6-24-2008

Approximate Reasoning in Hydrogeological Modeling

James Ross
University of Vermont

APPROXIMATE REASONING IN HYDROGEOLOGICAL MODELING

A Dissertation Presented

by

James L. Ross

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Civil and Environmental Engineering

May, 2008

Accepted by the Faculty of the Graduate College, The University of Vermont, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, Specializing in Civil and Environmental Engineering.

Dissertation Examination Committee:

George F. Pinder, Ph.D. — **Advisor**

Donna M. Rizzo, Ph.D.

Adel Sadek, Ph.D.

Richard Foote, Ph. D. — **Chairperson**

Frances E. Carr, Ph. D — **Vice President for Research and Dean of Graduate Studies**

Date: February 29, 2008

ABSTRACT

The accurate determination of hydraulic conductivity is an important element of successful groundwater flow and transport modeling. However, the exhaustive measurement of this hydrogeological parameter is quite costly and, as a result, unrealistic. Alternatively, relationships between hydraulic conductivity and other hydrogeological variables less costly to measure have been used to estimate this crucial variable whenever needed. Until this point, however, the majority of these relationships have been assumed to be crisp and precise, contrary to what intuition dictates.

The research presented herein addresses the imprecision inherent in hydraulic conductivity estimation, framing this process in a fuzzy logic framework. Because traditional hydrogeological practices are not suited to handle fuzzy data, various approaches to incorporating fuzzy data at different steps in the groundwater modeling process have been previously developed. Such approaches have been both redundant and contrary at times, including multiple approaches proposed for both fuzzy kriging and groundwater modeling. This research proposes a consistent rubric for the handling of fuzzy data throughout the entire groundwater modeling process. This entails the estimation of fuzzy data from alternative hydrogeological parameters, the sampling of realizations from fuzzy hydraulic conductivity data, including, most importantly, the appropriate aggregation of expert-provided fuzzy hydraulic conductivity estimates with traditionally-derived hydraulic conductivity measurements, and utilization of this information in the numerical simulation of groundwater flow and transport.

CITATIONS

Material from this dissertation has been published in the following form:

Ross, J., Ozbek, M., and Pinder, G.F. (2007). "Fuzzy inference of hydraulic conductivity from soil grain data and field observations." *Mathematical Geology*, 39(8), 765-780.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). "Fuzzy kalman filtering of hydraulic conductivity." *Computational Methods in Water Resources 26th International Conference*.

Material from this dissertation has been accepted by the *Journal of Hydrology*

Ross, J., Ozbek, M.M., and Pinder, G.F. "Expert judgment in hydrogeological site characterization." *Journal of Hydrology*.

Material from this dissertation has been submitted for publication to *Water Resources Research* on December 31, 2007 in the following form:

Ross, J., Ozbek, M.M., and Pinder, G.F. "Aleatoric and epistemic uncertainty in groundwater simulation." *Water Resources Research*.

Material from this dissertation has been submitted for publication to *Geoderma* on November 9, 2007 in the following form:

Ross, J., Ozbek, M.M., and Pinder, G.F. "General and site-specific means of defining fuzzy CPT-based soil classification." *Geoderma*.

DEDICATION

To my wife and our baby in her tummy,

To my parents,

and

To all the girls I loved before, who traveled in and out my door[1].

[1] Thanks to Willie Nelson for letting me borrow his lyrics.

ACKNOWLEDGMENTS

Where to begin? I'd like to thank the academy…

First of all, thanks to Dr. Walter Saukin at Manhattan College for deeming me worthy enough to recommend me to my advisor, Dr. Pinder (or, perhaps for wishing for me the most bitterly cold winters I have ever experienced).

My sincerest gratitude belongs to Dr. Pinder for accepting me as a student and mentee. Not only are you a distinguished leader in your field, but you are also an advisor of utmost skill and class.

To the Pinderita™, for evenings where I felt no pain.

I am grateful to my committee members, Dr. Richard Foote, Dr. Donna Rizzo and Dr. Adel Sadek, whose combined intellects and varied research backgrounds lend much credibility to this dissertation.

I will miss my officemates, past and present, Melissa, Xinyu, Bree, Ganesh and Hua. Nowhere can one find a group of individuals who can provide as fun and relaxed an office like these people can.

An officemate and dear friend, Zoe Dokou, showed me that humor transcends both cultures and international borders. Thank you for teaching me the English language and how to rock out on the drums to *Enter Sandman*.

My friends Lance, Jilby, Steph (Sweetums), Christine, Ashley and Ethan, have provided the perfect diversions and constant amusement over the past few years, from snow angels on the kitchen floor to Idol Tuesdays to endless chili. You are family for a couple whose real families are safely three states away.

I sincerely appreciate the support of the Crawfords, my inherited family. You are selfless individuals who have positively affected my life. See you at Thanksgiving. Do you have a hotel booked yet?

My sister Patty is one of the smartest people I know, hands down. Thank you for your persistant encouragement and style advice. With your guidance, this "geek" has endeavored to become somebody who transcends standard classifications. Get on the good foot, Moneypenny.

My niece Emily will certainly outrank me intellectually and athletically within a few short years. I enjoy every minute of the short time we get to hang out. I am extremely proud of your accomplishments, which are greater than any I have accomplished or could hope to accomplish in my life.

My parents are arguably the most caring, nurturing, encouraging and loving people a son could ever ask for. You are the ultimate safety net for a guy who oftentimes feels like he is walking on a tightrope. I can never repay you for all you have given me, though I will certainly try. Let's start with a bottle of wine. Perhaps one that smells like cat urine…or caviar?

Finally, my hot, talented and brilliant wife deserves utmost gratitude. "I've got 99 problems and my [wife] ain't one." You have not only suffered my antics, but also disregarded the gawking of others who silently ask, "What is *she* doing with *him*?" I love you dearly. Now you can say you've married a doctor.

To all those who have touched my life and whom I have unwittingly failed to acknowledge by name, see you in the paperback version.

This work was supported by the Connecticut River Airshed Watershed Consortium (CRAWC), the National Aeronautics and Space Administration (NASA), the Strategic Environmental Research and Development Program (SERDP), the letter **A** and by *Viewers Like You*.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

1: A REVIEW OF APPROXIMATE REASONING IN HYDROGEOLOGY

1.1. Introduction

The research contained within this dissertation predominantly addresses the appropriate role of fuzzy sets (Zadeh, 1965) in hydrogeologic applications. In addition, related mechanisms for the characterization of both expert knowledge and associated uncertainty are also addressed. These related frameworks include fuzzy logic (Zadeh, 1973), possibility theory (Zadeh, 1978), random set theory (Joslyn and Booker, 2004; Joslyn and Ferson, 2004; Joslyn and Kreinovich, 2005) and evidence theory (Dempster, 1968; Shafer, 1976). Though these tools perform distinct functions, they all share a common goal: the characterization of uncertainty. Traditionally, probability theory has been the "go to" framework for uncertainty characterization. As such, fuzzy set theory and fuzzy logic, especially, have faced significant criticism (Elkan, 1994; Garcia, 1994; Pelletier, 1994; de Silva and Attikiouzel, 1994) for attempting to challenge probability theory. In fact, fuzzy and probability theories play complementary roles in the characterization of uncertainty. In order to best comprehend how the mathematical tools listed above, collectively known as *approximate reasoning*, relate to probability theory and accomplish the task of uncertainty characterization from the perspective of hydrogeological investigations and modeling, the dual nature of uncertainty must first be introduced.

1.1.1. Uncertainty

Recent research into uncertainty has revealed a dichotomous nature to this property; uncertainty can be either aleatory or epistemic (Ferson and Ginzburg, 1996; Helton and Burmaster, 1996; Helton and Oberkampf, 2004). Aleatory uncertainty, also called irreducible or stochastic uncertainty arises due to the random nature of a system (Helton and Oberkampf, 2004). Epistemic, subjective, or reducible uncertainty is also referred to as state of knowledge and results, most often, from expert knowledge.

To best understand both types of uncertainty, it is most prudent and clarifying to consider them in concert. Consider the following simple scenario. Assume that the Environmental Protection Agency (EPA) wishes to determine the probability with which a certain contaminant will reach a target location within the next 50 years. Rather than run a sophisticated model, the EPA wishes only to rely upon the input of a very well-informed expert. Since the fate and transport of contaminants are governed by a naturally random system, the uncertainty surrounding the future event in question is aleatory (i.e. the contaminant will reach the target location with a probability P) (Helton et al, 2000a, b). However, the expert is additionally uncertain about the actual probability P. With more knowledge of the site, and as time progresses, the expert becomes more certain of the value P. For instance, perhaps at the present time, the expert is only able to specify an interval of values wherein he or she believes the exact probability value P lies. With increasing time and information, this expert-provided interval will likely narrow. As such, this latter uncertainty is epistemic in nature.

Helton and Oberkampf (2004) provide an excellent introduction to and summary of the dual nature of uncertainty. Moreover, in this dissertation, the chapter entitled *Aleatoric and Epistemic Uncertainty in Groundwater Fate and Transport Simulation* covers the topic in some depth.

Traditionally, probability theory has been employed in the characterization of both types of uncertainty. However, the ease with which a human agent is able to capture subjective knowledge using a probability framework is suspect (Ganoulis, 1996; O'Hagan and Oakley, 2004). As a result of this challenge rose the desire for a more facile manner of subjective knowledge characterization. Depending upon the nature of the source of uncertain information, however, different approximate reasoning tools apply.

The following describes and provides literature reviews of some of the mathematical constructs that developed out of the desire for an alternative theory for the lucid definition of subjective (expert) knowledge and epistemic uncertainty. For the sake of brevity, only those approximate reasoning theories and certain tools therein with relevance to this dissertation are introduced. References are provided to suggest avenues of more thorough investigation for the motivated reader.

In addition to the concise journeys into approximate reasoning, the remaining chapters of this dissertation are each given individual attention in the form of mini-literature reviews. These applications of approximate reasoning tools and theories to hydrogeology are, of course, presented in more detail in the chapters that follow.

1.1.2. Classical Sets and Finite Multivalued Logics

It's very warm today[2]. Until the mid-1960's (Zadeh, 1965), a declaration like this, formed by common parlance, provided little quantitative insight into the actual temperature outside. To make use of such information, it is desirable to know what constitutes *warm temperature* to the speaker. Intuitively, this human qualification (*warm*) of a measurable entity (*temperature*) is quite imprecise, and the manner in which such information was once quantified is quite awkward. For example, consider the quantification of warm temperature, as provided by classical set theory. By definition, a value $x \in X$ can only belong to a set A with full membership, $\mu_A(x) = 1$, or not belong at all, $\mu_A(x) = 0$:

$$\mu_A(x) = \begin{cases} 1 & if \quad x \in A \\ 0 & else \end{cases} \tag{1-1}$$

This formula is the mathematical representation of the *law of excluded middle* (Cooper, 1978), a philosophical concept that states that an item (i.e. temperature value) either belongs or doesn't belong in a set (i.e. the set *warm*). In terms of the example above, a temperature is either warm or not warm. A possible definition of warm temperature is provided in Figure 1.1, where any temperature in the interval *[21, 27]* is considered fully and equally representative of the notion *warm*. However, a brief inspection of this quantification of *warm* reveals a glaring inconsistency in the law of excluded middle: 21

[2] Since this dissertation was defended during a Vermont winter, this statement is both false and hopelessly optimistic.

degrees Celsius is considered to be warm, whereas 20.99 degrees Celsius is not defined as a warm temperature. This disconnect defies intuition simply because humans think in an imprecise manner. Thus, classical set characterizations of human knowledge imply a false sense of precision.

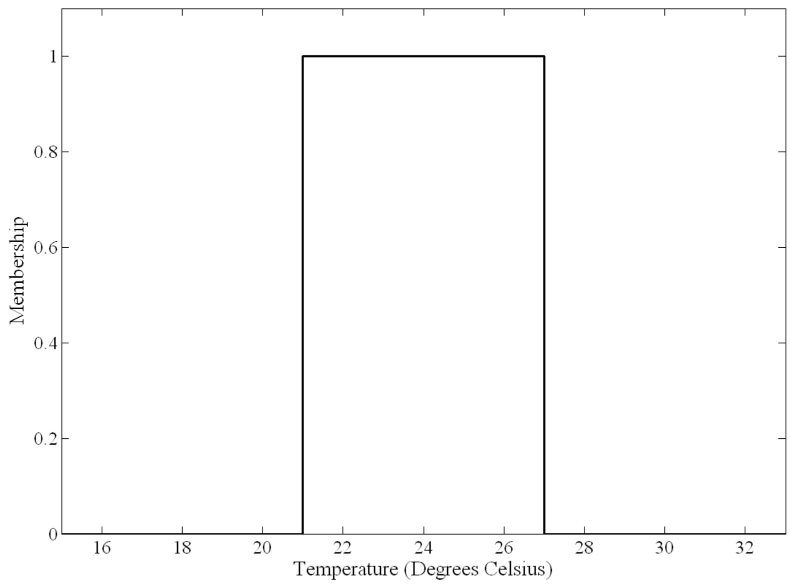


Figure 1.1. Classical set interpretation of *warm* temperature

Oddly enough, shortcomings of two-valued logic and the law of excluded middle, were first highlighted by Aristotle, the "father of [classical] logic" (Hurley, 2006), who noted that this law did not apply to future events; events that have not yet occurred are neither true nor false (Klir and Yuan, 1995). Nevertheless, it wasn't until the 20th century (almost 2500 years after Aristotle's death) that alternatives to Aristotelian logic were developed. Such alternatives were, for the most part, limited generalizations of two-valued logic, most notably three-valued logics, like those developed by Lukasiewicz (Klir and Yuan, 1995), and the n-valued logics of Lukasiewicz and Tarski, Reichenbach, and Godel (Malinowski, 2001). According to the three-valued logic, for example, a value could belong to a set with full, null or half membership.

Though certainly a step in the right direction these n-valued logics were still too unwieldy and unrealistic. The fact that that set memberships are still discrete precludes them from accurately portraying human knowledge. Fuzzy sets and fuzzy logic, however, introduced an infinite-valued logic that solved this problem.

1.1.3. Fuzzy Sets and Fuzzy Logic

The introduction of fuzzy sets by Lofti Zadeh (1965) provided an intuitive approach to the quantification of imprecision, especially in human-provided information. A fuzzy set is a natural extension of the classical set, where the constraint in Equation (1-1) no longer applies. Rather, the membership function μ_A permits the partial membership of elements $x \in X$ in a fuzzy set A that quantifies some human-provided notion. Consider the fuzzy set characterization of *warm* temperature in Figure 1.1. Note that the transition from temperature values with full membership to values with null membership is gradual. Specifically, using the example above, 21 degrees Celsius has membership of 0.75, and, intuitively, *20.99* has slightly less membership, $\mu_{warm}(20.99) = 0.7475$. In this example, temperature is a linguistic variable, because its values are qualitative (i.e. *warm*) rather than quantitative (Zadeh, 1975). Warm is called a linguistic value of the variable temperature. Linguistic values are generally specified by *fuzzy numbers*.

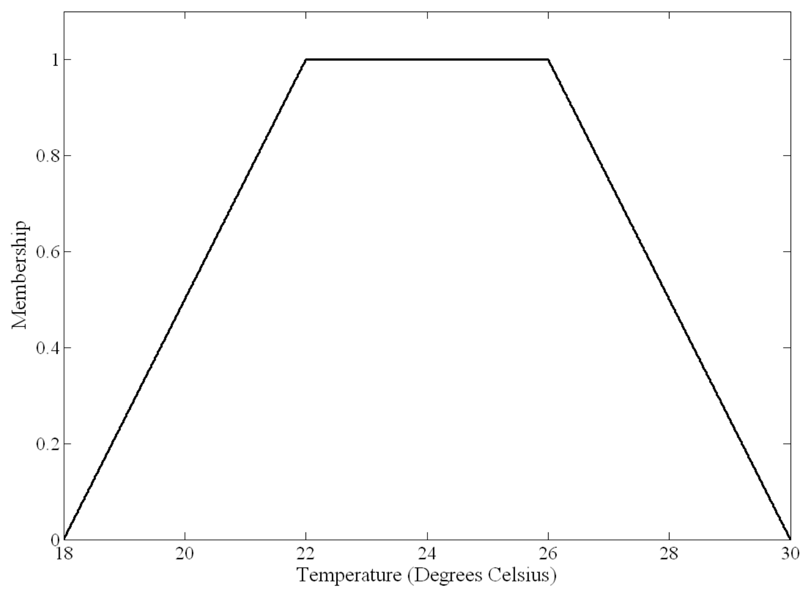


Figure 1.2. Fuzzy set representation of *warm* temperature

A fuzzy number is arguably the most common form of a fuzzy set and is defined on the set of real numbers. Though the most common fuzzy numbers are defined by triangular membership functions, the fuzzy set with the trapezoidal membership function in Figure 1.2 satisfies the three properties that define a fuzzy number (Klir and Yuan, 1995):

1. A must be normal,
2. α-cuts on A must be closed interval for $\alpha \in (0, 1]$,
3. the support of fuzzy set A must be bounded

The first property is easily understood: at least one value on the real number line must have full membership in the fuzzy set. It is intuitive to expect at least one number to fully embody the notion being defined. In the case of *warm* temperature, an interval of temperature values best exemplify that notion.

In order to understand the remaining two properties, an important aspect of the fuzzy set must be introduced. Fuzzy set membership values (the vertical axis in Figure 1.2) are also called α- (alpha) values. For every fuzzy number, associated with any α-value are two elements from the real number line (horizontal axis). For instance, in the fuzzy set in Figure 1.2, temperature values *20* and *28* degrees Celsius are associated with $\alpha = 0.5$. These two values comprise the lower and upper bounds, respectively, of an interval, or α-cut, created by horizontally cutting the fuzzy set at the *0.5* α-value. In other words, the *0.5* α-cut for the fuzzy set in Figure 1.2 is the interval [*20, 28*]. In general, an α-cut, ${}^{\alpha}A$, is an interval of values that have membership greater than or equal to α,

$$ {}^{\alpha}A = \{x \mid \mu_A(x) \geq \alpha\}. \qquad \textbf{(1-2)} $$

The second property of fuzzy numbers simply dictates that any α-cut in the interval (*0, 1*] must have real number upper and lower bounds. The reason the lower bound of the interval of α-values is open at zero is because the zero α-cut is the set of real numbers.

A strong α-cut, ${}^{\alpha+}A$, is defined as any real number that has membership greater than α in a particular fuzzy set,

$$^{\alpha+}A = \{x | \mu_A(x) > \alpha\}. \qquad \textbf{(1-3)}$$

A strong α-cut taken at $\alpha = 0$, is defined as the support of the fuzzy set. The support of the fuzzy set in Figure 1.2, for example, is the interval [*18, 30*]. The third property of fuzzy numbers, as defined above, states that a fuzzy set must have a support that is a closed interval.

While fuzzy numbers provide an intuitive means of subjective knowledge characterization, permitting one to model epistemic uncertainty surrounding human-provided information, the value of this tool is severely limited without any means of propagating this uncertainty through mathematical models to calculate output fuzzy numbers. Fortunately, mechanisms exist whereby fuzzy numbers can be operated upon just like certain, crisp numbers in deterministic problems. The extension principle (Dubois and Prade, 1991b), vertex method (Dong and Shah, 1987; Ross, 2004) and interval analysis (Moore, 1966) are tools that permit calculation with fuzzy numbers. Both interval analysis and the vertex method are predicated upon the discretization of a fuzzy set into a set of α-cuts and evaluation of the particular mathematical formulae via interval analysis. The extension principle, however, operates in unique manner.

Consider the extension of a mathematical model $Y = f(\boldsymbol{X})$, where $\boldsymbol{X}$ is a vector of fuzzy uncertain input parameters $\{A_1, \ldots, A_n\}$, whose membership functions are denoted by $A_1(x), \ldots, A_n(x)$, $x \in X = Reals$ and Y is the model output. The purpose of applying the extension principle to a mathematical model is the calculation of the membership function (i.e. $B(y),\ y \in Y = Reals$) of the fuzzy number output value given fuzzy number inputs. In order to best understand how the extension principle accomplishes this, the mathematical form of the principle must first be revealed:

$$B(y) = \sup_{y=f(x_1,\ldots,x_n)} [\min\{A_1(x_1),\ldots A_n(x_n)\}]$$

$$\forall y, x_1,\ldots,x_n \in \text{Reals} \qquad \textbf{(1-4)}$$

(Klir and Yuan, 1995). The following steps execute the extension principle:

1. Select a value y from the real number line;
2. find all input vectors $\boldsymbol{x_i} = \{x_1,\ldots,x_n\}$, $i = 1,\ldots,m$, such that $y = f(\boldsymbol{x}) = f(x_1,\ldots,x_n)$;
3. calculate the membership values $\{A_1(x_1),\ldots,A_n(x_n)\}$ for each of the m vectors found in Step 2;
4. select the minimum membership value from each of the above m vectors;
5. of the resulting m membership values, select the maximum (supremum);
6. this membership value is assigned to y in fuzzy set B.

Consider the function $f(x)$ and the input parameter defined by the discrete membership function A(x) in Figure 1.3.

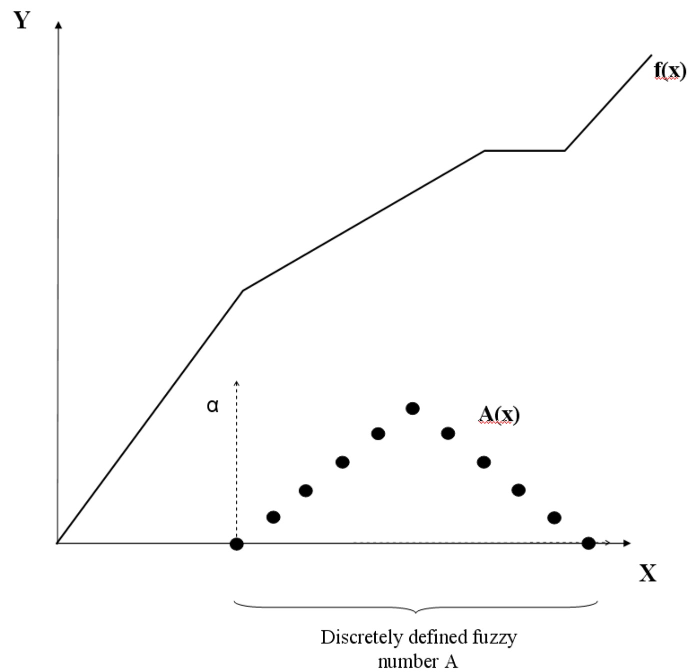
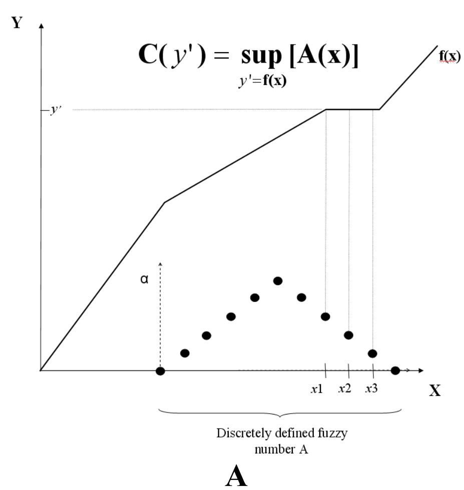


Figure 1.3. The extension principle can be applied to determine the membership function *B(y)* of a fuzzy set that results from executing *f(x)* on the fuzzy value *A*.

The steps to execute the extension principle are illustrated in Figure 1.4.



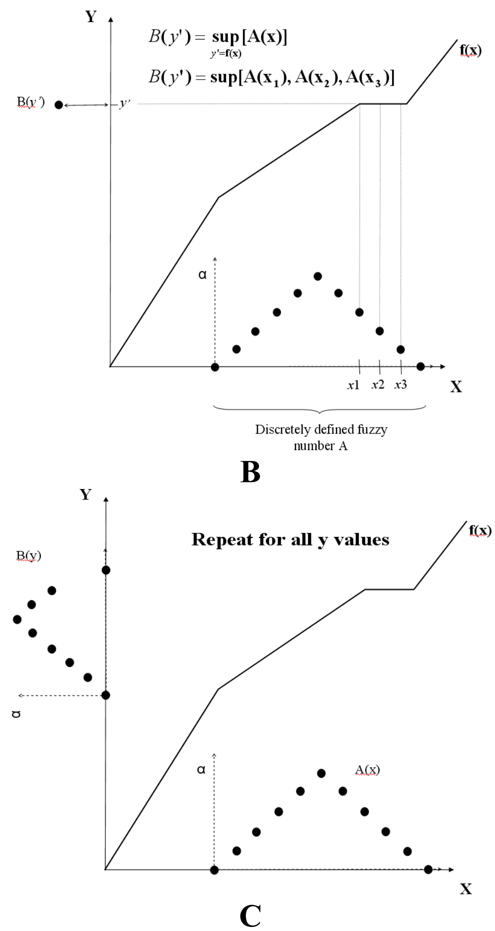


Figure 1.4. To apply the extension principle, (A) a possible output value y' ∈ Y is selected and all x ∈ X such that f(x)=y' are found; (B) B(y') is defined as the maximum membership value of {A(x_1), A(x_2), A(x_3)}; (C) repeat for all y ∈ Y

The application of the extension principle involves slightly more computation intricacy in cases where inputs number more than one. Consider the function $f(x) = x_1 + x_2$ and input parameters defined by the membership function $A_1(x)$ and $A_2(x)$ in Figure 1.5. In this case, the extension principle reduces to

$$B(y) = \sup_{y=x_1+x_2} [\min\{A_1(x_1), A_2(x_2)\}] \quad (1\text{-}5)$$

$$\forall y, x_1, x_2 \in \text{Reals}.$$

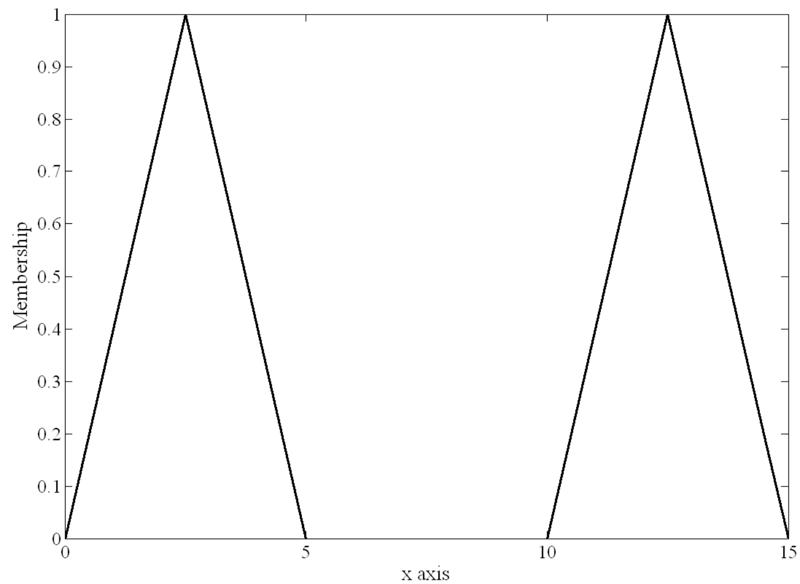


Figure 1.5. Fuzzy sets A_1 (left) and A_2 (right) are the fuzzy number input values to the function $f(x)=x_1+x_2$

Thus, one travels along the real line, selects a value *y* and finds all pairs $\{x_1, x_2\}$ such that $y = x_1 + x_2$. The fuzzy set memberships $A_1(x_1)$ and $A_2(x_2)$ are found for all pairs of values. For each pair, the minimum of the two memberships is selected and the supremum of these representative memberships is assigned as the membership of the value *y* in the fuzzy set *B*. Once the extension principle is executed on *f(x)*, the resulting fuzzy set *B* in Figure 1.6 is calculated. More information on the extension principle can be found in (Dubois and Prade, 1991b; Klir and Yuan, 1995) and in the chapter entitled *Aleatoric and Epistemic Uncertainty in Groundwater Fate and Transport Simulation*. It is further applied to both kriging and Kalman filtering equations in the chapters *Fuzzy Kalman Filtering of Hydraulic Conductivity* and *Kalman Filtering of Possibilistic Hydraulic Conductivity*.

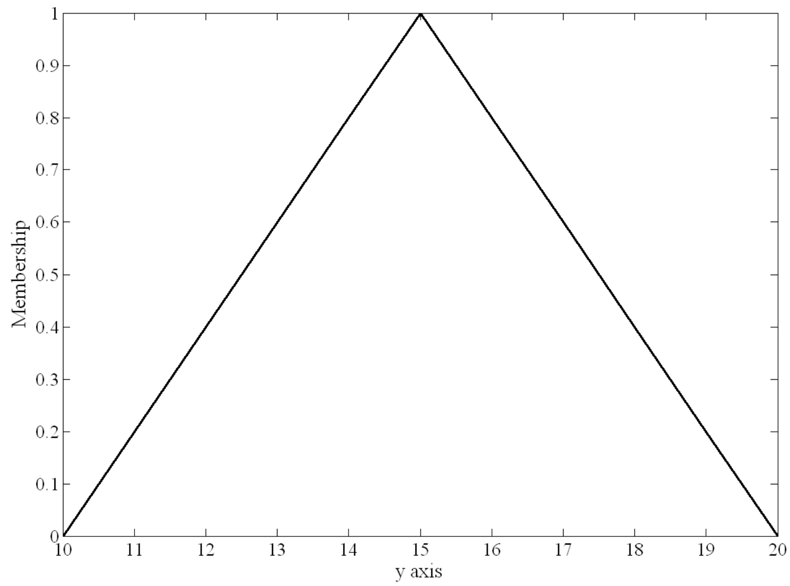


Figure 1.6. Fuzzy set *B* is the uncertain output parameter of the function $y=f(x)=x_1+x_2$ where A_1 and A_2 are the input fuzzy numbers in Figure 1.5

Where the extension principle effectively propagates fuzzy uncertain data through a traditional mathematical model, there exists an alternative approach for accomplishing this end whereby the model itself is rooted in fuzzy logic (Zadeh, 1973), rather than traditional mathematical equations. A fuzzy inference system is a collection of fuzzy rules that can be used to model a system. The most common fuzzy if/then rule, a mamdani fuzzy rule (Mamdani, 1975), is framed as a conditional statement, relating one or more fuzzy values of independent variables to a fuzzy value of an output variable,

$$\textit{If } X \textit{ is } A, \textit{ Then } Y \textit{ is } B. \qquad \textbf{(1-6)}$$

In this fuzzy rule, X and Y are the independent and dependent variables; A and B are fuzzy sets defined upon the respective variables.

A fuzzy rule of this form can have different interpretations depending upon the operators used to evaluate the rules. The most common interpretation is the conjunction based model (Klir and Yuan, 1995), where, for example, the rule above would be considered a fuzzy point (Zadeh, 1992) or patch (Kosko, 1993) $A \times B$ defined upon the Cartesian product $X \times Y$. Each patch is essentially a two-dimensional fuzzy set defining the the appropriateness of a particular rule for any pair of input and output values. The higher membership a pair $(x',y') \in X \times Y$ has in a two-dimensional fuzzy patch, the more representative this pair is of the mapping between *X* and *Y*. A collection of fuzzy rules, or fuzzy rule base, creates a collection of these patches that approximately models the relationship between variables *X* and *Y* (Figure 1.7). As with any mathematical relationship between independent and dependent variables, the fuzzy rule base, or series of overlapping patches (Figure 1.7), is used to estimate values of the dependent variable given values of the independent variables. Inputs to a fuzzy rule base can be both fuzzy and crisp in nature. Likewise, output estimates resulting from the execution of a fuzzy rule base can be fuzzy or crisp via defuzzification.

This fuzzy rule interpretation is most common in fuzzy control applications (Dubois and Prade, 1991a) that include the operation of systems such as cameras (So et al, 1993), washer/dryer systems (Berardinis, 1991), elevators (Zong et al, 2000), video games (Kosko, 1993), movies (Aitken et al, 2004), subway trains (Khanbaghi and Malhame, 1994) and helicopters (Amaral and Crisostomo, 2002). Applications to environmental sciences also abound, particularly in geohydrology (Bardossy and Disse, 1993; Nedungadi et al, 1994; Bardossy et al, 1995; Bagtzoglou et al, 1996; Bardossy, 1996; Fang and Chen, 1997; Schultz and Huwe, 1997; Dou et al, 1999; Coppolla et al, 2002; Finol and Jing, 2002; Vernieuwe et al, 2002a,b; Shokir, 2003; Demmico, 2004; Vernieuwe et al, 2007). Many such applications are summarized by Demmico and Klir (2004) and Ozbek and Pinder (2006), in chapter summaries below and in the three chapters entitled 1) *Hydraulic Conductivity Estimation via Fuzzy Analysis of Grain Size Data*, 2) *General and Site-specific Means of Defining Fuzzy CPT-based Soil Classification* and 3) *Fuzzy Spatial Reasoning for Reservoir Characterization*.

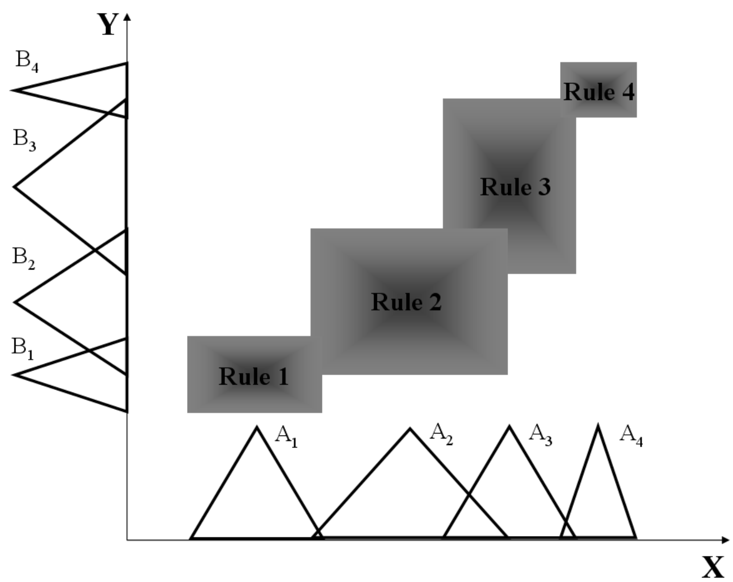


Figure 1.7. The above rule base approximates the relationship between independent variable X to dependent variable Y via four fuzzy rules

1.1.4. Evidence Theory

Uncertainty can result from a number of different sources (Klir, 1989), two of the most common being vagueness and randomness. The former is best handled by fuzzy sets and fuzzy logic, whereas the latter has been traditionally defined by probability theory. Where imprecision and randomness are coupled, however, subjective knowledge uncertainty characterization via probability theory can be awkward (O'Hagan and Oakley, 2004). As such, a quantitative framework of imprecise probabilities, called evidence theory, or Dempster-Shafer theory (Dempster, 1968; Shafer, 1976), has been developed for the definition of coupled epistemic and aleatoric uncertainties.

Evidence theory is primarily based upon the definition of focal elements (Wang and Klir, 1992). A focal element $\{A, m(A)\}$ has two components: a set A of elements $x \in X$, and a corresponding probability mass assignment $m(A)$, which is a probability value applied to a set A. This mass assignment can be interpreted as the proportion of evidence that gives credence to the claim that the unknown true value $x' \in X$ lies in set A (Klir and Yuan, 1995).

The properties of mass assignments should look familiar:

$$m : A \rightarrow [0,1]$$

$$\sum_{i=1}^{n} m(A_i) = 1, \qquad (1\text{-}7)$$

where n is the number of focal elements in a given body of evidence (*Wang and Klir, 1992*). These properties are quite similar to those of a probability distribution. In fact, the most significant difference between mass assignments and probability distributions is the form of the elements to which they assign probabilities; where X is the domain of interest, mass assignments operate upon subsets of the power set of X and probabilities are defined upon $x \in X$. As such, a probability distribution is a special case of the probability mass assignment. Likewise, probability theory is a special case of evidence theory (*Klir and Yuan, 1995*).

A collection of focal elements $\{A_i, m(A_i)\}$, $i = 1,\ldots,n$, such that Equation (1-7) holds, comprises a body of evidence, or *random set* (Joslyn and Booker, 2004; Joslyn and Ferson, 2004; Joslyn and Kreinovich, 2005), that characterizes the uncertainty regarding the true value $x' \in X$, essentially defining a probability distribution upon intervals $A_i \in 2^X$ that are elements of the power set of X. Thus, this uncertainty is both aleatoric (probability distribution) and epistemic (intervals). In the special case where each of the sets A_i contains a singleton, the uncertainty is aleatoric only.

Consider a collection of measurements of some particular variable. This set of measurements has a mean and variance, which, along with an assumption regarding the type of distribution, can be used to create a random variable representing the uncertainty in the set of measurements. Assume, however, there is a reducible error associated with each of the measurements, where the actual true value is unknown but expected to lie within an order of magnitude of the measured value. The resulting collection of uncertain measurements is a random set.

The information contained in a random set can be shown to bound from above and below the unknown true random variable. These upper and lower bounds on the probability are called plausibility and belief, respectively. Given the known probabilities

$m(A_i)$ for the sets $A_i \in 2^X$ in the body of evidence, the belief (*Bel*) and plausibility (*Pl*) for all subsets *B* of the power set of *X* are calculated as follows:

$$Pl(B) = \sum_{A_i | A_i \cap B \neq \emptyset} m(A_i) \quad \textbf{(1-8 a)}$$

$$Bel(B) = \sum_{A_i | A_i \subseteq B} m(A_i)\ . \quad \textbf{(1-8 b)}$$

Equations (1-8 a) and (1-8 b) reveal the meanings behind belief and plausibility. *Bel* is the total amount of evidence that supports the claim that unknown true value $x' \in X$ is contained in *B* or any of its subsets, where *Pl* represents the evidence that *x'* lies in *B*, its subsets and any intersecting sets. From this, it follows that for any set *B*, $Bel(B) \leq Pl(B)$. Moreover, since *Bel* and *Pl* are lower and upper bounds on the probability *P* of event B, the following is also true:

$$Bel(B) \leq P(B) \leq Pl(B)\ . \quad \textbf{(1-9)}$$

Cumulative *Bel* and *Pl* functions (Yager, 2004; Druschel et al, 2006; Mathon et al, 2008) are also calculable from a body of evidence, providing a transparent representation of the relationships between belief, plausibility and probability. Letting L_x denote a subset of *X* containing values less than or equal to *x*, the cumulative plausibility and belief of L_x can be calculated from focal elements $\{A_i,\ m(A_i)\}$, where the lower and upper bounds of A_i are a_i and b_i, respectively, using Equations (1-10 a) and (1-10 b) (Yager, 2004):

$$Pl_{Cumulative}(L_x) = \sum_{A_i | A_i \cap L_x = \varnothing} m(A_i) = \sum_{i | a_i \le x} m(A_i) \quad \textbf{(1-10 a)}$$

$$Bel_{Cumulative}(L_x) = \sum_{A_i | A_i \subseteq L_x} m(A_i) = \sum_{i | b_i \le x} m(A_i) \quad \textbf{(1-10 b)}$$

Given a certain body of evidence, the cumulative belief and plausibility in Figure 1.8 (from the chapter *Aleatoric and Epistemic Uncertainty in Groundwater Fate and Transport Simulation*) were constructed and shown to bound an independently calculated cumulative distribution function (CDF). In this case, the CDF was calculated from a set of concentration estimates, where the belief and plausibility functions resulted from a set of concentration focal elements.

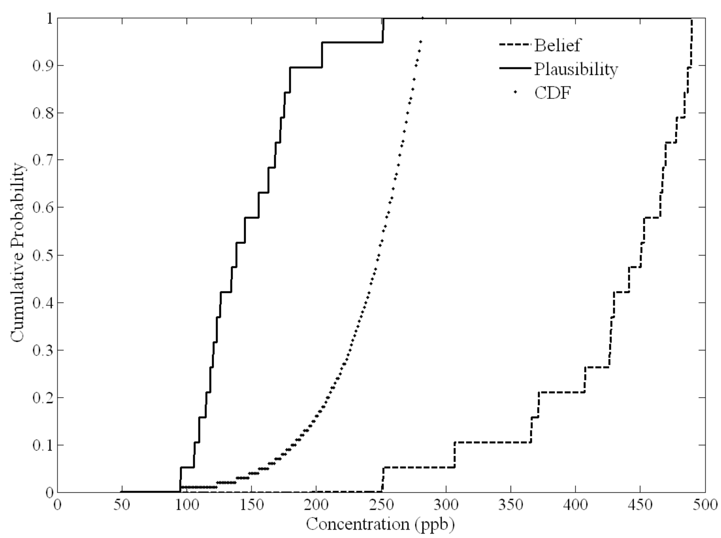


Figure 1.8. The cumulative belief and plausibility bound the known (in this case) cumulative distribution function (CDF) (from Ross et al, *submitted to WRR*)

1.1.5. Possibility Distributions

Where a body of evidence is comprised of a set of nested focal elements, like α-cuts of a fuzzy set, evidence theory reduces to one of its sub-theories called possibility theory (Zadeh, 1978). Though possibility theory is a special case of evidence theory, it is quite unique and warrants a short investigation, here, especially into its most commonly applied tool, the possibility distribution (Zadeh, 1978).

Figure 1.6 represents the uncertain output resulting from the application of the extension principle to the formula $f(x) = x_1 + x_2$, where the inputs to this function are fuzzy (Figure 1.5). Though the uncertain value in Figure 1.6 appears to be a fuzzy number, it is, in fact, a possibility distribution.

The possibility distribution, a non-probabilistic means of capturing uncertainty, is the principal contribution of possibility theory (Zadeh, 1978). Structurally identical to a fuzzy set, a possibility distribution can be expert-provided, though it represents a different facet of uncertainty. Whereas a fuzzy set typically models linguistic variables (Zadeh, 1975) that are certain yet imprecise notions, a possibility distribution represents a lack of knowledge regarding the true value of some variable.

Consider a simple example involving a clay lens in the subsurface. Two experts are available to opine upon the depth of this lens: one expert possessing substantial familiarity with the site and one expert who possesses only knowledge of a similar site, though not this site in particular. The first expert is asked to specify the depth of the clay lens, and responds that that clay lens is deep, using a fuzzy set to characterize this linguistic variable (Figure 1.9). The same information is requested of the second expert, who can merely define what depths he perceives are possible based upon his knowledge of similar sites, employing a possibility distribution to do so (Figure 1.9). In this case, the resulting characterizations of the expert knowledge are structurally identical, though intuitively, the experts are expressing two different sentiments. Expert One interprets Figure 1.9 as a membership function, where *50 meters* best represents his notion of deep. Expert Two, on the other hand, interprets *50 meters* as the most possible depth of the clay layer.

More commonly, though, possibility distributions result from the operation of an input fuzzy number by a mathematical model. Where fuzzy model inputs are most often certain yet imprecise expert-provided values, the execution of a mathematical model calculates uncertain values in the form of a possibility distribution.

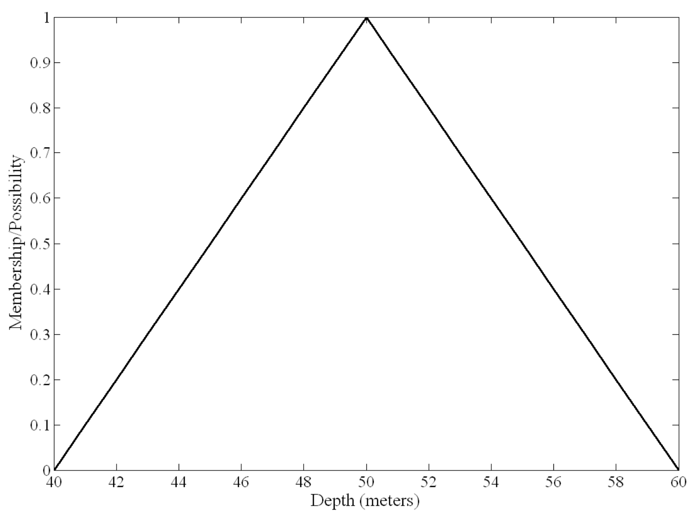


Figure 1.9. Expert One interprets this distribution as a membership function defining his notion *Deep*; Expert Two, on the other hand, provided this as a possibility distribution representing his knowledge of the clay lens' depth

Because possibility distributions model uncertainty that results from a lack of knowledge as opposed to a lack of precision, they are more similar to probability distributions than fuzzy sets. Aside from direct possibility-probability transformations (Jumarie, 1995; Oussalah, 2000; Klir, 2006), where a possibility distribution can be re-interpreted as a probability distribution (and vice versa) after an, oftentimes simple, functional transformation, possibility distributions are found to be naturally related to probability distributions via focal elements. A possibility distribution can be transformed into a set of nested α-cut-like intervals, which are, in fact, focal elements. Applying Equations (1-8 a), these nested focal elements are translated to necessity (Nec) and possibility (Pos), the possibility theory versions of *Bel* and *Pl*, respectively (Klir and Yuan, 1995).

1.2. Hydrogeological Uncertainty

The predominant source of uncertainty in hydrogeological modeling originates from model parameter measurements and estimates. This uncertainty can result from measurement error (Højberg, and Refsgaard, 2005), spatial and temporal heterogeneity (Gelhar, 1986; de Marsily et al, 1998; Nillson et al, 2007), error in the estimation procedures (Isaaks and Srivastava, 1989), expert knowledge (Piotrowski et al, 1996), amongst other sources. Regardless of the source, though, uncertainty characterization in groundwater fate and transport modeling has predominantly been accomplished by probability theory (Gelhar, 1986; Ozbek and Pinder, 2006). As such, groundwater model parameters are traditionally represented by stochastic, or random, variables. In order to accommodate the randomness of model parameters (i.e. boundary conditions and state variables), random versions of the groundwater model equations, called stochastic models (Freeze, 1975; Gelhar, 1986; Winter, 2004) are implemented. The execution of these stochastic models usually relies upon the implementation of Monte Carlo methods (Metropolis and Ulam, 1949), which, in general terms, sample the probability distributions of input parameters (i.e. hydraulic conductivity), execute model equations to calculate model estimates (i.e. head and concentration) and aggregate these estimates into random variables (Gelhar, 1993).

However, as mentioned above, the probabilistic approach is not always adequate or appropriate (Ganoulis, 1996; O'Hagan and Oakley, 2004). As a result I have applied the approximate reasoning methods introduced above throughout the entire groundwater fate and transport modeling process. This will now be discussed.

1.3. Groundwater Fate and Transport Modeling

The groundwater modeling process can be segmented into two broad stages: 1) hydrogeological aquifer characterization, and 2) implementation of a groundwater model. The first stage is comprised of hydrogeological parameter evaluation and spatial estimation. This includes the delineation of distinct soil formations and the identification of groundwater model input values. Arguably, the most commonly researched groundwater flow and transport model input parameter is hydraulic conductivity because of its significant impact upon the velocities with which water and contaminants travel (Bear, 1979; Sudicky, 1986; Domenico and Schwartz, 1990). Conventional approaches to hydraulic conductivity measurement include pumping tests, slug tests and inverse methods. However, such measurements are costly. The geological characterization of an aquifer is helpful because hydraulic conductivity can be predicted from soil properties that are, in general, less costly to measure than hydraulic conductivity.

The second stage of the groundwater modeling process is the definition and implementation of a flow and transport model. In practice, models are sets of differential equations whose solutions often require the application of numerical methods. Moreover, since model inputs are random variables, sampling regimes facilitate the execution of model equations upon realizations of discrete values, drawn from the random variables. The resulting realizations of crisp model estimates are aggregated into random variables. Software to execute these modeling tasks are prevalent (i.e. *Modflow/MT3D*, *Princeton Transport Code*) in practice. As such, innovative modeling techniques are fairly uncommon in real-world applications.

The bulk of this dissertation addresses the concepts behind the investigation of aquifer soil properties, the prediction of hydraulic conductivity from these properties, the accurate estimation of hydraulic conductivity throughout an aquifer, and the appropriate characterization of uncertainty throughout all of these properties. Significant attention is also given to the considerations of modeling with imprecise hydrogeological parameters. The chapter entitled *Fuzzy Spatial Reasoning for Reservoir Characterization* introduces a means of aquifer soil formation mapping that is predicated upon expert knowledge. A fuzzy logic-based method of hydraulic conductivity prediction from soil grain size and borehole log analyses is presented in *Hydraulic Conductivity Estimation via Fuzzy*

Analysis of Grain Size Data. The chapter, *General and Site-specific Means of Defining Fuzzy CPT-based Soil Classification*, introduces a fuzzy logic estimator of hydraulic conductivity that is reliant upon cone penetration test (CPT) data, rather than expert knowledge. The accurate estimation of hydraulic conductivity, where measurements and/or the aquifer's spatial continuity are imprecisely known, is addressed in the chapters *Fuzzy Kalman Filtering of Hydraulic Conductivity* and *Kalman Filter Updating of Possibilistic Hydraulic Conductivity*. Practical considerations are given to alternative characterization of uncertainty in hydraulic conductivity by random sets and the modeling with such uncertain parameters in *Aleatoric and Epistemic Uncertainty in Groundwater Simulation*. Finally, a method of modeling with imprecise hydraulic conductivity values, reminiscent of stochastic modeling, is introduced in *Respecting Correlation While Modeling with Possibilistic Data*.

1.3.1. Fuzzy Spatial Reasoning for Reservoir Characterization

Borehole logs (Chapellier, 1992; Temples and Waddell, 1996), soil grain size analyses and cone penetration tests are the hallmarks of soil-based aquifer investigations. In general, borehole logs provide a qualitative assessment of an aquifer's vertical geological make-up at various point locations. Cone penetration tests are similar to borehole logs in the sense that they provide insight into the vertical soil profile at a finite number of locations in an aquifer. On the other hand, a soil grain size analysis, or grain size distribution, provides quantitative information on the composition of a soil sample extracted from an aquifer, rather than an entire vertical profile.

Though these tools of soil investigations are relatively inexpensive, they are invasive and provide a limited amount of information relative to an entire aquifer's soil composition. Nevertheless, there are methods of expanding upon this information, such as kriging (Matheron, 1971; Isaaks and Srivastava, 1989) to provide a more complete characterization of an aquifer's soil properties. These procedures for spatially estimating soil data, however, require the crisp classification of properties whose boundaries are naturally gradual.

Alternative methods of mapping soil properties, or soil classification (Burrough et al, 1997) are, interestingly enough, predominantly fuzzy-based. Such soil mapping utilities are largely reliant upon fuzzy logic (Zhu et al, 1996; Zhu, 1997; Zhu et al, 2001) and fuzzy classification (De Gruijter and McBratney, 1988; McBratney and De Gruiter, 1992; McBratney et al, 1992; Odeh et al, 1992; Mazaheri et al, 1995; De Gruijter et al, 1997; Mazaheri et al, 1997; Bragato, 2004). The fuzzy logic-based methods utilize a soil similarity model (Zhu, 1997) to relate soil environmental conditions, available throughout an entire spatial domain, to predefined soil classes. Fuzzy classification approaches are quite similar to fuzzy logic based approaches, though they utilize fuzzy k-means classification (Bezdek and Pal, 1992) as well as kriging.

Whereas the above approaches depend upon measurements of a number of different soil properties, expert knowledge is, for the most part, merely used to map these soil properties to predefined soil classes. This is a limited use of expert knowledge, a vast resource, relative to the soil mapping method proposed herein. Rather than apply expert knowledge to determine soil classes and then interpolate these classes in a separate step, a novel approach would apply expert knowledge, in the form of fuzzy rules, to determine the location and extent of soil formations in one step. Such an algorithm would employ *fuzzy spatial reasoning*.

Fuzzy spatial reasoning is, quite simply, the application of fuzzy rules in a spatial context. Though the application of fuzzy techniques to the description of spatial objects is hardly novel (Leung, 1983; Leung, 1984; Altman, 1994; Wang and Keller, 1997; Wang and Keller, 1999; Keller and Wang, 2000; Bordogna and Chiesa, 2003; Li and Li, 2004) and the expression *fuzzy spatial reasoning* has been used to label algorithms for the fuzzy characterization of spatial concepts (i.e. *Near*) as linguistic variables (Robinson, 1990; Esterline et al, 1997; Schockaert et al, 2006), these approaches do not define logical relationships in the form of fuzzy rules.

In geohydrological terms, fuzzy spatial reasoning is a fuzzy rule-based approach to the delineation of soil formations based upon the location of known soil formations. The purpose of this methodology is to utilize soil data on a large scale, such as for the approximate location of an already-identified soil formation, in conjunction with generally applicable hydrogeological relationships, in order to identify the approximate soil composition of an entire aquifer. Aside from hydrogeological applications, this algorithm has utility in mineral potential mapping (Taylor and Steven, 1983; Carranza and Hale, 2001).

1.3.2. General and Site-specific Means of Defining Fuzzy CPT-based Soil Classification

Cone penetration tests have been applied to stratigraphy definition for over 80 years, though modern CPTs were developed in the 1950s (Begemann, 1953). A penetrometer is a cone-shaped device that is forced vertically downward into the subsurface, measuring the friction along the side of the device (sleeve friction) and the resistance experience by the device's tip (cone resistance).

Though the use of these two variables to classify soil types was first formalized by Begemann (1963), numerous methods were developed afterward (Schmertmann, 1978; Douglas and Olsen, 1981; Robertson, 1990). The resulting wealth of soil classification charts represents a variety of mappings between the CPT-based variables and soil type; little consensus is evident on the formulation of these charts. However, the soil classification charts possess at least one characteristic in common: it is assumed that the transitions between different soil types are crisp and abrupt.

