Hydrogeological parameter estimation using geophysical data: a review of selected techniques

Susan S. Hubbard a,*, Yoram Rubin b

a Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., MS 90-1116, Berkeley, CA 94720, USA
b Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA

Received 11 August 1999; received in revised form 14 February 2000; accepted 21 February 2000

Abstract

Subsurface environmental, engineering, and agricultural investigations often require characterization of hydraulic parameters. For example, groundwater flow modeling is often performed through an aquifer whose hydrological properties have been created using stochastic simulation techniques; these techniques use as input both hydraulic parameter point values and spatial correlation structure information. Conventional sampling or borehole techniques for measuring these parameters are costly, time-consuming, and invasive. Geophysical data can compliment direct characterization data by providing multi-dimensional and high resolution subsurface measurements in a minimally invasive manner. Several techniques have been developed in the preceding decade for using joint geophysical–hydrological data to characterize the subsurface; the purpose of this study is to review three methodologies that we have recently developed for use with geophysical–hydrological data to estimate hydrological parameters and their spatial correlation structures. The first two methodologies presented focus on producing high-resolution estimates of hydrological properties using densely sampled geophysical data and limited borehole data. Although we find that high-resolution geophysical data are useful for obtaining these estimates, in practice, geophysical profiles often sample only a small portion of the aquifer under investigation, and thus, the estimates obtained from geophysical data may not be sufficient to completely describe the hydraulic properties of the aquifer volume. The third and last section focuses on using high-resolution tomographic data together with limited borehole data to infer the spatial correlation structure of log-permeability, which can be used within stochastic simulation techniques to generate parameter estimates at unsampled locations. Our synthetic case studies suggest that collection of a few tomographic profiles and interpretation of these profiles together with limited wellbore data can yield hydrological point values and spatial correlation structure

* Corresponding author. Tel.: +1-510-486-5266; fax: +1-510-486-5686.
E-mail addresses: sshubbard@lbl.gov (S.S. Hubbard), rubin@ce.berkeley.edu (Y. Rubin).
information that can be used to aid numerical aquifer model construction, calibration, and flow simulation. As this information is typically only obtainable from extensive hydrological sampling, use of geophysical methods may offer a more efficient and less invasive approach than traditional characterization campaigns. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Hydrogeological parameter estimation; Geophysical data; Subsurface measurements

1. Introduction

Much of the problem and initial cost of subsurface remediation comes from field site characterization. Three-dimensional information about the heterogeneous subsurface is needed in order to identify the key controls on the flow and contaminant transport processes. It is widely recognized that natural heterogeneity and the large spatial variability of permeability predominantly control the flow field and hence, the transport. Moreover, natural heterogeneity exhibits variability over a wide range of scales and is difficult to characterize due to scarcity of data and the costliness of conventional field sampling techniques such as drilling.

With poor site characterization, remediation schemes are unnecessarily expensive, because costly over-design and characterization at a detailed scale may be required to compensate for uncertainty. Consider a flow variable $Z$ such as permeability. In an ideal situation, $Z$ can be inferred at each and every point in the domain at the necessary resolution. This situation is rarely realized, however, due to the scarcity of direct $Z$ measurements typically available using conventional sampling techniques. An alternative is to adopt a stochastic approach, whereby $Z$ is typically characterized by its mean and spatial covariance $C_Z(r)$, where $r$ denotes the lag or the separation distance between hydrogeological measurements. The spatial covariance indicates the distances (or length scale, $L_Z$) and directions over which the flow variable is correlated in space. Other important length scales in the characterization problem may include that associated with the support scale of the hydrogeological measurements, $L_{me}$, and that associated with the scale of the contaminant plume, $L_c$. As spatial phenomena are recognized as being comprised of variations that are characterized by different length scales, spatial correlation analysis is facilitated by working in the wave number domain where each length scale is associated with a wave number vector $k$ corresponding to $r$, and the power associated with each wave number is described by a spectral density function $S(k)$.

A characterization effort which ignores the significance of the length scales may provide information at too low of a resolution, which would not be sufficient for transport modeling, or at too high of resolution, which would be unnecessarily costly. The length scale associated with a contaminant plume, $L_c$, can be viewed as a cut-off length scale: smaller scale variability likely affects contaminant mixing and dispersion, while larger scale variability affects plume trajectory and rate of displacement. The significance of this distinction is in the fact that $L_c$ is of the order of $L_Z$, and most often much larger. Hence, the effects of the small-scale variability can be modeled using effective transport parameters, which can be determined using available high-resolution information from boreholes or outcrop data. This leaves us with the important task of
determining the large-scale variability, but this is a more manageable task. This concept is further explained in Fig. 1. Here, the entire spectral density function $S_{\chi}(k)$ is presented, as well as the cut-off wave number associated with the scale of a hypothetical contaminant plume $k_c$.

This figure illustrates that in this case, high frequency ($k > k_c$) core or outcrop information can be used to determine the effective parameters rather than for actual mapping of $Z$, while the low wave number variability $S_{\chi}(k < k_c)$, obtained from

![Fig. 1. A spectral density function $S_{\chi}(k)$ illustrates the importance of scale in characterization and modeling efforts. Due to the measurement sampling interval and length of sampling window, different types of measurement can reside in different portions of the wave number (k) range. The wave number range over which the measured data exist, compared to other existing wave number ranges (such as that associated with hydrological heterogeneity or with a contaminant plume) is important for determining the usefulness of the data for the characterization effort. In this example, $k_c \sim (1/L_c)$ is the cut-off wave number associated with the scale of a hypothetical contaminant plume. This figure illustrates that high frequency ($k > k_c$) borehole or outcrop information is most useful for determining the effective parameters rather than for actual mapping of $Z$, while lower wave number variability $S_{\chi}(k < k_c)$, such as those obtained from geophysical data, can be useful for direct characterization.](image-url)
geophysical and geological information, is the one needed for direct characterization. Because the resolution offered by geophysical data vary as a function of method and acquisition parameters employed, geophysical methods are exceptionally well suited for hydrological characterization efforts as they offer the potential for providing characterization information at various length scales.

Several recent studies have investigated the use of geophysical data to aid hydrological studies including Daily et al. (1992), Ramirez et al. (1993), Hyndman et al. (1994), Hyndman and Gorelick (1996), Teutsch et al. (1997), Poeter et al. (1997), and Eppstein and Dougherty (1998). These studies suggest that high-resolution geophysical data can be useful for estimating hydrological properties as well as for delineating the geometry of the aquifer at locations within the study volume, which are traversed by the geophysical profiles. A few investigations have also focused on extracting spatial correlation information from geophysical data, including Knight et al. (1996, 1997), Rea and Knight (1987), and McKenna and Poeter (1995). In this paper, we summarize three methodologies that we have developed for use with joint geophysical–hydrological data to aid in the characterization process. Our methods differ from those cited above in that our techniques were developed to either handle problems of non-uniqueness that are sometimes present in geophysical–hydrological relationships or to focus on problems related to integration of data having different measurement scales.

A review of the geophysical methods that are employed in the following case studies is presented in Section 2, and our three methodologies are presented in the subsequent sections. The first technique is presented in Section 3, where a Bayesian method is used to estimate log-permeability values using extensive seismic data and sparse wellbore data. A maximum likelihood (ML) technique is employed in Section 4 to tackle estimation problems that occur when the petrophysical relationships needed to link the geophysical and hydrological parameters are non-unique. In this particular study, log-permeability and saturation are estimated using ground penetrating radar (GPR) measurements. In circumstances where geophysical resolution is too low to provide useful estimates of hydrological values at a point in space, we find that these data can still be useful for inferring spatial correlation information. Section 5 focuses on the third technique, where log-permeability spatial correlation is estimated using tomographic and wellbore data. These methods provide frameworks that enable geophysical data to be used for hydrogeological applications under a variety of circumstances; we believe that systematic approaches such as these and the ones given by some of the authors cited above are vital for the successful integration of geophysical and hydrogeological data.

