

Regional-scale integration of multiresolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach

Paolo Ruggeri,¹ James Irving,¹ Erwan Gloaguen² and Klaus Holliger¹

¹*Applied Geophysics Group, Center for Research of the Terrestrial Environment, University of Lausanne, CH-1015 Lausanne, Switzerland.
E-mail: paolo.ruggeri@unil.ch*

²*Institut National de la Recherche Scientifique, Centre Eau, Terre et Environnement, G1K 9A9 Quebec, Canada*

Accepted 2013 February 14. Received 2013 February 14; in original form 2012 June 14

SUMMARY

Significant progress has been made with regard to the quantitative integration of geophysical and hydrological data at the local scale for the purpose of improving predictions of groundwater flow and solute transport. However, extending corresponding approaches to the regional scale still represents one of the major challenges in the domain of hydrogeophysics. To address this problem, we have developed a regional-scale data integration methodology based on a two-step Bayesian sequential simulation approach. Our objective is to generate high-resolution stochastic realizations of the regional-scale hydraulic conductivity field in the common case where there exist spatially exhaustive but poorly resolved measurements of a related geophysical parameter, as well as highly resolved but spatially sparse collocated measurements of this geophysical parameter and the hydraulic conductivity. To integrate this multi-scale, multi-parameter database, we first link the low- and high-resolution geophysical data via a stochastic downscaling procedure. This is followed by relating the downscaled geophysical data to the high-resolution hydraulic conductivity distribution. After outlining the general methodology of the approach, we demonstrate its application to a realistic synthetic example where we consider as data high-resolution measurements of the hydraulic and electrical conductivities at a small number of borehole locations, as well as spatially exhaustive, low-resolution estimates of the electrical conductivity obtained from surface-based electrical resistivity tomography. The different stochastic realizations of the hydraulic conductivity field obtained using our procedure are validated by comparing their solute transport behaviour with that of the underlying “true” hydraulic conductivity field. We find that, even in the presence of strong subsurface heterogeneity, our proposed procedure allows for the generation of faithful representations of the regional-scale hydraulic conductivity structure and reliable predictions of solute transport over long, regional-scale distances.

Key words: Probabilistic forecasting; Downhole methods; Tomography; Hydrogeophysics; Permeability and porosity.

1 INTRODUCTION

Knowledge of the detailed distribution of hydraulic conductivity within an aquifer is a key prerequisite for reliable predictions of solute transport (e.g. Delleur 1999; Chen *et al.* 2001). However, the hydraulic conductivity is an inherently challenging parameter to estimate because its value can vary over many orders of magnitude and it generally exhibits a pronounced degree of spatial heterogeneity (e.g. Ezzedine *et al.* 1999). Traditionally, aquifer characterization has been based on evidence from drill cores, hydraulic borehole logs and slug tests, as well as from tracer and pumping experiments.

Core and logging studies can provide highly detailed local information, but such information is inherently 1-D in nature and sparsely distributed throughout the aquifer volume. Tracer and pumping experiments, on the other hand, tend to capture only the gross average properties of the probed subsurface region (e.g. Rubin 2003). Because of the large gap in terms of spatial coverage and resolution between these techniques, they are often, without complementary information, inadequate for characterizing heterogeneous aquifers (e.g. Sudicky 1986; McKenna & Poeter 1995; Schreibe & Chien 2003; de Marsily *et al.* 2005). Geophysical methods have the potential of bridging the gap in resolution and coverage associated

with traditional hydrological measurements (e.g. Hubbard *et al.* 2001). This has resulted in much interest in the use of these methods for subsurface hydrogeological characterization as evidenced by the rapid growth in the recently established research domain of hydrogeophysics (e.g. Hubbard & Linde 2011). The drawback of using geophysical measurements in a hydrological context, however, is that the underlying material properties governing the geophysical measurements do not exhibit any straightforward link with the hydraulic conductivity (e.g. Hyndman & Tronicke 2005). Indeed, relationships between geophysical properties, such as the electrical conductivity or the seismic wave velocity, and the hydraulic conductivity, tend to be site-, scale- and/or facies-specific, and generally difficult to establish.

Despite a lack of clear and universal relationships between geophysical properties and the hydraulic conductivity, much work has been done at the local scale, typically involving lateral distances on the order of 10–20 m, on the problem of how geophysical data may be effectively used in hydrological investigations. Three types of approaches that have been successfully applied for local-scale aquifer characterization using a combination of downhole and crosshole geophysical methods include: (i) statistical analysis of collocated geophysical and hydrological parameters to establish site- and/or facies-specific petrophysical relationships (e.g. Chen *et al.* 2001; Chen & Rubin 2003; Dafflon *et al.* 2009a,b); (ii) the use of one or more geophysical data sets to perform aquifer zonation, the results of which are then used as a geometrical constraint for subsequent inversions of traditional hydrological test data (e.g. Hyndman *et al.* 1994; Tronicke *et al.* 2004; Paasche *et al.* 2006); and (iii) the use of time-lapse geophysical data to monitor saline tracer transport and/or water infiltration and subsequently infer hydraulic properties (e.g. Singha & Gorelick 2005; Looms *et al.* 2008; Irving & Singha 2010; Scholer *et al.* 2012). How to successfully utilize geophysical methods at a larger, more regional scale, however, remains one of the major ongoing research challenges in hydrogeophysics. At such scales, the large number of model parameters to be resolved, the lack of closely spaced boreholes for effective high-resolution crosshole tomographic imaging, and a scarcity of hydrological test data make direct application of the established local-scale approaches ineffective. Yet the regional-scale problem is of utmost importance because it is at this scale that the greatest benefits of improved predictions of flow and transport can be realized. What is thus needed is the development of practical, computationally feasible, regional-scale methodologies for the quantitative integration of geophysical and hydrological data to characterize subsurface hydrogeological heterogeneity. Such methodologies should be flexible enough to allow for the incorporation of a wide variety of prior information, and they should also allow for an assessment of uncertainty in the results obtained, such that we can most effectively cope with the relative paucity of subsurface information at the regional scale as compared to local-scale studies.

Geostatistics offers an effective and proven framework for integrating diverse sources of information for the purpose of characterizing spatial heterogeneity in subsurface properties (e.g. Isaaks & Srivastava 1989; Goovaerts 1997; Armstrong 1998; Chilès & Delfiner 1999; Deutsch 2002). 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 vastly different degrees of resolution and ‘hardness’ as well as to the deterministic or stochastic quantification of information (e.g. Journel & Huijbregts 1978; Kelkar & Perez 2002; Caers 2005). Further, through the use of conditional simulation, geostatistical methods can provide results in the form

of multiple plausible realizations of subsurface properties, which importantly allows for the assessment of parameter and prediction uncertainty. Although geostatistical methods have seen significant use in the field of groundwater hydrology, both at the local and regional scales (Gomez-Hernandez 2005), their use in the context of integrating geophysical and hydrological measurements to characterize heterogeneous aquifers, to the best of our knowledge, has so far been limited to local-scale investigations where high-resolution crosshole tomographic data and extensive borehole measurements are normally available (e.g., Dafflon *et al.* 2009a,b; Looms *et al.* 2010).

In this paper, we develop a geostatistical approach for the integration of hydrological and geophysical data at the regional scale. The overall objective of this approach is to provide, in a computationally efficient manner, stochastic realizations of the hydraulic conductivity field that are detailed and accurate enough to allow for reliable predictions of flow and transport phenomena over relatively large (~100 m to 1 km) distances. Our work is motivated by the fact that, when dealing with regional-scale problems, we often 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 or airborne regional geophysical data having extensive spatial coverage but with greatly reduced spatial resolution (e.g. Goldman *et al.* 2005). The standard regional-scale hydrogeophysical database is thus characterized by vast differences in coverage and resolution, which represents an additional challenge past the difficult-to-establish link between geophysical properties and the hydraulic conductivity. To integrate such a database, we develop herein a two-step approach based on a geostatistical conditional simulation technique known as Bayesian sequential simulation (Doyen & Boer 1996). The first step of this approach involves downscaling the regional-scale geophysical measurements to the resolution of the borehole data. The second step then involves linking the downscaled geophysical parameter field with the hydraulic conductivity. We begin by outlining the basic principle of our approach within a general methodological framework. We then proceed to apply it to a pertinent and realistic synthetic database consisting of collocated high-resolution borehole measurements of the hydraulic and electrical conductivities, as well as low-resolution estimates of the electrical resistivity obtained from surface-based electrical resistivity tomography (ERT). Finally, the overall validity of our approach is tested by comparing regional-scale tracer transport simulations conducted through the highly heterogeneous, known, ‘true’ hydraulic conductivity field with those conducted through the obtained stochastic realizations of the hydraulic conductivity. Please note that while this methodological study considers a 2-D example, an extension of the proposed algorithm to 3-D scenarios, which could be particularly interesting for the assimilation of vast airborne electromagnetic data and well data, is methodologically straightforward, albeit computationally costly.

2 METHODOLOGICAL BACKGROUND

2.1 Bayesian sequential simulation

The Bayesian sequential simulation technique was initially proposed by Doyen & Boer (1996) for the non-linear geostatistical interpolation and extrapolation of lithological data. The overall objective of this technique 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 the model space 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 methods, 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 Bayesian sequential simulation technique (Doyen & Boer 1996; Doyen 2007):

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

where A and B denote the primary and secondary variables, respectively, $p(\bullet)$ denotes a probability distribution and c is a normalization constant. The conditional distribution $p(A_n | A_1, \dots, A_{n-1})$ in this equation 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 it is obtained by simple kriging of those values to yield a Gaussian mean and variance at the chosen location. The conditional distribution $p(B_n | A_n)$ represents the Bayesian 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 $p(A_n | B_n, A_1, \dots, A_{n-1})$ represents the Bayesian 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 function at that location.