Rather, in reality, soil types change gradually. The transition from silty soils to sandy soils is imprecise, and can be best modeled by fuzzy sets (Demmico, 2004; Ross et al, 2007). As such, it makes sense to fuzzify the classification of soil when using CPT-based variables.

Artificial intelligence-based techniques applied to CPT-based soil classification have used fuzzy logic (Zhang and Tumay, 1999, 2003; Titi and Tumay; 2000), and artificial neural networks (Kurup and Griffin, 2006) or both (Romo and Garcia, 2003). Nevertheless, the fuzzy approaches classify soils into non-traditional classes, and though the soil classes predicted by the artificial neural networks are more common (i.e. *sand*, *silt*, *clay*), they are also crisp and non-overlapping, contradicting the true overlapping nature of soil types (Demmico, 2004; Ross et al, 2007).

A fuzzy logic approach to soil classification from CPT variables should apply fuzzy rules relating sleeve friction and cone resistance, or some functional transformation of those variables into the soil classes sand, silt and clay. Ideally, the fuzzy rule-based method would produce overlapping soil classes, like that which is evidenced in nature (see Figure 3.8 in the chapter entitled *General and Site-specific Means of Defining Fuzzy CPT-based Soil Classification*).

However, rather than rely upon expert knowledge, a novel approach to fuzzy set and fuzzy rule definition would rely strictly upon the wealth of CPT-based data available at any given site. Where CPT tests and borehole investigations were performed at identical locations, a data set is compiled where CPT data and corresponding soil types are available. Using this known input and output data, a genetic algorithm can be applied to optimize an ideal fuzzy rule base (Herrera et al, 1995; Reyes, 2004; Hong et al, 2006) to classify soil. A comparison to another data-driven classification tool, a kernel-based fuzzy c-means classifier (Zhang and Chen, 2003), is also warranted.

1.3.3. Hydraulic Conductivity Estimation via Fuzzy Analysis of Grain Size Data

The combination of the wealth of soil data and the relative dearth of hydraulic conductivity data at sites of hydrogeological investigation has given rise to the prediction of the hydraulic conductivity by soil analysis data. A number of different formulae have been developed toward this end. The most recognizable of these formulae for predicting hydraulic conductivity are the van Genuchten formula (van Genuchten, 1980), for unsaturated soil, and Kozeny-Carman equation (Carman, 1937; Carman, 1939), which calculates hydraulic conductivity from soil pore properties (Chapuis and Aubertin, 2003).

However, the majority of prediction equations are predicated upon the correlation between hydraulic conductivity and soil grain size (Masch and Denny, 1966; El-Kadi, 1985; Mishra et al, 1989; Shepherd, 1989; Uma et al, 1989; Sen, 1992; Alyamani and Sen, 1993; Koltermann and Gorelick, 1995; Arya et al, 1999; Boadu, 2000; Hwang and Powers, 2003; Cronican and Gribb, 2004). Though in some cases the prediction accuracies are quite promising, with R^2 values as high as *0.99* (Krumbein and Monk, 1943), where unity is perfect correlation, special attention must be given to both the range of soil grain sizes for which hydraulic conductivity is predicted and the size of the data sets used to create and evaluate the predicting equations. For example, Shepherd (1989) shows that the models with the most promising results were developed and tested with small data sets.

Neural network-based attempts to predict hydraulic conductivity from soil data are also prevalent (Tamari et al, 1996; Schaap and Leij, 1998; Schaap et al, 2001; Minasny et al, 2004; Parasuraman et al, 2006). However, these methods suffer from the same shortcoming as the regression models, above. They reduce the wealth of information available in a grain size distribution to, in many cases, a single representative variable (i.e. D_{10}, D_{50}, percent sand, percent clay). As yet, no methodology exists that utilizes a soil sample's entire grain size distribution to predict hydraulic conductivity.

The relationship between grain size and hydraulic conductivity is quite imprecise and cannot be captured by the consideration of only a handful of variables. Different texts reveal distinct mappings between soil type and hydraulic conductivity (Bear, 1979; Domenico and Schwartz, 1990). An intuitive means of capturing this imprecise relationship is with fuzzy logic. A set of fuzzy rules relating fuzzy grain size values to fuzzy hydraulic conductivity values has the potential to effectively and accurately predict hydraulic conductivity. Moreover, given that a grain size distribution is essentially a discrete cumulative distribution function and, as such, can easily be converted into a discrete probability density function, a possibilistic grain size distribution is calculable via the application of one of many probability-possibility transformations (Oussalah, 2000). An entire grain size possibility distribution, instead of one representative grain size value, can be an input value to a fuzzy rule-base relating grain size to hydraulic conductivity.

By modeling the imprecise soil-hydraulic conductivity correlation with fuzzy rules, qualitative information, such as borehole investigations, can also be viable input. Characterizing borehole descriptions of soil samples, such as, *Silt to Medium Clay Silt, trace Fine Sand*, as a fuzzy set permits the use of borehole data to estimate hydraulic conductivity.

The use of fuzzy logic to predict hydraulic conductivity has precedent (Fang and Chen, 1997; Finol and Jing, 2002; Shokir, 2003; Demmico, 2004). However, the development of a fuzzy rule base to facilitate the use of entire grain size distributions and qualitative borehole soil descriptions is novel.

1.3.4. Fuzzy Kalman Filtering of Hydraulic Conductivity and Kalman Filter Updating of Possibilistic Hydraulic Conductivity

As mentioned above, because the direct measurement of hydraulic conductivity can be costly, alternative methods of obtaining hydraulic conductivity values have been developed (Carman, 1937; Carman, 1939; Krumbein and Monk, 1943; Masch and Denny, 1966; van Genuchten, 1980; Mishra et al, 1989; Shepherd, 1989; Uma et al, 1989; Sen, 1992; Alyamani and Sen, 1993; Koltermann and Gorelick, 1995; Tamari et al, 1996; Schaap and Leij, 1998; Arya et al, 1999; Boadu, 2000; Schaap et al, 2001; Hwang and Powers, 2003; Cronican and Gribb, 2004; Minasny et al, 2004; Parasuraman et al, 2006; Ross et al, 2007). With the increasing popularity of fuzzy logic to capture expert knowledge, fuzzy set-based hydraulic conductivity values are becoming more prevalent in hydrogeological applications (Bardossy et al, 1988; Bardossy et al, 1989; Bardossy et al, 1990a,b,c; Piotrowski, 1996; Demicco, 2004; Ross et al, 2007) as is the use of fuzzy sets to describe other hydrogeological variables (Bardossy et al, 1995; Dou et al, 1995; Bagtzoglou et al, 1996; Piotrowski et al, 1996; Dou et al, 1997a,b).

The traditional mode of interpolating hydrogeological variable measurements is kriging (Matheron, 1971), ordinary kriging being the most commonly applied set of equations (Isaaks and Srivastava, 1989). However, aside from the developments of kriging algorithms for soft and uncertain data (Journel, 1983; Kulkarni, 1984), these equations were once only operable upon certain and precise (non-fuzzy) values. In order to accommodate data sets comprised of both crisp and fuzzy measurements, extensions of the kriging equations were developed and tested on such data sets (Bardossy et al, 1988; Bardossy et al, 1989; Diamond, 1989; Bardossy et al, 1990b,c; Piotrowski, 1996; Piotrowski et al, 1996; Bandemer and Gebhardt, 2000; Passarella et al, 2003; Pham, 2005;).

Collecting fuzzy and traditional crisp measurements in a single data set and estimating using kriging simultaneously inherently assumes that fuzzy and crisp data are of the same quality. However, it can be argued that the nature of a fuzzy measurement, especially one that is a fuzzy number provided directly by an expert, is that of an educated guess of the true value. If expert-provided fuzzy data is considered prior information, and crisp measurements updating information, a Bayesian framework seems appropriate. Attempts to merge Bayesian updating with kriging (Omre, 1987; Bandemer and Gebhardt, 2000) have failed to provide an intuitive means of combining fuzzy and crisp measurements.

A straightforward means of updating fuzzy data with crisp measurements is the fuzzy kalman filter. A number of fuzzy kalman filtering algorithms exists (Hsiao, 1997, 1999; Chen et al, 1998; Zhang and Wunsch, 2003; Matia et al, 2006), though the version most germane to this research is the application of the extension principle to the static kalman filter (Ross et al, 2006).

Consider the case where an expert provides fuzzy hydraulic conductivity values. A prior hydraulic conductivity field is created via fuzzy kriging (Bardossy et al, 1990b,c). Using the fuzzy kalman filter, crisp measurements are used to update the prior hydraulic conductivity field, creating a field of posterior hydraulic conductivity estimates that both reflects the expert's insight and respects the crisp measurements.

1.3.5. Respecting Correlation While Modeling with Possibilities

A running theme throughout this dissertation is the inappropriateness of characterizing expert knowledge with probability theory. The prevalence of applying fuzzy set theory to the expert definition of hydrogeological model parameters has also been demonstrated herein. The question of how to execute groundwater flow and transport models with fuzzy number inputs, however, still remains.

References for the application of the extension principle and fuzzy rules to the execution of mathematical models have been provided above. In addition to computational concerns, none of these methods respect the spatial correlation of the model inputs.

A more practical method of propagating fuzzy information should be based upon stochastic techniques. For example, Latin hypercube sampling (McKay et al, 1979) is a Monte Carlo method (Metropolis and Ulam, 1949) whereby random variables are sampled, the resulting realizations are passed through a mathematical model, and the model estimates are aggregated as random variables. The motivation behind an extended

Latin hypercube sampling algorithm (Zhang and Pinder, 2003) is to ensure that the realizations sampled from the random variables possess a correlation matrix approximately equal to the spatial correlation of the model inputs. By modifying this Latin hypercube sampling algorithm such that realizations are drawn from fuzzy data, a practical means of groundwater flow and transport modeling with fuzzy and possibilistic data is developed.

The fuzzy-numerical simulation algorithm (Chanas and Nowakowski, 1988) permits the repeated sampling of single values from fuzzy sets or possibility distributions such that mathematical models do not need to be extended to produce fuzzy outputs. Executing this algorithm M times produces M realizations of a particular model input parameter, such as hydraulic conductivity. Applying certain steps from Latin hypercube sampling ensures that this realization matrix respects the spatial correlation of hydraulic conductivity.

Each of the M crisp realizations can be used as input to groundwater flow and transport model equations in a deterministic fashion. This produces M realizations of the desired model output, such as concentration, throughout the spatial domain. A detailed step-by-step explanation of this algorithm is provided in Chapter 7. At each location, a concentration random variable can easily be calculated. However, because the model inputs are fuzzy and/or possibilistic, the uncertain concentration estimates should be modeled as possibility distributions.

A number of possibility-probability transformations are available in the literature (Jumarie, 1995; Oussalah, 2000; Klir, 2006). For the most part, though, they dictate the conversion of a continuous possibility distribution to a continuous probability distribution (or vice-versa). A method introduced by Dubois et al (1991, 2004) constructs a possibility distribution from empirical data. Aggregating the M concentration estimates at each location into a histrogram, the corresponding possibility distribution is readily constructed.

Thus, by coupling the theory behind Latin hypercube sampling with a novel fuzzy sampling algorithm, realizations may be drawn from a fuzzy hydraulic conductivity field such that they respect the spatial correlation of the fuzzy hydraulic conductivity field. These individual realizations may be operated upon via commonly applied numerical approximation codes to calculate concentration throughout a spatial domain. These concentrations are interpreted as possibility distributions; crisp concentration values can be deduced by defuzzification or by simply considering only the most possible concentration value at each node in the discretized spatial domain.

1.3.6. Aleatoric and Epistemic Uncertainty in Groundwater Fate and Transport Modeling

As mentioned previously, stochastic groundwater modeling is the traditional approach to propagating probabilistic uncertainty from model inputs to the models estimates such as concentration and head. However, the inappropriate use of probability theory to characterize some epistemic, or subjective, uncertainty has also been suggested above (Ganoulis, 1996; O'Hagan and Oakley, 2004). In cases where expert knowledge is most appropriately characterized by fuzzy hydrogeological variable values, two alternatives exist to propagate this fuzzy uncertainty through a model to output estimates: the application of extension principles to numerical approximations of groundwater flow and transport model equations (Dou et al, 1995; Dou et al, 1997a,b; Larue and Tyagi, 1997, 1998; Faybishenko, 2004) and the approximation of groundwater physics by a fuzzy logic-based model (Bardossy and Disse, 1993; Nedungadi et al, 1994; Bardossy et al, 1995; Bagtzoglou et al, 1996; Bardossy, 1996; Schultz and Huwe, 1997; Dou et al, 1999; Coppolla et al, 2002; Vernieuwe et al, 2002a,b; Vernieuwe et al, 2007).

Nevertheless, the fuzzy set characterization of epistemic uncertainty surrounding a model parameter is inappropriate where aleatory uncertainty also exists for the same parameter. Consider a simple case where a remediation site is subdivided into a small number of geohydrological zones, within each are available a small number of hydraulic conductivity measurements. The mean and variance of each zone's measurement data sets can be used to define a representative random variable for that zone.

However, in reality each of the hydraulic conductivity measurements are inherently epistemically uncertain. An expert can define such epistemic uncertainty for each measurement by specifying an interval about the measured value in which the true value is expected to lie. In a given zone, the resulting set of uncertain measurements gives rise to a random set, rather than a random variable. The random set characterizes both the aleatory uncertainty resulting from the stochasticity of hydraulic conductivity within the zone as well as the epistemic uncertainty about the accuracy of the individual measurements.

Because the extension of a groundwater flow and transport model to random sets (Dubois and Prade, 1991b) is computationally cumbersome, the most reasonable protocol for propogating the aleatory and epistemic uncertainty conveyed by the random sets involves approximating each random set (a set of non-nested focal elements) by a possibility distribution (a set of nested focal elements) (Dubois and Prade, 1990). The extension of groundwater flow and transport model equations to operate on possibility distributions is much less computationally intensive than the random set extension. Resulting model estimates can be represented as possibility distributions, upper and lower probability bounds, or defuzzified (Klir and Yuan, 1995) crisp values.

References

Aitken, M., Butler, G., Lemmon, D., Saindon, E., Peters, D., and Williams, G. (2004). "The Lord of the Rings: the visual effects that brought middle earth to the screen." *International Conference on Computer Graphics and Interactive Techniques*, *ACM SIGGRRAPH 2004 Course Notes*, 11.

Altman, D. (1994). "Fuzzy set theoretic approaches for handling imprecision in spatial analysis." *Journal of Geographical Information Systems*, 8(3), 271-289.

Alyamani, M.S., and Sen, Z. (1993). "Determination of hydraulic conductivity from complete grain-size distribution curves." *Ground Water*, 31(4), 551-555.

Amaral, T.G.B., Crisostomo, M.M. (2002). "Automatic helicopter motion control using fuzzy logic." *IEEE International Conference on Fuzzy Systems*, 2, 860-863.

Arya, L.M., Leij, F.J., Shouse, P.J., and van Genuchten, M.Th. (1999). "Relationship between the hydraulic conductivity function and the particle-size distribution." *Soil Science Society of America Journal*, 63, 1063-1070.

Bagtzoglou, A.C., Nedungadi, A., and Sagar, B. (1996). "A fuzzy rule-based model for flow simulation in heterogeneous media." *Proceedings of the 11st International Conference on Computational Methods in Water Resources*, 1, 629-637.

Bandemer, H., and Gebhardt, A. (2000). "Bayesian fuzzy kriging." *Fuzzy Sets and Systems*, 112, 405-418.

Bardossy, A. (1996). "The use of fuzzy rules for the description of elements of the hydrological cycle." *Ecological Modelling*, 85, 59-65.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1988). "Imprecise (fuzzy) information in geostatistics." *Mathematical Geology*, 20(4), 287-311.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). "Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy, A., Bogardi, I., and Duckstein, L. (1990a). "Fuzzy regression in hydrology." *Water Resources Research*, 26(7), 1497-1508.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990b). "Kriging with imprecise (fuzzy) variograms. I: Theory." *Mathematical Geology*, 22(1), 63-79.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990c). "Kriging with imprecise (fuzzy) variograms. II: Application." *Mathematical Geology*, 22(1), 81-94.

Bardossy, A., Bronstert, A., and Merz, B. (1995). "1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach." *Advances in Water Resources*, 18(4), 237-251.

Bardossy, A., and Disse, M. (1993). "Fuzzy rule-based models for infiltration." *Water Resources Research*, 29(2), 373-383.

Bear, J. (1979). *Hydraulics of Groundwater*, McGraw-Hill, New York.

Begemann, H.K.S. (1953). "Improved method of determining resistance to adhesion by sounding through a loose sleeve placed behind the cone." *Proceedings of the 3rd International Conference on Soil Mechanics and Foundation Engineering (ICSMFE)*, 213-217.

Begemann, H.K.S.P. (1963). "Use of static soil penetrometer in Holland." *New Zealand Engineering*, 18(2), 41-49.

Berardinis, L.A. (1991). "It washes! It rinses! It talks!" *Machine Design*, 63(18), 43-47.

Bezdek, J.C., and Pal, S.K. (1992). *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*, Institute of Electrical and Electronics Engineers, New York.

Boadu, F.K. (2000). "Hydraulic conductivity of soils from grain-size distribution: New models." *Journal of Geotechnical and Geoenvironmental Engineering*, 126(8), 739-746.

Bordogna, G., and Chiesa, S. (2003). "A fuzzy object-based data model for imperfect spatial information integrating exact objects and fields." I*nternational Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 11(1), 23-41.

Bragato, G. (2004). "Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain." *Geoderma*, 118, 1-16.

Burrough, P.A., van Gaans, P.F.M., and Hootsmans, R. (1997). "Continuous classification in soil survey: spatial correlation, confusion and boundaries." *Geoderma*, 77, 115-135.

Carman, P.C. (1937). "Fluid flow through granular beds." *Transactions, Institution of Chemical Engineers*, 15, 150-166.

Carman, P.C. (1939). “Permeability of saturated sands, soils and clays.” Journal of Agricultural Science, 29, 263-273.

Carranza, E.J.M., and Hale, M. (2001). “Geologically constrained fuzzy mapping of gold mineralization potential, Baguio District, Philippines.” *Natural Resources Research*, 10(2), 125-136.

Chanas, S., and Nowakowski, M. (1988). “Single value simulations of fuzzy variable.” *Fuzzy Sets and Systems*, 25, 43-57.

Chapellier, D. (1992). *Well logging in hydrogeology*, A.A. Balkema Publishers, Brookfield, MA.

Chapuis, R.P., and Aubertin, M. (2003). “On the used of the Kozeny-Carman equation to predict the hydraulic conductivity of soils.” *Canadian Geotechnical Journal*, 40, 616-628.

Chen, G., Xie, Q., and Shieh, L.S. (1998). “Fuzzy Kalman filtering.” *Information Sciences*, 109(1-4), 197-209.

Cooper, N. (1978). “The law of excluded middle.” *Mind*, 87(2), 161-180.

Coppola, E.A., Duckstein, L., and Davis, D. (2002). "Fuzzy rule-based methodology for estimating monthly groundwater recharge in a temperate watershed." Journal of Hydrologic Engineering, 7(4), 326-335.

Cronican, A.E., and Gribb, M.M. (2004). "Hydraulic conductivity prediction for sandy soils." *Ground Water*, 42(3), 459-464.

de Marsily, G., Delay, G., Teles, V., Schafmeister, M.T. (1998). "Some current methods to represent the heterogeneity of natural media in hydrogeology." *Hydrogeology Journal*, 6, 115-130.

De Gruijter, J.J., and McBratney, A.B. (1988). "A modified fuzzy k means for predictive classification." *Classification and Related Methods of Data Analysis*, Bock, H.H., ed., Elsevier, Amsterdam.

De Gruijter, J.J., Walvoort, D.J.J, and van Gaans, P.F.M. (1997). "Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models." *Geoderma*, 77, 169-195.

Demmico, R.V. (2004). "Fuzzy logic and earth science: An overview." *Fuzzy Logic in Geology*, Demmico, R.V., and Klir, G.J., eds., Elsevier, Amsterdam.

Demmico, R.V., and Klir, G. (2004). *Fuzzy Logic in Geology*, Elsevier, Amsterdam.

Dempster, A.P. (1968). “A generalization of Bayesian inference.” *Journal of the Royal Statistical Society,Series B,* 30, 205-247.

deSilva, C.J.S., and Attikiouzel, Y. (1994). “A better path to duplicating human reasoning.” IEEE Expert 9(4), 14.

Diamond, P. (1989). “Fuzzy kriging.” *Fuzzy Sets and Systems*, 33, 315-332.

Domenico, P.A., and Schwartz, F.W. (1990). *Physical and Chemical Hydrogeology*, John Wiley and Sons, New York.

Dong, W., and Shah, H.C. (1987). “Vertex method for computing functions of fuzzy variables.” *Fuzzy Sets and Systems*, 24, 65-78.

Dou, C., Woldt, W., and Bogardi, I. (1999). “Fuzzy rule-based approach to describe solute transport in the unsaturated zone.” *Journal of Hydrology*, 220, 74-85.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1995). “Steady state groundwater flow simulation with imprecise parameters.” *Water Resources Research*, 31(11), 2709-2719.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1997a). “Numerical solute transport simulation using fuzzy sets approach.” *Journal of Contaminant Hydrology*, 27, 107-126.

Dou, C., Woldt, W., Dahab, M., and Bogardi, I. (1997b). “Transient ground-water flow simulation using a fuzzy set approach.” *Ground Water*, 35(2), 205-215.

Douglas, B.J., and Olsen, R.S. (1981). “Soil classification using electric cone penetrometer.” *Proceedings of a Session at the ASCE National Convention: Cone Penetration Testing and Experience. ASCE*, 209-227.

Druschel, B., Ozbek, M.M., and Pinder, G.F. (2006). “An application of Dempster-Shafer theory to hydraulic conductivity.” *Proceedings of the 16th International Conference on Computational Methods in Water Resources*, Copenhagen, Denmark.

Dubois, D., Foulloy, L., Mauris, G., and Prade, H. (2004). “Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities.” *Reliable Computation*, 10, 273-297.

Dubois, D., and Prade, H. (1990). “Consonant approximations of belief functions.” *International Journal of Approximate Reas*oning, 4, 419-449.

Dubois, D., and Prade, H. (1991a). "Basic issues on fuzzy rules and their application to fuzzy control, *Proceedings IJCAI-91 Workshop on Fuzzy Control*, 5-17.

Dubois, D., and Prade, H. (1991b). "Random sets and fuzzy interval analysis." *Fuzzy Sets and Systems*, 42, 87-101.

Dubois, D., Prade, H., and Sandri, S. (1991). "On possibility/probability transformations." *Proceedings of the Fourth International Fuzzy Systems Association World Congress* (*IFSA '91*), 50-53.

El-Kadi, A. (1985). "On estimating the hydraulic properties of soil, Part 2. A new empirical equation for estimating hydraulic conductivity for sands." *Advances in Water Resources*, 8, 148-153.

Elkan, C. (1994). "The paradoxical success of fuzzy logic." IEEE Expert, 9(4), 3-8.

Esterline, A.C., Dozier, G., and Homaifar, A. (1997). "Fuzzy spatial reasoning." *NASA URC Technical Conference*.

Fang, J.H., and Chen, H.C. (1997). “Fuzzy modeling and the prediction of porosity and permeability from the compositional and textural attributes of sandstone.” *Journal of Petroleum Geology*, 20(2), 185-204.

Faybishenko, B.A. (2004). “Introduction to modeling of hydrogeologic systems using fuzzy differential equation.” *Fuzziness and Soft Computing*, 142, 267-284.

Ferson, S., and Ginzburg, L.R. (1996). “Different methods are needed to propagate ignorance and variability.” Reliability Engineering and System Safety, 54, 133-144.

Finol, J., and Jing, X.-D.D. (2002). “Permeability prediction in shaly formations: The fuzzy modeling approach.” *Geophysics*, 67(3), 817-829.

Freeze, R.A. (1975). “A stochastic-conceptual analysis of one dimensional groundwater flow in nonuniform homogeneous media, *Water Resources Research*, 11(5), 725-741.

Ganoulis, J. (1996). “Modeling hydrologic phenomena.” *Revue des Sciences de L’Eau*, 9(4), 421-434.

Garcia, O.N. (1994). “Known concerns about fuzzy logic.” IEEE Expert, 9(4), 22-24.

Gelhar, L.W. (1986). “Stochastic subsurface hydrology from theory to applications.” *Water Resources Research*, 22(9), 135S-145S.

Gelhar, L.W. (1993). *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, NJ.

Helton, J.C., and Burmaster, D.E. (1996). “Guest Editorial: Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems.” *Reliability Engineering and System Safety*, 54(2-3), 91-94.

Helton, J.C., Davis, F.J., and Johnson, J.D. (2000a). “Characterization of stochastic uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant.” *Reliability Engineering and System Safety*, 69, 167-189.

Helton, J.C., Martell, M.-A., and Tierney, M.S. (2000b). “Characterization of subjective uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant.” *Reliability Engineering and System Safety*, 69, 191-204.

Helton, J.C., and Oberkampf, W.L. (2004). “Alternative representations of epistemic uncertainty.” *Reliability Engineering and System Safety*, 85, 1-10.

Herrera, F., Lozano, M., and Verdegay, J.L. (1995). "Tuning fuzzy logic controllers by genetic algorithms." *International Journal of Approximate Reasoning*, 12, 299-315.

Højberg, A.L., and Refsgaard, J.C. (2005). "Model uncertainty – parameter uncertainty versus conceptual models." *Water Science Technology*, 52(6), 177-186.

Hong, T.P., Chen, C.H., Wu, Y.L. and Lee, Y.C. (2006). "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions." *Soft Computing*, 10, 1091-1101.

Hsiao, C.-Y. (1997). "Analysis and design of a fuzzy Kalman filtering algorithm." *Journal of Control Systems and Technology*, 5(2), 75-86.

Hsiao, C.-Y. (1999). "Modified gain fuzzy Kalman filtering algorithm." *IEEE Transactions on Systems, Man and Cybernetics, Series C*, 42(2), 363-368.

Hurley, P. (2006). *A Concise Introduction to Logic*, Thomson Wadsworth Publishing, California.

Hwang, S.I., Powers, S.E. (2003). "Using particle-size distribution models to estimate soil hydraulic properties." *Soil Science Society of America Journal*, 67, 1103-1112.

Issaks E.H., and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Joslyn, C., and J. M. Booker (2004), Generalized information theory for engineering modeling and simulation, in *Engineering Design and Reliability Handbook*, edited by E. Nikolaidis, D. Ghiocel, and S. Singhal, pp. 9-1 – 9-40, CRC Press, New York.

Joslyn, C., and S. Ferson (2004), Approximate representations of random intervals for hybrid uncertainty quantification in engineering modeling, in *Sensitivity Analysis of Model Output (SAM2004)*, edited by K. M. Hanson, and F. M. Hemez, pp. 453-469, Los Alamos National Laboratory.

Joslyn, C., and V. Kreinovich (2005), Convergence properties of an interval probabilistic approach to system reliability estimation, *Int. J. Gen. Syst*., 34(4), 465-482.

Journel, A.G. (1983). "Nonparametric-estimation of spatial distributions." *Journal of the International Association for Mathematical Geology*, 15(3), 445-468.

Jumarie, G. (1995). "Further results on possibility-probability conversion via relative information and informational invariance." *Cybernetics and Systems*, 26(1), 111-128.

Keller, J.M., and Wang, X. (2000). "A fuzzy rule-based approach to scene description involving spatial relationships." *Computer Vision and Image Understanding*, 80, 21-41.

Khanbagi, M., and Malhame, R.P. (1994). "Reducing travel energy costs for a subway train via fuzzy logic controls." *Proceedings of the 1994 IEEE International Symposium on Intelligent Control*, 99-104.

Klir, G.J. (1989). "Is there more to uncertainty than some probability theorists might have us believe?" *International Journal of General Systems*, 15, 347-378.

Klir, G.J. (2006). *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley & Sons, Hoboken, NJ.

Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River.

Koltermann, C.E., and Gorelick, S.M. (1995). "Fractional packing model for hydraulic conductivity derived from sediment mixtures." *Water Resources Research*, 31(12), 3283-3297.

Kosko, B. (1993). *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York.

Krumbein, W.C., and Monk, G.D. (1943). "Permeability as a function of the size parameters of sedimentary particles." *American Institute of Mining and Metallurgical Engineers Technical Publication*, 1492, 153-163.

Kulkarni, R.B. (1984). "Bayesian kriging in geotechnical problems." *Geostatistics for Natural Resources Characterizations*, Verly, G., David, M., Journel, A.G., and Marechal, A., eds., Reidel, Dordrecht, The Netherlands.

Kurup, P.U., and Griffin, E.P. (2006). "Prediction of soil composition from CPT data using general regression neural network." *Journal of Computing in Civil Engineering*, 20(4), 281-289.

Larue, J., and Tyagi, A.K. (1997). "Groundwater simulation using fuzzy numbers." *Groundwater: An Endangered Resource*, Findikakis, A.N., and Stauffer, F., eds., ASCE, New York.

Larue, J., and Tyagi, A.K. (1998). "Fate/transport modeling using fuzzy numbers." *After the Rain Has Fallen: Ground Water Management Symposium*, Smith, J.W., ed., ASCE, Reston, VA.

Leung, Y. (1983). “Fuzzy-sets approach to spatial-analysis and planning – A nontechnical evaluation.” *Geografiska Annaler Series B – Human Geography*, 65(2), 65-75.

Leung, Y. (1984). “Basic issues of fuzzy sets theoretic spatial analysis.” *24th European Congress of the Regional Science Association*, Milan, Italy.

Li, Y., and Li, S. (2004). “A fuzzy set theoretic approach to approximate spatial reasoning.” *IEEE Transactions on Fuzzy Systems*, 12(6), 745-754.

Malinowski, G. (2001). “Many-Valued Logics.” *The Blackwell Guide to Philosophical Logic*, Goble, L., ed.,. Blackwell.

Mamdani, E.H., and Assilian, S. (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." *International Journal of Man-Machine Studies*, 7(1), 1-13.

Masch, F.D., and Denny, K.J. (1966). “Grain size distribution and its effect on the permeability of unconsolidated sands.” *Water Resources Research*, 2(4), 665-677.

Matheron, G. (1971). “The theory of regionalized variables and its applications.” *Les Cahiers du Centre de Morphologie Mathematique de Fontainbleau*, 5, Ecole Nationale Supérieure des Mines de Paris.

Mathon, B., Ozbek, M.M., and Pinder, G.F. (2008) “Dempster-Shafer theory applied to uncertainty surrounding hydraulic conductivity.” *Mathematical Geology*, *in press*.

Matia, F., Jimenez, A., Al-Hadithi, B.M., Rodriguez-Losada, D., and Galan, R. (2006). “The fuzzy Kalman filter: State estimation using possibilistic techniques.” *Fuzzy Sets and Systems*, 157, 2145-2170.

Mazaheri, S.A., Koppi, A.J., and McBratney, A.B. (1995). “A fuzzy allocation scheme for the Australian Great Soil Groups classification system.” *European Journal of Soil Science*, 46, 601-612.

Mazaheri, S.A., McBratney, A.B., and Koppi, A.J. (1997). “Sensitivity of memberships to attribute variation around selected centroiods and intergrades in the continuous Australian Great Soil Groups classification system.” *Geoderma*, 77, 155-168.

McBratney, A.B., and De Gruijter, J.J. (1992). “A continuum approach to soil classification by modified fuzzy *k*-means with extragrades, *Journal of Soil Science*, 43, 159-175.

McBratney, A.B., De Gruijter, J.J, Brus, D.J. (1992). “Spacial prediction and mapping of continuous soil classes.” *Geoderma*, 54, 39-64.

McKay, M.D., Conover, W.J., and Breckman, R.J. (1979). “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code.” *Technometrics*, 21, 239-245.

Metropolis, N., and Ulam, S. (1949). "The Monte Carlo method." *Journal of the American Statistical Association*, 44(247), 335–341.

Minasny, B., Homans, J.W., Harter, T., Eching, S.O., Tuli, A., and Denton, M.A. (2004). “Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data.” *Soil Science Society of America Journal*, 68, 417-429.

Mishra, S., Parker, J.C., and Singhal, N. (1989). “Estimation of soil hydraulic properties and their uncertainty from particle size distribution data.” *Journal of Hydrology,* 108, 1-18.

Moore, R. (1966). *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Nedungadi, A., Bagtzoglou, A.C., and Sagar, B. (1994). “A fuzzy methodology to solve the groundwater flow problem in the presence of uncertain hydrogeological data.” Applications of Artificial Intelligence in Engineering IX, Adey, R.A., Rzevski, G., Russell, D.W., eds., Computational Mechanics Publications, Southampton, UK.

Nilsson, B., Højberg, A.L., Refsgaard, J.C., and Troldborg, L. (2007). "Uncertainty in geological and hydrogeological data." *Hydrology and Earth System Sciences*, 11, 1551-1561.

Nobre, M.M., and Sykes, J.F. (1992). "Application of Bayesian kriging to subsurface characterization." *Canadian Geotechnical Journal*, 29, 589-598.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J. (1992). "Fuzzy-*c*-means and kriging for mapping soil as a continuous system." *Soil Science Society of America Journal*, 56, 1848-1852.

Odeh, I.O.A., Todd, A.J., and Triantafilis, J. (2003). "Spatial prediction of particle size fractions as compositional data." *Soil Science*, 168(7), 501-515.

O'Hagan, A., and Oakley, J.E. (2004). "Probability is perfect, but we can't elicit it perfectly." Reliability Engineering and Risk Assessment, 85, 239-248.

Omre, H. (1987). "Bayesian kriging—Merging observations and qualified guesses in kriging." Mathematical Geology, 19(1), 25-39.

Oussalah, M. (2000). “On the probability/possibility transformations: A comparative analysis.” *International Journal of General Systems*, 29(5), 671-718.

Ozbek, M.M., and Pinder, G.F. (2006). “Non-probabilistic uncertainty in subsurface hydrology and its applications: An overview.” *Water, Air, and Soil Pollution: Focus*, 6, 35-46.

Parasuraman, K., Elshorbagy, A., and Si, B.C. (2006). “Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles.” *Soil Science Society of America Journal*, 70, 1851-1859.

Passarella, G., Vurro, M., D’Agostino, V., and Barcelona, M.J. (2003). “Cokriging optimization of monitoring network configuration based on fuzzy and non-fuzzy variogram evaluation.” *Environmental Monitoring and Assessment*, 82, 1-21.

Pelletier, F.J. (1994). “Fuzzy logic: A misplaced appeal.” IEEE Expert, 9(4), 29-31.

Pham, T.D. (2005). “Integration of fuzzy and geostatistical models for estimating

Piotrowski, J.A. (1996). “Estimation of hydrogeological parameters for groundwater modeling with fuzzy geostatistics: Closer to nature?” *Calibration and Reliability in Groundwater Modelling*, 237, 511-520.

Piotrowski, J.A., Bartels, F., Salski, A., and Schmidt, G. (1996). "Geostatistical regionalization of glacial aquitard thickness in northwestern Germany, based on fuzzy kriging." *Mathematical Geology*, 28(4), 437-452.

Reyes, C.A.P. (2004). *Coevolutionary fuzzy modeling*, Springer, Berlin.

Robertson, P.K. (1990). "Soil classification using the cone penetration test." *Canadian Geotechnical Journal*, 27(1), 151-158.

Robinson, V.B. (1990). "Interactive machine acquisition of a fuzzy spatial relation." *Computers and Geosciences*, 16(6), 857-872.

Romo, M.P., and Garcia, S.R. (2003). "Neurofuzzy mapping of CPT values into soil dynamic properties." *Soil Dynamics and Earthquake Engineering*, 23, 473-482.

Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, John Wiley & Sons, West Sussex.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). "Fuzzy kalman filtering of hydraulic conductivity." *Computations Methods in Water Resources Conference 26^th^ International Conference (CMWRXVI)*.

Ross, J., Ozbek, M., and Pinder, G.F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*. 39(8), 765-780.

Schaap, M.G., and Leij, F.J. (1998). "Database-related accuracy and uncertainty of pedotransfer functions." *Soil Science*, 163(10), 765-779.

Schaap, M.G., Leij, F.J., van Genuchten, M.Th. (2001). "ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions." *Journal of Hydrology*, 251, 163-176.

Schmertmann, J.H. (1978). "Guidelines for cone test, performance, and design." *Federal Highway Administration*, *Report FHWA-TS-78209*.

Schockaert, S., De Cock, M., and Kerre, E.E. (2006). "Towards fuzzy spatial reasoning in geographic IR systems." *Proceedings of GIR06 (Workshop on Geographic Information Retrieval), Workshop at SIGIR2006 (29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval),* 34-36.

Schultz, K., and Huwe, B. (1997). "Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach." Journal of Hydrology, 201, 211-229.

Sen, Z. (1992). “New statistical grain-size method for evaluating the hydraulic conductivity of sand aquifers – Comment.” *Journal of Hydrology*, 130, 399-403.

Shafer, G. (1976). “*A Mathematical Theory of Evidence*.” Princeton University, Princeton.

Shepherd, R.G. (1989). “Correlations of permeability and grain size.” *Ground Water*, 27(5), 633-638.

Shokir, E.M.E.-M. (2003). “A novel model identification for the prediction of heterogeneous reservoir permeability.” *Journal of Engineering and Applied Science*, 50(6), 1197-1214.

So, A.T.P., Li, K.C., Chan, W.L. (1993). “Implementation of fuzzy logic based automatic camera focuser.” *Proceedings of the 1993 IEEE Region 10 Conference on Computer, Communication, Control and Power Engineering*, 4, 292-295.

Sudicky, E.A. (1986). “Natural gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process.” *Water Resources Research*, 22(13), 2069-2082.

Tamari, S., Wösten, J.H.M, and Ruiz-Suárez, J.C. (1996). "Testing an artificial neural network for predicting soil hydraulic conductivity." *Soil Science Society of America Journal*, 60, 1732-1741.

Taylor, R.B., and Steven, T.A. (1983). "Definition of mineral resource potential." *Economic Geology*, 78, 1268-1270.

Temples, T.J., and Waddell, M.G. (1996). "Application of petroleum geophysical well logging and sampling techniques for evaluating aquifer characteristics." *Ground Water*, 34(3), 523-531.

Titi, H.H., and Tumay, M.T. (2000). "Cone penetration tests at the NGES-Texas A&M University: Clay site." *ASCE Geotechnical Special Publication*, 93, 186-205.

Uma, K.O., Egboka, B.C.E, and Onuoha, K.M. (1989). "New statistical grain-size method for evaluating the hydraulic conductivity of sandy aquifers." *Journal of Hydrology,* 108, 343-366.

van Genuchten, M. Th. (1980). "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils." *Soil Science Society of America Journal*, 44, 892-898.

Vernieuwe, H., De Baets, B., Verhoest, N.E.C. (2002a). "Simulation of one-dimensional water movement in the unsaturated zone by means of a first order Takagi-Sugeno model." IEEE International Conference on Fuzzy Systems, 2, 1069-1074.

Vernieuwe, H., Verhoest, N.E.C., and De Baets, B. (2002b). "A Takagi-Sugeno model of flow in the unsaturated zone." *Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality (Proceedings of ModelCARE 2002)*, 277, 79-85.

Vernieuwe, H., Verhoest, N.E.C., de Baets, B., Hoeben, R., and de Troch, F.P. (2007). "Cluster-based fuzzy models for groundwater flow in the unsaturated zone." *Advances in Water Resources*, 30, 701-714.

Wang, X., and Keller, J.M. (1997). "Fuzzy surroundedness." *IEEE International Conference on Fuzzy Systems*, 2, 1173-1178.

Wang, X., and Keller, J.M. (1999). "Human-based spatial relationship generalization through neural/fuzzy approaches." *Fuzzy Sets and Systems*, 101(1), 5-20.

Wang, Z., and Klir, G.J. (1992). Fuzzy Measure Theory, Springer, New York.

Winter, C.L. (2004). "Stochastic hydrology: practical alternatives exist." *Stochastic Environmental Research and Risk Assessment*, 18(4), 271-273.

Yager, R. (2004). "Cumulative distribution functions from dempster-shafer belief structures." *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 34(5), 2080-2087.

Zadeh, L. (1965). "Fuzzy sets", *Information Control*, 8, 338-353.

Zadeh, L. (1973). "Outline of a new approach to the analysis of complex systems and decision processes." *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(1), 28-44.

Zadeh, L. (1975). "The concept of a linguistic variable and its application to approximate reasoning – I." *Information Sciences*, 8, 199-249.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1(1), 3-28.

Zadeh, L. (1992). "The calculus of fuzzy if-then rules." *Artificial Intelligence Expert*, 7(3), 23-27.

Zhang, D.Q. and Chen, S.C. (2003). "Clustering incomplete data using kernel-based fuzzy C-means algorithm." *Neural Processing Letters*, 18(3), 155-162.

Zhang, Y., and Pinder, G. (2003). "Latin hypercube sample selection strategy for correlated random hydraulic conductivity fields." *Water Resources Research*, 39(8), doi:10.1029/2002WR001822.

Zhang, Z., and Tumay, M. (1999). "Statistical to fuzzy approach toward CPT soil classification." *Journal of Geotechnical and Geoenvironmental Engineering*, 125(3), 179-186.

Zhang, Z., and Tumay, M.T. (2003). "Non-traditional approaches in soil classification derived from the cone penetration test." *ASCE Geotechnical Special Publication*, 121, 101-149.

Zhang, N., and Wunsch, D.C. (2003). "An extended Kalman filter (EFK) approach on fuzzy system optimization problem." *The 12th IEEE Internation Conference on Fuzzy Systems, FUZZ '03*, 2, 1465-1470.

Zhu, A.-X. (1997). "A similarity model for representing soil spatial information." *Geoderma*, 77, 217-242.

Zhu, A.-X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. (2001). "Soil mapping using GIS, expert knowledge, and fuzzy logic." *Soil Science Society of America Journal*, 65, 1463-1472.

Zhu, A.-X., Band, L.E., Dutton, B., and Nimlos, T.J. (1996). “Automated soil inference under fuzzy logic.” *Ecological Modelling*, 90, 123-145.

missing multivariate observations.” *WSEAS Transactions on Systems*, 4(4), 233-237.

Zong, Q., Wang, Z., Yue, Y., and Lei, X. (2000). “The design of multi-objective fuzzy control algorithm in elevator group control system.” *Proceedings of the World Congress on Intelligent Control and Automation (WCICA),* 3, 2009-2012.

2: FUZZY SPATIAL REASONING FOR RESERVOIR CHARACTERIZATION

While both seismic and geologic data are utilized in the process of groundwater reservoir characterization, expert knowledge is, for the most part, a heretofore untapped resource. Such knowledge can offer insight into the extent and nature of soil formations, based upon both a geohydrologic understanding of a particular reservoir and a general knowledge of depositional processes. This paper introduces a fuzzy spatial reasoning approach employing expert knowledge for reservoir characterization. The discussed algorithm produces possibility distributions identifying the estimated location of soil formations. Relevant applications demonstrate the utility of the proposed strategy.

2.1. Introduction

Expert insight provides an intuitive means to describe the geologic structure of an aquifer. Historically, the use of expert insight in groundwater reservoir definition has been hampered by a lack of both formalism and structure.

With the introduction of fuzzy sets and fuzzy logic (Zadeh, 1965), however, the utilization of expert knowledge and reasoning has become formalized. Fuzzy sets provide a means by which imprecise linguistic variables, which are used in human discourse, can be characterized. In addition fuzzy rules can model human reasoning with conditional statements.

Applications of fuzzy set theory and fuzzy logic to hydrogeological problems have been documented (McBratney and Odeh, 1997). Most relevant to this article are those applicable to reservoir characterization and soil mapping. In those instances fuzzy sets and fuzzy logic have been applied to the definition of soil regions in the subsurface based upon the use of available hard geologic data (McBratney and De Gruijter, 1992; Odeh et al, 1992; Zhu et al, 1996; De Gruijter et al, 1997; Zhu et al, 2001; Sunila and Horttanainen, 2003; Triantifilis et al, 2003; Allen et al, 2007).

Where measured data is scarce, expert knowledge may be available. Appropriate expert knowledge may include both site-specific and general geologic and hydrologic information. It is presumed, herein, that knowledge is available regarding the relative locations and extents of soil regions. Such data can be quantified using fuzzy logic through the creation of rules aimed at identifying the location of a soil region based upon the known location of another soil region, or a known history of depositional processes. This formalism is referred to as *fuzzy spatial reasoning*.

The notion of applying fuzzy techniques to describe spatial objects has been reported in (Leung, 1983; Leung, 1984; Altman, 1994; Wang and Keller, 1997; Wang and Keller, 1999; Bordogna and Chiesa, 2003; Li and Li, 2004). In fact, fuzzy rules have been developed using arguments similar to those proposed in this paper (Keller and Wang, 2000) though they were applied to automated scene description in a given image, i.e. a photograph. Moreover, the term "fuzzy spatial reasoning" has been applied to at least two different algorithms (Robinson, 1990; Esterline et al, 1997; Schockaert et al, 2006). Herein, however, fuzzy spatial reasoning refers to the use of fuzzy rules to relate variables that imply distance or direction; the goal is to define the location and extent of objects, such as the areal extent of distinct lithologic units.

The paper is organized as follows. A brief discussion of fuzzy rules and how a fuzzy rule can have spatial meaning is presented in Section 2.2. In Section 2.3, fuzzy spatial reasoning is explained using very basic examples. Applications to hydrogeological reservoir characterization are introduced in Section 2.4. A summary of the important points introduced herein concludes the paper.

2.2. Multi-Dimensional Fuzzy Inference

Fuzzy reasoning is the process by which human logic, expressed in the form of language and thereby inherently imprecise, is formalized in a mathematical framework via fuzzy rules (Klir and Yuan, 1995). In this work the fuzzy-reasoning tools include the standard min and max operators for capturing the concepts of the conjunction `and' and disjunction `or' respectively and the Mamdani-type fuzzy inference formalism is used.

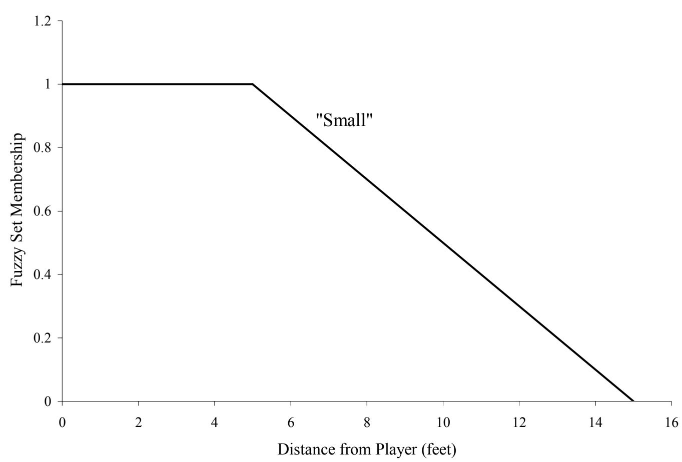
Fuzzy rules are comprised of two components: 1) conditional statements relating at least one independent variable to at least one dependent variable, and 2) fuzzy sets characterizing the values of the variables appearing in the conditional statements. For example, the conditional statement,

IF X is A THEN Y is B,

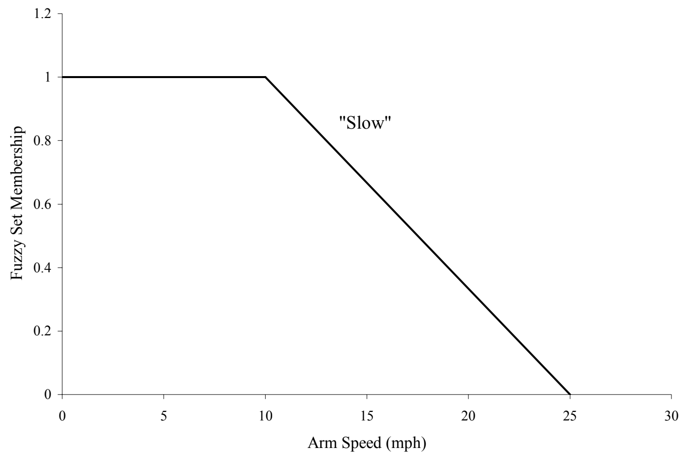
is a fuzzy rule as long as *A* is an antecedent fuzzy set defined by a membership function *A(x)*, $x \subseteq X$, and *B* is a consequent fuzzy set defined by the membership function *B(y)*, $y \subseteq Y$. A simple example of a fuzzy rule relates the distance one is standing from a baseball player to the speed with which one should throw the ball:

IF Distance from Player is Small THEN Arm Speed is Slow.

In this fuzzy rule, the independent antecedent and dependent consequence variables are *distance from player* and *arm speed*, respectively. The fuzzy sets are *Small* (Figure 2.1A) and *Slow* (Figure 2.1B). Employing fuzzy reasoning one can input a fuzzy distance and determine, via defuzzification (Klir and Yuan, 1995), a crisp speed value with which one's arm should hurl the ball.



A



B

Figure 2.1. Using the fuzzy sets A) "Small" and B) "Slow" a fuzzy rule can estimate the desired arm speed for a given distance between two baseball players

The independent variable *distance from player*, in the preceding example, implies a spatial component to the fuzzy rule. In this work, the fuzzy rules relating variables implying distance or direction are the basis for our fuzzy spatial reasoning.

2.3. Fuzzy Spatial Reasoning

2.3.1. One Dimension

In the above, it is implied that, in the context of this paper, the purpose of fuzzy spatial reasoning is to delineate the approximate location of an object relative to another object. The key to accomplishing this task is working with adjectives that define distance or direction, such as *near*, *far*, *about 20 feet*, *northeast*, *southwest*, *left*, *right*, etc. These adjectives are relative, which means that they can only provide the approximate, or fuzzy, location of an object relative to another object, the location of which is known.

Consider the statement,

The bike is near.

Unless one knows a reference object to which the *bike is near*, as assumed in Altman's (1994) fuzzy spatial relations, this statement is meaningless. A complete spatial relation is

The bike is near the tree.

Knowing the tree's location, as well as a fuzzy definition of the adjective *near*, it is possible for one to deduce the approximate distance to the bike. Consider the one-dimensional example in Figure 2.2, where the tree's location is known exactly (Figure 2.2A) and the fuzzy characterization of the adjective *near* (Figure 2.2B) is defined. Using the Euclidean distance and fuzzy arithmetic (Klir and Yuan, 1995), the approximate location of the bike can be calculated as a possibility distribution (Zadeh, 1978). The most commonly applied tool of fuzzy arithmetic is the extension principle (Klir and Yuan, 1995),

$$B(y) = \sup_{x|y=f(x)} A(x) \quad . \tag{2-1}$$

which permits the use of fuzzy numbers as arguments in otherwise traditional equations.

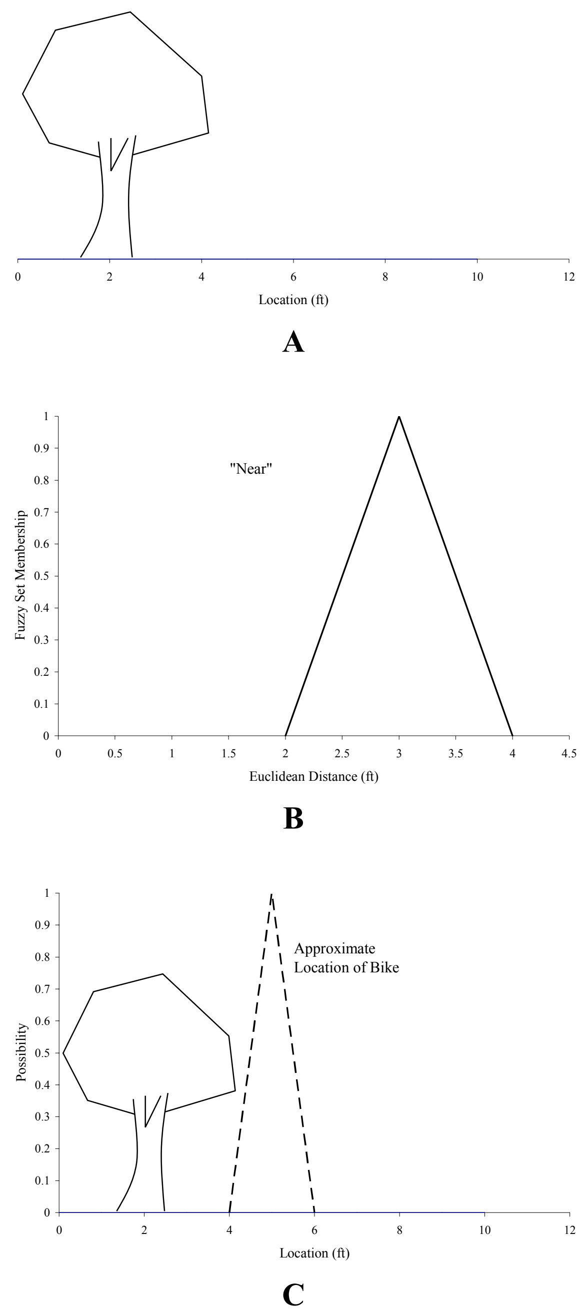


Figure 2.2. Given the location of the tree (A) and the characterization of the linguistic variable "Near" (B), knowing that the bike is near the tree can provide insight into an approximate quantitative location of the bike (C)

The principle in Equation **Error! Reference source not found.** dictates for each dependent variable value *y'*, one considers all values *x* of the independent variable *X* that, when evaluated by the function *y = f(x)*, produces the value *y'*. The greatest membership value of these values *x*, such that *f(x) = y'*, in fuzzy set *A(x)* is designated as the possibility value of *y'* in possibility distribution *B(y)*. In this case, the equation is the sum of the distance of the bike from the tree, *x*, to the location of the tree, *t*, in order to determine the location of the bike, *y*:

$$y = f(x) = t + x\,. \qquad \textbf{(2-2)}$$

The fuzzy argument, quantified in Figure 2.2C, is the distance of the bike from the tree, near, denoted here as *A(x)*. Thus, the purpose of Equation (2-2) becomes the determination of the possibility distribution, *B(y)*, defining the location of the bike. Applying the extension principle to Equation (2-2) produces the possibility distribution *B(y)* in Figure 2.2C, which characterizes, for each location along the horizontal axis, the possibility that the bike is situated there. For example, inspection of the distribution in Figure 2.2C reveals that the possibility is greatest that the bike is *5 feet* from the origin and it is impossible that the bike is more than *6 feet* or less than *4 feet* from the origin.

