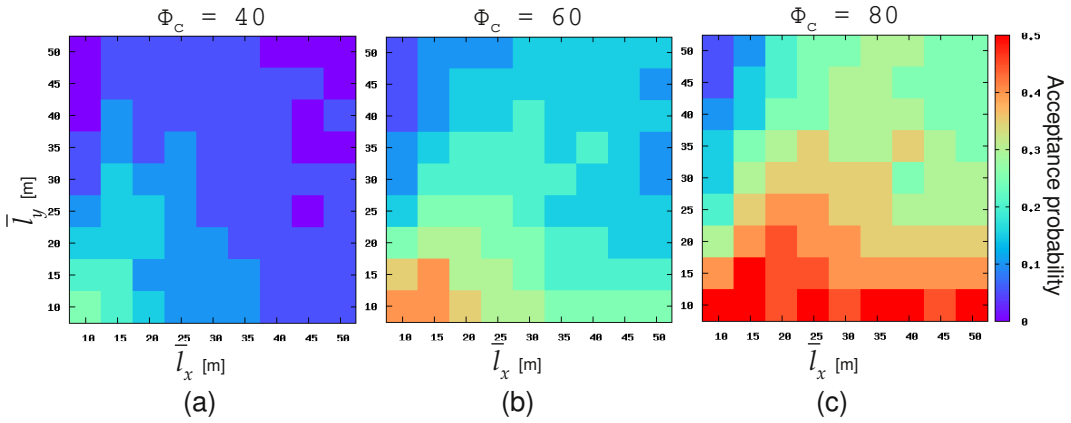

Figure 7: Maps of hydrologic acceptance probability on the parameter space of geostatistical model parameters. The effect of perceived conceptual model uncertainty are demonstrated by presenting maps for Φ_c of (a) 40, (b) 60, and (c) 80. Probabilities are estimated at the 95% confidence level for an interval length L of 0.05.

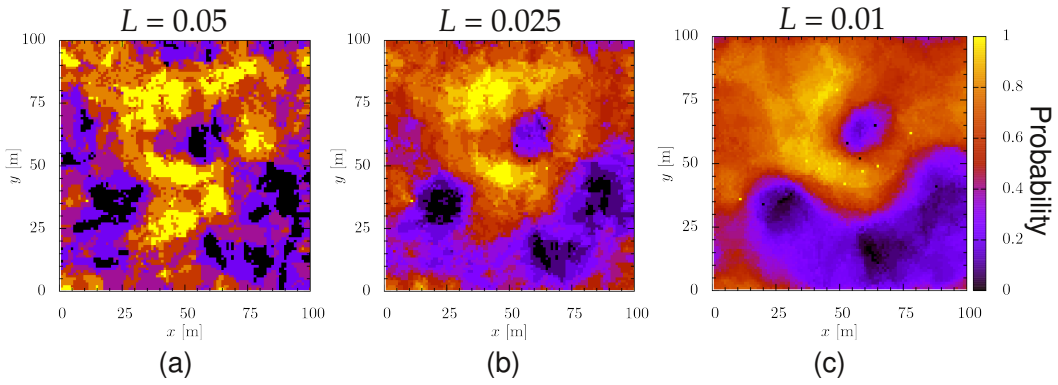

Figure 8: Facies probability maps for the high conductivity geologic unit ($K_1 = 10^-2$ m/s). Acceptance probabilities used to generate these maps were estimated with confidence interval (a) 0.050, (b) 0.025, and (c) 0.010 with 95% confidence and $\Phi_c = 10$.

by mapping the probability of the low conductivity unit alternatively. The maps are developed using realizations from the union of plausible sets from a matrix of geostatistical parameters (identical to the matrix of parameter values presented in Figure 6 (81 parameter sets)) with $\Phi_c = 10$. The acceptance probabilities are estimated within confidence interval lengths L of (a) 0.05, (b) 0.025, and (c) 0.01 at 95% confidence. The number of realizations increases as L decreases, resulting in a smoothing effect of the probability map. Comparisons can be made between these maps and the 'true' hydrogeologic model in Figure 3.

Figure 9 presents a similar analysis to Figure 8, in this case increasing Φ_c, effectively increasing the degree of uncertainty that the expert perceives in the conceptual model. For reference, Figures 8(a) and 9(a) are the same. It is apparent that as the Φ_c increases, indicating a greater level of epistemic uncertainty and that a greater number of realizations will be considered acceptable, the more that the information is restricted to the conditioning points, with locations far from the conditioning points having the greater uncertainty (facies probability ≈ 0.5).

