APPLICATION OF DEMPSTER-SHAFER THEORY TO HYDRAULIC CONDUCTIVITY

BREE R. DRUSCHEL[1], METIN OZBEK[2], GEORGE F. PINDER[1,2]

[1] *Research Center for Groundwater Remediation Design, The University of Vermont, Department of Mathematics and Statistics, 16 Colchester Avenue, Burlington, Vermont, USA*

[2] *Research Center for Groundwater Remediation Design, The University of Vermont, Department of Civil and Environmental Engineering, Votey Building, Burlington, Vermont, USA*

ABSTRACT

Dempster-Shafer theory can be used to combine probabilistic and subjective hydraulic conductivity data. While only probability theory is typically used to represent the uncertainty surrounding measured hydraulic conductivity, using Dempster-Shafer theory the opinions of hydrogeologists are quantified and incorporated into an alternative method for examining uncertainty.

1. INTRODUCTION

Uncertainty is an integral part of the mathematical representation of the environment. Moreover behavior forecasting using mathematical models requires the specification of the uncertainty associated with physically based parameters descriptive of the environment. In subsurface hydrology, for example, the hydraulic conductivity and its associated uncertainty (a measure of soil permeability) must be specified in equations descriptive of groundwater flow.

Traditionally probability is used to characterize uncertainty in hydraulic conductivity (K). It seeks to describe uncertainty arising from a lack of knowledge regarding concepts that are inherently crisp and well defined. However, classical probability itself is not applicable to situations where the concepts themselves are vague. In this situation one must consider other avenues for assessing the uncertainty.

One such technique is to employ evidence theory (Shafer, 1976) or Dempster-Shafer Theory (DST), which is a branch of the theory of monotone measures (a generalization of classical measure theory) (Klir, 2003). When there is insufficient information to define vague or fuzzy concepts both probability theory and DST can play a role. Effective husbanding of the environment requires that both forms of uncertainty be considered, both probability and evidence based. While the role of probability theory is well established, that of DST is not. In this paper we seek to explore the use of DST in the description of uncertainty surrounding hydraulic conductivity.

The impetus of this research is to develop a more cost effective method to provide an accurate approximation of a hydraulic conductivity field. Currently probability theory, usually within the framework of spatial interpolation (kriging), is used in an effort to generate a random field representation of a parameter (e.g. hydraulic conductivity). While some efforts have been made to accommodate subjective information (e.g. expert opinion) into these analyses via Bayesian methods, such efforts have been limited. Once a random field is generated Monte-Carlo simulation methods can be used to calculate state variable (e.g. hydraulic head) solutions under uncertainty.

It is our intention to use a DST framework to merge probabilistic and fuzzy (subjective) information in an effort to improve our ability to fully define a hydraulic conductivity field. Doing so will allow for a decrease in the number of hydraulic conductivity measurements that are needed to be obtained in the field. The advantages over using probability theory alone include 1) being able to use all available data to analyze hydraulic conductivity uncertainty (outliers are kept in the analysis) and 2) not having to make assumptions about distribution functions (e.g. typically a lognormal distribution is used to describe hydraulic conductivity of a site). The development of the above mentioned methods would improve our ability to properly describe subsurface heterogeneity and would result in improved models of subsurface environments.

The main focus of this paper will be on the application of Dempster-Shafer Theory to combine subjective information (expert defined uncertainty bounds) with objective hydraulic conductivity data sets measured in the Dakota Sandstone. In looking at three different methods to measure hydraulic conductivity, we are able to obtain expert opinion on the uncertainty surrounding each measurement technique from different experts and then combine their opinions to obtain a more comprehensive representation of the uncertainty surrounding the measured data.