2. Radar and seismic methods

Geophysical data complement direct characterization data by providing a denser grid of subsurface measurements than is obtainable from core point measurements or from volume-averaged pump test measurements alone. Geophysical data can be collected in a non-invasive manner and can be used to reduce the number of direct measurements needed to fully characterize a site. The volume of aquifer sampled and resolution of the
Geophysical measurements vary with each geophysical technique. This range is primarily governed by the physics and acquisition parameters of the geophysical technique (including source frequency and sample spacing) and the subsurface properties. In general, there is a trade-off between resolution and volume of aquifer sampled. The geophysical techniques used within the following studies include radar and seismic methods, both of which are commercially available and becoming more common in practice for detailed environmental site characterization (Rubin et al., 1998).

GPR methods use electromagnetic energy at frequencies of 10–1200 MHz to probe the subsurface (Davis and Annan, 1989). At these frequencies, the separation of opposite electric charges within a material that has been subjected to an external electric field typically dominates over the electrical conductive properties. These capacitive properties are described by the dielectric constant, which can be estimated using radar data. GPR performance is optimal in moderately to coarse-textured sediments, and can be poor in electrically conductive environments such as those dominated by the presence of expanding clays. At the high frequencies used for radar acquisition, and in geological environments amenable to radar acquisition, the electromagnetic wave velocities \( V \) obtained from radar data can be related to the real part of the dielectric constant \( \kappa \) as:

\[
\kappa = \left( \frac{c}{V} \right)^2
\]

where \( c \) is the plane-wave propagation velocity of electromagnetic waves in free space (Davis and Annan, 1989) \( (3 \times 10^8 \text{ m/s}) \).

High-resolution seismic methods employ high frequency pulses of acoustic energy (~50–10000 Hz) which are produced at a point and propagate out as a series of wavefronts. The passage of the wavefront creates a motion that can be detected by sensitive geophones. From the time of flight of this energy, information about the seismic velocity can be obtained.

In the following case studies, we consider information available from both surface and tomographic geophysical methods. For seismic and radar methods, surface profiles are typically presented as wiggle-trace displays, where signals correspond to the magnitude of acoustic or electromagnetic property variations across subsurface discontinuities. These profiles can be used to detect both lateral and vertical changes in subsurface physical properties. Analysis of surface seismic and GPR profile data acquired using a common-midpoint acquisition mode (Yilmaz, 1987) can yield two-dimensional information about P-wave velocity and dielectric constant, respectively. Tomographic acquisition methods transmit direct energy, electromagnetic in the case of radar and acoustic in the seismic case, from a transmitter in one borehole to a receiver in another borehole over several transmitting/receiving locations. The energy from the sources to the receivers can be envisioned as traveling along raypaths through the interwell area. Travel time information obtained from the recorded waveforms is processed to obtain estimates of acoustic or electromagnetic wave velocity between the tomographic boreholes; this procedure is known as tomographic inversion. The inversion discretization is governed by the data acquisition parameters and geometry (Williamson and Worthington, 1993). A straight raypath algebraic reconstruction technique inversion algorithm given by Peterson et al. (1985) was used in this study to obtain estimates of P-wave and electromagnetic wave velocity from seismic and radar tomographic data.
travel times, respectively. The electromagnetic wave velocities obtained from tomo-
graphic inversion can then be used with Eq. (1) to obtain dielectric constant estimates
between wells. In the following case studies, we consider the estimates of geophysical
attributes obtained from processing of surface geophysical data or from tomographic
inversion as data that are available to us for use in our hydrological estimation
procedures, and henceforth, we refer to these geophysical attribute estimates as our
“geophysical data”.

In the investigations to follow, we employ site-specific as well as published petro-
physical models to relate the attributes extracted from geophysical data, such as seismic
P-wave velocity and dielectric constants, to hydrological parameters such as log-permea-
bility and saturation. Development of petrophysical relationships for near-surface studies
is a current topic of research and there are few geophysical–hydrological relationships
that are considered to be applicable under a variety of geological environments.
Regression techniques assume a linear relationship between some parameterization of
the geophysical and hydrological variables and have been used extensively to aid
petroleum reservoir studies (e.g., Wyllie et al., 1956; Bourbie et al., 1987; Han et al.,
1986; Klimentos and McCann, 1990; Klimentos, 1991; Best et al., 1994). Regression
techniques have also been used for developing site-specific petrophysical relationships
between co-located geophysical and hydrological measurements for near-surface investi-
gations. In the following case studies, we employ petrophysical relationships that we
have developed using hydrogeological–geophysical data collected from near-surface
studies as well as the published petrophysical models of Rubin et al. (1992) and Knoll et
al. (1995). Use of these petrophysical relationships is not part of our inversion
methodologies, but a necessary step for building the synthetic data fields that are
subsequently sampled using wellbore and geophysical characterization tools.

3. Geophysical–hydrological identification of field log-permeabilities through
Bayesian updating

To overcome the scarcity of pressure and permeability data typically associated with
a hydrological characterization effort, a Bayesian technique, which jointly uses seismic
data and more conventional direct hydrological measurements, was developed to esti-
mate log-permeability. In this and the following numerical studies, log-permeability is
modeled as a space random function, which is characterized through its moments
(Dagan, 1989; Gelhar, 1993). Using the flow equation, pressure is also modeled as a
space random function.

This procedure consists of initially performing a hydrological inversion based solely
on the permeability and pressure data. The initial distribution of permeability obtained
from this inversion is referred to as the “prior” probability distribution function (pdf).
We subsequently attempt to improve this distribution using densely sampled seismic
P-wave velocity information and petrophysical relationships. For a water saturated
sand–clay system, the curves that we use in this case study to describe the relation
between permeability, effective overburden pressure and seismic velocity are shown in
Fig. 2a (Rubin et al., 1992). These curves suggest that the relationships between hydrological and geophysical parameters can be non-unique; for a given seismic velocity and pressure, Fig. 2a suggests that there are two possible values of log-permeability if pressure is known accurately and if not, there are many more possible values.

Fig. 2. (a) Relationships between log-permeability (darcies/1000), seismic velocity (m/s), and effective overburden pressure (MPa) for a saturated, sand–clay system given by Rubin et al. (1992), (b) Comparison of prior and posterior permeability pdfs at a single subsurface location (Copty et al., 1993).
3.1. Mathematical statement of the Bayesian approach

The goal is to estimate $Y(x)$ over the entire synthetic aquifer, where $Y(x) = \ln(k(x))$ is log-permeability and $x$ is a vector of coordinates along a vertical plane. Data that we use in this estimation procedure include wellbore permeability measurements, $Y(x)$, as well as sets of co-located geophysical data ($g(x)$), such as seismic velocity, seismic attenuation, electrical resistivity, dielectric constant, or radar attenuation measurements obtained from high-resolution seismic, electrical, and radar data. Log-permeability is treated as a spatial random function; the available measurements are assumed to be realizations of this random function and the distribution at each location within the aquifer is described by its pdf (Anderson, 1997). All attributes are modeled as second-order stationary random fields whose pdfs are assumed to be normally distributed. The log-permeability pdf is expressed as:

$$f_{Y(x)}(y) = \frac{1}{\sqrt{2\pi} \sigma_y} \exp \left[ -\frac{1}{2} \left( \frac{y - \langle Y \rangle}{\sigma_y} \right)^2 \right]$$  \hspace{1cm} (2)

where $y$ is a realization of the log-permeability and $\langle Y \rangle$ and $\sigma_y$ are the mean and standard deviation of $Y$ which completely describe the distribution.