Belief and plausibility maps are presented in Figure 10 for various membership functions. Two cases are presented, both starting with crisp acceptance metrics: (1) $\Phi_1 = \Phi_2 = 10$ and (2) $\Phi_1 = \Phi_2 = 20$, demonstrating that the facies belief and plausibility maps are equal to the facies probability maps in these cases (refer to Figure 9 (a) and (c), respectively). Below the plots with crisp acceptance metrics are facies belief and plausibility maps increasing Φ_2 while maintaining the same value for Φ_1. This increases the support (range of nonzero membership) of the membership functions. These

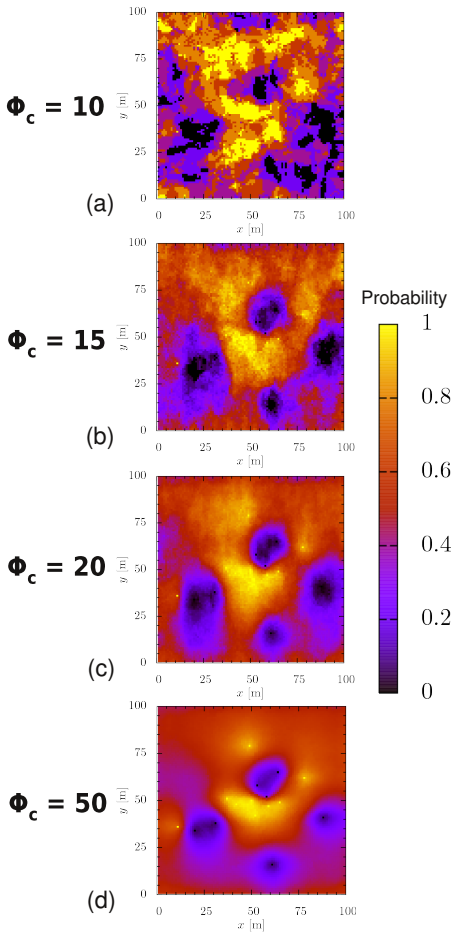

Figure 9: Facies probability maps for the high conductivity geologic unit ($K = 10^{-}2$ m/s). The plausibilities used to generate these maps were estimated with confidence interval 0.050 with 95% confidence. The effect of conceptual model uncertainty is evaluated using values of Φ_c of (a) 10, (b) 15, (c) 20, and (d) 50.

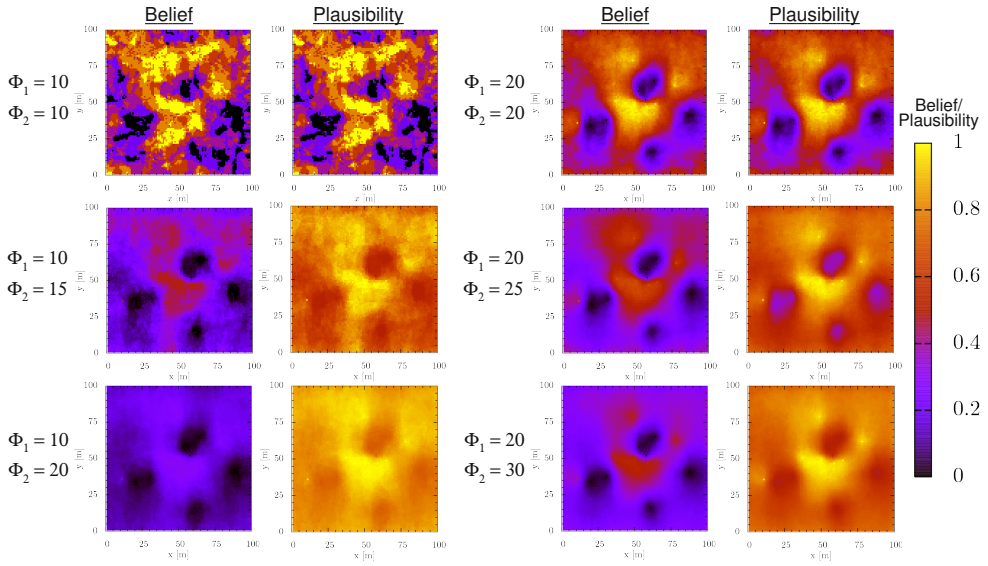

Figure 10: Facies belief and plausibility maps for various definitions of realization acceptance defined by Φ_1 and Φ_2 (refer to equations (2) and (3)).

plots indicate the increasing gap between $Pl(l;\mathbf{x})$ and $Bel(l;\mathbf{x})$ as the support increases, demonstrating how the increased epistemic uncertainty in the definition of the acceptance metric is included in the analysis.