In order to generalize fuzzy spatial relations such that they can be applied in many circumstances, they should be framed as fuzzy rules, rather than extended equations. Consider a scenario where one knows that there is a bike by a tree, but there are many trees in the area. The bike's location is contingent upon the identification of the correct tree. A statement like

IF the tree is two feet from the building THEN the bike is Near the tree,

provides a general description of the bike's location that holds true for some specified area. This statement is a fuzzy rule whose independent variable is *distance between the tree and building* and dependent variable is *distance between the bike and tree*. The antecedent fuzzy set is a crisp value *2 feet*, and the consequent fuzzy set is defined (by an appropriate agent) in Figure 2.2B.

2.3.2. Multiple Dimensions

The importance of this theory is more evident when applied to multiple spatial dimensions. However, in order to produce meaningful results, more information should be furnished in the fuzzy spatial relations. Consider the case of the bike and tree again. This time, the purpose of the rule is to approximately locate the bike's easting *and* northing.

Given the fuzzy rule, above, relating the bike and the tree, a two-dimensional possibility distribution, which delineates the approximate location of the bike, forms a ring around the tree (Figure 2.3).

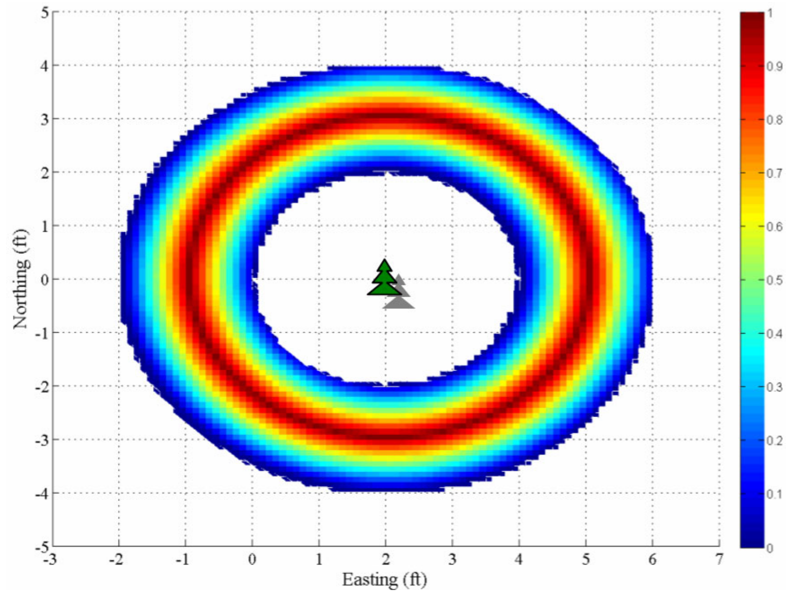


Figure 2.3. This two-dimensional ring-shaped possibility distribution provides a very imprecise idea of the true location of the bike compared to the one-dimensional version in Figure 2.2C

This region is created in the following manner. First, the spatial domain is discretized into a number of grids and a tree is found that satisfies the antecedent condition. Once this tree is found, the distance between each nodal location in the spatial domain and the tree is calculated using the Minkowski L_p-metric (Altman, 1994). In this case, p = 2, specifying conventional Euclidean distance:

$$dist((x_{tree}, y_{tree}), (x_i, y_j)) = \sqrt{(x_{tree} - x_i)^2 + (y_{tree} - y_j)^2}$$

for $i = 1,\ldots,m$ and $j = 1,\ldots,n$, where m is the number of nodes in the x direction and n is the number of nodes in the y direction. For each node, the membership value in the fuzzy set *Near*, associated with the node's distance from the tree, is assigned as that node's possibility value, $\pi_{Bike}(x_i, y_j)$, in the region defining the possible location of the bike (Figure 2.3).

By including just one more piece of information, this possibility distribution can be made considerably more insightful. For instance, by including direction in the fuzzy rule,

IF the tree is two feet from the building THEN the bike is Near to and Northwest of the tree,

along with a fuzzy set representing the term *Northwest* (Figure 2.4A), a more precise location of the bike is defined. The determination of this more informative possibilistic region is quite similar to that of the ring-shaped regions as described above. However, two regions must first be delineated. The ring-shaped region, describing a region that is *near* to the tree is calculated as above. Another region, representing areas northwest of the tree, is calculated in a similar manner. For each nodal location, the angle (in degrees) from the northing, or azimuth, is calculated, and the membership of this azimuth in the fuzzy set *Northwest* is tabulated. The membership value is assigned to that node and represents the possibility that the node is northwest of the tree, $\pi_{Bike}(x, y)$ (Figure 2.4B). In order to determine the approximate location of the bike, the two regions, one representing areas near the tree and the other representing areas northwest of the tree, must be combined. Since the bike is deemed, by the above rule, to be both near and northwest of the tree, the ideal combination of the possibilistic regions is intersection

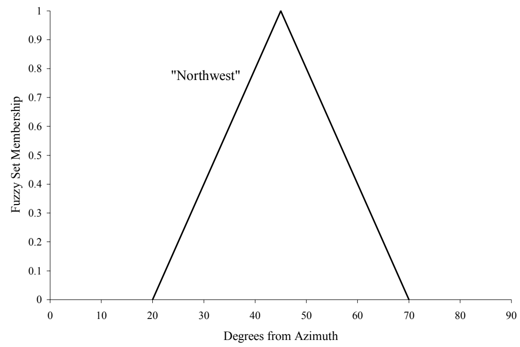
(Klir and Yuan, 1995). Intersection dictates that the minimum possibility value at each node is assigned to that node as the possibility that the bike is located there,

$$\pi_{Bike}(x_i, y_j) = \min\{\pi_{Near}(x_i, y_j), \pi_{Northwest}(x_i, y_j)\} \qquad i = 1,\ldots,m;\ j = 1,\ldots,n,$$

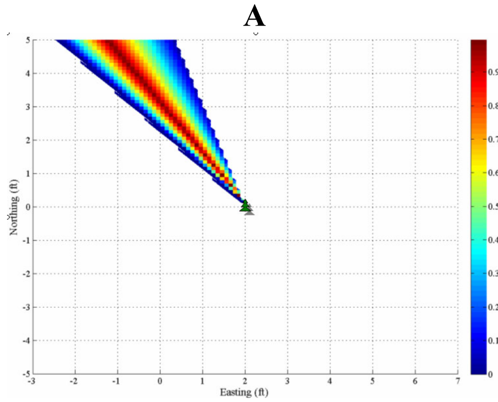
resulting in the bike's approximate location in Figure 2.4C.

2.4. Fuzzy Spatial Reasoning

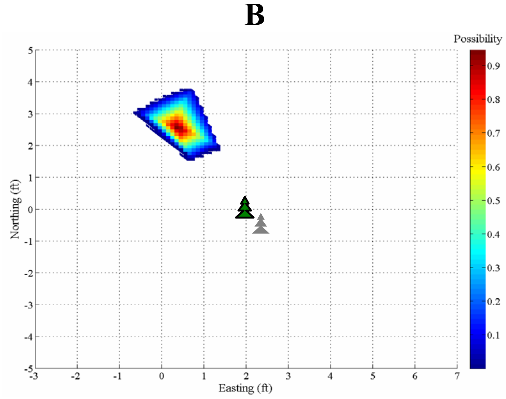
Though there are many problems to which this theory can be applied, the area of most interest to the authors is hydrogeological reservoir characterization. Specifically, we want to use fuzzy spatial reasoning to encapsulate and expert knowledge regarding the spatial relationships between different geologic formations and to express it quantitatively. We wish to show that as a complement to the Soil Land Inference Model (Zhu et al, 1996, 2001), which uses fuzzy logic to identify near-surface soil formations based upon environmental conditions, fuzzy spatial reasoning can identify the presence of one or more classes of soil for each cell in a spatial grid. Coupled with a mechanism to estimate hydraulic conductivity from soil data (Ross et al, 2007), the hydrological characteristics of a reservoir can be then identified.



A



B



C

Figure 2.4. The inclusion of additional information regarding direction (A), delineating a region that is northwest of the bike (B), to identify the bike's possible location provides a more insightful result (C) than that in Figure 2.3

2.4.1. Soil Spatial Relations

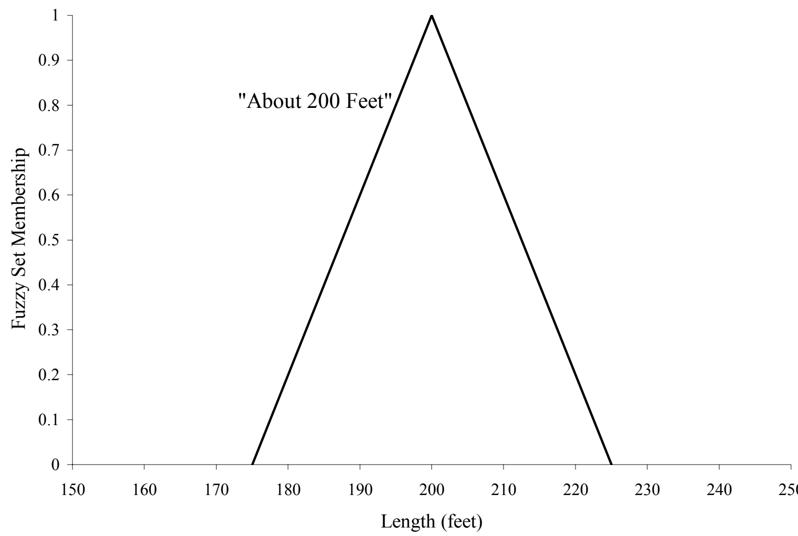
Although the deposition of soils in the subsurface is a function of many surface geologic processes, and the identification and modeling of these processes is a science unto itself, it is plausible, if not possible, to characterize expert knowledge of these processes as they are applicable to the geohydrologic composition of a site.

For instance, consider the identification of the extent of a coarse-grained soil deposit in the subsurface. Hydrologic intuition would suggest that finer-grained deposits would be carried further downstream by surface water than would coarse-grained deposits. Perhaps a fuzzy rule describing this process could read

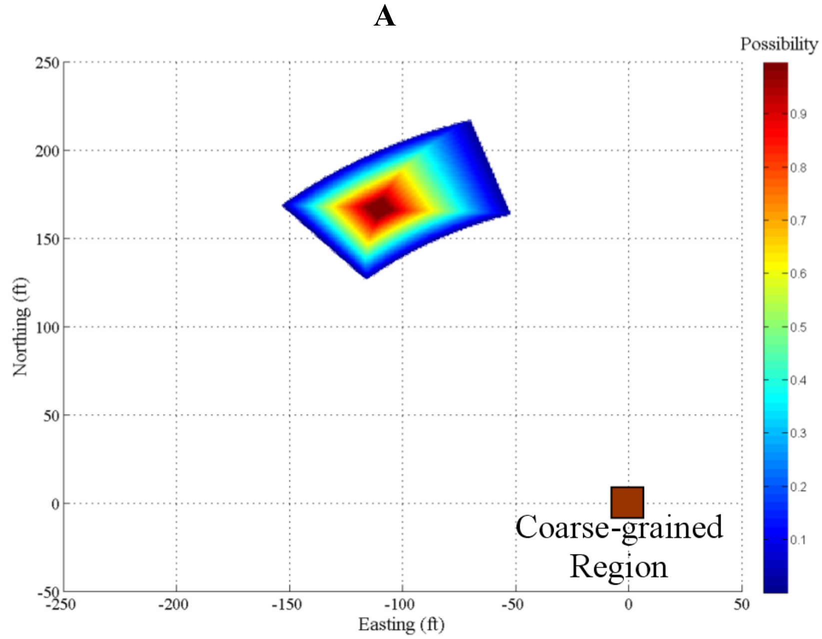
IF surface water historically flowed Northwest, THEN a fine-grained deposit should be located About 200 feet downstream of the coarse-grained deposit. **(2-3)**

Note that this rule is very precise and may be unrealistic. Nevertheless, it demonstrates the ability with which natural processes can be described by fuzzy rules. Here, the fuzzy sets are *Northwest* (Figure 2.4A) and *About 200 feet* (Figure 2.5A). Evaluation of the rule using the same procedure as above produces the possibility distribution of the fine-grained soil formation in Figure 2.5B.

Ideally, most sites will have a number of different rules stipulated to identify the locations of various soil regions. In such cases, it is quite possible that at least one grid location will have more than one possibility distribution identifying the existence of more than one soil. Where a cell grid has a vector of possibility values defining the existence of a number of different soils, Zhu et al (1996, 2001) introduced the notion of a soil similarity vector (SSV).



A



B

Figure 2.5. The characterizations of the notions "Northwest" in Figure 2.4A and "About 200" feet (A), in conjunction with the fuzzy rule in Equation (2-3), produces the approximate fine-grained soil region (B)

For instance, in addition to the rule above, another rule might define the location of a coarse-grained soil region. If this region were to overlap with the fine-grained soil regions, logically some grid locations will have two possibility values, one associated with fine-grained soil and associated with coarse-grained soil. The cell grid's SSV in this two rule example will have two possibility values. If it is ultimately desired that only one soil type be determined to exist at any one grid location, the soil corresponding to the greatest SSV will be identified as the soil at that location.

Intuitively, the geologic formations of interest are characterized by other than soil type. The occurrence of mineral deposits, for example may be a suitable application for fuzzy spatial reasoning such as presented herein. In fact, fuzzy techniques have already been applied to mapping gold mineralization potential (Carranza and Hale, 2001)

Expert knowledge-derived hydrologic regions are also realized in practical applications with multiple-point geostatistics (Strebelle, 2000). The geostatistical approach was developed to overcome the limitations of two-point geostatistics and is predicated upon the existence of training images. A training image is a Boolean representation, traditionally drawn by a geologist in order to represent structural patterns present in the subsurface, from which the geostatistics are derived, in combination with available parameter measurements. Some training images (Figure 2.6, from Caers and Zhang, 2002) contain hand-drawn regions that are very similar to that produced by fuzzy logic in Figure 2.4C. Thus, instead of relying upon an arbitrarily drawn image, a collection of fuzzy rules defining various soil type regions might be used to construct a training image.

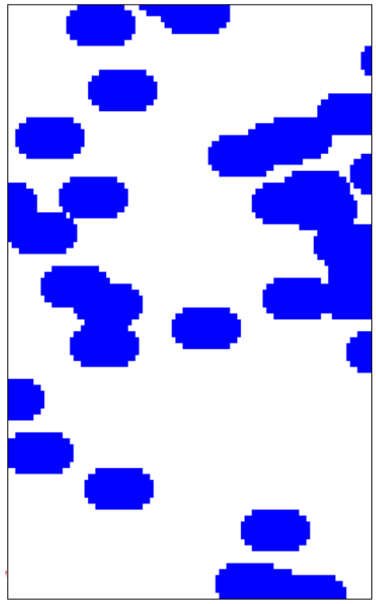


Figure 2.6. Example of a training image (from Caers and Zhang, 2002)

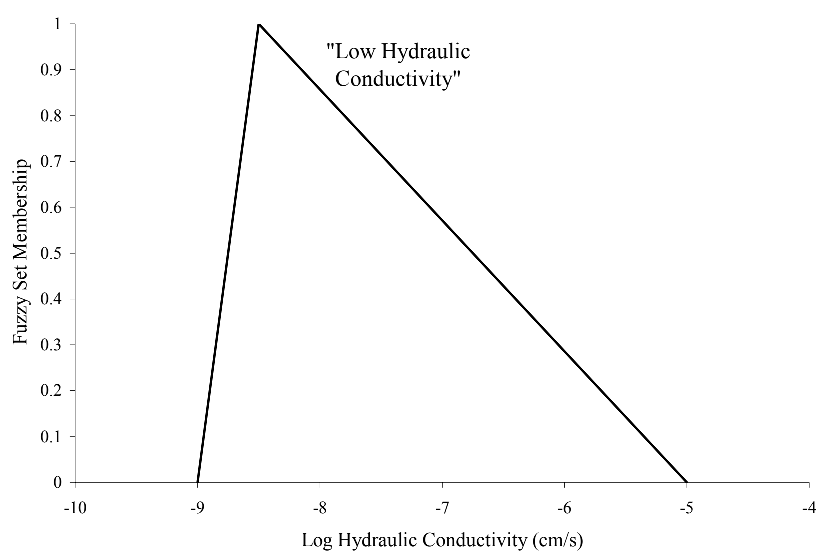
2.4.2. Hydrologic Extensions

Relationships between soil characteristics and parameters describing an aquifer are intuitive, and there are means by which hydrologic data can be estimated from soil data (Shepherd, 1989; Fogg et al, 1998; Arya et al, 1999; Boadu, 2000; Hwang and Powers, 2003; Cronican, 2004; Parasuraman, 2006;). However, it is not immediately evident how knowledge of the locations of soil formations can lead to quantitative hydrologic aquifer characterization. The question at hand is how to take advantage of the knowledge that, to a certain degree, a given cell grid is found to be occupied by one or more soil types, where the soil data is qualitative in nature only. Fuzzy logic provides an answer.

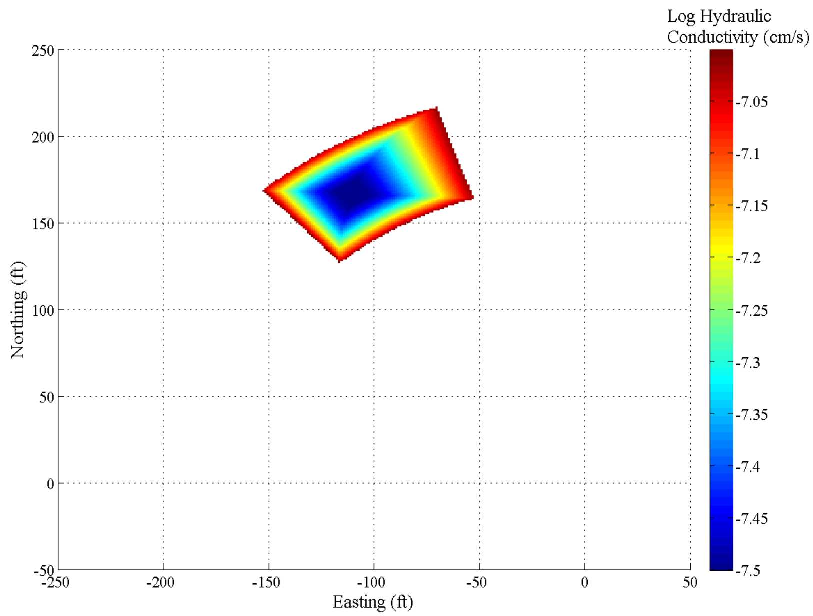
Ross et al (2007) have developed a means by which hydraulic conductivity can be estimated from knowledge of a sample's soil class. Relating fuzzy set soil classes to hydraulic conductivity, theirs is a mechanism by which the data output from fuzzy soil spatial reasoning can lead to estimates of hydraulic conductivity (*K*). Consider a rule of the form

IF soil type is Fine-Grained THEN hydraulic conductivity is Low **(2-4)**

where *low* hydraulic conductivity is defined as in Figure 2.7A. Given the possibilistic fine-grained soil region in Figure 2.5B, the possibility value at each grid node can be interpreted as the degree to which this rule pertains there. After defuzzification, the log hydraulic conductivity estimates are given in Figure 2.7B.



A



B

Figure 2.7. Using the fuzzy set in (A) defining "Low" hydraulic conductivity the fuzzy rule in Equation (2-4) produces the hydraulic conductivity estimates in (B)

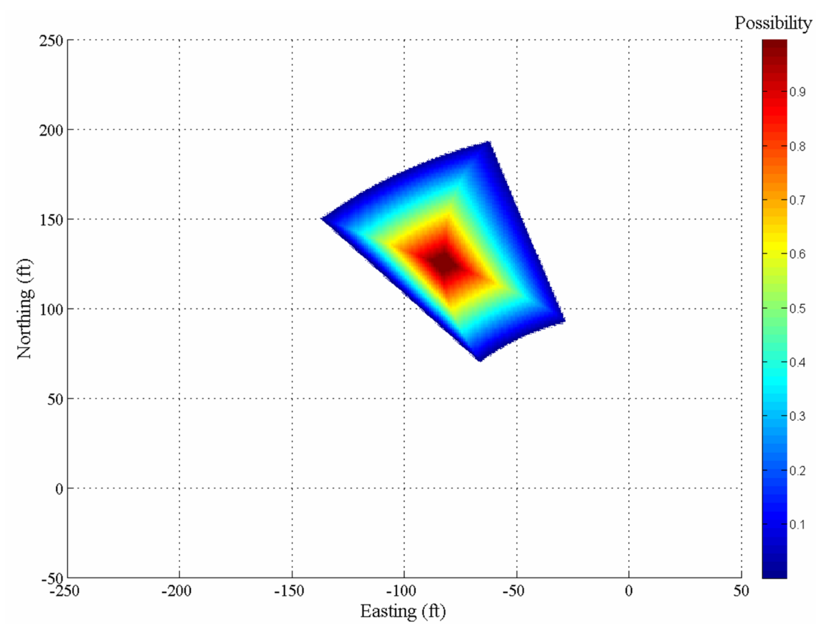


Figure 2.8. Two overlapping possibilistic regions representing fine-grained in Figure 2.5B and coarse-grained soil regions (pictured here), coupled with the fuzzy rules in Equations (2-4) and (2-5), produces estimates of hydraulic conductivity, like for that of location [E-100, N150] (*$K = 1.76 \times 10^{-5}$ cm/s*)

In a case where a grid location possessed a SSV with *n* possibility values corresponding to *n* different soil types, *n* different rules are applicable. The possibility value for each soil will then dictate the degree to which the corresponding rule will fire for that grid location. The outputs of the *n* rules are aggregated and a single estimated hydraulic conductivity value is deduced. Thus, for any one location, it is quite possible to have a rule base.

Assume that the two soil regions in Figure 2.5B and Figure 2.8 resulted from the determination and application of two rules, one identifying the locations of a fine-grained soil region and one delineating a coarse-grained soil region. At location (-100, 150), the rules found that it is possible that fine-grained soil exists to degree *0.35* and coarse-grained soil exists to degree *0.45*.

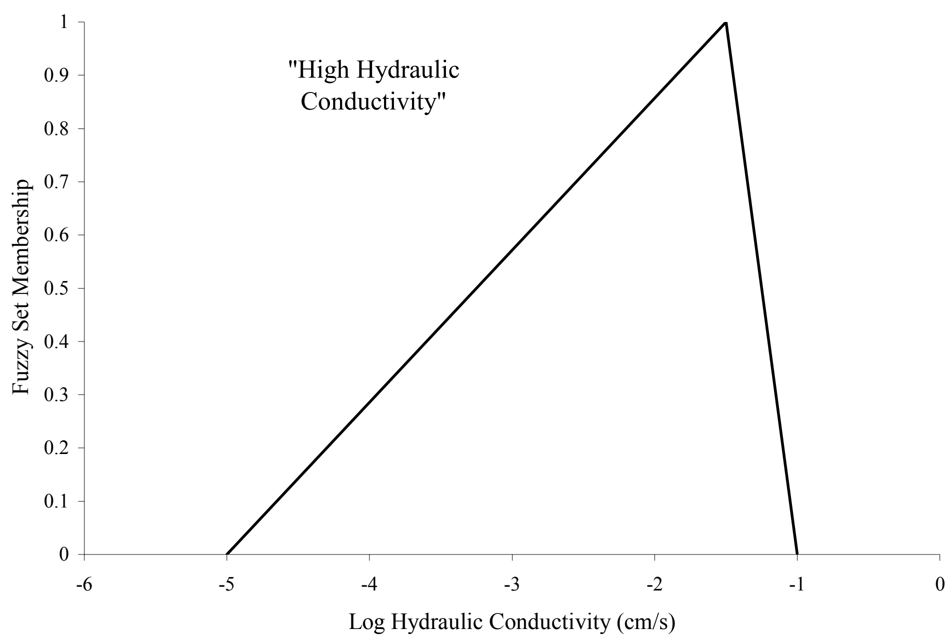


Figure 2.9. The characterization of "High" hydraulic conductivity can be used to estimate hydraulic conductivity when a component of a fuzzy rule like that in Equation (2-5)

Given the rule

IF soil type is Coarse-Grained THEN hydraulic conductivity is High, **(2-5)**

and using the *high* hydraulic conductivity fuzzy set in Figure 2.9, the evaluation of the rule base produces a hydraulic conductivity value of 1.76×10^{-5} *cm/sec* at location (-100, 150).

2.5. Conclusions

The application of fuzzy inference to the definition of spatial relations, herein referred to as fuzzy spatial reasoning, is well suited for reservoir characterization. It has been demonstrated that fuzzy sets and fuzzy logic can capture expert knowledge regarding the spatial relationship between objects as well as the relationship between hydrologic parameters and soil formations.

Arguably, the value of expert data relies greatly upon the experts who opine on the desired topic. Experience shows that it is important that the proper questions are posed to the experts in order to obtain the desired information in an efficient manner. In this vein, a considerable amount of research has been directed toward data acquisition for fuzzy system definition (Onisawa and Anzai, 1999).

The greatest value of fuzzy spatial reasoning would be realized in the application to the comprehensive modeling of the effect of glacial, aeolian and hydrologic dynamics upon geology. Again, the knowledge capture process toward this end would likely be an arduous task. Nevertheless, it is possible that a series of simple rules can approximately capture the physics of complex depositional processes. Such an endeavor is well beyond the scope of this paper, though the groundwork was certainly laid above.

References

Allen, D.M., Schuurman, N., and Zhang, Q. (2007). „Using fuzzy logic for modeling aquifer architecture." *Journal of Geographical Systems*, 9(3), 289-310.

Altman, D. (1994). "Fuzzy set theoretic approaches for handling imprecision in spatial analysis." *Journal of Geographical Information Systems*, 8(3), 271-289.

Arya, L.M., Leij, F.J., Shouse, P.J., and van Genuchten, M.Th. (1999). "Relationship between the hydraulic conductivity function and the particle-size distribution." *Soil Science Society of Americal Journal*, 63, 1063-1070.

Boadu, F.K. (2000). "Hydraulic conductivity of soils from grain-size distribution: new models." *Journal of Geotechnical and Geoenvironmental Engineering*, 126(8), 739-746.

Bordogna, G., and Chiesa, S. (2003). A fuzzy object-based data model for imperfect spatial information integrating objects and fields, International Journal of Uncertainty." *Fuzziness and Knowledge-Based Systems*, 11(1), 23-41.

Caers, J., and Zhang, T (2002). "Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models." *Integration of Outcrop and*

Modern Analog Data in Reservoir Models, Grammer, G. M., Harris, P.M., and Eberli, G. P., eds., American Association of Petroleum Geologists.

Carranza, E.J.M., and Hale, M. (2001). "Geologically constrained fuzzy mapping of gold mineralization potential, Baguio District, Philippines." *Natural Resources Research*, 10(2), 125-136.

Cronican, A.E., and Gribb, M.M. (2004). "Hydraulic conductivity prediction for sandy soils." *Ground Water*, 42(3), 459-464.

Esterline, A.C., Dozier, G., and Homaifar, A. (1997). "Fuzzy spatial reasoning." *NASA URC Technical Conference*.

Fogg, G.E., Noyes, C.D., and Carle, S.F. (1998). "Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting." *Hydrogeology Journal*, 6(1), 131-143.

de Gruijter, J.J., Walvoort, D.J.J., and van Gaans, P.F.M. (1997). "Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models." *Geoderma*, 77(2), 169-195.

Hwang, S.I., and Powers, S.E. (2003). “Using particle-size distribution models to estimate soil hydraulic properties.” *Soil Science Society of America Journal*, 67, 1103-1112.

Keller, J.M., and Wang, X. (2000). “A fuzzy rule-based approach to scene description involving spatial relationships.” *Computer Vision and Image Understanding*, 80, 21-41.

Klir, G., and Yuan, B. (1995). *Fuzzy sets and fuzzy logic*, Prentice Hall, Upper Saddle River, New Jersey.

Leung, Y. (1983). “Fuzzy-sets approach to spatial-analysis and planning – a nontechnical evaluation.” *Geografiska Annaler Series B – Human Geography*, 65(2), 65-75.

Leung, Y. (1984). “Basic issues of fuzzy sets theoretic spatial analysis.” *24th European Congress of the Regional Science Association*..

Li, Y., and Li, S. (2004). “A fuzzy sets theoretic approach to approximate spatial reasoning.” *IEEE Transactions on Fuzzy Systems*, 12(6), 745-754.

McBratney, A.B., and de Gruijter, J.J. (1992). “A continuum approach to soil classification by modified fuzzy k-means with extragrades.” *Journal of Soil Science*, 43(1), 159-175.

McBratney, A.B., and Odeh, I.O.A. (1997). "Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions." *Geoderma*, 77(2), 85-113.

Odeh, I.O.A, McBratney, A.B., and Chittleborough, D.J. (1992). "Fuzzy c-means and kriging for mapping soil as a continuous system." *Soil Science Society of America Journal*, 56, 1848-1854.

Onisawa, T., and Anzai, T. (1999) "Acquisition of intelligible fuzzy rules." *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 5, 268-273.

Parasuraman, K., Elshorbagy, A., and Si, B.C. (2006). "Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles." *Soil Science Society of America Journal*, 70, 1851-1859.

Robinson, V.B. (1990). "Interactive machine acquisition of a fuzzy spatial relation." *Computers and Geosciences*, 16(6), 857-872.

Schockaert, S., de Cock, M., and Kerre, E.E. (2006). "Towards fuzzy spatial reasoning in geographic IR systems." *Proceedings of GIR06* (*Workshop on Geographic Information Retrieval*), *Workshop at SIGIR2006* (*29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval*), 34-36.

Shepherd, R.G. (1989). "Correlations of permeability and grain size." *Ground Water*, 27(5), 633-638.

Strebelle, S. (2000). *Sequential simulation drawing structure from training images*, Ph.D. Thesis, Stanford University.

Sunila, R., and Horttanainen, P. (2003). "Fuzzy model of soil polygons for managing the imprecision." *Proceeding of Interfacing Geostatistics, GIS and Spatial Databases* (*STATGIS2003*).

Thore, P. (2006). "Accuracy and limitations in seismic modeling of reservoir." *SEG Technical Program Expanded Abstracts*, 1674-1677.

Triantifilis, J., Odeh, I.O.A., Minasny, B., and McBratney, A.B. (2003). "Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy k-means classification of EM34 data." *Environmental Modelling and Software*, 18(7), 667-680.

Wang, X., and Keller, J.M. (1997). "Fuzzy surroundedness." *IEEE International Conference on Fuzzy Systems*, 2, 1173-1178.

Wang, X., and Keller, J.M. (1999). „Human-based spatial relationship generalization through neural/fuzzy approaches." *Fuzzy Sets and Systems*, 101, 5-20.

Zadeh, L.A. (1965). "Fuzzy sets." *Information Control*, 8, 338-353.

Zadeh, L.A. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1, 3-28.

Zhu, A.X., Band, L.E., Dutton, B., and Nimlos, T.J. (1996). "Automated soil inference under fuzzy logic." *Ecological Modelling*, 90(2), 123-145.

Zhu, A.X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. (2001). "Soil mapping using GIS, expert knowledge, and fuzzy logic." *Soil Science Society of America Journal*, 65, 1463-1472.

3: GENERAL AND SITE-SPECIFIC MEANS OF DEFINING FUZZY CPT-BASED SOIL CLASSIFICATION

Though their primary use is to identify a soil's structural qualities, data from cone penetrometer tests are commonly used to delineate the soil composition in a vertical column of soil without the need of boreholes and grain size analyses. Soil classification is imprecise and traditional classification procedures often fail to produce accurate and consistent results. Two classification approaches using fuzzy set theory and fuzzy logic are presented whereby the soil composition is deduced using values of cone resistance and sleeve friction. One classification method relies upon field data to define a soil classification chart with fuzzy soil boundaries, while the other method applies fuzzy rules to model the physics governing the underlying relationship between CPT data and soil type.

3.1. Introduction

The in situ application of cone penetrometers has existed for over 80 years. The predecessor of current penetrometers, which allowed for the measurement of both cone resistance (Q_c) and sleeve friction (F_s), was developed in the early 1950s (Begemann, 1953). The application of these measured parameters to the classification of subsurface soil types was formalized shortly thereafter by Begemann (1963).

The ratio of sleeve friction to cone resistance, known as friction ratio (R_f), was also viewed as a viable classification tool (Begemann, 1963). Classification charts that followed plotted friction ratio against cone resistance (Schmertmann, 1978; Douglas and Olsen, 1981; Robertson, 1990) in order to deduce both the soil types and geotechnical characteristics of a sample.

A number of site specific approaches to soil classification based on cone penetration tests (CPT) were developed. The majority of these focused on determining soil properties, such as liquefaction and shear strength, (Romo and Garcia, 2003; Zhang and Tumay, 1999); only a few utilized CPT data for soil type classification (Zhang and Tumay, 1999; Kurup and Griffin, 2006; Mullarkey and Fenves, 1986; Vos, 1982).

Most relevant to this research are the CPT-based soil classifiers developed by Kurup and Griffin (2006) and Zhang and Tumay (1999), which produce soil classes through non-traditional transformations of cone resistance and sleeve friction data. The algorithms presented herein rely upon the most appealing aspects of these two classifiers, one offering a data-driven approach to soil classification, while the other utilizes expert logic to classify soils.

The artificial neural network (ANN) soil classification method, developed by Kurup and Griffin (2006), passed cone resistance and sleeve friction values through a general regression neural network to estimate percentages of clay, sand and silt. ANN training was predicated upon the existence of CPT soundings and grain size distribution analyses performed at identical locations. The use of a regression ANN to classify soils assumes that the process of soil classification is precise.

Zhang and Tumay (1999) accomplished classification by first redefining CPT sounding data using a soil engineering classification index *U*. Using this index fuzzy sets were defined, one for each of three distinct soil types, appropriately delineated by imprecise boundaries: Highly Probable Clayey soil, *Highly Probable Mixed soil* and *Highly Probable Sandy soil*. The membership value of a soil sample's index value $u \in U$ in the three fuzzy sets defined the composition of the soil sample. However, these classifications are not definitive declarations. Most notably, the imprecise term *Highly Probable* is never quantified, nor is the classification *Mixed Soil* explicitly defined.

CPT-based soil classification is inherently imprecise. Fuzzy set theory and fuzzy logic are two related frameworks by which imprecision can be modeled. The two soil classification algorithms presented herein rely on these two frameworks separately. The first algorithm presented applies a fuzzy kernel-based C-means classifier to develop site-specific fuzzy soil classification charts. This approach is purely data-driven. The second algorithm presented herein utilizes a genetic algorithm to develop a fuzzy logic-based classifier that defines soil composition based upon cone resistance and sleeve friction values. Unlike the first method, this approach is logic-driven rather than strictly data-driven. Borehole logs are used to evaluate the accuracy of both algorithms.

In the following section, arguments are presented for the non-generality of soil classification techniques based upon cone penetration data, accentuated by examples of the inaccuracies of common classification charts, and also for the modeling of imprecision in the soil classification process. In the third section, the data-driven fuzzy kernel-based classifier is presented along with results. In the fourth section the genetic algorithm-optimized fuzzy classifier is explained and results are provided. A discussion of the two classifiers concludes the paper.

3.2. Practical CPT-based Soil Classification

3.2.1. Conflicting Classifications

A number of distinct soil classification charts are available for use with cone resistance and sleeve friction data. As mentioned previously, Begemann (1963) devised the first classification chart by plotting sleeve friction in kiloPascals against cone resistance in MegaPascals (Figure 3.1A). More recently, Douglas and Olsen (1981) devised a classification chart using cone resistance and friction ratio data from an electrical cone penetrometer (Figure 3.1B).

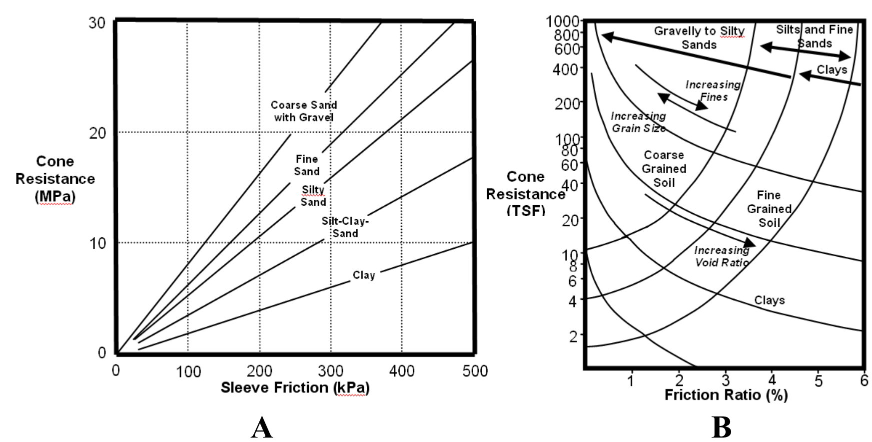


Figure 3.1. The Begemann (1963) classification chart (A) plots sleeve friction against cone resistance; and the Douglas and Olsen (1981) classification chart (B) considers the relationship between friction ratio and cone resistance, and identifies general relationships between fines content, grain size, void ratio and soil type

These two charts have been used interchangeably, suggesting the same general relationships between the independent variables cone resistance and sleeve friction and the dependent variable soil grain size. By inspection of these charts it is evident that grain size gradually increases with increasing cone resistance and decreasing sleeve friction. Nevertheless, assessment using some randomly selected cone resistance and sleeve friction data pairs reveals that the classification charts produce quite distinctly different results (Table 3.1). The classifications in Table 3.1 determined from the two charts are entirely at odds with each other.

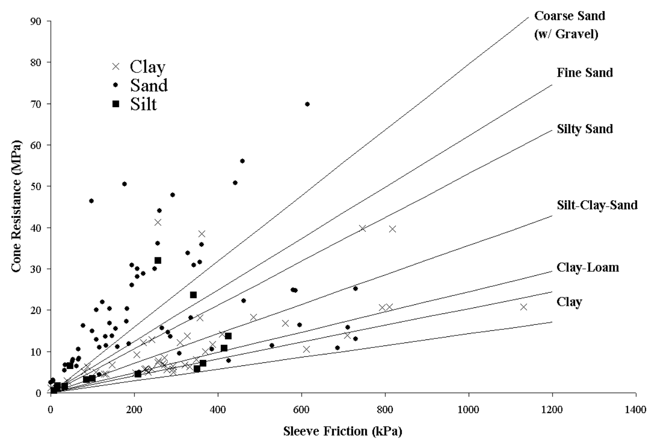
Table 3.1. Five randomly chosen CPT data (cone resistance Qc in tons per square foot and MegaPascals, sleeve friction Fs in kiloPascals and friction ratio Rf), the corresponding soil descriptions, Begemann classification and Douglas and Olsen classification. Note how neither classification chart produces consistently accurate classifications

Qc (tsf)	Qc (MPa)	Fs (kPa)	Rf (%)	Boring Log Description	Begemann Class.	Douglas/Olsen Class.
261.81	28.08	207.79	0.74	sand (90% fine sand, 10% fines)	No Classification (more coarse than coarse sand with gravel	Non-cohesive coarse-grained sand
112.78	12.10	309.65	2.56	sandy clay (70% fines, 30 fine sand)	silt-clay-sand	More coarse than coarse-grained soil
62.95	6.75	147.18	2.18	lean clay (80% fines, 20% fine sand)	Between Loam and silt-clay-sand	Non-cohesive coarse-grained sand
185.34	19.88	163.00	0.82	silty sand (55% fine sand, 45% fines)	No Classification (more coarse than than coarse sand with gravel	Non-cohesive coarse-grained sand

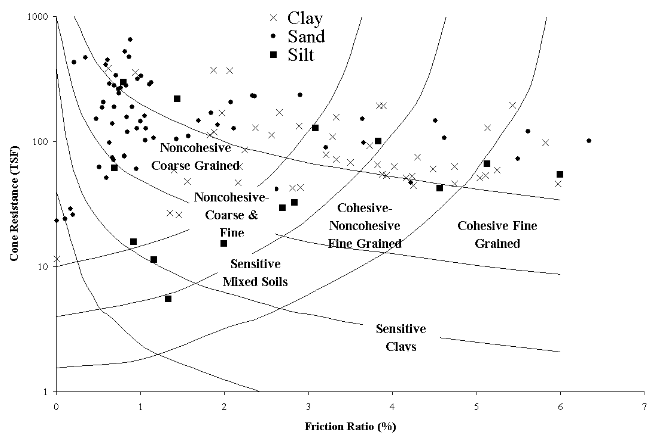
In applications where the soil classification is known for one or more CPT soundings, the most accurate classification method can be deduced. Note from Table 3.1 that Begemann's chart was more accurate than that of Douglas and Olsen in classifying a sandy clay sample as a sand-silt-clay.

In Figure 3.2A are plotted the cone resistance and sleeve friction values for soils that are predominantly (greater than *80%*) sand, silt or clay, overlain by the Begemann chart; and in Figure 3.2B are plotted the cone resistance and friction ratio values for the same soils overlain by the Douglas and Olsen chart. Note that the great majority of the sand samples are predominantly fine sand. The Begemann chart reasonably classifies clay samples and the Douglas and Olsen chart performs reasonably well with silt samples. However, neither chart provides an accurate means of soil classification overall; the Douglas and Olsen chart almost completely misclassifies this particular set of field data. Moreover, crisp classification of the same data performs poorly, classifying 119 out of 125 of the data pairs as clay soil samples.

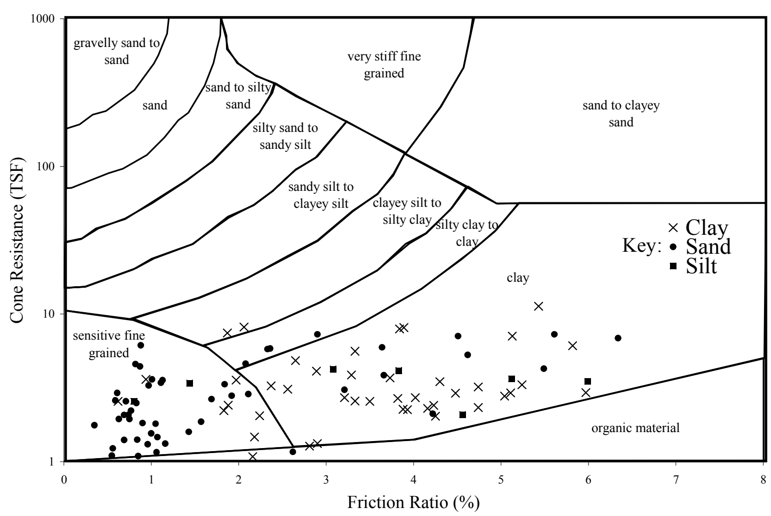
By employing more extensive data sets, improvements in soil classification charts were made by Robertson et al (1986) and Robertson (1990). For instance, consider the classification chart in Figure 3.2C, developed by Robertson (1986). It is evident that this classification chart correctly identifies the clay and silt soil samples from the data set in most cases. The Robinson classification chart is a marked improvement over older charts. Nevertheless, the sand soil samples are incorrectly classified as *sensitive fine grained* soils and 23 CPT data pairs do not lie within any of the 12 soil regions. Obviously, there is still room for improvement in soil classification with CPT data.



A



B



C

Figure 3.2. CPT data from clay, silt and sand samples plotted against (A) the Begemann chart and (B) the Douglas and Olsen chart. The Roberston chart (C) classifies more accurately.

3.2.2. Fuzzy Set Considerations

These above noted inaccuracies call into question the validity of broadly applied classification techniques. Alternatively, site- or region-specific classification algorithms admit the non-generality of CPT-based soil classification; but they generally fail to acknowledge inherent imprecision. In other words, false precision is forced upon the classification process when using crisp delineations between distinct soil types.

It has previously been recognized that the distinctions between soil types are more imprecise (McBratney and Odeh, 1997; Zhang and Tumay, 1999; Demmico and Klir, 2004; Lagacherie, 2005; Ross et al, 2007). For instance, there is no crisp line that separates silty sand from fine sand. Rather, one soil type should gradually transition into the other. Consider, for example, the cone resistance and sleeve friction field data plotted in Figure 3.2B, where the sand (●), silt (■) and clay (×) data clusters overlap each other to some degree. Such overlapping clusters are best modeled by fuzzy sets (Zadeh, 1965).

Fuzzy sets provide an intuitive means by which boundaries between imprecisely defined subjects can be modeled as gradual. The boundaries of a fuzzy set are defined by a membership function. Previous applications of this theory to soil type definition have been one-dimensional in nature (Ross et al, 2007) though the definition of soil type fuzzy sets for CPT-based classification would be two-dimensional.

This representation of a fuzzy classification chart is an improvement over traditional crisp charts because of the fuzzification of the boundaries between soil classes. The algorithms described below offer mechanisms for defining such fuzzy soil classes, one using site-specific data and the other applying fuzzy logic.

3.3. Fuzzy Kernel-Based Soil Classifying

3.3.1. Theory

Let us first consider fuzzy c-means clustering (Bezdek, 1981), a generalization of traditional statistical clustering whereby elements x_k ($k = 1, \ldots, n$) are classified into fuzzy groups with gradual, overlapping boundaries, defined by a membership function $A_i(x_k)$, as opposed to crisp groups with non-overlapping boundaries, like those seen in the classification charts in Figure 3.2. The algorithm dictates that a fuzzy partition P (or, pseudopartition, due to the overlapping nature of the clusters) of the problem space (i.e. the plot of the friction ratio and cone resistance data) is defined such that, for a given number of clusters c, a performance index $J_m(P)$ is minimized. The performance index is defined in Eq. (3-1),

$$J_m(P) = \sum_{k=1}^{n}\sum_{i=1}^{c}[A_i(x_k)]^m \left\| x_k - v_i \right\|^2, \qquad \textbf{(3-1)}$$

where $\boldsymbol{v}_i$ ($i = 1, \ldots, c$) are the cluster centers and <u>m</u> is a parameter related to the degree of fuzzification ($m=1.25$ provided the best results in this work). $J_m(P)$ measures the weighted sum of the squared Euclidean norm ($||x_k - v_i||$) between cluster centers and elements in the corresponding fuzzy cluster (Klir and Yuan, 1995).

Intuitively, the minimization of this performance index produces the most ideal spherically shaped fuzzy clusters for a given set of data. In reality, plausible data clusters are rarely spherical, the kernel-based fuzzy clustering algorithm (Zhang and Chen, 2003) permits the mapping of data into higher dimensional spaces that generate non-spherical fuzzy clusters.

The iterative process of kernel-based fuzzy c-means clustering (KFCM) begins with an initial, uninformed definition of the pseuodpartition. This entails defining a preliminary membership $A_i(x_k)$ for every data vector x_k ($k = 1, \ldots, n$) in each of the c fuzzy clusters. Care must be taken to ensure that the sum of the memberships for each data vector equals unity,

$$\sum_{i=1}^{c} A_i(x_k) = 1 \qquad \forall x_k (k = 1, \ldots, n) \; .$$

Based upon this initial fuzzy pseuodopartition, fuzzy clusters are calculated using the following equation, where the influence of kernel transformation is first evident:

$$v_i = \frac{\sum_{k=1}^{n} A_i(x_k)^m \kappa(x_k, v_i) x_k}{\sum_{k=1}^{n} A_i(x_k)^m \kappa(x_k, v_i)}, \quad i = 1, \ldots, c \; . \tag{3-2}$$

In Equation (3-2), the kernel function $\kappa(x_k, v_i)$ measures the similarity between the cluster centroid and a data vector. The further a data vector is from the cluster center, the lower the value of the corresponding kernel function. By using this function to weight the x values in both Equations (3-2) and (3-3), the clusters are transformed into non-spherical shapes. A number of functions can be applied as the kernel function. In this work the Gaussian function is used.

Once the cluster centers are calculated, the membership degrees for every data vector in all clusters are updated by applying Equation (3-3):

$$A_i(x_k) = \left(\sum_{i=1}^{c} \left(\frac{\kappa(x_k, v_m)}{\kappa(x_k, v_i)} \right)^{\frac{1}{l-1}} \right)^{-1}, \quad l = 1,\ldots,c; \; k = 1,\ldots,n, \qquad \textbf{(3-3)}$$

(Zhang and Chen, 2003). If, when this updated pseudopartition is compared to the previously defined pseuodpartition, the difference between the two is greater than some prescribed threshold ε, new cluster centers are calculated and the process is repeated until the absolute value of the difference between two consecutive pseudopartitions, $|P^{t+1} - P^t|$ is below the threshold,

$$\left| P^{t+1} - P^t \right| \le \varepsilon \, .$$

This difference may be obtained by finding the maximum difference in membership amongst all data vectors in all clusters, i.e. the formula in Equation (3-4),

$$\left|P^{t+1} - P^{t}\right| = \max_{\substack{i=1,\ldots,c \\ k=1,\ldots,n}} \left|A_i^{t+1}(x_k) - A_i^{t}(x_k)\right|, \tag{3-4}$$

as suggested by Klir and Yuan (1995). In this equation, the membership values of all data vectors in every cluster are compared to the previous time step. The maximum of all of these differences provides a measure of the difference between the pseuodpartitions of consecutive time steps.

3.3.2. Data

As alluded to in the introduction, both CPT soundings and borehole investigations were made at a number of coincident locations. Out of the *304* friction ratio-cone resistance data pairs, *125* were associated with soil samples that were comprised of at least 70% of one soil type: clay, silt or sand (boring log descriptions were limited to these three soil types). These data are plotted in Figure 3.3, along with an approximate outline of a crisp soil class that each defines. These boundaries were drawn by hand, since, as mentioned earlier, crisp classification results were sub-par.

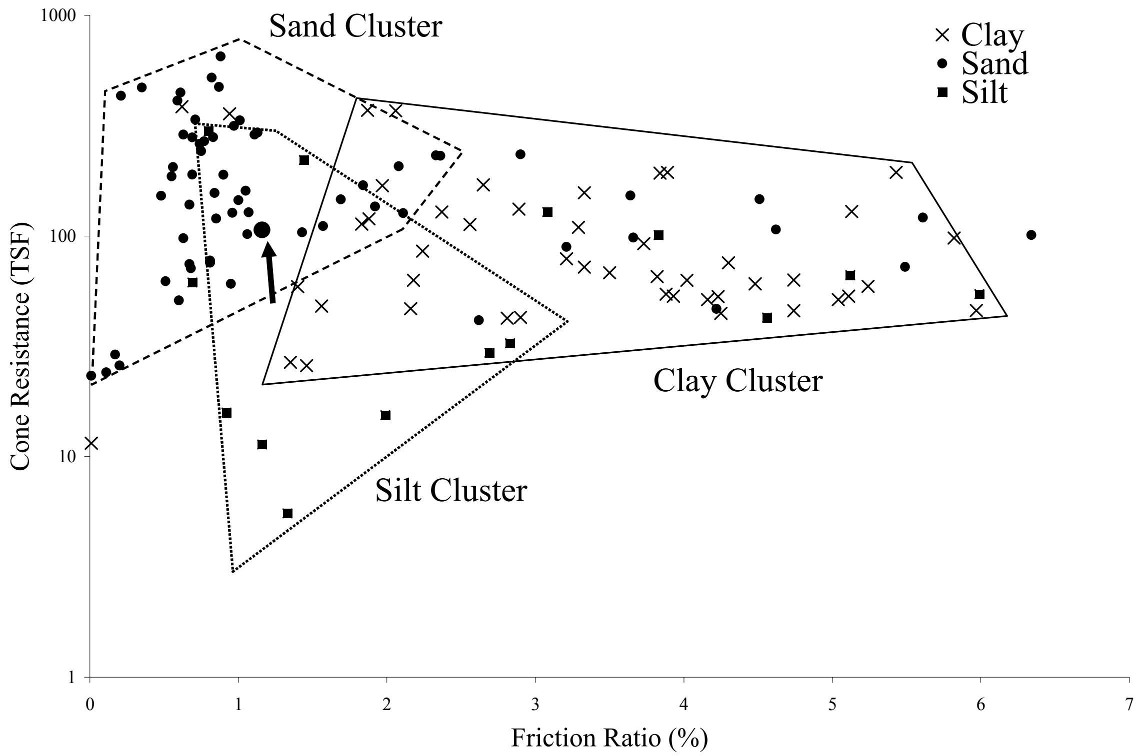


Figure 3.3. Overlapping crisp clusters are drawn upon the plot of friction ratios and cone resistance values. The overlapping clusters signify that some soil samples are comprised of more than one soil type (i.e. the large dot to which the arrow points is mostly sand with some silt)

Notice how these intuitively sketched clusters overlap each other and how some data belong to multiple clusters. This is likely due to the fact that most soil samples have grain sizes that would be interpreted as elements of different soils. For example, consider the data point (*1.16*, *106.57*) identified by the large dot to which the arrow is pointing. This data point resides in both the sand and silt clusters. Appropriately so, the borehole investigation found the soil to be a silty sand.

