

**Integration of hydrological and geophysical data beyond the local scale:
Application of Bayesian sequential simulation to field data from the
Saint-Lambert-de-Lauzon site, Québec, Canada**

Paolo Ruggeri^{*1}, Erwan Gloaguen², René Lefebvre², James Irving¹, and Klaus Holliger¹

¹ Applied Geophysics Group, Center for Research of the Terrestrial Environment, University
of Lausanne, CH-1015 Lausanne, Switzerland.

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

* Corresponding author: paolo.ruggeri@unil.ch

Draft for submission to *Journal of Hydrology*

March 28, 2014

ABSTRACT

Adequate characterization of aquifer heterogeneity is critically important for the sustainable use, protection, and remediation of groundwater resources. The combined use of hydrological and geophysical measurements is arguably the most effective means of achieving this objective. In this regard, significant progress has been made on the quantitative integration of geophysical and hydrological data at the local scale. However, the extension of such approaches to larger, more regional scales remains a major research challenge. In this paper, we demonstrate the application of a recently developed regional-scale hydrogeophysical data integration approach, which is based on Bayesian sequential simulation, to a field database from Quebec, Canada consisting of low-resolution, surface-based geoelectrical measurements as well as high-resolution direct-push and borehole-based measurements of the electrical and hydraulic conductivities. The results of our study, which involved the integration of data along an approximately 250-m-long survey line, confirm that this novel methodology, with suitable adaptation, is fully applicable to field data and has the potential of providing realistic estimates of the spatial distribution of hydraulic target parameters at the regional-scale. Equally importantly, through the generation of multiple stochastic realizations, the methodology allows for quantitative assessment of the uncertainty associated with the inferred subsurface models, which in turn is essential for interpreting subsequent predictions of the flow and transport characteristics of the studied region.

KEYWORDS

Hydrogeophysics; Downhole methods; Tomography; Probabilistic forecasting

1. INTRODUCTION

The protection, remediation, and sustainable management of the world's increasingly fragile groundwater resources require adequate models of the spatial distribution of hydraulic conductivity as prerequisites for realistic predictions of groundwater flow and contaminant transport (e.g., Delleur, 1999; Chen et al., 2001). The hydraulic conductivity is an inherently challenging material property to estimate because it varies over many orders-of-magnitude, typically exhibits a pronounced degree of spatial heterogeneity, and can in general only be measured through dedicated experiments (e.g., Domenico and Schwartz, 1998; Ezzedine et al., 1999; Rubin, 2003; Schön, 2004; Butler, 2005).

Traditionally, the hydrological characterization of aquifers has been based on evidence from drill cores, hydraulic borehole logs, and tracer and pumping experiments. Core- and borehole-based measurements can provide detailed local information, but such information is inherently 1D and spatially sparse in nature, while tracer and pumping experiments tend to capture only the gross average properties of the probed subsurface region. Correspondingly, there is a large gap in terms of spatial coverage and resolution between these conventional hydrological techniques and hence they are, without complementary information, often inadequate for characterizing heterogeneous aquifers (e.g., Sudicky, 1986; McKenna and Poeter, 1995; Schreibe and Chien, 2003; de Marsily et al., 2005). While geophysical methods have the potential of bridging this gap in resolution and coverage associated with traditional hydrological measurements, they do in general not exhibit any direct sensitivity to the hydraulic conductivity. Moreover, any potential rock physical relationships between geophysical parameters and the hydraulic conductivity tend to be site- and scale-specific (Purvance and Andricevic, 2000; e.g., Schön, 2004; Lesmes and Friedman, 2005; Hyndman and Tronicke, 2005; Linde, 2006).

To overcome these complications, a number of strategies have been proposed for local-scale aquifer characterization, that is, at lateral distances ranging from approximately 5 to 50 m, typical involving a combination of core- and/or borehole-based hydraulic measurements and high-resolution crosshole tomographic geophysical surveys (e.g., Hyndman et al., 1994; Hyndman and Harris, 1996; Chen et al., 2001; Singha and Gorelick, 2005; Paasche et al., 2006; Dafflon et al., 2009b; Dafflon et al., 2009a; Dafflon et al., 2010; Dubreuil-Boisclair et al., 2011; Lochbühler et al., 2013) Most of these local-scale data integration approaches are based on geostatistical methods, which are not only well suited for assimilating diverse sources of information of varying resolution and hardness, but also offer the possibility of constraining the uncertainty of the inferred models. These local-scale data integration approaches are reaching a certain degree of maturity. Due to the lack of closely spaced boreholes for effective crosshole tomographic imaging, the extension of these local-scale techniques to larger scales does, however, represent a major and until recently essentially unresolved challenge. This is unfortunate since in many, if not most, cases it is at these larger scales that the greatest benefits of improved flow and transport predictions can be reaped (Domenico and Schwartz, 1998).

To address this problem, Ruggeri et al. (2013a) recently proposed a novel method for the quantitative integration of larger-scale geophysical and hydrological data based on a geostatistical technique known as Bayesian sequential simulation or BSS (Doyen and Boer, 1996). This approach showed significant promise when applied to realistic synthetic data for heterogeneous larger-scale aquifer models, but its practical viability remained unproven. Some promising, albeit highly preliminary, initial results of the application of this novel data integration techniques to real data were recently presented by Ruggeri et al. (2013b) in the context of a recent broad-public review.

The objective of this paper is to extend and complement the previous work by Ruggeri et al. (2013a; 2013b) by rigorously testing this novel data integration approach on a typical sub-regional-scale geophysical and hydrological field database. More specifically, we wish to explore the method's capacity and robustness for generating, in a computationally efficient manner, realistic conditional stochastic realizations of the larger-scale hydraulic conductivity field as well as for assessing the uncertainties of the thus inferred stochastic aquifer models.

2. METHODOLOGICAL BACKGROUND

In the following, we briefly outline the methodological foundations of the BSS-based data integration approach of Ruggeri et al. (2013a) before proceeding to assess its practical potential by applying it to field measurements. The BSS method (Doyen and Boer, 1996) allows for the generation of multiple, spatially correlated realizations of some variable of interest, referred to as the primary variable, conditioned to (i) spatially extensive measurements of a related secondary variable, as provided for example by geophysical surveying; and (ii) sparsely distributed measurements of the primary variable, as provided for example by borehole data. The following simplified version of Bayes' theorem forms the basis for the technique:

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

where A and B denote the primary and secondary variables, respectively, $p(\bullet)$ denotes a probability distribution, and c is a normalization constant. Equation (1) is valid under the assumption of conditional independence of B_n with respect to A_1, A_2, \dots, A_{n-1} when given A_n . That is, we assume that $p(B_n | A_1, \dots, A_n) = p(B_n | A_n)$.

As with all sequential simulation procedures, the generation of each stochastic realization using BSS 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 this variable at subsequent points (e.g., Goovaerts, 1997). Before the simulation begins, the covariance matrix for the primary variable is defined based on the horizontal and vertical variograms computed from existing values. In each iteration of the procedure, a value for the primary variable at cell n is then randomly drawn from the posterior distribution $p(A_n | B_n, A_1, \dots, A_{n-1})$, which is obtained by multiplying the prior distribution $p(A_n | A_1, \dots, A_{n-1})$ with the likelihood function $p(B_n | A_n)$. The prior distribution is conditional to the measured and previously simulated values of the primary variable in cells 1 through $n-1$, and is obtained by simple kriging of those values to obtain a Gaussian distribution having the kriging mean and variance. 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, is determined from the joint probability density $p(A, B)$, which in turn is computed from collocated measurements using a non-parametric kernel-based smoothing approach (Silverman, 1986; Wand and Jones, 1995). The posterior distribution can be viewed as an updated state of information that accounts for the prior and likelihood information at the chosen location. One realization of the primary variable is generated when all unknown cells in the model space have been simulated. Quite importantly, multiple stochastic realizations can be readily obtained by changing the order of the visited cells and repeating the simulation procedure. The latter allows for an assessment of the posterior ensemble uncertainty.

It should be noted that the BSS method is highly flexible with regard to the relationship that exists between the primary and secondary variables, in the sense that the likelihood is estimated empirically from collocated measurements of these variables. The quality of the relationship between A and B is thus reflected in the variability of the output stochastic realizations. Also note that, unlike cokriging-based simulation methods, the BSS approach does not rely on a generalized linear regression model, which is not an appropriate

choice when the relationship between the primary and secondary variables strongly deviates from being linear and multi-Gaussian. On the other hand, when these conditions are satisfied, approaches such as collocated cokriging could be equally well implemented.

