Hydrogeophysical data integration at larger scales: Application of Bayesian sequential simulation for the characterization of heterogeneous alluvial aquifers

Paolo Ruggeri[1*], James Irving[1], Erwan Gloaguen[2], René Lefebvre[2], and Klaus Holliger[1]

[1] Applied and Environmental Geophysics Group, Center for Research of the Terrestrial Environment, University of Lausanne, Switzerland

[2] Institut National de la Recherche Scientifique, Centre Eau, Terre et Environnement, Quebec City, Canada

[*] Corresponding author: paolo.ruggeri@unil.ch

Draft for submission to *The Leading Edge*

May 8, 2013

1. INTRODUCTION

Knowledge of the spatial distribution of hydraulic conductivity (K) within an aquifer is critical for reliable predictions of solute transport and the development of effective groundwater management and/or remediation strategies. While core analyses and hydraulic logging can provide highly detailed information, such information is inherently localized around boreholes that tend to be sparsely distributed throughout the aquifer volume. Conversely, larger-scale hydraulic experiments like pumping and tracer tests provide relatively low-resolution estimates of K in the investigated subsurface region. As a result, traditional hydrogeological measurement techniques contain a gap in terms of spatial resolution and coverage, and they are often alone inadequate for characterizing heterogeneous aquifers. Geophysical methods have the potential to bridge this gap. The recent increased interest in the application of geophysical methods to hydrogeological problems is clearly evidenced by the formation and rapid growth of the domain of hydrogeophysics over the past decade (e.g., Rubin and Hubbard, 2005).

Perhaps the greatest challenge in using geophysical measurements in a hydrogeological context is the fact that the underlying physical properties governing nearly all geophysical responses do not exhibit any straightforward link with the hydraulic conductivity. Indeed, petrophysical relationships between K and geophysical parameters such as the electrical conductivity or seismic wave velocity are notoriously difficult to establish, and are most often site, scale, and/or facies specific. To deal with this issue, a number of strategies have been proposed for local-scale (~10 to 50 m) aquifer characterization, involving a combination of borehole and high-resolution crosshole tomographic geophysical methods (e.g., Hyndman and Gorelick, 1996; Chen *et al.*, 2001; Singha and Gorelick, 2005; Paasche *et al.*, 2006; Dafflon *et al.*, 2009). How to effectively utilize geophysical methods for larger-scale (>100 m) hydrogeological characterization, however, remains a major challenge. In this case, the large

domain to be characterized, the lack of closely spaced boreholes for effective crosshole tomographic imaging, and a scarcity of hydraulic test data make direct application of the established local-scale approaches problematic. What is critically needed are practical methodologies for the medium-to-regional-scale integration of geophysical and hydrological data to understand subsurface hydrogeological heterogeneity. Such methodologies must allow for the assessment of uncertainty in the results obtained because of the strong lack of information that exists at such scales as compared to local-scale studies.

Geostatistics provides a proven framework for integrating diverse sources of information for the purpose of characterizing spatial heterogeneity. Geostatistical methods are routinely applied to large regional-scale parameter fields in the petroleum and mining industries, and they naturally lend themselves to the assimilation of data having different degrees of resolution and hardness. In addition, through conditional simulation, such methods allow for an assessment of model parameter and prediction uncertainty. In this paper, we summarize recent work on the integration of hydrological and geophysical data at larger, flow-relevant scales through the application of a geostatistical technique known as Bayesian sequential simulation (Ruggeri *et al.*, 2013). The overall objective of this work is to generate, in a computationally efficient manner, conditional stochastic realizations of the hydraulic conductivity field that allow for assessment of flow and transport uncertainty over medium-to-regional-scale distances.

2. METHODOLOGY

2.1. Bayesian sequential simulation

The Bayesian sequential simulation (BSS) technique was originally developed by Doyen and Boer (1996) for the interpolation and extrapolation of lithological data. The goal of this method is to generate multiple feasible realizations of the spatial distribution of some variable

of interest, referred to as the primary variable, conditional to (i) measurements of a secondary variable, which are available extensively throughout subsurface and are statistically related in some way to the primary variable, and (ii) a smaller number of generally sparsely distributed measurements of the primary variable. As with all geostatistical sequential simulation approaches, the generation of each stochastic realization is accomplished iteratively, whereby previously simulated values for the primary variable at points along a randomly chosen path through the model space are treated as known "data" when simulating the primary variable at subsequent points (Goovaerts, 1997; Deutsch, 2002).

The following parameterization of Bayes' theorem forms the basis for the BSS technique:

$$p(A_n | B_n, A_1, ..., A_{n-1}) = c \cdot p(B_n | A_n) \cdot p(A_n | A_1, ..., A_{n-1}), \quad (1)$$

where A and B denote the primary and secondary variables, respectively, $p(\bullet)$ denotes a probability distribution, and c is a normalization constant. The conditional distribution $p(A_n | A_1, ..., A_{n-1})$ represents the prior for the primary variable in a chosen cell n in the model space. This prior is conditional to the measured and previously simulated values of the primary variable in cells 1 through n-1, and is obtained by kriging of these values to yield a Gaussian mean and variance at the chosen location. The distribution $p(B_n | A_n)$ is the likelihood function, which expresses the range of values for the primary variable in cell n that is consistent with a particular measured value of the secondary variable at the same location. Finally, the distribution $p(A_n | B_n, A_1, ..., A_{n-1})$ represents the posterior for the primary variable in cell n, which represents an updated state of knowledge that takes into account both the prior information and likelihood.