In our work, we determine the Bayesian likelihood function, $p(B_n | A_n)$, in eq. (1) 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 use a non-parametric density estimation approach where $p(A, B)$ is calculated as follows (Silverman 1986):

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

where N is the number of available collocated measurements, $k(\bullet)$ is a positive kernel density function, l_1 and l_2 denote the kernel bandwidths for the primary and secondary variables, respectively and A_i and B_i are the collocated data. Following Dubreuil-Boisclair *et al.* (2011), we choose l_1 and l_2 using the method of Silverman (1986) and we consider a Gaussian kernel for the density estimation procedure, which is given by (Wand & Jones 1995)

$$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \quad (3)$$

A key feature of our Bayesian sequential simulation approach in comparison to other related work (Doyen & Boer 1996; Gastaldi & Roy 1998; Dubreuil-Boisclair *et al.* 2011) is that, instead of evaluating $p(B_n | A_n)$ by simply extracting a single 1-D slice from the estimated joint distribution (i.e. by considering $p(A, B)$ where $B = B_n$), we account for uncertainties in the measured secondary variable by taking a weighted sum of all of the 1-D slices from this

distribution. The corresponding weights are defined by the estimated distribution of errors in the secondary variable at the chosen location (Fig. 1).

Multiplying the Bayesian likelihood function with the prior distribution yields the posterior probability for cell n (Eq. 1). 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. The general Bayesian sequential simulation algorithm used in our work is thus summarized as follows (Fig. 1):

- (1) Randomly select a cell in the discretized model space for which the value of the primary variable is not known.
- (2) Apply simple kriging to the measured and previously simulated values of the primary variable to obtain the mean and standard deviation of the assumed Gaussian prior distribution for the cell of interest.
- (3) Determine the Bayesian likelihood function from available collocated measurements of the primary and secondary variables using the methodology described above.
- (4) Estimate the posterior probability density for the primary variable at the selected location using eq. (1).
- (5) Draw a value for the primary variable from the posterior distribution and treat it as a new reference value.
- (6) Repeat steps (1)–(5) until all cells in the model domain have been simulated.

Note that multiple realizations of the primary variable are generated by repeating the above procedure. This importantly allows for estimation of the corresponding posterior ensemble uncertainty, which is discussed further below. Also note that, in using eq. (2) to determine the Bayesian likelihood function, our Bayesian sequential simulation procedure remains completely flexible with regard to the relationship that exists between the primary and secondary variables, in the sense that this relationship is estimated statistically based on available collocated data. In other words, the specific form and uncertainty of the relationship between the primary and secondary variables are determined empirically, and the quality of this relationship will be reflected in the variability of the generated stochastic realizations. Finally, it is important to emphasize that the Bayesian sequential simulation methodology can also be used in subsurface environments containing more than one relationship between the primary and secondary variables, such as, for example, in the case where different hydrological units exhibiting different relationships are present. The only caveat in this case is that $p(A, B)$ and the corresponding Bayesian likelihood function must be estimated separately for each unit, again based on collocated measurements within that unit (e.g. Dubreuil-Boisclair 2012).

2.2 Application to the regional-scale integration of hydrological and geophysical data

As mentioned previously, we consider in this paper the relatively common situation at the regional scale where we have access to vertically highly resolved, but spatially sparse, borehole-based measurements of both the hydraulic conductivity and a related geophysical parameter, as well as to extensive low-resolution estimates of the same geophysical parameter throughout the domain of interest obtained from surface-based or airborne geophysical data (e.g. Goldman *et al.* 2005; Siemon *et al.* 2009). For the sake of generality, our discussion throughout the rest of this section does not specify

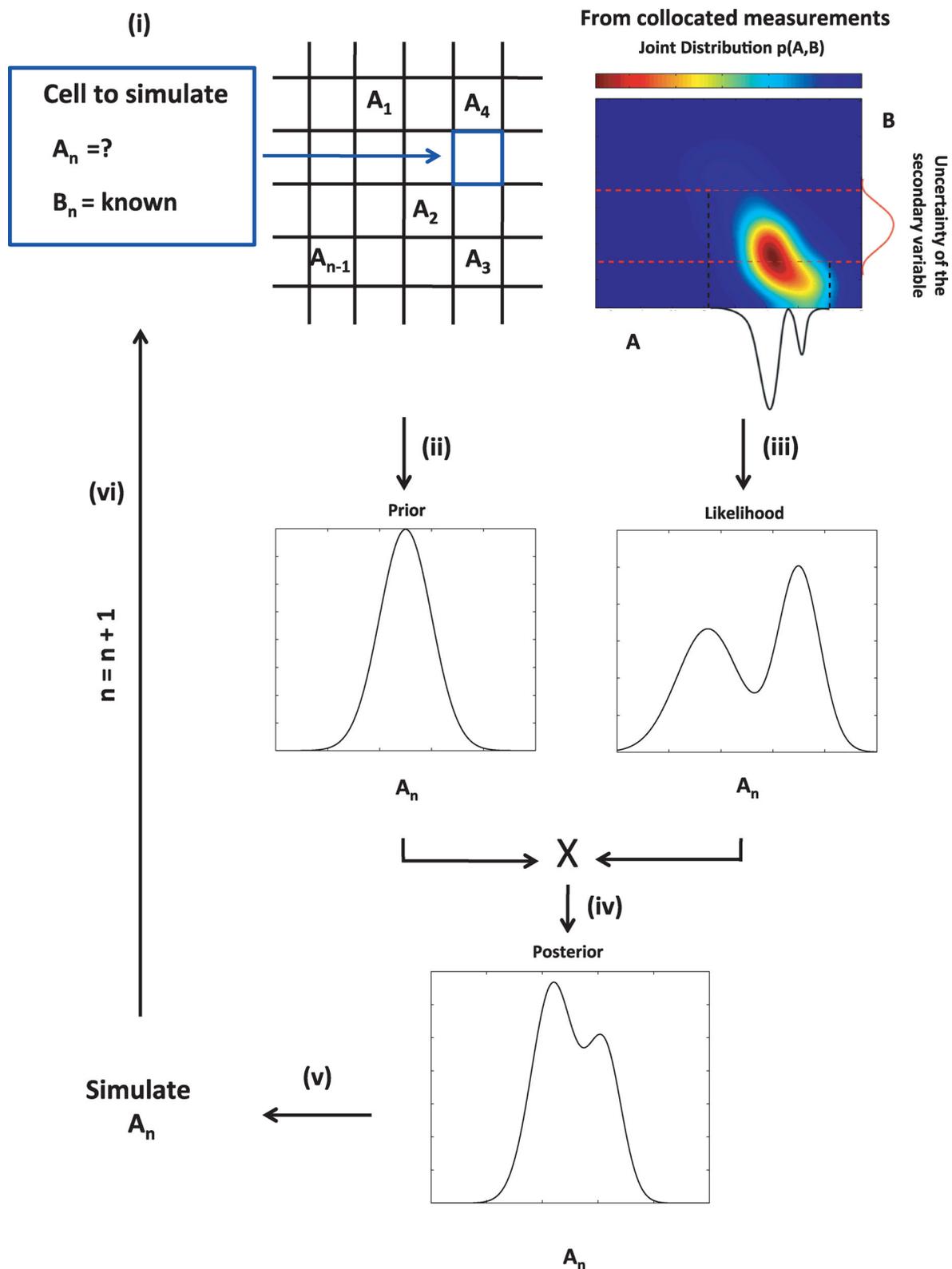


Figure 1. Schematic outline of the general Bayesian sequential simulation procedure considered in this study: (i) selection of the unknown cell to simulate; (ii) estimation of the prior distribution for that cell by kriging measured and previously simulated values of the primary variable; (iii) estimation of the joint probability distribution of the primary and secondary variables using Gaussian kernel density estimation based on collocated measurements and subsequent evaluation of the Bayesian likelihood function; (iv) determination of the posterior distribution by multiplying the prior and likelihood functions and (v) random simulation of the primary variable in the chosen cell according to the posterior distribution.

what geophysical parameter is being considered. Note, however, that arguably the most effective and common parameter in this regard is the electrical conductivity, 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 regional scale (e.g. Binley & Kemna 2005; Lesmes & Friedman 2005; Siemon *et al.* 2009). Below we describe the two-step Bayesian-sequential-simulation-based approach that we have developed in order to assimilate such a varied multiresolution database. This approach involves first linking the low- and high-resolution geophysical data via a stochastic downscaling procedure. The downscaled geophysical estimates are then used to generate stochastic realizations of the high-resolution hydraulic conductivity field.

2.2.1 Downscaling of the regional-scale geophysical information

In the first step of our approach, we perform Bayesian sequential simulation based on the high-resolution borehole geophysical measurements (primary variable) and low-resolution regional geophysical parameter estimates (secondary variable) in order to generate fine-scale realizations of the underlying geophysical parameter field. In other words, we aim in this step to quantify our uncertainty in the fine-scale field conditional to the low- and high-resolution geophysical data. This can be also regarded as stochastically downscaling the regional geophysical measurements to the resolution of the borehole logs. Eq. (1) is easily adapted to this objective and takes the form

$$p(g_n | G_n, g_1, \dots, g_{n-1}) = c \cdot p(G_n | g_n) \cdot p(g_n | g_1, \dots, g_{n-1}), \quad (4)$$

where g and G denote the high- and low-resolution geophysical measurements, respectively. It is important to emphasize that, for inclusion into the Bayesian sequential simulation procedure, we consider the regional geophysical data in the form of an already-inverted tomographic image of their governing geophysical parameter. That is, G in eq. (4) represents the results of inverting the regional data for a low-resolution geophysical image, which can be considered as a set of uncertain measurements of the spatially averaged ‘true’ fine-scale geophysical parameter field. In this case, the Bayesian likelihood function is estimated from values in the tomographic image that are collocated with the high-resolution geophysical measurements at the borehole locations.

Although other more formal strategies might be considered for assessing the uncertainty in the fine-scale geophysical parameter field conditional to the surface- and borehole-based geophysical measurements, one example being to invert multiple times the surface-based data using a fine spatial discretization within some kind of geostatistical framework (e.g. Johnson *et al.* 2007), we have found that there are a number of practical advantages to using Bayesian sequential simulation for this purpose. First and foremost, because we perform only one inversion of the surface-based geophysical data, which is generally done using a much coarser model parametrization than the vertical sampling of the borehole measurements, and because we subsequently use this result as conditioning information in the sequential simulation procedure, the generation of stochastic realizations is computationally efficient. Secondly, the Bayesian sequential simulation approach is naturally well suited to the incorporation of hard data, such as the borehole measurements, along with a wide variety of statistical constraints, into the output realizations through the use of a kriging-based prior. Finally, although the general geophysical inverse problem is non-linear and non-unique, especially with regard to a finely discretized geophysical

parameter field, the use of a relatively coarse discretization in the inversion procedure reduces these effects and provides us with a low-resolution image of average geophysical properties throughout the model space. The point of the Bayesian sequential simulation procedure is to then ‘fill in’ the missing high-frequency information based on the observed relationship between the low- and high-resolution data at the borehole locations, as well as information provided regarding the second-order statistics of the high-resolution target geophysical parameter field.

