

# Quantitative Integration of High-Resolution Hydrogeophysical Data: A Novel Approach to Monte-Carlo-Type Conditional Stochastic Simulations and Implications for Hydrological Predictions

Baptiste Dafflon, James Irving, Klaus Holliger\*

*Institute of Geophysics, University of Lausanne, CH-1015 Lausanne, Switzerland*

**ABSTRACT:** Geophysical techniques can help to bridge the inherent gap that exists with regard to spatial resolution and coverage for classical hydrological methods. This has led to the emergence of a new and rapidly growing research domain generally referred to as hydrogeophysics. Given the differing sensitivities of various geophysical techniques to hydrologically relevant parameters, their inherent trade-off between resolution and range, as well as the notoriously site-specific nature of petrophysical parameter relations, the fundamental usefulness of multi-method surveys for reducing uncertainties in data analysis and interpretation is widely accepted. A major challenge arising from such endeavors is the quantitative integration of the resulting vast and diverse database into a unified model of the probed subsurface region that is consistent with all available measurements. To this end, we present a novel approach toward hydrogeophysical data integration based on a Monte-Carlo-type conditional stochastic simulation method that we consider to be particularly suitable for high-resolution local-scale studies. Monte Carlo techniques are flexible and versatile, allowing for accounting for a wide variety of data and constraints of differing resolution and hardness, and thus have the potential of providing, in a geostatistical sense, realistic models of the pertinent target parameter distributions. Compared to more conventional approaches, such as co-kriging or cluster analysis, our approach provides significant advancements in the way that larger-scale structural information contained in the hydrogeophysical data can be accounted for. After outlining the methodological background of our algorithm, we present the results of its application to the integration of porosity log and tomographic crosshole georadar data to generate stochastic realizations of the detailed local-scale porosity structure. Our procedure is first tested on pertinent synthetic data and then applied to a field dataset collected at the Boise Hydrogeophysical Research Site. Finally, we compare the performance of our data integration approach to that of more conventional methods with regard to the prediction of flow and transport phenomena in highly

heterogeneous media and discuss the implications arising.

**KEY WORDS:** aquifer characterization, conditional simulation, geostatistics, hydrogeophysics, hydrology, quantitative data integration, Monte Carlo methods, simulated annealing, flow and transport modeling.

---

This study was supported by the Swiss National Science Foundation.

\*Corresponding author: klaus.holliger@unil.ch

Manuscript received October 3, 2008.

Manuscript accepted February 18, 2009.

## INTRODUCTION

Traditionally, aquifer characterization is based on the analysis of drill cores as well as the results of tracer and pumping experiments (e.g., Hubbard and Rubin, 2005). Core studies can provide detailed local information but are inherently 1D in nature, whereas tracer and pumping tests tend to capture the gross average properties of a probed region. Without complementary information, these techniques may thus be inadequate for reliably characterizing laterally heterogeneous aquifers. The inherent gap in spatial resolution and coverage between core analyses and pumping/tracer tests, which generally amounts to several orders of magnitude, can be bridged by high-resolution geophysical techniques. As a result, the corresponding new field of research, generally referred to as hydrogeophysics or groundwater geophysics, has been evolving strongly, as proven, for example, by the recent publication of several review-type textbooks on the subject (e.g., Kirsch, 2006; Vereecken et al., 2006; Rubin and Hubbard, 2005a).

Given the gaps in resolution and coverage between traditional hydrological techniques on one hand and the widely differing sensitivities to hydrologically relevant parameters as well as the notorious trade-off between resolution and range of the various geophysical techniques on the other hand, it is clear that optimal results are most likely obtained through the use of multiple methods. To obtain a detailed and internally consistent aquifer model, the geophysical data need to be integrated with all other hydrological data as well as any other kind of relevant a priori information. The quantitative integration of such a diverse database represents a major challenge due to the widely varying physical nature and hardness of the data, their differing scales of resolution and their deterministic or stochastic quantification (e.g., Paasche et al., 2006; Rubin and Hubbard, 2005b; Hyndman et al., 2000).

In recent years, significant progress regarding the quantitative characterization of hydrocarbon reservoirs has been made by integrating wide ranges of geophysical, petrophysical, and production data through geostatistical interpolation, classification, and simulation techniques (e.g., Deutsch, 2002; Kelkar and Perez, 2002). There are close analogies between hydrocarbon reservoirs and aquifers in general, and their geophysi-

cal characterization in particular (e.g., Szerbiak et al., 2001). However, an important difference between the two disciplines is that high-resolution crosshole tomographic methods, which are widely used in hydrogeophysical studies, are generally unavailable for the geophysical characterization of hydrocarbon reservoirs. It is therefore reasonable to assume that the integration of geophysical and hydrological data through suitably adapted geostatistical techniques will eventually prove to be at least as successful as their more established applications in the hydrocarbon industry. Given the rapidly growing need for managing and protecting groundwater resources, this clearly is an interesting and exciting prospect, which should be systematically explored (e.g., Tronicke and Holliger, 2005; Hubbard et al., 2001; Hyndman et al., 2000).

In this article, we pursue the above objective. After a brief review of commonly used geostatistical data integration techniques, we present a novel Monte-Carlo-type conditional simulation approach that we consider to be particularly promising for hydrogeophysical data integration. This approach is first tested on a realistic synthetic database and then applied to a pertinent observed dataset. Finally, we explore and discuss the potential implications arising for the prediction of hydrological flow and transport phenomena by comparing the results obtained with our approach with those obtained by more conventional techniques.

## DATA INTEGRATION

The potential of multi-method geophysical surveys to reduce uncertainties in data analysis and interpretation is widely recognized (e.g., Garambois et al., 2002; Hubbard et al., 2001; Dannowski and Yaramanci, 1999). A common way to integrate multiple and diverse geophysical surveys is by deriving independent subsurface property models, which are then jointly interpreted to obtain a single integrated model of the probed region. This approach is, however, largely qualitative in nature, and hence the outcome depends heavily on the background, experience, and preconceptions of the interpreter. Most importantly, this approach also largely excludes a rigorous quantitative assessment of the overall quality and internal consistency of the inferred integrated models.