Log-permeability estimates, obtained from geophysical tomographic data together with petrophysical relationships, can be used to update the prior pdf obtained from hydrological measurements in a Bayesian sense. The updated log-permeability pdf is referred to as a ‘first posterior’ pdf, and is denoted by $f'_{Y(x)}(y)$. The posterior pdf of $Y$ at a single location in space, $x_0$, given additional geophysical information at that location, $g(x_0)$, is given by:

$$f'_{Y(x_0)}(y) = f_{Y(x_0)}(y \mid g(x_0)) = f_{Y(x_0)}(y \mid \hat{y}(g(x_0)))$$  \hspace{1cm} (3)

where $\hat{y}(g(x_0))$ is the prediction of permeability based on geophysical data, henceforth abbreviated as $\hat{y}(x_0)$. This implies that the conditional distribution of log-permeability given geophysical measurements is a function of the predicted log-permeability based on those geophysical methods. Using Baye’s Theorem (Ang and Tang, 1975), the posterior pdf $f'_{Y(x_0)}(y)$ can be expressed as:

$$f'_{Y(x_0)}(y) = \frac{\int_{\mathbb{R}} f_{Y(x_0)}(\hat{y}(x_0) \mid Y(x_0) = y) f_{Y(x_0)}(y) \, dy}{\int_{\mathbb{R}} f_{Y(x_0)}(\hat{y}(x_0) \mid Y(x_0) = y) \, dy}$$  \hspace{1cm} (4)

If multiple types of co-located geophysical data are available, the Bayesian methodology presented above can be applied successively to improve the log-permeability estimation. For example, if estimates of log-permeability, obtained from a single tomographic data set ($g(x) = g(x)$), are available at a particular location $x_0$, the ‘first posterior’ obtained using this conditional information can be calculated using Eq. (3) and can in turn be updated using information available from a second co-located tomographic data set ($g_2(x)$) using:

$$f''_{Y(x_0)}(y) = f'_{Y(x_0)}(y \mid \hat{y}(g_2(x_0))).$$  \hspace{1cm} (5)
Eq. (5) yields a log-permeability distribution at one location in space that has been updated using two geophysical data sets and it is referred to as the ‘second posterior’. The second posterior field can in turn be updated using information from another co-located data set, and so on. Numerical simulations show that incorporation of multiple, co-located geophysical data into the log-permeability estimation procedure using the successive Bayesian technique presented in Eq. (5) decreases the error, variance, and entropy associated with the log-permeability estimate (Hubbard, 1998). Including multiple geophysical fields in the log-permeability estimation also reduces the ambiguity often associated with non-unique or weak petrophysical relationships.

3.2. Bayesian case study using seismic velocity information

Prior pdfs, calculated through hydrological inversion, can be updated using the Bayesian methodology described above. In this study (Copty et al., 1993), exhaustive seismic velocity information is considered to be available over the entire study domain. The petrophysical relationship shown in Fig. 2a is also considered to be known to us from previous investigations. Using this seismic velocity information, together with estimates of the overburden pressure and the petrophysical model that governs the relationship between seismic velocity and log-permeability at this site, the prior pdfs are updated using Eq. (4). These updated distributions are considered to be the posterior pdfs. Fig. 2b is a synthetic example that illustrates prior and posterior pdfs for a single subsurface location compared to the true value at that point. The posterior pdf is considerably sharper than the prior pdf. Its bimodality, resulting from the two limbs of the petrophysical curves shown in Fig. 2a, demonstrates the non-uniqueness of this inverse problem. The maximum likelihood (ML) estimate of the posterior pdf is the value corresponding to the larger of the two peaks; the ML formalism is explored in more detail in Section 4. In this case, the ML value is chosen at each subsurface location as the optimal estimate. The results of several numerical case studies (Copty et al., 1993) suggested that the ML estimates were considerably closer to the true log-permeability values than those estimates obtained from hydrological inversion alone, otherwise known as the prior estimates. If other, co-located geophysical data sets are also available for analysis, then the first posterior distribution shown in Fig. 2b can in turn be updated using this additional information and Eq. (5) to yield a second posterior.

Several synthetic case studies were conducted using this Bayesian methodology to investigate the utility of geophysical data whose quality ranged from error-free to severely corrupted (Copty et al., 1993). In all cases, the log-permeability images obtained from the joint geophysical–hydrological inversion showed improvement over the images obtained from hydrological data alone. Moreover, the uncertainty of the log-permeability estimates was also consistently lower than the uncertainty resulting from the hydrological inverse procedure. The improvement offered by incorporation of geophysical data with hydrological data is often most significant at locations between wellbores, where the prior estimates obtained from wellbore data are typically less informative.

This first study indicates that by joining seismic data and hydrological data into a common inverse procedure, improved and high-resolution permeability estimates can be obtained which have lower uncertainty than those obtained from wellbore data alone.
4. GPR-assisted saturation and permeability estimation in bimodal systems

In this section, we consider estimation of saturation and permeability, both of which are important to vadose zone studies, using GPR methods. We find that information potentially available from radar data collected in a common midpoint (Annan and Cosway, 1992) or tomographic acquisition geometries can be used to estimate these properties in bimodal systems. Numerical experiments were performed to investigate the general utility of the GPR-assisted estimation technique under a range of hydrogeological conditions (Hubbard et al., 1997a,b). Three bimodal systems were investigated, where each system was composed of a sand facies together with another facies having a larger clay volume fraction than the sand facies called the clay facies. Each facies was defined using characteristic values of clay content, porosity, and permeability. Using dielectric information obtained from radar data and a petrophysical model assumed to govern the relationship between dielectric constant and the hydrological parameters, degree of saturation and intrinsic permeability values at each location within the three geological systems were identified. For bimodal systems, a dielectric constant measurement corresponds to two possible values of saturation and intrinsic permeability at each location; single values of saturation and intrinsic permeability were estimated from these values using the principle of ML. Results from case studies demonstrate that a combination of GPR data with conventional borehole data significantly improves the estimates of saturation and has the potential to improve the estimates of permeability over those obtained from wellbore data alone.

Several numerical case studies were performed within synthetic bimodal geological systems using the following approach (Hubbard et al., 1997a,b). The geological domains were constructed using indicator simulation routines (Deutsch and Journel, 1992). Model saturation fields were obtained by numerically simulating two-dimensional water flow through the synthetic domains for a limited time period, at which time radar data acquisition and processing were simulated. The resulting synthetic two-dimensional dielectric constant fields represent geophysical field data potentially available for estimation of saturation and permeability. Fig. 3a illustrates an example of a dielectric constant field obtained from collecting synthetic radar data through an unsaturated bimodal system composed of sand and clay facies. Expected saturation values, shown in Fig. 3b, were obtained by ordinary kriging of saturation measurements obtained from four wellbores whose locations are annotated above the figure. The petrophysical curves that we assume govern the relationships between the facies (and corresponding permeability), saturation and dielectric constant for the sand–clay system are shown in Fig. 3c (reduced from Knoll et al., 1995). As was the case in Fig. 2a, these curves illustrate that non-uniqueness can exist when trying to map a geophysical attribute (dielectric constant obtained from radar data) into a hydrological estimate (saturation or permeability). For each dielectric constant value obtained from the radar data in this bimodal system, two possibilities of saturation ($S_1$ and $S_2$) are possible depending on whether the facies at that subsurface location is a clay (with associated permeability $k_1$), or a sand (with associated permeability $k_2$).

We now desire to estimate the most likely value of saturation at each subsurface location given the options presented from the radar data and petrophysical curves. To do
Fig. 3. (a) Dielectric field cross-section potentially available from radar data; (b) expected value of saturation obtained from kriging borehole saturation measurements; (c) petrophysical curves illustrating the relations between dielectric constant and fractional water saturation for a sand–clay system (reduced from Knoll et al., 1995); and (d) saturation estimates obtained using information displayed in (a)–(c) in an ML formalism. $S_w$ is water saturation.
this, we define saturation as a random function described by a binomial pdf. If \( S_1 \) is the saturation value associated with the clay facies and \( S_2 \) is the saturation value associated with the sand facies, the pdf for saturation \( (S) \) can be expressed as:

\[
f(S) = P\delta(S - S_1) + (1 - P)\delta(S - S_2),
\]

where \( \delta \) is the Dirac delta and \( P \) is the probability such that

\[
S = \begin{cases} 
S_1 & \text{prob} = P \\
S_2 & \text{prob} = 1 - P
\end{cases}
\]

From Eqs. (6) and (7), we obtain the relation between \( P \) and \( \langle S \rangle \), where \( \langle \rangle \) denotes the expected value operator:

\[
P = \frac{\langle S \rangle - S_2}{S_1 - S_2}.
\]

In Eq. (8), \( \langle S \rangle \) can be estimated using borehole measurements (Fig. 3b), and \( S_1 \) and \( S_2 \) can be obtained using the dielectric constants from radar data (Fig. 3a) together with the petrophysical relationship (Fig. 3c). Using these values, \( P \) can be calculated for each subsurface location and used in a ML formalism to choose which saturation option is optimal for each subsurface location. Fig. 3d displays the saturation estimates obtained using the information shown in Fig. 3a–c with this ML technique; when compared to the model saturation field these estimates were correct 97.8% of the time.