2. THEORY

2.1 Dempster-Shafer Theory (or Evidence Theory).

Dempster-Shafer Theory (DST) is based on belief measures, (Bel) and plausibility measures, (Pl). If we let X be a universal set (frame of discernment), which would be all the possible hydraulic conductivity values in our data set, and P(X) denote the set of all subsets of X; or all possible intervals of hydraulic conductivity, then given $A \in P(X)$, Bel(A) is interpreted as the degree of belief (based on total evidence) that a given element of X belongs to the set A. Pl(A) represents the total evidence that an element belongs to A or to any of its subsets as well as additional evidence associated with sets that overlap with A. Belief and plausibility measures can be characterized by the basic mass (probability) assignment function:

$$m : P(X) \rightarrow [0,1] \qquad \text{where } m(\emptyset) = 0 \text{ and } \sum_{A \in P(X)} m(A) = 1. \tag{1}$$

Belief, plausibility, and mass assignments are alternative representations of the same evidence. Belief and plausibility can be calculated from mass assignments using the following relations:

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \tag{2}$$

$$Pl(A) = \sum_{B|A \cap B \neq \emptyset} m(B). \tag{3}$$

Mass assignments characterize the degree of evidence that the element we are interested in belongs exactly to the set A. Every set $A \in P(X)$, for which $m(A) > 0$, is called a focal element. If we let F be the set of all focal elements induced by m, then (F,m) is called the body of evidence. Total ignorance is expressed as m(X) = 1 and m(A) = 0 for all A∈X. In the presence of fuzzy information we will interpret belief and plausibility measures as bounds of the unknown probability of A. These two measures are equal in the case of pure probabilistic information.

2.1.1 Dempster's Rule of Combination.

Dempster's rule of combination is a standard method to combine evidence obtained from two or more independent sources. Originally introduced by Arthur Dempster (1967) it was later improved upon by Shafer (1976). The combination rule is defined as follows:

$$m_{1,2}(J) = \frac{\sum_{B \cap C = J} m_1(B) m_2(C)}{1 - T} \qquad \text{this holds for all}$$

$$T = \sum_{B \cap C = \phi} m_1(B) m_2(C) \qquad J \neq \varnothing \text{ and } m_{1,2}(\varnothing) = 0. \tag{4}$$

Here, J is simply the resulting joint focal element formed from the nonempty intersections of the expert focal elements. The symbol $m_{1,2}(J)$ is referred to as a joint basic mass assignment and represents the degree of the combined evidence that the element of interest belongs exactly to the set J. The variable T represents the mass associated with conflict in the combined evidence. In other words, the denominator acts as a normalization factor since the mass assignments of the focal elements must sum to one.

2.2 Probability Boxes.

In order to convert the field measured hydraulic conductivities into structures that we can use in the Dempster-Shafer Theory framework, probability boxes are created from which we can obtain focal elements and the corresponding mass assignments (Ferson et al., 2003). The idea behind the probability box (p-box) is that given information (a random variable) one can construct an upper and lower bound on the unknown probability distribution, essentially forming a "box" around the true distribution, Figure 1a. Discretization of this box provides one with focal elements (intervals on the abscissa) and their corresponding basic mass assignments (determined from the step size on the ordinate), Figure 1b. The benefit of using this approach is that no assumption has to be made about the distribution used to fit the data.

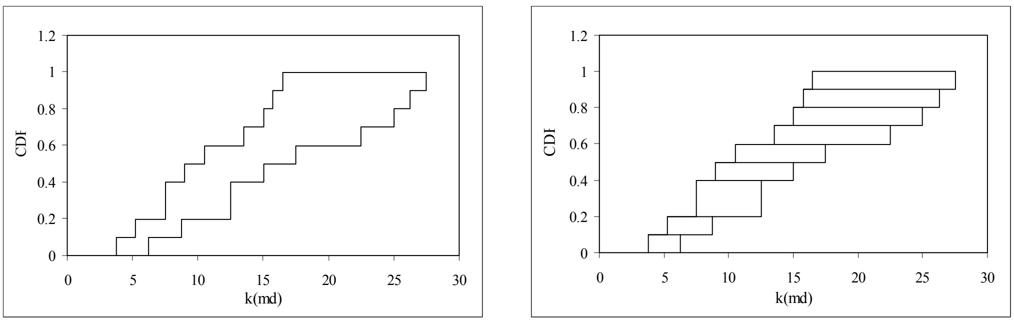

FIGURE 1. a) Example probability box. b) Probability box after discretization.

3. DATA SET

The data sets that are analyzed in this paper are permeability (k) values, with units of millidarcies (md), measured in the Dakota Sandstone within the Denver Basin (Belitz and Bredehoeft, 1988). Three different methods were used to obtain these permeability values; water-well pumping test, core analysis, and drill-stem analysis.

