SUBSURFACE CHARACTERIZATION USING GEOPHYSICAL DATA FUSION

METIN M. OZBEK[1] GEORGE F. PINDER[1] GILLES Y. BUSSOD[2]

[1]Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT
[2]New England Research Inc., White River Junction, VT

Abstract

We are introducing a novel technology applicable to the robust interpretation of the spatial distribution of permeability in heterogenous formations. The evidence theory approach is based on a combination of geophysical data together with expert opinion, and has its core strengths in 1) enabling the simultaneous use of probabilistic uncertainty and other (non-additive) representations of uncertainty (e.g., fuzzy or possibilistic), 2) integrating expert opinion with objective information, and 3) enabling a complementary (rather than sequential) use of geophysical data during the characterization process. A site application demonstrates the approach.

1. INTRODUCTION

Subsurface environmental, engineering, and agricultural investigations often require characterization of uncertain hydraulic parameters. Conventional sampling or borehole techniques for measuring these parameters are costly, time consuming, and invasive. Geophysical data can complement direct characterization by providing multi-dimensional and high resolution subsurface measurements in a minimally invasive manner.

Several techniques have been developed using joint geophysical-hydrological data to characterize the subsurface (e.g., Rubin et al. 1992, Copty et al. 1993). Hyndman et al. (1994) developed an inversion algorithm that uses both seismic cross-well travel times and solute tracer concentrations to estimate inter-well geology and hydraulic parameters. Copty and Rubin (1995) developed a stochastic approach that combined surface seismic data and well data to estimate the spatial arrangement of lithofacies and their mean hydrogeological parameters. The use of joint geophysical-hydrogeological data for parameter estimation in the unsaturated zone has been a focus in several studies (e.g., Mazac et al. 1988, Sheets and Hendricks 1995). Hubbard et al. (1997) investigated the joint use of ground penetrating radar (GPR) and borehole data for the estimation of vadose zone hydraulic parameters in bimodal systems. Some studies focused on extracting spatial correlation information from geophysical data, including Knight et al. (1996, 1997), Rea and Knight (1998) and Hubbard et al. (1999). Besides imaging subsurface structures, geophysical methods have also been used to monitor subsurface flow and transport processes. Time-lapse imaging has illustrated the potential for elucidating dynamic subsurface processes. A number of studies have illustrated the potential of electrical resistivity tomography (ERT) for monitoring tracer experiments in soils (e.g., Binley et al. 1996, 2002). The spatiotemporal information that is obtained using ERT can be used to calibrate flow and transport models (Binley et al., 2002).

Two important observations based on these studies are as follows: 1) No universal methods are available for converting geophysical attributes to hydrogeological ones. One of the most challenging problems is the issue of inconsistency in the methods of geophysical data acquisition and interpretation. Ezzedine et al. (1999) report on a related problem demonstrated by the fact that resistivity at a field site was measured along boreholes using several different tools, each characterized by a different support volume, sometimes leading to dramatically different results. The inherent nonlinear nature of the inversion problem associated with the GPR as well as ERT has been reported (Copty and Rubin et al., 1993; Vanderborght, 2005). 2) The complementary nature of the geophysical methods are not exploited. Indeed, there is no single geophysical method effective in most environmental and subsurface conditions, and all are strongly scenario-dependent. Thus it becomes essential to characterize the information that each individual geophysical method provides in combination. In other words, modeling the influence of various environmental conditions on these geophysical methods and on the information they provide in combination, in our opinion, is a very promising approach to improving our ability to determine subsurface parameters, especially permeability, which seems to have received very little attention.

Our goal in this paper is to develop a methodology that addresses both of the above mentioned issues. The strength of the approach is in explicitly quantifying and integrating into the characterization process the insight of a geophysicist on 1) the individual capabilities that these methods have and 2) what the meaning of the data is that they produce when interpreted collectively. The resulting model is based upon the mathematics of evidence or belief theory. In Section 2, we outline the essential mathematical background of the approach. In Section 3, we demonstrate an application followed by our concluding remarks in Section 4.

2. METHODOLOGY

We first introduce some definitions pertinent to the discussion in this section. We use capital letters to represent *sets*. We use lower case letters to denote *elements* in a set. For example, set X can represent an interval of permeability values and $x \in X$ denotes a particular permeability value in this set. We call set A a *subset* of X if it can contain only a portion of X and denote set inclusion by $A \subseteq X$. For example, if X represents an interval of log permeabilities $[-12, -8]$, then $A = [-10, -9]$ is a subset of K. Note that an element of a set is also considered a subset. The *intersection* of two sets is the set that they have in common. For example, the intersection of $B = [-11, -8]$ and $C = [-12, -10]$ is $D = [-11, -10]$ and it is denoted by $D = B \cap C$.