With regard to the fine-scale geophysical property realizations generated using this first application of Bayesian sequential simulation, one potential concern is that, in some cases, after filling in the high-frequency information, the realizations may no longer provide an acceptable fit to the original low-resolution geophysical measurements from which they were derived. That is, although the high-resolution geophysical parameter fields obtained using Bayesian sequential simulation will have been conditioned by the regional-scale tomographic geophysical image, there is no guarantee that they will honour the original geophysical data to an acceptable degree. As a result, it is possible that the uncertainty expressed by the multiple geophysical realizations will not be adequately represented. Through extensive testing on a wide range of electrical conductivity models and considering synthetic surface-based ERT measurements, we have found that this rarely occurs and that the stochastic realizations generated with Bayesian sequential simulation tend to offer a reasonable match to the underlying conditioning data. However, for completeness, we present here an additional step involving a gradual deformation procedure that allows us to generate stochastic realizations that fit, to a desired degree, the original regional geophysical measurements while keeping the overall large-scale structure, ensemble statistics and covariance properties of the Bayesian sequential simulation results.

Gradual deformation is a procedure by which different stochastic realizations of a parameter field, all corresponding to the same set of conditioning data and thus having the same overall statistical properties, are iteratively linearly combined in order to obtain a satisfactory match to some set of observed data that depends upon the parameter field. It is in essence an optimization process that allows us to iteratively ‘warp’ or ‘deform’ one realization towards another until a new realization is eventually obtained that allows for adequate predictions of the measurements of interest. The key to the gradual deformation method lies in the way that the realizations are combined and how the corresponding weights are chosen, which are done so as to preserve certain statistical properties of the input fields. The original form of the technique was developed by Roggero & Hu (1998) for wellfield history matching in hydrocarbon exploration under the constraint that the optimized permeability fields adhere to some specified mean and variogram function. Many variations on this approach have since been presented (e.g. Roggero & Hu 1998; Hu 2000, 2002; Hu *et al.* 2001; Ravelec-Dupin *et al.* 2002), but all follow the same basic principle that solutions to the inverse problem we seek to solve are contained in a subspace of linear combinations of the input realizations that preserves some of their characteristics. For our work, we consider the gradual deformation approach of Ying & Gomez-Hernandez (2000), which allows for conditioning of the output realizations to hard data in addition to preserving the mean and covariance properties of the input fields. For details on the specific implementation of their approach, please see the corresponding paper. We use the technique to obtain a set of fine-scale geophysical property realizations having the same overall structure and covariance properties as those obtained through the Bayesian sequential simulation procedure, but with the

additional advantage of adequately fitting the observed regional geophysical measurements. The point-by-point mean and variance obtained from these realizations can then be used as conditioning data in a second application of Bayesian sequential simulation in order to generate fine-scale stochastic realizations of the hydraulic conductivity field.

2.2.2 Linking the geophysical and hydrological parameters

In the second step of our regional-scale data integration approach, we again perform Bayesian sequential simulation, but this time the procedure is conditioned to the borehole measurements of the hydraulic conductivity (primary variable) and the point-by-point mean and variance of the fine-scale geophysical parameter fields obtained as described above (secondary variable) with the goal of generating high-resolution realizations of the hydraulic conductivity. In other words, after stochastically downscaling the regional geophysical measurements, we aim in this step to use the geophysical information to condition the fine-scale hydraulic conductivity field. Again, eq. (1) can be readily adapted for this purpose and takes the form

$$p(K_n | g_n, K_1, \dots, K_{n-1}) = c \cdot p(g_n | K_n) \cdot p(K_n | K_1, \dots, K_{n-1}), \quad (5)$$

where K denotes the hydraulic conductivity and g denotes the geophysical parameter. Note that, in this case, the Bayesian likelihood function is determined from the collocated high-resolution borehole measurements of K and g .

A major distinction of the methodological approach proposed in this study in comparison to previous related work (Doyen & Boer 1996; Gastaldi & Roy 1998; Dubreuil-Boisclair *et al.* 2011) is that the latter studies perform an upscaling procedure prior to a single application of Bayesian sequential simulation. This upscaling procedure is governed by the variable with the lowest resolution and thus the sequential simulation is performed on a correspondingly coarse grid. Conversely, our two-step Bayesian sequential simulation approach first downscales the low-resolution geophysical measurements to the scale of the borehole data. The resulting fine-scale discretization serves as the basis for estimating the relationship between the geophysical and hydrological variables, and for generating stochastic realizations of the high-resolution hydraulic conductivity field. Flow and transport simulations can be expected to benefit from this procedure, as we are able to better account for the lateral continuity and/or structural complexity of the underlying ‘true’ hydraulic conductivity distribution.

3 APPLICATION TO A SYNTHETIC EXAMPLE

In the following, we apply the regional-scale data integration approach outlined above to a synthetic, but arguably quite typical, multiple-scale geophysical and hydrological database. This database consists of spatially sparse, high-resolution borehole measurements of the hydraulic and electrical conductivities as well as spatially exhaustive, low-resolution estimates of the electrical conductivity obtained from surface-based ERT. The ultimate objective of integrating these data is to obtain an adequate stochastic description of the fine-scale spatial distribution of hydraulic conductivity throughout the probed subsurface region. To assess the overall success of our characterization effort in a hydrologically meaningful manner, we compare solute transport simulations performed through the multiple generated hydraulic conductivity realizations with the corresponding simulation performed through the original ‘true’ hydraulic conductivity field.

3.1 Hydrological and geophysical database

The 2-D heterogeneous synthetic hydraulic conductivity field that we consider for this example, which can be regarded as a realistic first-order abstraction of many surficial alluvial aquifers, is shown in Fig. 2a (e.g. Gelhar 1993; Hubbard *et al.* 2001; Heinz *et al.* 2003). The considered model 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. The mean and standard deviation of the $\log_{10}(K)$ [m/s] distribution are -3.18 and 0.36 , respectively. To generate this highly heterogeneous model, we used a spectral stochastic simulation approach (e.g. Goff & Jennings 1999) and we assumed an exponential variogram model for $\log_{10}(K)$ having horizontal and vertical correlation lengths of 27 and 2.7 m, respectively. The chosen values for the horizontal and vertical correlation lengths thus correspond to a structural aspect ratio of 10, which is consistent with the predominantly layered nature of surficial alluvial deposits (e.g. Gelhar 1993). Full saturation is assumed to prevail throughout this aquifer model.

To simulate the spatial distribution of the porosity ϕ in the considered subsurface region, we assumed a linear relation between porosity and the logarithm of hydraulic conductivity of the form $\log_{10}(K) = 6.66\phi - 4.97$ (Heinz *et al.* 2003). We then added correlated zero-mean Gaussian random noise characterized by an exponential variogram model having horizontal and vertical correlation lengths of 3.5 and 1 m, respectively, and a standard deviation of

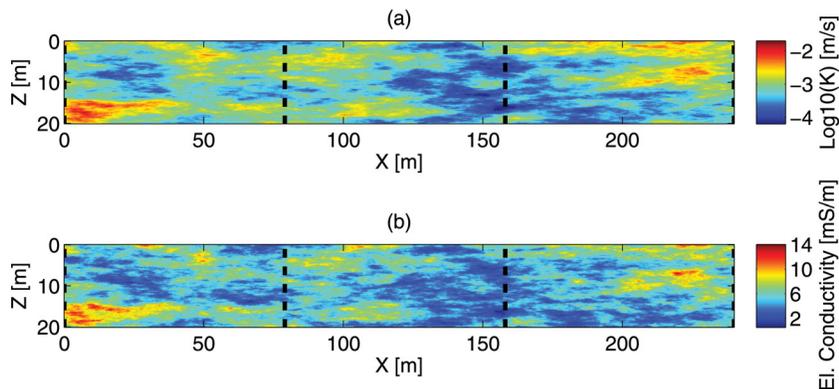


Figure 2. Heterogeneous spatial distributions of (a) the hydraulic conductivity and (b) the electrical conductivity for the synthetic aquifer model considered in this study. The considered borehole locations are shown as dashed lines.

0.025, in order to reflect more realistic conditions (Dafflon *et al.* 2010). Next, to simulate the spatial distribution of the electrical conductivity σ , we used Archie's (1942) law for saturated media

$$\sigma = \sigma_w \phi^m, \quad (6)$$

where σ_w denotes the electrical conductivity of the pore water and m is the Archie cementation exponent. We set $\sigma_w = 43$ mS/m and $m = 1.4$, which are realistic values for unconsolidated alluvial aquifer environments (e.g. Keller & Frischknecht 1966; Schön 2004; Lesmes & Friedman 2005). Again, correlated zero-mean Gaussian random noise characterized by an exponential variogram model having horizontal and vertical correlation lengths of 3.5 and 1 m, respectively, and a standard deviation of 0.25 mS/m, was added to the corresponding results such that a significant degree of uncertainty was introduced into the final relationship between the hydraulic and electrical conductivities. The resulting electrical conductivity field is shown in Fig. 2b.

It is important to note that the petrophysical relationships assumed above between the hydraulic conductivity and the porosity, as well as between the porosity and the electrical conductivity, are only valid in the absence of significant amounts of clay (Freeze & Cherry 1979; Schön 2004). We chose to work with this admittedly idealized petrophysical scenario in our synthetic example because of its conceptual simplicity and strong relevance to the case of surficial alluvial aquifers (e.g. Gelhar 1993; Hubbard *et al.* 2001; Heinz *et al.* 2003; Schön 2004; Lesmes & Friedman 2005; Dafflon *et al.* 2009b). Please note, however, that the proposed data integration approach is by no means limited to situations where there exists a well-defined relationship between the considered geophysical parameter and the hydraulic conductivity. Indeed, our approach has the distinct advantage of not relying on any specific parametric relationship between these parameters because their joint distribution is determined empirically from collocated data using the previously described non-parametric density estimation approach. In this regard, as mentioned previously, the only requirement needed in the estimation process is to work within a single hydrological unit, where stationarity of the relationship between the geophysical property and the hydraulic conductivity can be assumed. In cases where great uncertainty is found to exist in this relationship, a high degree of variability will be observed in the output realizations obtained with the Bayesian sequential simulation technique.