Water-well pump-test data were compiled from state water reports in the regions of South Dakota, southwestern Kansas, and southeastern Colorado. The sandstone here is a source of water and the measurements are taken at relatively shallow depths, less than 3000 feet. There are 74 points in this set, all of which are included in this analysis.

Next, there are core data that were compiled from state petroleum reports and other literature pertaining to regions of northeastern Colorado, southeastern Wyoming, and the Nebraska panhandle. Here the sandstone is primarily used as a source of oil, so the measurements are taken at depths from approximately 3,200 feet to 8,400 feet. All 161 data points from this set were analyzed.

The final data set consists of drill-stem data that were interpreted by Belitz and Bredehoeft using data from the USGS Petroleum Library in Denver. The data were obtained from the regions of northeastern Colorado, southeastern Wyoming, and the Nebraska panhandle. This was the largest data set at 453 data points, all of which were included in the analysis.

Plots of depth versus permeability demonstrate that there is not a strong correlation between the two, see Figure 2. A correlation coefficient of 1 implies the data is perfectly correlated and the closer the correlation coefficient is to 0 the more it becomes uncorrelated (Ghahramani, 2000). The correlation coefficient for depth versus hydraulic conductivity for the water-well pump-test, core, and drill-stem data are 2×10^{-5}, 0.180, and 0.023, respectively. Such small values suggest no correlation between depth and hydraulic conductivity in these data sets.

4. RESULTS

4.1 Probability Boxes.

The first step towards constructing our p-boxes is to obtain the uncertainty surrounding the hydraulic conductivity measurements taken using the three different methods. Two experts in the field of hydrogeology and familiar with the Denver Basin were asked to provide a range of uncertainty for each of the three methods. Neither expert had knowledge of the responses of the other. The values are given in Table 1. The uncertainty values were then used to create two p-boxes for each data set. P-boxes are constructed such that the upper and lower bounds (boundaries of the "box") are obtained from subtracting and adding the uncertainty from the empirical data.

The resulting p-boxes for the pump-test data (the smallest data set with 74 data points) yield 36 focal elements for both Experts. Given an uncertainty of +/- one order of magnitude, Expert 1's focal elements cover a range from 2 to 148,000 md. With an uncertainty of +/-0.5 orders of magnitude, Expert 2's focal elements range from 7 to 46,902 md.

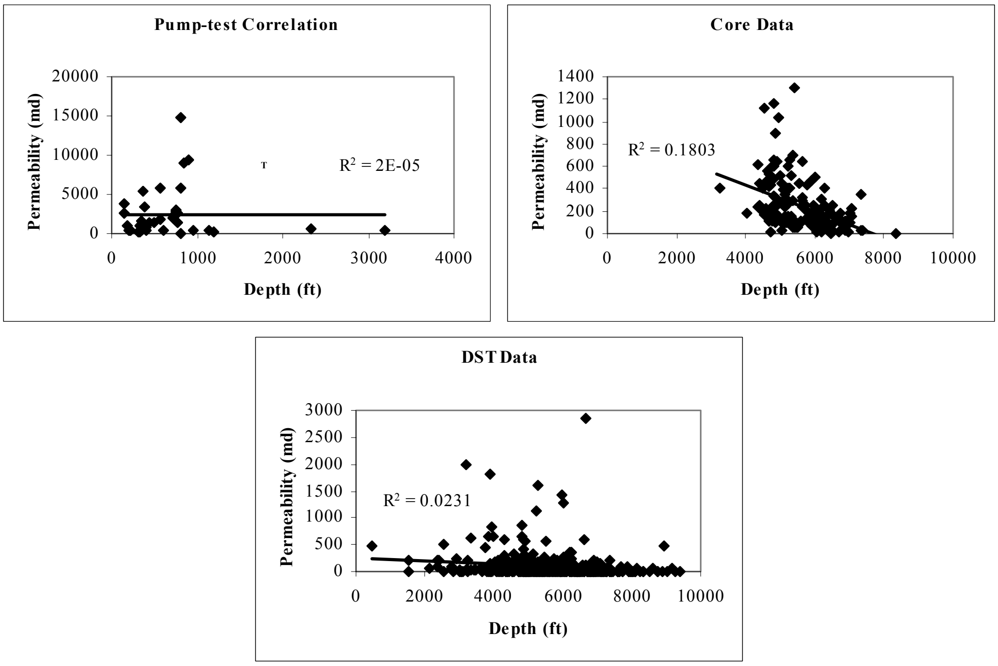

FIGURE 2. Correlation plots for hydraulic conductivity data.