For our work, we determine the likelihood function $p(B_n | A_n)$ by first estimating the joint probability density for the primary and secondary variables $p(A,B)$. This is accomplished

using collocated measurements of these variables under the assumption that, within a given hydrological unit, the relationship between them is statistically stationary and thus does not depend on the chosen cell location. To this end, we calculate $p(A,B)$ using a non-parametric density estimation approach (Silverman, 1986):

$$p(A,B) = \frac{1}{N l_1 l_2} \sum_{i=1}^{N} k\left(\frac{A - A_i}{l_1}\right) k\left(\frac{B - B_i}{l_2}\right), \tag{2}$$

where N is the number of collocated measurements, $k(\bullet)$ is a positive kernel density function, l_1 and l_2 denote the chosen kernel bandwidths for the primary and secondary variables, respectively, and A_i and B_i are the collocated data. Uncertainties in the secondary variable can be accounted for in determining $p(B_n \mid A_n)$ by taking a weighted sum of the marginal distributions from $p(A,B)$. Full details can be found in Ruggeri et al. (2013).

Multiplying the likelihood function with the prior distribution yields the posterior probability for cell n. Within the framework of sequential simulation, a value for the primary variable can be drawn from this posterior distribution and treated as a known or reference value in subsequent iterations of the procedure involving different cells until all unknowns have been simulated. Multiple stochastic realizations of the primary variable can then be generated by repeating the entire process. It is important to note that the BSS procedure as described above is highly flexible with regard to the relationship that exists between the primary and secondary variables, in the sense that this relationship is estimated empirically based on collocated data, and the quality of the relationship is thus reflected in the variability of the stochastic realizations.

2.2. Application to aquifer characterization

In hydrogeophysical studies at the medium-to-regional scale, it is relatively common to have access to locally highly resolved but spatially sparse borehole logs of a variety of geophysical

parameters and the hydraulic conductivity, as well as to surface-based and/or airborne regional geophysical data having extensive spatial coverage but with greatly reduced spatial resolution (e.g., Goldman *et al.*, 2005). Motivated by the availability of such a database, we have developed an aquifer characterization approach based on two applications of the BSS procedure outlined above.

In the first step of our approach, we use the high-resolution borehole geophysical measurements (primary variable) and low-resolution geophysical parameter estimates (secondary variable) to generate fine-scale realizations of the underlying geophysical parameter field. The aim of this step to quantify the uncertainty in the fine-scale field conditional to these two data sets, thus effectively downscaling the low-resolution geophysical parameter estimates. Note that, for inclusion into the BSS procedure, the low-resolution geophysical data are considered in the form of an already-inverted tomographic image, which can be regarded as a set of uncertain measurements of the spatially averaged "true" geophysical parameter. The likelihood function is estimated from values in this image that are collocated with the high-resolution measurements at the borehole locations.

In the second step of our characterization approach, we again perform BSS, but this time based on the borehole measurements of the hydraulic conductivity (primary variable) and point-by-point statistics of the high-resolution geophysical parameter field derived from the realizations obtained above (secondary variable) in order to generate high-resolution K realizations. In other words, after stochastically downscaling the low-resolution geophysical measurements, we aim in this step to use the geophysical information to condition the fine-scale hydraulic conductivity field. In this case, the likelihood function is determined from the collocated borehole measurements of K and the geophysical parameter.

Although the discussion above has been general with regard to the low- and high-resolution geophysical data considered, it is important to note that arguably the most effective

and common geophysical parameter in large-scale hydrogeological investigations is the electrical conductivity σ, which can be readily constrained through borehole logs at the local scale and through the inversion of geoelectric and/or electromagnetic survey data at the medium-to-regional scale (Goldman *et al.*, 2005; Siemon *et al.*, 2009). With reference to σ as the governing geophysical parameter, Figure 1 is a flowchart summarizing our two-step BSS-based aquifer characterization approach. Details regarding its application will become clear through the following synthetic and field examples.

3. SYNTHETIC DATA EXAMPLE

3.1. Subsurface model and data

We first show the application of our approach to a synthetic data example where there exist high-resolution borehole logs of the hydraulic and electrical conductivities, as well as low-resolution estimates of σ obtained from the inversion of surface-based geoelectrical data. Figure 2a shows the "true" heterogeneous K field considered for this example. The field is 240 m long by 20 m deep and is discretized on a 0.20 m grid, yielding a total number of 120,000 model parameters. It was generated geostatistically assuming an exponential variogram for $\log_{10}(K)$ having horizontal and vertical correlation lengths of 27 m and 2.7 m, respectively, and can be regarded as a realistic first-order abstraction of many surficial alluvial aquifers. The aquifer sediments are assumed to be fully saturated. In Figure 2b, we show the corresponding "true" distribution of the electrical conductivity. This was obtained by first simulating the spatial distribution of porosity ϕ throughout the subsurface region assuming a linear relation with $\log_{10}(K)$ of the form

$$\log_{10}(K) = 6.66\phi - 4.97 \quad . \tag{3}$$

After adding correlated zero-mean Gaussian random noise to simulate more realistic conditions, σ was then calculated using Archie's (1942) equation for saturated media

$$\sigma = \sigma_w \phi^m , \tag{4}$$

where σ_w = 43 mS/m and m = 1.4 were used as typical values for the pore water conductivity and Archie cementation exponent in alluvial sediments, respectively (Schön, 2004). Again, correlated random noise was added to the obtained conductivity field, resulting in a significant amount of uncertainty in the overall relationship between σ and K throughout the model domain.

Having specified the detailed distribution of the hydraulic and electrical conductivities, we next simulated the acquisition of high-resolution σ and K measurements along four boreholes located at lateral positions of 0, 80, 160, and 240 m (Figure 2). The vertical resolution of the borehole measurements was set equal to one grid cell, or 0.2 m. Surface-based geoelectrical measurements were then simulated over the model domain using a dipole-dipole acquisition geometry with an electrode spacing of 2.4 m. After adding 5% uncorrelated noise to the resulting apparent resistivity values, they were tomographically inverted on a coarse grid having a horizontal discretization of 2.4 m and a vertical discretization ranging from 0.8 to 3.2 m with increasing depth. The inversion was carried out using least-squares data fitting and model smoothness constraints. A comparison of the resulting image (Figure 2c) with the "true" σ structure in Figure 2b clearly illustrates the noticeable smoothing and decrease in model resolution with increasing depth that are typical of surface-based tomographic geoelectrical reconstructions. To estimate the uncertainty in the inverted values, we used the method of Alumbaugh and Newman (2000) based on the diagonal elements of the model covariance matrix. The estimated uncertainties, expressed as a percentage of the corresponding σ values, are shown in Figure 2d.