A more quantitative approach to this problem is

to link multiple datasets during the inversion and model generation process. As is the case for essentially all quantitative data integration procedures, an important advantage of such joint inversion approaches is that the inherent ambiguity of each dataset is reduced due to the additional constraints offered by the other datasets. However, joint inversion techniques tend to require far-reaching a priori assumptions on the relations between the various parameters involved (e.g., Linde et al., 2008, 2006; Kowalsky et al., 2005; Bosch, 2004; Gallardo and Meju, 2003). The quality and consistency of the resulting multi-parameter models thus depends critically on the reliability of these assumptions and on the way they are accommodated in the inversion procedure. Unfortunately, many, if not most, of these parameter relations tend to be non-unique, site-specific, spatially variable, and/or scale-dependent (e.g., Schoen, 1996). Nevertheless, ongoing algorithmic and computational improvements can be expected to further enhance the attractiveness and applicability of joint inversion approaches.

An entirely different class of approaches to quantitative data integration, which we consider in this study, is based on geostatistical principles. In its original definition, geostatistics refers to the interpolation or extrapolation of sparsely sampled data based on the observed or inferred covariance structure of these data. The corresponding techniques are referred to as kriging or, if multiple datasets and their interrelations are considered, co-kriging. For convenience and consistence, we adopt the considerably broader definition of the term “geostatistics” as practiced in hydrocarbon reservoir characterization, which essentially comprises all multi-variate statistical and stochastic techniques that can be used for spatial data analyses (e.g., Deutsch, 2002; Kelkar and Perez, 2002).

Probably the simplest geostatistical approach *sensu stricto* to integrate multiple datasets is through classical co-kriging based on the auto- and cross-covariances of the various parameters and observations (e.g., Gloaguen et al., 2001; Cassiani et al., 1998). This approach can be quite effective for providing smoothed reconstructions of the spatial distributions of the target parameters. An inherent shortcoming is, however, that autocovariance functions of the inferred

models differ substantially from those of the actual data. Moreover, co-kriging assumes the relationships between the various datasets to be linear and unique, which in practice is often not the case.

Bayesian statistics is another approach for the quantitative integration of multiple and/or diverse datasets (e.g., Gelman et al., 2003). This approach requires an initial estimate of the probability distribution of the target parameter, as could be obtained, for example, by kriging or co-kriging of the hydrological database alone. This prior distribution is complemented through estimates of the joint probability distribution of collocated hydrological and geophysical data, which is then used to update the prior distribution into a posterior distribution of the hydrological target parameter given additional constraints provided by the geophysical data. Bayesian approaches for the integration of hydrogeophysical and/or hydrological data have been successfully used in a number of studies (e.g., Chen et al., 2001; Ezzedine et al., 1999). A recent comprehensive review of this topic is provided by Rubin and Hubbard (2005b).

In many cases, an important objective of hydrological and geophysical data integration is to detect relevant zoning within the probed aquifer, for example, to distinguish between gravelly or sandy and clay-rich deposits. An effective approach to achieve this objective is to compare collocated data and use multivariate statistical data classification techniques, such as cluster analysis, support vector engines, or neural networks, to divide them into a lithologically and/or hydrologically meaningful number of groups (e.g., Tronicke et al., 2004; Hoepfner et al., 1999; Hyndman and Harris, 1996; Kaufman and Rousseeuw, 1990). These data groups can then be mapped back into the probed aquifer region to define zones characterized by some mutual correlation of the various inferred parameters as well as by corresponding mean values and standard deviations (e.g., Paasche et al., 2006).

Finally, conditional stochastic simulations aim at finding models that reproduce all available data, constraints and a priori information by randomly drawing from an ensemble of models that fulfill the “conditioning” deterministic and/or stochastic constraints. There are basically two avenues in conditional simulation: sequential Gaussian co-simulations and Monte-

Carlo-type approaches (e.g., Deutsch, 2002; Kelkar and Perez, 2002; Sen and Stoffa, 1995). Sequential Gaussian cosimulations have been successfully used by Hyndman et al. (2000) to integrate crosshole seismic and hydrological data. An important limitation of Gaussian cosimulations is, however, that due to their close relation to the classical co-kriging mentioned above, they are based on the assumption of linear relationships between the various parameters. Conversely, Monte-Carlo-type approaches, which will be discussed in more detail in the following, are characterized by a very large degree of flexibility and allow for accounting for nonlinear and/or nonunique parameter relations.

### Monte-Carlo-Type Conditional Simulations: A Novel Approach Based on Simulated Annealing

Monte-Carlo-type conditional stochastic simulations aim at producing models that reproduce all available data and constraints, as well as the inferred geostatistical characteristics where no data or deterministic constraints are available (e.g., Deutsch, 2002; Kelkar and Perez, 2002; Sen and Stoffa, 1995). Monte Carlo approaches are essentially guaranteed to find the optimal solution but tend to be computationally impractical for multi-method data integration. Instead, directed Monte-Carlo-type approaches, such as genetic algorithms or simulated annealing, which are more efficient but more prone to getting stuck in local minima of the solution space, are being used. While for most intents and purposes genetic algorithms and simulated annealing seem to be largely equivalent, the latter is clearly the preferred approach for conditional stochastic simulations as it is conceptually considerably simpler and easier to parameterize.

The central idea behind simulated annealing is adapted from the thermodynamics of a cooling metallic melt (e.g., Sen and Stoffa, 1995). Atoms can move freely throughout a melt at high temperatures, but as the temperature is lowered, their mobility progressively decreases. Eventually, the system reaches its minimum energy state, and the atoms assume fixed positions within a crystal lattice. In simulated annealing, we therefore begin with a very large number of possible initial states for the target parameter distribution, but during the cooling or annealing process, all

possible states converge to a few acceptable final states. When applying this idea to conditional stochastic simulation, we require a suitable global objective function  $O$  that quantifies the overall closeness of the modeled data to the observed data. A common way to achieve this is to formulate the global objective function as a weighted sum of several different components  $O_i$

$$O = \sum_i w_i O_i \quad (1)$$

where  $w_i$  denote the weights of the various components. A typical objective function could, for example, have three components  $O_1$ ,  $O_2$ , and  $O_3$ :  $O_1$  controlling the reproduction of a specified geostatistical model,  $O_2$  controlling the reproduction of the conditioning borehole data, and  $O_3$  controlling the correlation between the target parameter and the conditioning geophysical data, for example, seismic and/or georadar crosshole tomographic images (e.g., Tronicke and Holliger, 2005). The composite nature of the global objective function is a key characteristic of Monte-Carlo-type optimization approaches and contributes substantially to their flexibility and versatility. That said, it is, however, also important to note that multi-component objective functions of the form given in equation (1) are associated with a number of drawbacks and inconveniences. In our experience, the most important of these are (i) the inherently subjective choice of the weights attributed to the various components and (ii) the increasing width and flatness of the region in the vicinity of the global minimum. Moreover, the computational effort naturally increases with the complexity of the global objective function and the number of its components.