Permeability can also be defined as a random function and described by a binomial pdf. The methodology presented above can be used to choose the optimal permeability value at each subsurface location given the options presented from the radar data and petrophysical relationships. Using this ML technique, the corresponding permeability estimates obtained (for the same system and using the dielectric data shown in Fig. 3a) yielded permeability estimates that were correct 93% of the time when compared to the model permeability field.

In the preceding example, it was assumed that the dielectric constant information available from radar data was known completely and exactly at all locations; these data were in turn used to estimate saturation and permeability. In reality, there are errors associated with assumptions about the bimodality of the geologic system and our representation of the facies using characteristic values of clay content, porosity, and permeability. There are also errors associated with GPR data collection and processing, use of a published or developed petrophysical model, and the transfer of radar information into dielectric constant estimates. To account for these errors, we repeated the numerical study described as well as other experiments in different geologic domains using radar data corrupted with different levels of spatially correlated and uncorrelated noise. The results of the case study presented here and these additional case studies (Hubbard et al., 1997a,b) suggested that by incorporating GPR information into the estimation process the following are achieved.

(a) Saturation estimates are improved over those estimates obtained from borehole information even when there is a significant level of noise in the data.
Saturation estimates are optimal when the system is most homogeneous. In this study, we presented only results from investigation of a sand–clay system. However, as the two facies that comprise the geological system become more similar in hydrogeological properties, such as a sand and a sandy loam, the petrophysical curves that describe the system, such as those shown for the sand–clay case in Fig. 3c, also become more similar. When this occurs, the dielectric constant will yield a reasonable saturation estimate regardless of which facies is present at a particular location.

Permeability estimation is improved using GPR over those estimates obtained using wellbore data alone when the system is heterogeneous. As the system becomes homogeneous, the method loses the capability to distinguish between the components of the system. This is because when the hydrological properties of the facies in a system are dramatically different from each other (such as the sand–clay system investigated here), the petrophysical curves that describe the facies (such as those shown in Fig. 3c) are distinct from one another and there is minimal ambiguity in facies or permeability estimation given a single dielectric constant. As the facies that describe the system become more hydrologically similar, this ambiguity increases and facies or permeability estimation becomes more error-prone. As we are generally interested in permeability estimation in heterogeneous environments, under these circumstances the GPR-assisted permeability estimation procedure is useful.

Permeability estimates are more sensitive to the presence of noise than saturation estimates. As data error increases, the GPR-assisted procedure cannot distinguish between the facies (and corresponding permeability) that describe the system, while the saturation values associated with the different facies are still meaningful.

This study suggests that GPR can be a viable tool for obtaining multi-dimensional cross-sections of both saturation and intrinsic permeability. The method should be especially advantageous in areas favorable for GPR data acquisition where detailed resolution is required, but drilling of numerous boreholes is prohibited due to the time and cost involved as well as specific site limitations. Although collection and analysis of GPR data in the format necessary for dielectric constant estimation are relatively new and laborious processes, and are still topics of research, this study suggests that GPR-assisted vadose zone hydraulic parameter estimation is a method worthy of further investigation.

5. Spatial correlation structure inference using geophysical and hydrological data

Flow and transport modeling requires as input a regular grid of hydraulic conductivity values. Because exhaustive sampling of the permeability field is not realistically possible, stochastic simulation methods are often used to produce these regular grids given spatial correlation information. Wellbore data are often useful for obtaining information about the correlation structure in the vertical direction, but the measurements are rarely dense enough in the horizontal direction to provide sufficient information about the horizontal spatial correlation structure. Further exacerbating the quest for correlation structure information, experimental and numerical studies have suggested
that the integral scales of hydraulic conductivity that are estimated from measured data are a function of the scale of the observation relative to the scale of the heterogeneity (Gelhar, 1986; Russo and Jury, 1987). Different measuring devices sample the subsurface with different support scales (Baveye and Sposito, 1984; Cushman, 1986), which are a function of both the resolution of the measuring device, or the measurement scale, and the spatial separation of the measurements, or the network scale (Beckie, 1996).

In this section, we investigate the use of tomographic data as a supplement to hydrogeological measurements in the spatial correlation structure inference procedure. As has been illustrated by the last two studies presented as well as by other researchers, tomographic data can be used together with limited wellbore data to obtain detailed estimates of permeability. Although these data can be used to build a numerical model for flow simulation, for a typical detailed characterization project involving tomographic data, the geophysical profiles only traverse a small portion of the aquifer under consideration. Permeability values for the remaining portions of the study volume can be obtained using stochastic simulation routines with the spatial correlation structure estimated from the tomographic data, hence, the significance of using high-resolution geophysical data for spatial correlation structure estimation. Furthermore, we have found that when the geophysical data resolution is too low to offer point estimates of log-permeability at the scale necessary for flow and transport modeling, that these data can still be useful for providing information about spatial variability that occurs at larger spatial scales of organization.

Our inference procedure is performed in the spectral domain, where analysis of data having different support scales and spatial sampling ranges are facilitated. As data with different support scale and sampling ranges occupy different wave number ranges, the ability to estimate the spatial statistical structure from the spectral density curves varies as a function of the measurement data support scale relative to the scale of the heterogeneity. For example, core data collected from a short core section at a fine sampling interval could be used to determine the high wave number variations of the measured property, while geophysical data, typically sampled at a coarser interval but over longer sampling windows could be used to determine variations that occur at the lower end of the wave number spectrum. The concept of analyzing different data types to investigate variabilities that occur at different spatial scales was illustrated in Fig. 1.

We choose to model the hydraulic properties, such as log-conductivity, as stochastic processes that can be characterized by their spatial covariance structures (Rubin et al., 1998). An example of a two-dimensional log-conductivity exponential covariance structure is given by:

$$C(r) = \sigma_r^2 \exp \left[ - \left( \frac{r_1^2}{I_1} + \frac{r_2^2}{I_2} \right)^{1/2} \right],$$

where $\sigma_r^2$ is the variance of the log-conductivity, $r_1$ and $r_2$ are the log-conductivity measurement separation distances, $I_1$ and $I_2$ are the corresponding integral scales, and the subscripts 1 and 2 refer to the directions perpendicular and parallel to bedding, respectively. The variance indicates dispersion of the properties around the mean value of the distribution, and the integral scale is a measure of the separation distance at which...
the measurements become weakly correlated. Anisotropy, $a$, of the hydrogeological properties is a function of the geometric organization that commonly exists in geologic strata. In our two-dimensional studies, the anisotropy ratio is defined as $I_2/I_1$. Our spectral domain procedure involves the estimation of the variance, integral scales, and anisotropy ratio of the hydraulic properties using hydrological and tomographic data along two-dimensional traverses coincident with the tomographic profiles. Analysis of tomographic profiles along different directions, such as parallel and perpendicular to geologic strike, permits estimation of the three-dimensional spatial correlation structure of the aquifer.