3.2. Data integration procedure

In the first step of our large-scale data integration procedure, BSS was used to generate multiple high-resolution realizations of the σ field that are consistent with the borehole electrical conductivity logs and low-resolution surface-based geoelectrical image. Figure 3a shows the joint distribution of high- and low-resolution electrical conductivity estimated from the collocated measurements of these variables at the borehole locations. Uncertainty associated with the tomographic estimates (Figure 2d) was accounted for during the evaluation of the likelihood function, thus ensuring that the realizations generated through BSS were not too strongly conditioned by the tomographic image in regions poorly informed by the data. In Figure 4, we show nine random σ realizations that were obtained. A comparison of these realizations with the "true" field indicates that the BSS procedure has provided reasonable estimates of both the local- and large-scale electrical conductivity structure, in the sense that the overall large-scale patterns are well reproduced and the style of small-scale variability has been adequately represented. The realizations are tied to the logging measurements at the borehole locations, and they are less constrained at depth where the tomographic image is less reliable.

Next, using the collocated, high-resolution borehole logs of the hydraulic and electrical conductivities as well as the point-by-point statistics of the ensemble of downscaled σ realizations, a second application of BSS was performed to generate high-resolution stochastic realizations of the hydraulic conductivity field. Figure 3b shows joint distribution of σ and K that was estimated from the borehole data. Nine random K realizations that were obtained are shown in Figure 5. The results are largely comparable with those observed for the electrical conductivity, and suggest that the two-step BSS procedure has properly captured the "true" statistics of the underlying K field. This is confirmed in Figures 6a and 6b, which

compare the histogram and vertical experimental variogram for the borehole hydraulic conductivity measurements with those calculated globally from 20 stochastic *K* realizations.

To further validate the characterization results, we simulated the transport of a conservative tracer through the original *K* field in Figure 2a and compared the behavior with transport predicted through 20 stochastic realizations. For the simulations, steady-state groundwater flow was assumed with no-flow boundary conditions at the top and bottom of the model domain and fixed head conditions on each side, providing a lateral hydraulic gradient of 0.013. The transient advection-dispersion equation was solved assuming a fixed tracer concentration in the left-most borehole. The values for the longitudinal and transverse dispersivity in each grid cell were specified to be 0.02 and 0.002 m, respectively. Measurements of the average tracer concentration in the right-most borehole as a function of time were simulated. Figure 6c shows the tracer breakthrough curves that were obtained. Quite importantly, we see that the first-arrival times for the tracer predicted by our stochastic realizations are in good agreement with the "true" tracer breakthrough time, which suggests that our large-scale data integration procedure allows for adequate predictions of this key characteristic of the transport behavior and its uncertainty. We do, however, observe a gradual deterioration in the match between the curves as time increases. Analysis has shown that this is because the obtained *K* realizations do not adequately capture the poorly hydraulically conductive zone located in the lower part of the model between lateral distances of 160 and 240 m (Figure 2a). Please see Ruggeri *et al.* (2013) for further details.

4. FIELD STUDY

4.1. Field site and available data

We now show the application of our proposed characterization approach to field data. The considered study region is located in the municipality of Saint-Lambert-de-Lauzon,

approximately 40 km south of Québec City, Canada (Figure 7). It consists of a ~12 km^2 surficial aquifer upon which there exists a sanitary landfill. The aquifer is semi-confined and consists of Quaternary sediments ranging in thickness from 5 to 20 m. The water table is located at a depth of ~1-2 m. A detailed description of the surficial geology of the site can be found in Paradis *et al.* (2011).

The data considered for our analysis were acquired in the eastern part of the study region and, as in our synthetic example, consist of low-resoluiton σ estimates inferred from surface-based geoelectrical measurements along with borehole measurements of σ and K. The geoelectrical data were collected using a dipole-dipole acquisition geometry and an electrode spacing of 2 m. Figure 8a shows the corresponding tomogram, which is 267 m long by 9 m deep and discretized on a 1 m grid, whereas Figure 8b shows the estimated uncertainties. The borehole σ measurements were acquired at five locations along the considered section using a cone penetrometer system (Figure 8a). Collocated flowmeter measurements of K were only available in one of these boreholes (CPT1). However, these data were kept for model validation purposes and were not used in the characterization effort. Instead, collocated measurements of σ and K at nearby off-profile locations were used to establish the joint distribution between these parameters.

4.2. Data integration procedure

As in the synthetic example, BSS was used to first stochastically downscale the tomographic electrical conductivity estimates, and then to generate realizations of the hydraulic conductivity field. Figures 9 and 10 show ten stochastic realizations obtained for σ and K, respectively. As expected, the electrical conductivity realizations appear to honor the large-scale structure imposed by the geoelectrical image along with the fine-scale structure provided by the borehole measurements. With regard to K, the realizations were intentionally

not conditioned to the data from borehole CPT1, such that these data could be used to assess the different stochastic realizations obtained. In Figure 11, we compare the K profile measured along CPT1 with a single prediction provided by one of the stochastic realizations, as well as the overall statistics regarding K versus depth at this location from a collection of 1000 realizations. We see that the key characteristics of the small-scale variability have been adequately captured, and that the measured K profile is well within the uncertainties expressed by the realizations. Note that the variability between the different K realizations in Figure 10 is significantly higher than that seen in our synthetic example (Figure 5). This results because the empirical relationship established between σ and K based on the high-resolution borehole measurements exhibits greater uncertainty for the case of the field data, as well as because no borehole measurements of K along the profile were used to condition the output stochastic realizations.