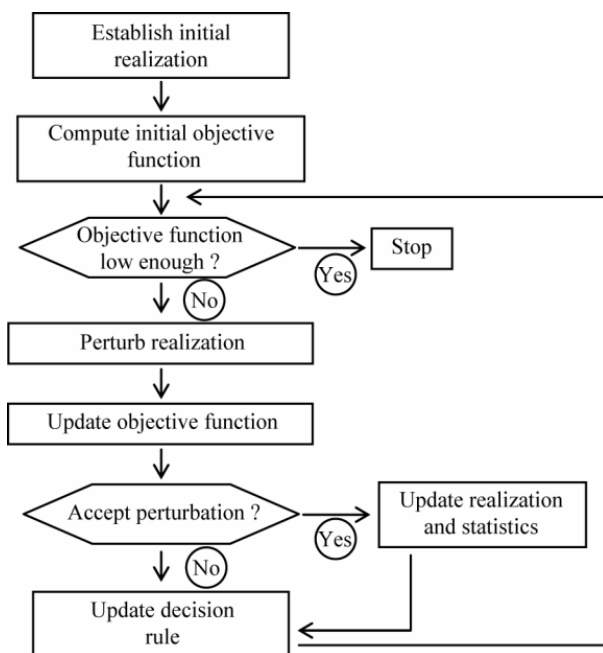
In simulated annealing, the global objective function is then gradually minimized following the thermodynamic analogy mentioned above. In doing so, the acceptance probability  $P$  for a new parameter distribution is calculated as

$$P = \begin{cases} 1, & \text{if } O_{\text{new}} \leq O_{\text{old}} \\ \exp\left(\frac{O_{\text{old}} - O_{\text{new}}}{T}\right), & \text{otherwise} \end{cases} \quad (2)$$

where  $O_{\text{old}}$  and  $O_{\text{new}}$  denote the global objective functions before and after the perturbation of the model parameters, respectively, and  $T$ , in reference to the Boltzmann distribution in thermodynamics, is often referred to as the temperature parameter. Thus,

all favorable perturbations are accepted, whereas unfavorable ones are accepted according to the exponential probability distribution controlled by  $T$ . For deciding whether an unfavorable perturbation is accepted or not, we generate a uniform random number  $0 \leq r \leq 1$  and compare it with  $P$ . The perturbation is accepted if  $r < P$ ; otherwise, it is rejected. Lowering  $T$  results in lower probabilities for the acceptance of unfavorable perturbations (Fig. 1). How and when  $T$  is lowered is controlled by the annealing scheme or “cooling path”, the choice of which may have a significant impact on the computational efficiency and/or convergence characteristics of the simulation process (e.g., Deutsch, 2002; Sen and Stoffa, 1995). Unfortunately, there are no universally valid recipes regarding the optimal choice of starting value of  $T$  and the subsequent cooling path, and hence these parameters have to be constrained empirically for any given problem. Clearly, setting the initial value of  $T$  too high and/or lowering  $T$  too slowly has a negative impact on the computational efficiency, whereas a too low initial value of  $T$  and/or a too rapid lowering of  $T$  will not allow for adequate model perturbations.

Traditionally, simulated-annealing-based conditional simulations start off with a random field of



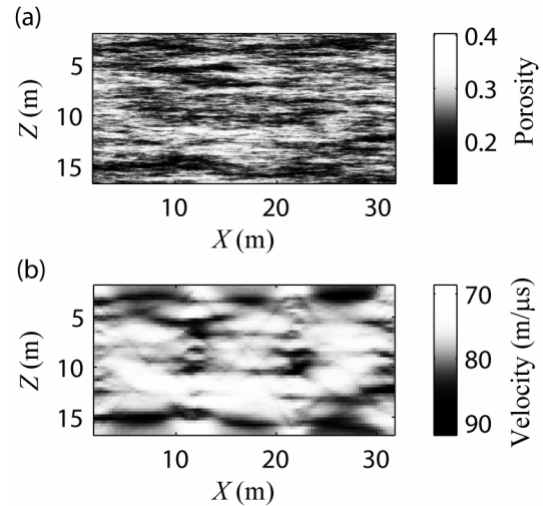
**Figure 1.** Flowchart illustrating the overall concept of simulated annealing for conditional stochastic simulation.

values based on the inferred/assumed probability density function of the target parameter. This initially uncorrelated structure is then gradually “organized” by repeated random swapping of values to meet the structural and/or petrophysical constraints imposed by the available geophysical data. Although this conventional approach is quite robust and flexible, it also suffers from a number of shortcomings. In particular, deterministic information with regard to the larger-scale structure as provided, for example, by geophysical data, is difficult to incorporate into the constraints imposed on the stochastic simulation process, and hence, the lateral continuity tends to be systematically underestimated by the resulting models (e.g., Tronicke and Holliger, 2005). Moreover, such swapping-based approaches tend to be computationally expensive. These problems can be alleviated by (i) linking each value of the larger-scale geophysical models, as provided, for example, by crosshole georadar and/or seismic tomography, to a conditional distribution of the target parameter, as provided, for example, by borehole logging data and (ii) fitting an autocovariance function based on an inferred parametric model only at shorter lags while leaving the structural information contained in the geophysical data at larger lags untouched (Dafflon et al., 2009). It is also important to note that this approach reduces the number of components of the objective function to one, and hence we have  $O = O_1$ , with  $O_1$  denoting the reproduction of the target autocovariance structure at short lags. It should be noted that the choice of the cut-off lag, that is, the lag beyond which we do not constrain the realization based on the parametric covariance model, is based on the estimated resolution of the geophysical image in the vertical and horizontal directions. This, in turn, implies that we naturally avoid all the potential inconveniences and problems associated with a multi-component objective function of the form given by equation (1), such as the question of the weights and the potentially poor definition of the global minimum. Reducing the number of components of the objective function also comes along with a significant reduction of the computational cost by a factor of two to three.