3.3.3. Results

Applying the kernel-based fuzzy c-means algorithm with *3* clusters to the *125* data vectors plotted in Figure 3.3 results in the fuzzy soil classes plotted in Figure 3.4. The identification of the soil classes as identifiers for the clusters (i.e. sand as the cluster in Figure 3.4 (top left), silt as the cluster in Figure 3.4 (top right) and clay as the final cluster in Figure 3.4 (bottom left)) is accomplished by a simple examination of the plotted data in Figure 3.3 with the clusters. Note, in Figure 3.4, how the sand samples are associated with low friction ratios and moderate-to-high cone resistance values, clay sample with moderate-to-high friction ratios and high cone resistance values, and silt with low-to-moderate friction ratios and low-to-moderate cone resistance value. Thus, by using these facts, for example, the fuzzy cluster whose maximum memberships are associated with low friction ratios and high cone resistance value is identified as sand. The remaining two clusters were identified in a similar manner. Figure 3.5 shows the 0.6 alpha-cuts for the 3 clusters. It is interesting to note that the 0.6 alpha-cuts resemble, to a reasonable degree, the hand-drawn clusters drawn about the CPT data in Figure 3.3. The degree of overlap between the clusters changes with alpha-cuts, lower alpha-cuts showing to more overlap.

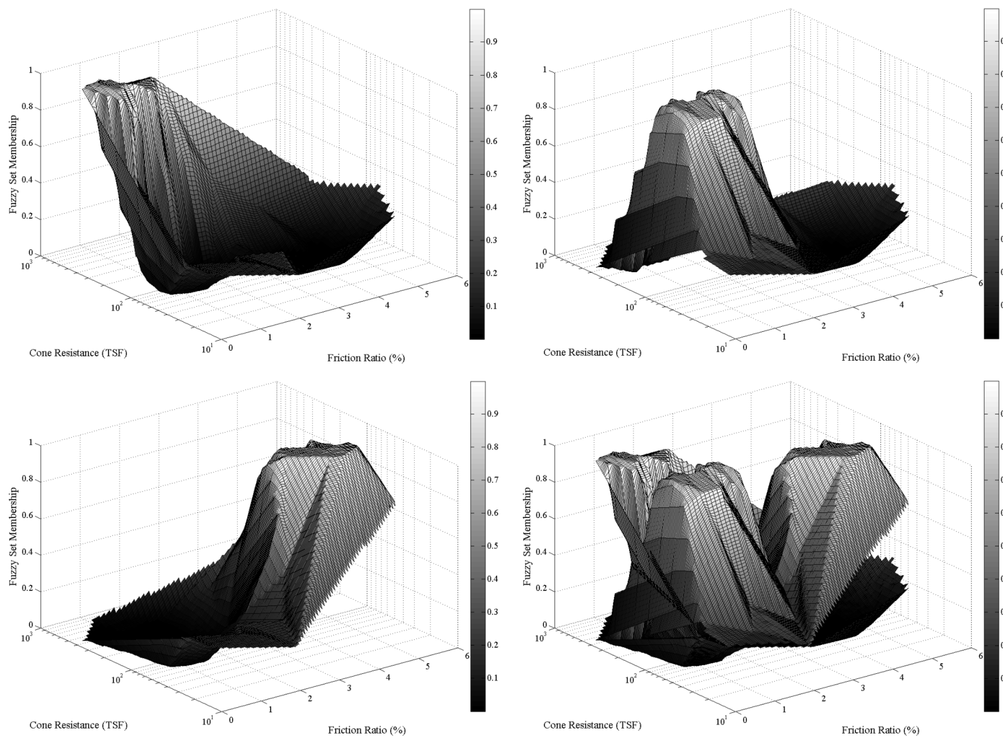


Figure 3.4. The fuzzy clusters representing (A) sand, (B) silt and (C) clay were produced by kernel-based fuzzy c-means clustering. All of the clusters are plotted together (D) to demonstrate a structure similar to the overlapping clusters in Figure 3.3

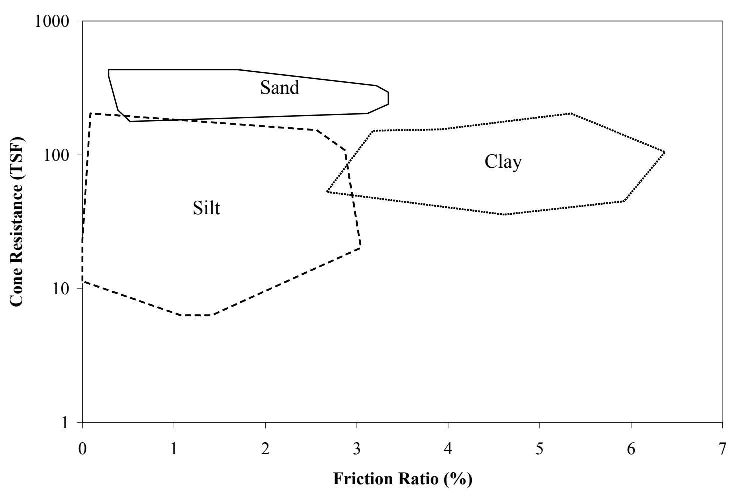


Figure 3.5. Approximate 0.6 degree alpha-cuts for the fuzzy clusters in Figure 3.4

When tested on the data used to form the fuzzy clusters, the fuzzy classification chart identified the predominant soil *50%* of the time, which is significantly better than the classification charts in Figure 3.2. That is, in cases where a soil was classified as a silty sand via a borehole investigation, the corresponding friction ratio – cone resistance data pair belongs to the fuzzy sand cluster with the greatest membership. In addition, like the neural network approach (Zhang and Tumay, 1999), this fuzzy classification chart provides an estimate of the percentages of clay, silt and sand comprising the soil sample, by interpreting the membership values in each of the three regions as percentages. When the remaining *174* data pairs are tested, the accuracy of the classifier in identifying the predominant soil type reduces to *32%*.

3.4. Genetic Algorithm Optimized Fuzzy Logic Soil Classifier

3.4.1. Theory

While data-driven fuzzy c-means clustering results in a fuzzy version of a soil classification chart (Figure 3.4 (bottom right)), it is accomplished without any formal expert logic. Douglas and Olsen (1981) identified relationships between soil properties (grain size, fines content) and the CPT parameters on their classification chart, as signified by the arrows (Figure 3.1B). Similar logic can be captured by fuzzy rules relating sleeve friction and cone resistance to the soil types clay, silt and sand in order to define soil classifications.

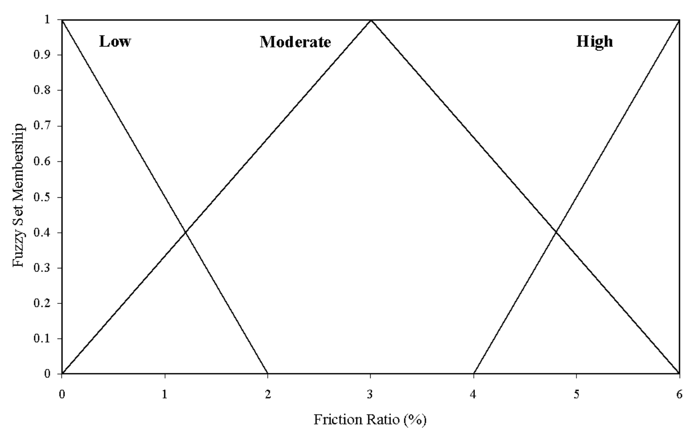
While the classification charts in Figure 3.1 are visibly dissimilar, there are some consistencies. Clayey soils are associated with higher friction ratio (sleeve friction) values and low cone resistance values, moderate friction ratio and cone resistance values result from penetration of silty soils, and sandy soils lead to low friction ratio and high cone resistance values. These observations can be framed as conditional statements, such as

If friction ratio is high and cone resistance is low, then soil type is clay,

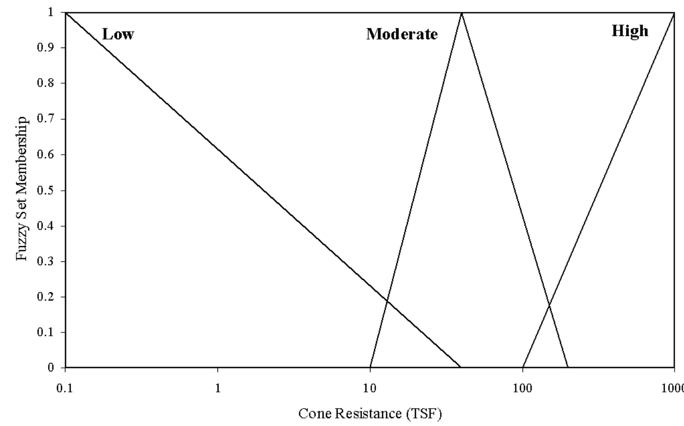
If friction ratio is moderate and cone resistance is moderate, then soil type is silt,

If friction ratio is low and cone resistance is high, then soil type is sand.

These statements act as general guidelines by which CPT-based soil classification can be governed.



A



B

Figure 3.6. Arbitrary pseudopartitioning of the variables (A) friction ratio and (B) cone resistance by fuzzy sets *low*, *medium* and *high*. Because the membership functions were arbitrarily defined, the estimates produced by the corresponding fuzzy rules are inaccurate

Consider the partitioning by fuzzy sets of the two CPT parameters in Figure 3.6A and Figure 3.6B. By combining these fuzzy sets with the conditional statements, fuzzy rules (Klir and Yuan, 1995) are constructed (Figure 3.7).

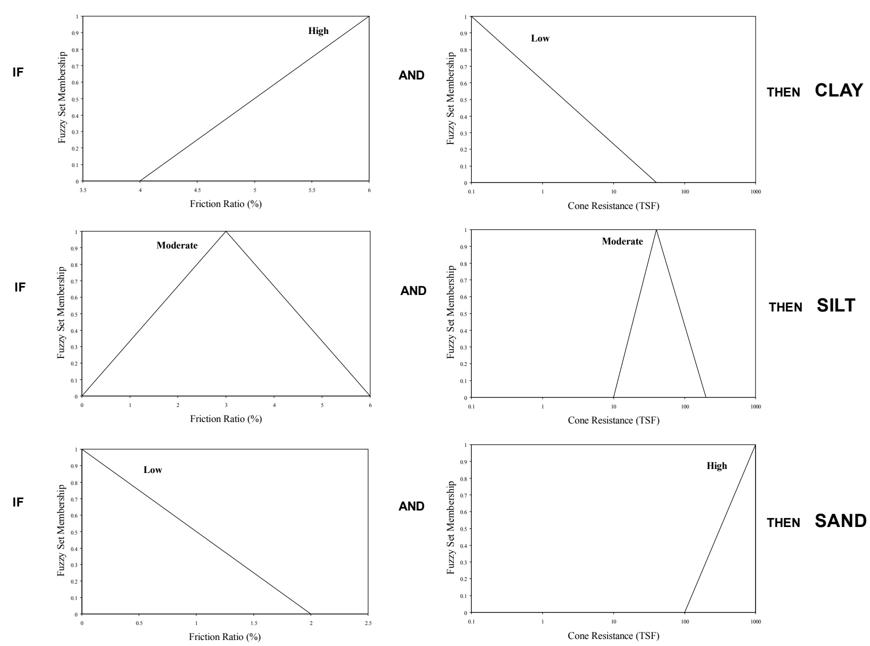


Figure 3.7. The fuzzy rules corresponding to the fuzzy partitions in Figure 3.6 and conditional statements that capture the general physics underlying CPT-based soil classification

When this fuzzy rule base is executed for all plausible values of cone resistance and friction ratio, clay, sand and silt fuzzy soil regions are delineated (Figure 3.8). For any given pair of values, one simply reads the membership value of the data pair in each fuzzy region and normalizes the three soils' membership values so they sum to one. The soil compositions are estimated for the same pairs of values as in Table 3.1, showing that the fuzzy rule base is a poor classifier (Table 3.2), achieving only one correct classification.

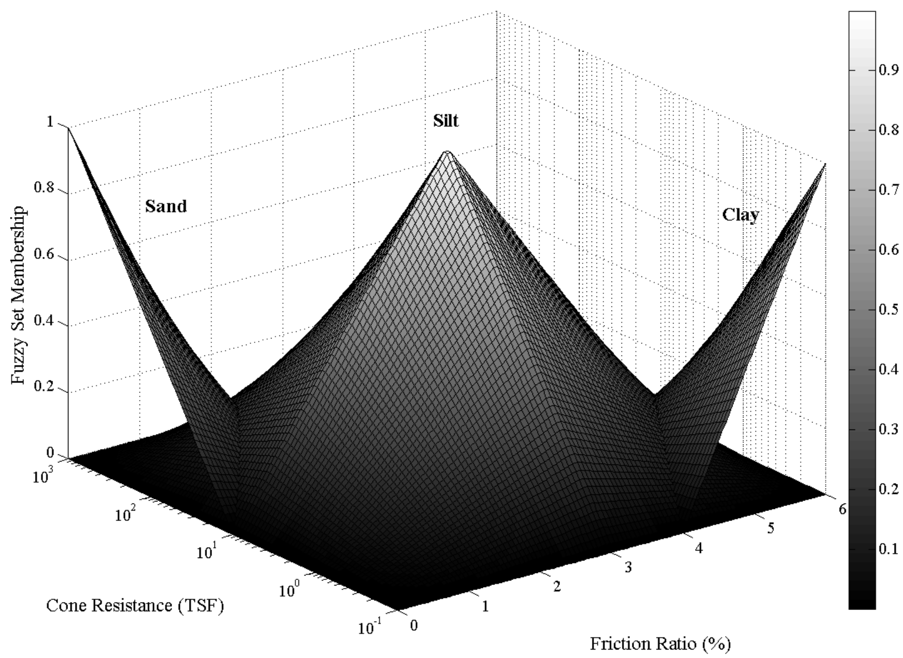


Figure 3.8. Fuzzy soil clusters produced by the evaluation of the fuzzy rules in Figure 3.7

The reason for the inaccuracies is primarily caused by the shape and location of the membership functions in Figure 3.6. In order to improve estimates, the membership functions must be fine-tuned to the particular data set of interest. In order to optimize the membership functions for best results, a genetic algorithm is employed.

The genetic algorithm, so named by Holland (1975) because of its simulation of biological evolutionary processes, is a general optimization tool that has applications in a number of facets of applied science and engineering. A number of such applications have been directed toward the optimization of fuzzy logic-based models (Herrera et al, 1993, 1995; Ho et al, 2000; Reyes, 2004; Castro and Camargo, 2005; Hong et al, 2006). A thorough review of such applications is provided by Cordon et al (2004).

Table 3.2. The same data vectors from Table 3.1 are classified using the fuzzy rule base approach

Qc (tsf)	Friction Ratio (%)	Boring Log	Fuzzy Estimated Composition
261.81	0.74	sand (90% sand, 10% fines)	sandy silt (0% clay, 62% silt, 38% sand)
112.78	2.56	sandy clay (70% fines, 30% sand)	silt (0% clay, 100% silt, 0% sand)
62.95	2.18	lean clay (80% fines, 20% sand)	silt (0% clay, 100% silt, 0% sand)
185.34	0.82	silty sand (55% sand, 45% fines)	silty sand (0 % clay, 21% silt, 79% sand)

The application of genetic algorithms in this work is focused upon the determination of the parameters (vertices of the triangular fuzzy sets) that govern the shape and location of the fuzzy sets *low*, *moderate* and *high* friction ratio and *low*, *moderate* and *high* cone resistance such that the soil composition estimates are accurate. A particular instance of these fuzzy set parameters comprises an *individual*, in genetic algorithm terms. For each individual, the fuzzy rule base is evaluated and the difference between the estimated soil composition is compared to the actual soil composition. A *fitness score* is assigned to each individual in this initial population, signifying the accuracy of the estimates it produced. Individuals with high fitness scores are permitted to reproduce via cross breeding, producing a second generation of individuals (Beasley et al, 1993). This process is repeated until a prescribed maximum number of generations is reached or until the genetic algorithm reaches an optimal solution, which in this case is a minimum difference between estimated and actual soil composition percentages. The parameters encoded in the optimal individual define the fuzzy sets in the above three fuzzy rules.

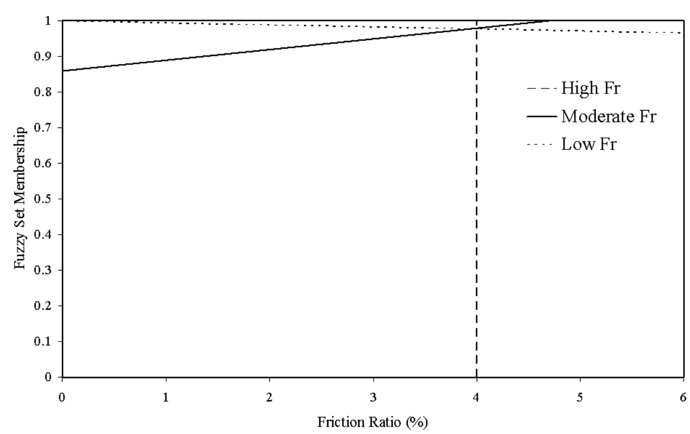
3.4.2 Data

The cone resistance and friction ratio data values with corresponding soil composition percentages were used for genetic algorithm fuzzy rule base tuning were gleaned from borehole logs. The ideal source of soil composition data are grain size distributions, as utilized in Zhang and Tumay (1999). However, borehole logs are generally far more numerous in practical applications, and as such, are applied herein. In some cases the borehole log soil descriptions are also accompanied by clay, silt and sand percentages. In this case, however, the data did not possess such specificity, merely providing percentages of fines and sand in each soil sample. Thus, the qualitative soil descriptions derived from the borehole logs were translated into soil percentages via the Burmeister classification.

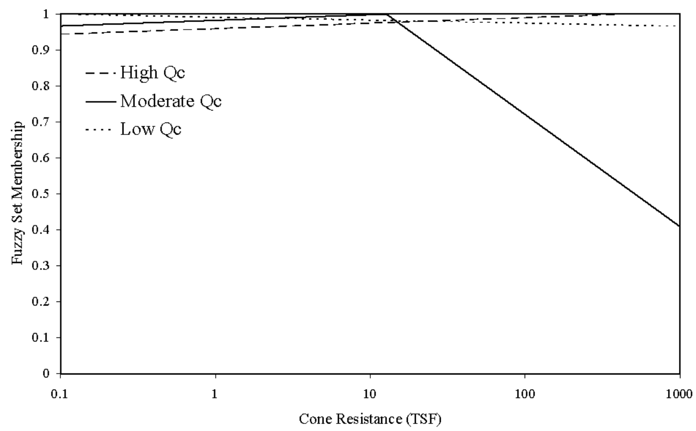
Over *300* cone resistance and friction ratio data pairs were known for the soil samples described in *6* boreholes. Data from two boreholes, roughly *40%* of the entire data set, were used to optimize the fuzzy set parameters in the fuzzy rule base. The percentage of clay, silt and sand were deduced from the borehole log descriptions. Using the optimal fuzzy rules, the soil compositions were estimated for the remaining *4* boreholes and compared to the actual soil composition values.

3.4.3 Results

The optimal fuzzy sets for *low*, *moderate* and *high* friction ratio and cone resistance are shown in Figure 3.9. The fuzzy sets are far from intuitive and, with such elevated possibility values (Zadeh, 1978), convey a certain degree of ignorance. For friction ratio, every value in the range [*0, 6*] has high membership in each fuzzy set.



A



B

Figure 3.9. Genetically optimized membership function for the fuzzy sets low, medium and high partitioning the variables (A) friction ratio and (B) cone resistance

Such a circumstance is likely due to the data used to optimize the fuzzy set parameters. The data are plotted in Figure 3.10 showing the percentage of clay, silt and sand in each data pair. Notice that there is little organization in the data and the soil types associated with each, though areas of high sand percentages are distinguishable in Figure 3.10 (bottom). Had the data exhibited more distinct clay, silt and sand regions on the plot, the membership functions of the fuzzy sets in Figure 3.8 would provide more appealing partitions of the CPT variables.

In spite of the data quality, the fuzzy logic soil classifier correctly identified the soil types present in a sample 82% of the time (*143* correct out of *174* total samples). Nevertheless, the percentage of each soil type was less often correctly identified. Though a certain degree of this error is due to the fuzzy classifier, other sources of error include the quantification of borehole logs and the omnipresence of silty sand in the borehole logs.

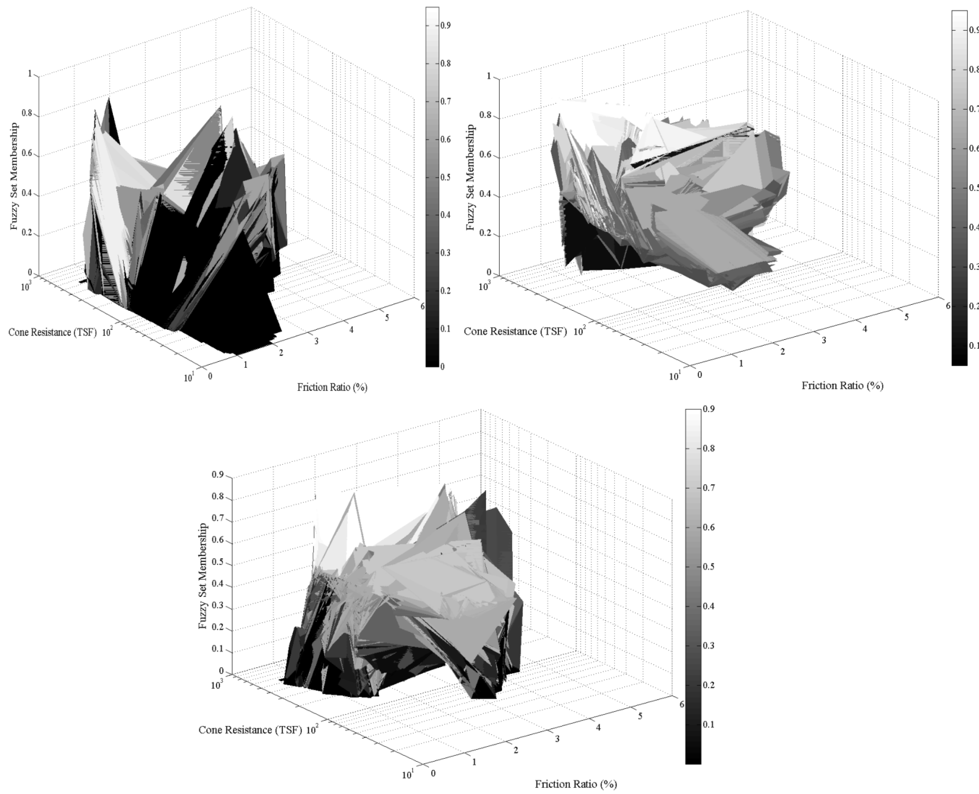


Figure 3.10. These plots illustrate the percentages of clay (top left), silt (top right) and sand (bottom), as determined by borehole investigations, for the over-300 friction ratio – cone resistance data pairs. Notice how difficult it would be to differentiate clusters of clay, silt and sand

3.5. Discussion of Results

In this paper, two new approaches to identifying soil classes based upon CPT sounding data are introduced. The site-specific approach utilizes kernel-based fuzzy clustering, whereas a more another soil classifier was attempted using fuzzy rules.

The improvements of the new approaches over traditionally employed charts for the classification of soil samples are meaningful and demonstrate promise in fuzzy-based classification tools. The most significant characteristic of the new methods, however, is the representation of soil classes with gradually changing boundaries. With such boundaries, one can represent a number of the soil types in the Robertson chart (i.e. *silty clay*, *sandy silt*) with just three fuzzy soil classes (clay, silt and sand). For example, a data point that has membership in both the clay and silty classes, greater in the former than the latter, can be classified as a *silty clay*.

While the performance of the kernel-based fuzzy classification was inferior to that of the fuzzy rule-based classification, it is appealing in its strict dependence upon the site data. At any site, a unique fuzzy classification chart can be constructed. Moreover, the kernel-based approach marks a first attempt to fuzzify soil classification charts in the vein of Begemenn (1963) and Douglas and Olsen (1981).

The fuzzy rule-based approach, however, is considered more promising due to its higher classification success rate over both the kernel-based approach and traditional classification charts. It is expected that, with the consideration of more data, the rules would frame a logic that provides a more universal appeal. Furthermore, the generality and transparency of the fuzzy rule-based approach make it an attractive classification method.

References

Beasley, D., Bull, D.R., and Martin, R.R., (1993). "An overview of genetic algorithms: Part 1, fundamentals." *University Computing*, 15(2), 58-69.

Begemann, H.K.S. (1953). "Improved method of determining resistance to adhesion by sounding through a loose sleeve placed behind the cone." *Proceedings of the 3rd International Conference on Soil Mechanics and Foundation Engineering* (*ICSMFE*), 213-217.

Begemann, H.K.S.P. (1963). "Use of static soil penetrometer in Holland." *New Zealand Engineering*, 18(2), 41-49.

Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.

Castro, P.A.D., and Camargo, H.A. (2005). "Focusing on interpretability and accuracy of a genetic fuzzy system." *14th International Conference on Fuzzy Systems*, 696-701.

Cordon, O., Gomide, F., Herrera, F., Hoffmann, F., and Magdalena, L. (2004). "Ten years of genetic fuzzy systems: Current framework and new trends." *Fuzzy Sets and Systems*, 141, 5-31.

Demicco, R.V., and Klir, G.J. (2004). *Fuzzy Logic in Geology*, Elsevier, Amsterdam.

Douglas, B.J., and Olsen, R.S. (1981). "Soil classification using electric cone penetrometer." *Proceedings of a Session at the ASCE National Convention: Cone Penetration Testing and Experience*, 209-227.

Herrera, F., Lozano, M., and Verdegay, J.L. (1993). "Genetic algorithms applications to fuzzy logic based systems." *Proceedings of the 9th Polish-Italian and 5th Polish-Finnish Symposium on Systems Analysis and Decision Support in Economics and Technology*, 125-134.

Herrera, F., Lozano, M., and Verdegay, J.L. (1995). "Tuning fuzzy logic controllers by genetic algorithms." *International Journal of Approximate Reasoning*, 12, 299-315.

Ho, S.Y., Cheng, T.K., and Ho, S.J. (2000). "Designing an efficient fuzzy classifier using an intelligent genetic algorithm." *Proceedings of the IEEE Computer Society's International Computer Software and Applications Conference*, 293-298.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Michigan.

Hong, T.P., Chen, C.H., Wu, Y.L., and Lee, Y.C. (2006). "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions." *Soft Computing*, 10, 1091-1101.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, New Jersey.

Kurup, P.U., and Griffin, E.P. (2006). "Prediction of soil composition from CPT data using general regression neural network." *Journal of Computing in Civil Engineering*, 20(4), 281-289.

McBratney, A.B., and Odeh, I.O.A. (1997). "Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions." *Geoderma*, 77, 85-113.

Mullarky, P.W., and Fenves, S.J. (1986). "Fuzzy logic in a geotechnical knowledge based system: CONE." *Civil Engineering Systems*, 3(2), 58-81.

Reyes, C.A.P. (2004). *Coevolutionary fuzzy modeling*, Springer, Berlin.

Robertson, P.K. (1990). "Soil classification using the cone penetration test." *Canadian Geotechnical Journal*, 27(1), 151-158.

Robertson, P.K., Campanella, R.G., Gillespie, D., and Grieg, J. (1986). "Use of piezometer cone data." *Proceedings*, *In-situ '86*, ASCE Specialty Conference, Blacksburg, Virginia.

Romo, M.P., and Garcia, S.R. (2003). "Nerofuzzy mapping of CPT values into solid dynamic properties." *Soil Dynamics and Earthquake Engineering*, 23(6), 473-482.

Ross, J., Ozbek, M., and Pinder, G.F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*, 39(8), 765-780.

Schmertmann, J.H. (1978). "Guidelines for cone test, performance, and design." *Federal Highway Administration*, Report FHWA-TS-78209.

Vos, J.D. (1982). "The practical use of CPT in soil profiling." *Proceedings of the 2nd European Symposium on Penetration Testing*, 933-939.

Zadeh, L. (1965). "Fuzzy sets." *Information and Control*, 8(3), 338-353.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1, 3-28.

Zhang, D.Q., and Chen, S.C. (2003). "Clustering incomplete data using kernel-based fuzzy C-means algorithm." *Neural Processing Letters*, 18(3), 155-162.

Zhang, Z., and Tumay, M.T. (1999). "Statistical to fuzzy approach toward CPT soil classification." *Journal of Geotechnical and Geoenvironmental Engineering*, 125(3), 179-186.

4: HYDRAULIC CONDUCTIVITY ESTIMATION VIA FUZZY ANALYSIS OF GRAIN SIZE DATA

Accepted for publication in Mathematical Geology on February 8, 2007; Published in November 2007;

Authors: Ross, J., Ozbek, M., and Pinder, G.F.

A measure of hydraulic conductivity is arguably the most important variable to practicing hydrogeologists. However, the amount of readily available hydraulic conductivity data at any site is generally small, given the resources required to adequately sample a spatial domain. However, other hydrogeologic data, such as grain size distributions and soil descriptions, are often rather easy to determine. Here, a fuzzy reasoning algorithm is used to define a relationship between soil grain size and hydraulic conductivity. By introducing soil grain distributions and qualitative borehole log descriptions into this fuzzy inference system, hydraulic conductivity can be estimated. Herein the theory is defined, and an application to data from a Superfund site is provided, where the inference procedure produces accurate hydraulic conductivity estimates.

4.1. Introduction

Large amounts of information often result from a hydrogeologic investigation, whereas hydraulic conductivity data comprise a relatively small subset because the tests required to determine it are generally both time consuming and expensive. Consequently, regional estimates of hydraulic conductivity are inherently uncertain. Since significant uncertainty in hydraulic conductivity estimates lead to uncertainty in groundwater flow and transport models, less costly means of acquiring hydraulic conductivity data have been explored. Specifically, several attempts have been made to relate other more easily measured hydrogeologic parameters to hydraulic conductivity. Porosity, particle sorting, texture and soil type have all been related to hydraulic conductivity. Such relationships have intuitive appeal.

The determination of porosity requires little effort and expense when compared to that of hydraulic conductivity, and relationships between it and hydraulic conductivity have been modeled (Selley, 1985). The degree of sorting can be evaluated from the soil's grain size distribution curve and related to hydraulic conductivity through an empirical equation (Demmico and Klir, 2004). Textural characteristics (i.e. grain size, ductility) are determined by the examination of a soil sample and have been related to hydraulic conductivity in sandstone through fuzzy relations (Fang and Chen, 1997). In a very general sense, an understanding of the ranges of likely hydraulic conductivity values for various soil types has been used to map these two variables (Domenico and Schwartz, 1990). Further, fuzzy sets have been used to characterize an expert's notion of the hydraulic conductivity at unsampled locations (Bardossy et al, 1989), essentially providing opinion-based imprecise values of hydraulic conductivity at different locations in a spatial domain. More appropriate to the theory presented herein, attempts have been made to relate porosity and grain size distributions to hydraulic conductivity through petrophysical relations, the Kozeny-Carman Equation and other pedotransfer functions (Uma et al, 1989; Sen, 1992; Alyamani and Sen, 1993; Koltermann and Gorelick, 1995).

Though the above methods differ in their approaches, they share a common goal: to expand available hydraulic conductivity data with that derived from complementary sources. However, these approaches suffer from various shortcomings. Porosity data, although easily measured in the laboratory, are not commonly available in practical applications; the degree of sorting is simply a convenient measure of considerably greater information from a soil grain distribution curve; the relationship of the textural measure to hydraulic conductivity has only been evaluated for sandstone; the soil type-to-hydraulic-conductivity relationship is highly uncertain; fuzzy set hydraulic conductivities provide little help, as fuzzy kriging algorithms produce fuzzy estimates (Diamond, 1989; Bardossy et al, 1989, 1990). Finally, attempts to relate hydraulic conductivity to grain size distributions and porosity have not considered the high degree of uncertainty implicit in such a relationship, as they have, thus far, relied upon deterministic equations. Such a relation is however generally recognized as fuzzy and should be modeled as such.

In this paper, a method is defined whereby hydraulic conductivity values are estimated from grain size distribution curves and qualitative descriptions of soil samples (i.e. describing a soil sample as "a silty fine sand with some clay"). The approach recognizes that the grain size distribution curve implicitly contains some of the measures noted above, such as grain size and sorting. Soil type is also often implicitly defined when common soil classifications are plotted along the horizontal axis of the grain size distribution graph.

However, this crisp representation of soil types, which states that a grain size diameter of 0.002 mm characterizes clay, but a grain size diameter of 0.0021 mm is characteristic of a silt, implies unwarranted accuracy. If the soil grain size axis is partitioned so that the soil classification avoids sharp boundaries (*i.e.* is fuzzy), and there is a fuzzy rule base that defines an imprecise relationship between soil type to hydraulic conductivity, the grain size distribution curve can be used to estimate a hydraulic conductivity value for the sample.

The above defined estimate of hydraulic conductivity can be enhanced further using an independent qualitative field description of a soil sample provided by a groundwater professional. Such descriptions by, for example, a geologist are done by inspecting soil samples taken at borehole locations. The adjectives used to describe a soil sample in the field, such as sandy or clayey, are inherently imprecise. However, such subjective information reflects valuable insight into the sample's makeup and complements the information obtained by a grain size distribution. This qualitative information, converted into fuzzy sets, will produce results similar to, and that augment the soil grain distribution data. The information from the grain size distribution curves and field assessments can be combined to produce a knowledge base that facilitates hydraulic conductivity estimates. Such an integration of subjective and objective fuzzy geological data is well founded in mineral potential mapping (Cheng and Agterberg, 1999; Brown et al, 2003; Porwal et al, 2003). This method is developed herein and tested using data from the CIBA-Geigy Superfund site in Toms River, New Jersey.

In the following sections, the creation of the fuzzy sets is discussed first. The introduction and transformations of the qualitative field data (Table 4.1) and grain size distributions into fuzzy data are explained next, as is the juxtaposition of the fuzzy data with fuzzy rules. Finally, an example with results from the application of this reasoning mechanism to Toms River data is provided.

Table 4.1. Qualitative field data in the form of borehole log descriptions of the four soil samples used in the application, below

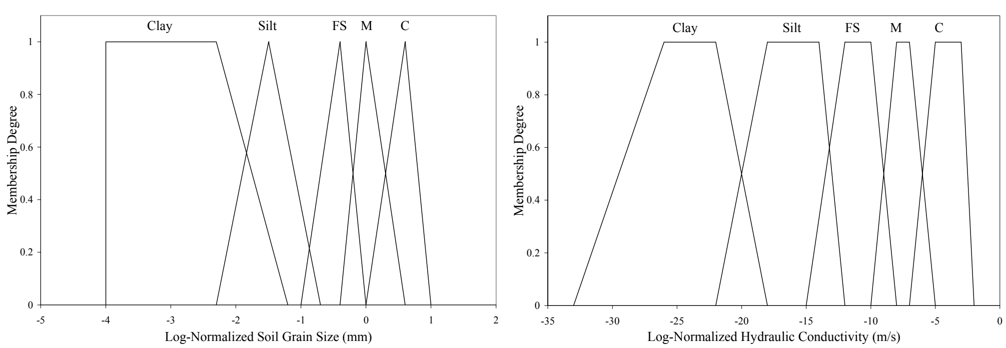
Sample	Borehole Log Description
1	*Silty Fine Sand*
2	*Silt to Medium Clay Silt, trace Fine Sand*
3	*Fine Sand, some Clayey Silt, trace Medium Sand*
4	*Fine Sand, some Silt, trace Medium Sand*

4.2. Defining Sets and Distributions

The definition of soil grain fuzzy sets is integral to defining the relationship between soil grain size and hydraulic conductivity. If the soil grain fuzzy sets are poorly defined, then the entire relationship is poorly defined and will produce disappointing estimates. The fuzzy sets defined on the domain soil grain size (mm) will be named after the soil types that are generally found on soil grain distribution curves (Coarse Sand, Medium Sand, Fine Sand, Silt, Clay).

The fuzzy sets that are provided herein are defined by membership functions and aim to capture the imprecision inherent in soil type names, like *fine sand* and *clay*. These functions, typically piecewise linear to create triangular or trapezoidal fuzzy sets, represent, for every soil grain size on the abscissa, the degree to which that soil grain size belongs to a particular fuzzy set. Given that the fuzzy sets represent soil types, the membership functions stipulate the degree to which a given soil grain size belongs to the concepts *coarse sand*, *medium sand*, etc.

As a rule of thumb, fuzzy sets should both overlap each other and, together, cover the range of reasonable domain values. This serves two purposes. First is semantic continuity, where a realistic definition of soil types must have gradual and overlapping boundaries. Secondly, overlapping fuzzy sets make for continuous results when used in a rule base. That is, for gradually changing input soil grain size values, the hydraulic conductivity estimates that result from the rule-based reasoning mechanism also gradually change.



A **B**

Figure 4.1. Fuzzy sets defined by an expert to characterize A, common soil types, and B, their corresponding hydraulic conductivities. The soil types considered are clay, silt, fine sand (FS), medium sand (MS), and coarse sand (CS)

The soil type fuzzy sets based on soil grain size are shown in Figure 4.1A. Notice that they both overlap and overlay the range of reasonable soil grain sizes. These soil type fuzzy sets, defined with an understanding of the aforementioned rules of thumb, relied upon traditional crisp delineations of soil types as guidelines. Though there is precedent for fuzzy soil type definitions (Hsieh et al, 2005), such fuzzy sets did not guide the definitions of the soil types in Figure 4.1A. The hydraulic conductivity fuzzy sets are defined in a similar manner (Figure 4.1B), though the expertise of a hydrogeologist was the only guide in their construction. Information regarding the soil type and hydraulic conductivity fuzzy sets are provided in Table 4.2.

Table 4.2. Parameters for the soil type and hydraulic conductivity fuzzy sets in Figure 4.1

Fuzzy Set Name	Soil Type Parameters (mm)	Log Soil Type Parameters (log mm)	Log Hydraulic Conductivity Parameters (log m/s)
Clay	[0.0001, 0.0001, 0.005, 0.063]	[-4, -4, -2.3, -1.2]	[-33, -26, -22, -18]
Silt	[0.005, 0.032, 0.2]	[-2.3, -1.5, -0.7]	[-22, -18, -14, -12]
Fine Sand	[0.1, 0.4, 1]	[-1, -0.4, 0]	[-15, -12, -10, -8]
Medium Sand	[0.4, 1, 4]	[-0.4, 0, 0.6]	[-10, -8, -7, -5]
Coarse Sand	[1, 4, 10]	[0, 0.6, 1]	[-7, -5, -3, -2]

The data in this table describe the parameters, described below, of the fuzzy sets. For triangular fuzzy sets the middle value is the median (full membership), while the left and right parameters are the upper and lower bounds of the fuzzy set (zero membership). Trapezoidal fuzzy sets differ by having two median parameters that define an interval of values with full membership.

Fuzzy rules are the means by which an estimate is actually produced for a given input value. Each rule relating a soil type fuzzy set to a hydraulic conductivity fuzzy set defines a vaguely defined “point” that, when plotted along with a number of other rules relating the same variables, defines an imprecise relationship between these variables.

The fuzzy rules provided are very simple in this case. Since the soil grain fuzzy sets share the same names as the hydraulic conductivity fuzzy sets, the rules essentially match like-named fuzzy sets. The rules are structured as follows:

If Soil Grain Size is Coarse Sand THEN Hydraulic Conductivity is that of Coarse Sand.

If Soil Grain Size is Medium Sand THEN Hydraulic Conductivity is that of Medium Sand.

etc.

The collection of the five rules provided is the rule base, and the heart of the fuzzy reasoning process. The reasoning will be discussed later.

The input into each rule is derived from a grain size distribution curve, resulting from sieve and hydrometer analyses. A grain size distribution is a representation of the distribution of soil grain sizes for a given sample of soil (Figure 4.2). They are cumulative distribution functions (CDF), which are structurally and theoretically distinct from fuzzy sets. The vertical axis is analogous to a probability value in that it records what fraction of the soil sample (by weight) is less or equal in size to each grain size value along the horizontal axis. Thus, the grain size distribution must be transformed from a probabilistic form to one that is consistent with a fuzzy set representation, without changing or losing the information contained in the original distribution.

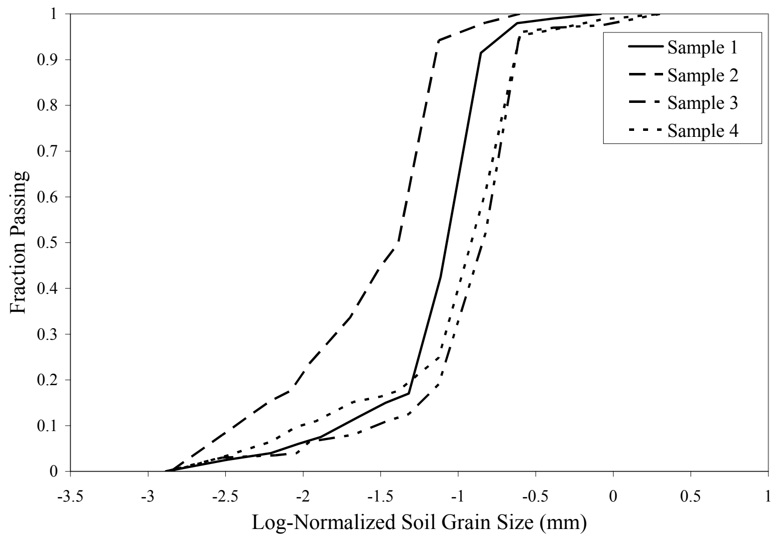


Figure 4.2. Grain size distribution curves for the four samples used in the application section, below. The vertical axis is analogous to probability

To achieve this goal, the discrete CDF must first be converted to a discrete probability density function (PDF). The probability values represent the prevalence of each soil grain size.

Next, a tool from Possibility Theory, a probability-possibility transformation (Oussalah, 2000) is used to transform the PDF to a possibility distribution (PD), $\pi(x)$, on the variable X (Zadeh, 1978). Such a transformation converts uncertain data in *probabilistic* form to uncertain data in *possibilistic* (related to fuzzy logic) form. The PD is the sister distribution to the PDF, as it is able to characterize the same data with a different interpretation.

The possibility value for a soil grain size represents the degree to which the soil grain size is possible, given some background information. Possibility distributions are generally used to characterize a person's knowledge regarding a matter about which there is no available evidence from which to make a certain conclusion. Thus, while fuzzy sets characterize certain but imprecise information, possibility distributions define uncertain information. Total uncertainty (or ignorance) would be represented by a possibility distribution where every value along the abscissa has a possibility of one (everything is possible and nothing can be discounted).

However, interpretation is not of significance in this algorithm, as only the structure of the PD is important here. Fuzzy sets and possibility distributions are generally defined on the same scale, and the two differ only in interpretation. As such, they are defined on the same range, since their scales are consistent.

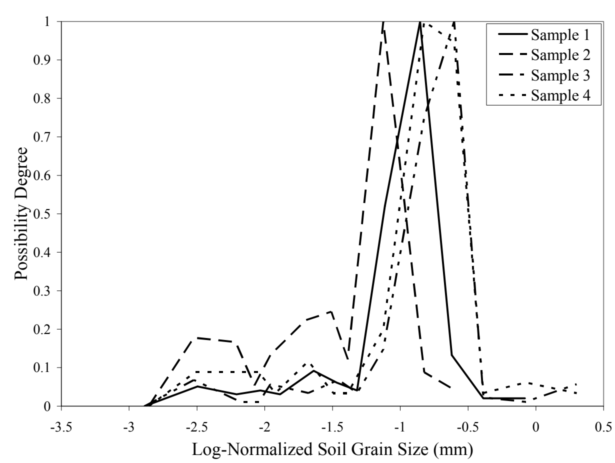


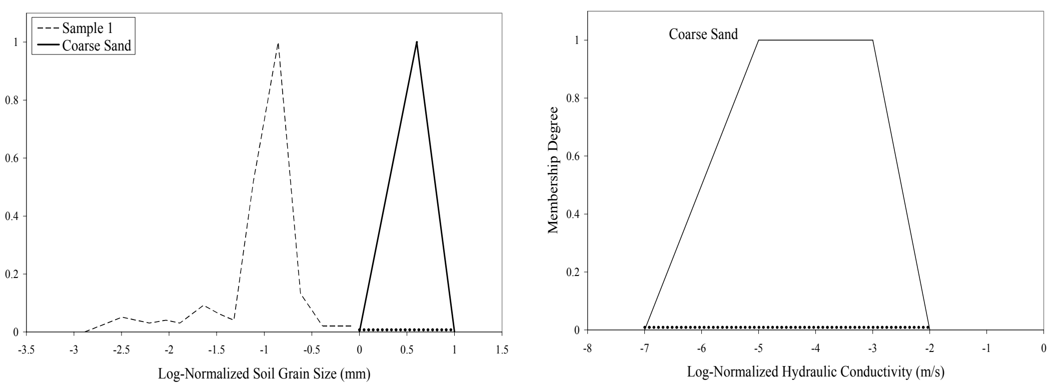
Figure 4.3. Possibility distribution of the four samples used in the application, below, transformed from the grain size distribution data

The transformation used is the most obvious and most common PDF-PD transformation. Each probability value is merely divided by the maximum probability value, which satisfies the constraint that a possibility distribution needs to have at least one domain value that is entirely possible, $\pi(x)=1$, (Dubois and Prade, 1988). The results of the transformation of the soil grain PDFs into soil grain PDs for the four soil samples considered in the Application Section are in Figure 4.3. With the soil grain data now characterized by a PD, reasoning may begin.

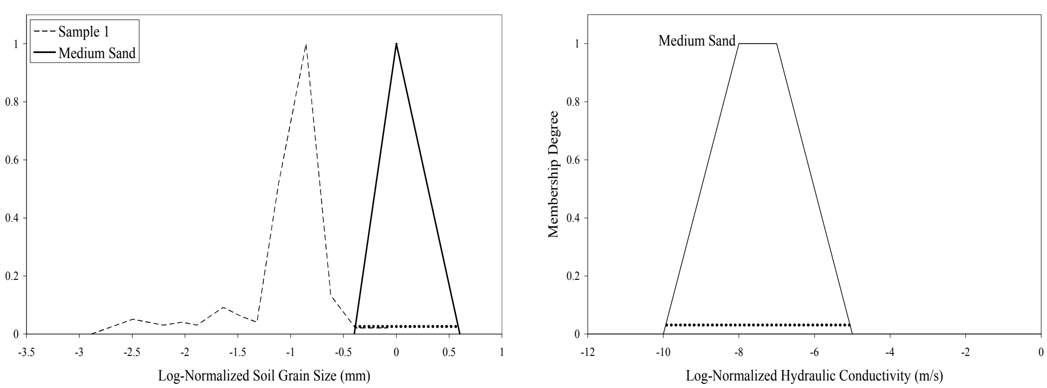
Juxtaposing the PD with the soil grain fuzzy sets, it is evident that the former intersects many of the fuzzy sets. Figure 4.4 provides insight into how the PD for one sample intersects each of the five soil grain fuzzy sets, which, in turn, reveals how each of the rules, described earlier, will operate when estimating hydraulic conductivity. The maximum degree of intersection with each fuzzy set is important. That is to say, if one visualizes the intersection of the PD and a certain fuzzy set, the maximum degree of membership of the intersection is the degree to which the corresponding rule will operate. This is further explained in the following section.

It is interesting to note first, however, looking at the left-hand graphs in Figure 4, that it seems rather straightforward to qualitatively describe the soil. For the example given, this soil sample maximally intersects the Fine Sand and Silt fuzzy sets more so than the others. Knowing this, one might describe the soil sample as fine sandy silt, with trace amounts of clay and medium sand. Consider now an alternative strategy to describing this soil.

Often soil samples are described via visual inspection by the groundwater professional resident during drilling. Based on this a soil type identification is recorded in a borehole log. This information is helpful provided that it can be presented in a manageable framework.

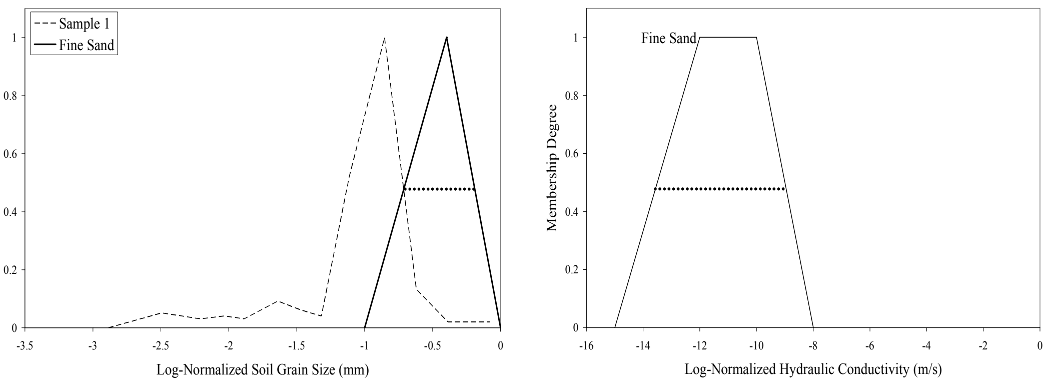


A

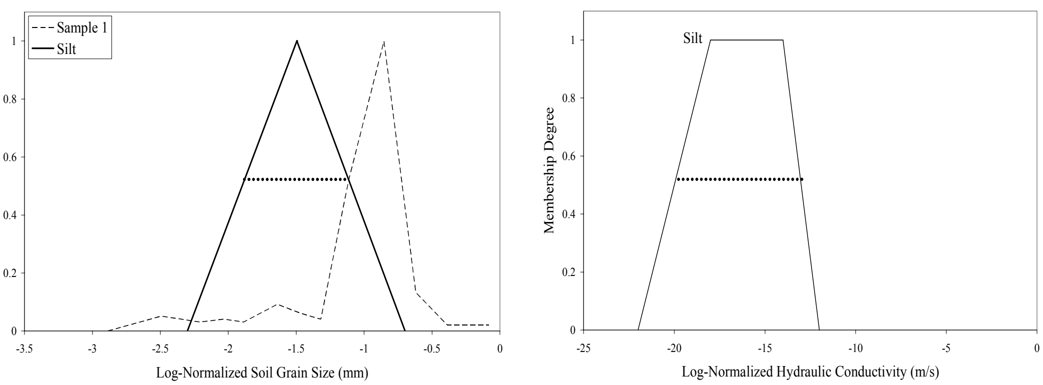


B

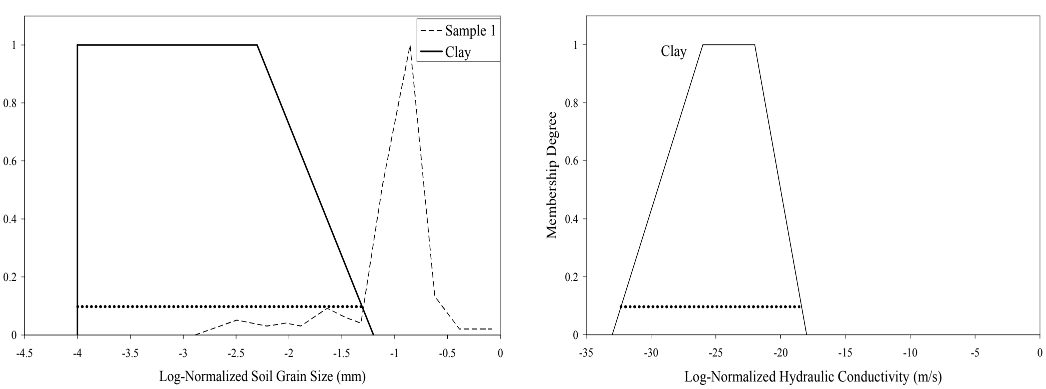
(figure continued on next page)



C



D



E

Figure 4.4. The possibility distribution representing the grain size distribution data for sample 1 is intersected with the fuzzy sets (A) coarse sand; (B) medium sand; (C) fine sand; (D) silt; and (E) clay (left-most figures. The maximum degree of intersection between the PD and each fuzzy set defines where each fuzzy set is cut and as well as the corresponding cut hydraulic conductivity fuzzy sets (shaded regions in right-most figures)

Fuzzy sets offer a unique approach to exploiting these subjective descriptions, as they generally define linguistic variables. The geologist's description of the soil sample is a linguistic variable of sorts. It may be converted into a fuzzy set by the person preparing the log or an appropriate expert. For instance, if a soil sample is described as being a silty fine sand, a fuzzy set resembling that in Figure 4.5 (situated between the Silt and Fine Sand fuzzy sets) may be defined by an expert using a basic understanding of membership functions. In general, one may feel as though the soil sample is best represented by: 1) a single soil grain size (full membership), 2) may be characterized by other grain size values to a lesser degree (membership decreases from unity), and 3) would not be characterized by other soil grain sizes (zero membership). Thus, a geologist, knowing the predefined Fine Sand and Silt fuzzy sets in Figure 4.1, could feel comfortable characterizing a soil sample deemed to be a silty fine sand by sketching the triangular membership function labeled Sample 1 (Figure 4.5).

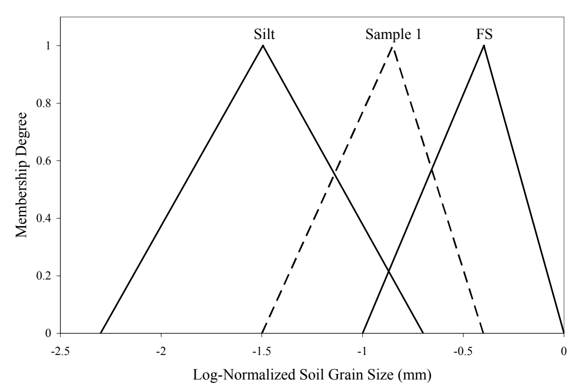


Figure 4.5. Expert's interpretation of borehole log data for sample 1 (Table 4.1), shown in reference to fuzzy sets Silt and Fine Sand (FS)

Training and experience would be needed to avoid arbitrary definitions. Therefore, experts rather than novices should create such fuzzy sets from qualitative descriptions.