The central element of our framework is the *basic probability mass assignment* that we denote as m. For any subset A of X, $m(A)$ is the part of the belief that supports A (e.g., the claim that the (unknown) soil type x_0 is in the subset A of all soil types in X), and that, due to lack of information, does not support any subset of A. In the following we will refer to $m(A)$ as the *basic probability mass* of A. The initial total belief is scaled to

1, and thus $m(A) \in [0,1]$, with the sum of the beliefs $m(A)$ of all subsets equal to 1. The set X is the *universal set* and any subset of it with nonzero basic mass is a *focal element* of the evidence. We will follow the approach of Smets and Kennes (1994) to belief theory and do not require the empty set $\varnothing$ to have zero basic mass as is the case in the original theory of evidence outlined by Shafer (1976).

Additional functions that are essential to the theory are the set function of belief and plausibility. The *degree of belief* $bel(A)$ is defined over all subsets $A \subseteq X$ as

$$bel(A) = \sum_{B \subseteq A,\ B \neq \varnothing} m(B) \tag{1}$$

and it quantifies the total amount of "justified specific" support given to the claim that, for example, the unknown soil type x_0 is in the subset A of all soil types. The term "justified" means that B supports A, thus $B \subseteq A$, and the term "specific" means that B does not support A^c, the complement of A, thus $B \nsubseteq A^c$ and $B \neq \varnothing$.

Similarly, the *degree of plausibility* $pl(A)$ is defined as,

$$pl(A) = \sum_{B \subseteq X,\ B \cap A \neq \varnothing} m(B) = bel(X) - bel(A^c) \tag{2}$$

The plausibility of A in (2) quantifies the maximum amount of "potential specific" support that could be given to the claim that the soil type x_0 is in A. The term "potential" means that B might come to support A without supporting A^c if a further piece of evidence is taken into consideration, thus $B \cap A \neq \varnothing$.

The functions m, bel and pl are always in one to one correspondence. They all describe the same information but seen from different points of view. In particular, $bel(A)$ and $pl(A)$ can be viewed as the *lower* and *upper bounds* on a *family of probability measures* such that $bel(A) \leq P(A) \leq pl(A)$ where $P(A)$ denotes the (unknown) probability that the unknown soil type x_0 is in A. Note that using the belief framework to represent the uncertainty associated with measurements or estimates of soil types, or any other relevant variable, gains us a significant practicality: we are able to convey both randomness (as a probability value is assigned to each focal element) and imprecision (as a focal element can be a collection of soil types or an interval of permeability values). In other words, we are in the position of synthesizing in a single framework for both probabilistic and fuzzy information where the latter essentially is shown to be a body of evidence with nested focal elements (Klir and Wierman, 1999). We also note that the least committed belief function (analogous to probability density with maximum entropy in probability theory) defined on a universal set X is the *vacuous belief function* defined by $m(A) = 1$, if $A = X$, and 0, otherwise. It represents *total ignorance* as to the true value of the unknown soil type x_0. We will revisit this notion in the next section.

3. APPLICATION

We obtained a typical finely-bedded sample block of Berea sandstone (32 cm x 14.8 cm x 5.8 cm) from the Amherst quarry, Ohio. Berea sandstone has long been regarded as a laboratory standard in rock properties studies, owing to its uniformity and 'typical' physical properties. Although commonly described as 'homogeneous' throughout the literature, subtle heterogeneities due to mineral layering are visible from the sub-millimeter

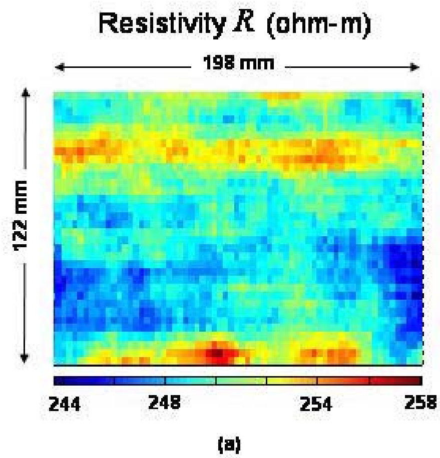

FIGURE 1. *AutoScan* data maps of Berea sandstone used in the study: (a) resistivity. (b) acoustic velocity (shear wave).

scale to decimeter scale. Similar heterogeneities are visible in the field from the decimeter scale to the meter scale.