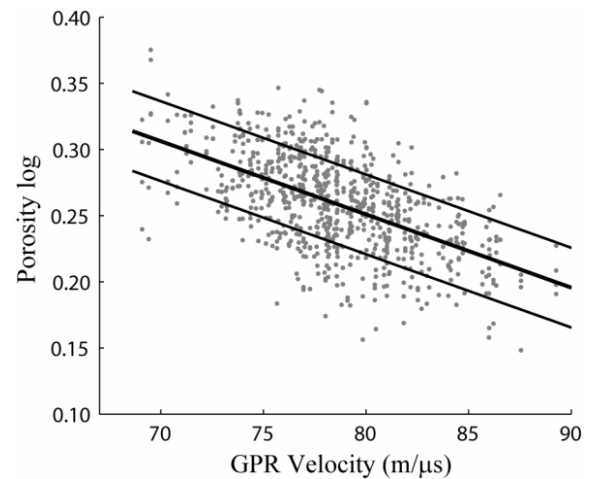
The potential of our approach is illustrated on a realistic porosity model for a heterogeneous alluvial

aquifer, which is characterized by a scale-invariant layered structure with a horizontal-to-vertical aspect ratio of  $\sim 10$  and a power spectrum decaying as  $\sim 1/f$  with  $f$  denoting the spatial frequency (Fig. 2a). This aquifer model represents a particularly challenging test case, as it is statistically non-stationary, exhibiting pronounced structural complexity at both the small and large scales. It is probed by crosshole georadar data and porosity logs with boreholes located at 0, 10, 20, and 30 m lateral distance from the left-hand edge of the model. Synthetic crosshole georadar data are generated using a finite-difference solution of Maxwell's equations (Ernst et al., 2006). After picking the traveltimes of the first arrivals, these data are then tomographically inverted for the spatial distribution of the electromagnetic velocity (Fig. 2b). Together with the synthetic porosity logs, this represents the basis for the reconstruction of the detailed porosity structure using our novel conditional simulation approach outlined above. To this end, we first infer the relation between the porosity logs and the georadar velocity along the boreholes to establish a probability distribution of possible porosity values for each georadar velocity present in the tomogram (Fig. 3). Next, we need to determine the horizontal and vertical target autocovariance functions. We use the porosity logs to infer a parametric covariance model at shorter lags. Assuming that the parametric form of the covariance function is the same in all directions, the horizontal covariance function is then constrained by estimating the structural aspect ratio, that is, the ratio between the horizontal and vertical correlation lengths, from the tomograms. The cut-off lags, below which we constrain the realizations to the covariance model, were set to 2 and 5 m in the vertical and horizontal directions, respectively. The corresponding results are shown in Fig. 4 and compared with porosity models based on simple kriging interpolation of the porosity logs and the tomographic inversion of the crosshole georadar data. The conditional simulation does indeed compare quite favorably with the original porosity model in terms of its small- and large-scale structures and their lateral continuity, as well as with regard to the overall distribution of the porosity values (Figs. 4a and 4d), whereas the models resulting from crosshole georadar tomography alone and kriging interpolation bear only

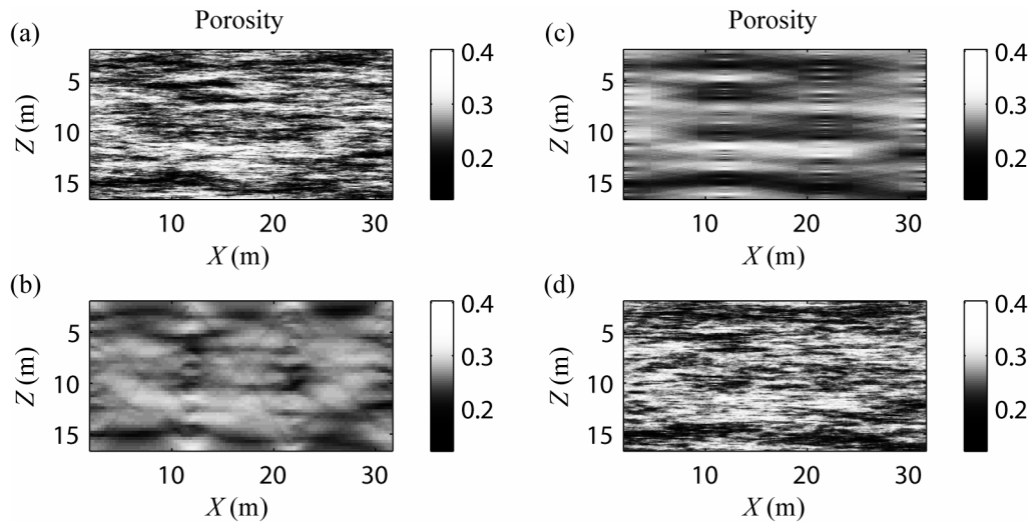
a faint resemblance with the overall character of the original porosity structure and definitely fail to adequately capture its inherent heterogeneity and complexity (Figs. 4a–4c).