5. CONCLUSIONS

We have developed a two-step aquifer characterization procedure based on Bayesian sequential simulation that is targeted to the common case where there exist low-resolution, but spatially exhaustive, surface-based and/or airborne geophysical measurements over a large region and high-resolution, but spatially sparse, borehole measurements of the governing geophysical parameter and the hydraulic conductivity. Testing on a realistic synthetic aquifer model as well as field data has indicated that, given adequate prior information, the proposed approach allows for faithful estimates of the large-scale hydraulic conductivity structure and reliable characterization of solute transport. Future work will explore the application of this methodology to even larger domains, as well as 3D scenarios.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Swiss National Science Foundation.

REFERENCES

Alumbaugh, D.L. & Newman, G.A., 2000. Image appraisal for 2-D and 3-D electromagnetic inversion, *Geophysics,* 65**,** 1455-1467.

Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics, *Trans. Americ. Inst. Mineral. Met.,* 146**,** 54-62.

Chen, J., Hubbard, S.S. & Rubin, Y., 2001. Estimating the hydraulic conductivity at the South Oyster site from gephysical tomographic data using Bayesian techniques based on normal linear regression model, *Water Resources Research,* 37**,** 1603-1613.

Dafflon, B., Irving, J. & Holliger, K., 2009. Use of high-resolution geophysical data to characterize heterogeneous aquifers: Influence of data integration method on hydrological predictions, *Water Resources Research,* 45**,** W0940.

Deutsch, C., 2002. *Geostatistical reservoir modelling,* Oxford University Press, New York.

Doyen, P.M. & Boer, D.L.D., 1996. Bayesian sequantial Gaussian simulation of lithology with non-linear data, United States Patent. Western Atlas International, Inc.

Goldman, M., Gvirtzman, H., Meju, M. & Shtivelman, V., 2005. Hydrogeophysical case studies at the regional scale. *in Hydrogeophysics*, pp. 361-389, eds. Rubin, Y. & Hubbard, S. Springer, Dordrecht.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation,* Oxford University Press, New York.

Hyndman, D.W. & Gorelick, M., 1996. Estimating lithologic and transport properties in three dimension using seismic and tracer data: The Kesterson aquifer, *Water Resources Research,* 32**,** 2659-2670.

Paasche, H., Tronicke, J., Holliger, K., Green, A.G. & Maurer, H., 2006. Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses, *Geophysics,* 71**,** H33-H44.

Paradis, D., Lefebvre, R., Morin, R.H. & Gloaguen, E., 2011. Permeability profiles in granular aquifers using flowmeters in direct-push wells, *Ground Water,* 49**,** 534-547.

Rubin, Y. & Hubbard, S., 2005. *Hydrogeophysics,* Springer, Dordrecht.

Ruggeri, P., Irving, J., Gloaguen, E. & Holliger, K., 2013. Regional-scale integration of multi-resolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach, *Geophysical Journal International*.

Schön, J.H., 2004. *Physical properties of rocks: Fundamentals and principles of petrophysics,* Elsevier, Oxford.

Siemon, B., Christiansen, A.V. & Auken, E., 2009. A review of helicopter-borne electromagnetic methods for groundwater exploration, *Near Surface Geophysics,* 7**,** 629-646.

Silverman, B.W., 1986. *Density estimation for statistics and data analysis,* Chapman and Hall, New York.

Singha, K. & Gorelick, M., 2005. Saline tracer visualized with three-dimensional electrical resistivity tomography: Field-scale spatial moment analysis, *Water Resources Research,* 41**,** W05023.

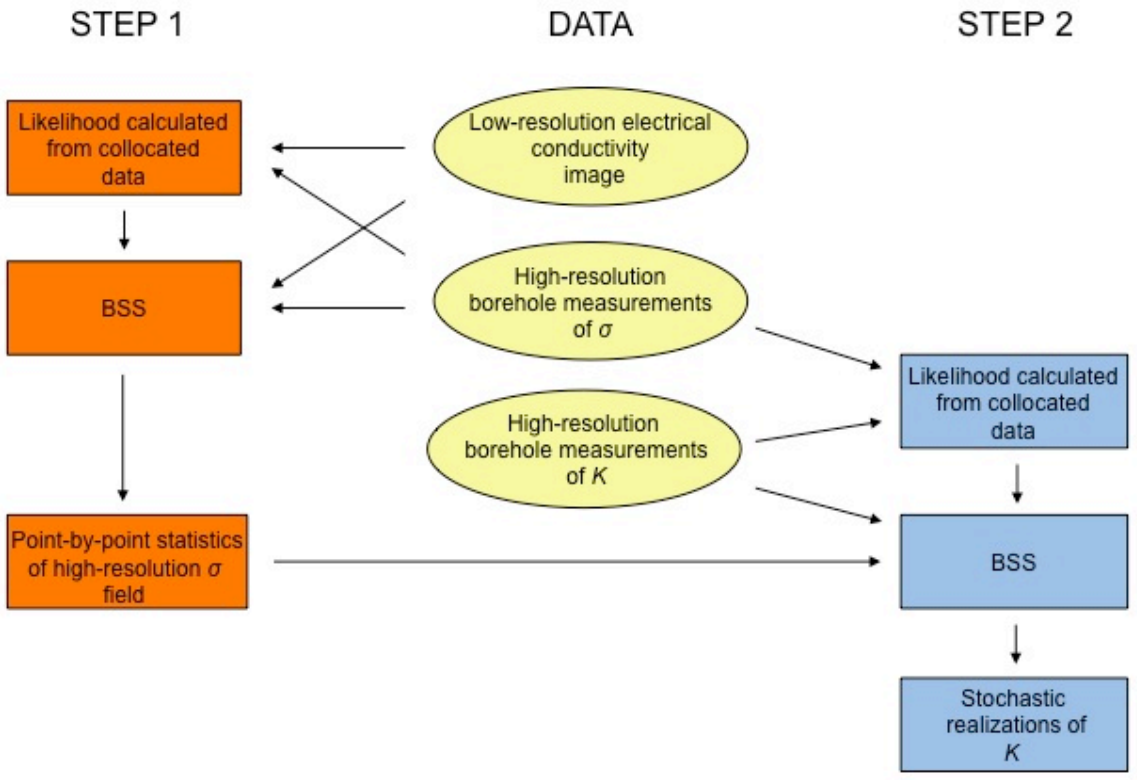

Figure 1: Schematic outline of the two-step Bayesian sequential simulation (BSS) aquifer characterization procedure.