TABLE 1. Expert assigned uncertainty to the three different methods for measuring hydraulic conductivity.

	Water-well Pump-test	Core Analysis	Drill Stem Analysis
Expert 1	± 1 order of magnitude	± 2 orders of magnitude	± 0.75 orders of magnitude
Expert 2	± 0.5 orders of magnitude	± 1 order of magnitude	± 0.5 orders of magnitude

The core data was the second largest data set with 161 points. While pump-tests give an average hydraulic conductivity value for a relatively large area, core analyses are done at a very local scale. Hence, they are quite accurate for a small area, however when applied to the aquifer the uncertainty is large. The p-boxes from Expert's 1 and 2 both yield 123 focal elements along with their respective mass assignments. The range of permeability values that result from Expert 1 (+/- 2 orders of magnitude uncertainty) and Expert 2 (+/- 1 order of magnitude uncertainty) are 0 to 130,000 md and 0 to 13,000 md, respectively.

Finally, for the drill-stem data, Expert 1 assigned an uncertainty of +/-0.75 orders of magnitude. The resulting p-box provides 258 focal elements ranging in permeability values from 0 to 16,004 md. Expert 2 assigned the drill-stem data an uncertainty of +/-0.5 orders of magnitude. This resulted in a p-box that yielded 218 focal elements ranging in permeability values from 0 to 9,000 md.

4.2 Dempster's Rule of Combination.

Once all the focal elements and corresponding mass assignments were determined, the calculations necessary to combine the information were performed. Analysis using

Dempster's rule of combination yielded conflict values for the water-well pump-test, core, and drill-stem data of T = 0.065741, 0.00027, and 0.38053, respectively.

Applying Dempster's rule of combination to the water-well pump-test data yielded 637 joint focal elements. In order to determine the effect the combination of information process has on the uncertainty, cumulative belief and plausibility were plotted for the individual Experts (Figure 3a and b) and the joint data (Figure 3c). A comparison of the three plots clearly shows a decrease in the size of the "box" formed from the belief (lower bound curve) and plausibility (upper bound curve). This translates to a decrease in uncertainty upon combination of information from the sources. The lognormal curve of the original data is plotted as well, since any probability distribution curve should fit within the bounds of the box. However, in this case there is a violation.

The second data set to be analyzed, core data, produced 1003 joint focal elements when combined using Dempster's rule of combination. The cumulative belief and plausibility plots (Figure 4) for this data set show a less dramatic decrease in the size of the "box" upon combination of the data. Here, however, the lognormal curve fits within the bounds for all three cases.

Finally we will assess the results of our largest data set, the drill-stem data, which also proved to have the most conflict, 0.38053. Dempster's rule of combination applied to the focal elements provided by the Experts resulted in 13,928 joint focal elements. Dempster's rule of combination reports no probability of permeability values past ~2,400 md (illustrated here by a cumulative belief and plausibility value of 1, Figure 5c). However, both Experts suggest that it is probable to have permeability values past this point. This perhaps demonstrates the weakness of Dempster's rule of combination when there is an increase in conflict between the information sources. The lognormal curve also violates the bounds in Figure 5c.