**Figure 2. (a) Synthetic porosity model representing a heterogeneous alluvial aquifer; (b) result of tomographic inversion of three synthetic crosshole georadar datasets simulated between boreholes located at 0, 10, 20 and 30 m lateral distance from the left-hand edge of the model.**



**Figure 3. Scatter plot of the near-borehole georadar velocity  $v$  versus porosity  $\phi$  for all four boreholes for our synthetic example shown in Fig. 2. The thicker central line denotes the conditional expectation  $E(\phi|v)$ , estimated assuming a linear relationship and minimizing the squared error. The two finer peripheral lines indicate the conditional standard deviation  $\sigma(\phi|v)$ .**

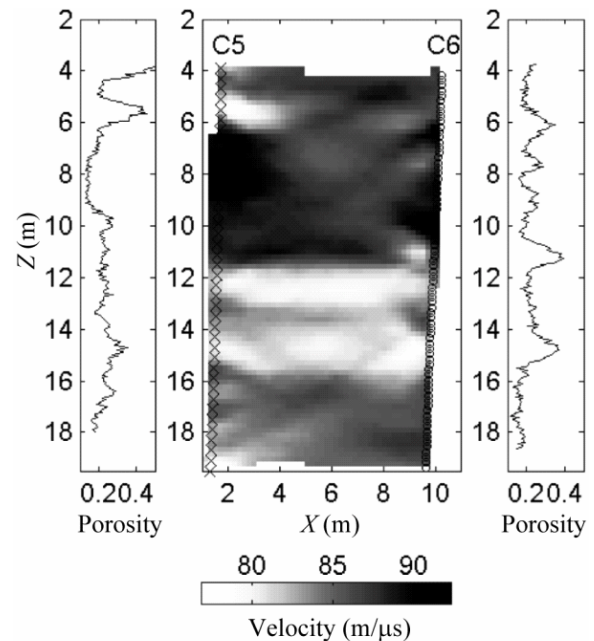


**Figure 4.** (a) Synthetic porosity model from Fig. 2a; (b) conversion of the crosshole georadar velocity image shown in Fig. 2b to porosity; (c) porosity distribution obtained by kriging interpolation of porosity logs; (d) stochastic realization of the porosity distribution constrained by the tomographic georadar data, porosity logs, and target covariance function at short lags. The vertical target autocovariance function, which also served as a basis for the kriging interpolation of the porosity logs, was inferred from the porosity logs and the structural aspect ratio was estimated from the tomographic image.

#### APPLICATION TO OBSERVED DATA

We now show the application of our new simulated-annealing-based conditional stochastic simulation approach to the integration of crosshole georadar and porosity log data collected at the Boise Hydrogeophysical Research Site (BHRS) near Boise, Idaho, USA, with the objective of constraining the detailed porosity distribution in this heterogeneous alluvial aquifer (e.g., Tronicke et al., 2004). The subsurface at the site is characterized by an approximately 20-m-thick alluvial layer consisting predominantly of gravel and sand with minimal fractions of silt and clay and underlain by a layer of red clay with a thickness of at least 3 m. At the time of the geophysical measurements (October 1998), the water table was at a depth of 2.96 m. Figure 5 shows the tomographic inversion of the considered crosshole georadar dataset together with neutron porosity logs acquired along the boreholes (Tronicke et al., 2004). The velocity tomogram is distinguished by predominantly subhorizontal structures, which is consistent with stratigraphic layering in the gravel and sand deposits at the BHRS (Barrash and Clemo, 2002).

The integration of the porosity log data with the crosshole georadar velocity tomograms then

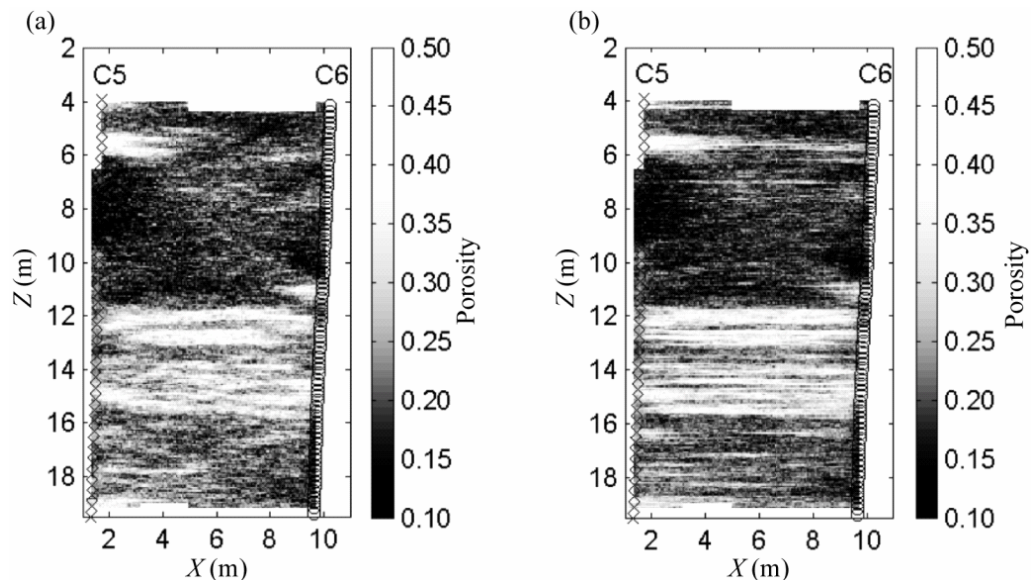


**Figure 5.** Crosshole georadar tomogram and porosity logs acquired at the Boise Hydrogeophysical Research Site (BHRS) near Boise, Idaho.

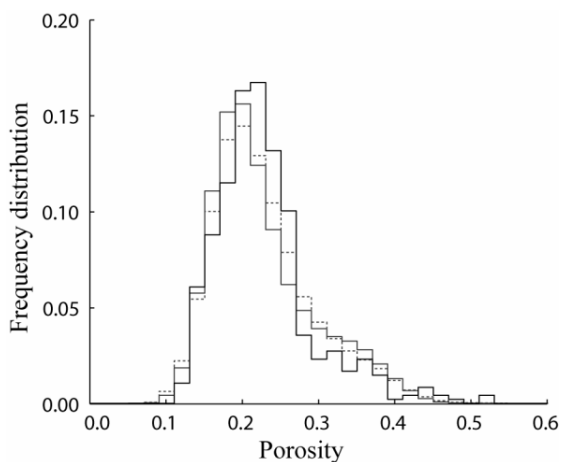
proceeded in the same fashion as outlined above for the synthetic data. Due to lacking constraints regarding the ratio between the horizontal and vertical correlation lengths, we, however, decided to generate conditional stochastic simulations for a range of as-

pect ratios. A selection of the corresponding results is shown in Fig. 6. For each of the realizations shown, the target parameters are well fitted, and the realizations can be considered, in terms of its appearance, as representative of a large number of identically con-

strained conditional stochastic simulations. Moreover, it is important to note that, despite their clearly non-Gaussian nature, the histograms of the porosity logs are faithfully reproduced by those of the resulting porosity models (Fig. 7).