The hydrogeophysical data integration approach of Ruggeri *et al.* (2013a), which again showed significant potential in the course of its initial testing on synthetic data, consists of two key steps, both of which are based on the general BSS methodology outlined above. Figure 1 summarizes the overall procedure. In the first step, high- and low-resolution geophysical parameter estimates (primary and secondary variable, respectively) are used to generate fine-scale realizations of the underlying geophysical property. The aim of this step is to effectively downscale the low-resolution geophysical parameter estimates and quantify the corresponding uncertainty with regard to the fine-scale grid. In this procedure, the high-resolution data are considered to be measurements of the geophysical parameter at a small number of sparsely distributed borehole locations throughout the aquifer volume. The low-resolution data, on the other hand, are considered to be a tomographic image obtained through the inversion of geophysical survey data, which can be regarded as set of uncertain spatially averaged measurements of the “true” subsurface geophysical parameter field. The likelihood function is thus estimated from the joint probability inferred from collocated or quasi-collocated high- and low-resolution geophysical parameter estimates at the borehole locations.

In the second step of the data integration approach of Ruggeri *et al.* (2013a), 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 in the first step (secondary variable) are used within BSS to generate multiple high-resolution realizations of the spatial distribution of the hydraulic conductivity. In other words, after stochastically downscaling the low-resolution geophysical parameter estimates, this second

step aims at using the geophysical information to condition the fine-scale hydraulic conductivity field. The likelihood function in this case is determined from collocated borehole measurements of the hydraulic conductivity and the geophysical parameter.

3. FIELD SITE AND DATA

The field site considered in this study is located in the municipality of Saint-Lambert-de-Lauzon, approximately 40 km south of Quebec City, Canada. It consists of a ~ 12 km² sub-watershed surficial granular aquifer over which there exists a decommissioned sanitary landfill (Figure 2). The aquifer is semi-confined and made up of Quaternary sediments with a thickness ranging from 5 to 20 m. The sediments were deposited in an east-west trending littoral environment, which finds its expression in a corresponding spatial variation of the sedimentation character. In particular, the aquifer sediments in the eastern region of the study area are dominated by medium- to fine-grained sandy deposits, which give way to alternating layers of fine sand and silt towards the west (Paradis et al., 2011b). The study area is generally flat and groundwater flow is governed by the presence of natural streams and drainage networks. The groundwater table is located below a surficial layer of marine clayey silt at a depth of ~ 1 -2 m (Bolduc, 2003). Further details regarding the geological setting of the site can be found in Paradis *et al.* (2011b).

The geophysical and hydrological data considered in this study were acquired in the north-eastern part of the St-Lambert site (Figure 2) and consist of 2D estimates of the electrical conductivity inferred from surface-based electrical resistivity tomography (ERT) along with high-resolution vertical measurements of the electrical and hydraulic conductivities. The orientation of the chosen ERT profile, which is 440 m long, is along the sedimentary depositional trend of the coastal environment in which sediments composing the superficial granular aquifer system were formed. This orientation is therefore optimal to

capture the variations in hydrofacies related to different parts of the depositional environment. Moreover, a vast variety of additional measurements were available along this profile, most importantly of which were observation wells and cone penetrometer data that could be used to establish a site-specific relationship between the electrical and hydraulic conductivities.

The surface-based geoelectrical data were acquired using an Iris Instruments SYSCAL Pro resistivity meter with a dipole-dipole acquisition geometry and a minimum electrode spacing of 2 m and a maximum electrode spacing of 94 m. The resulting 8394 resistance measurements were tomographically inverted on a coarse grid having a constant horizontal discretization of 2 m and a variable vertical discretization increasing from 1.0 m to 2.7 m with increasing depth. The tomographic inversion was performed using the commercial RES2DINV software (Loke, 2012) with regularization imposed through model smoothness constraints in the horizontal and vertical directions (Gloaguen et al., 2012). The final RMS error between calculated and measured apparent resistivities after 5 iterations was approximately 8.7%. Figure 3a shows the entire inverted ERT profile, which is 440 m long by 20 m deep, whereas Figure 3b shows the central portion considered in this study, which is only 267 m long by 9 m deep. To estimate the uncertainty in the electrical conductivity estimates shown in Figure 3b, we followed the approach of Alumbaugh and Newman (2000) based on the diagonal elements of the calculated model covariance matrix. Figure 3c shows the corresponding spatial distribution of the estimated standard deviation for the \log_{10} conductivity values which, as expected, tend to be relatively high in regions of the subsurface that are poorly constrained by the geoelectrical data, most notably in the deeper parts of the ERT profile. Note, however, that the estimated uncertainty at depth near the center of the profile is significantly less than that towards the edges due to the local absence of the highly conductive surficial layer of clayey silt (Figure 3b).

High-resolution direct-push measurements of the electrical conductivity were collected at five locations (CPT1, CPT2, CPT3, CPT4, and CPT5) along the considered ERT profile, as well as at two off-profile locations (P3 and P16) (Figures 2 and 3b). These measurements were acquired using a Geotech 605-D cone penetrometer system equipped with a pore pressure probe (CPTu) and a soil moisture resistivity (SMR) probe. The SMR probe consists of four ring electrodes spaced 3 cm apart (Shinn et al., 1998), the inner two of which are used to measure the dielectric permittivity and the outer two of which are used to measure the bulk electrical resistivity. The probe operates at a frequency of 1000 Hz to avoid soil polarization effects (Paradis et al., 2008). The combined CPTu/SMR probe was pushed vertically into the subsurface at a constant rate of ~ 2 cm/s, which resulted in a vertical spacing of ~ 2 -3 cm between the measurements. Multi-level slug test measurements with a vertical spacing of ~ 15 cm were also carried out in fully-screened observation wells at locations CPT1, P3, and P16, which provided high-resolution estimates of the hydraulic conductivity (Paradis et al., 2011b). Quality control on multi-level slug tests was done following the best practices recommended by Butler (1998) and included multiple tests with varying initial hydraulic head changes at the onset of slug tests. This was done for about 15 tests covering the range of hydraulic conductivity values found at the site. The relative heads recorded through time obtained from multiple tests were perfectly recovered. A comparison between measurements obtained by flowmeter measurements and multi-level slug tests along the same vertical profiles showed difference in $\log_{10}K$ estimates below 10% (Paradis et al., 2011b).

To allow for a direct comparison of the electrical and hydraulic conductivity data, the CPTu/SMR probe measurements were decimated to the same 15 cm measurement interval (Paradis et al., 2008).

4. RESULTS

4.1 Stochastic simulation of the fine-scale electrical conductivity distribution

Using the high-resolution direct-push measurements of the electrical conductivity collected at locations CPT1 through CPT5 along with the low-resolution ERT-based estimates available everywhere throughout the model space (Figure 3b-c), we first used BSS to generate multiple realizations of the fine-scale distribution of electrical conductivity conditioned to these data. As outlined earlier, a critical component of the BSS procedure is the evaluation of the likelihood function $p(B_n | A_n)$, which statistically relates the primary and secondary variables. This is accomplished by first estimating the joint probability density $p(A, B)$ from collocated measurements (Ruggeri et al., 2013a). In that regard, the left-hand column of Figure 4 compares the SMR- and ERT-based electrical conductivity estimates at locations CPT1 through CPT5 along the profile. Notice that the overall trend of the SMR-probe data is clearly seen in the corresponding ERT-based estimates, which finds its quantitative expression in a relatively high correlation coefficient of 0.83. However, we also observe that there is a bias between the two curves in each case, in that the magnitudes of the SMR-probe data are generally higher than those of the ERT image. We must consider the possibility that edge/skin effects around the SMR probe as well as inaccuracies in its calibration have affected the low-frequency content of the data. There is thus good reason to believe that the low-frequency variations of the electrical conductivity are better captured by the ERT image. In view of this, we used the low-frequency trend in the ERT image to effectively rescale the SMR-probe data before inferring $p(A, B)$. This was done in the wavenumber domain by replacing the very lowest frequency components of the SMR-probe data with those from the ERT image followed by an inverse Fourier transform back into the space domain. The corresponding results are shown in the right-hand column of Figure 4 and indicate that the adopted rescaling procedure was effective at removing the bias. Figure 5a shows the joint probability density that was subsequently obtained from 196 collocated ERT- and SMR-based conductivity

estimates using a kernel-based smoothing approach (Silverman, 1986; Wand and Jones, 1995).