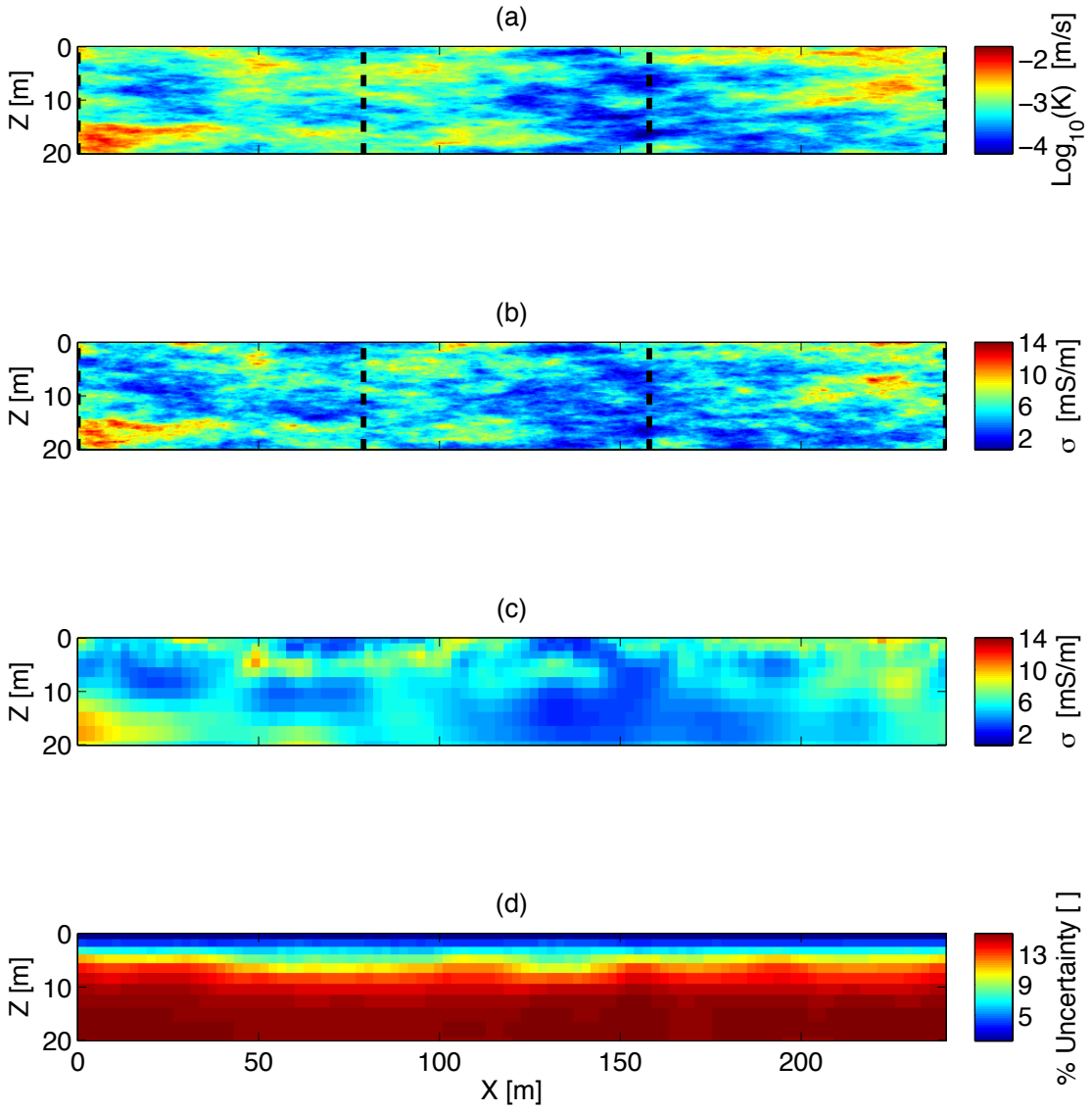

Figure 2: (a and b) "True" spatial distribution of the hydraulic and electrical conductivities considered for the synthetic example, respectively. (c) Electrical conductivity tomogram obtained from the inversion of surface-based geoelectrical measurements simulated over (b). (d) Estimated percentage uncertainty in the values in (c).