**Figure 6.** Stochastic realizations of the aquifer porosity distribution at BHRS constrained by the georadar tomogram and porosity logs shown in Fig. 5. The parametric form of the target autocovariance function at short lags was inferred from the porosity logs. The only difference between the two realizations is the horizontal-to-vertical aspect ratio, which was assumed to be (a) 7 and (b) 20.



**Figure 7.** Histogram of the porosity log data shown in Fig. 5 (solid black line) and the stochastic models shown in Fig. 6a (dotted grey line) and Fig. 6b (solid grey line).

The general structure of the conditional simulations is also consistent with information that we have about the true structure of the BHRS aquifer. In particular, we have the presence of three major geological

units in depth having low, high, and medium porosities, respectively. In addition, all of the results show a striking degree of structural resemblance with a recently published full-waveform inversion of the crosshole data that have an expected resolution of less than half a meter in both the vertical and horizontal directions (Ernst et al., 2007). Based on this comparison, it appears that the models characterized by intermediate aspect ratios around 10 are likely to be more realistic than those characterized by the end members of the considered range shown in Fig. 6. Albeit circumstantial, this evidence is consistent with information on the ratios of the horizontal-to-vertical correlation lengths in alluvial aquifers compiled by Gelhar (1993), which also tend to cluster around a value of 10.

#### IMPLICATIONS FOR THE PREDICTION OF LOCAL-SCALE FLOW AND TRANSPORT PHENOMENA

Unfortunately, geophysical data are in general



not directly sensitive to hydraulic conductivity, which arguably represents the key parameter of interest to hydrologists. As illustrated by the synthetic and observed data examples above, some geophysical data do, however, have a pronounced sensitivity to the water-saturated porosity. Porosity represents a prerequisite for hydraulic conductivity, and hence, these two key hydrogeological properties are in general systematically and causally linked, albeit on a site-specific basis. Still, more and more commonly, such relations can be reliably established provided that coincident high-resolution hydraulic conductivity and porosity measurements are available. To conclude this study, we explore the potential hydrological value of our novel data integration approach by assuming perfect knowledge of the underlying relation between porosity and hydraulic conductivity.

To this end, we return to the original porosity model (Fig. 4a), the porosity model based on the crosshole georadar tomography alone (Fig. 4b), the kriging interpolation of the porosity logs (Fig. 4c), and the conditional stochastic simulation based on the approach outlined above (Fig. 4d). The relation between porosity and hydraulic conductivity we employed is of the form (e.g., Schoen, 1996)

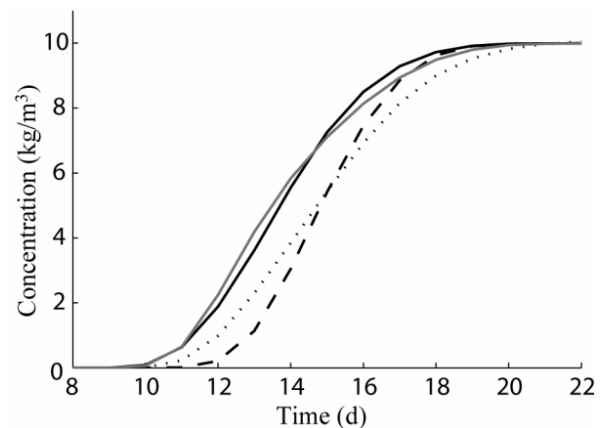
$$^{10} \log K = a + b\phi \quad (3)$$

where  $\phi$  is the porosity;  $K$  is hydraulic conductivity in m/s; and  $a = -4.97$  and  $b = 6.66$  are constants. In practice, such a relation could, for example, be established through flowmeter logging and/or densely sampled slug tests along the boreholes and comparing the resulting hydraulic conductivity profiles with the coincident neutron porosity logs. Alternatively, direct-push permeameter measurements could be acquired anywhere between the boreholes. In this case, reconstructed stochastic porosity models of the type shown in Fig. 4d, or corresponding averages of multiple stochastic realizations, would serve as a basis for establishing a meaningful relationship between porosity and hydraulic conductivity.

After converting the porosity distributions shown in Fig. 4 to their corresponding hydraulic conductivity values, we model propagation of a conservative tracer through these models assuming a constant hydraulic gradient. The tracer is continuously injected along the entire length of the borehole along the left-hand model

edge and recovered in an integrative manner through pumping at the borehole at the right-hand model edge. Figure 8 shows the resulting breakthrough curves corresponding to the tracer concentration recovered as a function of time. We see that the model based on the novel Monte-Carlo-type data integration approach presented in this study does indeed result in a breakthrough curve that is remarkably close to that of the underlying model, whereas the hydraulic conductivity models based on the kriging of borehole logs or crosshole georadar tomography alone produce breakthrough curves that differ markedly from that of the underlying model. It should, however, also be noted that the superior results obtained with Monte Carlo approach come at a computational cost that is roughly three orders-of-magnitude higher than that of the more conventional techniques.