5.1. Description of available data

With our spectral technique, we investigate the spatial correlation information available from both wellbore log-permeability measurements and high-resolution two-dimensional tomographic data. In practice, even if wellbore measurements are abundant, they are typically most prevalent along the vertical direction and are rarely sufficient to yield complete information about horizontal spatial variability. To either provide more information about horizontal variability or to investigate variability at different spatial scales, we supplement the wellbore data with tomographic information in the spatial correlation estimation procedure. The direct geophysical information, referred to as $g$, may include geophysical attributes such as P-wave velocity obtained from seismic tomographic data or dielectric constants obtained from radar tomographic data. Geophysical measurements are often co-located with wellbore measurements; this overlap permits development of petrophysical relationships that can be used to transform the direct geophysical measurements into estimates of hydrological properties.

In general, as well as in the case study to follow, geophysical and wellbore data have different support scales. The support scale of the geophysical data is defined for this case study as the “geophysical-scale”, and the support scale of the wellbore data is referred to as the “core-scale”. To develop petrophysical relationships, it is often necessary to average several of the core measurements so that the core and co-located tomographic data have the same sampling interval. The scale matching between field measurements is referred to here as upscaling.

5.2. Spectral representation of data

Spectral domain analysis facilitates integration between data which have different support scales and spatial sampling windows by representing the spatial phenomena as being comprised of variations that are characterized by different length scales. The integral scale associated with a one-dimensional, uniformly spaced measurement series can be investigated using the amplitudes obtained from Fourier transformation of the data, $a(k_m)$, where $k_m$ denotes the wave number in the Cartesian direction $m = 1, 2$ (Bracewell, 1965; Gardner, 1988). For situations where evenly spaced data are not available, interpolation can be performed prior to transformation (Press et al., 1992). In
this study, we work with uniformly spaced samples, and express the energy spectrum obtained from these amplitudes as a spectral density function, $S(k_m)$:

$$S_{\eta}(k_m) = \langle a_{\eta}(k_m) a_{\eta}(k_m)^* \rangle$$

(10)

where $\eta = Y$ or $G$ refers to the type of data employed, $m$ denotes the spatial direction, the asterisk denotes the complex conjugate, and $\langle \rangle$ denotes the expected value. Neither summation over repeated indices is implied in Eq. (10) nor in subsequent expressions. Spatial correlation functions such as that given by Eq. (9) can be expressed as spectral density functions using Fourier transformations as discussed by Dagan (1989) and Gelhar (1993).

If more than one measurement series is available in the same direction, averaging of the individual spectra can be performed to obtain an average spectral density curve. For example, if several core sections of the same length and sampling interval are available from several wellbores, the average vertical log-permeability spectral density function can be obtained by averaging the wellbore data spectral density functions. Similarly, space-averaged spectral density functions in the vertical and horizontal directions can be calculated from two-dimensional data, such as from a geophysical tomogram, by calculating the average one-dimensional spectral density function along the vertical direction and then repeating the procedure along the horizontal direction. The space-averaged spectral density function can be calculated over $d$ individual data measurement series (such as core sections and tomographic rows or columns) using:

$$\overline{S}_{\eta}(k_m) = \frac{1}{d} \sum_{j=1}^{d} S_{\eta}^{(j)}(k_m)$$

(11)

where the superscript $j$ denotes row, column, or core section over which the averaging is taking place, and the horizontal bar denotes space averaging.

Spectral density functions are calculated from measured data as a function of wave number, or spatial frequency. For all data types, the low wave number cut-off value of the spectral density function is proportional to the inverse of the sampled domain size, $1/L$ (in units of $1/distance$). For data sampled with a spatial interval of $l$ (in units of samples/distance), the spatial Nyquist frequency ($l_N = 1/2$) dictates the high wave number cut-off, or the highest frequency that can be detected using that sampling interval. Thus, the frequencies associated with the hydrological heterogeneity must be lower than the measured data Nyquist frequency but higher than the frequencies associated by the low wave number cut-off in order to be detected by that particular set of measurements. In terms of angular wave number, the low cut-off ($k_{min}$) is expressed as $2\pi/L$ and the high angular Nyquist wave number cut-off ($k_{N}$) is expressed as $2\pi l_N$ (Ababou and Gelhar, 1990). Tomographic data generally have fewer samples per length than direct wellbore measurements ($l_{geophysical} < l_{wellbore}$); the spatial interval of the tomographic data is a function of the discretization chosen for the geophysical inversion which is a function of acquisition geometry and excitation frequency (Williamson and Worthington, 1993). In addition to having fewer samples than the wellbore measurements, the tomographic data typically sample a greater extent of the spatial domain than...
the wellbore data \((L_{\text{geophysical}} > L_{\text{wellbore}})\). As a result, the spectral information obtained from geophysical data typically resides in the low wave number portion of the spectrum relative to the position of the spectra obtained from wellbore data as was illustrated in Fig. 1.

5.3. Estimation procedure

We present two approaches for estimating spatial correlation information from the spectral density curves obtained from the measured data. In the first approach, no assumption about the type of spatial correlation model is necessary to obtain estimates of the variance and integral scale. In the second approach, a spatial correlation model is assumed, presumably known from prior investigations at the study site or borrowed from geologically similar formations (Rubin et al., 1998). The choice of which of the following approaches to use in practice depends on the information existing about the spatial correlation model and the confidence that the interpreter has in this model. We find that if the correct model is assumed, then the model-based spatial correlation parameters are better than those estimated by making no model assumptions. This is of course not surprising since a-priori knowledge of the spatial correlation model brings a significant amount of information to the estimation process.

5.3.1. Approach #1

In the first approach, we do not assume that the structural correlation model type is known. In this case, the log-permeability correlation function in the direction \(m\) can be found by integrating the spectral density calculated from measured data using Eq. (10) or Eq. (11):

\[
C_y(r_m) = \int_{-\infty}^{+\infty} S_y(k_m) e^{ik_m r_m} dk_m
\]

(Gelhar, 1993). Note that when \(r_m = 0\):

\[
C_y(0) = \sigma_y^2 = \int_{-\infty}^{+\infty} S_y(k_m) dk_m.
\]

Using a representation of the integral scale, \(I_m\):

\[
I_m = \frac{1}{\sigma_y^2} \int_{-\infty}^{+\infty} C_y(r_m) dr_m.
\]

(Dagan, 1989) together with the expressions given in Eqs. (12) and (13), we can estimate the integral scale from the spectral density function. This approach requires no assumption about the spatial correlation model that governs the log-permeability distribution in order to obtain estimates of the variance and integral scale. However, in order to use these correlation parameters for measurement interpolation, a correlation model must be invoked.
5.3.2. **Approach #2**

An alternate approach is to assume that the spatial correlation function can be represented using a particular structural model obtained from previous investigations within the aquifer under investigation or within another geologically similar aquifer. For example, if it is believed that the covariance can be adequately represented with an exponential covariance structure, then a spectral representation of this model can be found and the spectra calculated from the data can be fit with that model. For example, Fourier transformation of a one-dimensional, exponential covariance structure yields the log-permeability exponential spectral density function:

\[
S_f(k_m) = \frac{\sigma^2_I Y_m}{\pi (1 + k_m^2 I_m^2)}
\]  

(15)

where \( \sigma^2_I \) is the log-permeability variance and \( I_m \) is the integral scale in the direction \( m = 1, 2 \) (Gelhar, 1993). To implement this approach, the experimental spectral density function, obtained from the measured data using Eq. (10) or Eq. (11), can be equated to the right-hand side of Eq. (15) and fit to estimate \( I_m \) and \( \sigma^2_I \). This fit can be performed using a number of algorithms; for the case study to follow, we used a Gauss–Newton nonlinear least squares procedure. This statistical fitting approach has an advantage over the integration approach presented in Eqs. (12)–(14) in that some statistical procedures, such as the nonlinear least squares technique used herein, provide both estimates of the parameter and estimates of parametric uncertainty (Ratkowsky, 1983). To improve the stability of the fit, the variance in Eq. (15) can be estimated using Eq. (13) prior to fitting for the integral scale.