Having specified the detailed subsurface hydraulic and electrical conductivity distributions, we next simulated the acquisition

of high-resolution electrical and hydraulic conductivity measurements along four boreholes, which were considered to be located at lateral distances of 0, 80, 160, and 240 m from the left-hand model edge (Fig. 2). The spacing between the boreholes is significantly larger than the lateral correlation length of the heterogeneous model structure, which means that these data will contain little to no information regarding the lateral variability. The vertical resolution of the borehole measurements was assumed to be equal to one grid cell, or 20 cm, yielding a total number of 400 data for the electrical and hydraulic conductivities. Note that the number of borehole data is substantially smaller than the total number of model parameters in the fine-scale grid (120 000) upon which we wish to generate stochastic realizations of the electrical and hydraulic conductivities.

Finally, we simulated the acquisition of low-resolution geoelectrical measurements over the model domain. To this end, we used the spatial distribution of the electrical conductivity shown in Fig. 2b to simulate a surface-based ERT survey having a dipole-dipole acquisition geometry with a minimum electrode spacing of 2.4 m. After adding 5% uncorrelated Gaussian random noise to the resulting apparent resistivity values, they were tomographically inverted on a coarse grid having a constant horizontal discretization of 2.4 m and a variable vertical discretization ranging from 0.8 to 3.2 m with increasing depth. The forward simulation of the electrical resistivity data was carried out using the finite-element-based R2 code (Binley & Kemna 2005), whereas the tomographic inversion was performed using the commercial RES2DINV software (Loke 2012) using least-squares data fitting and model smoothness constraints. A comparison of the resulting low-resolution ERT image (Fig. 3a) with the 'true' heterogeneous electrical conductivity structure (Fig. 2b) clearly illustrates the noticeable spatial smoothing and decrease in model resolution with increasing depth that are typical of tomographic geoelectrical reconstructions (Binley & Kemna 2005). To estimate the uncertainty in the inverted resistivity values, we used the method proposed by Alumbaugh & Newman (2000), which is based on the diagonal elements of the inverted model covariance matrix, approximated by

$$C_m \cong (J^T J + \lambda F)^{-1}, \quad (7)$$

where J is the Jacobian or sensitivity matrix about the inverted model, T denotes the matrix transpose, F is the model regularization or smoothing matrix and λ is the considered trade-off parameter which, at a given iteration, weighs model smoothness against data fit. The estimated uncertainties obtained using eq. (7), expressed

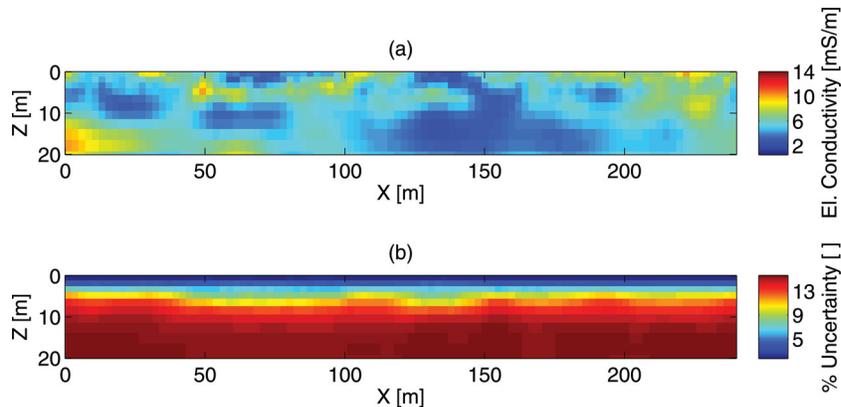


Figure 3. (a) Low-resolution electrical conductivity tomogram obtained from the inversion of surface-based dipole-dipole geoelectrical measurements simulated over the 'true' electrical conductivity model shown in Fig. 2b. (b) Estimated percentage uncertainty in each of the values in Fig. 3a based on the method of Alumbaugh & Newman (2000).

as a percentage of the corresponding electrical conductivity values (Fig. 3b), can be seen to be highest in the parts of the model space that are characterized by particularly low-spatial resolution.

3.2 Regional-scale data integration procedure

Using the high-resolution borehole measurements of the electrical conductivity and the low-resolution surface-based ERT conductivity estimates available everywhere throughout the model space, the first step of our regional-scale data integration procedure is to generate multiple high-resolution realizations of the electrical conductivity field that are consistent with these observations. Again, this is done through a combination of Bayesian sequential simulation and gradual deformation. As described previously, the Gaussian prior distribution $p(g_n|g_1 \dots g_{n-1})$ at each iteration of the Bayesian sequential simulation procedure is obtained by simple kriging of the borehole electrical conductivity measurements along with any previously simulated conductivity values. To this end, we inferred the vertical variogram of the high-resolution electrical conductivity field from the borehole measurements themselves. The best-fitting parametric model was found to be a spherical variogram function with a correlation length of 2.6 m and a standard deviation of 1.47 mS/m. The horizontal correlation length, although not directly constrained by the available data, can be estimated, for example, based on the structural aspect ratio of the ERT image or additional geological information (e.g. Gelhar 1993; Troncke & Holliger 2005; Dafflon *et al.* 2009a). Using the former approach, we estimated the aspect ratio to have a value of 9.4. The Bayesian likelihood function $p(G_n|g_n)$ in eq. (4) was obtained from the joint probability density estimated from collocated high- and low-resolution electrical conductivity measurements at the borehole locations. Uncertainty associated with the ERT estimates (Fig. 3b) was accounted for during the evaluation of this likelihood function, thus ensuring that the realizations generated through Bayesian sequential simulation were not too strongly conditioned by the ERT image in regions of significant uncertainty.

Using stochastic realizations of the high-resolution electrical conductivity field generated through the Bayesian sequential simulation procedure, the gradual deformation method was then utilized to create, while preserving the mean and covariance properties of these realizations and conditioning to the borehole data, new realizations for which the predicted apparent resistivity measurements were a suitable match to the original ‘observed’ apparent resistivity data from which the ERT image in Fig. 3b was derived. A suitable fit was defined using the chi-squared statistic

$$\chi^2 = \sum_{j=1}^N \frac{(\rho_j^o - \rho_j^p)^2}{\sigma_j^2}, \quad (8)$$

where N is the number of measurements, ρ_j^o and ρ_j^p are the observed and predicted apparent resistivity, respectively, and σ_j is the measurement error which was defined in this case to be equal to 5% of the observed value. Iterations of the gradual deformation optimization procedure were carried out until $\chi^2 \leq N$, after which point the resulting electrical conductivity field was saved as a suitable ‘optimized’ realization. This procedure was repeated until 100 of such realizations had been created.

A first comparison between the original ‘true’ electrical conductivity field (Fig. 4a) and two stochastic realizations of the corresponding downscaled field obtained through the above procedure (Figs 4b and c) indicates that our Bayesian sequential simulation procedure allows for obtaining good estimates of both the local- and

regional-scale electrical conductivity structure, in the sense that the overall large-scale conductivity patterns are well reproduced and the style of small-scale variability has been properly represented. This observation is also reflected in the point-by-point ensemble mean (Fig. 4d) and variance (Fig. 4e) of the 100 output realizations, the latter of which effectively illustrates how the generated realizations are tied to the conductivity measurements at the borehole locations, and are slightly less constrained at greater depths where the ERT image is less reliable. The statistical accuracy of the generated realizations is further confirmed through a comparison of the histogram and vertical experimental variogram calculated from the electrical conductivity measurements at the four borehole locations (i.e. our target statistics) with those calculated globally from 20 different stochastic realizations (Fig. 5). For each of the chosen realizations, we see that the overall histogram of the borehole electrical conductivity measurements is preserved (Fig. 5a), and that a good fit is obtained to the target vertical variogram (Fig. 5b).

Finally, in order to assess the value of the gradual deformation procedure and its effect on the output realizations obtained, we compare the histogram of the point-by-point ensemble variance, calculated from 100 stochastic realizations that were generated using Bayesian sequential simulation alone, with the corresponding histogram calculated from the realizations obtained with the addition of gradual deformation (Fig. 6). Although the stochastic realizations generated with Bayesian sequential simulation already offer a reasonable match to the conditioning data and a correspondingly reasonable estimate of the model parameter uncertainty because they are based on the tomographic ERT image, we see an overall reduction in variability when the additional step of gradual deformation is considered. That is, we see that there is value to ensuring, through the gradual deformation step, that the stochastic realizations fit the original apparent resistivity data, as these data contain important information that allows us to further reduce the posterior uncertainty as compared to using Bayesian sequential simulation alone.

A potential point of concern with regard to the first application of Bayesian sequential simulation in our regional-scale data integration procedure is how the generated high-resolution stochastic realizations of the governing geophysical parameter, in this case the electrical conductivity, will be influenced by the choice of tomographic image that is considered as secondary information. Tomographic images obtained from the inversion of geophysical data are highly dependent upon a number of different factors, most notably the data acquisition geometry and the style and amount of regularization used. Changes in these factors will clearly result in different joint relationships being established between the low- and high-resolution geophysical data because the set of low-resolution collocated measurements derived from the tomographic geophysical image will change. In turn, differences in the determined joint relationship will affect the generated stochastic realizations. It is important to note, however, that this does not represent a problem because the subsurface variability expressed by the stochastic realizations is based upon what we empirically observe to be a suitable relationship between the low- and high-resolution geophysical data. In other words, 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 worse resolution tomographic image should be compensated by greater variability in the output realizations (i.e. the image provides us with less concrete information regarding the high-resolution properties, and so the output uncertainty will be increased). To illustrate this point, we simulated, in addition to the