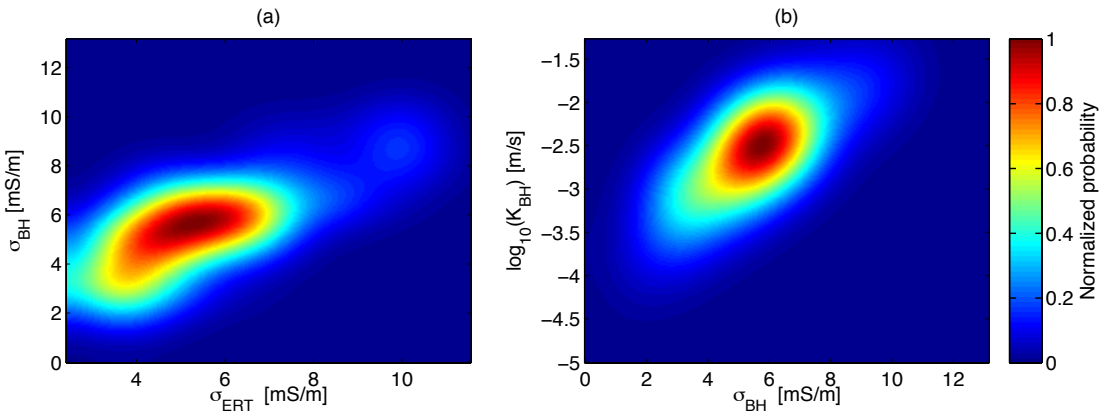

Figure 3: (a) Joint probability distribution estimated between the low- and high-resolution electrical conductivity from collocated data at the borehole locations. (b) Joint probability distribution estimated between the high-resolution electrical and hydraulic conductivities from borehole logs.

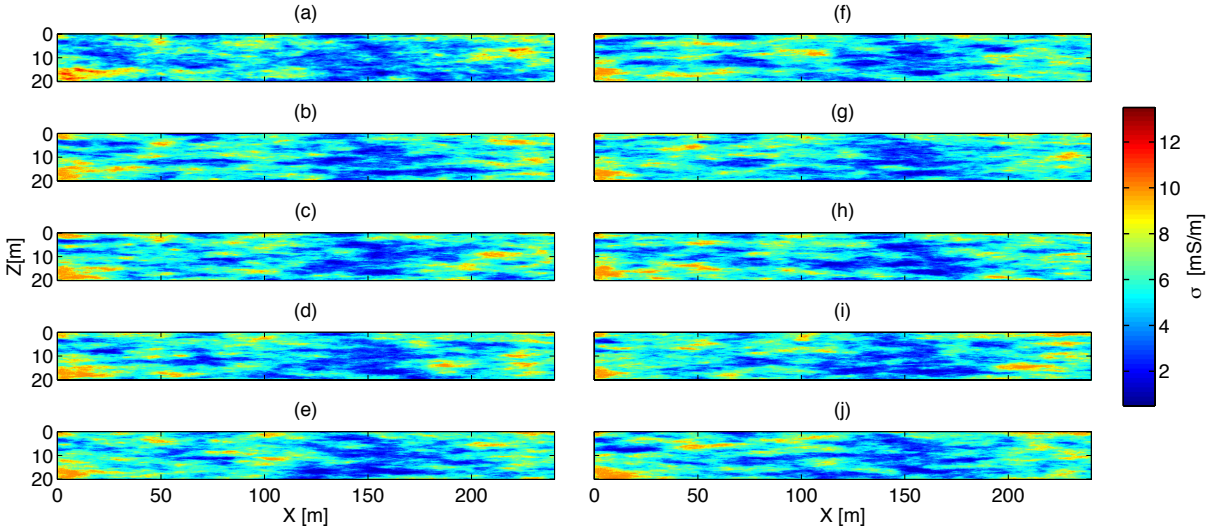

Figure 4: (a) "True" electrical conductivity distribution from Figure 2b. (b-j) Nine stochastic realizations of the electrical conductivity field obtained using Bayesian sequential simulation. The realizations are conditioned to the geoelectrical inversion results in Figure 2 (c and d) and to the high-resolution σ measurements at the four borehole locations.

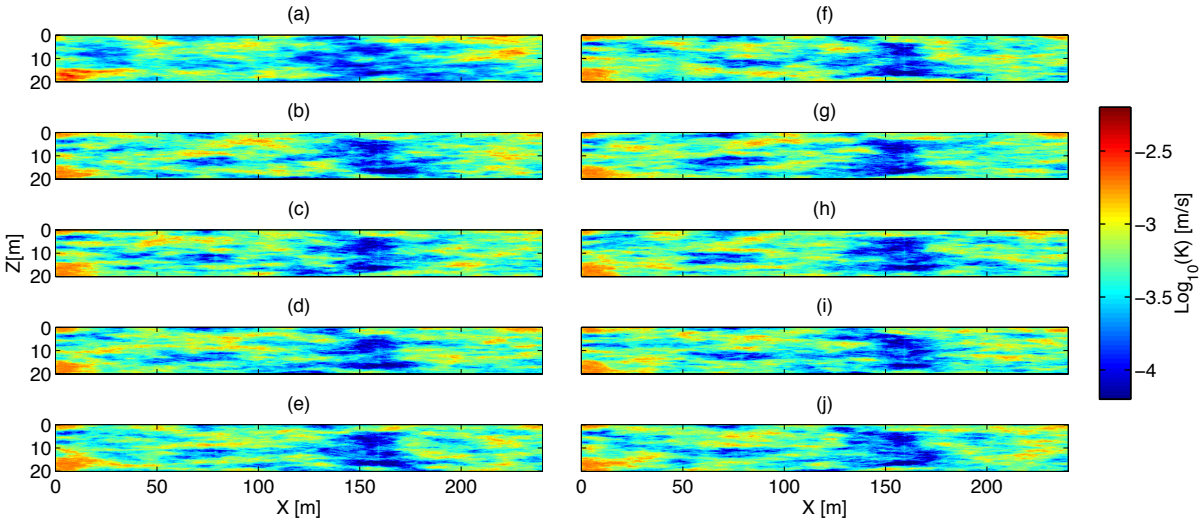

Figure 5: (a) "True" hydraulic conductivity distribution from Figure 2a. (b-j) Nine stochastic realizations of the hydraulic conductivity field obtained using Bayesian sequential simulation. The realizations are conditioned to the point-by-point statistics of the high-resolution electrical conductivity field (Figure 4) and to the measurements of σ and K at the four borehole locations.