In summary, our correlation structure parameter estimation procedure consists of the following steps:

(i) obtain log-permeability measurements or estimates from hydrological or tomographic data;
(ii) transfer this information into Fourier space and calculate the one-dimensional experimental spectral density functions. If two-dimensional data are available, the average spectral density function in each direction is calculated;
(iii) estimate the covariance and variance from the resulting spectral density data using discrete forms of Eqs. (12) and (13) and calculate the integral scale by integration using Eq. (14) (Approach #1); or
(iv) fit the experimental spectral density curve to solve for the integral scale and variance using Eq. (15). Alternatively, the variance can be calculated using Eq. (13) and substituted into Eq. (15) prior to fitting for the integral scale (Approach #2).

5.4. **Case study using synthetic seismic tomographic data**

We present a numerical case study that was designed to investigate the usefulness of high-resolution cross-hole seismic data as a supplement to wellbore data in the statistical
structure inference procedure. In this case study, neither the wellbore data nor the geophysical data alone are capable of capturing the spatial variability of the system. For this study, we simulated a model permeability field with an exponential log-permeability covariance function using a sequential simulation routine (Deutsch and Journel, 1992). These data were simulated at the "core-scale", with square pixels having 10-cm sides. The structural covariance of this model field is described using:

$$C_Y(r_1, r_2) = 0.63 \exp \left[ -\left( \frac{r_1^2}{13.28} + \frac{r_2^2}{49.1} \right)^{1/2} \right],$$

(16)

where $r_1$ and $r_2$ denote the separation distances in the vertical and horizontal directions, respectively. The mean permeability value of this synthetic aquifer is $10^{-9}$ cm$^2$; the range of permeability values was designed to represent a clayey-sand system. In our water-saturated system, we assumed that the sand content varied between 60% and 90%, and that the effective pressure was 0.1 MPa. Regression analysis was performed over the portion of the petrophysical curves given by Rubin et al. (1992; Fig. 2a) that corresponds to these parameters to yield the following relationship between the geophysical seismic velocity values ($G$) and log-permeability ($Y$):

$$Y(cm^2) = 3.55 - 0.0078G(m/s).$$

(17)

The sand content range was chosen so that the petrophysical model designed to govern the system is simple and unimodal. The model permeability field that was simulated using Eq. (16) was transformed into a core-scale velocity field using the relationship given by Eq. (17). Forward geophysical modeling was performed through the resulting model velocity field using seismic acquisition parameters common for detailed characterization projects. Ten percent random noise was added to the raw geophysical data to simulate error that could occur during data acquisition and reduction. These "corrupted data" were inverted using a geophysical-scale discretization pixel size of 20 cm square; the inversion process resulted in additional error that was spatially correlated. The resulting two-dimensional seismic velocity field is shown in Fig. 4; this field represents data available to us for spatial correlation inference.

To show the benefits of this technique when core data are sparse (a typical field condition), in this case study, only a limited number of permeability values were obtained from two half-meter core sections extracted from a single borehole. These two core sections, in addition to the tomographic seismic data P-wave velocity information, were considered to be the data available for spatial structural inference. Additionally, it is assumed that the petrophysical relationship between seismic velocity and permeability given by Eq. (17) is known from previous laboratory studies or field calibrations. We perform this analysis in three parts: first trying to recover the spatial correlation parameters of the model log-permeability field given in Eq. (16) using only limited information from the core data, then using only information from seismic data, and finally, using core and seismic information together.

The seismic and core data were transformed into the wave number domain using fast Fourier transforms. Spectral density curves were obtained for the core data in the
vertical direction using Eq. (11) and are shown in Fig. 5a. The tomographic velocity field shown in Fig. 4 was transformed into the wave number domain, and the average velocity spectral density functions, $\hat{S}_i(k)$, $i = 1, 2$, were calculated using Eq. (11). The velocity spectra were used together with the petrophysical relationship, which we assume governs the system, to estimate permeability spectra as follows. Using the petrophysical linear relationship between log-permeability ($Y$) and seismic velocity ($G$) given in Eq. (17), the covariance of the estimated log-permeability ($C_Y$) and seismic velocity ($C_G$) can be found using the corresponding relation $C_Y = (0.0078)^2 C_G$. Since $C_Y$ and $C_G$ are Fourier transform pairs with $S_Y$ and $S_G$, respectively, then by using the linearity property of Fourier transforms, we can estimate the spectral density of the hydrological measurements given the spectral density of the geophysical data using $\hat{S}_Y = (0.0078)^2 \hat{S}_G$. Using this Fourier property, the estimated average log-permeability spectral density functions were calculated from the seismic velocity spectral density functions and are shown in Fig. 5b for the vertical direction.

Integral scales were calculated from the core and seismic spectral curves, first by using the integration approach given in Eqs. (12)–(14), and then by assuming an exponential structural covariance model and applying the second approach using Eq. (15) together with the variance calculated using Eq. (13). These integral scales were found for the vertical direction using the core data and for both the vertical and horizontal directions using the seismic data. Parametric uncertainty is also reported for the analysis performed using the second approach. The results of these estimations are shown in Table 1, under the headings ‘Wellbore data’ and ‘Seismic data’.
Comparison of the results shown in Table 1 with the true model parameters given by Eq. (16) shows that the integral scale estimation from core and seismic data differ from that of the model log-permeability field, and that this difference is greater using the first approach. In both cases, the core data provide an underestimation of the integral scale and the seismic data provide an overestimation of the integral scale. Because neither the spectra obtained from the core data nor the spectra obtained from the seismic data traverse the entire wave number range, the actual model log-permeability integral scale...
Table 1
Spatial correlation parameter estimates obtained from data sampling of the model field. The spatial correlation function of the model field is exponential with $\sigma^2 = 0.63$, $I_1 = 13.28$ cm, $I_2 = 49.1$ cm and an anisotropy of 3.697.

<table>
<thead>
<tr>
<th>Data used to construct spectral density function</th>
<th>$1/I$ (cm/sample)</th>
<th>$\tilde{\sigma}^2$</th>
<th>Approach #1 $I_1$ (cm)</th>
<th>$I_2$ (cm)</th>
<th>Anisotropy</th>
<th>Approach #2 $I_1$ (cm) and $\tilde{\delta}_l^2$</th>
<th>$I_2$ (cm) and $\tilde{\delta}_l^2$</th>
<th>Anisotropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wellbore data</td>
<td>10</td>
<td>0.12</td>
<td>~0</td>
<td>N/A</td>
<td>N/A</td>
<td>5.1, 0.172</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Seismic data</td>
<td>20</td>
<td>0.40</td>
<td>21.01</td>
<td>166.64</td>
<td>7.93</td>
<td>14.8, 0.046</td>
<td>54.5, 1.003</td>
<td>3.68</td>
</tr>
<tr>
<td>Seismic and wellbore data</td>
<td>20, 10</td>
<td>0.52</td>
<td>16.12</td>
<td>127.86</td>
<td>7.93*</td>
<td>13.03, 0.008</td>
<td>47.9, N/A</td>
<td>3.68*</td>
</tr>
</tbody>
</table>

*Anisotropy estimated using seismic data only.
cannot be accurately estimated using either data set alone. To improve the estimation, the spectral curves obtained from the seismic and core data were combined as shown in Fig. 5c to yield the spectral density function that spans the entire wave number range. This combined spectral data set was again analyzed using both estimation approaches, and the results are reported in Table 1 under the heading ‘Seismic and wellbore data’. This analysis illustrates that the correlation structure parameter estimates obtained from the combined data set were improved over those obtained using core data or seismic data alone. Additionally, as shown in Table 1, the estimated variance associated with the integral scale estimate decreases as the two data sets are combined prior to analysis.