The BSS procedure was performed on a 267-m-long by 9-m-deep simulation grid, identical in size to Figure 3b, that was uniformly discretized at 0.15 m in order to match the spatial sampling interval of the direct-push measurements. This provided high-resolution electrical conductivity realizations that honour the large-scale structure imposed by the ERT image along with the fine-scale structure of the SMR-probe measurements. Before beginning the simulations, we computed the best-fitting vertical variogram model from the high-resolution data, which was found to be a linear combination of spherical and Gaussian functions with vertical correlation lengths of 0.75 m and 5.1 m, respectively, and sill values of 0.08 and 0.92, respectively (Figure 5b). The corresponding horizontal correlation lengths, although not directly constrained by the available data, were quantified based on variogram analysis of the structural aspect ratio of the ERT image and found to be 7.5 m for the spherical and 100 m for the Gaussian variogram functions (e.g., Gelhar, 1993; Tronicke and Holliger, 2005; Dafflon et al., 2009b). Following Ruggeri et al. (2013a), the likelihood $p(B_n | A_n)$ was evaluated at each iteration from the joint probability density in Figure 5a, taking into account the previously inferred uncertainty associated with the ERT image (Figure 3c). Accounting for this uncertainty ensures that the output realizations are not overly constrained by the ERT image in places where the geoelectrical data provide little information.

The output realizations generated using BSS have the desired property of honouring both the large-scale structure imposed by the ERT image as well as the fine-scale structure provided by the borehole electrical conductivity measurements. In addition, they are tied to the high-resolution electrical conductivity data at the borehole locations. However, there is no guarantee that they will still honour within a suitable degree the original geoelectrical data from which the ERT image was derived (i.e., the measured voltages or apparent resistivities).

For this reason, an additional step involving gradual deformation can be included after BSS in order to generate stochastic realizations that provide an acceptable degree of fit between observed and predicted geoelectrical measurements while keeping the large-scale structure, ensemble statistics, and covariance properties of the BSS realizations (Ruggeri et al., 2013a). Gradual deformation is an optimization procedure by which different stochastic realizations of a target parameter are iteratively combined in order to obtain a satisfactory match to a set of observed data that depends upon that parameter (Roggero and Hu, 1998; Ying and Gomez, 2000; Hu et al., 2001; Ravelec-Dupin, 2005). The procedure effectively “deforms” one stochastic realization towards another until an adequate match to the data is obtained, while preserving through the optimization process the mean, covariance, and fit to the conditioning data of the original stochastic model. In this study, we used gradual deformation as a means of combining the stochastic realizations of the fine-scale electrical conductivity field generated through BSS to create realizations that provided a suitable fit to the original geoelectrical measurements at the St-Lambert site, which corresponds to essentially the same RMS as that of the ERT image. The forward simulations of the electrical conductivity data corresponding to the fine-scale grid were carried out using a 2D finite-element-based R2 code (Binley and Kemna, 2005). The procedure was repeated until 150 of such “optimized realizations” had been created.

Figure 6a-c shows three representative stochastic realizations of the fine-scale electrical conductivity field inferred through a combination of BSS and gradual deformation. We see that the simulation results appear to honour both the large-scale structure imposed by the ERT image as well as the fine-scale structure provided by the borehole measurements. The statistical accuracy of the results is further confirmed through comparisons of the vertical experimental variogram and histogram obtained from the high-resolution electrical conductivity measurements with those calculated from the entire set of 150 stochastic

realizations (Figures 5b and 5c). One of the advantages of using a stochastic approach in the given context is that the uncertainty of each model parameter can be readily evaluated through the ensemble of generated realizations. Assuming that the parameters are normally distributed, the point-by-point means and variances can be used to completely describe the marginal distributions, as proposed by Doyen and Boer (1996) and Gastaldi and Roy (1998). However, in our case, the Gaussian assumption is not warranted and thus using such an approach would be largely meaningless. This is shown in Figures 7a and 7b, where we see substantial differences between the cumulative distribution functions computed for two model parameters located at $[x = 109.5\text{m}, z = 0.45\text{m}]$ and $[x = 160.5\text{m}, z = 2.1\text{m}]$ with those corresponding to the Gaussian distributions obtained from the ensemble mean and variance at these locations. For this reason we display in Figure 6d the point-by-point estimates of the 5th-95th inter-percentiles of the set of 150 electrical conductivity realizations to best visualize uncertainty. Notice in these images that a greater degree of variability is present in the central part of the profile and also at depth, as evidenced by a larger difference between the 5th and 95th percentiles at these locations. At first glance, the observation of greater variability in the central part of the profile would appear to be in contrast with the lower uncertainty in the ERT image in this region (Figure 3c). It is, however, a consequence of the fact that joint probability density function in Figure 5a predicts greater uncertainty in the fine-scale conductivity values when provided with lower values of the ERT-based conductivity, even if the latter can be estimated with high precision.

It is important to note that choices made during the production of the ERT image in Figure 3 will clearly have an influence on the high-resolution stochastic realizations of the electrical conductivity obtained using our BSS procedure. Indeed, the ERT inverse problem is highly non-unique and the tomographic images obtained are strongly affected by different factors such as the parameterization of the model space and the type of regularization used.

As is always the case with geophysical inversion, we must make choices in this regard that are most consistent with our existing knowledge regarding the subsurface environment. However, it is important to emphasize that the BSS-based integration approach links indirect measurements of the electrical conductivity (from the ERT image) to in-situ hydraulic conductivity measurements, through their relationship with high-resolution electrical conductivity measurements. As a result, all one needs for the approach to work is that there exists a certain level of correlation between the low- and high-resolution electrical conductivity data. In this perspective, even with a different spatial configuration of the electrical conductivity corresponding to a different inversion, the same workflow could be applied, as long as a reasonably good correlation between these variables was still observed. It is important to note, however, that because the form and uncertainty of the joint relationship are determined empirically based on collocated measurements in the Bayesian sequential simulation procedure, any issues with having, for example, a different tomographic image should be implicitly accounted through the variability of the output realizations.

4.2 Stochastic simulation of the fine-scale hydraulic conductivity distribution

In a second application of BSS, we next generated high-resolution stochastic realizations of the hydraulic conductivity based on the marginal posterior statistics determined from the high-resolution electrical conductivity realizations (i.e., point-by-point histograms) and the relationship established between these two variables from collocated borehole measurements. In doing so, it is important to note that we assume in this study to work within a given hydrogeological unit for which the relationship between the hydraulic and electrical conductivities is statistically stationary (Prasad, 2003; Cosentino, 2001) and thus does not depend on the chosen location within the model domain. Given that the water table was located at a depth of 1.5-2 m at the time of the measurements and that the lower boundary of

the aquifer is between 8-10 m depth along the profile, this assumption was satisfied by restricting the vertical extent of our simulation domain for this application of BSS to between 2 and 8 m depth. The considered domain was therefore 267 m long and 6 m deep and was again discretized at 0.15 m. Figure 8a shows the joint probability distribution that was estimated between the hydraulic and electrical conductivities in the corresponding aquifer sediments using 130 high-resolution data acquired at well locations CPT1, P3, and P16.

Before starting the BSS procedure, we computed the best-fitting vertical variogram model from the borehole hydraulic conductivity data, which was found to be a linear combination of spherical and Gaussian functions having vertical correlation lengths of 0.6 m and 3.3 m, respectively, and sill values of 0.5 and 0.5, respectively (Figure 8b). To estimate the corresponding horizontal correlation lengths, we again used the aspect ratio inferred from the ERT image. The uncertainty associated with the high-resolution electrical conductivity at each location, as quantified by the previously evaluated marginal distributions for each cell of the model space, was taken into account during evaluation of the likelihood based on the joint density in Figure 8a. Please note that, as opposed to the stochastic simulation of the electrical conductivity, which was conditioned to the high-resolution electrical conductivity measurements collected at locations CPT1 through CPT5, we intentionally generated 150 unconditional realizations of the hydraulic conductivity field in this case to allow for validating our results against the hydraulic conductivity measurements made at well CPT1 (Kelkar and Perez, 2002; Gastaldi and Roy, 1998).