In this example, most experts would agree that a sample described as a silty fine sand would be comprised predominantly of fine sand, although the influence of silt should shift the resulting fuzzy set slightly towards the Silt set. It is important to note that an expert-defined fuzzy number must have at least one value with full membership and any horizontal cuts through the fuzzy set must create closed and finite intervals along the domain. For example, the Silt fuzzy set in Figure 4.1 has at least one log grain size value (-1.5) with full membership, and a horizontal cut at the zero degree membership results in the closed and finite interval [-2.7, -0.3]. A horizontal cut through the fuzzy set at any other membership value would produce subsets of this interval. The manner in which this borehole log information is combined with the grain size distribution data is explained in the next section.

4.3. Reasoning

The approach taken to reasoning is that which is most common in interpreting fuzzy IF-THEN rules, the 'conjunction-based' model. Such models are desirable because they treat a rule as defining a fuzzy point in a fuzzy relationship between two or more variables. A collection of such rules becomes a series of fuzzy points that, together, define a complete imprecise relationship such that any feasible input value will produce a reasonable output value. This is the precise intention of this algorithm, where the two variables are soil grain size and hydraulic conductivity.

Once the possibility distribution is intersected with each fuzzy soil type, the degree to which each rule operates is determined by the maximum point of intersection between the possibility distribution and each predefined fuzzy set. Since the rules relate soil grain size to hydraulic conductivity, the membership of the horizontal ‘cuts’ of the soil grain fuzzy sets direct the ‘cuts’ of the hydraulic conductivity fuzzy sets. For example, if the PD intersects the Silt soil fuzzy set to a degree of 0.6, then the Silt hydraulic conductivity fuzzy set is cut at degree 0.6 membership. This membership at which maximum intersection occurs between the PD and a soil type fuzzy set dictates the degree to which the corresponding rule pertains to the soil sample represented by the PD. That this degree of intersection between the PD and soil grain fuzzy sets guides the degree to which the corresponding hydraulic conductivity fuzzy set operates should be intuitive. If one views the intersection as the degree to which the soil type applies to the given soil sample, then the hydraulic conductivity fuzzy set should only apply, or contribute, to the soil sample's true hydraulic conductivity value to the same degree (Figure 4.4).

Each rule will operate to a degree between zero and unity, depending upon how strongly each soil grain fuzzy set matches the soil sample's PD. What results are five hydraulic conductivity fuzzy sets, whose peaks are cut off at various degrees (right-hand side of Figure 4.4). These abridged fuzzy sets are combined to produce a fuzzy hydraulic conductivity value for the soil sample, using the union operator. The union operator works by assigning to each hydraulic conductivity value along the abscissa the maximum membership degree from the five cut sets for that hydraulic conductivity value (Klir and Yuan, 1995). The formula is

$$K_U(k) = \max_i \{K_i(k)\}, \quad \textbf{(4-1)}$$

for all k, $i = 1,...,5$. K_U represents the union of the five cut hydraulic conductivity fuzzy sets, K_i, resulting from the rules, and k are the hydraulic conductivity values along the abscissa over which this operation is performed. For example, given the five rules relating the soil types and hydraulic conductivities in Figure 4.1, the union operator in Equation (4-1) produces the fuzzy hydraulic conductivity value in Figure 4.6A based on the cut hydraulic conductivity fuzzy sets on the right-hand side of Figure 4.4.

The fuzzy input from the borehole log data provided by the groundwater professional resident at the drilling will produce another fuzzy hydraulic conductivity result (Figure 4.6B) when its soil grain fuzzy set is transposed over the soil type fuzzy sets, as had been done with the PD in Figure 4.4.

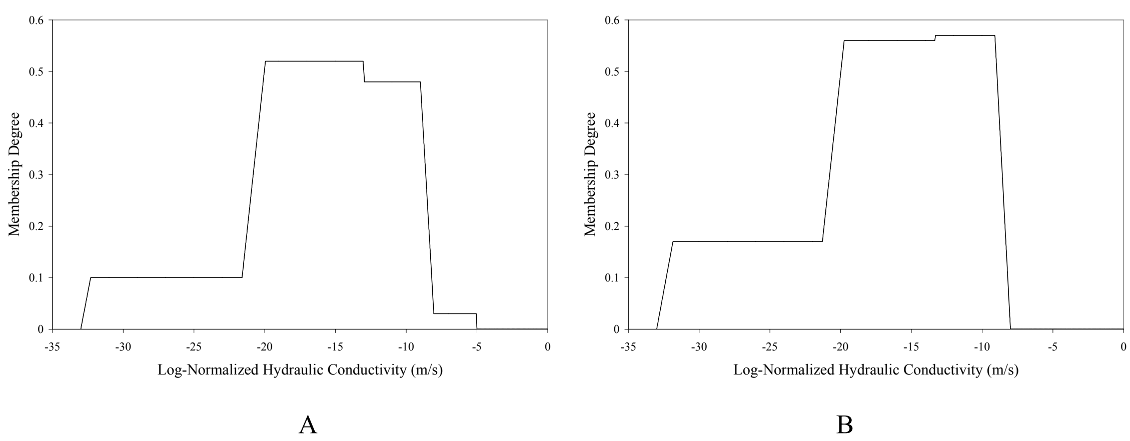


Figure 4.6. Fuzzy sets resulting from inputting A, the grain size distribution PD, and B, the fuzzy characterization of the borehole log data, into each fuzzy rule and aggregating the resulting cut hydraulic conductivity fuzzy sets using the union operator

The two hydraulic conductivity fuzzy results need to be combined to produce a single result. This is accomplished through a weighted mean, which aggregates the two fuzzy sets to produce a single representative fuzzy value. The weighted mean equation, acting on fuzzy sets, is a function on a domain element's (k) membership values ($\mu_K(k)$) from different fuzzy sets (K_i):

$$\mu_K(k) = \sum_i \sigma_i \cdot \mu_{K_i}(k), \tag{4-2}$$

for all hydraulic conductivity values k along the domain, where σ_i represents the weight given to the i^{th} fuzzy set. This weight measures how influential each information source (borehole log and soil grain distribution) will be to the final answer. These weights must have a value of between zero and unity, and they must sum to one.

Intuitively, the soil grain distribution data, an objective characterization of the soil sample should be weighted more heavily than the subjective characterization from the borehole log data. However, the latter should not be discounted completely, as it is viable data. As such, reasonable weights would be the following:

$$\sigma_{Borehole\ Log} = 0.4,$$

$$\sigma_{Grain\ Size\ Distribution} = 0.6.$$

Applying these weights and Equation (4-2), a fuzzy hydraulic conductivity value that represents both sources of hydraulic conductivity values, a weighted average of the fuzzy sets in Figure 4.6, is generated (Figure 4.7).

While this fuzzy result in Figure 4.7 provides a vague understanding of the estimated hydraulic conductivity value for the given soil sample, and is, itself, considered a fuzzy estimate, its use in geostatistical estimation is not obvious and requires non-traditional estimation procedures (Bardossy et al, 1989, 1990; Diamond, 1989).

Inasmuch as ordinary and simple kriging are considerably more prevalent and popular approaches to estimation than those founded in fuzzy logic, it is convenient if the hydraulic conductivity value that results from the above reasoning mechanism is a crisp value. To achieve this goal the fuzzy estimate must be defuzzified, or reduced to a crisp (non-fuzzy) value. The most common defuzzification method is called the Center of Gravity (COG) method. In this approach the fuzzy set is considered to be a uniformly thin sheet of metal, and the horizontal component of the center of gravity is the defuzzified value. The finite form of the COG equation is

$$COG = \frac{\sum_j \mu_K(k_j) \cdot k_j}{\sum_j \mu_K(k_j)} \tag{4-3}$$

where μ_K is the membership function of the hydraulic conductivity fuzzy set, and k_j are the finite number of domain values used in the calculation. In the case of the fuzzy set in Figure 4.7, the defuzzified value, or log hydraulic conductivity estimate for this soil sample is -16.85. The exponential applied to this value ($e^{-16.85}$) give a hydraulic conductivity estimate of 4.8×10^{-8} *m/s*. The defuzzified value will change with the discretization of the K-axis ($k_j \in K$) in Equation (4-3).

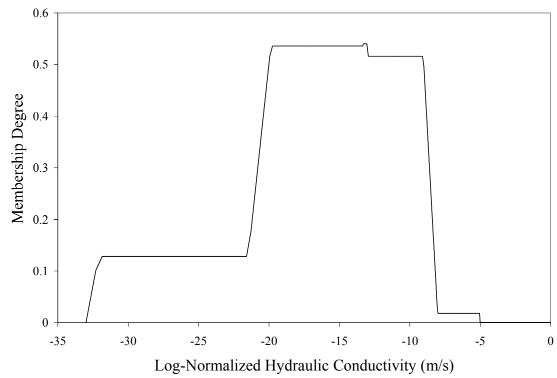


Figure 4.7. Aggregate fuzzy hydraulic conductivity estimate for sample 1, from the combination of the individual fuzzy results from grain size data and borehole log data (Figure 4.6)

4.4. Application

The reasoning described above was applied to four soil samples from the CIBA-Geigy Superfund site in Toms River, New Jersey, for which both soil grain distribution data and hydraulic conductivity values were available. At this site, many borehole logs were available, a number of which were accompanied by soil grain distribution curves. However, hydraulic conductivity measurements were few compared to the number of soil grain distribution curves. Since there were numerous sample locations with soil grain distribution curves but without hydraulic conductivity values, it was an ideal site for the practical application of the method described herein.

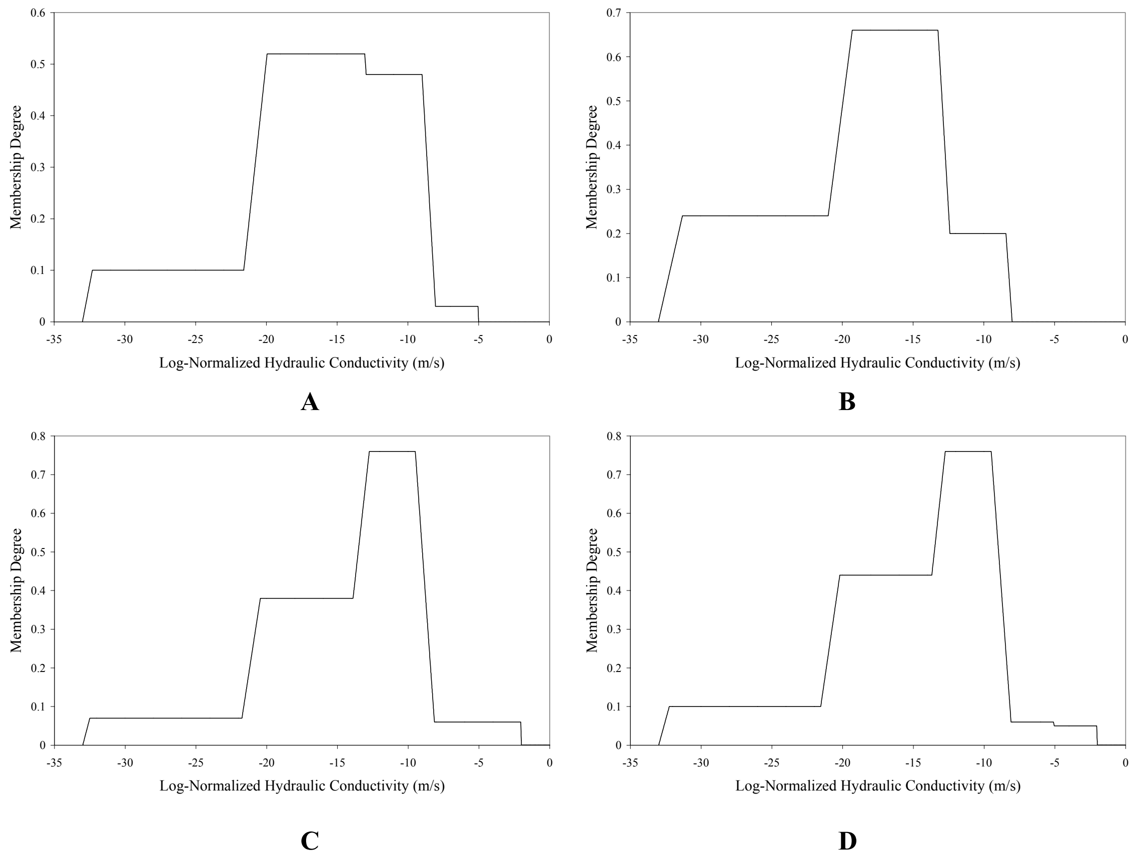


Figure 4.8. Fuzzy sets that result from inputting the transformed soil grain distribution from (A) sample 1; (B) sample 2; (C) sample 3; and (D) sample 4, into the fuzzy rule base

The particle size distributions of the four soil samples for which hydraulic conductivity measurements were available had been determined by both sieve and hydrometer methods, so that the soil grain distribution curves were complete. In cases where hydrometer analysis has not been performed on a soil sample, a geologist can use expert judgment to complete the curve.

The soil type and hydraulic conductivity fuzzy sets have been defined, therefore, the fuzzy relationship is already in place. The input, however, needed to be transformed into the appropriate possibility distributions for the inferential reasoning. Each of the four soil grain distribution curves (Figure 4.2) were changed into PDF form and then PD form (Figure 4.3) using the aforementioned transformation. Each of these inputs activated the five rules in different ways, resulting in a set of four fuzzy hydraulic conductivity output values (Figure 4.8).

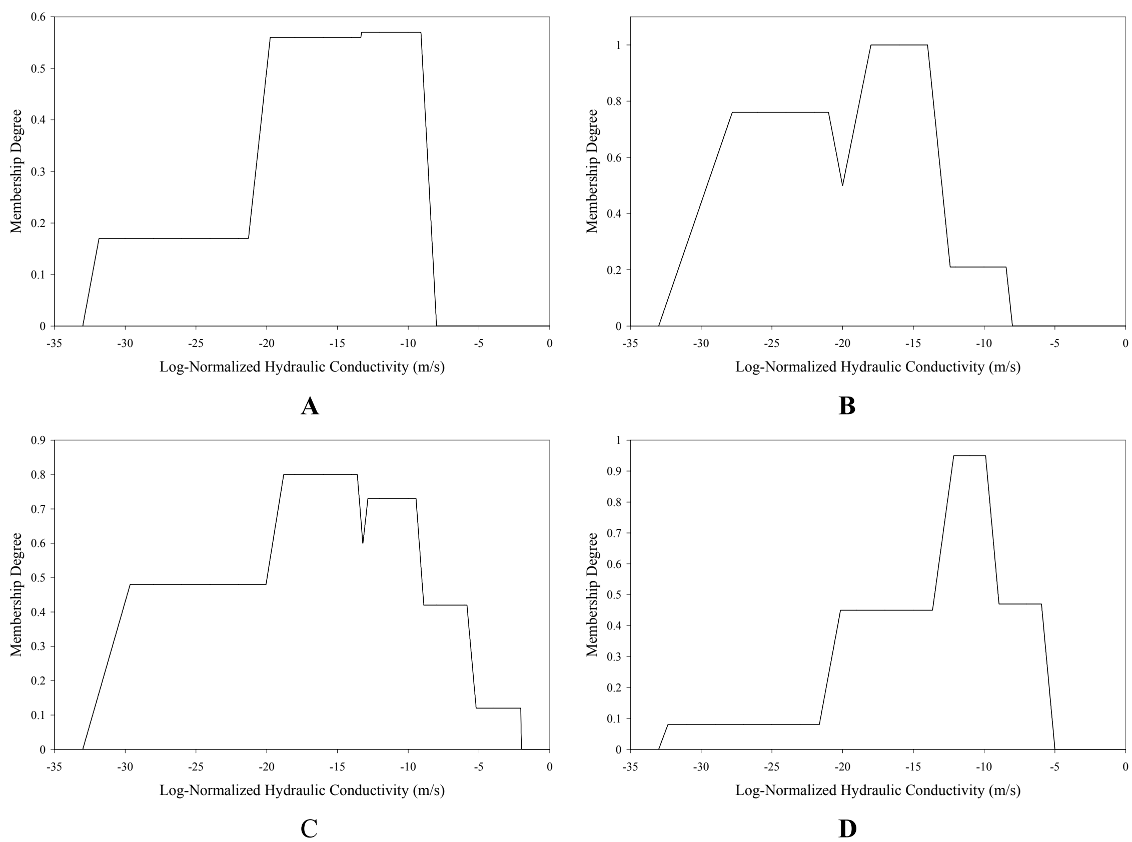


Figure 4.9. Fuzzy sets that result from inputting the fuzzy characterizations (Table 4.3) of the borehole log data from (A) sample 1; (B) sample 2; (C) sample 3; and (D) sample 4, into the fuzzy rule base

Similarly, an expert characterized the borehole log descriptions of these soil samples using fuzzy sets (Table 4.3), simply by applying his or her understanding of such descriptions and a basic understanding of triangular membership functions. These fuzzy set versions of borehole log data were also fed through the fuzzy rule base to produce another set of four fuzzy hydraulic conductivity estimates (Figure 4.9). Corresponding fuzzy hydraulic conductivity estimates (resulting from the grain size distribution curve and borehole log description of each soil) were combined (Figure 4.10) using the weighted mean approach, described in the previous section.

Table 4.3. Shapes and parameters of expert-provided fuzzy sets interpreted from borehole log descriptions

Sample	Fuzzy Set Shape	Parameters (log Grain Size)
1	Triangular	[-1.5, -0.85, -0.4]
2	Trapezoidal	[-2.8, -1.8, -1.5, -0.7]
3	Trapezoidal	[-2.3, -1.1, -0.85, 0.2]
4	Trapezoidal	[-1.35, -0.7, -0.45, 0]

4.5. Results and Conclusions

Crisp estimates were obtained by defuzzification of the fuzzy estimates (Table 4.4). TThe table also gives the measured hydraulic conductivity values for the samples. All estimated values are within two orders of magnitude of the corresponding measured value, which is considered reasonable inasmuch as hydraulic conductivity values may vary by up to five orders of magnitude if there are small differences in the percentages of the fine fractions (Koltermann and Gorelick, 1995).

Table 4.4. Hydraulic conductivity estimates from the combination of two types of soil data and the corresponding measurements from pump tests

Sample	Aggregate Estimate (m/s)	Measurement (m/s)
1	4.8×10^{-8}	2.5×10^{-7}
2	2.3×10^{-9}	1.6×10^{-9}
3	7.8×10^{-8}	1.8×10^{-6}
4	3.6×10^{-7}	1.6×10^{-6}

Also relevant is the interpretability of the results. With the fuzzy reasoning process, one can understand how the estimates are calculated. For instance, it is evident that the second hydraulic conductivity estimate in Table 4.5 is lower-valued than the other three estimates because the aggregate estimated hydraulic conductivity fuzzy set in Figure 4.10B has high memberships for lower hydraulic conductivity values.

Table 4.5. Hydraulic conductivity estimates from borehole log and grain distribution data

Sample	Borehole Log	Grain Distribution	Measurement
1	3.2×10^{-8}	6.5×10^{-8}	2.5×10^{-7}
2	1.1×10^{-9}	5.2×10^{-9}	1.6×10^{-9}
3	2.6×10^{-8}	3.5×10^{-7}	1.8×10^{-6}
4	7.5×10^{-7}	1.9×10^{-7}	1.6×10^{-6}

Thus, when this fuzzy set is defuzzified, the lower values of hydraulic conductivity are weighted more heavily than in the other samples' estimated hydraulic conductivity fuzzy sets in Figure 4.10. The higher memberships for lower-valued hydraulic conductivity values in this fuzzy set directly result from the significant overlapping of the grain size distribution and borehole log description fuzzy sets with the Clay soil type fuzzy set during inference.

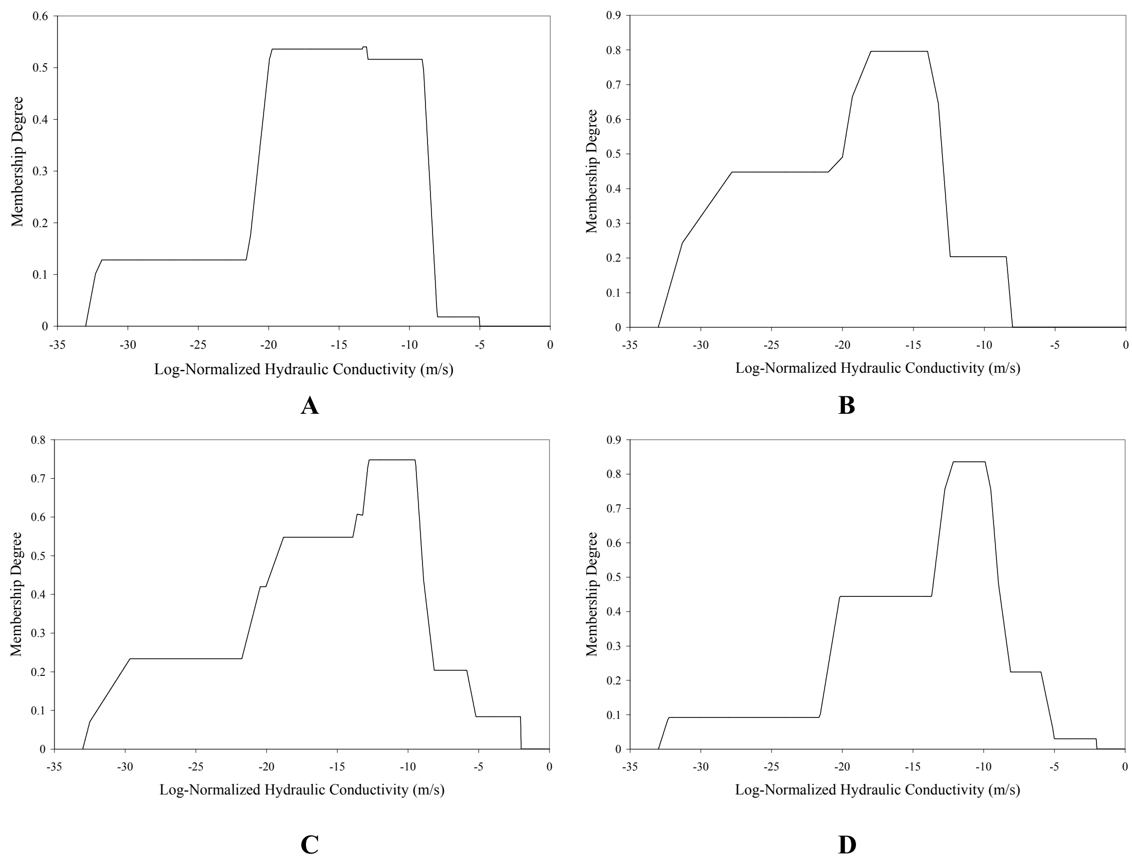


Figure 4.10. Fuzzy sets that result from the combination of the fuzzy sets from borehole log data and grain size distribution data for (A) sample 1; (B) sample 2; (C) sample 3; and (D) sample 4

In many applications, borehole log data are not only more prevalent than grain size data, but often are the only types of data available. Considering this, an investigation was made into the reliability of estimates resulting from just the expert fuzzy characterization of borehole log descriptions of the soil samples. For completeness, the scenario where only grain size distribution data are available was also evaluated. The results of these investigations are summarized in Table 4.5. In general, for this set of data, it is difficult to distinguish which form of data is more accurate. However, estimates stemming from both forms of data are still within two orders of magnitude of the measurements. Both types of data are not required to produce acceptable estimates. It has been shown that the use of only one data type in the fuzzy reasoning process can still lead to relatively accurate results.

Ideally, for the generation of realizations in groundwater modeling, measures of uncertainty should accompany hydraulic conductivity estimates. However, the issue of uncertainty in the results is not addressed since this is a budding area of fuzzy logic, and definitive answers regarding the uncertainty of the fuzzy estimates are unavailable. A thorough description of uncertainty of fuzzy sets would be requisite and is beyond the scope of this research.

The soundness of this approach to hydraulic conductivity estimation relied greatly upon the definition of the original soil type and hydraulic conductivity fuzzy sets that were used for reasoning. More accurate estimates could be produced had more data from the CIBA-Geigy Superfund site been available. With more data, the soil type and hydraulic conductivity fuzzy sets could have been constructed to respect the site data, thus producing better hydraulic conductivity estimates for that particular site. This would be true for any site with sufficient soil data. However, given that experts in hydrogeology defined these fuzzy sets, the reasoning process is considered sound and generally applicable to any site with reasonable accuracy.

References

Alyamani, M.S., and Sen, Z. (1993). "Determination of hydraulic conductivity from complete grain-size distribution curves." *Ground Water*, 31(4), 551-555.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). "Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31(3), 311-328.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990). "Kriging with imprecise (fuzzy) variograms." *Mathematical Geology*, 22(1), 63-94.

Brown, W., Groves, D., and Gedeon, T. (2003). "Use of fuzzy membership input layers to combine subjective geological knowledge and empirical data in a neural network method for mineral-potential mapping." *Natural Resources Research*, 12(3), 183-200.

Cheng, Q., and Agterberg, F.P. (1999). "Fuzzy weights of evidence method and its application in mineral potential mapping." *Natural Resources Research*, 8(1), 27-35.

Demmico, R. (2004). "Fuzzy logic and earth science: An overview." *Fuzzy Logic in* Geology, Demmico, R., and Klir, G., eds., Elsevier, Amsterdam.

Diamond, P. (1989). "Fuzzy kriging." *Fuzzy Sets and Systems*, 33(3), 315-332.

Domenico, P.A., and Schwartz, F.W. (1990). *Physical and chemical hydrogeology*, John Wiley & Sons, New York.

Dubois, D., and Prade, H. (1988). *Possibility theory: An approach to computerized processing of uncertainty*, Plenum Press, New York.

Fang, J. H., and Chen, H.C. (1997). "Fuzzy modeling and the prediction of porosity and permeability from the compositional and textural attributes of sandstone." *Journal of Petroleum Geology*, 20(2), 185-204.

Hsieh, B., Lewis, C., and Lin, Z. (2005). "Lithology identification of aquifers from geophysical well logs and fuzzy logic analysis: Shui-Lin Area, Taiwan." *Computers and Geosciences*, 31, 263-275.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey.

Koltermann, C.E., and Gorelick, S. M. (1995). "Fractional packing model for hydraulic conductivity derived from sediment mixtures." *Water Resources Research*, 31(12), 3283-3297.

Oussalah, M. (2000). "On the probability/possibility transformations: A comparative analysis." *International Journal of General Systems*, 29(5), 671-718.

Porwal, A., Carranza, E.J.M., and Hale, M. (2003). "Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping." *Natural Resources Research*, 12(1), 1-25.

Selley, R.C. (1985). *Elements of Petroleum Geology*, W. H. Freeman and Company, New York.

Sen, Z. (1992). "New statistical grain-size method for evaluating the hydraulic conductivity of sand aquifers – Comment." *Journal of Hydrology*, 130, 399-403.

Uma, K.O., Egboka, B.C.E., and Onuoha, K.M. (1989). "New statistical grain-size method for evaluating the hydraulic conductivity of sandy aquifers." *Journal of Hydrology*, 108, 343-366.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1(1), 3-28.

5: FUZZY KALMAN FILTERING OF HYDRAULIC CONDUCTIVITY

This paper addresses the utilization of expert information in groundwater hydrology. Given that a fuzzy model of hydraulic conductivity can be provided by appropriate experts for some spatial domain, crisp measurements can be used to update the fuzzy model using a fuzzy Kalman filter. The motivation behind this algorithm is to respect the relative integrity of each data source, while still taking advantage of both forms of data.

5.1. Introduction

The characterization of imprecise information in hydrogeological applications has frequently been achieved using fuzzy logic, a non-frequentist methodology; and, in a number of these applications, fuzzy sets are used to define imprecise values of a variable. Too often, though, fuzzy data are assumed to have the same relative reliability as crisp (non-fuzzy) data. The former are generally provided by an appropriate expert, and can be rather imprecise. Such data should be used to provide a starting point for most problems. In hydrogeological applications, crisp data are acquired through in- or ex-situ testing. If an initial fuzzy guess is available, the crisp data should update the fuzzy data, rather than be used in conjunction.

Consider the problem of estimating hydraulic conductivity. Such a task is as crucial to groundwater modeling as it is difficult to perform accurately. This difficulty is tied to the high cost identified with groundwater sampling, which often limits the number of hydraulic conductivity measurements at a given site. A small number of crisp measurements, combined with preferential sampling can make estimates, namely kriged estimates, relatively uninformed and unrealistic. Thus, data derived from expert knowledge to supplement the available measured data could be quite valuable in procuring hydraulic conductivity estimates of considerable accuracy.

However, in contrast to methods where fuzzy and crisp data are kriged together to develop a fuzzy hydraulic conductivity model for the spatial domain, the work described herein views fuzzy and crisp data in different lights, utilizing each toward an end respective of its relative reliability. Namely, expert-derived data are used to define a starting point for a hydraulic conductivity model, namely an imprecise idea of what the hydraulic conductivity field is for a site. Available crisp data update this fuzzy model using a fuzzy Kalman filter. Presuming that the fuzzy model is sufficiently accurate, the crisp data fine-tune the model and produce accurate, yet fuzzy, estimates of hydraulic conductivity. An example is provided to illustrate the usefulness of this algorithm.

5.2. Fuzzy Logic

Fuzzy logic provides a theory by which non-traditional sources of data can be characterized to facilitate use with more traditional data. Specifically, in this case, expert knowledge is considered to be a non-traditional data type that, without fuzzy logic, is difficult to accurately capture. The fuzzy set is the trademark feature of fuzzy logic and is the mechanism by which imprecise human input is reflected.

A fuzzy set is defined by a membership function, $\mu(x)$, which, for the variable of interest, X, stipulates to what degree certain variable values belong to a given imprecise notion. For instance, it is possible for an appropriate expert to claim with certainty, based upon some evidential motivation, that the hydraulic conductivity at some location is *about 3* millidarcies. The expert is certain but imprecise.

Without fuzzy sets, defining the notion of *about 3* (Figure 5.1) would be a dubious enterprise. In fact, Figure 1 is an example of a particular type of fuzzy set, called the fuzzy number. A fuzzy number is a fuzzy set that satisfies three criteria (Klir and Yuan, 1995):

(i) The fuzzy set must have at least one value with full membership (normality);

(ii) α-cuts (see below) are closed intervals for $0 < \alpha \leq 1$;

(iii) The support of the fuzzy set, the zero-level cut, must be bounded.

It should be evident, from the *about 3* example that fuzzy numbers are the imprecise counterpart of traditional (crisp) numbers.

Definitions of fuzzy sets are quite subjective. In scientific applications, only appropriate experts should be petitioned to provide data in the form of fuzzy sets. By utilizing expert knowledge in scientific applications, it is expected that variability in fuzzy set definition, due to subjectivity, is reduced to an acceptable amount.

An important feature of the fuzzy set is the α-cut. Membership values, along the vertical axis in Figure 1, are also referred to as α-values. Horizontally cutting a fuzzy set at a given α-value produces an interval of values, called an α-cut.

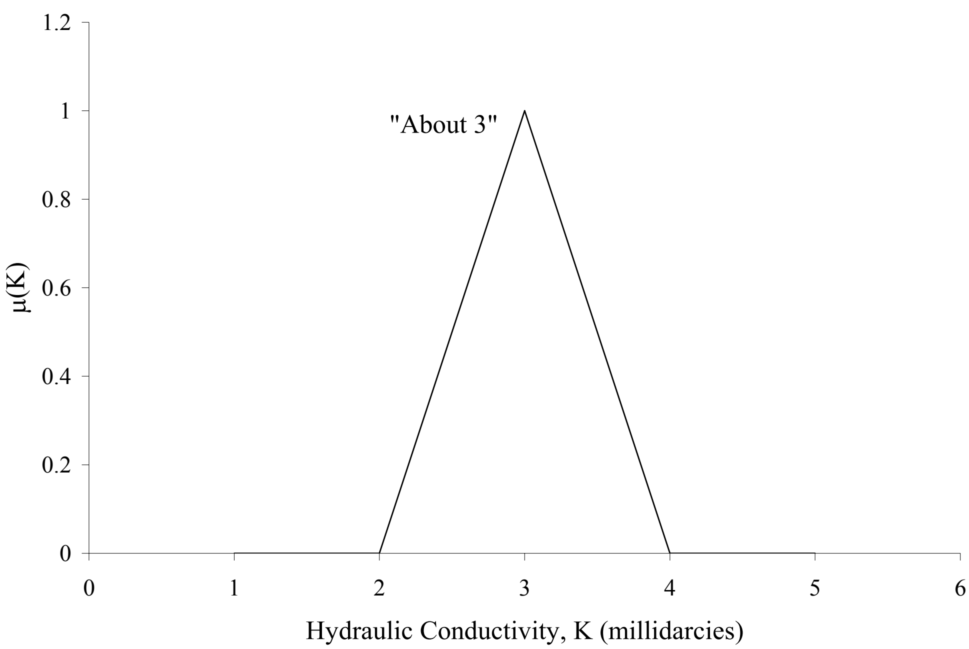


Figure 5.1. One expert's definition of "About 3," defined for hydraulic conductivity

Fuzzy numbers can be operated upon using the same equations that operate on crisp numbers. This is possible through the extension principle (Zadeh, 1965). The extension principle retrofits traditional formulae for use with fuzzy numbers.

5.3. Fuzzy Variogram

For a given site, knowledgeable hydrogeologists can provide their impression of hydraulic conductivity values as fuzzy numbers at various locations throughout the spatial domain. Geostatistical kriging is used to create estimates at regular intervals throughout the entire domain. As long as the measurements are fuzzy numbers, fuzzy kriging is required. The kriging approach uses weighted sums of available fuzzy "measurements" to calculate fuzzy estimates at un-sampled locations.

The fuzzy kriging method used herein is rather straightforward, and requires only a slight change to the determination of the sample semivariogram (Bardossy et al, 1989), which is a representation of the spatial covariance structure. This slight change is, in fact, the calculation of fuzzy sample semivariance values, rather than traditional crisp values. The most important step in constructing the sample variogram is the calculation of the squared differences between pairs of measurements. However, one must apply the extension principle in order to find the fuzzy squared difference between a pair of fuzzy measurements.

For each pair of measurements, there is a calculated squared difference and a known physical distance between their measurement locations. Measurement pairs with sufficiently similar separation distances are grouped together, their squared differences and separation distances averaged. The resulting averaged distance, h, and average fuzzy squared difference, $\gamma(h)$, are plotted to produce the sample variogram. To this sample semivariogram is fit a crisp model semivariogram selected from a prescribed set of possible positive definite models. The significance of this is that the weights calculated by the kriging equations are crisp values. Multiplying the crisp weights by the fuzzy values produces a fuzzy estimate at each desired location in the spatial domain.

5.4. Kalman Filter

The Kalman filter is a series of predicting and updating steps generally employed to estimate the state of a dynamic system, such as the location of some projectile object. The filter, named for Rudolf Kalman, relies upon measurements of the state variable (location, in the projectile object scenario) to perform the updating steps, while an understanding of the processes that govern the dynamics are requisite for prediction.

Typically, the Kalman filter algorithm begins with an educated guess as to what the state of the system is. After acquiring measurements of the state variable, the educated guess is refined (updating step). Employing an understanding of the system dynamics, the state of the system is forecasted for the next time step (prediction step). A new set of measurements incites more updating, and the process ensues until measurements are no longer made (Welch and Bishop, 2001).

However, in this particular application, the system is static (no prediction step required), only one updating step is performed for the one set of crisp measurements, and the educated guess that incites the algorithm is the set of fuzzy hydraulic conductivity estimates that result from the fuzzy kriging. Thus, crisp measurements are used to update the fuzzy model in this fuzzy Kalman filter.

There has been a significant amount of research into the use of fuzzy logic in Kalman filtering (Chen et al, 1998; Chiang, 2003). However, no evident research has applied the extension principle strictly to the updating step, especially in a hydrogeological setting.

In a Kalman filter, the prior model ($\hat{\boldsymbol{y}}^-$) is updated ($\hat{\boldsymbol{y}}$) by adding to the prior model values a weighted difference between the measurements and the prior model values at those measurement locations:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}^- + \kappa(\mathbf{z} - H\hat{\mathbf{y}}^-) \tag{5-1}$$

where κ is the Kalman gain, a matrix that minimizes the covariance of the estimation error, $\boldsymbol{z}$ is a vector of measurements, and H is a matrix of ones and zeros that pulls the prior model values at measurement locations out of $\hat{\boldsymbol{y}}^-$. The algorithm for extending this equation requires individually updating each element of $\hat{\boldsymbol{y}}^-$ by the measurement values. Thus, there are n implementations of the above equation, one for each value updated. For each element $\hat{y}_i^-$ ($i = 1,..,n$) of $\hat{\boldsymbol{y}}^-$, Equation (5-1) becomes

$$\hat{y}_i = \hat{y}_i^- + \kappa_i(\mathbf{z} - H\hat{\mathbf{y}}^-) \tag{5-2}$$

where κ_i represents the i^{th} row of the Kalman gain matrix. The extension principle is applied to Equation (5-2) in order to produce an a posteriori fuzzy number estimate. Repeating this for every $\hat{y}_i^-$ in $\hat{\boldsymbol{y}}^-$ provides one with the complete updated fuzzy model.

5.5. Example

The example application provided here is comprised of synthetic values. To begin, an appropriate expert provides fuzzy hydraulic conductivity values at sparsely sampled areas. Possible information the expert could use to opine upon the hydraulic conductivity values might pertain to soil type, geologic history, or knowledge of a similar site. Regardless of the evidential motivation behind the expert "measurements," it is assumed that they were provided independent of the crisp measurements at sampling locations. Figure 5.2 shows a plan view of the site, signifying the locations of fuzzy expert values with triangles and measurement locations with circles. Units of length are given in meters. All samples are taken at the same depth in this two-dimensional example.

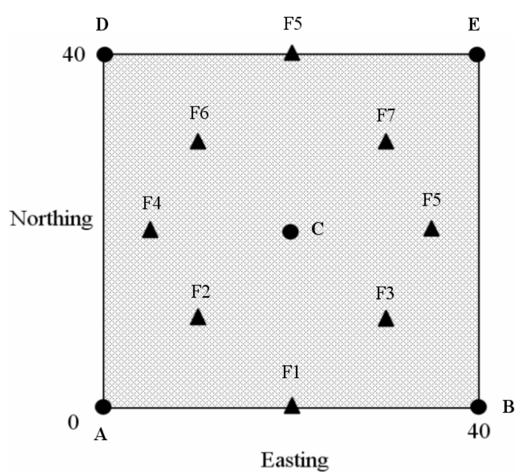


Figure 5.2. Plan view of example problem; triangles are fuzzy value locations, circles are crisp measurement value locations

The fuzzy numbers used to describe the expert's knowledge of the hydraulic conductivity values are all defined by triangular membership functions. As such, the fuzzy numbers can be defined by three parameters: 1) the lower bound of the base of the fuzzy set, 2) the upper bound of the base, and 3) the median value (value with full membership). The locations and said parameters of the fuzzy hydraulic conductivity values are given in Table 5.1, while the locations and values of the crisp measurements are provided in Table 5.2.

Table 5.1. Fuzzy number IDs, locations, and parameters

ID	Easting	Northing	Log Lower Bound (m/s)	Log Median (m/s)	Log Upper Bound (m/s)
F1	20	0	-12	-11	-10
F2	10	10	-17	-15	-14
F3	30	10	-9	-7	-6
F4	5	20	-13	-12	-11
F5	35	20	-20	-19	-18
F6	10	30	-10	-8	-7
F7	30	30	-9	-8	-7
F8	20	40	-5	-3	-2

Table 5.2. Crisp measurement IDs, locations, values and errors (variance)

ID	Easting	Northing	Log Measurement (m/s)	Variance (m^2/s^2)
A	0	0	-9	3
B	40	0	-6	4
C	20	20	-25	4
D	0	40	-3	4
E	40	40	-4	3

A sample semivariogram was constructed from the available fuzzy hydraulic conductivity measurements. This semivariogram, with fuzzy semivariance values for different separation distances, *h*, is shown in Figure 5.3. The α-cuts plotted for each fuzzy number are $\alpha = \{0, 0.5, 1\}$, the actual plotted values being the upper and lower bounds of each fuzzy set's α-cuts. The model semivariogram fit to this sample is Gaussian with a sill of *30* and a range of *40*. It is superimposed upon the sample variogram in Figure 5.3. No nugget is given to the model, because no "measurement" variance is assumed for the expert's input.

With the crisp model variogram defined, the fuzzy values are kriged and a field of fuzzy hydraulic conductivity estimates is generated. This fuzzy model is best represented by a series of contour plots for various α-cuts, as in (Bardossy et al, 1990). These contour plots, for $\alpha = \{0, 1\}$ are provided in Figure 5.4.

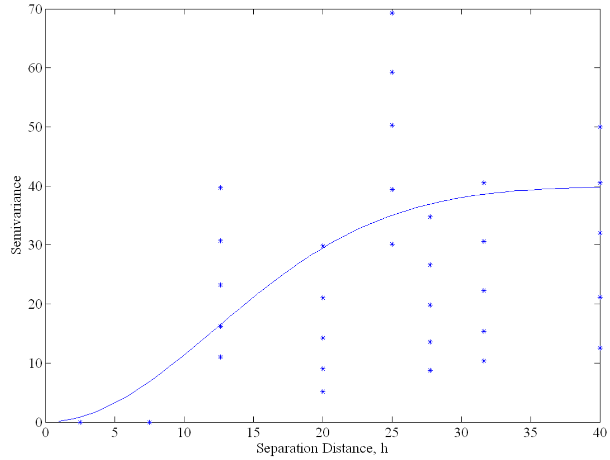


Figure 5.3. Fuzzy sample and crisp model variogram, as prescribed by the fuzzy hydraulic conductivity data

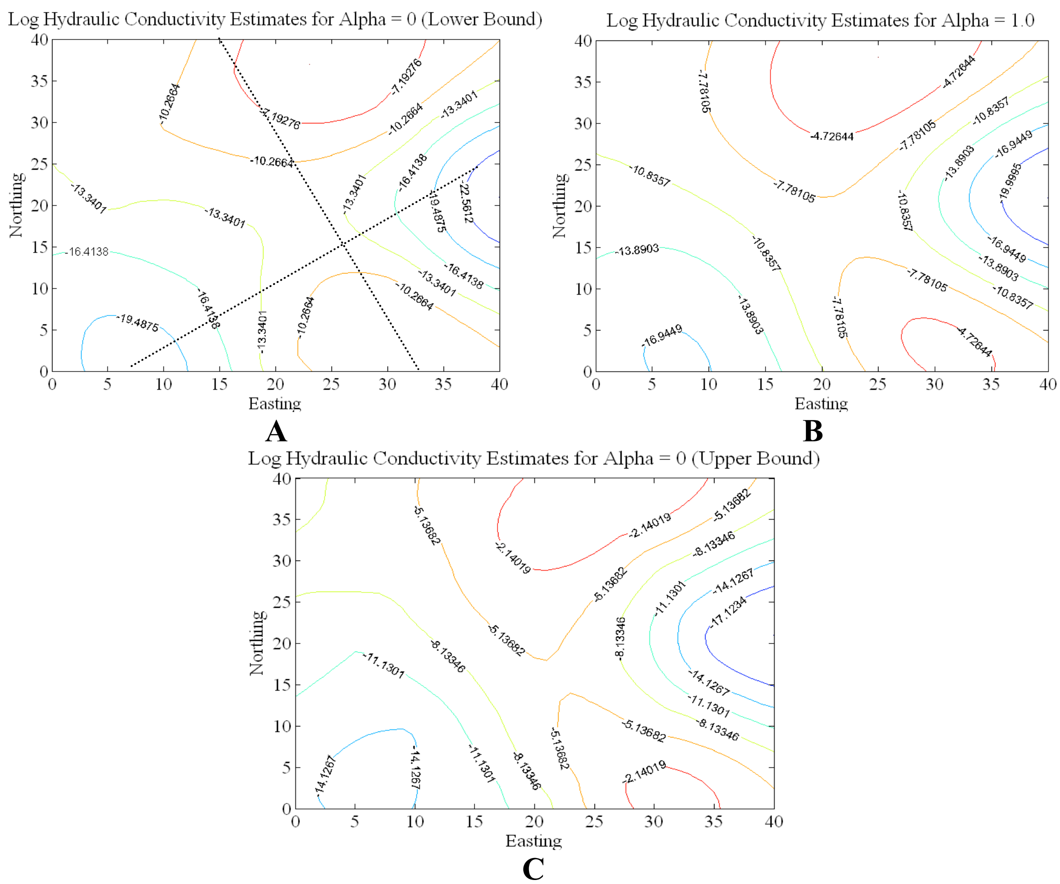


Figure 5.4. Contour plots of kriged fuzzy estimates of log hydraulic conductivity for (A) lower bound of the zero α-cut, (B) the median values (α = 1), (C) upper bound of the zero α-cut

Having the fuzzy hydraulic conductivity model in hand, the next step is to incorporate the crisp measurements in order to update the fuzzy model. Applying the updating step of the fuzzy Kalman filter, the fuzzy model is updated by those crisp measurements in Table 5.2. The resulting contours corresponding to those in Figure 5.4 are shown in Figure 5.5.

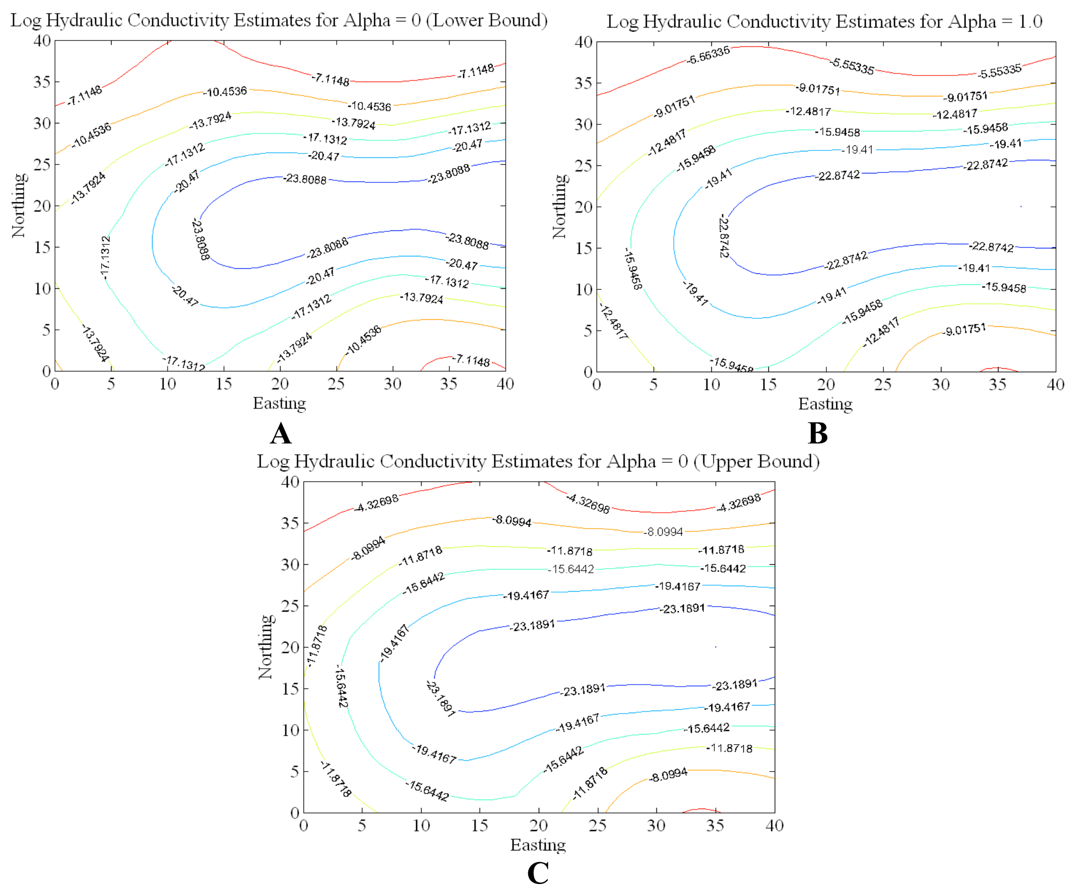


Figure 5.5. Contour plots of updated fuzzy estimates for (A) lower bound of the zero α-cut, (B) the median values ($\alpha = 1$), (C) upper bound of the zero α-cut

5.6. Results and Conclusions

By contrasting the corresponding contours in Figure 5.4 and Figure 5.5, two differences are evident. The first difference is intuitive. When crisp measurements are introduced to update the fuzzy estimates, the fuzziness of the estimates (how wide the α-cuts are) can only be reduced since the crisp measurements contain no fuzziness, making the fuzzy model more precise.

Secondly, in the contour plots of Figure 5.4, there are axes of symmetry, roughly illustrated by the dashed black lines in Figure 5.4A. This symmetry, however, is missing from the corresponding contour plot in Figure 5.5A. This is due to the influence the crisp measurements have upon the final estimates, presumably defining a posterior model that more accurately represents the true trends present at the site.

At measurement locations, the fuzzy estimates are made very precise and close in value to the corresponding crisp measurements. However, unlike the traditional Kalman filter algorithm, these posterior estimates at measurement locations are not made identical in value to the crisp measurements. This is a byproduct of the difference in data types between the measurements and prior estimates, and is not considered an inconsistency. As Figure 5.6 shows, the posterior fuzzy estimates are quite similar to the corresponding crisp values.

Presented here was an alternative to using fuzzy information in conjunction with crisp data in a hydrogeological setting. By using Kalman filtering, each data type is used in a manner that respects its relative credibility. What results is a rather precise fuzzy field that presumably captures the trends inherent in the model. This fuzzy model may be defuzzified to create a crisp hydraulic conductivity field for input into traditional groundwater flow models, or may be left fuzzy to be used in fuzzy-friendly groundwater flow models (Dou et al, 1995; Dou et al, 1999). Further investigation is provided in the next chapter.

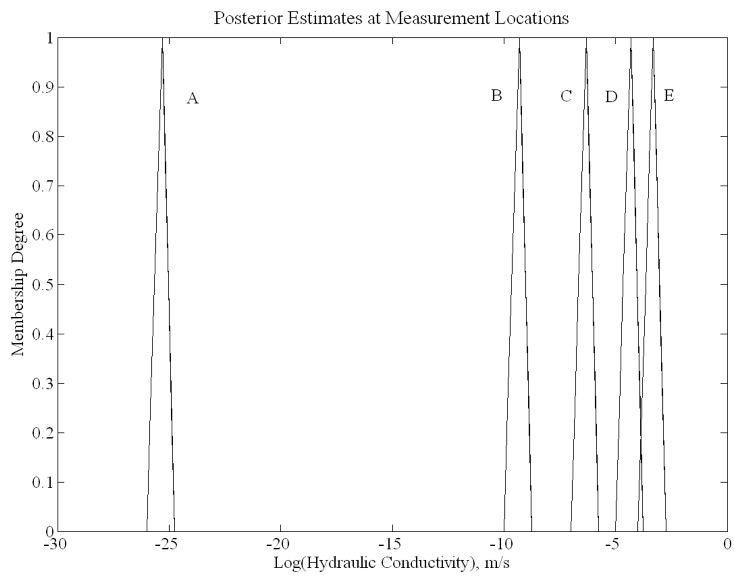


Figure 5.6. Posterior estimates at all measurement locations; Measured values are: -25(A), -9(B), -6(C), -4(D), -3(E)

References

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990). "Kriging with imprecise (fuzzy) variograms II: Application." *Mathematical Geology*, 22(1), 81-94.

Chen, G., Xie, Q., and S. Shieh, L.S. (1998). "Fuzzy kalman filtering." *Information Sciences*, 109(1-4), 197-209.

Chiang, H. H., Wu, S.J., and Lee, T.T. (2003). "Application of fuzzy sets theory in process data filtering." *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems*, 325-330.

Dou, C., Woldt, W., Dahab, M., and Bogardi, I. (1995). "Steady state groundwater flow simulation with imprecise parameters." *Water Resources Research*, 31(11), 2709-2719.

Dou, C., Woldt, W., Dahab, M., and Bogardi, I. (1997). "Transient ground-water flow simulation using a fuzzy set approach." *Ground Water*, 35(2), 205-215.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey.

Welch, G., Bishop, G. (2001). “An introduction to the kalman filter.” *SIGGRAPH 2001 course 8*, *In Computer Graphics*, *Annual Conference on Computer Graphics & Interactive Techniques*. ACM Press, Addison-Wesley, Los Angeles, CA, USA.

Zadeh, L. (1965). “Fuzzy sets.” *Information and Control*, 8(3), 338-353.

6: KALMAN FILTER UPDATING OF POSSIBILISTIC HYDRAULIC CONDUCTIVITY

Accepted for Publication in Journal of Hydrology on March 5, 2008;

Authors: Ross, J., Ozbek, M.M., and Pinder, G.F.

Preferential sampling regimes too often leave hydrogeological site characterization in need of hydraulic conductivity data. Furthermore, faith in a kriged hydraulic conductivity field (K-field) wanes when viable data are few. A straightforward solution to this problem is the application of expert knowledge to the identification of hydraulic conductivity at specified locations. However, the nature of this supplementary data suggests that its use should be limited to the initial estimation of a K-field. Available hydraulic conductivity measurements are used to update the initial estimate of the K-field in a fuzzy Kalman filtering procedure. The results are fuzzy and can be defuzzified into traditional crisp estimates. An application of these protocols to the CIBA-Geigy site in New Jersey is provided to illustrate the ease of the procedure and the impact of viable expert input.