As continuous core coverage in the horizontal direction is not available for this case study, it is impossible to combine wellbore and seismic information together to obtain a composite horizontal spectral density function as was performed for the vertical direction. By assuming that the anisotropy ratio inferred from the seismic data (Table 1) is valid, we can use this value with the vertical integral scale estimate \( \hat{I}_1 \) obtained from the composite wellbore-seismic spectra (Fig. 5c) to estimate the composite horizontal integral scale:

\[
\hat{I}_2 = \alpha \hat{I}_1. 
\]

(18)

This assumption implies that there is only one scale of spatial variability and that upscaling is linear, or that the properties estimated by the seismic tomographic data are linear averages of the wellbore point property measurements. The results of this estimation using both approaches and Eq. (18) are given in Table 1 under the heading ‘Seismic and wellbore data’. Alternatively, if horizontal core data were available, spectral information from the core and tomographic data could be combined prior to analysis as was performed in the vertical direction.

The results reported in Table 1 reveal that, for all cases, the second approach yielded the best estimates and also gave an estimation of parametric uncertainty. In this case study, where we (correctly) assumed that the structural correlation could be described using an exponential function, this additional information served to improve the estimates over the first approach, where no model was assumed to be known. Additionally, as the spectral density function is obtained from measured data, it exists only over a finite wave number range dictated by the size of the sampled domain and the sampling interval; this potentially truncated spectra may negatively influence the results obtained from integration over the spectral curve using Eqs. (12)–(14). This simple case study illustrates the benefits of including all available information, such as core data, geophysical data and correlation structural model type, into a single estimate of the correlation structure parameters.

5.5. Spatial correlation estimation results

The results of the numerical case study presented here, as well as additional numerical case studies (Hubbard, 1998; Hubbard et al., 1999a,b) suggest the following.

(a) If good information is available about the type of correlation structure model, incorporation of that information into the estimation procedure improves the correlation.
structural parameter estimates over those estimates obtained without assuming any structural model.

(b) The integration approach given by Eqs. (12)–(14) can be used to obtain reasonable estimates of the structural correlation parameters without assuming a spatial correlation model.

(c) Geophysical data alone can be used successfully to infer the spatial correlation structure of an aquifer when the geophysical resolution is high compared to the scale of the heterogeneity and the petrophysical relationships between the geophysical and hydrogeological parameters are unique.

(d) If the geophysical data resolution is low compared to the scale of the heterogeneity, measurements with a smaller support scale, such as core measurements, can be incorporated with the geophysical data to more accurately infer the spatial correlation structure.

(e) Parametric uncertainty can be reduced by combining data that exist in different portions of the wave number domain prior to analysis.

(f) Geophysical data are capable of yielding critical information about the correlation structure in the horizontal direction. This information is extremely difficult to obtain given sparse conventional borehole data.

(g) In hierarchical geological systems, which are systems where the hydrogeological variabilities occur at different spatial scales, geophysical data can be used to estimate the spatial correlation structure of the level of heterogeneity that shares a similar support scale of the geophysical data measurement. An example of a hierarchical system is an aquifer that is composed of several geologic facies (i.e., sand, silt, and clay facies) where there is a large-scale spatial variability associated with the geometry of the facies packages and a smaller scale heterogeneity associated with the spatial variability within each facies unit. If the resolution of the geophysical data is low compared to the within-facies variability, but is on the same order as the variability of the facies geometry, it may be possible to extract information about the larger scale of heterogeneity using the geophysical data.

6. Application of methodologies to field data set

In the preceding sections, three methodologies for using joint geophysical–hydrological data sets for estimating hydrological parameters and their spatial correlation structure were presented. In each case, the methodology was illustrated using a synthetic data set. Here, we apply two of the methodologies to a real field data set to obtain estimates of log-conductivity and the spatial correlation of log-conductivity. The data set under consideration was collected at the US Department of Energy (DOE) bacterial transport site near Oyster, VA. The Oyster site is located on the southern Delmarva Peninsula situated on the eastern coast of the US between the Chesapeake Bay and the Atlantic Ocean. The sediments at the Oyster site consist of unconsolidated to weakly cemented, well-sorted, medium- to fine-grained Late Pleistocene sands and pebbly sands that were deposited in tide and upper shoreface settings (Mixon, 1985; Parsons et al., 1997).
Bioremediation, the use of microorganisms that can degrade organic wastes or immobilize inorganic contaminants in the subsurface, is among the many technologies that are currently being evaluated for the restoration of contaminated subsoils and groundwater. Although bioremediation is considered a viable technique, microbial transport in the presence of complex subsurface heterogeneities is not well understood. The investigations at the Oyster site are geared toward understanding the relative importance of hydrogeological and chemical heterogeneities in controlling bacterial transport at the field scale (DeFlaun et al., 1997). The following data that we analyze to illustrate our methodologies are a portion of the data collected at the Oyster site to characterize the hydrogeological heterogeneities and to aid in the building of a numerical flow model used to predict bacterial transport at the Oyster site (Scheibe et al., 1999; Hubbard et al., 1999a).

The data available for this example include hydraulic conductivity measurements from one-dimensional flowmeter logs as well as two-dimensional radar tomographic profiles. The borehole electromagnetic flowmeter (Molz and Young, 1993; Molz et al., 1994) data that were used in this study were collected in a saturated section between the depths of 3.75 and 9.0 m below ground surface and with a 0.15 m sampling interval. The tomographic data considered here were collected using a 1/8 m transmitter/receiver spacing in the wellbores. The first arrival travel times were picked for each transmitter/receiver location in each well pair, and inversion was performed using an algebraic reconstruction technique (as described in Section 2) to obtain radar velocity values. These radar velocity values were converted to dielectric constant estimates following Eq. (1). We consider here data available from a tomographic radar profile that was collected along the geologic strike direction between two wellbores located 8.9 m apart (wellbores T3 and T1), as well as a tomographic radar profile collected in the geological dip direction between two wellbores located 9.49 m apart (wellbores B2 and M3). Both radar profiles were collected from 0.5 to 9 m below ground surface; however, in this example, we examine only the radar tomographic responses in the saturated zone, from 3–9 m below ground surface.

The flowmeter measurements and tomographic radar information were initially used to provide point values of hydraulic conductivity following the procedure that was presented in Section 3.1 and that was applied to synthetic data in Section 3.2. The flowmeter data were initially used within an ordinary kriging algorithm to obtain two-dimensional cross-sections of log-conductivity ‘prior’ pdfs along the tomography profiles. Likelihood relationships, needed to link the radar attribute of dielectric constant to log-conductivity, were developed using upscaled flowmeter measurements and the dielectric constant information obtained from the radar tomograms adjacent to the wellbore location. The ‘prior’ pdfs were then updated at the flowmeter discretization using the radar tomography data together with the likelihood relationships following Eq. (4); details of the Bayesian analysis of the Oyster data are given by Chen et al. (1999, 2000). As discussed in Section 3, this procedure yields a complete log-conductivity distribution for each location along the two-dimensional tomograms that is conditional to the wellbore data. Fig. 6 displays the means of the estimated log-conductivity pdfs along the strike direction (between wells B2–M3) and along the dip direction (between wells T3–T1). Also superimposed on this figure are hydraulic conductivity measure-
Fig. 6. Geophysical and borehole data available for analysis from the Oyster site in Virginia. The means of the estimated log-conductivity probability density functions created using limited flowmeter data and two radar tomography profiles within a Bayesian estimation framework are shown. Profile T1–T3 is aligned along geologic strike, while profile B2–M3 is perpendicular to strike. Superimposed on profile T1–T3 are log-conductivity measurements collected from flowmeter data at well S7. For the spatial correlation analysis, log-conductivity information from both profiles was considered, as well as flowmeter measurements from wells S24, S18, S9 and S7.

ments from one of the flowmeter logs (well S9) that was used to both develop the petrophysical relationships and to create the prior distributions. The values obtained from this flowmeter log and from the tomographic data display a range hydraulic conductivity is representative of clean sands to silty sands (Freeze and Cherry, 1979).