The above results illustrate that, provided that a meaningful relation between the hydrogeophysical target parameter and the hydraulic conductivity can be established, the novel Monte-Carlo-type data integration approach presented in this study does indeed provide a formidable basis for the construction of detailed



**Figure 8. Breakthrough curves for the porosity models shown in Fig. 4 assuming a known and error-free relation between porosity and hydraulic conductivity of the form given in equation (3). Black solid line: underlying stochastic hydraulic conductivity model (Fig. 4a); dashed line: model based on crosshole georadar tomography only (Fig. 4b); dotted line: model based on kriging interpolation of porosity logs (Fig. 4c); grey solid line: model based on Monte-Carlo-type data integration approach presented in the study (Fig. 4d).**

hydrological models that allow for a reliable prediction of contaminant transport at the local scale. In practice, however, the highly resolved hydraulic conductivity measurements required to build a relationship between porosity and hydraulic conductivity of the form given by equation (3) are often not available. In this case, the question arises as to whether we could estimate such a relation based on the porosity model inferred through our Monte-Carlo-type data integration approach (Fig. 4d) and trying to match the observed breakthrough curve (Fig. 8). We have found that this is indeed the case that the estimated coefficients of  $a = -4.75$  and  $b = 5.85$ , inferred through a grid search approach, are indeed very close to those of underlying petrophysical model, and that, thus, the inferred empirical relationship between porosity and hydraulic conductivity was valid for different realizations of the considered stochastic model. The latter implies that as long as there are no fundamental hydrogeological facies changes, the inferred relation between porosity and hydraulic conductivity is likely to hold throughout the entire aquifer. Interestingly, we have found that similarly adequate and generic field relations can also be developed for the porosity distributions inferred from crosshole tomography (Fig. 4b) and kriging interpolation of the borehole logs (Fig. 4c). Having said this, it is important to note that an inherent prerequisite for the successful kriging of borehole data is that the horizontal correlation lengths of the underlying structures are larger than the spacing between the individual boreholes. Here, although the coefficients  $a$  and  $b$  constraining the considered relation between porosity and hydraulic conductivity (equation (3)) differ fundamentally from the underlying petrophysical relation, such “field-scale relations” are found to hold remarkably well for different realizations of the considered stochastic model and thus can also be expected to hold throughout the considered aquifer. On the one hand, these observations point to the inherently diffusive nature of flow and transport phenomena, as well as the integrative nature of conventional hydrological observation techniques. On the other hand, these admittedly preliminary results indicate that, even in the presence of very strong heterogeneity, viable field relations between a pertinent and highly resolved hydrogeophysical target pa-

rameter and hydraulic conductivity can be obtained, which allows for the reliable prediction of contaminant transport at the local scale.

## CONCLUSIONS

We have presented a novel Monte-Carlo-type approach to conditional stochastic simulation based on a simulated annealing concept that demonstrates significant potential for generating highly detailed and realistic aquifer models integrating a variety of hydrogeophysical data and prior information. The major advantage of our approach compared with related previous efforts is that deterministic information with regard to the larger-scale subsurface structure, as provided, for example, by tomographic geophysical images, is incorporated into the resulting realizations in an efficient and effective manner. An additional benefit of this approach is that, because of a dramatically simplified global objective function consisting essentially of the parametric autocovariance model at short lags, our algorithm exhibits very favorable characteristics with regard to convergence and computational efficiency compared to more conventional methods.

Our new algorithm is defined by two key features. First, instead of swapping values within the simulation grid, we perturb the target field during the stochastic simulation procedure by drawing from a probability distribution for the output parameter conditioned to the geophysical data. For the synthetic and observed data sets considered in this study, we estimated these conditional distributions using collocated georadar velocity and porosity measurements at the borehole locations. Clearly, any appropriate alternative analysis technique could be used. Second, instead of constraining the stochastic simulations to match a parametric geostatistical model over a wide range of lags, we release this constraint at large lags and only enforce adherence to the covariance model at short lags that are not resolved by the geophysical data. In doing so, we let drawing from the conditional distribution correctly incorporate the larger-scale deterministic structure contained in the geophysical data.

We found that the deterministic information with regard to the larger-scale subsurface structure, as provided by geophysical data, was properly incorporated into the output realizations, while the smaller-scale

stochastic fluctuations were realistically represented by the parametric autocovariance models employed at short lags. Indeed, it can be argued that the resulting conditional stochastic simulations are characterized by an unprecedented degree of realism at all scales, which in turn is indicative of the fact that the proposed method has the potential of optimally capitalizing on the complementarities of the various data and constraints. The results obtained by applying our new algorithm to both synthetic and observed data sets are quite encouraging and should allow for faithful predictions of hydrological transport phenomena in highly complex and heterogeneous aquifers provided that the relation between porosity and hydraulic conductivity is known and well constrained. Finally, it is important to note that the new conditional stochastic simulation approach presented here is quite generic and highly flexible and hence can be expected to be equally applicable to the quantitative integration of data from a wide range of geophysical, hydrological, and environmental techniques.

When exploring the potential implications arising for the local-scale prediction of flow and transport phenomena in highly heterogeneous media, we found, not surprisingly, that our algorithm provides the greatest benefits to the prediction of flow and transport phenomena when a detailed relation between the hydrogeophysical target parameter and the hydraulic conductivity is available. If such a relation is unavailable, corresponding field relations based on the reproduction of the observed breakthrough curves using more conventional approaches to reconstructing the aquifer structure, such as crosshole seismic or georadar images alone, or provided that the horizontal correlation length is sufficiently large, even simple kriging of closely spaced borehole data may indeed provide surprisingly reliable predictions.

#### ACKNOWLEDGMENTS

This research has been supported by grant from the Swiss National Science Foundation. We thank the Center for the Geophysical Investigation of the Shallow Subsurface (CGISS) at Boise State University for allowing us to work with crosshole georadar and neutron porosity log data collected at the BHRS, Jens Tronicke at the University of Potsdam for comments

and suggestions as well as providing access to the results of earlier work, and Jan van der Kruk and anonymous referee for reviewing this manuscript.