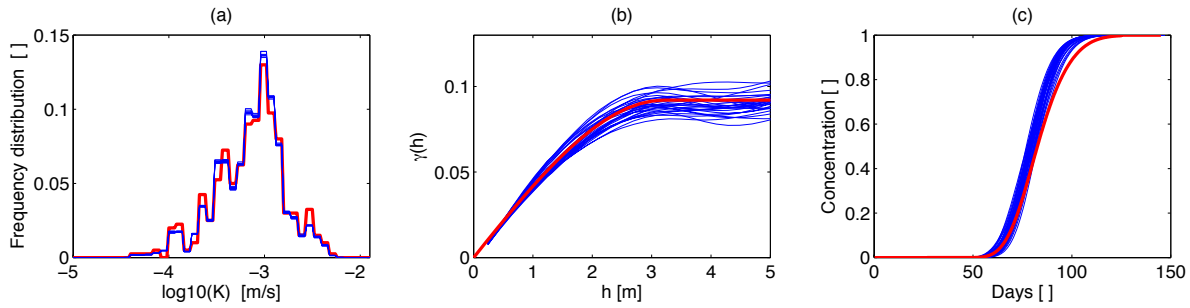

Figure 6: (a) Histogram and (b) experimental vertical variogram for the *K* measurements at the borehole locations (red) versus those corresponding to 20 stochastic realizations obtained using Bayesian sequential simulation (blue). (c) Breakthrough curves showing the normalized average tracer concentration in the right-hand borehole as a function of time for the "true" *K* field (red), and for the 20 stochastic realizations (blue).

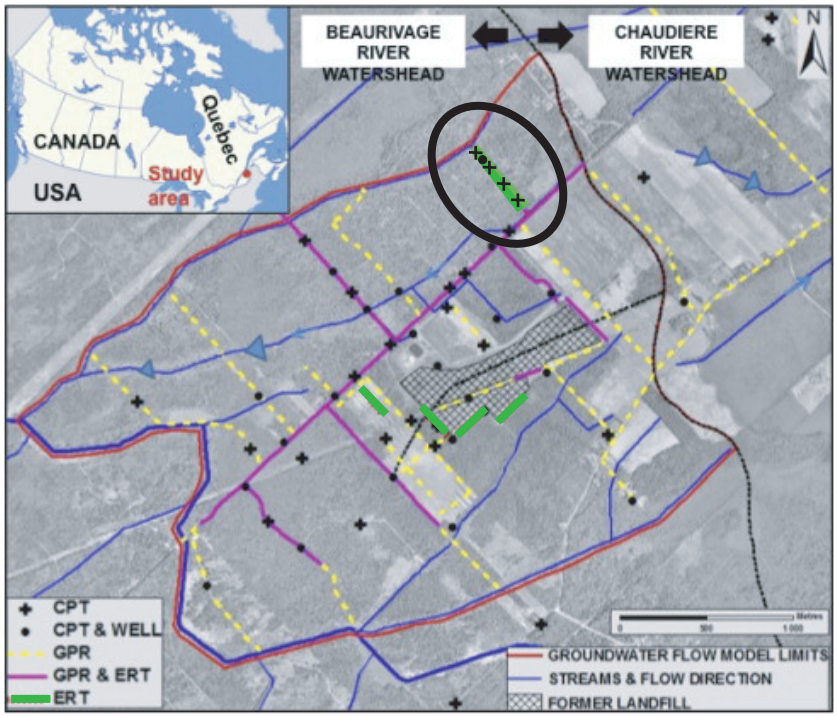

Figure 7: Location of the St-Lambert study area and the geophysical surveys and borehole logging that have been conducted there. The geoelectrical profile circled in black is the focus of our field application of the Bayesian sequential simulation aquifer characterization procedure. Positions marked "CPT" denote locations where borehole σ measurements were taken. Collocated measurements of K are available at locations marked "CPT & Well".

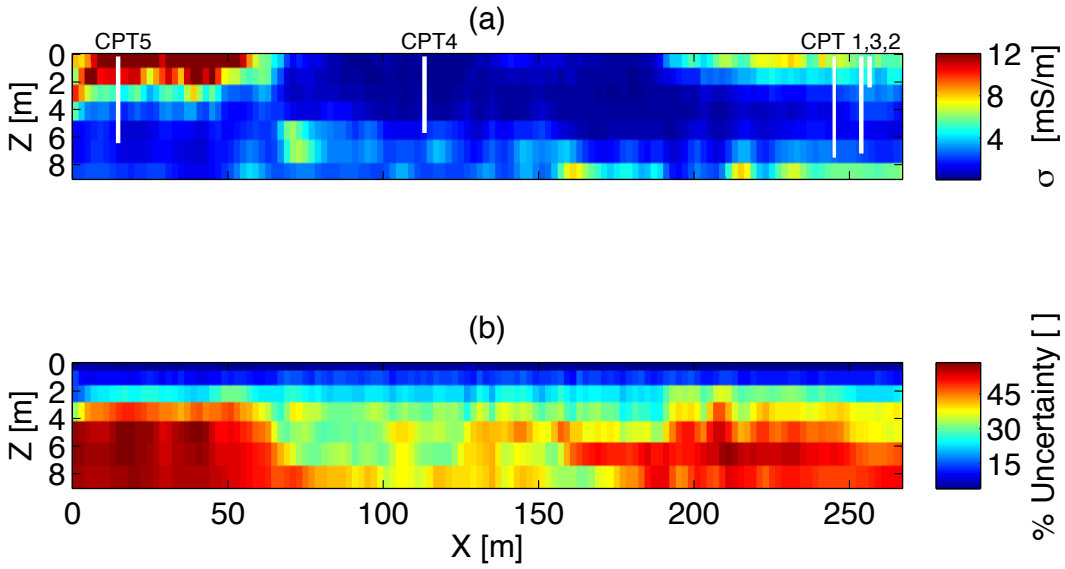

Figure 8: (a) Electrical conductivity tomogram obtained from the inversion of surface-based dipole-dipole geoelectrical measurements along the chosen profile (Figure 7). The boreholes are shown in white. (b) Estimated percentage uncertainty in the values shown in (a).