The starting point in our methodology involves constructing detailed petrophysical maps of the sample, as exemplified by the resistivity and acoustic velocity (shear wave) data on a Berea slab shown in Figures 1 a and b, respectively. Each map consists of a collection of individual point measurements (> 2000) made with an *AutoScan* probe, a multi-probe physical properties scanner, developed by New England Research Inc. *AutoScan* allows millimeter-scale mapping of geophysical properties on a slabbed sample or core of rock or soil. The properties that can be measured by *AutoScan* in addition to resistivity and velocity include: gas permeability, complex electrical impedance (4 electrode, wide frequency coverage), and ultrasonic reflection (ultrasonic impedance and permeability).

Once the resistivity and velocity maps have been generated, we can apply our methodology to estimate the permeability of this Berea slab. In order to use these data simultaneously such that they complement each other, the geophysicist identified the ratio of resistivity to velocity, R/V_p, as a reasonable metric that he could relate to the permeability of the Berea slab (see Figure 2a).

This was followed by the geophysicist providing us with statements that related ranges of permeability values to R/V_p. A sample set of such statements that we will further analyze to demonstrate the methodology are as follows:

1. If the ratio R/V_p is too low, then permeability is unknown.
2. If the ratio R/V_p is low, then permeability is very low or low.
3. If the ratio R/V_p is high, then permeability is medium or high
4. If the ratio R/V_p is very high, then permeability is very high.

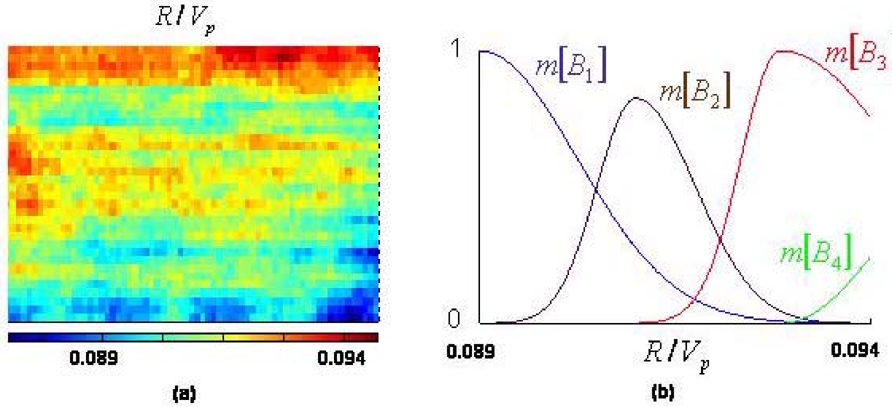

FIGURE 2. (a) Ratio of Resistivity and Velocity data (R/V_p). (b) Basic mass assignments. At a given R/V_p, $m[B_i](R/V_p)$ represents the the basic mass of focal B_i, which is a subset of possible conductivity ranges that the geophysicist thinks R/V_p implies.

In these statements above, observe that the geophysicist uses $X = \{$ very low ; low ; medium ; high ; very high $\}$, a collection of pre-determined permeability ranges (associated with soil types) as his set of universe. The estimation of a permeability range, which we demonstrate here, is the starting point of our ongoing work which involves the conditioning of a priori estimate of the permeability field as we will discuss later again.

The next step involved quantifying the meaning of these statements as shown in Figure 2b. Each of the curves $m[B_i]$ is associated with a subset B_i of X and reflects the meaning of the i'th statement above for given values of the ratio R/V_p. In the first statement, for example, nothing is known about the permeability range, thus $B_1 = X$ (i.e., total ignorance). In the second statement, the ratio is not informative enough to distinguish between low and very low permeability ranges whereas the geophysicist can conclude that it is not in the higher ranges, thus $B_2 = \{$ very low ; low $\}$. A similar argument applies to $B_3 = \{$ medium ; high $\}$ and finally, the last subset $B_2 = \{$ very high $\}$ is a trivial one.

We note that at any given R/V_p, we have

$$\sum_{i}^{n} m[B_i](R/V_p) = 1$$

thus, $m[B_i](R/V_p)$ is a basic probability mass assigned to focal B_i, and their collection for $i = 1, \ldots, n$ forms a body of evidence suggesting what the permeability range ought to be based on available geophysical data and the insight of the geophysicist.