#### REFERENCES CITED

- Barrash, W., Clemo, T., 2002. Hierarchical Geostatistics and Multifacies Systems: Boise Hydrogeophysical Research Site, Boise, Idaho. *Water Resour. Res.*, 38: 1196. DOI: 10.1029/2002WR001436
- Bosch, M., 2004. The Optimization Approach to Lithological Inversion: Combining Seismic Data and Petrophysics for Porosity Prediction. *Geophysics*, 69: 1272–1282
- Cassiani, G., Boehm, G., Vesnaver, A., et al., 1998. A Geostatistical Framework for Incorporating Seismic Tomography Auxiliary Data into Hydraulic Conductivity Estimation. *J. Hydrol.*, 206: 58–74
- Chen, J., Hubbard, S., Rubin, Y., 2001. Estimating the Hydraulic Conductivity at the South Oyster Site from Geophysical Tomographic Data Using Bayesian Techniques Based on the Normal Linear Regression Model. *Water Resour. Res.*, 37: 1603–1613
- Dafflon, B., Irving, J., Holliger, K., 2009. Simulated-Annealing-Based Conditional Simulation for the Local-Scale Characterization of Heterogeneous Aquifers. *J. Appl. Geophys.*, 68: 60–70
- Dannowski, G., Yaramanci, U., 1999. Estimation of Water Content and Porosity Using Combined Radar and Geoelectrical Measurements. *Europ. J. Env. Eng. Geophys.*, 4: 71–85
- Deutsch, C. V., 2002. Geostatistical Reservoir Modeling. Oxford University Press, Oxford
- Ernst, J. R., Holliger, K., Maurer, H. R., et al., 2006. Realistic FDTD Modeling of Borehole Georadar Antenna Radiation: Methodology and Application. *Near Surface Geophysics*, 4: 19–30
- Ernst, J. R., Maurer, H. R., Green, A. G., et al., 2007. Application of a New 2-D Time-Domain Full-Waveform Inversion Scheme to Crosshole Radar Data. *Geophysics*, 72: J53–J64
- Ezzedine, S., Rubin, Y., Chen, J., 1999. Bayesian Method for Hydrogeological Site Characterization Using Borehole and Geophysical Survey Data: Theory and Application to the Lawrence Livermore National Laboratory Superfund Site. *Water Resour. Res.*, 35(9): 2671–2684
- Gallardo, L. A., Meju, M. A., 2003. Characterization of Heterogeneous Near-Surface Materials by Joint 2D Inversion

- of Resistivity and Seismic Data. *Geophys. Res. Lett.*, DOI: 10.1029/2003GL017370
- Garambois, S., Sénéchal, P., Perroud, H., 2002. On the Use of Combined Geophysical Methods to Assess Water Content and Water Conductivity of Near-Surface Formations. *J. Hydrol.*, 259: 32–48
- Gelhar, L. W., 1993. Stochastic Subsurface Hydrology. Prentice-Hall, Englewood Cliffs
- Gelman, A., Carlin, J. B., Stern, H. S., et al., 2003. Bayesian Data Analysis. 2nd ed.. Chapman & Hall, Boca Raton
- Gloaguen, E., Chouteau, M., Marcotte, D., et al., 2001. Estimation of Hydraulic Conductivity of an Unconfined Aquifer Using Cokriging of GPR and Hydrostratigraphic Data. *J. Appl. Geophys.*, 47: 135–152
- Hoepfner, F., Klawonn, F., Kruse, R., et al., 1999. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. Wiley, Chichester
- Hubbard, S. S., Rubin, Y., 2005. Introduction to Hydrogeophysics. In: Rubin, Y., Hubbard, S. S., eds., Hydrogeophysics. Springer, Dordrecht. 3–21
- Hubbard, S. S., Chen, J., Peterson, J., et al., 2001. Hydrogeological Characterization of the South Oyster Bacterial Transport Site Using Geophysical Data. *Water Resour. Res.*, 37: 2431–2456
- Hyndman, D. W., Harris, J. M., 1996. Traveltime Inversion for the Geometry of Aquifer Lithologies. *Geophysics*, 61: 1728–1737
- Hyndman, D. W., Harris, J. M., Gorelick, S. M., 2000. Inferring the Relation between Seismic Slowness and Hydraulic Conductivity in Heterogeneous Aquifers. *Water Resour. Res.*, 36: 2121–2132
- Kaufman, L., Rousseeuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York
- Kelkar, M., Perez, G., 2002. Applied Geostatistics for Reservoir Characterization. Society of Petroleum Engineers
- Kirsch, R., 2006. Groundwater Geophysics. Springer, Berlin
- Kowalsky, M., Finsterle, S., Peterson, et al., 2005. Estimation of Field-Scale Soil Hydraulic and Dielectric Parameters through Joint Inversion of GPR and Hydrological Data. *Water Resour. Res.*, 41: W11425. DOI: 10.1029/2005WR004237
- Linde, N., Finsterle, S., Hubbard, S., 2006. Inversion of Tracer Test Data Using Tomographic Constraints. *Water Resour. Res.*, 42: W04410. DOI: 10.1029/2004WR003806
- Linde, N., Tryggvason, A., Peterson, J. E., et al., 2008. Joint Inversion of Crosshole Radar and Seismic Traveltimes Acquired at the South Oyster Bacterial Transport Site. *Geophysics*, 73(4): G29–G37
- Paasche, H., Tronicke, J., Holliger, K., et al., 2006. Integration of Diverse Physical-Property Models: Subsurface Zonation and Petrophysical Parameter Estimation Based on Fuzzy C-Means Cluster Analyses. *Geophysics*, 71: H33–H44
- Rubin, Y., Hubbard, S. S., 2005a. Hydrogeophysics. Springer, Dordrecht
- Rubin, Y., Hubbard, S. S., 2005b. Stochastic Forward and Inverse Modelling: The “Hydrogeophysical” Challenge. In: Rubin, Y., Hubbard, S. S., eds., Hydrogeophysics. Springer, Dordrecht. 487–511
- Schoen, J. H., 1996. Physical Properties of Rocks: Fundamentals and Principles of Petrophysics. Pergamon, Oxford
- Sen, M., Stoffa, P. L., 1995. Global Optimization Methods in Geophysical Inversion. Elsevier, Amsterdam
- Szerbiak, R. B., McMechan, G. A., Corbeanu, R. M., et al., 2001. 3-D Characterization of a Clastic Reservoir Analog: From 3-D GPR Data to a 3-D Fluid Permeability Model. *Geophysics*, 66: 1026–1037
- Tronicke, J., Holliger, K., Barrash, W., et al., 2004. Multivariate Analysis of Cross-Hole Georadar Velocity and Attenuation Tomograms for Aquifer Zonation. *Water Resour. Res.*, DOI: 10.1029/2003WR002031
- Tronicke, J., Holliger, K., 2005. Quantitative Integration of Hydrogeophysical Data: Conditional Geostatistical Simulation for Characterizing Heterogeneous Alluvial Aquifers. *Geophysics*, 70: H1–H10
- Vereecken, H., Binley, A., Cassiani, G., et al., 2006. Applied Hydrogeophysics. Springer, Dordrecht