6.1. Introduction

The characterization and propagation of uncertainty in hydrogeological applications is important in groundwater flow system modeling. Though inappropriate groundwater model selection is a significant source of model uncertainty, so to is the uncertainty surrounding hydraulic conductivity (K) that is introduced primarily via measurement and interpolation.

Traditionally, uncertainty in hydrogeology has been framed in probabilistic terms. Nevertheless, fuzzy set and possibility theories, as well as fuzzy logic, have recently provided more intuitive means of uncertainty characterization (Zadeh, 1965; Zadeh, 1978). In fact, the introduction of such frameworks has revealed new sources and enhanced upon traditional sources from which viable hydrogeological data may be extracted, including the facilitation of direct expert knowledge extraction (Bardossy et al, 1989; Piotrowski, 1996). In this latter approach, an expert opines upon the value of hydraulic conductivity at specific locations, utilizing only an implicit understanding of the particular site's characteristics drawn from available field information, education and experience, information which cannot otherwise be explicitly and mathematically defined. The use of possibility theory to accomplish this task is considerably more flexible and intuitive than permitting the petitioned expert to frame his/her knowledge in a probabilistic framework.

The hydrological site characterization process using non-traditional data forms has two distinct issues. Hydraulic conductivity estimates determined through the site characterization process using non-traditional data forms can incorporate soil physical measurements as well as expert knowledge. These different data sources possess different degrees of integrity. While the physically-based methods are reliant upon relatively precise hard data, expert knowledge is a very imprecise data source, perhaps implying a lower degree of quality.

This paper suggests an approach to hydrogeological site characterization that not only accommodates the use of imprecise data, but also sets guidelines governing how data from physical measurements and expert knowledge can be properly utilized. Specifically, the authors propose that 1) hydraulic conductivity measurements resulting from direct expert input be interpolated using fuzzy kriging (Bardossy et al, 1989; Bardossy et al, 1990a,b) to produce an initial estimate of the K-field; 2) a static fuzzy Kalman filter (Ross et al, 2006) be used to update this first estimate with available measurement data to produce the final K-field for a site. This fuzzy Kalman filtering approach is considered to be an alternative to Bayesian fuzzy kriging (Bandemer and Gebhardt, 2000). In the latter, experts are required to frame their knowledge as expectation values and covariance matrices of trend parameters, rather than membership functions and possibility distributions representing uncertain hydraulic conductivity values (Figure 6.2).

In the next section, the application of possibility theory to hydrogeological site characterization is discussed. Section 6.3 highlights the major implications of fuzzy kriging. The fuzzy Kalman filter is introduced in Section 6.4, followed by a field application of the algorithm presented herein, in Section 6.5. Introductions to pertinent tools of fuzzy set and possibility theories are provided in the Appendix.

While fuzzy sets and fuzzy numbers are inherent in the groundwater flow modeling process, there has been no discussion, as far as the authors are aware, regarding the best use and propagation of fuzzy data relative to the data source. Such a discussion follows.

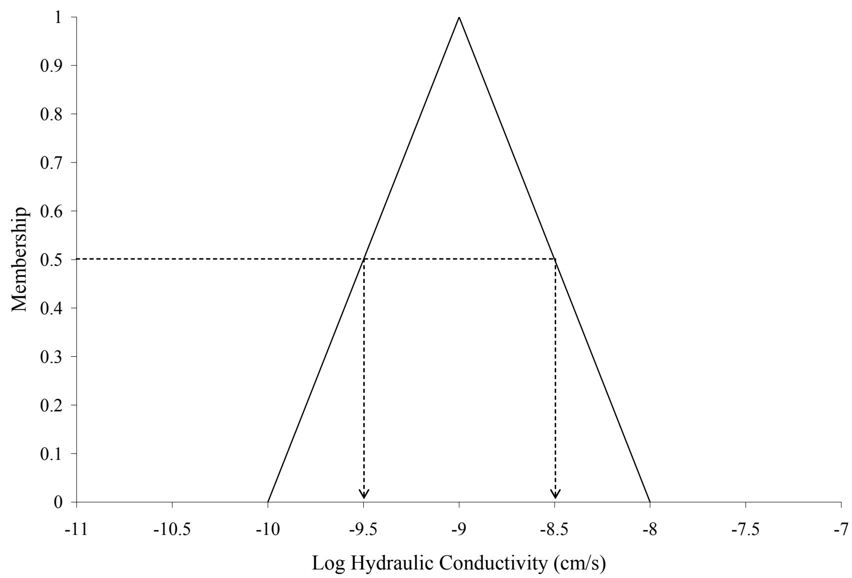


Figure 6.1. This fuzzy set is a possible characterization of the expert's notion of *low* log hydraulic conductivity (cm/sec). The 0.5 level α-cut produces the interval [-9.5, -8.5].

6.2. Expert Judgment

As noted, data sources in hydrogeological site characterization can be partitioned into two groups: physical measurement and expert knowledge. The former data source produces hydraulic conductivity values based upon direct measurements of that variable or through translations of secondary variable measurements, while the latter dictates that an expert armed with geological knowledge of a given site, can identify, via his/her understanding, the hydraulic conductivity values at various spatial locations. Hydraulic conductivity data derived from expert knowledge may be considered somewhat less reliable than physics-based data and perhaps should be treated as such.

Traditionally, hydraulic conductivity measurements are made via indirect methods. Indirect methods include both inverse approaches, where head measurements lead to hydraulic conductivity values, and functional transformations of secondary data, most often derived from soil sample analyses (Alyamani and Sen, 1993; Uma et al, 1989; Sen, 1992; Masch and Denny, 1966). Estimated K-fields rely upon the number and location of hydraulic conductivity measurements made throughout the site under study. However, the spatial distribution of measurement locations is often not conducive to accurate spatial estimation regionally, and measurements are often too few for the accurate estimation of a K-field (Piotrowski, 1996).

Numerous approaches have been devised to correct some of the deficiencies resulting from sparse sampling (Gotway and Hartford, 1996; Odeh et al, 1999; Asli and Marcotte, 1995). Of these, cokriging (Isaaks and Srivastava, 1989) is most often encountered in practice. Though the use of secondary data in this context improves an understanding of the spatial correlation of the site, it requires the modeling of multiple semi-variograms and may not positively impact the quality of the values being interpolated. Moreover, in cases where the primary variable is hydraulic conductivity (as it is in this research) and secondary data results from soil analysis, certain information must either be discarded or simplified. For instance, grain size distributions may need to be reduced to a single representative value and qualitative data resulting from borehole investigations must be transformed into crisp soil classes. As such, the quality of primary variable estimates is arguably more positively impacted by the availability of more primary data rather than secondary data.

A much more straightforward solution to the problem of too few measurement data is the application of expert intuition to the identification of hydraulic conductivity values at various locations throughout the spatial domain. In this work, an expert is any agent familiar with the hydrogeological aspects of the site of interest. It is assumed here that such an appropriate expert could roughly estimate, without sampling, the approximate hydraulic conductivity value or representative soil grain size at any number of locations throughout the spatial domain. Expert knowledge applied toward this end is arguably quite valuable. With an assumed minimal loss of accuracy, a well-informed expert can provide much the same information as would be obtained through the installation of a pumping or observation well.

Obviously, varying amounts of uncertainty cloud the expert-provided hydraulic conductivity values. Traditionally, uncertainty in hydrogeological applications is characterized via probability theory. However, humans do not think probabilistically and generally, the delineation of a probability density function (PDF) from scratch is an arduous task (Ganoulis, 1996; O'Hagan and Oakley, 2004), requiring the expert to know what mean, variance and distribution model together define the best stochastic (random) number. Because humans do not explicitly characterize their understanding with these parameters in mind, possibility distributions seem to be a logical choice for uncertainty characterization. The use of possibility distribution in this circumstance is reasonable since they are easy to define and implicitly bracket the PDF that best represents the stochasticity of hydraulic conductivity at a given location (Klir and Yuan, 1995).

In sparsely sampled domains, expert-provided fuzzy number or possibilistic hydraulic conductivity values can be applied at locations that would minimize the overall uncertainty, or kriging variance, in a kriged K-field (Piotrowski, 1996). Also, the expert can apply his/her knowledge to identify zones of anomalously high or low hydraulic conductivity. Consider Figure 6.2, showing a horizontal slice of the subsurface. The black dots represent well locations, where hydraulic conductivity measurements are available.

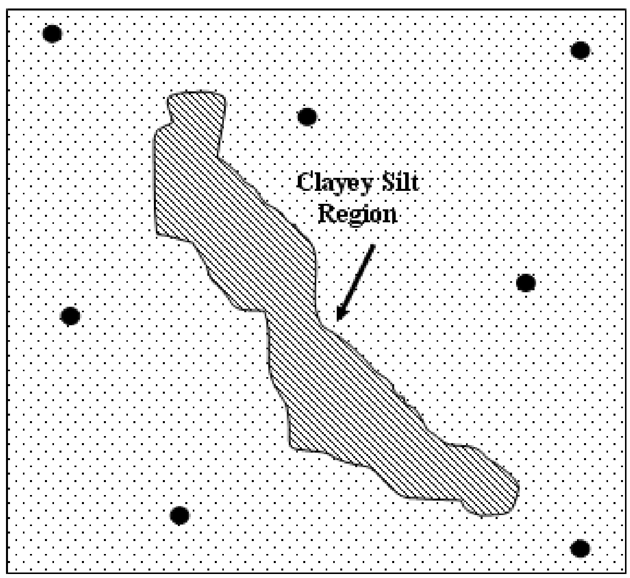


Figure 6.2. In this image of a hypothetical subsurface slice, the silty clay unit (diagonal lined region) has gone unpenetrated (and unsampled) by the pumping and/or observation wells (black dots)

Notice that in the clayey silt region, no measurements are made, though this is a region of potentially low hydraulic conductivity. An appropriate agent, armed with an implicit understanding of the site geology may know that such a region exists; and, assuming additional wells cannot be drilled, using fuzzy numbers, an expert can identify, at one or more locations in this region, hydraulic conductivity values representative of that soil type. Alternatively, borehole investigations may be available in that region; and a hydrogeologist can deduce approximate hydraulic conductivity values from the qualitative borehole information.

The value of expert insight is not limited to hydraulic conductivity data. In the previous example referencing Figure 6.2, the petitioned expert may feel that it is much more straightforward and intuitive to identify the soil grain size representative of the unsampled geologic units than a value of hydraulic conductivity. In essence, this entails the specification of a fuzzy number grain size. A mechanism by which such information can be translated into hydraulic conductivity data is provided by Ross et al (2007). This approach produces hydraulic conductivity estimates from fuzzy soil grain size data by means of a fuzzy rule base.

Though the benefit to such K-value specification is the identification of hydraulic conductivity at any particular location throughout the spatial domain with reasonable accuracy and precision, hydraulic conductivity values (both fuzzy and crisp) strictly based upon an expert's insight regarding the site may, in fact, be flawed. For instance, an expert-based estimate at a particular location may not be a fuzzy version of that location's true hydraulic conductivity value. It can be assumed without risk, however, that the expert-derived values are, at the very least, similar to the corresponding true values. Consider a case where the expert defines the fuzzy number in Figure 6.1 as the log K-value at some location. In reality, the unknown true log hydraulic conductivity value may be *-10 cm/s*, though the expert claimed a value of *-9 cm/s* was most plausible. Note that two orders of magnitude of variation in hydraulic conductivity, as shown in Figure 6.1, is a reasonable scenario (Druschel et al, 2006; Mathon et al, 2008).

Based upon the potential error associated with estimates of hydraulic conductivity data and soil data provided by experts, these expert-based estimates should be treated quite differently from measurements achieved through more traditional means. Specifically, expert-provided data are essentially educated estimates, and an estimate K-field kriged from such data is at best, a first approximation of a site's true K-field. Any available measurements resulting from both direct and indirect approaches should be used to update the best first estimate of the K-field. Fuzzy kriging is used to perform spatial interpolation of the expert-provided data and provide a possibilistic K-field, while a fuzzy Kalman filter updates the approximate K-field.

6.3. Interpolating Fuzzy Measurements

6.3.1. Variogram Calculations

The expert-provided fuzzy and crisp hydraulic conductivity values can be spatially interpolated using a fuzzy kriging algorithm. Fuzzy kriging is the result of applying the extension principle to the standard kriging system. By this algorithm, fuzzy number values are interpolated spatially to produce possibilistic estimates everywhere throughout the spatial domain. Via kriging each log hydraulic conductivity estimate, $K(x_j)$, is a weighted linear combination of the b available values $K(x_i)$, ($i = 1,\dots, b$),

$$K(x_j) = \sum_{i=1}^{b} w_i \cdot K(x_i) \qquad \textbf{(6-1)}$$

In Equation (6-1), the individual weights w_i ($i = 1,\dots, b$) depend upon the spatial continuity of the field and are recalculated for each estimate $K(x_j)$. The spatial continuity, or covariance, is measured by the semivariogram, which is a plot of distance h versus semivariance $\gamma(h)$

$$\gamma(h) = \frac{1}{2 \cdot N_h} \sum_{(x_a, x_b) | h_{ab} \approx h} (K(x_a) - K(x_b))^2 \qquad \textbf{(6-2)}$$

The separation distance h_{ab} in Equation (6-2) is the spatial distance between two measurements $K(x_a)$ and $K(x_b)$. The number of values separated by a distance h is denoted N_h. As the distance between two values increases, their correlation decreases, and the semivariance increases. Because available measurements are finite, the semivariance calculated from these values generates a discontinuous relationship and must be modeled in order to provide a smooth curve representing the semivariance for any separation distance and to produce a positive definite covariance matrix.

Variogram models are the continuous positive definite functions fit to sample variograms in order to provide a complete understanding of a site's spatial continuity. However, measurements are often too few to instill confidence in any one variogram model; the larger the number of measurements, the more reliably a variogram model can be fit to the sample (Webster and Oliver, 1992). Kravchenko (2003) and Wang and Qi (1998) have illustrated that the number of available measurements and their locations can have significant effects upon the sample variogram's functional form and, in turn, upon the accuracy of the model variogram. Few and randomly sampled measurements produce chaotic sample variograms that prove difficult to model.

For example, Figure 6.3 shows the locations of 17 hydraulic conductivity measurements. The log measurements do not vary much with a mean log measurement of -1.628 and a variance of 0.246. However, the sample variogram produced from these measurements (Figure 6.4) does not suggest any obvious model, as would be expected with so few measurements. A conservative solution would be a pure nugget model, which admits no spatial correlation. However, this model fails to respect the spatial correlation implicit in Figure 6.4 at low lag distances.

Despite this, in practice, a crisp model would be fit to this data in order to crisply model the spatial correlation. Because these data do not lend themselves to such accuracy, an informative variogram should reflect uncertainty resulting from the incomplete sample data. In this archetypical case of an *uncertain* sample variogram, the precision implied by a crisp model variogram is unwarranted and must be considered a possible misstep in the kriging process.

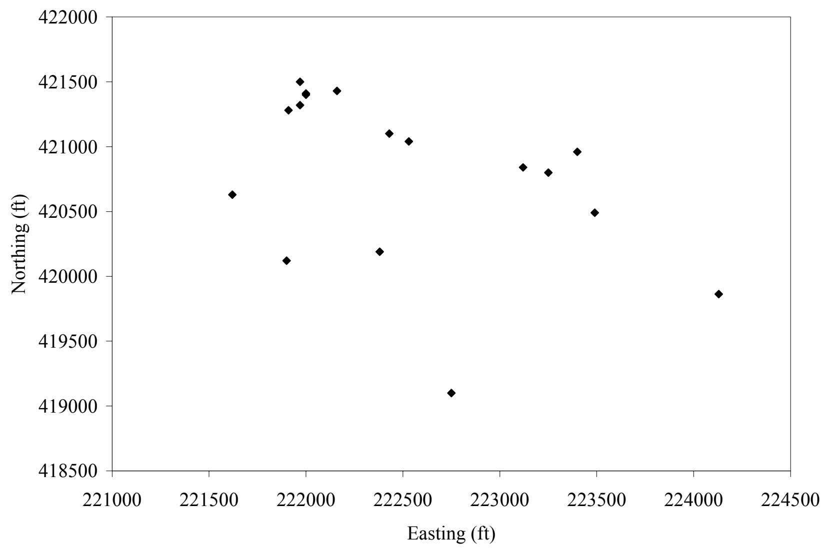


Figure 6.3. The 2-D spatial locations of hydraulic conductivity measurements at a remediation site show a relatively thorough sampling of the field

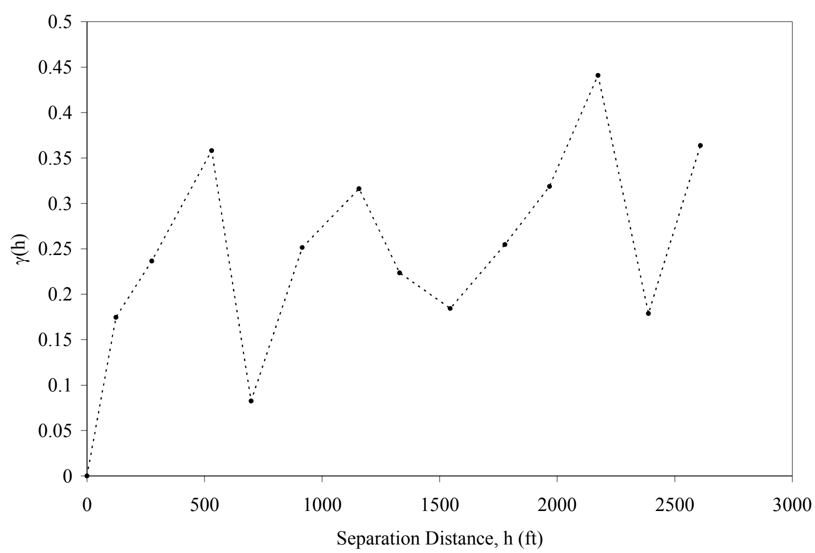


Figure 6.4. The sample variogram produced by the log hydraulic conductivity measurements, whose locations are plotted in Figure 6.3, does not imply any crisp variogram model

6.3.2. Fuzzy Variogram

A *second level* of uncertainty is introduced to the sample variogram when the hydrogeological parameter measurements at selected locations are expert-provided fuzzy numbers, rather than crisp values. The expert determines a value at a point and its inherent uncertainty based largely upon 1) the distance the point of interest is from locations where geohydrological observations have been made, 2) knowledge of the anticipated heterogeneity in the subsurface, 3) knowledge of the lithofacies in the unit of interest, and other similar qualitative information. The sample variogram in Figure 6.4 was computed using crisp measurements, whereas the sample variogram in Figure 6.5 was computed from fuzzy values. In such cases where measurements are fuzzy numbers, the sample variogram can be considered *second-level* uncertain due to the uncertainty imposed by the dispersion of the sample variogram data cloud compounded by the fuzziness of the individual semivariance values introduced by the expert fuzzy parameter estimates.

A model (pre-specified mathematical function) must be fit to the fuzzy second-level uncertain data like that in Figure 6.5 so as to assure that the variogram exhibits certain attractive mathematical properties. An expert fits the variogram to the fuzzy binned values and in the process generates a *third level* of uncertainty that is associated with the adequacy of the fit of the selected model to the discontinuous sample variogram. The bounding curves seen in Figure 6.6 illustrate this concept. This process of defining of fuzzy variogram model is accomplished by the modeler specifying fuzzy variogram parameters (nugget, sill, range).

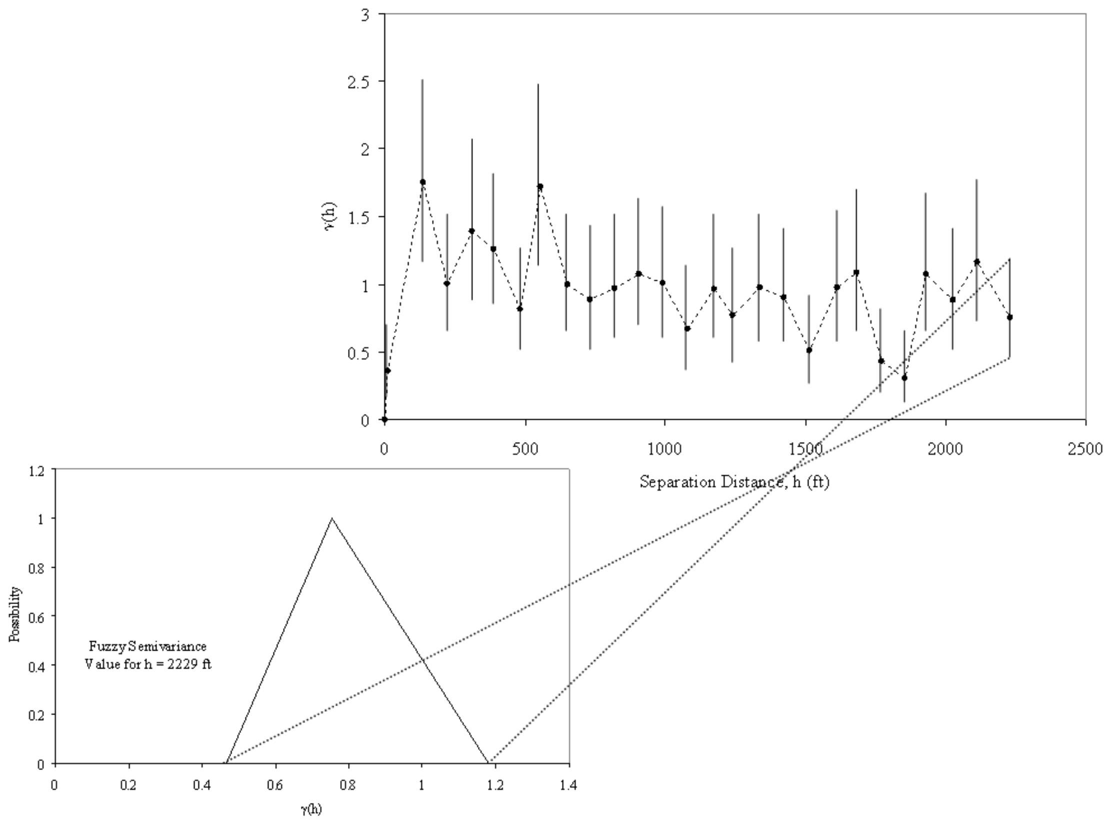


Figure 6.5. This sample variogram in the upper right is comprised of fuzzy semivariance values. The modal value for each fuzzy number is represented as a black dot. The interval created by the upper and lower bounds of each triangular fuzzy number is delineated by the vertical black line. An example of what the fuzzy numbers look like is provided in the lower left.

Admittedly, like traditional crisp variogram modeling, fuzzy variogram modeling is a trial and error process. Fuzzy variogram parameters are refined until an adequate model (elaborated below) is determined. Applying the extension principle to Gaussian model equation with fuzzy parameters produces the fuzzy variogram model (Bardossy et al, 1989, 1990a,b) represented by the median and bounding curves in Figure 6.6.

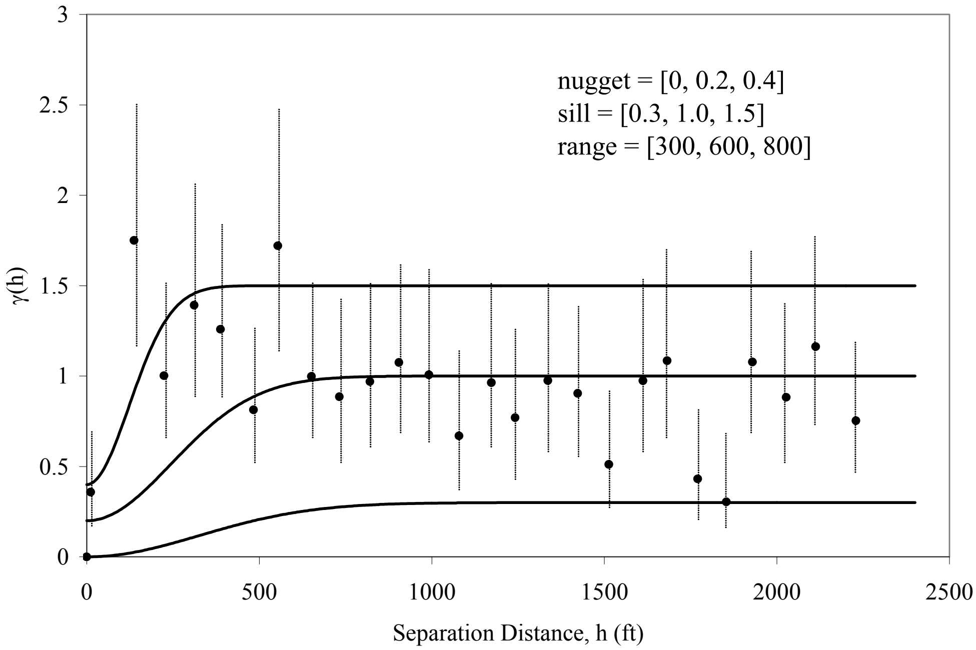


Figure 6.6. The upper bound, mode, and lower bounds of a fuzzy Gaussian variogram model for the sample variogram in Fig. 5. The nugget is a fuzzy number with parameters *[0, 0.2, 0.4]*; the sill is a fuzzy number with parameters *[0.3, 1.0, 1.5]*; and the range is a fuzzy number with parameters *[300, 600, 800].*

According to Uddameri (2004) a model is deemed adequate as long it bounds a majority of the data. While accomplishing this, the Gaussian model in Figure 6.6 both reveals the modeler's intuition regarding the spatial continuity of the field and neither imposes gratuitous precision nor excessive imprecision. An example of excessive imprecision, where the variogram model provides no information, is a fuzzy nugget model with parameters [0, 1.0, 2.0] that would totally encompass the sample data yet abandon all precision and reveal nothing of the perceived spatial continuity of the field.

Using the variogram model, with the total uncertainty expressed in terms of the fuzzy variogram model parameters listed in Figure 6.6, one can compute the kriging weights w_i

that appear in Equation (6-1). To obtain estimates at locations other than those points where initial *a priori* estimates are available one uses the formula shown in Equation (6-1). The values at the initial sites impose level-two uncertainty and are interpolated to other locations using the fuzzy kriging algorithm that exhibits level-three uncertainty (which implicitly also contains level-one and level-two uncertainty). Thus, the interpolated values exhibit all forms of uncertainty known to the problem. Such uncertainty is expressed in terms of possibility distributions. The interested reader is directed to (Bardossy et al, 1989, 1990a,b) , (Diamond, 1989) and (Piotrowski et al, 1996) for more information on fuzzy kriging.

6.4. Fuzzy Kalman Filtering

6.4.1. Updating Equations

The first estimate of the K-field is produced using expert generated estimates at selected points without the use of any physically-based (hard) hydraulic conductivity data. Hard data is based upon either traditional measurements or empirical transformations of measured secondary data into hydraulic conductivity values, and can be used to update the prior guess (or simply, *prior*) K-field and the prior spatial correlation structure. This updating is accomplished through the application of the corrector equations of a Kalman filter.

A Kalman filter (Kalman, 1960) is comprised of two sets of equations, one set predicting the state at the next time step, and the other set correcting, or updating, these predicted values with available state measurements. In this work, the state is a particular site's K-field, and does not change with time. Thus, only the updating equations are applied in this work (static Kalman filter). There are three steps to Kalman filter updating, the first being the calculation of the Kalman gain κ in Equation (6-3). The Kalman gain is merely an error minimization tool, whereby calculation of the optimal Kalman gain ensures that the updated estimates have the least mean squared error possible.

$$\kappa = C^{-}H^{T}(HC^{-}H^{T} + V)^{-1} \qquad \textbf{(6-3)}$$

If m is the number of updating measurements and n is the number of prior estimates, then C^{-} is the covariance matrix *[n×n]* for the entire prior K-field, V is a diagonal matrix *[m x m]* of updating measurement variances and H is a matrix *[m x n]* of ones and zeros that identifies the measurement locations in the prior K-field (Welch and Bishop, 2001). The form of Equation (6-3) results from the minimization of the estimate error covariance, and produces a matrix *[n x m]* requisite for covariance and estimate updating.

The estimate vector K *[n x 1]* and covariance matrix are updated in Equations (6-4) and (6-5), where Z is the vector *[m x 1]* of updating measurements and I is an identity matrix *[n x n]*.

$$K = K^{-} + \kappa(Z - HK^{-}) \qquad \textbf{(6-4)}$$

$$C = (I - \kappa H)C^{-} \tag{6-5}$$

6.4.2. Fuzzy Considerations

The method of fuzzy Kalman filtering applied as an extension of standard static Kalman filtering is quite similar in its function as that developed by Oussalah and De Schutter (2000), whereby state variable values are possibilistic as opposed to deterministic values. In this work, though, the Kalman filtering updating equations are simply extended by Equation (6-6) in order to operate on the fuzzy number and possibilistic prior estimates K^{-} and covariance matrix C^{-} in Equations (6-4) and (6-5). The posterior estimates K and the elements of the covariance matrix C are, in turn, characterized by possibility distributions. However, with regard to the calculation of the Kalman gain κ in a fuzzy Kalman filter, there are certain considerations.

Notice in Equation (6-3) that the Kalman gain matrix is determined knowing the prior covariance matrix C^{-} and the vector of updating measurement variances V. In the aforementioned geostatistical protocols, this work recommends the application of fuzzy model variograms, which lead to prior covariance matrices with possibilistic elements, a *fuzzy prior covariance matrix*. Where the prior covariance is comprised of possibilistic matrix elements, so to is the resulting Kalman gain a possibilistic matrix; all resulting uncertainty in the Kalman gain is due solely to the admitted uncertainty in the spatial continuity defined by the model variogram.

Examination of Equation (6-4) reveals that calculation of the posterior estimates matrix relies upon the Kalman gain matrix, as well as the prior estimates and updating measurements. The uncertainties in the Kalman gain and prior estimates, as modeled by possibility distributions, are conveyed to the posterior estimates, which are likewise calculated as possibility distributions.

However, while the Kalman gain transfers uncertainty in the spatial continuity of the field to the posterior estimates, the prior estimates convey the same uncertainty in addition to expert measurement uncertainty. Thus, the uncertainty due to the fuzzy model variogram is unnecessarily double counted. The same is true for the calculation of the posterior covariance in Equation (6-5), where uncertainty from the prior covariance matrix is passed on to the posterior covariance by both the possibilistic Kalman gain and the possibilistic prior covariance.

Since the Kalman gain is simply a means to minimum square error estimation and due to the superfluous double-counting of uncertainty in the updating equations, in this approach the Kalman gain is calculated with a defuzzified covariance matrix. That is, before calculation of the Kalman gain in Equation (6-3), the possibilistic elements of the prior covariance matrix C^- are reduced to crisp numbers by defuzzification (Klir and Yuan, 1995). Though uncertainty is eliminated from the Kalman gain, the posterior estimates and covariance matrix still exhibit the intended uncertainty due to the imprecision in spatial continuity and expert-provided measurements.

6.4.3. Posterior Estimates

Commensurate with traditional Kalman filter theory, the updated K-field will respect the updating measurement values. In fact, nearly the entire prior K-field will change as a result of the updating process. Consider a case with a prior estimate K-field comprised of *n* possibilistic numbers and a set of *m* updating measurements that are all crisp in form. Based upon the precision of the updating measurements, the uncertainty in the updated estimates should, in general, be less than that of the prior estimates. This means that throughout the estimated K-field, the possibilistic posterior estimates should have narrower possibility distributions than their corresponding possibilistic prior estimates. At the very least, the average possibility distribution should be narrower in the updated field due to the presence of the crisp updating measurements at specified locations. This is investigated in the field application, below.

6.5. Field Application

Superfund remediation has been ongoing at the CIBA-Geigy (Figure 6.7) site since 1982. The equalization basins on site were a part of the wastewater treatment process and were contaminated predominantly by 1,2,4-trichlorobenzene and 1,2-dichlorobenzene. The contaminated aquifer, comprised, with relative homogeneity, of yellow sand, is called the Primary Cohansey.

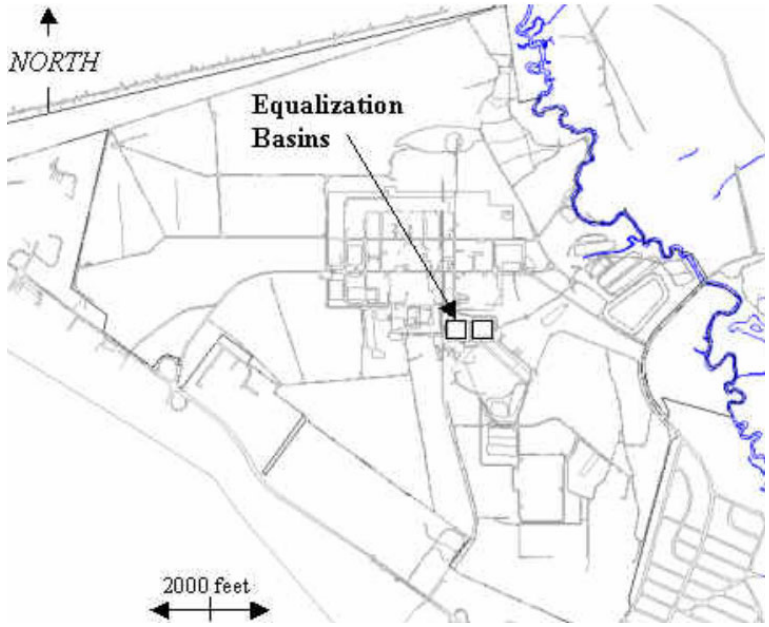


Figure 6.7. Map of the CIBA-Geigy site; the equalization basins are delineated (from www.cibageigysite.org)

The area overlying the Primary Cohansey is roughly 2 000 feet by 4 000 feet, from which 17 viable hydraulic conductivity measurements are available in addition to 41 borehole logs. The locations of the measurements and boreholes are shown in Figure 6.8, where it is clear that boreholes fill in many areas of the spatial domain where measurements are unavailable. Out of the 17 hydraulic conductivity measurements, 9 have corresponding borehole logs. Alone, the 17 hydraulic conductivity measurements would fail to adequately estimate the true hydraulic conductivity. However, the consideration of the additional soil data via fuzzy Kalman filtering can produce more accurate hydraulic conductivity estimates.

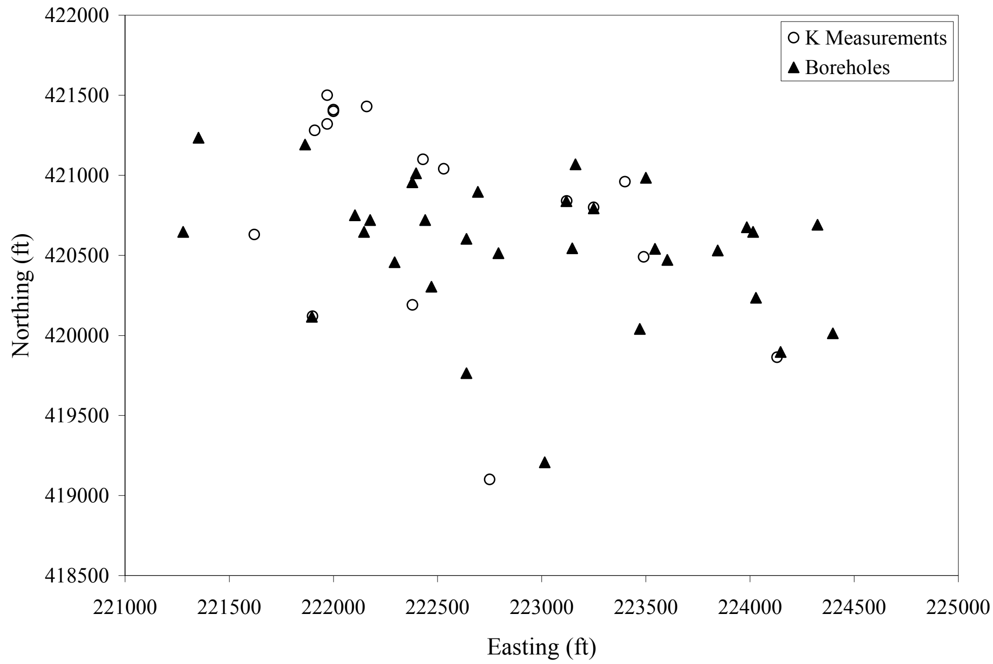


Figure 6.8. The locations of hydraulic conductivity measurements (.) are plotted along with the more numerous and more dispersed borehole locations (*)

Borehole quantification is the process by which a geologist interprets borehole soil descriptions and determines the percent contribution of various soil types to the soil samples described in the borehole log. Considering 7 soil types (*Clay*, *Silt*, *Fine Sand*, *Medium Sand*, *Coarse Sand*, *Fine Gravel*, *Medium Gravel* and *Coarse Gravel*), the borehole description,

Light gray coarse-medium-fine SAND, little Silt and Clay,

was translated by the geologist into the data vector in Table 6.1, where each value represents the proportion each soil class contributes to the soil sample described by the borehole entry. This was repeated for all 41 borehole descriptions.

Table 6.1. Expert-provided quantification of a borehole log description; each value represents the contribution of the soil class to the overall soil sample's composition

Clay	Silt	Fine Sand	Med Sand	Coarse Sand	Fine Gravel	Med Gravel	Coarse Gravel
1	3	32	32	32	0	0	0

A fuzzy rule base, relating soil type to hydraulic conductivity, was developed by Ross et al (2007) in order to estimate hydraulic conductivity from soil data. This rule base was tailored for the study site using a genetic algorithm (Pena Reyes, 2004) guided by the 9 quantified borehole log-hydraulic conductivity pairs. With this fuzzy rule base, hydraulic conductivity was estimated for the remaining 32 quantified borehole descriptions. One of the resulting fuzzy numbered hydraulic conductivity values is shown in Figure 6.9.

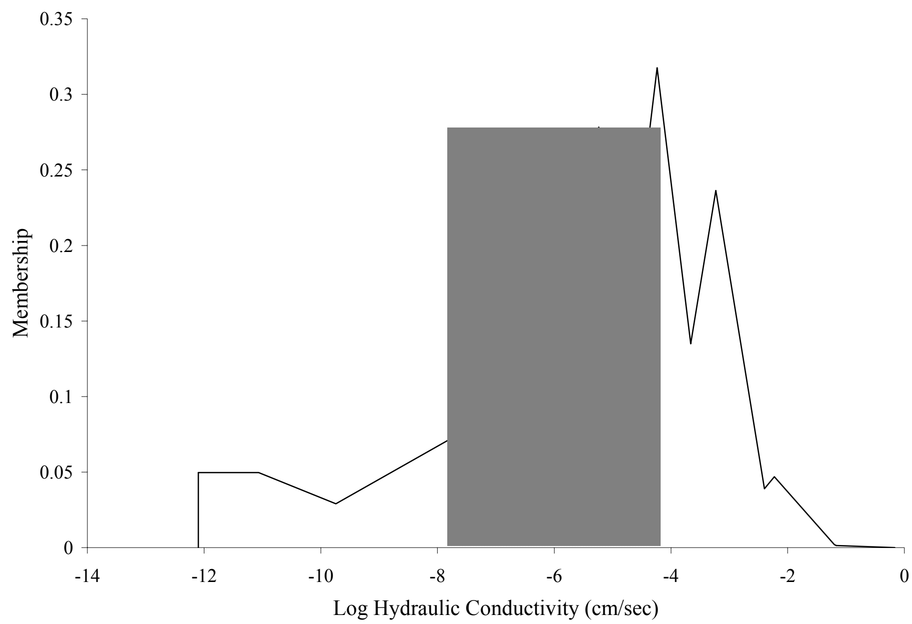


Figure 6.9. The fuzzy hydraulic conductivity estimate that results from a fuzzy rule base (pictured) does not resemble traditional fuzzy sets

In order to apply the Kalman filtering algorithm, as intended above, these fuzzy K values need to be interpolated to produce a possibilistic estimated K-field. As such, the fuzzy K values were interpolated throughout the entire spatial domain, using fuzzy kriging (Bardossy et al, 1989, 1990a,b), resulting in a possibilistic K value at each estimations location. These resulting possibilistic estimates were updated by 17 hydraulic conductivity measurements (plotted in Figure 6.3) using the fuzzy Kalman filtering algorithm described above. As anticipated, the posterior possibilistic estimates were, in general, more narrow than the corresponding prior estimates (Figure 6.10).

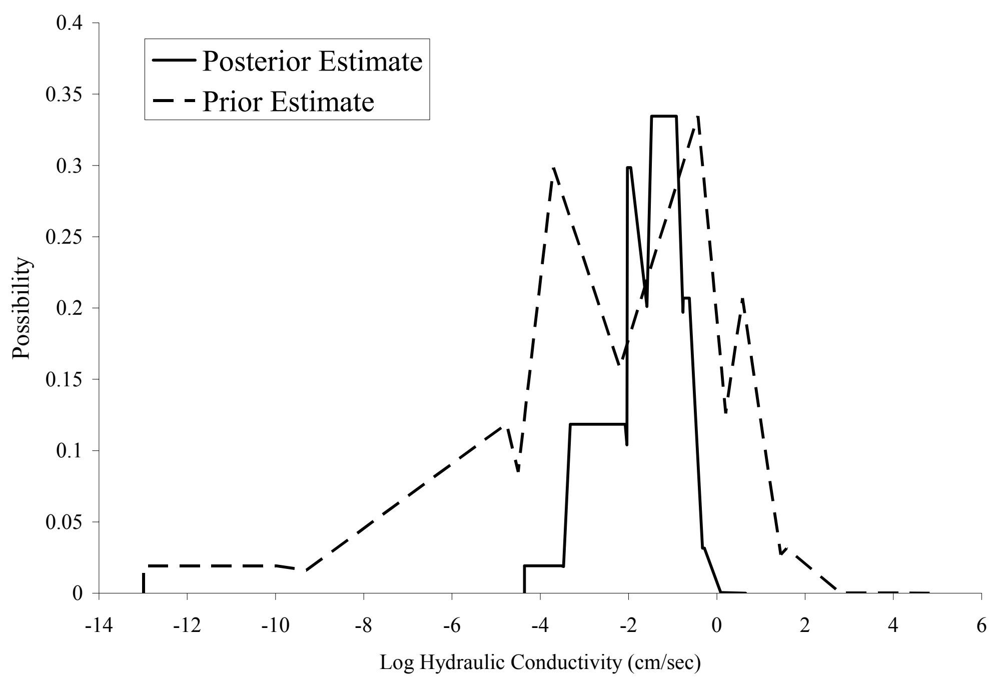


Figure 6.10. The posterior estimate (solid lines) at Easting 221 620 and Northing 420 800 was much less fuzzy (had a more narrow base) than the corresponding prior fuzzy estimate (dashed lines)

These posterior possibilistic estimates were defuzzified to produce a crisp estimated K-field (Figure 6.11A). In situations where prior hydraulic conductivity possibility distributions are symmetric, the defuzzified K-field would be the same as the posterior K-field resulting from non-fuzzy priors.

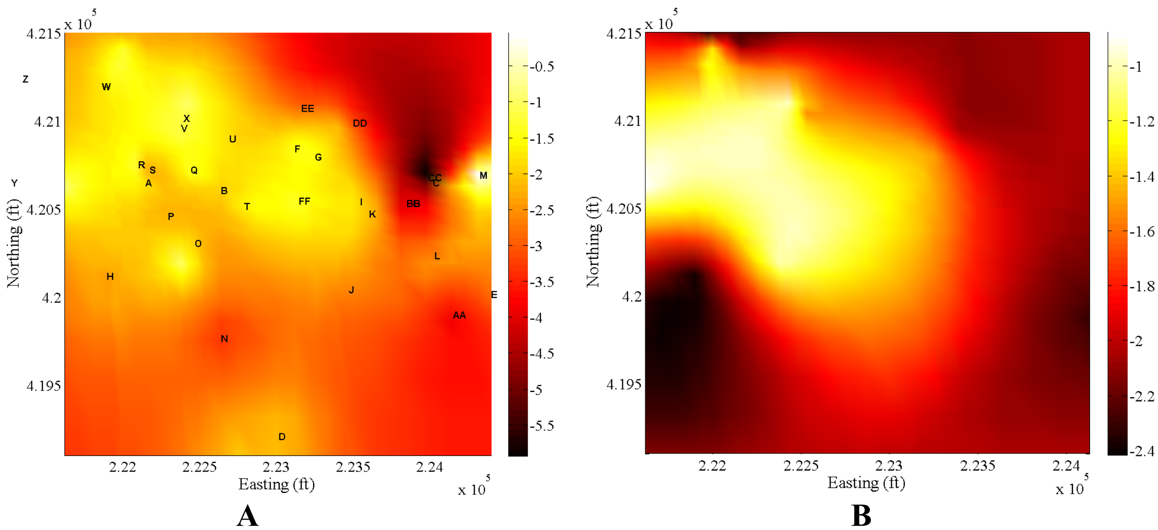


Figure 6.11. The log K-field (cm/sec) produced by fuzzy Kalman filtering, (left), exhibits phenomena not present in the K-field produced by kriging the 16 hydraulic conductivity measurements, (right)

6.6. Results and Conclusions

The posterior estimated crisp K-field (Figure 6.11A) reveals some phenomenon that are absent from the estimated K-field that resulted from simply kriging the 17 hydraulic conductivity measurements (Figure 6.11B). The fuzzy Kalman filtering K-field shows that the area of low hydraulic conductivity in the northeast section of the spatial domain is larger than that produced by the K-field produced by kriging crisp measurements. Furthermore, the updated K-field reveals less continuity of higher hydraulic conductivity than is evident in the kriging-only K-field. An additional patch of high hydraulic conductivity is evident in the east of the updated figure. These results are confirmed by the soil samples evaluated in boreholes in those respective areas of the domain (Table 6.2).

Table 6.2. Borehole descriptions used to predict hydraulic conductivity; locations plotted in Figure 6.11

Well	Borehole Description
A	Light gray c-m-f SAND, little Silt & Clay, slightly micaceous, medium dense
B	Very dark gray m-f SAND, trace Silt, slightly micaceous, medium dense
C	Light gray to brown, c-m-f SAND, trace f Gravel, trace Silt and Clay, loose.
D	Orange c-m-f SAND
E	Black organic m-f SAND, some to little Silt
F	f Sand
G	f Sand
H	f Sand, little f Gravel, Clay
I	Light gray c - f Sand, some m - f Gravel, trace Silt
J	Gray brown fine Sand, little Silt
K	Light gray c - f Sand and m - f Gravel, little Silt
L	Wet brown Sand and Peat with trace Gravel
M	Wet coarse orange Sand
N	Medium coarse Sand and Gravel, streaks yellow, white, blue Clay
O	f brown Sand
P	f brown Sand
Q	f brown Sand
R	f brown Sand
S	f brown Sand
T	f brown Sand
U	f brown Sand
V	f brown Sand
W	brown Sand and grits
X	reddish brown Sand
Y	yellowish Sand
Z	reddish brown Sand
AA	dark brown fine Sand
BB	Light brown Sandy Clay
CC	Dark brown Silt
DD	medium brown Sand
EE	Light brown Sandy Clay
FF	Brown Sand and Gravel

The summary statistics mathematically confirm the differences between the K-fields in Figure 6.11 (Table 6.3). Lower K values implied by the boreholes lead to a lower mean value for the updated K-field than for the traditionally determined K-field. Nevertheless, expert knowledge suggests a greater range of K values for the K-field in Figure 6.11A. Certainly, the differences between a traditionally determined K-field and a K-field augmented by expert knowledge should be greater where soil data and expert knowledge reveal hydrogeological formations not identified by traditional hydraulic conductivity measurements.

Table 6.3. Summary statistics for the posterior K-field and the K-field estimated using only crisp measurements.

	Posterior K-Field	Crisp K-Field
Mean	-2.4711	-1.6848
Standard Deviation	0.9474	0.4204
Minimum	-5.9320	-2.4164
Q_1 (0.25 quartile)	-3.0386	-2.0506
M	-2.1809	-1.7745
Q_3 (0.75 quartile)	-1.7274	-1.2775
Maximum	-0.0357	-0.8761

Nevertheless, the consideration of additional data sources can certainly enhance site characterization. By incorporating the additional soil data in the form of quantified boring logs in this case, fuzzy Kalman filtering permitted both an appropriate approach to aggregate fuzzy and crisp data and a more accurate hydrogeological characterization of the CIBA-Geigy remediation site.

Previous approaches to incorporating fuzzy and crisp data in hydrogeological site characterization have suggested that fuzzy and crisp data be used in conjunction. Generally, fuzzy data results from expert knowledge. The nature of expert knowledge suggests that it should be used to produce a best first guess of the hydraulic conductivity field to be updated by more certain and precise data resulting from direct measurements. The fuzzy Kalman filtering algorithm presented herein takes this more intuitive approach to hydrogeological site characterization with expert knowledge. Nevertheless, it is not the first and only method to attempt to treat expert knowledge in this manner.

Algorithms have been developed to update kriged priors with additional data in a Bayesian context (Omre, 1987; Bandemer and Gebhardt, 2000). While a comparison between the proposed Kalman filtering approach and the Bayesian fuzzy kriging method herein may seem prudent, the two methods are quite dissimilar in their intentions. Whereas the fuzzy Kalman filtering algorithm uses expert knowledge of hydraulic conductivity directly in the form of fuzzy numbers, Bayesian fuzzy kriging is built for expert knowledge regarding the first and second moments of trend parameters. As such, further research is necessary to develop a Bayesian updating scheme that revises expert-provided fuzzy K values with crisp measurements.

A6.1. Fuzzy Sets

The fuzzy set, A, a generalization of the classical set, is defined by a membership function, A(x), on the particular variable of interest X ($x \in X$). The membership function defines, for each value within the variable's domain, the belongingness of that value in the notion conveyed by the fuzzy set. Fuzzy sets are traditionally applied to the characterization of imprecise notions and linguistic variables (Zadeh, 1965). For instance, if an appropriate expert were to specify the notion of the hydraulic conductivity at some location in a spatial domain as *low*, this notion is an imprecise linguistic variable and could be characterized with a membership function like that shown in Figure 6.1. In this fuzzy set, the log hydraulic conductivity value *-9* has full membership in the prescribed hydraulic conductivity fuzzy set. Membership steadily decreases toward *-10* on the lower bound and *-8* on the upper bound. Alternatively, the fuzzy set can be represented as a collection of nested intervals, or α-cuts. An α-cut of fuzzy set A, $^{\alpha}A$, is an interval of elements that have membership degrees greater than or equal to α, created by horizontally cutting the fuzzy set at some membership, or α, value. In Fig. 1, the *0.5* level α-cut is found to be *[-9.5, -8.5]*.

Though membership functions can take many shapes, that which is given in Fig. 1 is triangular. Triangular fuzzy sets are commonly employed, and they can be represented by three parameters, a lower bound (LB), upper bound (UB), and modal value (M), [LB, M, UB]. The fuzzy set in Fig. 1 can be denoted *[-10, -9, -8]*.

A6.2. Fuzzy Numbers

While fuzzy sets define adjectives and partition a variable's domain, fuzzy numbers are a particular type of fuzzy set that satisfy certain conditions:

a.) At least one value has full membership, $\exists x \text{ s.t. } A(x) = 1$;

b.) ${}^{\alpha}A$ must be closed and bounded $\forall \alpha \in (0, 1]^3$.

Fuzzy numbers are fuzzy sets defined on the real number line, prescribed by a possibility distribution, thereby representing imprecise numerical values. For example a fuzzy *-9*, perhaps labeled *about -9*, could be defined by a possibility distribution identical to the membership function in Figure 6.1, as an expert-provided solution to the question, "What is the hydraulic conductivity?" for some specified location. As a fuzzy number, the possibility distribution, $\pi(k)$, stipulates, for each element on the domain ($k \in K$), the possibility that k is the true log hydraulic conductivity at that particular location. Thus, while similar in structure, fuzzy numbers and fuzzy sets differ in interpretation and use. Fuzzy sets define imprecise notions, most often as a means of fuzzy inference via fuzzy rule bases, while fuzzy numbers are most appropriate for the possibilistic quantification of a numerical value.

A6.3. Possibility and Probability

The constraints that govern the definition of a possibility distribution are those provided above for fuzzy numbers; and the information that guides such definition is generally born of the intuition of an appropriate expert.

The possibility distribution, a non-probabilistic means of capturing uncertainty, is the principal contribution of possibility theory (Zadeh, 1978). Structurally identical to a

---

1 An α-cut at the zero level is the real number line, and, as such, is not a closed interval

fuzzy number, a possibility distribution can be expert-provided, though it represents a different facet of uncertainty. Whereas a fuzzy set typically models linguistic variables (Zadeh, 1975), which are certain yet imprecise notions, a possibility distribution represents a lack of knowledge regarding the true value of some variable.

Because possibility distributions model uncertainty that results from a lack of knowledge as opposed to a lack of precision, they are more similar to probability distributions than fuzzy sets. Aside from direct possibility-probability transformations (Jumarie, 1995; Oussalah, 2000; Klir, 2006), where a possibility distribution can be re-interpreted as a probability distribution (and vice versa) after an, oftentimes simple, functional transformation, possibility distributions are found to be naturally related to probability distributions via evidence theory (Yager, 2004; Klir and Yuan, 1995).