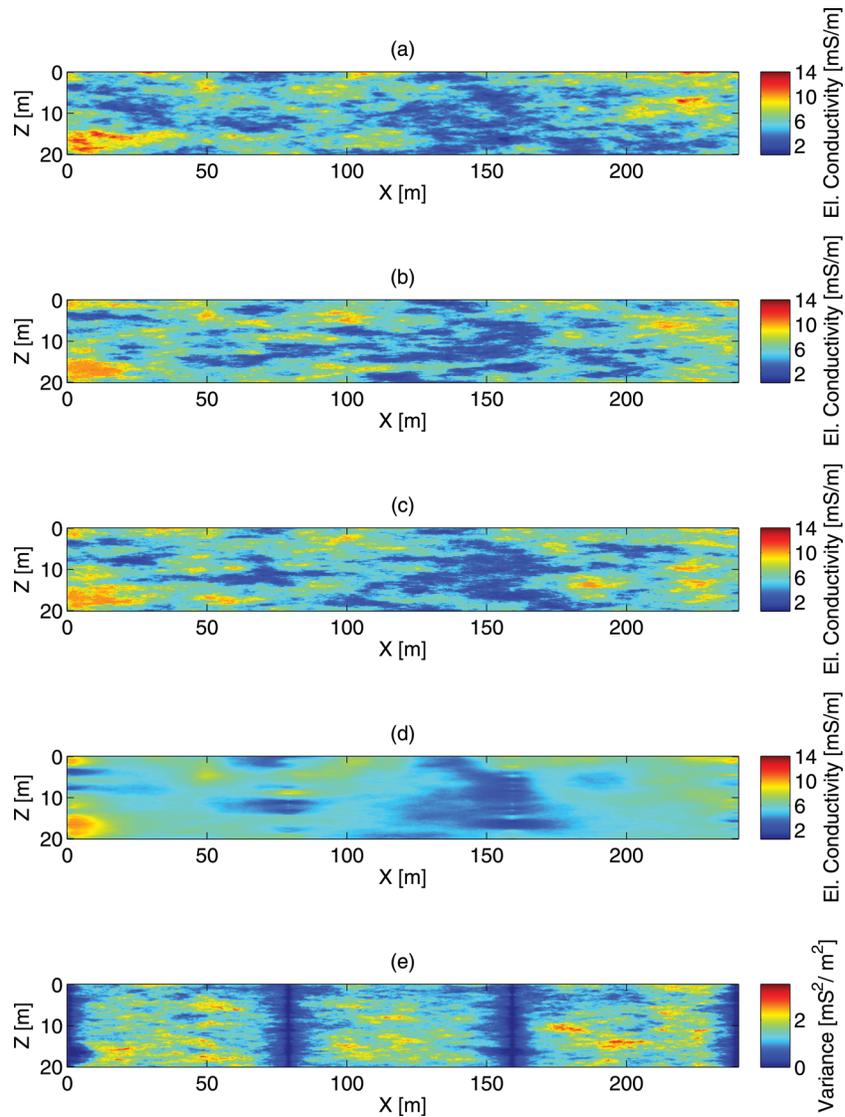


Figure 4. (a) Original ‘true’ high-resolution electrical conductivity field from Fig. 2b. (b) and (c) Two stochastic realizations of the electrical conductivity field that were obtained through Bayesian sequential simulation based on the ERT inversion results in Fig. 3 and the high-resolution conductivity measurements at the four borehole locations, followed by a gradual deformation procedure (d) and (e) Ensemble mean and variance inferred from all of the 100 obtained electrical conductivity realizations, respectively.

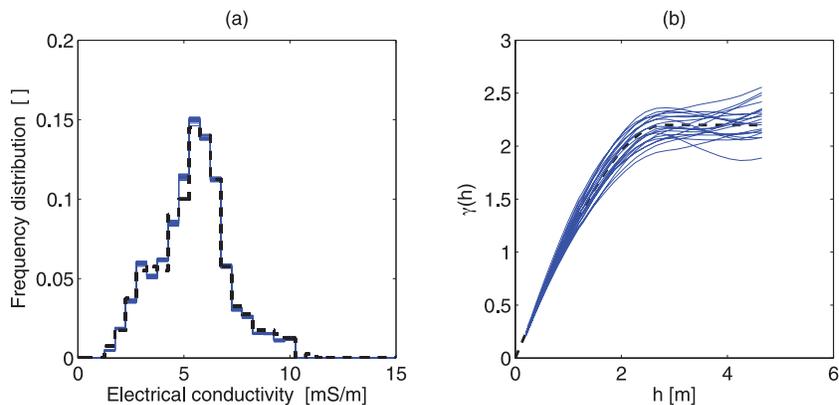


Figure 5. (a) Histogram and (b) vertical variogram calculated from the electrical conductivity measurements at the four borehole locations (dashed black lines), versus those corresponding to 20 stochastic realizations of the electrical conductivity obtained using Bayesian sequential simulation followed by gradual deformation (solid blue lines).

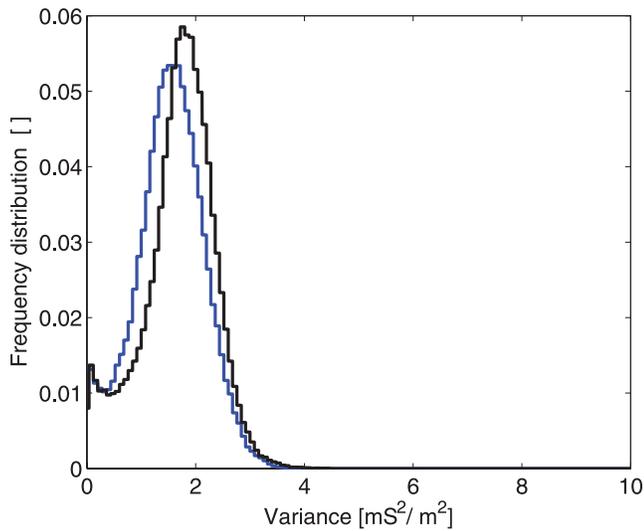


Figure 6. Histogram of the point-by-point ensemble variance of 100 high-resolution electrical conductivity realizations obtained using the Bayesian sequential simulation procedure alone (solid black line), compared with that corresponding to 100 optimized realizations obtained with the additional gradual deformation step (solid blue line).

dipole–dipole ERT data considered above, geoelectrical measurements corresponding to a Wenner–Schlumberger acquisition geometry over the model domain. In Fig. 7, we observe that the electrical conductivity tomogram obtained from the Wenner–Schlumberger data (Fig. 7b) exhibits a lower lateral resolution than the tomographic image obtained from the dipole–dipole data (Fig. 7a). This results from the fact that dipole–dipole configurations are more sensitive to horizontal changes in resistivity (e.g. Dahlin & Loke 1998; Oldenburg & Li 1999). Note, however, that the estimated joint relationship between the low- and high-resolution electrical conductivity is altered between the Wenner–Schlumberger (Fig. 7d) and dipole–dipole configurations (Fig. 7c) to compensate for these image differences. Specifically, the uncertainty in the high-resolution conductivity given a particular measured value of the low-resolution conductivity is greater for the Wenner–Schlumberger configuration than for the dipole–dipole configuration, as indicated by the larger spread of the joint distribution in the horizontal direction in Fig. 7d. Again, this greater uncertainty will be reflected in a higher degree of variability in the generated stochastic realizations.

Using the collocated, high-resolution borehole measurements of the hydraulic and electrical conductivities as well as the point-by-point mean and variance fields of the ensemble of downscaled electrical conductivity realizations, the second step of our regional-scale data integration procedure aims at generating multiple

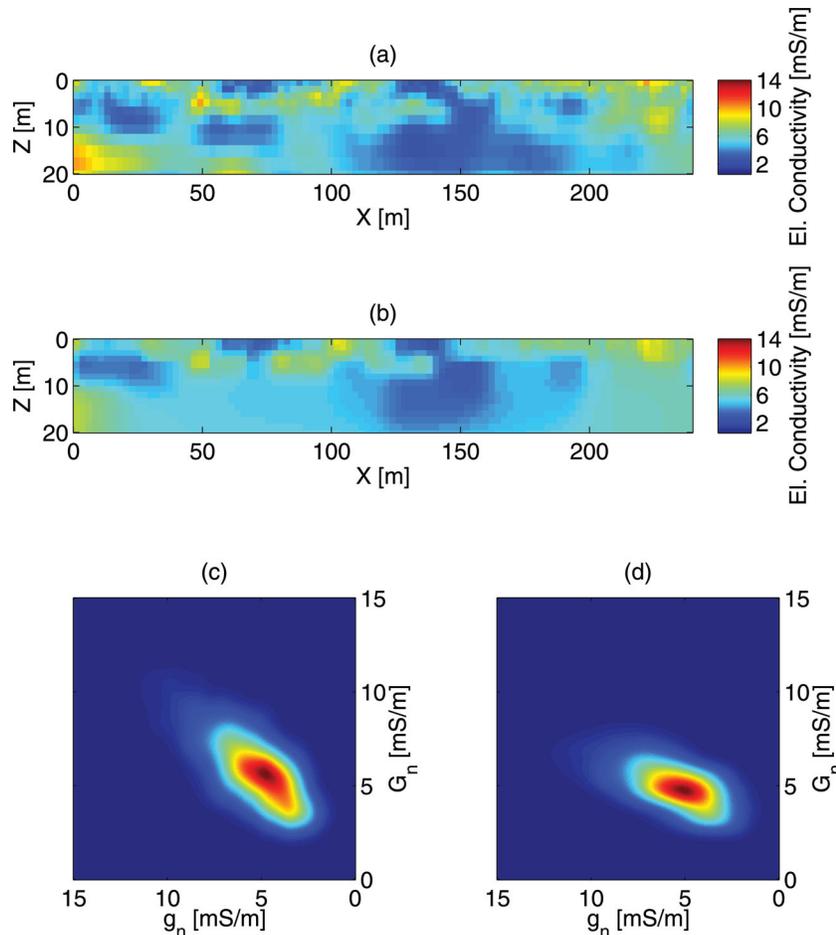


Figure 7. (a) Electrical conductivity tomogram from Fig. 3a, obtained from the inversion of synthetic surface-based geoelectrical measurements having a dipole-dipole acquisition geometry. (b) Electrical conductivity tomogram obtained from the inversion of geoelectrical measurements simulated using a Wenner–Schlumberger acquisition geometry. (c) Joint probability distribution $p(g_n, G_n)$ estimated using the image in (a) and collocated high-resolution electrical conductivity measurements from Fig. 2b. (d) Corresponding joint probability distribution $p(g_n, G_n)$ estimated using the image in (b).