6. Conclusions

Accounting for the equifinality of solutions requires the inclusion of all solutions deemed acceptable by their consistency with observations. In geostatistical models, this means that model ranking can not be used to identify the 'true' model of statistical characteristics of heterogeneity. This is an example of asking a question that cannot be answered given the available information. We have proposed to instead ask the question what is the propensity of the geostatistical models to produce acceptable realizations within the context of

the chosen flow simulator. Estimation of the acceptance probability provides this information.

Based on the available information, all acceptable realizations with respect to observations (e.g. water levels or concentrations) must be considered in a comprehensive uncertainty analysis. While a complete sampling of all possible geostatistical models may not be feasible, an appropriate sampling should be selected. A crisp definition of acceptance will allow the calculation of facies probability maps within the context of the sampled geostatistical models. An imprecise definition of acceptance will produce bounds on the facies probability map in the form of facies belief and plausibility maps, incorporating epistemic uncertainty in the definition of realization acceptability in the analysis.

This paper has demonstrated that:

1. Geostatistically-equally-probable realizations from a geostatistical model can produce extremely varied flow simulations,

2. Considering epistemic and aleatoric uncertainty, many geostatistical models will produce acceptably consistent realizations with respect to observations,

3. Acceptance probability can be used to rank geostatistical models according to their propensity to produce acceptable realizations,

4. Estimation of acceptance probability at a specified confidence level and interval length can be preemptively terminated based on the current estimate,

5. Varying the confidence level and/or interval length can be used to obtain information at the level of detail and computational expense desired,

6. Expert opinion of epistemic uncertainty provides the appropriate uncertainty in facies distribution (increasing Φ_c increases the uncertainty in facies distribution as more realizations become acceptable),

7. Imprecise definitions of epistemic uncertainty can be accounted for by utilizing imprecise probabilities (DST), providing bounds of the facies probability map in the form of facies belief and plausibility maps.

Acknowledgements

This work was supported by various projects within the Environmental Programs directorate of the Los Alamos National Laboratory. The authors are grateful for constructive reviews of this work by Greg Chavez, Kay H. Birdsell, Zhiming Lu, Timothy J. Ross, Lowell E. Harp,...

References

[1] J. L. Ross, M. M. Ozbek, G. F. Pinder, Aleatoric and epistemic uncertainty in groundwater flow and transport simulation, Water Resources Research 45. doi:10.1029/2007WR006799.

[2] K. Beven, A. Binley, The future of distributed models: Model calibration and uncertainty prediction, Hydrologic Processes 6 (1992) 279–298.

[3] K. Beven, J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology 249 (2001) 11–29.

[4] S. Neuman, Maximum likelihood bayesian averaging of uncertain model predictions, Stochastic Environmental Research and Risk Assessment 17 (2003) 291–305.

[5] P. Gaganis, L. Smith, A Bayesian approach to the quantification of the effect of model error on the prediction of groundwater models, Water Resources Research 37 (9) (2001) 2309–2322.

[6] R. Rojas, L. Feyen, A. Dassargues, Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and bayesian model averaging, Water Resources Research 44. doi:10.1029/2008WR006908.

[7] C. Deutsch, A. Journel, GSLIB Geostatistical Software Library and User's Guide, Oxford University Press, New York, 1992.

[8] S. Strebelle, Conditional simulation of complex geologic structures using multiple-point statistics, Mathematical Geology 34 (2002) 1–21. doi:10.1023/A:1014009426274.

[9] S. Carle, G. Fogg, Transition probability-based indicator geostatistics, Mathematical Geology 28 (4) (1996) 453–476.

[10] G. Matheron, Eléments pour une théorie des milieux poreux [elements for a theory of porous media], Masson, Paris (1967).

[11] H. Ben Ameur, G. Chavent, J. Jaffre, Refinement and coarsening indicators for adaptive parametrization: Application to the estimation of hydraulic transmissivities, Inverse Problems 18 (2002) 775–794.