It is worth mentioning that the numerical representation of basic probability mass assignments assumes that we can assign numbers representing belief. The general shapes and tendencies of the curves in Figure 2b are derived from the knowledge that we gather from the geophysicist. There remain some additional choices, which might initially appear as a drawback of the approach. However, it is not necessary to have precise estimates of these values inasmuch as Milisavljevic and Bloch (2003) also observed good robustness experimentally. There are two reasons for this robustness: first, the representations are used for rough information hence do not have to be precise themselves, and secondly, several metrics (although we have in our application only one, namely, R/V_p) are combined in the whole reasoning process, which decreases the influence of each individual value (of individual information). What is important is that the ranking is preserved, as well as the shape of these curves, and these are derived from the knowledge of the geophysicist.

The above outlined process applies to every pixel in Figure 2a resulting in pixel-based bodies of evidence. At each pixel, the final step is in making a decision as to what type of soil (i.e., permeability range) or object is driving the pattern in the data at a given location. Several decision rules are suggested in the literature. Two well known decision rules use the basic probability mass assignment function m to calculate the belief $bel(x_i)$ and the plausibility $pl(x_i)$, respectively, for each element x_i in X. The decision is to pick the element with the maximum belief or plausibility. Another alternative by Harmanec and Klir (1994) identifies the representative probability density as the one with maximum Shannon Entropy among all the probability densities bounded by the belief and plausibility functions of the available evidence. Here we consider the so called pignistic transformation of Smets (1998) that constructs a probability density function on X that satisfies certain rationality requirements. This transformation is given as:

$$P(x_i) = \sum_{A: x_i \in A \subseteq X} \frac{m(A)}{|A|\,(1 - m(A))} \tag{3}$$

where $P(x_i)$ represents the probability of permeability range or soil type x_i explaining the pattern in the data, and $|A|$ is the cardinality of subset A. The decision is to pick x_i with the maximum probability.

The result of the pignistic transformation at each pixel is shown in Figure 3b. We observe a good agreement between the estimated permeability ranges and the true permeability values in Figure 3b.

As part of our ongoing work on the general methodology for permeability field characterization, such permeability range or soil type estimates are being used to condition a prior estimate of the permeability field in the belief theory framework where the prior field estimate is based on site measurements and expert-based field information.

4. CONCLUSION

We have presented part of an ongoing work on a methodology to estimate a permeability field. This methodology is based on the theory of belief and it facilitates the integration of geophysical expert insight on the relation between permeability and a multiple set of data with diverse geophysical properties. Statements of a geophysicist on this relationship are translated into basic probability mass assignment functions which in turn facilitated the prediction of permeability ranges. An application of the methodology to a Berea sandstone slab demonstrated good agreements between the known permeability field and the estimated permeability ranges.

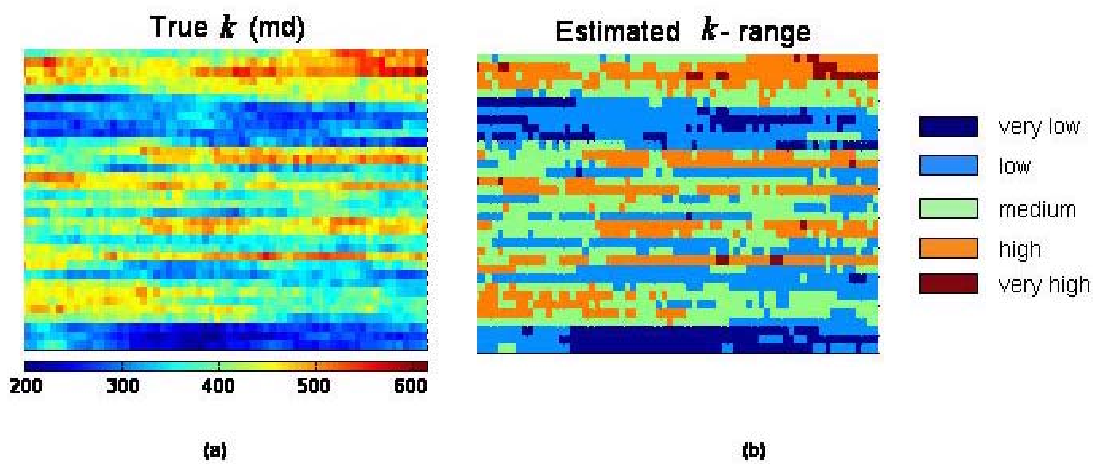

FIGURE 3. (a) True permeability field of the Berea sandstone slab obtained using *AutoScan*. (b) Estimated permeability-range field obtained using the proposed methodology based on belief theory.