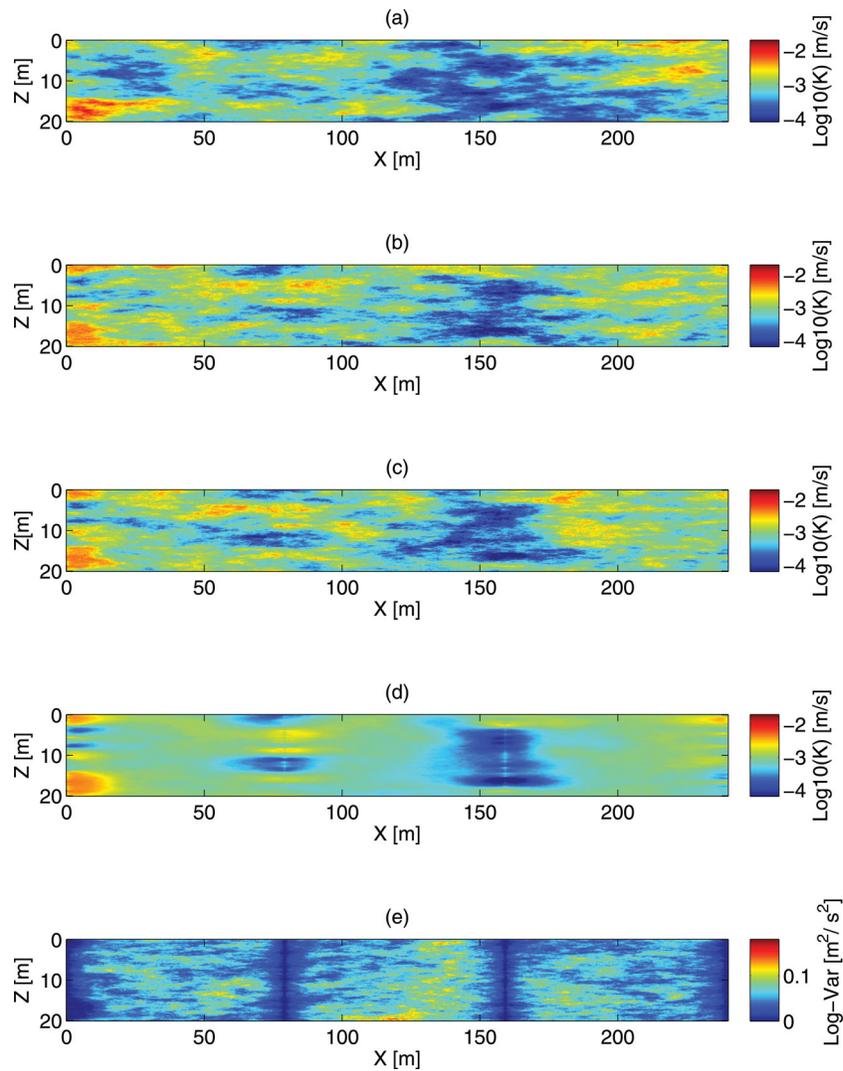


Figure 8. (a) Original ‘true’ high-resolution hydraulic conductivity field from Fig. 2a. (b) and (c) Two stochastic realizations of the hydraulic conductivity field that were obtained through Bayesian sequential simulation based on the electrical conductivity mean and variance in Fig. 4 and high-resolution collocated measurements of the electrical and hydraulic conductivities at the four borehole locations. (d) and (e) Ensemble mean and variance inferred from the 100 obtained hydraulic conductivity realizations, respectively.

high-resolution stochastic realizations of the hydraulic conductivity field. The vertical variogram model used in this case to obtain the kriging-based prior for Bayesian sequential simulation was determined from the borehole measurements of the hydraulic conductivity. This yielded a best-fitting spherical variogram function having a correlation length of 3.1 m and a log-standard deviation of 0.3. The structural aspect ratio of 9.4 previously inferred from the ERT image was again used to constrain the horizontal correlation length. The Bayesian likelihood function in this case was determined from the joint distribution of electrical and hydraulic conductivity estimated from the borehole data.

In Fig. 8, we show the original ‘true’ hydraulic conductivity field (Fig. 8a) together with two representative stochastic realizations obtained through Bayesian sequential simulation (Figs 8b and c) and the point-by-point mean and variance of an ensemble of 100 realizations (Figs 8d and e). These results are largely comparable with those obtained for the electrical conductivity (Fig. 4), whose capacity of reproducing the essential elements of the underlying model has been thoroughly discussed above. A comparison of the overall histogram and vertical experimental variogram of the borehole hy-

draulic conductivity measurements with those calculated globally from 20 different stochastic realizations indicates that the general distribution of the borehole hydraulic conductivity measurements is preserved (Fig. 9a) and a good fit to the target vertical variograms is obtained (Fig. 9b). In their combination, the results shown in Figs 8 and 9 indicate that our Bayesian sequential simulation algorithm allows for obtaining adequate estimates of both the small- and large-scale components of the hydraulic conductivity structure.

3.3 Model validation

To validate the results of our aquifer characterization effort, we simulated the transport of a conservative tracer through the original hydraulic conductivity field with the aim of comparing this ‘true’ transport behaviour with that predicted through 40 stochastic realizations of the hydraulic conductivity that were obtained as described above. 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 boundary conditions

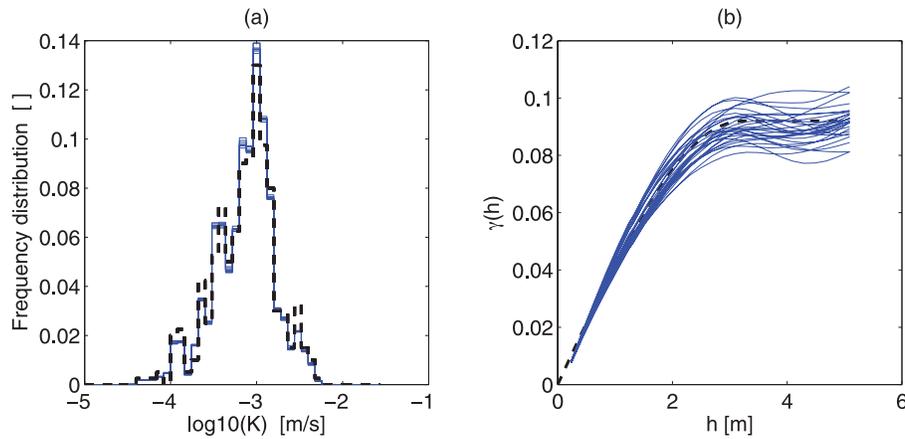


Figure 9. (a) Histogram and (b) vertical variogram calculated from the hydraulic conductivity measurements at the four borehole locations (dashed black lines), versus those corresponding to 20 stochastic realizations of the hydraulic conductivity obtained using Bayesian sequential simulation (solid blue lines).

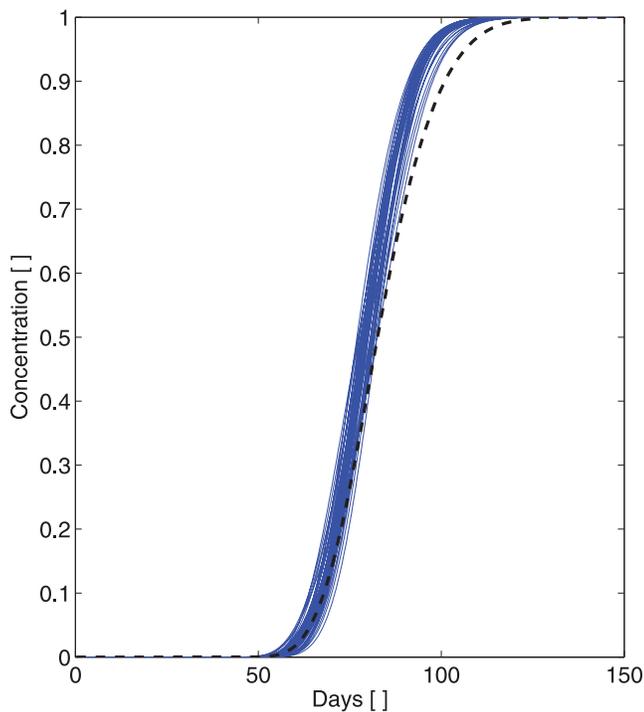


Figure 10. Breakthrough curves showing the normalized average tracer concentration in the borehole at the right-hand model edge as a function of time. The black dashed lines corresponds to the breakthrough curve for the true hydraulic conductivity field in Fig. 2a, whereas the solid blue lines correspond to breakthrough curves for 40 hydraulic conductivity realizations obtained using the proposed Bayesian sequential simulation procedure.

on either side, such that a lateral hydraulic head gradient of 0.013 was produced. A natural-gradient tracer test through the subsurface region was considered. To perform the transport simulations, we solved the transient advection–dispersion equation assuming a fixed tracer concentration of 1000 mg L^{-1} along the left-hand model edge. Measurements of the tracer concentration were then assumed to be available all along the borehole on the right-hand side of the model at 240 m.

The normalized tracer breakthrough curves obtained from the transport simulations are shown in Fig. 10. These curves were calculated for each considered hydraulic conductivity field by taking

the mean of the tracer concentration in the right-most borehole and plotting this as a function of time. Quite importantly, we see that the first-arrival times for the tracer predicted by our stochastic realizations are in good agreement with the corresponding ‘true’ tracer breakthrough time. In other words, the realizations obtained through our regional-scale data integration procedure appear to allow for adequate predictions of this key characteristic of the transport behaviour, and we are able to further assess our uncertainty in the arrival time by considering the range of predictions for the ensemble of stochastic realizations. We do, however, observe a gradual deterioration in the match between the ‘true’ and predicted average tracer concentration curves as time increases. Detailed analysis has shown that this results from the fact that our obtained hydraulic conductivity realizations do not adequately reproduce the poorly hydraulically conductive zone located in the lower part of our subsurface model between lateral distances of 160 and 240 m (Fig. 8). This in turn means that the ‘true’ propagation of the tracer will be slower than that predicted by the stochastic realizations.