[12] F.-C. Tsai, N.-Z. Sun, W.-G. Yeh, Geophysical parameterization and parameter structure identification using natural neighbors in groundwater inverse problems, Journal of Hydrology 308 (2005) 269–283, doi:10.1016/j.jhydrol.2004.11.004. doi:10.1016/j.jhydrol.2004.11.004.

[13] Z. Lu, B. Robinson, Parameter identification using the level set method, Geophysical Research Letters 33, l06404, doi:10.1029/2005GL025541. doi:10.1029/2005GL025541.

[14] B. Ramarao, A. Lavenue, G. de Marsily, M. Marietta, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1. theory and computational experiments, Water Resources Research 31 (3) (1995) 475–493.

[15] P. Kitanidis, On the geostatistical approach to the inverse problem, Adv. Water Resour. 19 (6) (1996) 333–342.

[16] T.-C. Yeh, S. Liu, Hydraulic tomography: Development of a new aquifer test method, Water Resources Research 36 (8) (2000) 2095–2105.

[17] V. Vesselinov, S. Neuman, W. Illman, Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff 2. equivalent parameters, high-resolution stochastic imaging and scale effects, Water Resources Research 37 (12) (2001) 3019–3041.

[18] A. Hernandez, S. Neuman, A. Guadagnini, J. Carrera, Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion, Stochastic Environmental Research and Risk Assessment 17 (5) (2003) 329–338.

[19] E. Pardo-Igùzquiza, M. Chica-Olmo, M. J. Garcia-Soldado, J. A. Luque-Espinar, Using semivariogram parameter uncertainty in hydrogeological applications, Ground Water 47 (1) (2009) 25–34.

[20] W. Nowak, F. de Barros, Y. Rubin, Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain, Water Resources Research 46. doi:10.1029/2009WR008312.

[21] P. M. Meier, A. Medina, J. Carrera, Geostatistical inversion of cross-hole pumping tests for identifying preferential flow channels within a shear zone, Ground Water 39 (1) (2001) 10–17.

[22] D. Harp, Z. Dai, A. Wolfsberg, J. Vrugt, B. Robinson, V. Vesselinov, Aquifer structure identification using stochastic inversion, Geophysical Research LettersL08404, doi:10.1029/2008GL033585. doi:10.1029/2008GL033585.

[23] D. Harp, V. Vesselinov, Stochastic inverse method for estimation of geostatistical representation of hydrogeologic stratigraphy using borehole logs and pressure observations, Stochastic Environmental Research and Risk Assessment.

[24] T. J. Ross, Fuzzy Logic with Engineering Applications, 2nd Edition, John Wiley & Sons, West Sussex, England, 2004.

[25] G. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, John Wiley & and Sons, Hoboken, NJ, 2006.

[26] G. Aronica, B. Hankin, K. Beven, Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, Advances in Water Resources 22 (4) (1998) 349–365.

[27] S. W. Franks, K. J. Beven, Estimation of evapotranspiration at the landscape scale: A fuzzy disaggregation approach, Water Resources Research 33 (12) (1997) 2929–2938.

[28] A. Bàrdossy, A. Bronstert, B. Merz, 1-,2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach, Advances in Water Resources 18 (4) (1995) 237–251.

[29] B. R. Mathon, M. M. Ozbek, G. F. Pinder, Transmissivity and storage coefficient estimation by coupling the Cooper-Jacob method and modified fuzzy least-squares regression, Journal of Hydrology 253 (2008) 267–274. doi:10.1016/j.jhydrol.2008.02.004.

[30] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[31] B. R. Mathon, M. M. Ozbek, G. F. Pinder, Dempster-Shafer theory applied to uncertainty surrounding permeability, Mathematical Geosciences 42 (2010) 293–307. doi:10.1007/s11004-009-9246-0.

[32] A. J. Hayter, Probability and Statistics for Engineers and Scientists, 2nd Edition, Duxbury, 2002.

[33] S. Carle, G. Fogg, Modeling spatial variability with one and multidi-

mensional continuous-lag markov chains, Mathematical Geology 29 (7) (1997) 891–918.

[34] G. Zyvoloski, B. Robinson, Z. Dash, L. Trease, Summary of the models and methods for the fehm application—a finite-element heat- and mass-transfer code, Tech. Rep. LA-13307-MS, Los Alamos Natl. Lab., Los Alamos, NM (1997).