As such, a possibility distribution is not necessarily a replacement for a probability distribution. Rather, a possibility distribution is more general. For instance, the area contained within the possibility distribution need not equal unity. In fact, the probability distribution is a special case of a possibility distribution approximation of the random number that best defines the true value in question (Klir and Yuan, 1995). A possibility distribution actually induces upper and lower bounds on the unknown true cumulative probability distribution (Yager, 2004; Druschel et al, 2006; Mathon et al, 2008). When information, requisite for probability distribution definition, is absent, a possibility distribution appears to be an intuitive means of providing a representation of the uncertain value, implicitly bracketing a probability distribution. Klir and Yuan (1995), Klir (2006) and Yager (2004) provide a thorough treatment of the equivalence between a possibility distribution and upper and lower probability bounds.

A6.4. Extension Principle

Just like crisp numbers, fuzzy numbers can arise as arguments in functions. In such circumstances, the extension principle (Klir and Yuan, 1995) must be applied in order for the desired functions to operate upon fuzzy numbers. The frequent application of the extension principle reinforces its appropriateness for propagating fuzzy information through both linear and non-linear functions (Dubois and Prade, 1991). If the independent variables (X_i, $i = 1,\ldots, n$) are each valued by a fuzzy number (A_i, $i = 1,\ldots, n$), the extension principle stipulates the corresponding fuzzy number B of the dependent variable Y, the values of which are calculated using the particular function being extended. The form of the extension principle is

$$B(y) = \sup_{x_i,\ldots,x_n \mid f(x_i,\ldots,x_n)=y} [\min\{A(x_i),\ldots,A(x_n)\}] \qquad \textbf{(6-6)}$$

$\forall y \in Y$, where both X and Y are often $\Re$, the real number line. Eq. (6-6) states for a given element of Y, $y \in Y$ all combinations of $x_i, \ldots, x_n$ are found that satisfy $f(x_i, \ldots, x_n) = y$. For each combination, the membership values $A(x_i)$ are determined and the minimum membership value is selected. Then, over all the combinations, the maximum membership value is attributed to the dependent variable value y. This is repeated for all

$y \in Y$; thereby a membership function is obtained for the variable *Y*. The extension principle, though intricate in its operation, has greatly simplified the means by which fuzzy numbers can be utilized in otherwise non-fuzzy functions. The fuzzy kriging algorithm summarized above results from the application of the extension principle to the crisp kriging system of equations

References

Alyamani, M.S., and Sen, Z. (1993). "Determination of hydraulic conductivity from complete grain-size distribution curves." *Ground Water*, 31(4), 551-555.

Asli, M., and Marcotte, D. (1995). "Comparison of approaches to spatial estimation in a bivariate context." *Mathematical Geology*, 27(5), 641-658.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). "Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy, A., Bogardi, I., and Kelly, W.E., (1990a). "Kriging with imprecise (fuzzy) variograms I: Theory." *Mathematical Geology*, 22(1), 63-79.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990b). "Kriging with imprecise (fuzzy) variograms II: Application." *Mathematical Geology*, 22(1), 81-94.

Gotway, C.A., and Hartford, A.H. (1996). "Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables." *Journal of Agricultural, Biological, and Environmental Statistics*, 1(1), 17-39.

Issaks, E.H., and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Kalman, R.E. (1960). “A new approach to linear filtering and prediction problems.” *Transactions of the ASME - Journal of Basic Engineering* (*Series D*), 82, 35-45.

Klir, G.J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, New Jersey.

Kravchenko, A.N. (2003). “Influence of spatial structure on accuracy of interpolation methods.” *Soil Science Society of America Journal*, 67, 1564-1571.

Masch, F.D., and Denny, K.J. (1996). „Grain size distribution and its effect on the permeability of unconsolidated sands.” *Water Resources Research*, 2(4), 665-677.

Odeh, I.O.A, McBratney, A.B., and Slater, B.K. (1999). “Predicting soil properties from ancillary information: Non-spatial models compared with geostatistical and combined methods.” *Geostatistics Wollongong*, 2, 1008-1019.

Oussalah, M., and de Schutter, J. (2000). “Possibilistic kalman filtering for radar 2-D tracking.” *Information Sciences*, 130, 85-107.

Piotrowski, J.A. (1996). "Estimation of hydrogeological parameters for groundwater modeling with fuzzy geostatistics: closer to nature?." *Calibration and Reliability in Groundwater Modeling*, 237, 511-520.

Reyes, C.A.P. (2004). *Coevolutionary Fuzzy Modeling*, Springer, Berlin, Germany.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). "Fuzzy kalman filtering of hydraulic conductivity." *Computations Methods in Water Resources Conference 26th International Conference (CMWRXVI)*.

Ross, J., Ozbek, M., and Pinder, G.F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*, 39(6), 765-780.

Sen, Z. (1992). "New statistical grain-size method for evaluating the hydraulic conductivity of sand aquifers – Comment." *Journal of Hydrology*, 130, 399-403.

Uddameri, V. (2004). "Relationships of longitudinal dispersivity and scale developed from fuzzy least-squares regressions." *Environmental Geology*, 45(8), 1172-1178.

Uma, K.O, Egboka, B.C.E., and Onuoha, K.M. (1989). "New statistical grain-size method for evaluating the hydraulic conductivity of sandy aquifers." *Journal of Hydrology*, 108, 343-366.

Wang, X.J., and Qi, F. (1998). "The effects of sampling design on spatial structure analysis of contaminated soils." *The Science of the Total Environment*, 224, 29-41.

Webster, R., and Oliver, M.A. (1992). "Sample adequately to estimate variograms of soil properties." *Journal of Soil Science*, 43, 177-192.

Welch, G., and Bishop, G. (2001). "An introduction to the kalman filter." *SIGGRAPH 2001 course 8*, *In Computer Graphics*, *Annual Conference on Computer Graphics & Interactive Techniques*. ACM Press, Addison-Wesley, Los Angeles, CA, USA.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1, 3-28.

Zadeh, L. (1965). "Fuzzy sets." *Information and Control*, 8(3), 338-353.

7: RESPECTING CORRELATION WHILE MODELING WITH POSSIBILISTIC DATA

Stochastic groundwater modeling involves the propagation of probabilistic uncertainty from model input parameters to model estimates usually via a Monte Carlo method. However, with the increasing reliance upon expert knowledge to define model inputs, such as hydraulic conductivity, and fuzzy set theory to characterize this expert knowledge, alternative means of executing model equations were developed. In this approach, a novel methodology for propagating fuzzy-based uncertainty is presented. Fuzzy sets are sampled to produce a number of realizations in a manner analogous to Latin Hypercube sampling. The resulting realizations are passed through numerically approximated groundwater flow and transport equations. The resulting concentration estimates are assembled as possibility distributions.

7.1. Introduction

The natures of the measurement, estimation, and natural spatial variability of an aquifer's hydrogeological properties account for some of the uncertainty in the groundwater flow and transport modeling process (Gelhar, 1986; Isaaks and Srivastava, 1989; de Marsily et al, 1998; Højberg, and Refsgaard, 2005; Nillson et al, 2007;). Such uncertainty has traditionally been characterized by probabilities (Gelhar, 1993). As a result, generally employed *stochastic groundwater modeling* (Gelhar, 1993; Winter,

2004) has combined groundwater model equations with Monte Carlo analyses. The result is a commonly accepted method of uncertainty propagation in groundwater fate and transport modeling.

Expert knowledge is a potentially important source of data on aquifer properties (Bardossy et al, 1988; Bardossy et al, 1989; Bardossy et al, 1990a,b,c; Piotrowski, 1996; Dhiman and Keshari, 2003; Demicco, 2004; Ross et al, 2006; Ross et al, 2007;). Though, as noted, parameter uncertainty has traditionally been characterized by probability theory, it is cumbersome to describe expert knowledge probabilistically (Ganoulis, 1996; O'Hagan and Oakley, 2004). As a result, alternative methodologies have been used to capture expert knowledge and associated uncertainties. Evidence theory (Dempster, 1968; Shafer, 1976) characterizations of expert knowledge in groundwater applications constitutes one approach (Druschel et al, 2006; Mathon et al, 2008), but more common are fuzzy set based characterizations. A thorough review of such applications is provided by Ozbek and Pinder (2006).

In spite of the ease with which expert knowledge can be framed using fuzzy sets, groundwater model equations cannot be solved directly when fuzzy numbered input parameters are used. Two popular methods for accomodating fuzzy-based uncertainty in a groundwater model are the extension principle (Dou et al, 1995; Dou et al, 1997a,b; Larue and Tyagi, 1997, 1998; Faybishenko, 2004; Prasad and Mathur, 2007) and fuzzy logic model approximations (Bardossy and Disse, 1993; Nedungadi et al, 1994; Bardossy et al, 1995; Bagtzoglou et al, 1996; Bardossy, 1996; Schultz and Huwe, 1997; Dou et al, 1999; Coppolla et al, 2002; Vernieuwe et al, 2002a,b; Vernieuwe et al, 2007). The extension principle wraps around a crisp equation, or set of equations, generalizing them to operate upon fuzzy numbers. A fuzzy rule-based model, on the other hand, is comprised of a set of conditional statements relating fuzzy values of an independent variable, such as hydraulic conductivity, to fuzzy values of a dependent variable, such as head.

Whereas the application of the extension principle leaves the groundwater model equations intact, the development of a fuzzy rule-based groundwater model not only requires significant insight into the physics of groundwater and contaminant flow, but an in depth understanding of fuzzy set theory and fuzzy logic. A drawback of these approaches is that both fail to respect the spatial correlation of model input parameters.

Commonly applied flow and transport codes (i.e. Princeton Transport Code) lack fuzzy set and fuzzy logic-based computations, in spite of the increasing use of fuzzy and possibilistic data in groundwater flow and transport modeling. Thus, the propagation of fuzzy-based parameter uncertainties by the extension principle and fuzzy rule-based models is impractical. A novel method that facilitates the use of fuzzy set groundwater parameters in a traditional groundwater model is presented herein. The new method, introduced below, modifies an extended Latin Hypercube sampling algorithm (Zhang and Pinder, 2003) to effectively sample a fuzzy set. In this application realizations are drawn from a set of hydraulic conductivity possibility distributions. These realizations are passed through a groundwater flow and transport model (the Princeton Transport Code), resulting in a set of uncertain concentration estimates. Due to the possibilistic nature of the input hydraulic conductivity values, the computed concentration values are framed as possibility distributions.

This paper is organized in the following manner. A brief introduction into the construction of fuzzy sets and their relationship to possibility distributions is provided first. Next, the new Latin Hypercube sampling (LHS) algorithm is described including a brief illustration. The proposed algorithm is then applied to the simplified Woburn, MA site. Included is an explanation of the characterization of the concentration estimates as possibility distributions. Conclusions follow the field application.

7.2. Fuzzy Sets and Possibility Distributions

The fuzzy set was introduced by Zadeh (1965) as a generalization of the classical set, in order to more appropriately characterize the imprecision inherent in human-provided information. In hydrogeological applications, fuzzy sets have often been applied to the characterization of expert knowledge regarding hydraulic conductivity (Bardossy et al, 1988; Bardossy et al, 1989; Bardossy et al, 1990a,b,c; Piotrowski, 1996; Demicco, 2004; Ross et al, 2006; Ross et al, 2007). In these instances, the fuzzy number hydraulic conductivity values were regarded as measurements to be used in a kriging model to spatially estimate a fuzzy hydraulic conductivity field. The kriged estimates of hydraulic conductivity are most appropriately interpreted as possibility distributions (Zadeh, 1978). Whereas fuzzy sets are intended to measure the imprecision inherent in a human notion, a possibility distribution, structurally identical to a fuzzy number, identifies uncertainty resulting from a certain level of ignorance. The distinction lies in the information used to construct the distribution. An expert defines a fuzzy set if he/she is certain of the true approximate value of hydraulic conductivity, and specifies a possibility distribution when little to no information is available regarding the true value of hydraulic conductivity. A kriged estimate of hydraulic conductivity is a possibility distribution, since no information about the true approximate value of hydraulic conductivity at the estimate location was used to determine the uncertain hydraulic conductivity value. Moreover, in cases where fuzzy groundwater parameter values are employed in a groundwater flow and transport model (Dou et al, 1995; Dou et al, 1997a,b; Larue and Tyagi, 1997, 1998; Faybishenko, 2004; Prasad and Mathur, 2007), the resulting state variable estimates are possibility distributions for the same reason – no prior insight into the state variable values is used to estimate them.

Though the extension principle is a common approach to sampling a fuzzy set for the purpose of propagating fuzzy data through a mathematical model, an alternative sampling regime, introduced by Chanas and Nowakowski (1988), inspires the sampling procedure in the fuzzy Latin Hypercube sampling algorithm to be presented in this paper. By this *fuzzy-numerical simulation method*, realizations comprised of crisp numbers can be drawn from a set of fuzzy numbers and ultimately used directly in a crisp mathematical model. In this approach, for a fuzzy number A, the following steps are repeated for each of the desired number of realizations M. A value t is sampled from the uniform random variable over the interval $(0, 1]$. An α-cut is taken at the membership value t, and from a uniform distribution over this α-cut is sampled a value x, such that the membership of x in A is greater than or equal to t. This value is considered to be a representative of the fuzzy set A and is passed through the set of equations comprising the mathematical model. These steps are modified for the fLHS algorithm, as presented in the next section.

7.3. Fuzzy Latin Hypercube Sampling (fLHS)

McKay et al (1979) introduced LHS in order to deterministically estimate a random variable Y resulting from the evaluation of some function $Y = f(\boldsymbol{X})$ with an input vector of n random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, by sampling M realizations from the stochastic independent variables. Zhang and Pinder (2003) extended LHS to ensure that the correlation of the M realizations approximates the correlation of the independent variables. The probability density function of each random variable is segmented into M non-overlapping equi-probable areas (Figure 7.1, from Dokou, 2008).

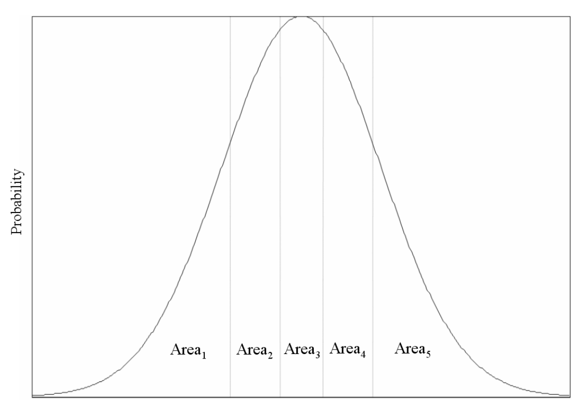


Figure 7.1. A probability density function segmented into *M* = *5* areas of equal probability, from Dokou, 2008

From each of these areas is sampled a single value, creating a discrete random variable comprised of M sampled values.

Likewise, a possibility distribution can be segmented into M equi-probable areas. Analogous to an α-cut of a fuzzy set, a horizontal cut through a possibility distribution produces an interval called a focal element (Klir and Yuan, 1995). A focal element is comprised of an interval and associated probability mass assignment. Where the unit interval of possibilities is discretized into M equally-spaced values, a possibility distribution is segmented into M equi-probable intervals (Figure 7.2), or M intervals each with an associated probability equal to $1/M$. On each of these intervals is defined a uniform probability distribution.

Whereas the equi-probable areas of a probability density function do not overlap, a possibility distribution's focal elements are nested. In other words the interval corresponding to a possibility value t contains the interval created by a cut at $t + \varepsilon$, where ε is some small value. If one were to follow the fuzzy-numerical sampling algorithm, it is possible to sample the same value from each of the M nested focal elements. Thus, the fuzzy-numerical sampling algorithm, as applied to fuzzy Latin Hypercube sampling, must be amended. Instead of sampling just once over each focal element, fuzzy Latin Hypercube sampling dictates that one randomly samples J values from a focal element. By sampling in this manner, the resulting $L = M \times J$ realizations adequately represent the sampled possibility distribution. The entire algorithm can be summarized by the following steps, some of which are illustrated in Figure 7.2, for a field of N possibility distributions with a spatial correlation defined by $\boldsymbol{C^*}$:

a. Specify the number of desired realizations M.

b. Segment the N possibility distributions into M equi-probable nested intervals.

c. Define a uniform probability distribution over each interval, and randomly select J values.

d. Assemble the $L = M \times J$ sampled values from each of the N possibility distributions in a realization matrix $\boldsymbol{X}$ $[L \times 4]$.

e. Calculate the correlation matrix $\boldsymbol{C}$ for $\boldsymbol{X}$.

f. Determine the cholesky decompositions for $\boldsymbol{C}$ and $\boldsymbol{C^*}$ and use these decompositions to calculate a new realization matrix $\boldsymbol{X^*}$. The correlations for $\boldsymbol{X^*}$ should closely approximate the target correlation matrix $\boldsymbol{C^*}$.

g. Permute the values in each column of $\boldsymbol{X}$ according to the rank order of the corresponding column of $\boldsymbol{X^*}$. This newly permuted realization matrix D is that which is employed in stochastic modeling. Each row represents a unique realization of the parameter field.

h. Realizations are passed through the numerical model as in a Monte Carlo approach.

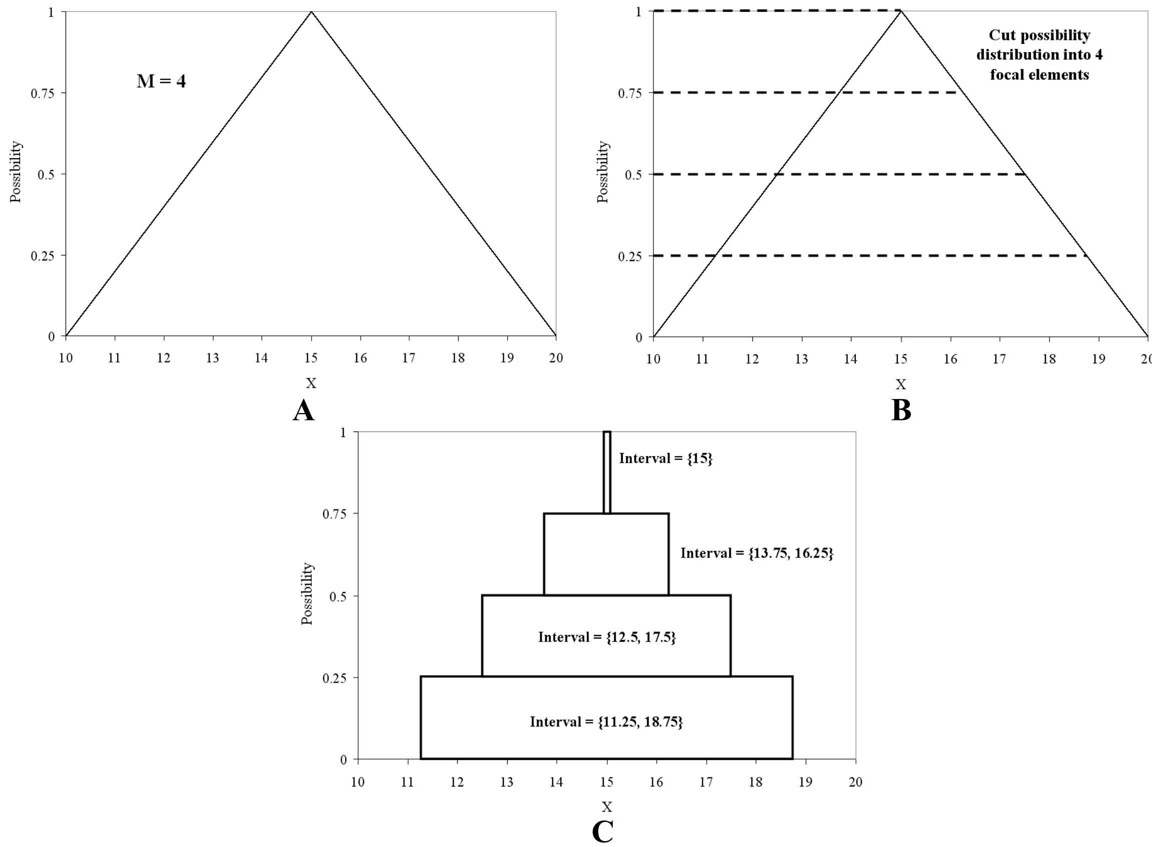


Figure 7.2. Once the desired number of realizations is selected (A), the possibility distribution is cut *M* times (B), creating *M* equiprobable intervals (C)

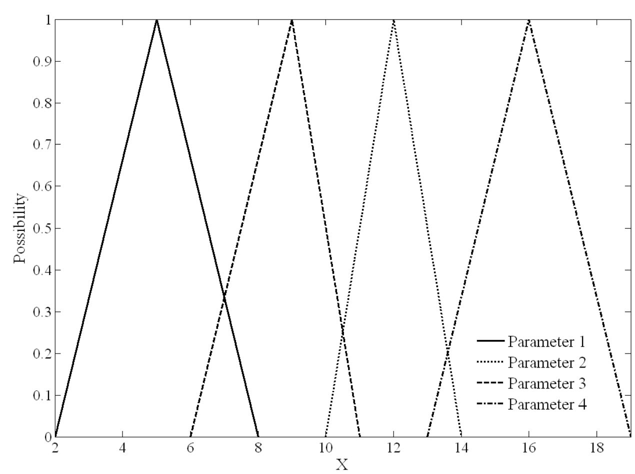


Figure 7.3. Possibilistic uncertain parameters used for the brief demonstration of the efficacy of the fLHS; their correlation matrix is given in Table 7.1

Consider a simple example (interpreted from Zhang and Pinder, 2003) comprised of four possibilistic variables (Figure 7.3) with associated correlation matrix (Table 7.1).

Table 7.1. Correlation matrix for triangular fuzzy values in Figure 7.3

	1	**2**	**3**	**4**
1	1.0	0.702	0.702	0.606
2	0.702	1.0	0.606	0.702
3	0.702	0.606	1.0	0.702
4	0.606	0.702	0.702	1.0

Table 7.2. Ranks for realization matrix *X*

X1	**Rank(X1)**	**X2**	**Rank(X2)**	**X3**	**Rank(X3)**	**X4**	**Rank(X4)**
6.1104	7	7.5179	8	6.2355	4	9.5814	8
6.6611	8	8.662	10	6.4742	5	7.387	4
7.3574	9	5.592	5	8.5267	8	9.0127	7
4.3499	4	5.3516	4	9.3041	9	6.5987	3
3.0495	2	3.1713	1	7.1762	6	6.085	2
3.6422	3	8.2007	9	4.2103	1	10.116	9
4.6855	5	4.0303	2	5.6789	3	10.6198	10
2.0148	1	4.7727	3	8.0275	7	5.2887	1
5.1863	6	6.0738	6	5.0687	2	8.1518	6
7.4761	10	6.9277	7	9.554	10	7.5058	5

Each possibility distribution is discretized into M nested intervals. Beginning with the base interval of each possibility distribution, a unique value ($J = 1$ and $L = 4$ in this simple example) is selected for each variable. These four values are placed into the first row of a realization matrix [$L \times 4$]. This is repeated for the remaining focal elements. With each interval, a new row of the realization matrix is filled in, resulting in realization matrix X (Table 7.2). Through traditional LHS computation (step f., above), a matrix X^* is calculated (Table 7.3), whose correlation matches that of the target correlation matrix (Table 7.1).

Table 7.3. Ranks for *X matrix**

X1*	Rank(X1*)	X2*	Rank(X2*)	X3*	Rank(X3*)	X4*	Rank(X4*)
-12.1419	3	2.7076	7	7.6613	7	14.6802	9
-9.9547	9	4.7016	10	7.5007	6	13.2059	5
-10.4627	6	3.1094	8	9.9854	9	15.2794	10
-11.1128	4	2.1448	5	7.8495	8	11.4371	3
-9.4333	10	0.5283	2	6.2353	3	9.4371	2
-14.4532	1	0.9471	4	4.8713	1	13.2452	6
-12.4346	2	-0.306	1	7.3249	5	14.2562	7
-10.9687	5	0.9022	3	5.4411	2	8.4645	1
-10.0293	7	2.1538	6	6.4597	4	12.391	4
-10.0116	8	4.547	9	10.0086	10	14.378	8

The ranks of each column of *X** are recorded and used to permute the original realization matrix *X*, creating a new matrix *D* (Table 7.4). The correlation matrix of this new realization matrix *D* closely approximates that of the original possibilistic variables. The target correlation matrix (Table 7.1) is approximated well by the matrix is in Table 7.5.

Table 7.4. New realization matrix *D*; note that the ranks match those of *X in Table 7.3**

D1	Rank(D1)	D2	Rank(D2)	D3	Rank(D3)	D4	Rank(D4)
3.6422	3	6.9277	7	8.0275	7	10.116	9
7.3574	9	8.662	10	7.1762	6	7.5058	5
5.1863	6	7.5179	8	9.3041	9	10.6198	10
4.3499	4	5.592	5	8.5267	8	6.5987	3
7.4761	10	4.0303	2	5.6789	3	6.085	2
2.0148	1	5.3516	4	4.2103	1	8.1518	6
3.0495	2	3.1713	1	6.4742	5	9.0127	7
4.6855	5	4.7727	3	5.0687	2	5.2887	1
6.1104	7	6.0738	6	6.2355	4	7.387	4
6.6611	8	8.2007	9	9.554	10	9.5814	8

While it has already been demonstrated through the above example that the correlation matrix of the realizations closely approximates that of the original variables, other facets of LHS must be considered in order to validate the fLHS algorithm as an appropriate proxy where uncertainty is either fuzzy or possibilistic. For one, traditional

LHS is an unbiased sampling algorithm for probabilistic data (Zhang and Pinder, 2003). Figure 7.4 shows that fLHS is an unbiased sampling technique for fuzzy and possibilistic data.

Table 7.5. Approximation of the target correlation matrix in Table 7.1

	1	**2**	**3**	**4**
1	1.0	0.704	0.695	0.606
2	0.704	1.0	0.585	0.688
3	0.695	0.585	1.0	0.680
4	0.609	0.675	0.680	1.0

In this figure, it is evident that the means of realizations of the parameters are very good approximations of median values (values with full membership or possibility). Moreover, it is expected that as the number of realizations increases, the correlation matrix for the realizations will more closely approximates the true correlation matrix. The root mean square error between the the approximate and target correlation matrices remains relatively constant regardless of the number of values drawn from each interval, though does decrease as the total number of intervals decreases.

7.4. Field Application

A simplified version of the Woburn, Massachusetts site was employed to demonstrate the practicality of the proposed algorithm. The domain is shown in Figure

7.5 and is roughly *3900 feet* north to south and about *3500 feet* west to east at its widest expanse . Two contaminant source locations are present, in the northeast (*4000 ppb*) and southwest (*1500 ppb*) areas of the domain. Two pumping wells, G and H, pump *150,000* ft^3/day and *250,000* ft^3/day, respectively. The northern boundary has a constant head of *65 feet* and the southern boundary of *50 feet*.

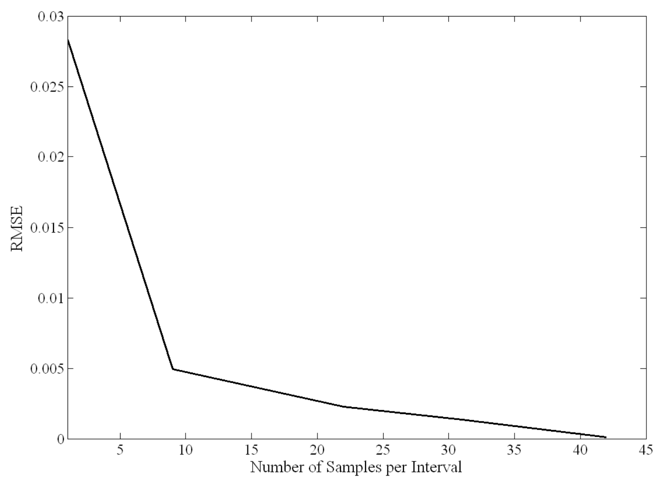


Figure 7.4. The errors in the means resulting from the fLHS algorithm decrease as the number of realizations increases, demonstrating unbiasedness

The locations of crisp hydraulic conductivity measurements are shown in Figure 7.5 (●). Notice that, for such a large domain, the available hydraulic conductivity measurements in this simplified case are quite few. Because hydraulic conductivity is so sparsely sampled (*11* measurements), without additional data, kriged estimates of hydraulic conductivity will likely be highly uncertain. Moreover, estimates of concentration would be of little value. As such, an additional *40* boring logs (▲ in Figure 7.5) were provided to a hydrogeologist charged with the task of translating the qualitative boring log data into fuzzy hydraulic conductivity values (Ross et al, 2007).

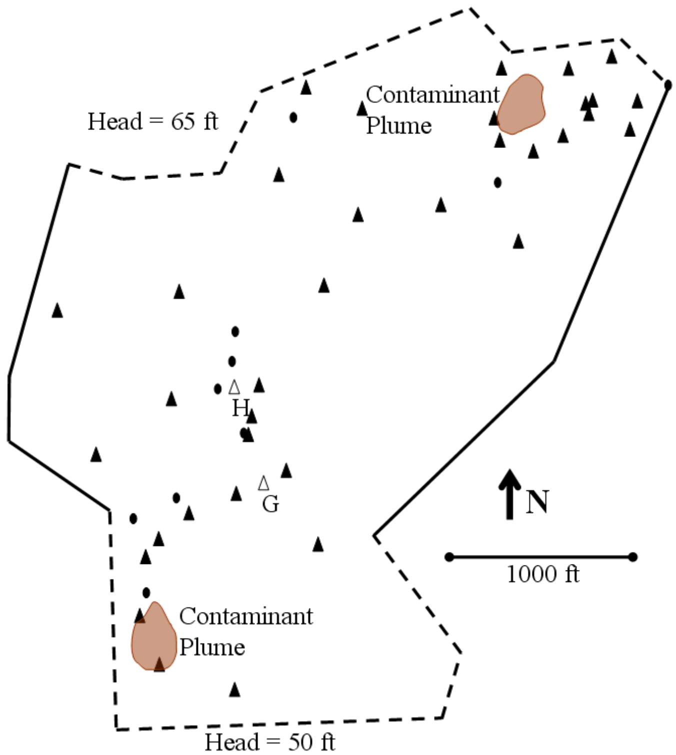


Figure 7.5. Domain outline for the simplified Woburn site; crisp hydraulic conductivity measurements (●) are few relative to the fuzzy measurements (▲)

Using a fuzzy Kalman Filtering algorithm (Ross et al, 2006), the resulting fuzzy hydraulic conductivity measurements were kriged in order to produce an estimated fuzzy hydraulic conductivity field (K-field) for the spatial domain. This fuzzy K-field was then updated with the *11* crisp measurements in order to produce a more precise posterior estimated K-field.

The numerical discretization of the areal, two-dimensional spatial domain produced *434* triangular finite element nodes, each of which is associated with a possibility distribution representing the uncertain hydraulic conductivity value at that location. Each of these possibility distributions was segmented into $M = 50$ nested intervals each sampled $J = 10$ times for $L = 500$ realizations, according to the fLHS algorithm described above. The mean hydraulic conductivity field is shown in Figure 7.6 (left) alongside the contours of the most possible hydraulic conductivity field (right). As can be seen in the figure, the two plots are practically identical, illustrating the unbiasedness of the fLHS algorithm (RMSE = 0.2). Each realization of *434* hydraulic conductivity values was used as input to groundwater fate and transport model (run for *5000 days*), which, in turn, was used to estimate concentration at all nodes. Ultimately, *500* values of concentration are estimated at each node.

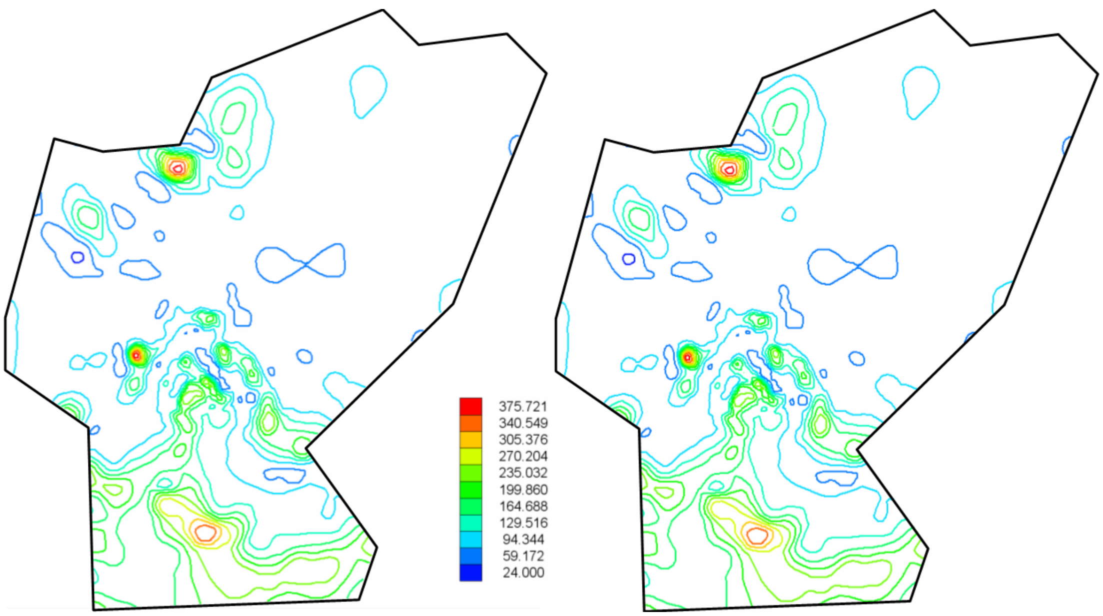


Figure 7.6. The means of the hydraulic conductivity realizations (left) are identical to the most possible values of the hydraulic conductivity possibility distributions (right)

Though the mean concentrations determined from the stochastic sampling of the possibility distributions can be calculated and plotted. However, because the original input parameters are characterized by possibility distributions it is appropriate that the uncertain concentration estimates be characterized in a possibilistic manner.

A number of formulae are available to calculate a possibility distribution from a probability distribution (Delgado and Moral, 1987; Dubois et al, 1991, 2004; Oussalah, 2000; Klir, 2006). However, the problem of inferring a possibility distribution from a data set is uncommon. An approach to accomplish such an inference was provided by Masson and Denœux (2006) when data are relatively few. Fortunately, in most groundwater flow and transport models, the number of nodes, and consequently, the number of sampled values per node, is many.

As such, a less computationally intensive alternative approach to possibility distribution calculation (Dubois et al, 1991, 2004) from histrogram data is employed herein. A histrogram, representing the uncertain concentration at node 300 (approximately *300 feet* southeast of well H) and given in Figure 7.7, is comprised of vectors of bin values $\Omega = \{\omega_1, \ldots, \omega_K\}$ and frequencies associated with each bin $\{n_1, \ldots, n_K\}$. In order to calculate the possibility π_i for each bin value, one follows the following steps:

a. Normalizes the frequencies of the bin values to probabilities $\{p_1, \ldots, p_K\}$
b. Defines a permutation σ of the indices $1, \ldots, J$ such that $p_{\sigma(1)} < p_{\sigma(2)} < \ldots < p_{\sigma(K)}$.
c. The reverse transformation σ^{-1} of the permutation facilitates the calculation of the possibility values. A reverse transformation records the ascending rank of

the probabilities in $\{p_1, \ldots, p_K\}$ (Masson and Denœux, 2006). For instance, if $K = 10$, and p_1 is the greatest probability, then $\sigma^{-1}(1) = 10$.

$$\pi_i = \sum_{\{j | \sigma^{-1}(j) \leq \sigma^{-1}(i)\}} p_j \tag{7-1}$$

d. Apply Equation (7-1) to calculate the possibilities associated with each bin.

Following these steps for the histogram in Figure 7.7 produces the discrete possibility distribution in Figure 7.8. The most possible concentration estimates at each node are plotted in Figure 7.9.

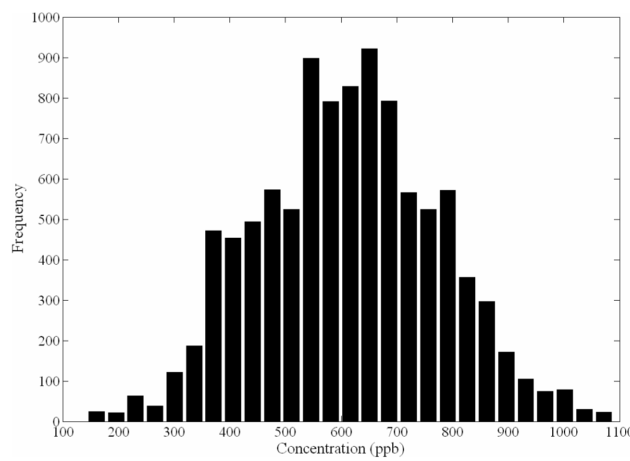


Figure 7.7. Histogram of concentration estimates for node *300* with *27* bins (model run for *5000 days*)

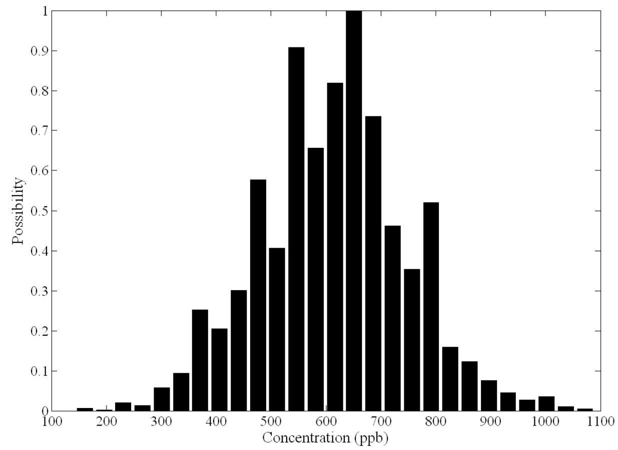


Figure 7.8. Discrete possibility distribution characterizing the uncertainty in the concentration estimate at node *300*, as calculated from the histogram in Figure 7.7 (model run for 5000 days)

7.5. Conclusions

The use of expert knowledge in engineering applications is easily used when such information is characterized with fuzzy sets. This is evident in groundwater flow and transport modeling, where fuzzy valued model input parameters such as hydraulic conductivity have been defined by experts. Whereas fuzzy set- and fuzzy logic-based approaches to propagating the uncertainty characterized by fuzzy sets and possibility distributions are adequate in some applications, they are impractical for groundwater modeling because they are strongly rooted in fuzzy set theory and fuzzy logic. The fLHS algorithm offers an alternative approach to uncertainty modeling that borrows some techniques from stochastic modeling.

By taking advantage of the properties of a possibility distribution, the fLHS algorithm used a modified fuzzy numerical sampling scheme to sample from equi-probable focal elements. In cases where the number of realizations is high, the fLHS algorithm essentially amounts to sampling the convex hull of the possibilistic model parameters. Though this process sounds similar to the vertex method and interval analysis, these methods, like the extension principle, fail to respect the correlation of the spatially distributed model inputs. In order to respect this spatial correlation, steps were borrowed from LHS, whereby the sample realization matrix is operated upon and permuted. The result is M individual realizations that are consistent with the physics of the field.

It is clear that the error between the most possible values of the model parameters and the calculated means of the sampled realizations is minimal after a small number of realizations (Figure 7.4). The convergence of the model concentration estimates was confirmed by sampling *10000* realizations for the simplified Woburn example; the contour plot of the most possible nodal concentration values was identical to Figure 7.9.

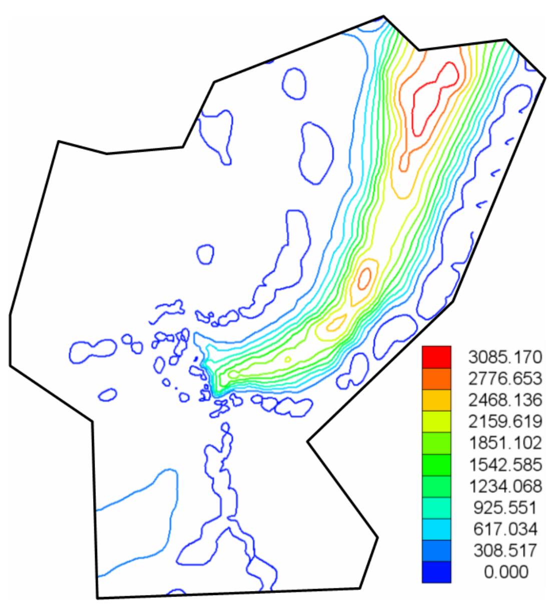


Figure 7.9. Most possible concentration values (model run for *5000 days*)

Even though concentration estimates are easily reduced to means and variances, this is inconsistent with the manner in which input model uncertainty is characterized. As such, each node's concentrations estimates are interpreted in a possibilistic manner. Since concentration estimates defined by possibility distributions may not be ideal, the most possible value at each node is taken, in order to plot the resulting concentration field. In stochastic modeling, this is analogous to considering only the most probable value at each node, as determined from a histogram. The analogy to mean values in possibility distribution amounts to defuzzifying (Klir and Yuan, 1995) the possibility distributions.

References

Bagtzoglou, A.C., Nedungadi, A., and Sagar, B. (1996). "A fuzzy rule-based model for flow simulation in heterogeneous media." *Proceedings of the 11th International Conference on Computational Methods in Water Resources*, 1, 629-637.

Bardossy, A. (1996). "The use of fuzzy rules for the description of elements of the hydrological cycle." *Ecological Modelling*, 85, 59-65.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1988). "Imprecise (fuzzy) information in geostatistics." *Mathematical Geology*, 20(4), 287-311.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). "Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy, A., Bogardi, I., and Duckstein, L. (1990a). "Fuzzy regression in hydrology." *Water Resources Research*, 26(7), 1497-1508.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990b). "Kriging with imprecise (fuzzy) variograms. I: Theory." *Mathematical Geology*, 22(1), 63-79.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990c). "Kriging with imprecise (fuzzy) variograms. II: Application." *Mathematical Geology*, 22(1), 81-94.

Bardossy, A., Bronstert, A., and Merz, B. (1995). "1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach." *Advances in Water Resources*, 18(4), 237-251.

Bardossy, A., and Disse, M. (1993). "Fuzzy rule-based models for infiltration." *Water Resources Research*, 29(2), 373-383.

Chanas, S., and Nowakowski, M. (1988). "Single value simulations of fuzzy variable." *Fuzzy Sets and Systems*, 25, 43-57.

Coppola, E.A., Duckstein, L., and Davis, D. (2002). "Fuzzy rule-based methodology for estimating monthly groundwater recharge in a temperate watershed." *Journal of Hydrologic Engineering*, 7(4), 326-335.

Delgado, M., and Moral, S. (1987). "On the concept of possibility-probability consistency." *Fuzzy Sets and Systems*, 21, 311-318.

Demmico, R.V. (2004). "Fuzzy logic and earth science: An overview." *Fuzzy Logic in Geology*, Demmico, R.V., and Klir, G.J., eds., Elsevier, Amsterdam.

Dhiman, S.D., and Keshari, A.K. (2003). “Quantifying uncertainties using fuzzy logic for groundwater driven contaminant exposure assessment.” *Probabilistic Approaches to Groundwater Modeling Symposium*, 236-247.

Dokou, Z. (2008). *Optimal Search Strategy for the Definition of a DNAPL Source*, Ph.D. Dissertation, University of Vermont.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1995). “Steady state groundwater flow simulation with imprecise parameters.” *Water Resources Research*, 31(11), 2709-2719.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1997a). “Numerical solute transport simulation using fuzzy sets approach.” *Journal of Contaminant Hydrology*, 27, 107-126.

Dou, C., Woldt, W., Dahab, M., and Bogardi, I. (1997b). “Transient ground-water flow simulation using a fuzzy set approach.” *Ground Water*, 35(2), 205-215.

Dou, C., Woldt, W., and Bogardi, I. (1999). “Fuzzy rule-based approach to describe solute transport in the unsaturated zone.” *Journal of Hydrology*, 220, 74-85.

Dubois, D., Foulloy, L., Mauris, G., and Prade, H. (2004). “Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities.” *Reliable Computation*, 10, 273-297.

Dubois, D., Prade, H., and Sandri, S. (1991). “On possibility/probability transformations.” *Proceedings of the Fourth International Fuzzy Systems Association World Congress* (*IFSA ’91*), 50-53.

Faybishenko, B.A. (2004). “Introduction to modeling of hydrogeologic systems using fuzzy differential equation.” *Fuzziness and Soft Computing*, 142, 267-284.

Ganoulis, J. (1996). “Modeling hydrologic phenomena.” *Revue des Sciences de L’Eau*, 9(4), 421-434.

Gelhar, L.W. (1986). “Stochastic subsurface hydrology from theory to applications.” *Water Resources Research*, 22(9), 135S-145S.

Gelhar, L.W. (1993). *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, NJ.

Højberg, A.L., and Refsgaard, J.C. (2005). “Model uncertainty – parameter uncertainty versus conceptual models.” *Water Science Technology*, 52(6), 177-186.

Issaks E.H., and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River.

Klir, G.J. (2006). *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley & Sons, Hoboken, NJ.

Larue, J., and Tyagi, A.K. (1997). "Groundwater simulation using fuzzy numbers." *Groundwater: An Endangered Resource*, Findikakis, A.N., and Stauffer, F., eds., ASCE, New York.

de Marsily, G., Delay, F., Teles, V., and Schafmeister, M.T. (1998). "Some current methods to represent the heterogeneity of natural media in hydrogeology." *Hydrogeology Journal*, 6, 115-130.

Masson, M., and Denœux, T. (2006). "Inferring a possibility distribution from empirical data." *Fuzzy Sets and Systems*, 157, 319-340.

Mathon, B, Ozbek, M., and Pinder, G.F. (2008). "Dempster-Shafer theory applied to uncertainty surrounding hydraulic conductivity." *Mathematical Geology*, in press.

McKay, M.D., Conover, W.J., and Breckman, R.J. (1979). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." *Technometrics*, 21, 239-245.

Nedungadi, A., Bagtzoglou, A.C., and Sagar, B. (1994). "A fuzzy methodology to solve the groundwater flow problem in the presence of uncertain hydrogeological data." *Applications of Artificial Intelligence in Engineering* IX, Adey, R.A., Rzevski, G., Russell, D.W., eds., Computational Mechanics Publications, Southampton, UK.

Nilsson, B., Højberg, A.L., Refsgaard, J.C., and Troldborg, L. (2007). "Uncertainty in geological and hydrogeological data." *Hydrology and Earth System Sciences*, 11, 1551-1561.

O'Hagan, A., and Oakley, J.E. (2004). "Probability is perfect, but we can't elicit it perfectly." *Reliability Engineering and System Safety*, 85, 239-248.

Oussalah, M. (2000). "On the probability/possibility transformations: A comparative analysis." *International Journal of General Systems*, 29(5), 671-718.

Ozbek, M.M., and Pinder, G.F. (2006). "Non-probabilistic uncertainty in subsurface hydrology and its applications: An overview." *Water, Air, and Soil Pollution: Focus*, 6, 35-46.

Piotrowski, J.A. (1996). "Estimation of hydrogeological parameters for groundwater modeling with fuzzy geostatistics: Closer to nature?" *Calibration and Reliability in Groundwater Modelling*, 237, 511-520.

Prasad, R. K., and Mathur, S. (2007), "Groundwater flow and contaminant transport simulation with imprecise parameters." *Journal of Irrigation and Drainage Engineering.,* 133(1), 61-70.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). "Fuzzy kalman filtering of hydraulic conductivity." *Computations Methods in Water Resources Conference 26th International Conference (CMWRXVI)*.

Ross, J., Ozbek, M. and Pinder, G.F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*. 39(8), 765-780.

Schultz, K., and Huwe, B. (1997). "Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach." *Journal of Hydrology*, 201, 211-229.

Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University, Princeton.

Vernieuwe, H., De Baets, B., Verhoest, N.E.C. (2002a). "Simulation of one-dimensional water movement in the unsaturated zone by means of a first order Takagi-Sugeno model." *IEEE International Conference on Fuzzy Systems*, 2, 1069-1074.

Vernieuwe, H., Verhoest, N.E.C., and de Baets, B. (2002b). "A Takagi-Sugeno model of flow in the unsaturated zone." *Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality (Proceedings of ModelCARE 2002)*, 277, 79-85.

Vernieuwe, H., Verhoest, N.E.C., de Baets, B., Hoeben, R., and de Troch, F.P. (2007). "Cluster-based fuzzy models for groundwater flow in the unsaturated zone." *Advances in Water Resources*, 30, 701-714.

Winter, C.L. (2004). "Stochastic hydrology: practical alternatives exist." *Stochastic Environmental Research and Risk Assessment*, 18(4), 271-273.

Zadeh, L. (1965). "Fuzzy sets", *Information Control*, 8, 338-353.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1(1), 3-28.

Zhang, Y., and Pinder, G. (2003). “Latin hypercube sample selection strategy for correlated random hydraulic conductivity fields.” *Water Resources Research*, 39(8), doi:10.1029/2002WR001822.

8: ALEATORIC AND EPISTEMIC UNCERTAINTY IN GROUNDWATER SIMULATION

The characterization of aleatory hydrogeological parameter uncertainty has traditionally been accomplished using probability theory. However, when consideration is given to epistemic, as well as aleatory uncertainty, probability theory is not necessarily appropriate for capturing the uncertainty in parameter uncertainty.

This is especially the case where expert opinion is regarded as an appropriate source of information. When experts opine upon the uncertainty of a parameter value, both aleatoric and epistemic uncertainties are introduced and must be modeled appropriately.

Herein, we propose a novel approach to expert-provided parameter uncertainty characterization, where random sets formalize expert knowledge and fuzzy sets are used to propagate this uncertainty to model estimates of contaminant transport. The resultant random set-based concentration estimates are shown to be more general than the corresponding random variable estimates. In some cases, the random set-based results are shown as upper and lower probabilities that bound the corresponding random variable's cumulative distribution function.

8.1. Introduction

Uncertainty in groundwater flow and transport modeling comes in two forms: aleatory and epistemic. Such distinctions in uncertainty are most often identified in risk assessment and reliability engineering (Helton et al, 2000a,b; Hofer et al, 2002; Helton and Oberkampf, 2004; Helton et al, 2004; Oberkampf et al, 2004). Aleatory uncertainty, also called stochastic or variable uncertainty, refers to uncertainty that cannot be reduced by more exhaustive measurements or a better model. Epistemic uncertainty, or subjective uncertainty, on the other hand, refers to uncertainty that can be reduced.

Despite these apparent distinctions in uncertainty, probability theory alone has traditionally been used to characterize both forms of uncertainty in engineering applications (Apostolakis, 1990; Helton et al, 2004); and, while it is commonly accepted that probability theory is ideal for the characterization of aleatory uncertainty (Ganoulis, 1996), the facility with which probability theory effectively captures epistemic uncertainty has been called into question (O'Hagan and Oakley, 2004), especially with the introduction of a number of alternative methods of epistemic uncertainty characterization (Choquet, 1954; Zadeh, 1965; Shafer, 1976; Zadeh, 1978). Of the alternate methods, random set theory (Zadeh, 1965) provides an intuitive means for both epistemic and aleatory uncertainty characterization.

Whereas probability theory's basic tool for uncertainty characterization is the probability density function (PDF), random set theory is predicated upon the assignment of probabilities to intervals, rather than point values. The PDF has been commonly applied to the characterization of subjective knowledge, despite the fact that human thought and discourse do not operate in the form of means, variances and probabilistic models. While some probability theorists suggest that the Dutch Book, an alternative framework for identifying subjective probabilities, is more intuitive, it does not necessarily apply to all engineering problems.

Random set theory (Helton and Oberkampf, 2004; Joslyn and Kreinovich, 2005), however, is a general approach to subjective knowledge characterization, and, in addition, random sets can be transformed into fuzzy sets (Joslyn and Booker, 2004; Joslyn and Ferson, 2005), with little difficulty. The transformation from random sets to fuzzy sets facilitates solution of groundwater flow and transport model equations characterized by uncertainty.

Though a few applications of fuzzy set theory to expert knowledge characterization in geohydrologic applications have been identified (Bardossy et al, 1989; Bardossy et al, 1990a,b,c; Dou et al, 1995; Bagtzoglou et al, 1996; Dou et al, 1997a,b; Fang and Chen, 1997; Dou et al, 1999; Demmico and Klir, 2004; Guan and Aral, 2004; Ross et al, 2006; Ross et al, 2007), an explanation of the distinctions between probability and fuzzy set theories in the characterization of epistemic knowledge is lacking. Moreover, the clarification of the roles played by aleatory and epistemic uncertainties in geohydrology is needed.

The purpose of this paper is to frame uncertain hydraulic conductivity information in terms of aleatory and epistemic uncertainty, propagate this uncertainty in space and time using a groundwater flow and transport model and compare the resulting two forms of uncertainty in the calculated concentration values. In the aleatory case, traditional probability theory is used to characterize the uncertainty in hydraulic conductivity measurements. In the epistemic case, random set theory captures expert knowledge regarding this uncertainty in addition to the epistemic uncertainty implicit in expert knowledge. The alternative uncertainty characterizations are discussed in this paper in terms of the model's hydraulic conductivity input and concentration output values.

8.2. Probabilistic Uncertainty

In groundwater modeling problems where few hydraulic conductivity measurements are few, a hydraulic conductivity field is assumed to be comprised of a few large subdomains of equal hydraulic conductivity, like the representation of the Woburn, Massachusetts site presented in Figure 8.1, simplified from the model developed by the U.S Geological Survey (Lima and Olimpio, 1989).