The mismatch mentioned above points to an important issue regarding our Bayesian sequential simulation approach that deserves some further discussion. With this approach, realizations of the primary variable are generated by sampling from the Bayesian posterior distribution, which in turn is obtained from the product of the prior distribution and the Bayesian likelihood function. As with any Bayesian method, the quality of the final results obtained is thus strongly dependent upon the adequacy and reliability of the prior information. In general, the accuracy of such prior information will increase with the amount of borehole data that are available. However, if the prior is based, as in our study, on relatively few borehole measurements, then its effectiveness will critically depend upon the statistical representativeness of those data. The poorly hydraulically conductive zone described above, for example, could be readily reproduced by our generated stochastic realizations if the conditioning borehole data had contained enough information for a representative prior distribution. Hence, for improving the hydrological characterization effort, we could increase the number of conditioning data by considering additional boreholes, or we could optimize the positions of the four existing boreholes to be more representative. Whereas the former option would make our test scenario less realistic, there commonly exists pertinent complementary information from previous geological and geophysical surveys that may allow for optimizing of the choice of borehole locations. From a methodological point of view, this indicates that geophysical

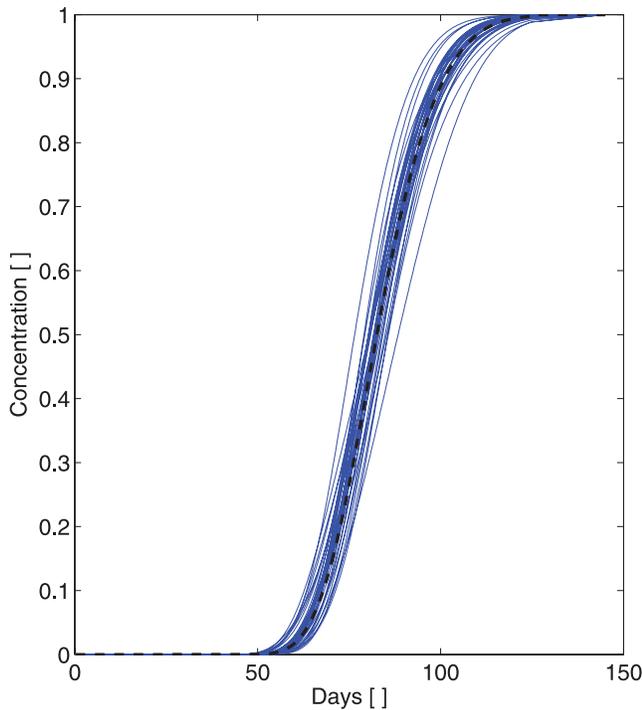


Figure 11. Effect of changing the position of the conditioning borehole measurements on the predicted breakthrough curves. For this set of simulations, the third and fourth boreholes were moved from $x = 160$ m to $x = 155$ m and from $x = 240$ m to $x = 225$ m, respectively.

surveys should, if possible, be completed before drilling boreholes such that this can be accomplished. To illustrate this concept, we moved the third and the fourth boreholes in our synthetic example to 155- and 225-m lateral distance from the left-hand model edge, respectively, and we regenerated several stochastic realizations of the hydraulic conductivity field based on this more representative prior information. Fig. 11 shows the corresponding predicted average tracer concentration in the right-hand borehole as a function of time, which now shows good agreement with that obtained for the ‘true’ hydraulic conductivity model.

4 DISCUSSION

Over the past two decades, a number of geostatistically based downscaling approaches have been developed. In close conceptual analogy with the methodology presented in this study, these approaches have aimed at integrating multiple parameters measured at different scales and resolutions in order to constrain the distribution of one or more subsurface target parameters. Most of the corresponding studies have been geared towards hydrocarbon exploration and hence the key geophysical parameter considered has been the seismic impedance, with the target variable being the porosity. In particular, to integrate regional-scale seismic-based impedance estimates with high-resolution but spatially sparse borehole porosity-log data, the typical workflow has involved (i) upscaling the porosity logs through relatively straightforward local averaging approaches, (ii) establishing the relationship between upscaled porosity and seismic impedance via some form of linear regression, (iii) estimation of the regional-scale porosity field through the use of cokriging-type algorithms and (iv) use of the latter information to constrain the stochastic simulation of the corresponding high-resolution porosity

field (Gorell 1995; Behrens *et al.* 1996; Doyen *et al.* 1997; Lee *et al.* 2002; Gilbert & Andrieux 2003; Malallah *et al.* 2003).

While the above general data integration methodology has become relatively well established in hydrocarbon exploration, it is important to note that its applicability for the integration of typical near-surface geophysical and hydrological data is severely limited by the fact that the upscaling and downscaling of parameters like the hydraulic and electrical conductivities is highly measurement dependent, non-linear and hence not well suited to any kind of straightforward averaging approach. Furthermore, in contrast to when working with seismic impedance, significant differences in the resolution and uncertainty of geophysically derived properties in the near surface (e.g. electrical conductivities estimated from ERT) require the development of more sophisticated techniques for their successful incorporation. In the approach presented in this study, the regional-scale geophysical parameter is not considered as a local average, but rather it is used directly, together with its inferred uncertainty, to constrain the simulation of the corresponding fine-scale geophysical parameter field. As such, the local uncertainty of the inverted geophysical parameter directly determines the strength with which the simulations are conditioned by these data. By assessing the interrelations between parameters based on a joint probability function computed from collocated measurements, we importantly avoid the need to specify *a priori* any kind of analytical scaling relationships.

Another important point with regard to our regional-scale data integration approach is that, through the use of gradual deformation to ensure an adequate fit to the original geophysical measurements, the approach actually represents a kind of stochastic inversion procedure. Stochastic inversion has been widely tested in the field of hydrocarbon exploration with the overall goal of finding realizations of physical properties that are consistent with a variety of geophysical measurements and wellfield production data (Haas & Dubrule 1994; Lamy *et al.* 1999; Buland & Omre 2003a,b; Escobar *et al.* 2006; Grana & Della Rossa 2010). A major limitation of such techniques, however, is that, without suitable prior information, they tend to be extremely computationally expensive. It is for this reason that the use of gradual deformation within the Bayesian sequential simulation methodology is highly advantageous. By starting with realizations that have been already conditioned to existing borehole data and regional-scale geophysical measurements, we greatly reduce the number of subsurface configurations to be explored to only those that are reasonable. Indeed, our results indicate that convergence towards optimal solutions is reached relatively rapidly because the fields generated by the Bayesian sequential simulation algorithm are already well constrained by the existing data. Finally, it should be noted that gradual deformation is certainly not the only stochastic inversion/optimization strategy that could be employed in this particular stage of the data integration procedure. For forward models that can be computed in a reasonably computationally efficient manner, Bayesian sequential simulation could be effectively used to deliver priors for a global stochastic exploration of the model space, for example, by means of a Markov chain Monte Carlo approach (Eidsvik *et al.* 2004; Larsen *et al.* 2006; Gunning & Glinsky 2007; Rimstad & Omre 2010; Ulvmoen & Omre 2010; Hansen *et al.* 2012). With any of such techniques, the speed of generation of posterior realizations can be significantly increased by running in parallel different optimizations and/or Markov chains.

A final important issue that deserves some discussion is the assumption made when relating the borehole electrical conductivity measurements with collocated measurements of the regional-scale ERT-based electrical conductivity in our procedure. Admittedly, we

have assumed in this paper a rather ideal scenario where the borehole measurements are considered as representative of the 'true' small-scale properties, and we have not investigated the potential bias that may result from the measurement support scale of the borehole-logging tools. While this is an important practical issue that will be further investigated in future work, particularly with regard to field applications of our methodology, it is important to note that any such biases with regard to the high-resolution electrical conductivity will be at least in part addressed in the second step of the Bayesian sequential simulation algorithm. That is, because the relationship between the high-resolution electrical and hydraulic conductivities is established empirically based on the estimated joint distribution of collocated measurements of these parameters, such biases will be taken into account when generating the hydraulic conductivity realizations.

5 CONCLUSIONS

The objective of this study was to develop a practical technique for the integration of diverse multiscale geophysical and hydrological data in order to generate regional-scale hydraulic conductivity models that are sufficiently detailed and accurate enough to allow for reliable predictions of flow and transport phenomena. To this end, we have developed a two-step data assimilation procedure based on Bayesian sequential simulation that is targeted to the common case where there exist low-resolution, but spatially exhaustive, surface-based geophysical measurements over a large region along with high-resolution, but spatially sparse, borehole measurements of the governing geophysical parameter and the hydraulic conductivity. Numerical testing on a realistic heterogeneous aquifer model considering surface-based geoelectrical measurements has indicated that, given adequate prior information, the proposed data-integration approach should allow for faithful estimates of the regional-scale hydraulic conductivity structure and reliable predictions of solute transport behaviour. Critical future work will need to explore the application of this methodology to field data, as well as its extension to 3-D scenarios.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Swiss National Science Foundation. We thank René Lefebvre for inspiring discussions and the Fondation Herbet for supporting a visit of Erwan Gloaguen to the University of Lausanne.