The flowmeter and interpreted tomograms were subsequently used to obtain estimates of log-conductivity spatial correlation parameters following the procedure that was presented in Section 5.3 and that was applied to synthetic data in Section 5.4. The measured log-conductivity values from the four flowmeter logs were transformed into the wave number domain using fast Fourier transforms. An average spectral density curve in the vertical direction was obtained from the four logs using Eq. (11) and is shown in Fig. 7. Spectral density information in the vertical direction was then
calculated using the tomographic information. If only the estimated means of the log-conductivity prior distribution function, obtained from tomographic data and shown in Fig. 6, were used to estimate spatial correlation information, the high-frequency variations that may be present in the actual fields would not be detectable. To ensure that all elements of the model are present during spatial correlation estimation, the differences between the estimated means and the log-permeability values near the flowmeter locations were calculated. These error residuals were modeled, and random error with the modeled distribution was added to the log-permeability mean estimates that are shown in Fig. 6. The average spectral density function in the vertical direction was then calculated from the resulting two data sets that correspond to the tomographic profiles using Eq. (11) and is shown in Fig. 7. This figure suggests that the average vertical spectral density curves, obtained from the flowmeter and tomographic data, are similar. The tomographic data had a longer sampling window and thus, were able to capture more low-frequency information than the flowmeter data. As the sampling interval is the same for both the flowmeter and interpreted tomograms, the high frequency cut-off is the same. We expect, then, that since both data sets traverse a similar sampling window and have a similar sampling interval, that the spatial correlation information obtained from the disparate data types will be similar.

The spectral density curves shown in Fig. 7 were analyzed following the two approaches presented in Section 5.3. Vertical integral scales were calculated from the flowmeter and average tomographic spectral curves, first by using the integration approach given in Eqs. (12)–(14), and then by assuming an exponential structural covariance model and applying the second approach using Eq. (15) together with the variance calculated using Eq. (13). The results of the spatial correlation parameter estimations in the vertical direction are shown in Table 2, under the headings ‘Flowmeter data’ and ‘Radar tomography data’. Parametric uncertainly is also reported for the analysis performed using the second approach.
Table 2
Variance and vertical integral scale estimated using flowmeter and tomographic data at the Oyster, VA Bacterial Transport Site

<table>
<thead>
<tr>
<th>Data used to construct spectral density function</th>
<th>1/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowmeter data</td>
<td>15</td>
</tr>
<tr>
<td>Radar tomography data</td>
<td>15</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.33</td>
</tr>
<tr>
<td>Approach #1: $l_1$ (cm)</td>
<td>18.9</td>
</tr>
<tr>
<td>Approach #2: $l_1$ (cm) and $\tilde{\sigma}_2^2$</td>
<td>20.3, 0.79</td>
</tr>
</tbody>
</table>

The results presented in Table 2 show that analysis of both the tomographic and flowmeter spectral data yield vertical integral scale estimates of about 20 cm. Comparison of the results obtained using the different approaches suggests that a larger integral scale is obtained when an exponential covariance model is imposed on the data (Approach #2). The slightly larger integral scale values obtained from the tomography data may be due in part to low frequency information that the tomography data were able to sample compared to the flowmeter data, which did not traverse as long of a sampling window. Even with highly sampled borehole data available at the Oyster site, it was difficult to extract reliable information about spatial correlation parameters in the horizontal direction. A similar analysis of both tomography profiles in the horizontal direction suggested that the horizontal integral scale is on the order of 1.5 m, which yields an anisotropy estimate of approximately 7.5. This analysis suggests that the tomography data can be integrated with borehole data to provide denser estimates of hydraulic conductivity as well as more complete information about conductivity spatial correlation structure. At the Oyster study site, the hydraulic conductivity pdf estimates obtained from the tomography data will be used with the integral scale and anisotropy estimates to help to constrain the stochastic numerical flow studies which are being performed to understand bacterial transport in granular media (Scheibe et al., 1999).

7. Summary

Natural geologic systems exhibit large spatial variability of hydrogeologic properties such as hydraulic conductivity over a wide range of scales. Numerical modeling is often performed to gain an understanding of the hydraulic nature of the aquifer and to predict contaminant transport; these models require spatial distributions of hydraulic properties as inputs. It is typically not feasible to describe the hydraulic properties over the entire flow domain at the resolution needed for accurate flow modeling using conventional hydraulic property measurement techniques alone; the ability to adequately describe these properties is strongly dependent on the availability and distribution of field measurements and on the support scale of the measurement relative to the scale of hydrological heterogeneity. Incorporation of two- and three-dimensional densely sampled geophysical data with conventional hydrological data increases the amount of data available for the characterization and thus, has the potential to significantly improve the estimates of hydraulic properties and their spatial correlation over those estimates.
obtained from borehole data alone, and to provide this information at the relevant spatial scale.

The three methodologies that have been reviewed were all developed to investigate the usefulness of geophysical data for characterization problems. The first study presented a Bayesian methodology for incorporating geophysical data with hydrogeological data to provide log-permeability estimates. This study showed that the estimates were dramatically improved when this additional data were incorporated into the estimation procedure. The petrophysical relationship that was assumed to govern the system in this case study was non-unique; however, the Bayesian approach used in this study would be useful for other types of petrophysical relationships as well. The amount of improvement offered by this approach is a function of the information content of the prior estimates; the poorer the prior estimates (obtained from wellbore data), the more the potential for improvement offered by incorporation of geophysical data. In practice, the improvement offered by geophysical data is typically greatest in between wellbores, where the prior estimates obtained from interpolating wellbore data are the poorest.

The second investigation focused on using ML techniques for estimating saturation and permeability with in a bimodal system using GPR methods. In this case study, two options of saturation and permeability existed at each location due to the bimodality of the system and the petrophysical relationships that were assumed to govern the system. Using prior information from wellbores, geophysical data and the governing petrophysical relationships, an ML technique was employed to choose between the available hydrogeological parameter options at each location. Although we formulated our methodology to investigate saturation and permeability variations within a bimodal system, the methodology could be expanded to include more facies or to distinguish between other factors that effect the geophysical signal such as the type of fluid present in the pore space. Results of this case study suggest that saturation estimates should always improve and that permeability estimates may improve when incorporating radar tomographic information into the estimation procedure.

The last methodology focuses on using geophysical data for hydraulic parameter spatial correlation estimation. The results of this last study suggest that geophysical data can be extremely useful for spatial correlation estimation, especially when the geophysical data measurement spacing is small compared to the scale of hydrogeologic heterogeneity. When the geophysical data measurement spacing is large compared to the scale of the hydrogeological heterogeneity, or when multiple levels of variability exist within an aquifer, then borehole data can be incorporated with the geophysical data to gain an understanding of the spatial correlation structure. A key benefit of incorporating geophysical data into the spatial correlation estimation procedure is to obtain information about variability in the horizontal direction. This information, which is one of the most crucial pieces of information needed to predict contaminant flow in saturated systems, is extremely difficult to capture using conventional wellbore measurements.

Two of the presented methodologies were applied to a hydrological–geophysical data set collected to help characterize the DOE NABIR bacterial transport site in Oyster, VA. Although more field work of this nature is necessary to validate the usefulness and cost-effectiveness of including geophysical data in hydraulic parameter estimation procedures, the numerical and real field studies presented here suggest that incorporation
of geophysical data with limited hydrological data may provide valuable information about permeability estimates and their spatial correlation structure that is traditionally only obtainable by performing extensive and intrusive hydrological sampling.

Acknowledgements

We extend our appreciation to Eileen Poeter and to an anonymous reviewer for their thorough reviews and insightful comments. We sincerely thank Jinsong Chen and Nadim Copty for providing data and figures used within this study. This research was supported by NSF Grant EAR 9628306 to Yoram Rubin, by the Assistant Secretary for Energy Research, Office of Health and Environmental Research, Department of Energy contracts DE-FG07-96ER14726 to Yoram Rubin and DE-AC03-76SF00098 to Ernie Majer. Some of the computations were carried out at the Center for Computational Seismology (CCS) at Lawrence Berkeley National Laboratory.

References


Hubbard, S.S., Rubin, Y., Majer, E., Peterson, J., Chen, J., 1999a. Log-permeability estimation using multiple, co-located geophysical data sets within a Bayesian framework. EOS 80 (17), S120.