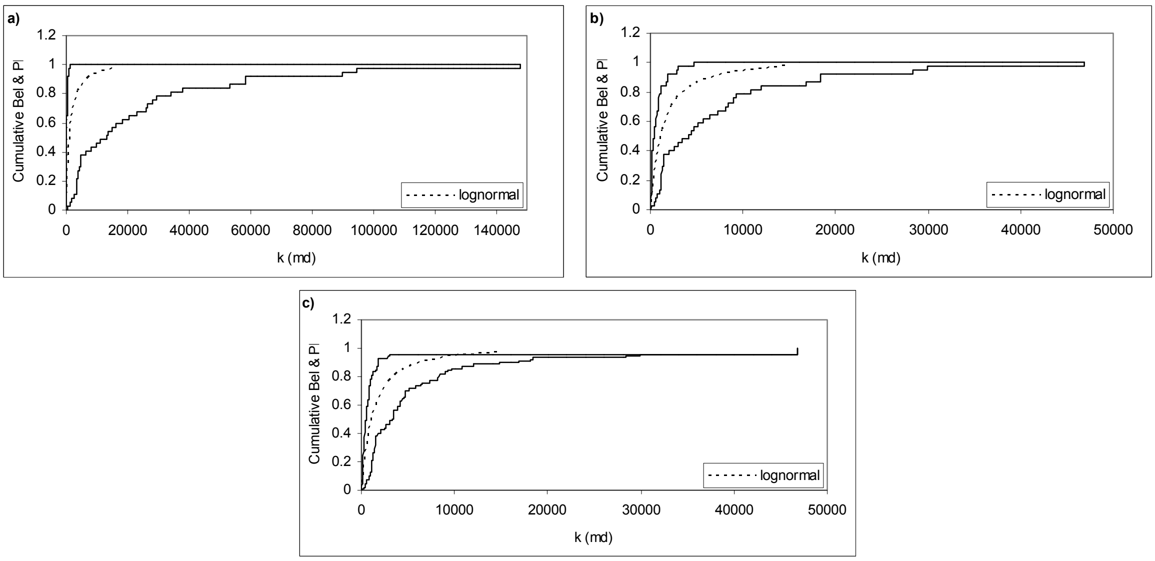

FIGURE 3. Cumulative belief and plausibility plots for the water-well pump data. a) Expert 1, b) Expert 2, and c) Dempster's rule of combination.

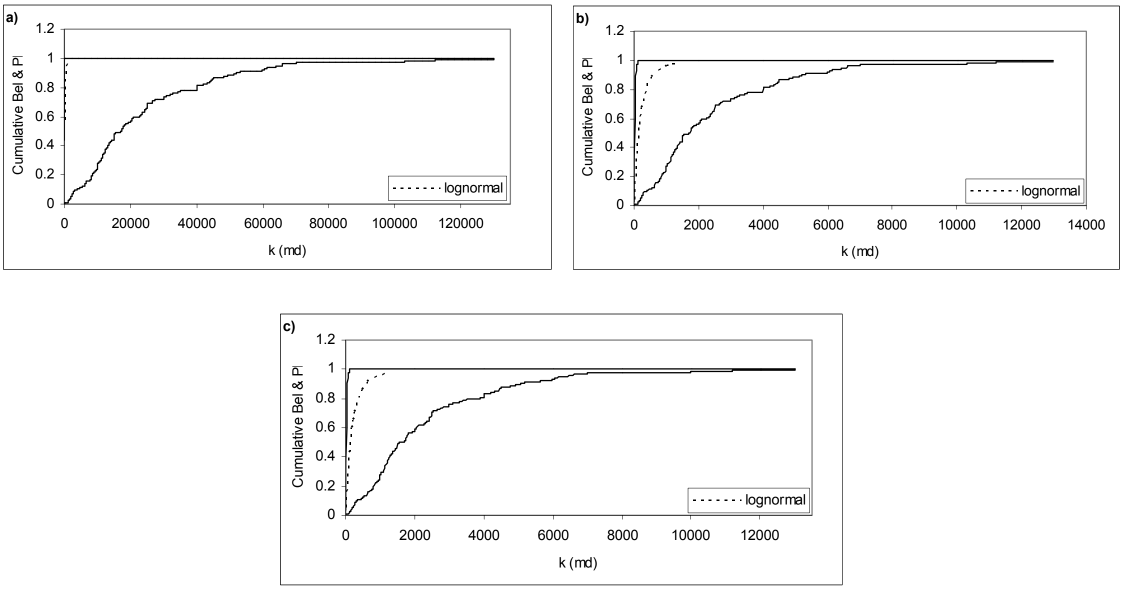

FIGURE 4. Cumulative belief and plausibility plots for the core data. a) Expert 1, b) Expert 2, and c) Dempster's rule of combination.

The cases where the lognormal curve violates the bounds of the cumulative belief and plausibility suggest that further investigation should be conducted to determine the appropriateness of the lognormal curve to describe the hydraulic conductivity data.