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. Am. Inst. Miner. Met.*, **146**, 54–62.
- Armstrong, M., 1998. *Basic Linear Geostatistics*, Springer-Verlag, Berlin Heidelberg, New York.
- Behrens, R.A., MacLeod, M.K. & Tran, T.T., 1996. Incorporating seismic attribute maps in 3D reservoir models, in *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, Denver.
- Binley, A. & Kemna, A., 2005. DC resistivity and induced polarization methods, in *Hydrogeophysics*, pp. 129–156, eds Rubin, Y. & Hubbard, S., Springer, Dordrecht.
- Buland, A. & Omre, H., 2003a. Bayesian linearized AVO inversion, *Geophysics*, **68**, 185–198.
- Buland, A. & Omre, H., 2003b. Bayesian wavelet estimation from seismic and well data, *Geophysics*, **68**, 2000–2009.
- Caers, J., 2005. *Petroleum Geostatistics*, Wiley and Sons, New York.
- 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 Resour. Res.*, **37**, 1603–1613.
- Chen, J. & Rubin, Y., 2003. An effective Bayesian model for lithofacies estimation using geophysical data, *Water Resour. Res.*, **39**, 1118, 11pp.
- Chilès, J.P. & Delfiner, P., 1999. *Geostatistics: Modelling Spatial Uncertainty*, Wiley and Sons, New York.
- 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, *J. appl. geophys.*, **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 Resour. Res.*, **45**, W0940.
- Dafflon, B., Irving, J. & Holliger, K., 2010. Calibration of high-resolution geophysical data with tracer test measurements to improve hydrological predictions, *Adv. Water Resour.*, **33**, 55–68.
- Dahlin, T. & Loke, M.H., 1998. Resolution of 2D Wenner resistivity imaging as assessed by numerical modelling, *J. appl. Geophys.*, **38**, 237–249.
- de Marsily, G., Delay, F., Gonav, J., Renard, P., Teles, V. & Violette, S., 2005. Dealing with spatial heterogeneity, *Hydrogeol. J.*, **13**, 161–183.
- Delleur, J.W., 1999. Elementary groundwater flow and transport processes, in *The Handbook of Groundwater Engineering*, pp. 2.1–2.40, ed. Delleur, J.W., CRC Press, Boca Raton.
- Deutsch, C., 2002. *Geostatistical Reservoir Modelling*, Oxford University Press, New York.
- Doyen, P.M., 2007. *Seismic Reservoir Characterization*, EAGE, Houten.
- Doyen, P.M. & Boer, D.L.D., 1996. *Bayesian sequential Gaussian simulation of lithology with non-linear data*, United States Patent, Western Atlas International, Inc.
- Doyen, P.M., Psaila, D.E., den Boer, L.D. & Jans, D., 1997. Reconciling data at seismic and well log scales in 3-D earth modelling, in *Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, San Antonio, USA.
- Dubreuil-Boisclair, C., 2012. Non-Gaussian gas hydrate simulation at the Mallik site, Mackenzie Delta, Canada, *Mar. Petrol. Geol.*, **35**, 20–27.
- Dubreuil-Boisclair, C., Gloaguen, E., Marcotte, D. & Giroux, B., 2011. Heterogeneous aquifer characterization from Ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear bayesian simulations, *Geophysics*, **76**, J13–J25.
- Eidsvik, J., Avseth, P., Omre, H., Mukerji, T. & Mavko, G., 2004. Stochastic reservoir characterization using prestack seismic data, *Geophysics*, **69**, 978–993.
- Escobar, I., Williamson, P., Cherrett, A., Doyen, P.M., Bornard, R., Moya, R. & Crozat, T., 2006. Fast geostatistical stochastic inversion in a stratigraphic grid, in *SEG Annual Meeting*, Society of Exploration Geophysicists, New Orleans, Louisiana.
- Ezzedine, S., Rubin, Y. & Chen, J., 1999. Hydrological-geophysical Bayesian method for subsurface characterization: theory and application to LLNL suerfund site, *Water Resour. Res.*, **35**, 2671–2683.
- Freeze, A.R. & Cherry, A., 1979. *Groundwater*, Prentice Hall, New York.
- Gastaldi, C. & Roy, D., 1998. Using Bayesian simulations to predict reservoir thickness under tuning conditions, *Leading Edge*, 589–593.
- Gelhar, L.W., 1993. *Stochastic Subsurface Hydrology*, Prentice Hall, Englewood Cliffs.
- Gilbert, F. & Andrieux, B., 2003. High resolution reservoir property modeling using multi-frequency geostatistical simulations, in *SEG Annual Meeting*, Society of Exploration Geophysicists, Dallas, Texas.
- Goff, J.A. & Jennings, J.W., 1999. Improvement of Fourier-based unconditional and conditional simulations for band limited fractal (von Kármán) statistical models, *Math. Geol.*, **31**, 627–649.

- Goldman, M., Gvirtzman, H., Meju, M. & Shtivelman, V., 2005. Hydrogeophysical case studies at the regional scale, in *Hydrogeophysics*, pp. 361–369, eds Rubin, Y. & Hubbard, S., Springer, Dordrecht.
- Gomez-Hernandez, J.J., 2005. Geostatistics, in *Hydrogeophysics*, pp. 59–83, eds Rubin, Y. & Hubbard, S., Springer, Dordrecht.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.
- Gorell, S.B., 1995. Creating 3-D reservoir models using areal geostatistical techniques combined with vertical well data, in *SPE Western Regional Meeting*, Society of Petroleum Engineers, Bakersfield, California.
- Grana, D. & Della Rossa, E., 2010. Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion, *Geophysics*, **75**, O21–O37.
- Gunning, J. & Glinsky, M., 2007. Detection of reservoir quality using Bayesian seismic inversion, *Geophysics*, **72**, R37–R49.
- Haas, A. & Dubrule, O., 1994. Geostatistical inversion—a sequential method of stochastic reservoir modeling constrained by seismic data, *First Break*, **12**, 561–569.
- Hansen, T.M., Cordua, K.S. & Mosegaard, K., 2012. Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling, *Comput. Geosci.*, **16**, 1–19.
- Heinz, J., Kleinedam, S., Teutsch, G. & Aigner, T., 2003. Heterogeneity patterns of quaternary glaciofluvial gravel bodies (SW Germany): application to hydrogeology, *Sediment. Geol.*, **158**, 1–23.
- Hu, L.Y., 2000. Gradual deformation and iterative calibration of gaussian-related stochastic models, *Math. Geol.*, **32**, 87–108.
- Hu, L.Y., 2002. Combination of dependent realizations within the gradual deformation method, *Math. Geol.*, **34**, 953–963.
- Hu, L.Y., Blanc, G. & Noetinger, B., 2001. Gradual deformation and iterative calibration of sequential stochastic simulations, *Math. Geol.*, **33**, 475–489.
- Hubbard, S., Chen, J., Peterson, J.E., Mayer, E.L., Williams, K.H., Swift, D.J., Mailloux, B. & Rubin, Y., 2001. Hydrogeological characterization of South Oyster Bacterial Transport Site using geophysical data, *Water Resour. Res.*, **37**, 2431–2456.
- Hubbard, S.S. & Linde, N., 2011. Hydrogeophysics, in *Treatise on Water Science*, pp. 401–434, ed. Wilderer, P., Academic Press, Oxford, UK.
- Hyndman, D.W., Harris, J.M. & Gorelick, S.M., 1994. Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, **30**, 1965–1977.
- Hyndman, D.W. & Tronicke, J., 2005. Hydrogeophysical case studies at the local scale: the saturated zone, in *Hydrogeophysics*, pp. 391–412, eds Rubin, Y. & Hubbard, S., Springer, New York.
- Irving, J. & Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resour. Res.*, **46**, W11514, 16pp.
- Isaaks, E.H. & Srivastava, R.M., 1989. *An Introduction to Applied Geostatistics*, Oxford University Press, New York.
- Johnson, T.C., Routh, S., Clemo, T., Barrash, W. & Clement, W.P., 2007. Incorporating geostatistical constraints in nonlinear inversion problems, *Water Resour. Res.*, **43**, W10422, 18pp.
- Journel, A.G. & Huijbregts, C.J., 1978. *Mining Geostatistics*, Academic Press, New York.
- Kelkar, M. & Perez, G., 2002. *Applied Geostatistics for Reservoir Characterization*, SPE, Richardson.
- Keller, G.V. & Frischknecht, F.C., 1966. *Electrical Methods in Geophysical Prospecting*, Pergamon Press Inc, Oxford.
- Lamy, P., Swaby, P.A., Rowbotham, P.S. & Dubrule, O., 1999. From seismic to reservoir properties with geostatistical inversion, *SPE Reserv. Eval. Eng.*, **2**, 334–340.
- Larsen, A.L., Ulvmoen, M., Omre, H. & Buland, A., 2006. Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model, *Geophysics*, **71**, R69–R78.
- Lee, S.H., Malallah, A., Datta-Gupta, A. & Higdon, D., 2002. Multiscale data integration using Markov random fields, *SPE Reserv. Eval. Eng.*, **5**, 68–78.
- Lesmes, D.P. & Friedman, S., 2005. Relationships between electrical and hydrogeological properties of rocks and soils, in *Hydrogeophysics*, pp. 129–156, eds Rubin, Y. & Hubbard, S., Springer, Dordrecht.
- Loke, M.H., 2012. Available at: <http://www.geotomosoft.com>.
- Looms, M.C., Binley, A., Jensen, K.H., Nielsen, L. & Hansen, T.M., 2008. Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data, *Vadose Zone J.*, **7**, 238–248.
- Looms, M.C., Hansen, T.M., Cordua, K.S., Nielsen, L., Jensen, K.H. & Binley, A., 2010. Geostatistical inference using crosshole ground-penetrating radar, *Geophysics*, **75**, J29–J41.
- Malallah, A., Perez, H., Datta-Gupta, A. & Alamoudi, W., 2003. Multiscale data integration using Markov Random fields and Markov Chain Monte Carlo: a field application in the Middle-East, *SPE Reservoir Evaluation & Engineering*, **7**, 416–426.
- McKenna, S.A. & Poeter, E.P., 1995. Field example of data fusion in site characterization, *Water Resour. Res.*, **31**, 3229–3240.
- Oldenburg, D.W. & Li, Y., 1999. Estimating depth of investigation in DC resistivity and IP surveys, *Geophysics*, **64**, 403–416.
- 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.
- Ravelec-Dupin, M.L., Noetinger, B. & Hu, L.Y., 2002. The fft moving average (fft-ma) generator: an efficient numerical method for generating and conditioning gaussian simulations, *Math. Geol.*, **32**, 701–723.
- Rimstad, K. & Omre, H., 2010. Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction, *Geophysics*, **75**, R93–R108.
- Roggero, F. & Hu, L.Y., 1998. Gradual deformation of continuous geostatistical models for history matching. Paper SPE 49004 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 27–30 September, doi: 10.2118/49004-MS.
- Rubin, Y., 2003. *Applied Stochastic Hydrology*, Oxford University Press, New York.
- Scholer, M., Irving, J., Looms, M.C., Nielsen, L. & Holliger, K., 2012. Bayesian Markov-Chain-Monte-Carlo inversion of time-lapse crosshole GPR data to characterize the vadose zone at the Arenas site, Denmark, *Vadose Zone J.*, **11**, 19pp.
- 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**, 128–141.
- Siemon, B., Christiansen, A.V. & Auken, E., 2009. A review of helicopter-borne electromagnetic methods for groundwater exploration, *Near Surface Geophys.*, **7**, 629–646.
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.
- Singha, K. & Gorelick, S.M., 2005. Saline tracer visualized with electrical resistivity tomography: field scale moment analysis, *Water Resour. Res.*, **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 Resour. Res.*, **22**, 2069–2089.
- Tronicke, J. & Holliger, K., 2005. Quantitative integration of hydrogeophysical data: conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers, *Geophysics*, **70**, H1–H10.
- Tronicke, J., Holliger, K., Barrash & Knoll, M., 2004. Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation, *Water Resour. Res.*, **40**, W01519, 14pp.
- Ulvmoen, M. & Omre, H., 2010. Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: part 1—methodology, *Geophysics*, **75**, R21–R35.
- Wand, M.P. & Jones, M.C., 1995. *Kernel Smoothing*, 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).