Figure 9a-c shows three representative stochastic realizations of the fine-scale hydraulic conductivity field that were obtained through the BSS procedure. Given the structural similarities between these images and Figure 3b, it is clear that the realizations have been successfully conditioned by the low-resolution electrical conductivity data. Notice, however, that the hydraulic and electrical conductivities are anti-correlated, in the sense that

lower hydraulic conductivities are found in regions of higher electrical conductivity and vice versa. This results because the joint probability distribution established between these two variables exhibits significant anti-correlation (Figure 8a). Comparing the vertical experimental variogram and histogram obtained from the high-resolution hydraulic conductivity measurements with those calculated from the entire set of 150 stochastic realizations (Figures 8b and 8c), we also see excellent agreement, which helps to confirm the statistical validity of our simulation results. Figure 9d shows the point-by-point estimates of the 5th-95th inter-percentiles calculated from the ensemble of 150 BSS-generated realizations. Here, we observe a noticeable degree of variability throughout the entire model domain with greater differences between the 5th and 95th percentile images in parts of the hydraulic conductivity models constrained by higher values of the fine-scale electrical conductivity (Figure 6a-c). This results because the joint relationship in Figure 8a is highly uncertain and predicts greater variability in the fine-scale hydraulic conductivity for higher values of the electrical conductivity.

4.3 Model validation

We now seek to validate the output realizations of the hydraulic conductivity field obtained through our BSS-based data integration approach with data that were not used in the conditioning of these realizations. To this end, we first compare the hydraulic conductivity measured along the borehole at CPT1 with the five best predictions provided by the stochastic realizations. The corresponding results are shown in Figure 10a, along with 5th and 95th percentile bounds obtained from the marginal distributions of hydraulic conductivity along the vertical profile based on all 150 realizations. We see in this figure that the measured profile of the hydraulic conductivity is well within the 5th and 95th percentile bounds at

almost every depth, and that the overall characteristics of the fine-scale variability have been adequately captured.

In addition to comparing our simulations with the measurements of hydraulic conductivity acquired along the borehole at CPT1, we have also compared them with predictions of hydraulic conductivity obtained using a relevance vector machine (RVM) approach based on a combination of CPTu/SMR data at locations CPT4 and CPT5 (Paradis et al., 2011a; Paradis et al., 2014). RVM is a powerful machine learning technique that uses Bayesian inference for regression and classification tasks (Tipping, 2001). The “machine” was first trained on data collected from all boreholes within the study region with collocated measurements of the hydraulic conductivity and various CPTu/SMR data, such as sleeve friction, tip mechanical stress, dielectric permittivity, and electrical resistivity and then used to predict the hydraulic conductivity in locations where only CPTu/SMR data were available. Figures 10b and 10c show the corresponding comparison at locations CPT4 and CPT5, respectively. Again, for comparison with the RVM-based predictions, we selected five best predictions provided by the stochastic realizations. Although it can be seen that our stochastic realizations of the hydraulic conductivity generally exhibit more small-scale variability than the RVM-based predictions, the qualitative comparison between the estimates obtained by these fundamentally different and entirely independent approaches shows good agreement. In this context it is particularly interesting and important to note that the low values of hydraulic conductivity predicted by the RVM technique between 2 and 3 m depth are well reproduced by our stochastic realizations (Figure 10c). For CPT4, the marginal distribution of the hydraulic conductivity obtained through the BSS procedure suggests a bimodal distribution between 2 and 4 m depth. For this reason, we show two sets of representative stochastic realizations drawn from the bimodal ensemble (Figure 10b). Further inspections revealed that the bimodal distribution between 2 and 4 m depth is characterized by two “peaks” with

probabilities of $\sim 30\%$ and $\sim 70\%$ and values ranging between -4.9 and $-4.5 \log_{10}(K[\text{m/s}])$ and -4.5 and $-3.8 \log_{10}(K[\text{m/s}])$, respectively. Quite interestingly, the corresponding RVM-based prediction is consistent with the former, less likely range of values (Figure 10b), which, in conjunction with the other results, again indicates that the considered data BSS approach is capable of producing realistic distributions of the fine-scale hydraulic conductivity through the integration of hydraulic and geophysical data typically used for regional-scale studies.

5. CONCLUSIONS

The purpose of this study was to explore the practical applicability of a recently developed data integration procedure to a relatively novel type of geophysical and hydrogeological database with the objective of obtaining stochastic realizations of hydraulic conductivity of the probed subsurface region of a granular aquifer. The approach consists of a two-step Bayesian sequential simulation technique that is targeted to the common case where there exist low-resolution, but spatially exhaustive, surface-based or airborne geophysical measurements along with high-resolution, but spatially sparse, borehole measurements of the governing geophysical parameter as well as the hydraulic conductivity. This novel data integration method has the particular advantage of using the larger-scale geophysical parameter, together with its inferred uncertainty, to constrain the simulation of the corresponding fine-scale geophysical parameter field. This means that the uncertainty of the inverted geophysical parameter directly determines the strength with which the simulations are conditioned by these data. Moreover, the interrelations between fine-scale geophysical and hydrogeological parameters are based on joint probability functions computed from collocated measurements, thus avoiding the need of having to specify, explicitly or implicitly, any such relationships. Quite importantly, this data integration procedure provides solutions in terms of multiple realizations of the target parameter. This characteristic is of great

importance because it always allows the assessment of the uncertainty of the corresponding predictions.

We found that the information regarding the larger-scale subsurface structure, as provided by the ERT-based electrical conductivity image, was properly incorporated into the stochastic realizations with the smaller-scale stochastic fluctuations realistically superimposed. Indeed, the method capitalizes well on the complementarity of the various data and constraints, which in turn contributes to the consistency and realism of the resulting stochastic simulations. The results indicate that, for the considered case study, our data integration approach adequately reproduces the small-scale heterogeneity as well as the global trends of the prevailing hydraulic conductivity field. The results of this study also indicate that this novel data integration approach is flexible and robust and hence can be expected to be applicable to a wide range of geophysical and hydrological data having different spatial resolutions, and opens the perspective of allowing for the integration of such data at the regional-scale.

ACKNOWLEDGMENTS

This work was supported by a grant from the Swiss National Science Foundation. We also thank the Fondation Herbette for supporting a visit of EG to the University of Lausanne. The field data considered in this paper were acquired through the support of the *Régie Intermunicipale de gestion des déchets des Chutes-de-la-Chaudière*, along with NSERC Discovery Grants to EG and RL.