References

[1] Binley, A.., S. Henry-Poulter, and B. Shaw (1996), Examination of solute transport in an undisturbed soil column using electrical resistance tomography, *Water Resources Research,* 32, 763-769.

[2] Binley, A., G. Cassiani, R. Middleton, and P. Winship (2002), Vadose zone flow model parameterization using cross-borehole radar and resistivity imaging, *Journal of Hydrology*, 267, 147-159.

[3] Copty, N. and Y. Rubin (1995), A stochastic approach to the characterization lithofacies from surface seismic and well data, *Water Resources Research*, 31, 1673-1686.

[4] Copty, N., Y. Rubin, and G. Mavko (199), Geophysical-hydrological identification of field permeabilities through Bayesian updating, *Water Resources Research*, 29, 2813-2825.

[5] Ezzedine, S., Y. Rubin, and S. Chen (1999), Bayesian method for hydrogeological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund site, *Water Resources Research*, 35, 2671-2683.

[6] Harmanec, D. and G. Klir (1994), Measuring total uncertainty in Dempster-Shafer theory: A novel approach, *International Journal of General Systems*, 22, 405-419.

[7] Hubbard, S.S., Y. Rubin, and E. Majer (1997), Ground-penetrating-radar-assisted saturation and permeability estimation in bimodal systems, *Water Resources Research*, 33, 971-990.

[8] Hubbard, S., Y. Rubin, and E. Majer (1999), Spatial correlation structure estimation using geophysical and hydrogeological data, *Water Resources Research*, 35, 1809-1825.

[9] Hyndman, D.W., J.M. Harris, and S.M. Gorelick (1994), Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resources Research*, 30, 1965-1977.

[10] Klir, G. and M.J. Wierman (1999), *Uncertainty-Based Information: Elements Of Generalized Information Theory.* Studies in Fuzziness and Soft Computing. 15. J. Kacprzyk (Ed.), Physica, New York.

[11] Knight, R., P. Tercier, and H. Jol (1997), The role of ground penetrating radar and geostatistics in reservoir description, *Leading Edge*, 16, 1576-1583.

[12] Knight, R., J. Rea, P. Tercier, and H. Jol (1996), Geostatistical analysis of ground penetrating radar data: A means of characterizing the correlation structure of sedimentary units. *Eos Trans. AGU,* 77(46), Fall Meet. Suppl., F237.

[13] Mazac, O., M. Cislerova, and T. Vogel (1988), Applications of geophysical methods in describing spatial variability of saturated hydraulic conductivity in the zone of aeration, *Journal of Hydrology*, 103, 117-126.

[14] Milisavljevic, N. and I. Bloch (2003), Sensor fusion in anti-personel mine detection using a two-level belief function model, *IEEE Transactions on Systems, Man, and Cybernetics. Part C: Applications and Reviews*, 33, 269-283.

[15] Rea, J. and R. Knight (1998), Geostatistical analysis of ground penetrating radar data: A means of describing spatial variations in the subsurface, *Water Resources Research*, 34, 329-341.

[16] Rubin, Y., G. Mavko, and J. Harris (1992), Mapping permeability in heterogeneous aquifers using hydrological and seismic data. *Water Resources Research*, 28, 1809-1816.

[17] Shafer, G (1976), *A Mathematical Theory of Evidence*. Princeton University Press.

[18] Sheets, K.R. and J.M.H. Hendricks (1995), Noninvasive soil water content measurement using electromagnetic induction, *Water Resources Research*, 31, 2401-2409.

[19] Smets, P. and R. Kennes (1994), The transferable belief model, *Artificial Intelligence*, 66, 191-234.

[20] Smets, P. (1998), The transferable belief model for quantified belief representation, in *Handbook of Defeasible Reasoning and Uncertainty Management Systems, v.1: Quantified Representation of Uncertainty and Imprecision,* edited by D.M. Gabbay and P. Smets, pp. 267-301, Kluwer Academic Publishers.

[21] Vanderborght, J., A. Kemna, H. Hardelauf, and H. Vereecken (2005), Potential of electrical resistivity tomography to infer aquifer transport characteristics from tracer studies: A synthetic case study, *Water Resources Research*, 41, W06013.