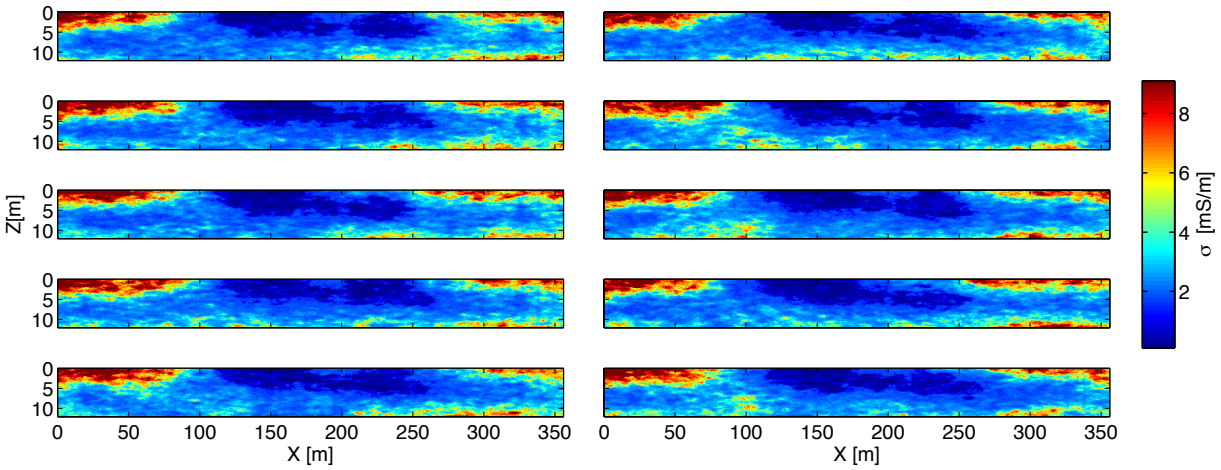

Figure 9: Ten stochastic realizations of the electrical conductivity field obtained using Bayesian sequential simulation. The realizations are conditioned to the geoelectrical inversion results in Figure 8 and the high-resolution σ measurements at the five borehole locations.

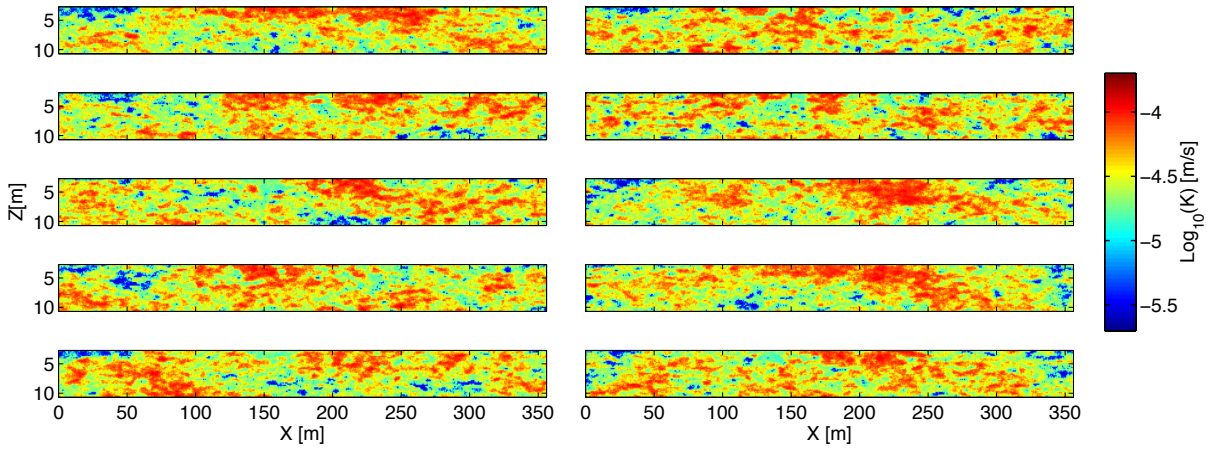

Figure 10: Ten stochastic realizations of the hydraulic conductivity field obtained using Bayesian sequential simulation. The realizations are conditioned to the point-by-point statistics of the high-resolution electrical conductivity field (Figure 9) and the relationship between σ and K established from collocated borehole measurements.

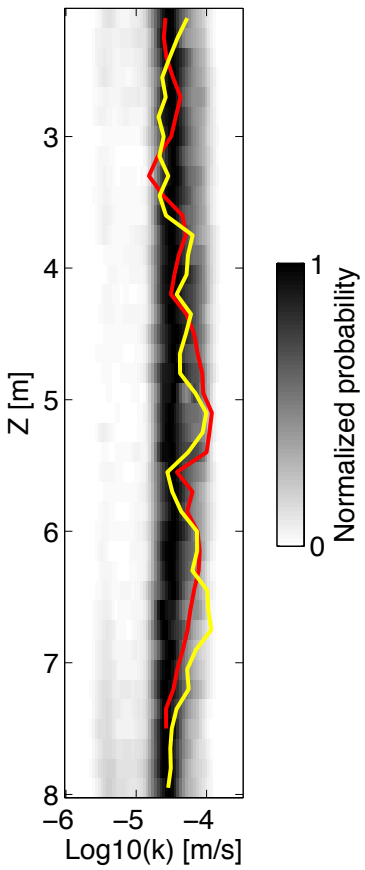

Figure 11: Measured hydraulic conductivity profile along borehole CPT1 (red) and one simulated profile at the same location extracted from a stochastic *K* realization (yellow). Also shown in black and white is the probability density image of *K* versus depth at this location based on 1000 stochastic realizations.