REFERENCES

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

- Binley, A., Kemna, A., 2005. DC resistivity and induced polarization methods. In: Rubin, Y., Hubbard, S. (Eds.), *Hydrogeophysics*. Springer, Dordrecht, pp. 129-156.
- Bolduc, A., 2003. *Géologie des formations superficielles*, Charny, Québec, Commission Géologique du Canada.
- Butler, J.J., 2005. Hydrogeological methods for estimation of spatial variations in hydraulic conductivity. In: Rubin, Y., Hubbard, S. (Eds.), *Hydrogeophysics*. Springer Netherlands, pp. 23-58.
- Butler Jr., J.J., 1998. *The Design, Performance and Analysis of Slug Tests*. Lewis
- Chen, J., Hubbard, S.S., Rubin, Y., 2001. Estimating the hydraulic conductivity at the South Oyster site from geophysical tomographic data using Bayesian techniques based on normal linear regression model. *Water Resources Research*, 37: 1603-1613.
- Cosentino, L., 2001. *Integrated reservoir studies*. Editions Technip, Paris, 336 pp.
- Dafflon, B., Irving, J., Holliger, K., 2009a. Simulated-annealing-based conditional simulation using high-resolution geophysical data for the local-scale characterization of heterogeneous aquifer. *Journal of applied geophysics*, 68: 60-70.
- Dafflon, B., Irving, J., Holliger, K., 2009b. Use of high-resolution geophysical data to characterize heterogeneous aquifers: Influence of data integration method on hydrological predictions. *Water Resources Research*, 45: W0940.
- Dafflon, B., Irving, J., Holliger, K., 2010. Calibration of high-resolution geophysical data with tracer test measurements to improve hydrological predictions. *Advances in Water Resources*, 33(1): 55-68.
- de Marsily, G. et al., 2005. Dealing with spatial heterogeneity. *Hydrogeological Journal*, 13: 161-183.
- Delleur, J.W., 1999. Elementary groundwater flow and transport processes. In: Delleur, J.W. (Ed.), *The handbook of groundwater engineering*. CRC Press Boca Raton, pp. 2.1-2.40.
- Domenico, P.A., Schwartz, F.W., 1998. *Physical and chemical hydrogeology*. , 44. Wiley, New York.
- Doyen, P.M., Boer, D.L.D., 1996. *Bayesian sequential Gaussian simulation of lithology with non-linear data*, Western Atlas International, Inc.
- Dubreuil-Boisclair, C., Gloaguen, E., D.Marcotte, B.Giroux, 2011. Heterogeneous aquifer characterization from Ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear bayesian simulations. *Geophysics*, 76(4): J13-J25.
- Ezzedine, S., Rubin, Y., Chen, J., 1999. Hydrological-geophysical Bayesian method for subsurface characterization: theory and application to LLNL suerfund site. *Water Resources Research*, 35: 2671-2683.
- Gastaldi, C., Roy, D., 1998. Using Bayesian simulations to predict reservoir thickness under tuning conditions. *The Leading Edge*: 589-593.
- Gelhar, L.W., 1993. *Stochastic subsurface hydrology*. Prentice Hall, Englewood Cliffs.
- Gloaguen, E. et al., 2012. Inference of the two dimensional GPR velocity field using collocated cokriging of direct push permittivity and conductivity logs and GPR profiles. *Journal of Applied Geophysics*, 78: 94-101.
- Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.
- Hu, L.Y., Blanc, G., Noetinger, B., 2001. Gradual deformation and iterative calibration of sequential stochastic simulations. *Mathematical Geology*, 33(4): 475-489.
- 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., 1994. Coupled seismic and tracer test inversion for aquifer property characterization. *Water Resources Research*, 30: 1965--1977.
- Hyndman, D.W., Tronicke, J., 2005. Hydrogeophysical case studies at the local scale: The saturated zone. In: Rubin, Y., Hubbard, S. (Eds.), *Hydrogeophysics*. Springer, New York, pp. 391-412.
- Kelkar, M., Perez, G., 2002. Applied geostatistics for reservoir characterization. SPE, Richardson, 264 pp.
- Lesmes, D.P., Friedman, S., 2005. Relationships between electrical and hydrogeological properties of rocks and soils. In: Rubin, Y., Hubbard, S. (Eds.), *Hydrogeophysics*. Springer, Dordrecht, pp. 129-156.
- Linde, N., 2006. Hydrogeophysical parameter estimation approaches for field scale characterization. *Applied Hydrogeophysics*. Springer Netherlands, pp. 9-44.
- Lochbühler, T., Doetsch, J., Brauchler, R., Linde, N., 2013. Structure-coupled joint inversion of geophysical and hydrological data. *Geophysics*, 78(3): ID1-ID14.
- Loke, M.H., 2012. Available at <http://www.geotomosoft.com/>
- McKenna, S.A., Poeter, E.P., 1995. Field example of data fusion in site characterization. *Water Resources Research*, 31(12): 3229-3240.
- 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., Gloaguen, E., Lefebvre, R., Rivera, A., 2011a. Tying geophysics to hydrogeology: a learning machine approach to characterize heterogeneous granular aquifers. *Geohydro2011*, Quebec City, Canada, pp. 8.
- Paradis, D. et al., 2008. Multivariate integration of CPTu/SMR and hydraulic conductivity measurements for the definition of hydrofacies in unconsolidated sediments. *GeoEdmonton '08, 61st Canadian Geotechnical Conference and 9th Joint CGS/IAH Conference*. Canadian Geotechnical Society, Edmonton, pp. 1470-1477.
- Paradis, D., Lefebvre, R., Morin, R.H., Gloaguen, E., 2011b. Permeability profiles in granular aquifers using flowmeters in direct-push wells. *Ground Water*, 49(4): 534-547.
- Prasad, M., 2003. Velocity-permeability relations within hydraulic units. *Geophysics*, 68(1): 108-117.
- Purvance, D.T., Andricevic, R., 2000. On the electrical-hydraulic conductivity correlation in aquifers. *Water Resources Research*, 36(10): 2905-2913.
- Ravelec-Dupin, M.L., 2005. Inverse stochastic modeling of flow in porous media: applications to reservoir characterization. IFP Publications, 1.
- Roggero, F., Hu, L.Y., 1998. Gradual deformation of continuous geostatistical models for history matching.
- Rubin, Y., 2003. *Applied Stochastic Hydrology*. Oxford University Press, New York.
- Ruggeri, P., Irving, J., Gloaguen, E., Holliger, K., 2013a. Regional-scale integration of multi-resolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach. *Geophysical Journal International*, 194(1): 289-303
- Ruggeri, P., Irving, J., Gloaguen, E., Lefebvre, R., Holliger, K., 2013b. Hydrogeophysical data integration at larger scales: Application of Bayesian sequential simulation for the characterization of heterogeneous alluvial aquifers. *TLE*, 32(7): 766-774.
- Schön, J.H., 2004. *Physical properties of rocks: Fundamentals and principles of petrophysics*. Elsevier, Oxford.
- Schreibe, T.D., Chien, Y.J., 2003. An evaluation of conditioning data for solute transport prediction. *Ground Water*, 41(2): 128-141.

- Shinn, J.D. et al., 1998. Development of a CPT deployed probe for in situ measurements of volumetric soil moisture content and electrical resistivity. *Field Analytical Chemistry and Technology*, 2(2): 103-110.
- Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. Chapman and Hall, New York, 175 pp.
- Singha, K., Gorelick, S.M., 2005. Saline tracer visualized with electrical resistivity tomography: Field scale moment analysis. *Water Resources Research*, 41: W05023.
- Sudicky, E.A., 1986. A natural gradient experiment on solute transport in sand aquifers: spatial variability of hydraulic conductivity and its role in the dispersion process. *Water Resources Research*, 22: 2069-2089.
- Tipping, M.E., 2001. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1: 211-244.
- Tremblay, L., Lefebvre, R., Paradis, D., Gloaguen, E., 2013. Conceptual model of leachate migration in a granular aquifer derived from the integration of multi-source characterization data (St-Lambert, Canada). *Hydrogeology journal*: 1-22.
- Tronicke, J., Holliger, K., 2005. Quantitative integration of hydrogeophysical data: conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers. *Geophysics*, 70: H1-H10.
- Wand, M.P., Jones, M.C., 1995. *Kernel smoothing*. Monographs on Statistics and Applied Probability. Chapman and Hall, New York.
- Ying, Z., Gomez, J.H., 2000. An improved deformation algorithm for automatic history matching, Stanford Center for Reservoir Forecasting (SCRF).

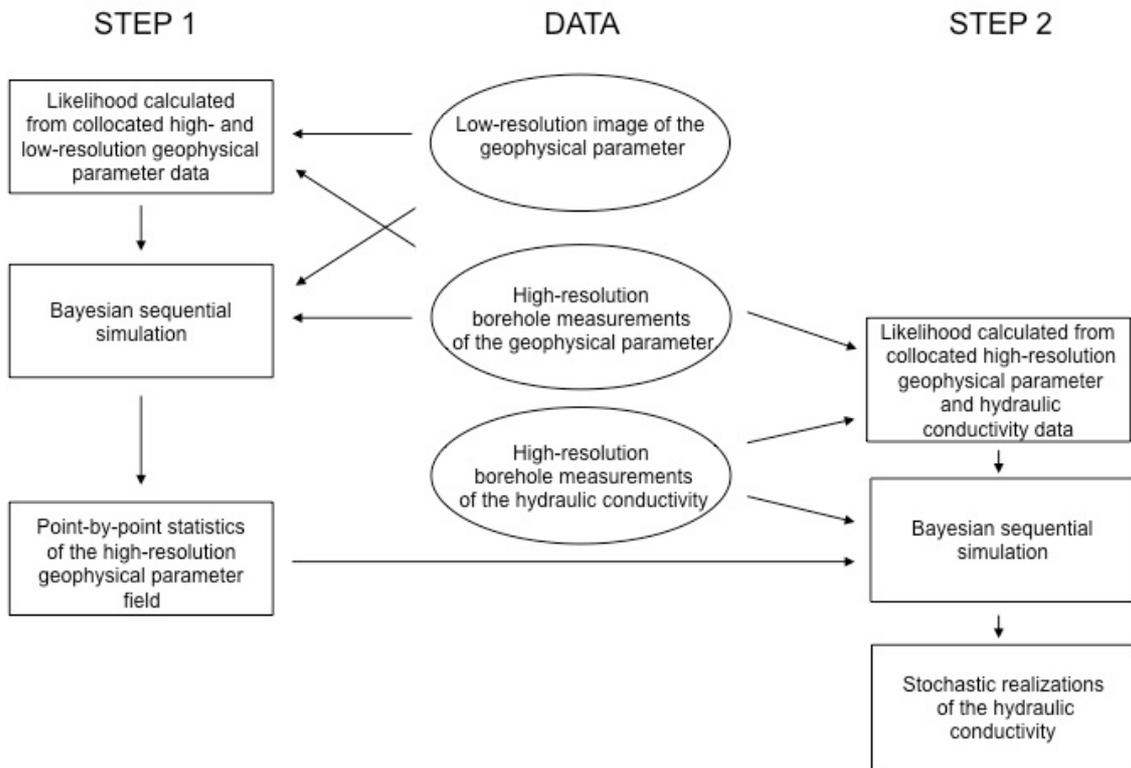


Figure 1: Schematic outline of the two-step, BSS-based aquifer characterization procedure of Ruggeri et al. (2013a). Adapted from Ruggeri et al. (2013b).