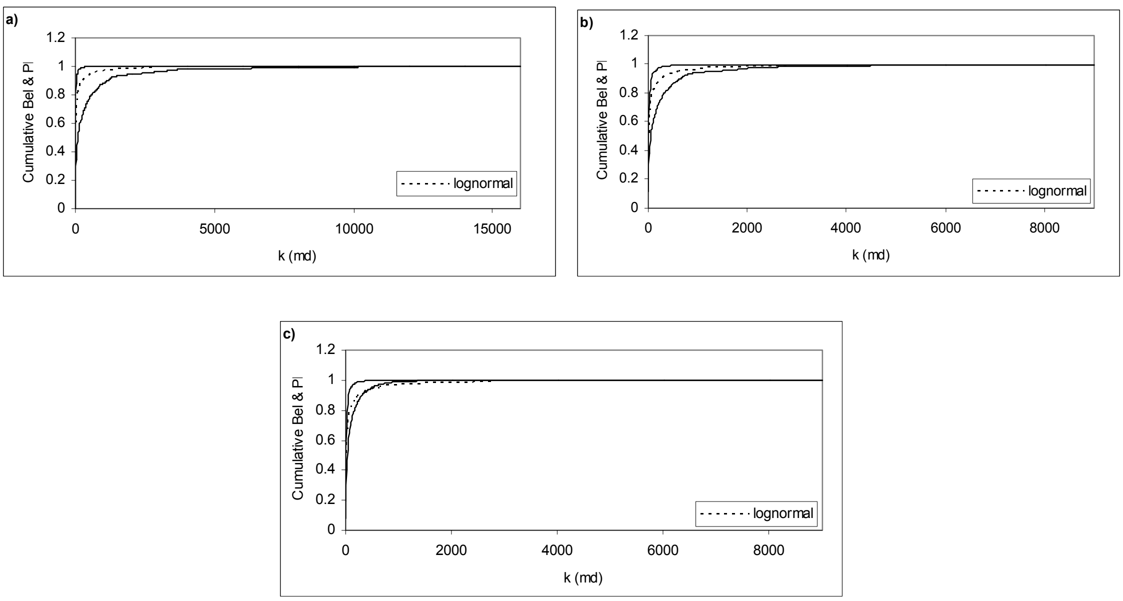

FIGURE 5. Cumulative belief and plausibility plots for the drill-stem data. a) Expert 1, b) Expert 2, and c) Dempster's rule of combination.

5. CONCLUSIONS

In this paper we have successfully taken hydraulic conductivity data, transformed it into Dempster-Shafer structures, and combined the empirical data with subjective data from two different experts. Typically, if one is given the same empirical data as used here, the uncertainty would be examined by assuming a lognormal curve since this has been the standard for years. Here we do not have to choose a distribution that may not be a best fit for the data (note Figures 3c, and 5c). Also, the process allows all available data to be used to describe the uncertainty of hydraulic conductivity. Being able to define the uncertainty of hydraulic conductivity using Dempster-Shafer Theory is a viable alternative to using the lognormal distribution.

6. ACKNOWLEDGEMENTS

We would like to thank our experts for their opinions: Ken Belitz and John Bredehoeft. This research is funded by the EPA under the Connecticut River Airshed Watershed Consortium grant.

REFERENCES

Belitz, K. and J. D. Bredehoeft (1988), Hydrodynamics of Denver Basin - Explanation of subnormal fluid pressures, *American Association of Petroleum Geologists Bulletin,* 72, 11, 1334-1359.

Dempster, A. P. (1967), Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics,* 38, 325-339.

Ferson, S., V. Kreinovich, L.Ginzburg, D.S. Myers, and K. Sentz (2002), Constructing Probability Boxes and Dempster-Shafer Structures, Sandia National Laboratories, *Technical Report SAND2002-4015*, Albuquerque, NM. Available at: http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf.

Ghahramani, S. (2000), Fundamentals of Probability, Prentice Hall, Upper Saddle River, New Jersey.

Klir, G.J. (2003), Uncertainty, *Encyclopedia of Information Systems*, 4, 511-521.

Shafer, G. (1976), *A mathematical theory of evidence*, Princeton University Press, Princeton, New Jersey.