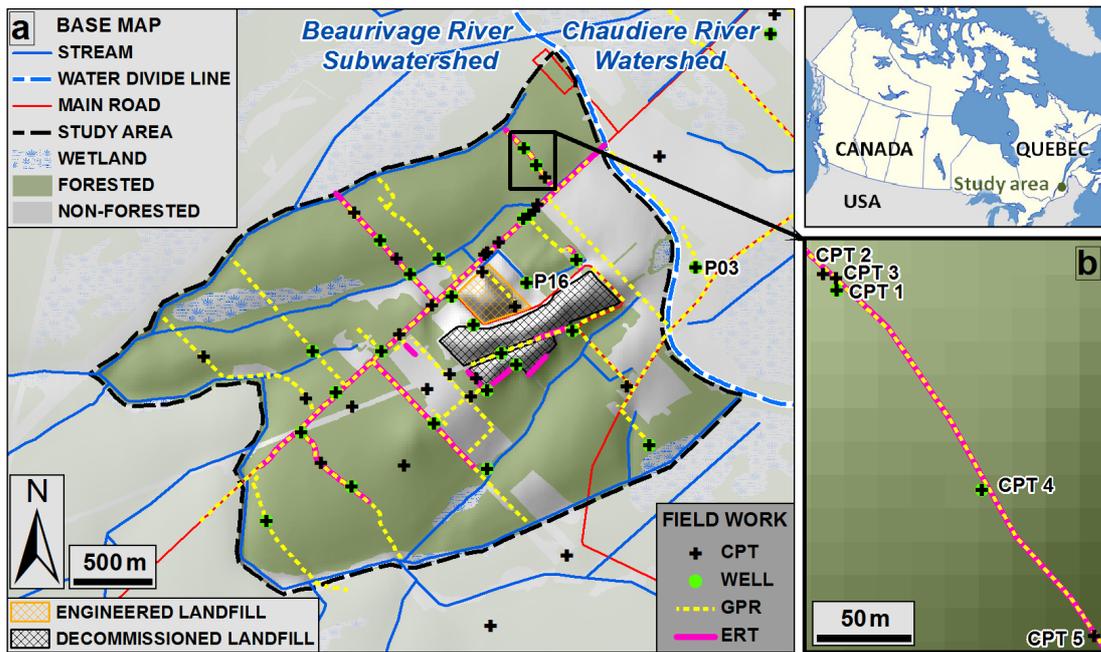


Figure 2: Location of the St. Lambert study area and the various measurements acquired there. The limits of the domain (dashed black line) correspond to the sub-watershed underlying the decommissioned St-Lambert sanitary landfill. The black rectangle, which is enlarged on the bottom right, contains the 440-m-long geoelectrical profile and five cone penetrometer soundings considered in this study. Hydraulic conductivity measurements from wells P03 and P16, as well as from another well coincident with CPT1, were also utilized. Adapted from Tremblay et al. (2013).

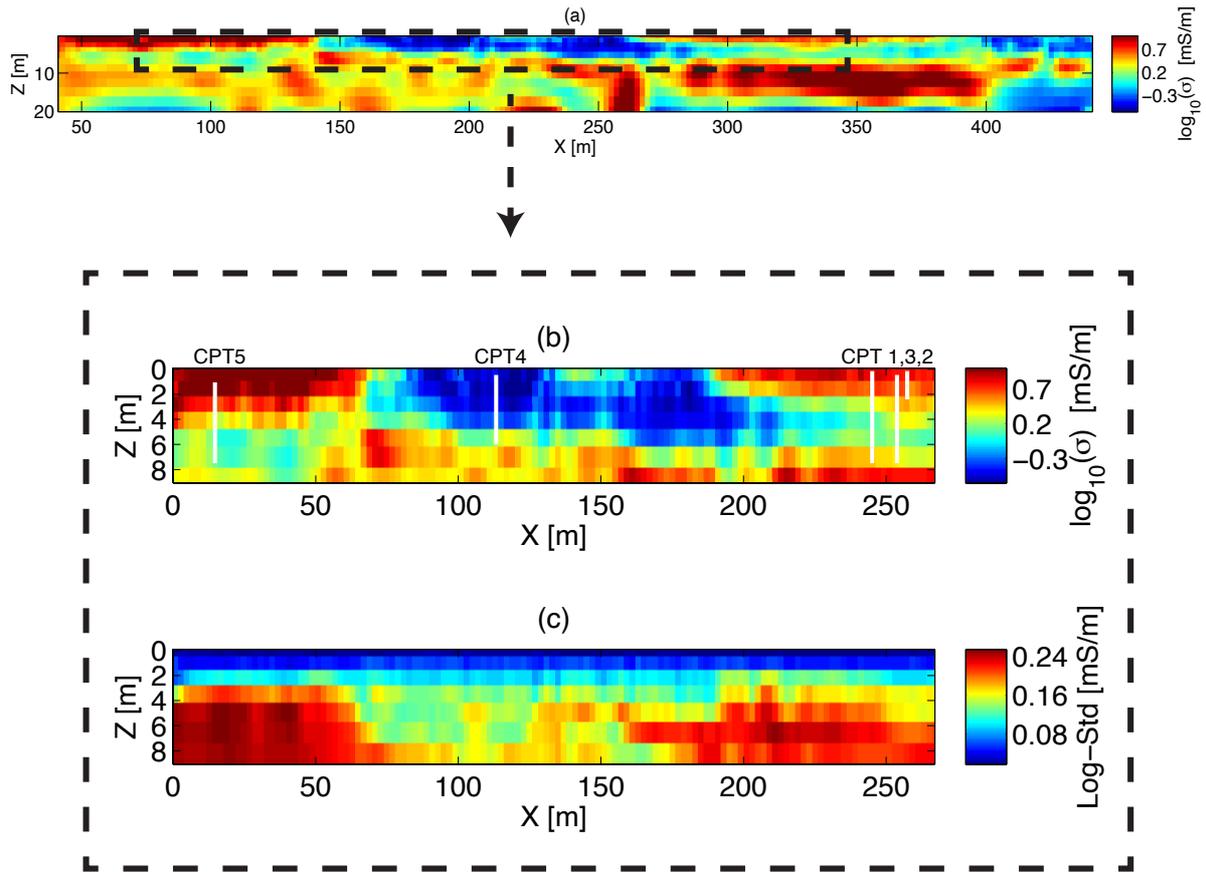


Figure 3: (a) Electrical conductivity tomogram obtained from the inversion of surface-based dipole-dipole geoelectrical measurements acquired along the profile line shown in Figure 2b; (b) Zoom of the region of interest considered in our study, with cone penetrometer sounding locations shown in white; (c) Estimated standard deviation for each \log_{10} conductivity value in Figure 3b using the method of Alumbaugh and Newman (2000).

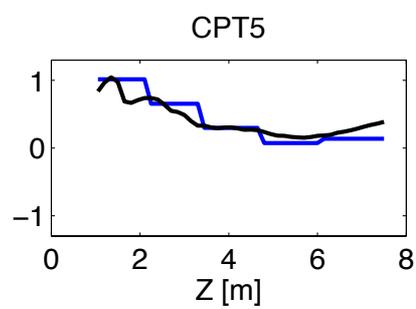
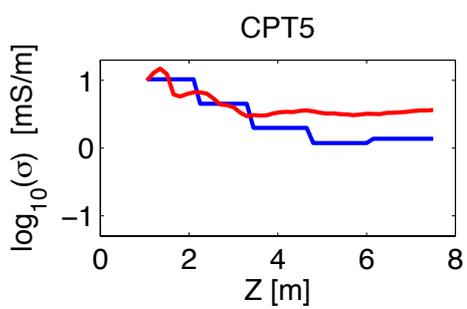
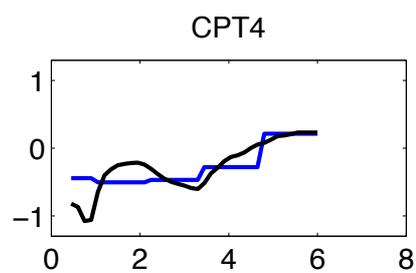
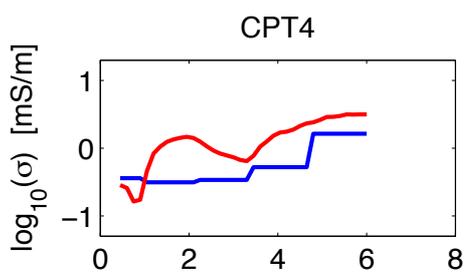
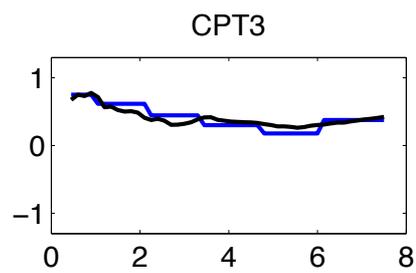
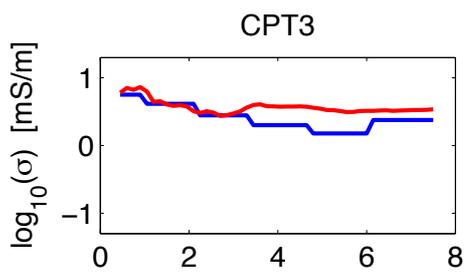
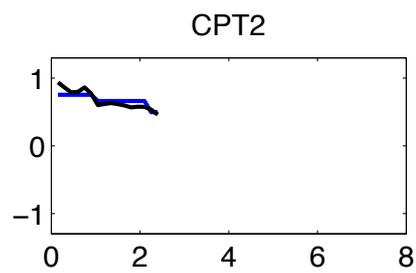
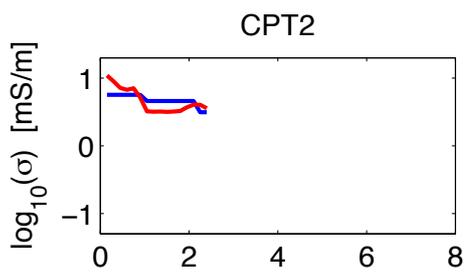
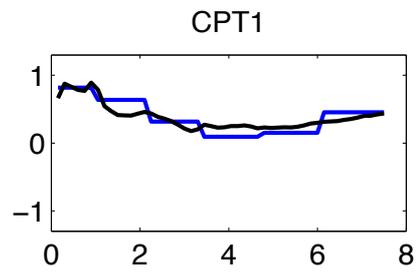
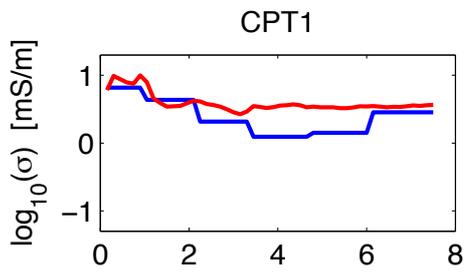


Figure 4: (Left column) Comparison of high-resolution cone-based electrical conductivity measurements at locations CPT1 through CPT5 (red) with collocated estimates of the electrical conductivity from the tomographic image in Figure 3a (blue); (Right column) Comparison after frequency-domain rescaling of the cone-based measurements (black).

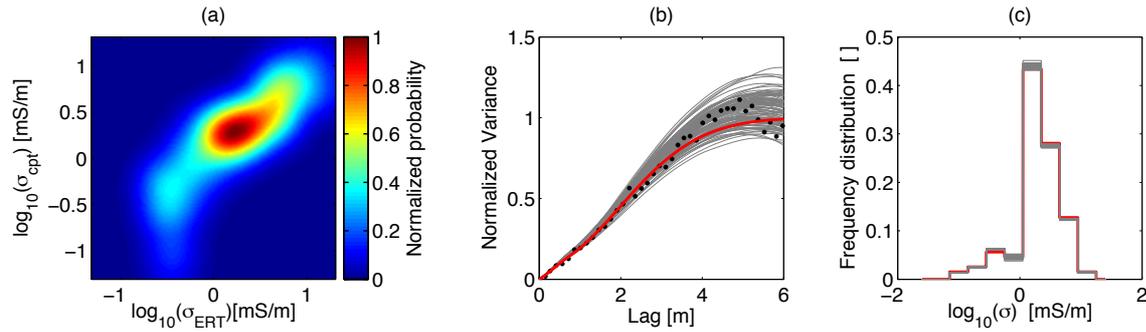


Figure 5: (a) Joint probability distribution for the low- and high-resolution electrical conductivity inferred from 196 collocated ERT and SMR-probe data; (b) Vertical variogram and (c) histogram for the high-resolution electrical conductivity measurements at the borehole locations (red), versus those corresponding to 150 stochastic realizations of the electrical conductivity field obtained using BSS (gray).

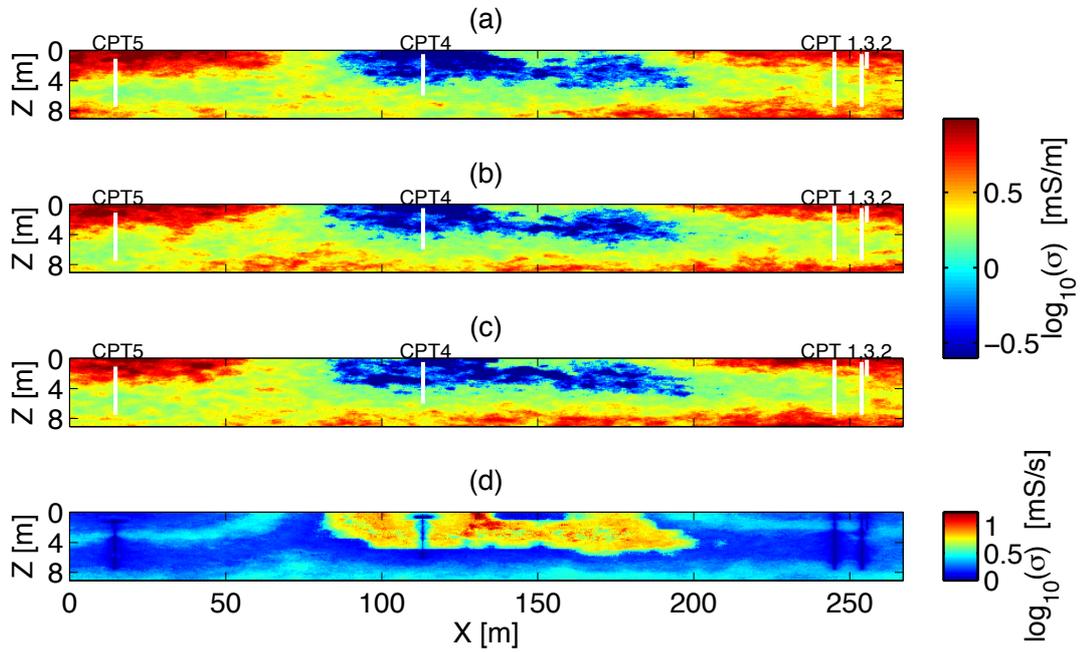


Figure 6: (a-c) Three stochastic realizations of the high-resolution electrical conductivity field obtained using BSS followed by gradual deformation. The BSS procedure was based on the low-resolution ERT inversion results (Figure 3), the high-resolution cone-based electrical conductivity measurements (Figure 4), and the joint relationship established between them (Figure 5a). Cone penetrometer sounding locations are shown in white; (d) image of the 5th-95th inter-percentiles for the high-resolution electrical conductivity, respectively, inferred point-by-point from all 150 output stochastic realizations.

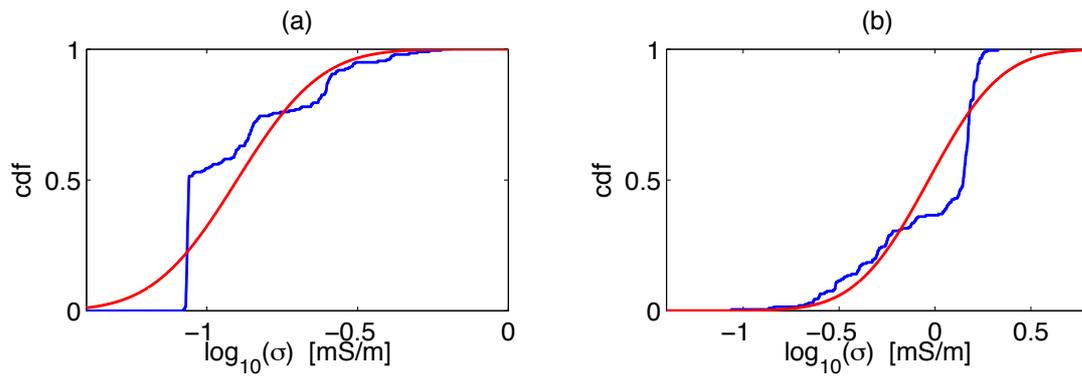


Figure 7: (a and b) Cumulative distribution functions for the high-resolution electrical conductivity at model cell locations ($x = 109.5 \text{ m}$, $z = 0.45 \text{ m}$) and ($x = 160.5 \text{ m}$, $z = 2.1 \text{ m}$), respectively (blue), versus those corresponding to the Gaussian distributions obtained from the ensemble mean and variance at these locations (red).

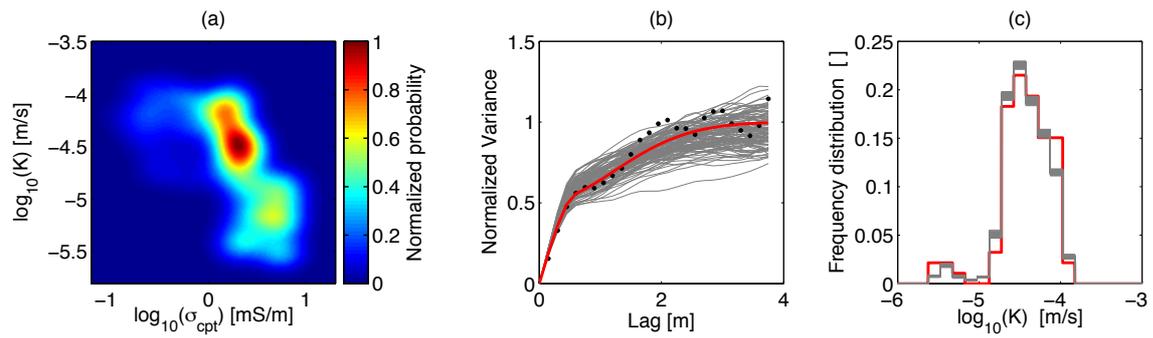


Figure 8: (a) Joint probability distribution for the high-resolution electrical and hydraulic conductivity inferred from 130 collocated measurements. (b) Vertical variogram and (c) histogram for the hydraulic conductivity measurements at the borehole locations (red), versus those corresponding to 150 stochastic realizations of the hydraulic conductivity field obtained using BSS (gray).

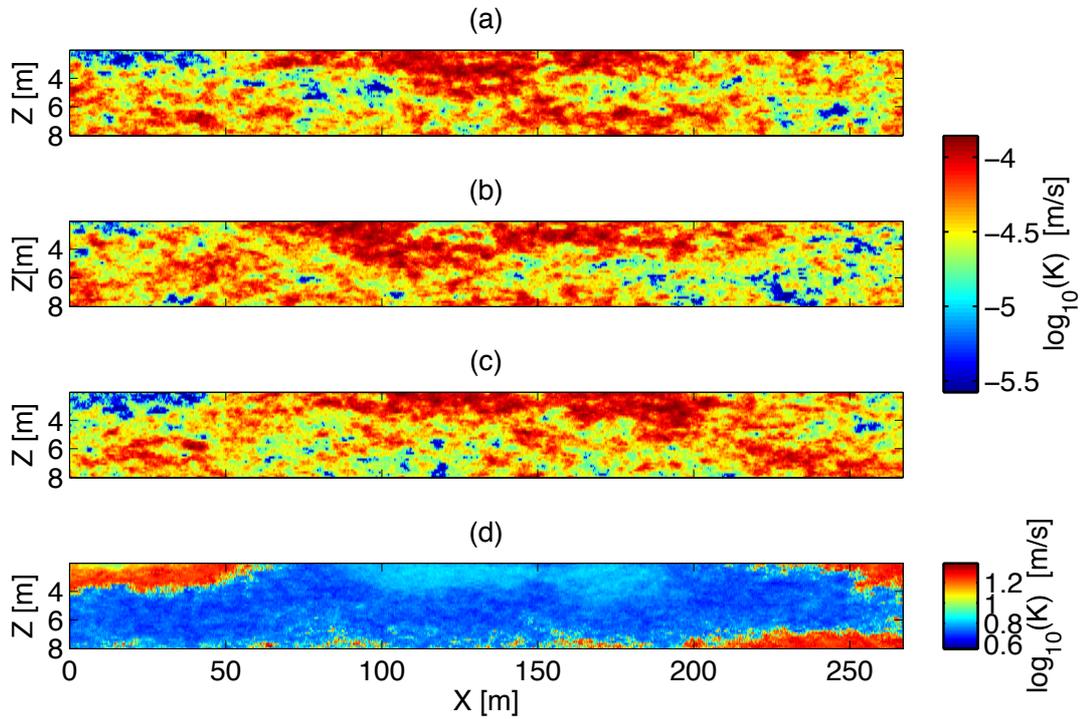


Figure 9: (a-c) Three stochastic realizations of the high-resolution hydraulic conductivity (K) field, obtained using BSS based on the inferred statistics of the high-resolution electrical conductivity field (Figures 6 and 7) and the joint relationship established between the hydraulic and electrical conductivities (Figure 8a). Note that borehole measurements of K along the profile were not used to condition the realizations, such that these measurements could be used for validation purposes; (d) image of the 5th-95th inter-percentiles for the high-resolution hydraulic conductivity, respectively, inferred point-by-point from all 150 output stochastic realizations.

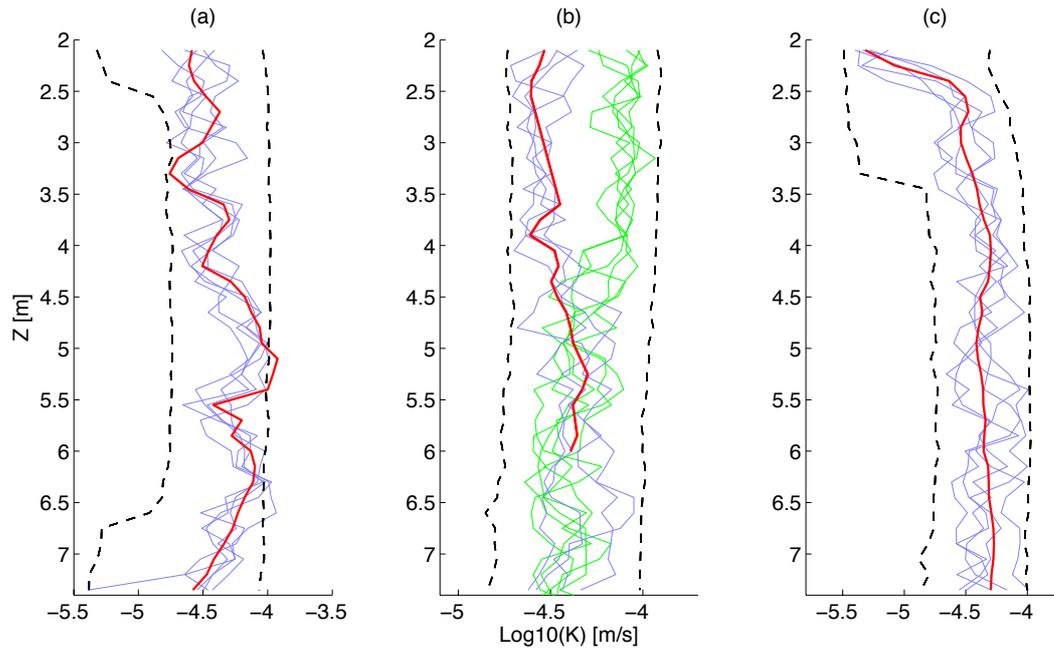


Figure 10: (a) Hydraulic conductivity (K) profile measured at location CPT1 (red) together with five K profiles from stochastic realizations showing the best fit with these data in a root-mean-square sense (blue); (b and c) K profiles predicted using the RVM technique at locations CPT4 and CPT5, respectively (red), along with five profiles from stochastic realizations having the best fit with these results (blue). In Figure 10b, five additional stochastic profiles are provided as examples of the bimodal nature of the K distribution between 2 and 4.5 m depth (green). In all cases, the 5 and 95 percentile bounds for the entire set of K realizations are indicated with dashed black lines.