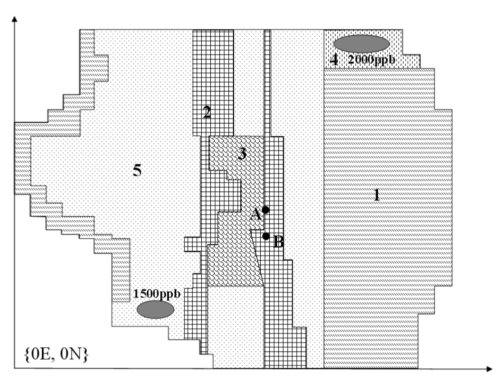


Figure 8.1. Inversely estimated hydraulic conductivity field; different numbers identify fields with distinct hydraulic conductivity values; gray ovals represent contaminant source locations. Dots denoted as A and B are pumping wells

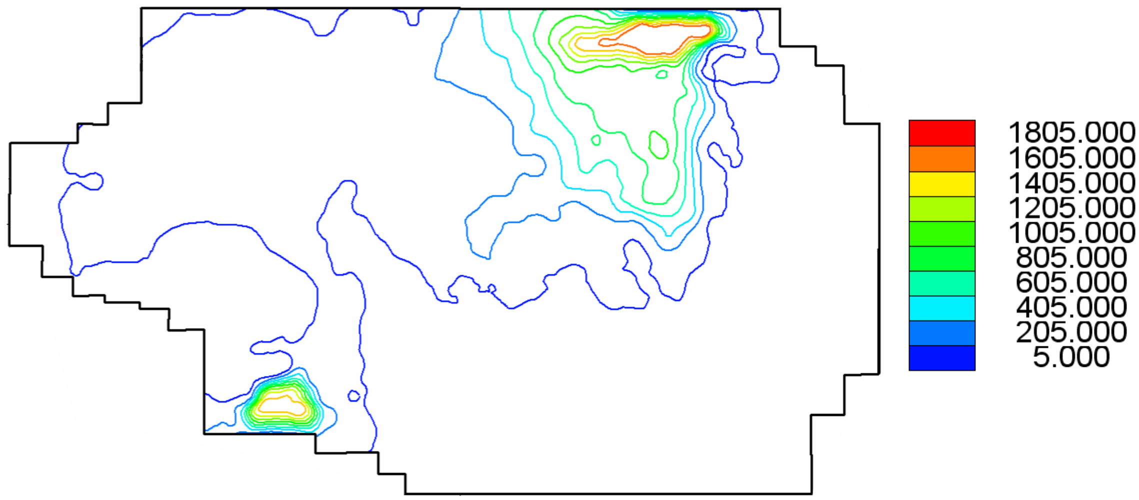
A small number of hydraulic conductivity measurements are available in each of these hydraulic conductivity subdomains. For a given patch, a representative random variable hydraulic conductivity value can be constructed from the mean and variance of that subdomains's measurement data set. However, these measurements are inherently uncertain and both the mean and variance of the measurement data set may not be representative of the subdomain's random variable hydraulic conductivity. That is, because the measurements are uncertain, the variance of the measurement data set may likely be a low estimate of the uncertainty of the subdomain's hydraulic conductivity random variable. In order to obtain a better estimate of the uncertainty surrounding a subdomain's representative hydraulic conductivity value, an appropriate expert is asked to provide some measure of uncertainty in the form of a 95% confidence interval. This interval defines two values between which the expert is 95% certain the true value of hydraulic conductivity lies. Given a mean value, the assumption of hydraulic conductivity's lognormality and this 95% confidence interval a PDF defining the hydraulic conductivity random variable can be constructed.

The mean hydraulic conductivity values associated with the domain in Figure 8.1 are provided in Table 8.1. Also provided are the expert provided confidence intervals for the five formations and the resulting calculated variances. The intervals for three of the formations were specified to approximate two orders of magnitude variation in hydraulic conductivity in three of the formations. The remaining two formations were assigned zero variation for computational convenience. Total correlation was assumed within each formation and zero correlation is assumed between the units. Constant head boundaries were specified on the left (*75 ft*) and right (*125 ft*) boundaries, and no-flow boundary conditions were specified along the top and bottom of the domain. Contaminant sources (grey ovals in Figure 8.1) are located in formations four and five at concentrations of *2000 ppb* and *1500 ppb*, respectively. Finally, two pumping wells were placed in formation two. Wells A and B pump at *2500 feet^3/day* and *1500 feet^3/day*, respectively.

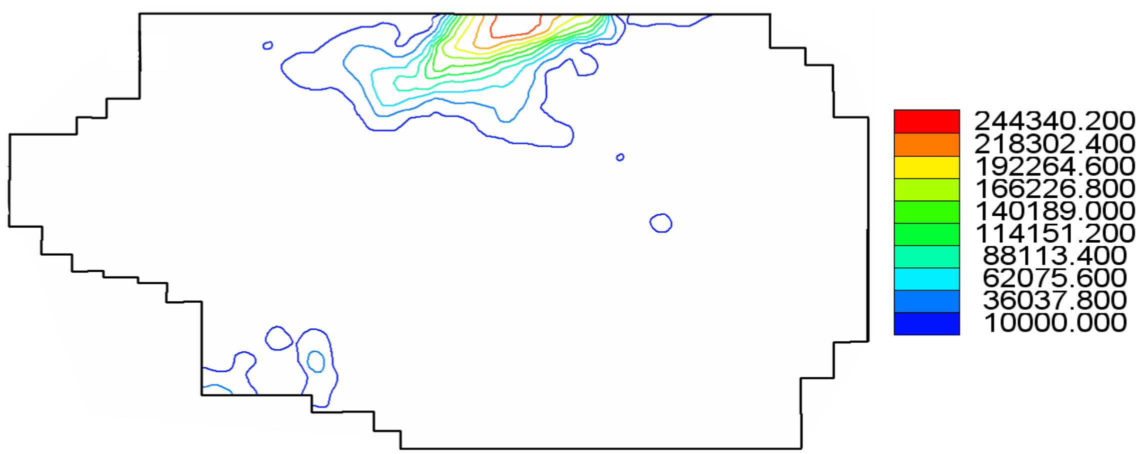
Table 8.1. Hydraulic conductivity random variable properties for the domain in Figure 8.1

Unit Number	Mean Ln(K), ft/day	Confidence Interval	Variance
1	-6.101874888	[-7.9, -4.3]	0.834001
2	0.693147181	[0.69, 0.69]	0
3	0.311704003	[-1.4, 2.1]	0.794899
4	4.514645016	[2.3, 6.7]	1.255079
5	7.60090246	[7.6,7.6]	0

Monte Carlo analysis (Metropolis and Unam, 1949) was performed to propagate uncertainty through the groundwater model and produce stochastic concentration values throughout the spatial domain. The concentration mean and variance values are contoured in Figure 8.2. Four of the calculated cumulative distribution functions (CDF) of the concentration random variables are plotted in Figure 8.3 (Case 1), along with the CDFs for two other cases with different variances (see below). The pertinent data for these random variables are provided in Table 8.2.



A



B

Figure 8.2. Concentration mean (A) and variance (B) contours show the contaminant source in the upper right-most spreading more than the lower left contaminant source; variances are greater for concentrations resulting from this source

The probabilistic form of these results is significantly dependent upon the certainty with which the expert can define the confidence intervals and the appropriateness of the lognormality assumption, amongst other factors. In fact, the longstanding assumption of lognormality for hydraulic conductivity may not be correct in all cases (Ricciardi et al, 1998; Mathon et al, 2008). The true random variable may actually be best defined by any one of an infinite number of alternative probability functions, two of which are plotted in Figure 8.3 (* and + curves). The taller, steeper distributions (Case 2, * curves) result from the smaller hydraulic conductivity confidence intervals in Table 8.2.

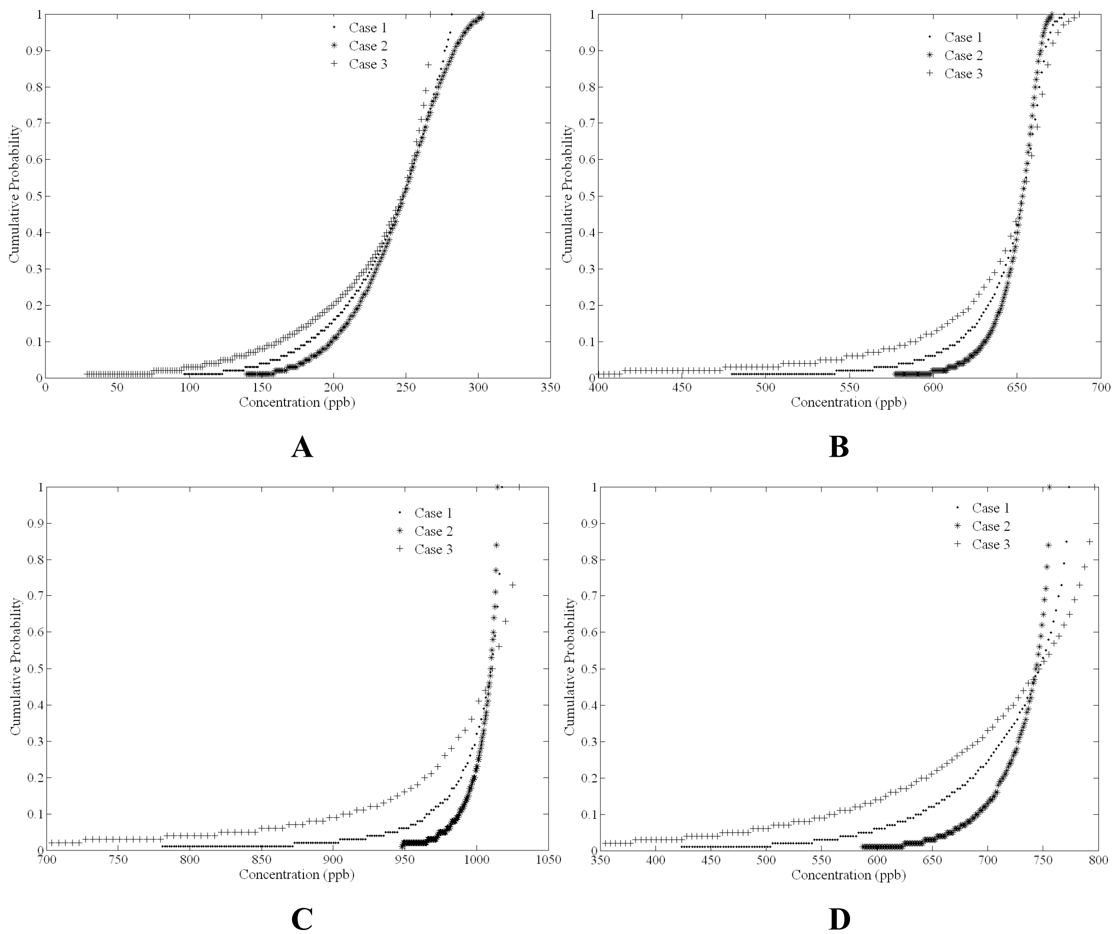


Figure 8.3. Random variables representing the concentrations at locations {1107E, 635N} (A), {1590E, 1030N} (B), {1450E, 682N} (C), {1420E, 634N} (D). Intuitively, the variance changes through out space and as the expert-provided hydraulic conductivity confidence intervals are narrow (Case 2) relative to the case of interest (Case 1) and relatively wide (Case 3). The data for these random variables are given in Table 8.1. The confidence intervals for the three cases are provided in Table 8.2

Table 8.2. Expert-provided confidence intervals for the base case (Case 1), a high-certainty case (Case 2) and a low-certainty (Case 3)

Unit Number	Confidence Interval, Case 1	Confidence Interval, Case 2	Confidence Interval, Case 2
1	[-7.9, -4.3]	[-6.88, -5.32]	[-9.04, -3.16]
2	[0.69, 0.69]	[0.69, 0.69]	[0.69, 0.69]
3	[-1.4, 2.1]	[-0.47, 1.10]	[-2.43, 5.06]
4	[2.3, 6.7]	[2.92, 6.08]	[0.98, 8.04]
5	[7.6,7.6]	[7.6,7.6]	[7.6,7.6]

The longer, wider distributions (Case 3, + curves) are the random variables resulting from a less certain expert, who provided wider confidence intervals. It is evident that variances depend significantly upon the confidence intervals provided by the expert. Lower variances are associated with tighter confidence intervals provided by the expert for the hydraulic conductivity data, and higher variances with wider confidence intervals. More importantly, small variations in the expert's certainty regarding model input parameters can produce larger changes in the model output random variables (Figure 8.3).

8.3. Random Set Uncertainty

An approximation of a hydraulic conductivity subdomain's uncertain hydraulic conductivity is the discrete CDF calculated from the available measurements in that subdomain. However, as stated above, these individual measurements are, themselves, uncertain. Intuitively, then, this measurement error warrants characterization before the subdomain's hydraulic conductivity random variable can be defined. Rather than restrict an expert to a single interval in an attempt to capture both the measurement error as well as the stochasticity of hydraulic conductivity throughout a particular hydraulic conductivity zone, it is more intuitive to permit the expert to opine upon the uncertainty in the measurement values that helped determine the random variable in the Monte Carlo approach, above. It has been demonstrated that an appropriate expert can opine upon the uncertainty of hydrogeological measurements simply by assigning an interval in which the true value is expected to lie (Mathon et al, 2008). For example, knowing the measuring device (*i.e.* pump test, slug test), and expert can simply define such an interval by stating that the true value lies within *x* orders of magnitude of the measured value. Thus, in any one hydraulic conductivity subdomain with *m* equiprobable measurements, a collection of *m* equiprobable intervals is defined, representing *m* uncertain measurements. Where the measurements are used to construct a discrete CDF, these expert-provided intervals essentially bracket the unknown true random variable hydraulic conductivity (*i.e.* Figure 8.4). A collection of these random intervals (Helton and Oberkampf, 2004; Joslyn and Kreinovich, 2005), or random sets, are isomorphic to a Dempster-Shafer body of evidence (Joslyn and Booker, 2004). In this framework, a random set *A* and associated probability function *m* are collectively referred to as a focal element {*A, m)*}. A random set can be transformed into upper and lower probability bounds, around the unknown true random variable, called belief, *bel(A)*, and plausibility, *pl(A)*.

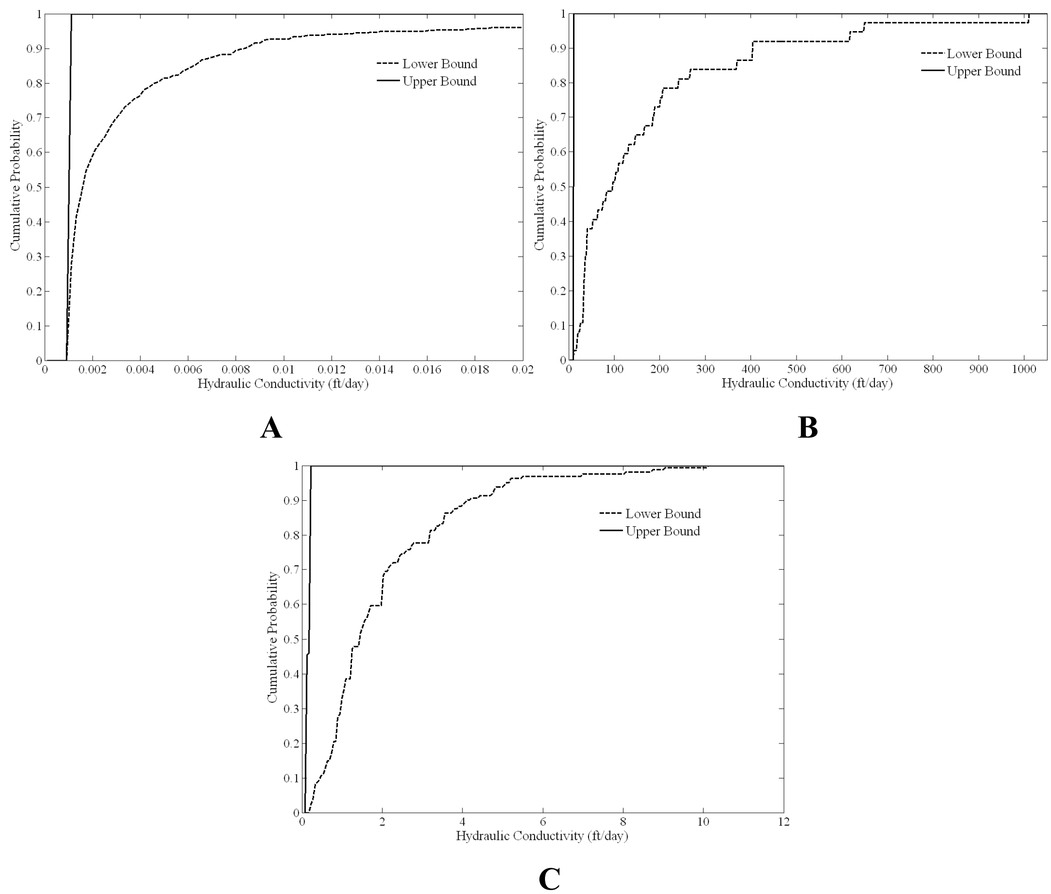


Figure 8.4. Cumulative random set hydraulic conductivity values for Zone 1 (A), Zone 2 (B) and Zone 5 (C), determined by the +/- two orders of magnitude uncertainty on available measurements

Because the random set definition requires less precision from the opining expert, the resulting belief and plausibility functions do not impose false precision in the parameter uncertainty characterization and avoid any inaccuracies that result from forcing an expert to provide a single random variable. Moreover, no probability model need be selected or assumed, which is desirable in light of the above-mentioned inaccuracies in the lognormality assumption for hydraulic conductivity.

Formally, a hydraulic conductivity random set (A, m_A), is defined on the Cartesian product $\boldsymbol{K}=K_1\times\ldots\times K_{n1}$, where K_j denotes the domain of the hydraulic conductivity value and $n1$ is the number of uncertainty hydraulic conductivity zones (in the trial case, $n1 = 3$). In this formal definition, m_A is a function mapping elements, $I_A \in A$, of A to the interval $[0, 1]$,

$$m(I_A) :\to [0,1] \quad I_A \in A$$

$$\sum_{I_A | I_A \in A} m_A(I_A) = 1$$

Since the conductivity zones are assumed uncorrelated in the trial case presented herein, the n1-dimensional random sets are marginalized to random sets (A_j, m_A^j), $j = 1,\ldots,n1$, defined solely upon the individual domains K_j. In other words, one can specify (A, m_A) by means of n1 stochastically independent random sets.

Each of these random sets is comprised of a finite number of intervals I_{Aj}, $I_{A_j} \in A_j$, or focal elements, each associated with a probability mass assignment $m(I_{A_j})$. The propagation of the random set-based hydraulic conductivity values necessitates the use of a tool to extend the flow and transport model such that it can operate upon these focal elements. Such an extension would permit the calculation of concentration focal elements that can be aggregated into a concentration random set at each location throughout the spatial domain.

Consider the extension of the transport model $\boldsymbol{y} = f(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1,\ldots,x_{n1})$ is the vector of uncertainty conductivities in n_1 zones and $\boldsymbol{y} = (y_1,\ldots,y_{n2})$ is the vector of uncertain concentration values at n_2 nodes. Assuming, without loss of generality, that $n1 = 1$ and $K_1 = \Omega = \{k_1,\ldots,k_L\}$, the domain of L possible hydraulic conductivity values, where focal elements are subsets of Ω. The random set extension principle (Dubois and Prade, 1991) defines the random set concentration (B, m) at node i as

$$B = \{f(I_A) | I_A \in A\}$$

and

$$m(I_B) = \{m(I_A) | I_B = f(I_A)\}$$

where I_B represents a focal set of concentrations that is an image of a hydraulic conductivity focal element I_A through the transport model f. This concentration focal element is defined by

$$I_B = f(I_A) = \{f(k) | k \in I_A\}$$

A significant drawback to the random set extension principle is its computational intensity, since a random set uses the power set of its domain rather than the domain itself. Therefore, an approximation of the random set hydraulic conductivities, such that the extended transport model can be carried out over Ω rather than its power set, is prudent.

In general, simplifying or approximating a random set means approximating it by another random set, in which the number of subsets containing relevant informations is reduced (Bauer, 1996). One approach to approximate a random set that, presented in Dubois and Prade (1990), is adopted in this paper. This approximation uses the following steps: 1) formation of a collection of subsets of the original hydraulic conductivity domain Ω, and 2) allocation of basic probability masses to the sets from the first step using a process that is optimal in the sense that the resulting hydraulic conductivity focal elements of the approximating random set are the smallest in size, effectively minimizing the amount of imprecision introduced by the approximation.

Approximating a random set by the method above results in nested focal elements, which comprise a fuzzy set F (Klir and Yuan, 1995), a special type of random set. This fact permits the solution of the transport model over the domain Ω, rather than its power set. For the exact representation of the approximating conductivity fuzzy set on Ω, the membership function is calculated by the one-point coverage function (Goodman and Nguyen, 1985) for random sets:

$$\textit{for all } k \in \Omega\text{: } \mu_F(k) = \sum_{C_i : k \in C_i} m(C_i)$$

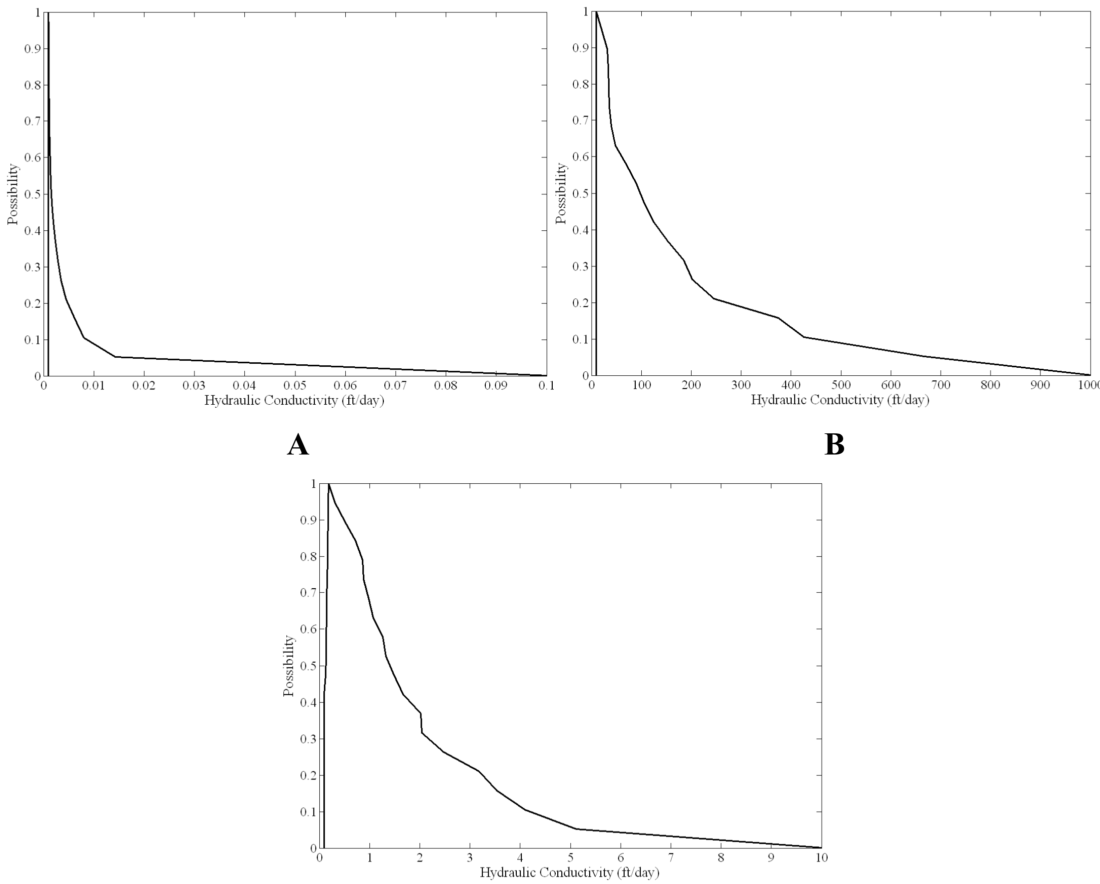
where C_i are the focal elements of the approximating hydraulic conductivity random set and μ_F denotes the membership function of the fuzzy set F. This fuzzy set representation of the uncertain hydraulic conductivity values allows for the use of the special case of the extension principle, described above, for fuzzy sets (Dubois and Prade, 1991):

$$\textit{for all } y \in f(\Omega)\textit{:}\ \mu_{f(F)}(y) = \sup\{\mu_F(k) \mid y = f(k)\}$$

where $\mu_{f(F)}$ represents the membership function of the fuzzy concentration at a given node.

Where fuzzy sets are used to approximate random sets, the application of the extension principle (Klir and Yuan, 1995) to the model equations is relatively straightforward and has precedent in hydrogeological applications (Dou et al, 1995; Dou et al, 1997a,b; Prasad and Mathur, 2007). Nevertheless, in this case the vertex method (Ross, 2004), an approximation to the extension principle, was applied. The vertex method results in, for each α-cut (an interval created by the horizontal cut of a fuzzy set at a given α, or membership value), eight concentration values at each location due to the eight possible permutations of the three lower and three upper hydraulic conductivity bounds (from three uncertain hydraulic conductivity values) of the α-cut. The minimum of these eight values is taken as the lower bound of the concentration α-cut, and the maximum is assigned as the upper bound. If the function being extended (i.e. the transport model) is nonlinear with respect to its variables (in our case three hydraulic values) then there is a possibility that the function will take smaller or larger values for a combination of three hydraulic conductivity values that are not necessarily a permutation of the bounds, but rather sampled anywhere within their respective α-cuts. If one computes with the eight concentration values, one essentially assumes that the concentration at that node cannot get any smaller than the minimum of the eight values and it can not get any larger than the maximum of the eight values. In other words one admits that there is no need to investigate the entire hydraulic conductivity α-cuts in order to determine the α-cut bounds of the concentration at any node. In this exercise, it is assumed that the concentration variation over the α-cuts of the fuzzy hydraulic conductivities is essentially monotonic. Therefore, it suffices to evaluate the concentrations at the bounds of the hydraulic conductivity α-cuts, as accomplished by the vertex method.

Figure 8.5 provides examples of a fuzzy set defining an uncertain hydraulic conductivity value. The interpretation of these fuzzy sets in Figure 8.5 that is most pertinent to this topic is that of a possibility distribution (Zadeh, 1978). Where model inputs are defined as fuzzy sets, model estimates of concentration are interpreted as possibility distributions. A possibility distribution defines for each value along the horizontal axis the degree to which that value is possible, given available evidence. In Figure 8.5 (*top left*), for instance, the hydraulic conductivity value 9.9×10^{-4} *feet/day* is most possible. This is similar to probability theory, whereupon inspection of the discrete probability distribution in Figure 8.3 reveals that the concentration at location {1107E, 635N} is most probably *237.71 ppb*.



C

Figure 8.5. Possibilistic approximations for the random set provided by the expert for the hydraulic conductivity values associated with Zone 1 (A), Zone 2 (B) and Zone 5 (C). The hydraulic conductivity values for Zones 3 and 4 are considered certain and precise

Whereas, through the transformation to fuzzy sets, random sets offer a relatively facile strategy to computation with uncertainty, the most significant advantage to the random set approach is its potential to characterize both aleatory and epistemic uncertainty. Their foundation lies in probability theory, which, as mentioned above, is ideally suited for aleatory uncertainty characterization. On the other hand, random sets are less specific, or less precise, than random variables (Joslyn and Booker, 2004). This imprecision is a form of epistemic uncertainty. An expert who provides a body of evidence to characterize the uncertainty regarding some hydraulic conductivity measurement actually admits these two forms of uncertainty. The natural randomness (aleatory uncertainty) of hydraulic conductivity is captured by the stochastic nature of random sets. Due to an expert's inability (epistemic uncertainty) to precisely define this natural randomness, a random set merely provides bounds on the exact random variable, imprecisely defining the random variable.

The starting point for the application of random set-based uncertainty characterization is similar to the probabilistic approach above. Given the measurements in each zone in Figure 8.1, the expert, armed with knowledge of the measurement technique and aquifer characteristics, provides the afore-mentioned intervals on the each measured value by specifying that the true hydraulic conductivity value lies within +/- two orders of magnitude of the measurement. As mentioned above, where the measurements in a particular zone are used to construct an approximate CDF, these expert-provided intervals become upper and lower bounds on the zone's true random variable hydraulic conductivity. Figure 8.4 shows these upper and lower bounds for zones one, two and five. As in the Monte Carlo approach, the hydraulic conductivity values for zones three and four are considered certain and precise. The possibility distribution approximations of these random sets are shown in Figure 8.5.

Though the set of intervals and associated probabilities provided by the expert are a natural extension of the confidence interval in the Monte Carlo approach above, they comprise a greater amount of information. As such, the bodies of evidence in provided by the expert capture both the uncertainty surrounding the mean hydraulic conductivity values and the imprecision with which the expert can truly characterize this uncertainty.

Using the transformations summarized above, the bodies of evidence for each formation were ultimately converted to possibility distributions (Figure 8.5). The vertex method (Ross, 2004), an approximation to the extension principle, was applied to the finite element approximation equations of the groundwater flow and transport model in order to propagate the possibilistic uncertainty through to the concentrations values. As a result, uncertain concentration estimates are described by possibility distributions. The possibilistic concentration values can be transformed into upper and lower bounds on the unknown random variable. The resulting bounds for the nodes of interest are plotted in Figure 8.6 with the corresponding random variables in cumulative distribution function form (dashed lines) from Figure 8.3.

8.4. Discussion of Results

The outcomes of the two approaches to uncertainty characterization are significantly distinct. Intuitively, the strictly probabilistic approach produces random variable concentration values, whereas the random set-based approach results in upper and lower probability bounds. Assuming that uncertainty in both approaches was characterized by the same expert or different experts with the same understanding of pertinent data, the upper and lower probabilities should bracket the corresponding probability distribution (produced by the stochastic approach) at a given location.

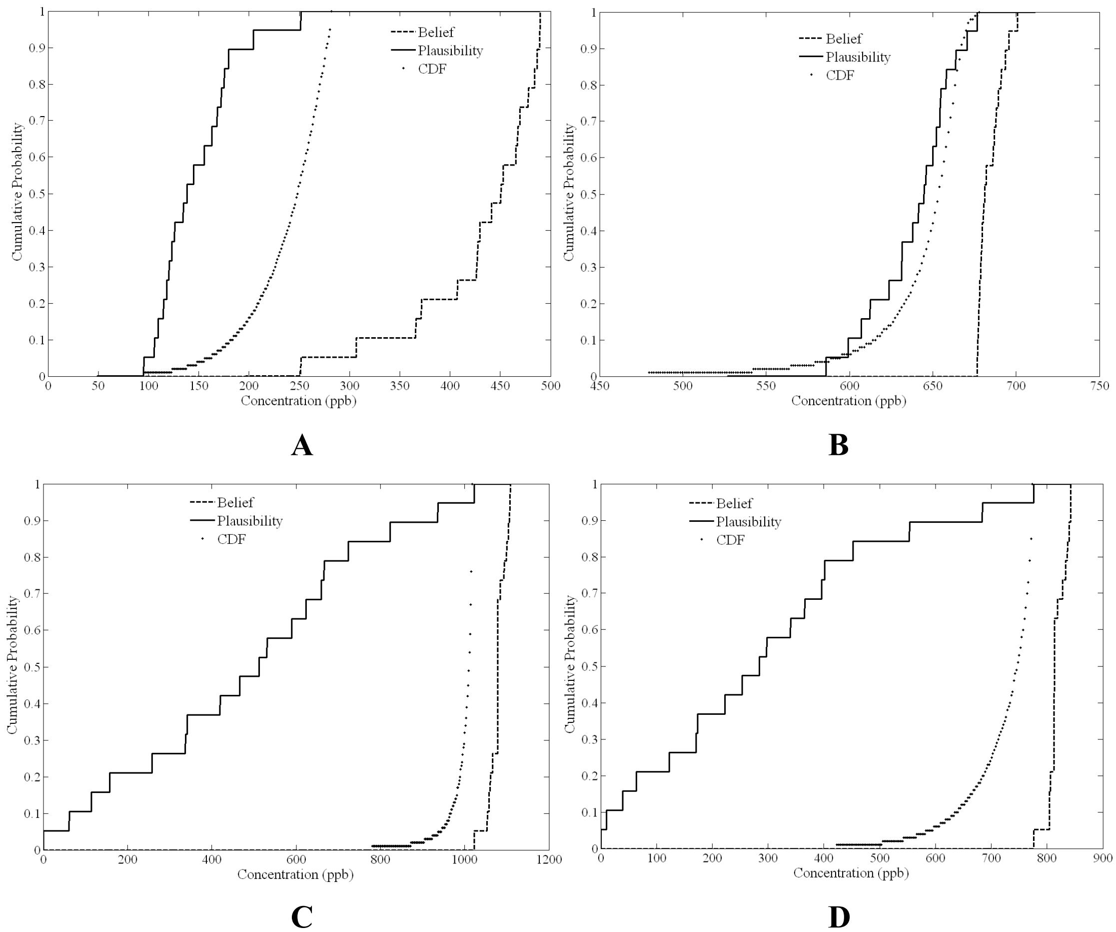


Figure 8.6. The cumulative belief and plausibility curves are quite narrow and do not entirely bound the corresponding cumulative distribution function for nodes 234 (B) and 162 (C), but are rather wide and do bound the cumulative distribution functions for nodes 400 (A) and 190 (D)

Consider the same nodal locations whose random variables are plotted in Figure 8.3. The upper and lower probabilities for these same nodal locations are plotted in Figure 8.6. Note that, at some nodes, the upper and lower probabilities entirely bound the corresponding random variable, whereas, other nodes do not entirely bound the corresponding cumulative distribution function.

Upon inspection of the variance contours in Figure 8.2 and the probability distributions in Figure 8.3, it is evident that the uncertainty (variance) associated with concentration values changes throughout the spatial domain. Those concentration values with lower variances are presumably more certain than those values with high variances, though higher variances may also be associated with means of greater magnitude. Random set-based probability bounds, however, are independent of the magnitude of the concentration values and provide a true means of uncertainty identification. Wider intervals signify more uncertainty in concentration estimates, and, as a result, locations where more data is warranted. For instance, the estimate in Figure 8.6D is more uncertain than that in Figure 8.6B and, as such, is in need of additional data.

8.5. Conclusions

Because thorough hydrogeological investigations cost significant amounts of money and time, an efficient means of data acquisition is valuable. One such means is expert knowledge extraction for uncertainty characterization. Uncertainty has two forms, aleatory and epistemic, and expert insight has generally applied to the characterization of aleatory, or irreducible, uncertainty. However, inherent in human thought, is a certain degree of epistemic, or reducible, uncertainty. Thus, appropriate characterizations of expert knowledge should represent both uncertainty types.

Probability theory requires extraordinary precision of an expert and is not well suited for uncertainty characterization. The hazard associated with applying traditional probability theory is that the opining expert may provide inaccurate and artificially precise characterizations of the random variable that best captures the naturally stochastic nature of hydraulic conductivity.

Random sets were introduced in this paper as a more appropriate means to the characterization of both aleatory (the random variable) and epistemic (expert-characterized measurement error) uncertainty. This method of uncertainty characterization provides a means of bounding this true, yet unknown random variable, accurately capturing the imprecise nature of expert knowledge. It is important to reiterate, as well, that uncertainty characterization via random sets eliminates the need for any model definition or assumption.

Computationally, the application of the vertex method provides a direct approach to operating upon fuzzy-based transformations of random set data with standard modeling equations. Because the vertex method is based upon interval arithmetic, little knowledge of fuzzy set theory is required for its implementation. Though the extension principle provides greater accuracy, it is more computationally intensive than the vertex method, and the latter approach is a fair approximation to the extension principle given the perceived monotonicity of the concentration values with α-cuts of the input hydraulic conductivity values.

While the representation of model concentration estimates as upper and lower probabilities provides a transparent comparison between the random variable and random set approaches defined above, the utility of data in such a form may not be immediately obvious. Nevertheless, the representation of concentration estimates as possibility distributions, a form equivalent to the probability bounds, user-friendly. Inspection of a possibility distribution reveals not only the most possible concentration value, but also a range of concentrations that are also possible to lesser and varying degrees. In fact, algorithms have been developed to complement and update possibilistic model estimates with updated information (Fruhwirth-Schnatter, 1993; Pan and Klir, 1996; Yang, 1997; Ross et al, 2007).

References

Apostolakis, G. (1990). "The concept of probability in safety assessments of technological systems." *Science*, 250(4986), 1359-1364.

Bardossy, A., Bogardi, I., and Kelly, W. E. (1989). "Geostatistics utilizing imprecise (fuzzy) information." *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy A., Bogardi, I., and Duckstein, L. (1990a). "Fuzzy regression in hydrology." *Water Resources Research*, 26(7), 1497-1508.

Bardossy, A., Bogardi, I., and Kelly, W. E. (1990b). "Kriging with imprecise (fuzzy) variograms I: Theory." *Mathematical Geol*ogy, 22(1), 63-79.

Bardossy, A., Bogardi, I., and Kelly, W. E. (1990c). "Kriging with imprecise (fuzzy) variograms II: Application." *Mathematical Geology*, 22(1), 81-94.

Bauer, M. (1996). "Approximation algorithms and decision making in the Dempster-Shafer theory of evidence: An empirical study." *International Journal of Approximate Reasoning*, 17, 217-237.

Choquet, G. (1954). "Theory of capacities." *Annales de L'Institut Fourier*, 5, 131-295.

Demmico, R. V., and Klir, G. J. (2004). *Fuzzy Logic in Geology*, Elsevier, Amsterdam.

Dou, C., Woldt, W., Bogardi, I., and Dahab, M. (1995). "Steady state groundwater flow simulation with imprecise parameters." *Water Resources Research*, 31(11), 2709-2719.

Dou, C., Woldt, W., Bogardi, I., and Daheb, M. (1997a). "Numerical solute transport simulation using fuzzy sets approach." *Journal of Contaminant Hydrol*ogy, 27(1-2), 107-126.

Dou, C., Woldt, W., Daheb, M., and Bogardi, I. (1997b). "Transient ground-water flow simulation using a fuzzy set approach." *Ground Water*, 35(2), 205-215.

Dou, C., Woldt, W., and Bogardi, I. (1999). "Fuzzy rule-based approach to describe solute transport in the unsaturated zone." *Journal of Hydrology*, 220(1-2), 74-85.

Dubois, D. and Prade, H. (1990). "Consonant approximations of belief functions." *International Journal of Approximate Reasoning*, 4, 419-449.

Dubois, D., and Prade, H. (1991). "Random sets and fuzzy interval analysis." *Fuzzy Sets and Systems*, 42, 87-101.

Fang, J. H., and Chen, H. C. (1997). "Fuzzy modeling and the prediction of porosity and permeability from the compositional and textural attributes of sandstone." *Journal of Petroleum Geology*, 20(2), 185-204.

Fruhwirth-Schnatter, S. (1993). "On fuzzy bayesian inference." *Fuzzy Sets and Systems*, 60(1), 41-58.

Ganoulis, J. (1996). "Modeling hydrologic phenomena." *Revue des Sciences de L'Eau*, 9(4), 421-434.

Goodman, I. R., and Nguyen, H. T. (1985). *Uncertainty Models for Knowledge-based Systems: A Unified Approach to the Measurement of Uncertainty*, Elsevier, Amsterdam.

Guan, J., and Aral, M. M. (2004). "Optimal design of groundwater remediation systems using fuzzy set theory." *Water Resources Research*, 40, W01518, doi: 10.1029/2003WR002121.

Helton, J. C., Davis, F. J., and Johnson, J. D. (2000a). "Characterization of stochastic uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant." *Reliability. Engineering and System Safety*, 69(1-3), 167-189.

Helton, J. C., Johnson, J. D., and Oberkampf, W. L. (2004). "An exploration of alternative approaches to the representation of uncertainty in model predictions." *Reliability Engineering and System Safety*, 85(1-3), 39-71.

Helton, J. C., Martell, M.-A., and Tierney, M. S. (2000b). "Characterization of subjective uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant." *Reliability Engineering and System Safety*, 69(1-3), 191-204.

Helton, J. C., and Oberkampf, W. L. (2004). "Alternative representations of epistemic uncertainty (Guest Editorial)." *Reliability Engineering and System Safety*, 85(1-3), 1-10.

Hofer, E., Kloos, M., Krzykacz-Hausmann, B., Peschke, J., and Woltereck, M. (2002). "An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties." *Reliability Engineering and System Safety*, 77(3), 229-238.

Joslyn, C., and Booker, J. M. (2004). "Generalized information theory for engineering modeling and simulation." *Engineering Design and Reliability Handbook*, Nikolaidis, E., Ghiocel, D., and Singhal, S., eds., CRC Press, New York.

Joslyn, C., and Ferson, S. (2004). "Approximate representations of random intervals for hybrid uncertainty quantification in engineering modeling." *Sensitivity Analysis of Model*

Output (SAM2004), Hanson, K. M., and Hemez, F. M., eds., Los Alamos National Laboratory.

Joslyn, C., and Kreinovich, V. (2005). "Convergence properties of an interval probabilistic approach to system reliability estimation." *International Journal of General Systems*, 34(4), 465-482.

Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River.

de Lima, V., and Olimpio, J.C. (1989). *Hydrogeology and Simulation of Ground-Water Flow at Superfund-Site Wells G and H, Woburn, Massachusetts*, U.S. Geological Survey: Water Resources Investigations Report 89-4059, Boston, Massachusetts.

Mathon, B, Ozbek, M., and Pinder, G.F. (2008). "Dempster-Shafer theory applied to uncertainty surrounding hydraulic conductivity." *Mathematical Geology*, in press.

Metropolis, N., and Ulam, S. (1949). "The monte carlo method." *Journal of the American Statistical Association*, 44(247), 335-341.

Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S. (2004). “Challenge problems: uncertainty in system response given uncertain parameters.” *Reliability Engineering and System Safety*, 85(1-3), 11-19.

O’Hagan, A., and Oakley, J. E. (2004). “Probability is perfect, but we can’t elicit it properly.” *Reliability Engineering and System Safety*, 85(1-3), 239-248.

Pan, Y., and Klir, G. J. (1996). “Bayesian inference of fuzzy probabilities.” *International Journal of General Systems*, 26(1-2), 73-90.

Prasad, R. K., and Mathur, S. (2007). “Groundwater flow and contaminant transport simulation with imprecise parameters.” *Journal of Irrigation and Drainage Engineering,* 133(1), 61-70.

Ricciardi, K., Pinder, G.F., and Karatzas, G.P. (1998). “A new probability density function for hydraulic conductivity in optimal design.” *Eos*, *Transactions*, AGU. 79(45), Suppl., Page F280.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). “Fuzzy kalman filtering of hydraulic conductivity.” *Computations Methods in Water Resources Conference 26^th^ International Conference (CMWRXVI)*.

Ross, J., Ozbek, M. M., and Pinder, G. F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*, 39(8), 765-780.

Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, John Wiley & Sons, West Sussex.

Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University, Princeton

Yang, C.C. (1997). "Fuzzy bayesian inference." *IEEE Transactions of Systems, Man, and Cybernetics*, 3, 2707-2712.

Zadeh, L. (1965). "Fuzzy sets." *Information. Control*, 8, 338-353.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1(1), 3-28.

CONCLUSIONS

(These conclusions are the individual conclusions copied from each of the Chapters 2 through 8)

Fuzzy Spatial Reasoning for Reservoir Characterization

The application of fuzzy inference to the definition of spatial relations, herein referred to as fuzzy spatial reasoning, is well suited for reservoir characterization. It has been demonstrated that fuzzy sets and fuzzy logic can capture expert knowledge regarding the spatial relationship between objects as well as the relationship between hydrologic parameters and soil formations.

Arguably, the value of expert data relies greatly upon the experts who opine on the desired topic. Experience shows that it is important that the proper questions are posed to the experts in order to obtain the desired information in an efficient manner. In this vein, a considerable amount of research has been directed toward data acquisition for fuzzy system definition (Onisawa and Anzai, 1999).

The greatest value of fuzzy spatial reasoning would be realized in the application to the comprehensive modeling of the effect of glacial, aeolian and hydrologic dynamics upon geology. Again, the knowledge capture process toward this end would likely be an arduous task. Nevertheless, it is possible that a series of simple rules can approximately capture the physics of complex depositional processes. Such an endeavor is well beyond the scope of this paper, though the groundwork was certainly laid above.

General and Site-Specific Means of Defining Fuzzy CPT-Based Soil Classification

In this paper, two new approaches to identifying soil classes based upon CPT sounding data are introduced. The site-specific approach utilizes kernel-based fuzzy clustering, whereas a more another soil classifier was attempted using fuzzy rules.

The improvements of the new approaches over traditionally employed charts for the classification of soil samples are meaningful and demonstrate promise in fuzzy-based classification tools. The most significant characteristic of the new methods, however, is the representation of soil classes with gradually changing boundaries. With such boundaries, one can represent a number of the soil types in the Robertson chart (i.e. *silty clay*, *sandy silt*) with just three fuzzy soil classes (clay, silt and sand). For example, a data point that has membership in both the clay and silty classes, greater in the former than the latter, can be classified as a *silty clay*.

While the performance of the kernel-based fuzzy classification was inferior to that of the fuzzy rule-based classification, it is appealing in its strict dependence upon the site data. At any site, a unique fuzzy classification chart can be constructed. Moreover, the kernel-based approach marks a first attempt to fuzzify soil classification charts in the vein of Begemenn (1963) and Douglas and Olsen (1981).

The fuzzy rule-based approach, however, is considered more promising due to its higher classification success rate over both the kernel-based approach and traditional classification charts. It is expected that, with the consideration of more data, the rules would frame a logic that provides a more universal appeal. Furthermore, the generality and transparency of the fuzzy rule-based approach make it an attractive classification method.

Hydraulic Conductivity Estimation via Fuzzy Analysis of Grain Size

Crisp estimates were obtained by defuzzification of the fuzzy estimates (Table 4.4). TThe table also gives the measured hydraulic conductivity values for the samples. All estimated values are within two orders of magnitude of the corresponding measured value, which is considered reasonable inasmuch as hydraulic conductivity values may vary by up to five orders of magnitude if there are small differences in the percentages of the fine fractions (Koltermann and Gorelick, 1995). Also relevant is the interpretability of the results. With the fuzzy reasoning process, one can understand how the estimates are calculated. For instance, it is evident that the second hydraulic conductivity estimate in Table 4.5 is lower-valued than the other three estimates because the aggregate estimated hydraulic conductivity fuzzy set in Figure 4.10B has high memberships for lower hydraulic conductivity values. Thus, when this fuzzy set is defuzzified, the lower values of hydraulic conductivity are weighted more heavily than in the other samples' estimated hydraulic conductivity fuzzy sets in Figure 4.10. The higher memberships for lower-valued hydraulic conductivity values in this fuzzy set directly result from the significant overlapping of the grain size distribution and borehole log description fuzzy sets with the Clay soil type fuzzy set during inference.

In many applications, borehole log data are not only more prevalent than grain size data, but often are the only types of data available. Considering this, an investigation was made into the reliability of estimates resulting from just the expert fuzzy characterization of borehole log descriptions of the soil samples. For completeness, the scenario where only grain size distribution data are available was also evaluated. The results of these investigations are summarized in Table 4.5. In general, for this set of data, it is difficult to distinguish which form of data is more accurate. However, estimates stemming from both forms of data are still within two orders of magnitude of the measurements. Both types of data are not required to produce acceptable estimates. It has been shown that the use of only one data type in the fuzzy reasoning process can still lead to relatively accurate results.

Ideally, for the generation of realizations in groundwater modeling, measures of uncertainty should accompany hydraulic conductivity estimates. However, the issue of uncertainty in the results is not addressed since this is a budding area of fuzzy logic, and definitive answers regarding the uncertainty of the fuzzy estimates are unavailable. A thorough description of uncertainty of fuzzy sets would be requisite and is beyond the scope of this research.

The soundness of this approach to hydraulic conductivity estimation relied greatly upon the definition of the original soil type and hydraulic conductivity fuzzy sets that were used for reasoning. More accurate estimates could be produced had more data from the CIBA-Geigy Superfund site been available. With more data, the soil type and hydraulic conductivity fuzzy sets could have been constructed to respect the site data, thus producing better hydraulic conductivity estimates for that particular site. This would be true for any site with sufficient soil data. However, given that experts in hydrogeology defined these fuzzy sets, the reasoning process is considered sound and generally applicable to any site with reasonable accuracy.

Fuzzy Kalman Filtering of Hydraulic Conductivity

By contrasting the corresponding contours in Figure 5.4 and Figure 5.5, two differences are evident. The first difference is intuitive. When crisp measurements are introduced to update the fuzzy estimates, the fuzziness of the estimates (how wide the α-cuts are) can only be reduced since the crisp measurements contain no fuzziness, making the fuzzy model more precise.

Secondly, in the contour plots of Figure 5.4, there are axes of symmetry, roughly illustrated by the dashed black lines in Figure 5.4A. This symmetry, however, is missing from the corresponding contour plot in Figure 5.5A. This is due to the influence the crisp measurements have upon the final estimates, presumably defining a posterior model that more accurately represents the true trends present at the site.

At measurement locations, the fuzzy estimates are made very precise and close in value to the corresponding crisp measurements. However, unlike the traditional Kalman filter algorithm, these posterior estimates at measurement locations are not made identical in value to the crisp measurements. This is a byproduct of the difference in data types between the measurements and prior estimates, and is not considered an inconsistency.

As Figure 5.6 shows, the posterior fuzzy estimates are quite similar to the corresponding crisp values.

Presented here was an alternative to using fuzzy information in conjunction with crisp data in a hydrogeological setting. By using Kalman filtering, each data type is used in a manner that respects its relative credibility. What results is a rather precise fuzzy field that presumably captures the trends inherent in the model. This fuzzy model may be defuzzified to create a crisp hydraulic conductivity field for input into traditional groundwater flow models, or may be left fuzzy to be used in fuzzy-friendly groundwater flow models (Dou et al, 1995; Dou et al, 1999). Further investigation is provided in the next chapter.

Kalman Filter Updating of Possibilistic Hydraulic Conductivity

The posterior estimated crisp K-field (Figure 6.11A) reveals some phenomenon that are absent from the estimated K-field that resulted from simply kriging the 17 hydraulic conductivity measurements (Figure 6.11B). The fuzzy Kalman filtering K-field shows that the area of low hydraulic conductivity in the northeast section of the spatial domain is larger than that produced by the K-field produced by kriging crisp measurements. Furthermore, the updated K-field reveals less continuity of higher hydraulic conductivity than is evident in the kriging-only K-field. An additional patch of high hydraulic conductivity is evident in the east of the updated figure. These results are confirmed by the soil samples evaluated in boreholes in those respective areas of the domain (Table 6.2). The summary statistics mathematically confirm the differences between the K-fields in Figure 6.11 (Table 6.3). Lower K values implied by the boreholes lead to a lower mean value for the updated K-field than for the traditionally determined K-field. Nevertheless, expert knowledge suggests a greater range of K values for the K-field in Figure 6.11A. Certainly, the differences between a traditionally determined K-field and a K-field augmented by expert knowledge should be greater where soil data and expert knowledge reveal hydrogeological formations not identified by traditional hydraulic conductivity measurements. Nevertheless, the consideration of additional data sources can certainly enhance site characterization. By incorporating the additional soil data in the form of quantified boring logs in this case, fuzzy Kalman filtering permitted both an appropriate approach to aggregate fuzzy and crisp data and a more accurate hydrogeological characterization of the CIBA-Geigy remediation site.

Previous approaches to incorporating fuzzy and crisp data in hydrogeological site characterization have suggested that fuzzy and crisp data be used in conjunction. Generally, fuzzy data results from expert knowledge. The nature of expert knowledge suggests that it should be used to produce a best first guess of the hydraulic conductivity field to be updated by more certain and precise data resulting from direct measurements. The fuzzy Kalman filtering algorithm presented herein takes this more intuitive approach to hydrogeological site characterization with expert knowledge. Nevertheless, it is not the first and only method to attempt to treat expert knowledge in this manner.

Algorithms have been developed to update kriged priors with additional data in a Bayesian context (Omre, 1987; Bandemer and Gebhardt, 2000). While a comparison between the proposed Kalman filtering approach and the Bayesian fuzzy kriging method herein may seem prudent, the two methods are quite dissimilar in their intentions. Whereas the fuzzy Kalman filtering algorithm uses expert knowledge of hydraulic conductivity directly in the form of fuzzy numbers, Bayesian fuzzy kriging is built for expert knowledge regarding the first and second moments of trend parameters. As such, further research is necessary to develop a Bayesian updating scheme that revises expert-provided fuzzy K values with crisp measurements.

Respecting Correlation While Modeling with Possibilistic Data

The use of expert knowledge in engineering applications is easily used when such information is characterized with fuzzy sets. This is evident in groundwater flow and transport modeling, where fuzzy valued model input parameters such as hydraulic conductivity have been defined by experts. Whereas fuzzy set- and fuzzy logic-based approaches to propagating the uncertainty characterized by fuzzy sets and possibility distributions are adequate in some applications, they are impractical for groundwater modeling because they are strongly rooted in fuzzy set theory and fuzzy logic. The fLHS algorithm offers an alternative approach to uncertainty modeling that borrows some techniques from stochastic modeling.

By taking advantage of the properties of a possibility distribution, the fLHS algorithm used a modified fuzzy numerical sampling scheme to sample from equi-probable focal elements. In cases where the number of realizations is high, the fLHS algorithm essentially amounts to sampling the convex hull of the possibilistic model parameters. Though this process sounds similar to the vertex method and interval analysis, these methods, like the extension principle, fail to respect the correlation of the spatially distributed model inputs. In order to respect this spatial correlation, steps were borrowed from LHS, whereby the sample realization matrix is operated upon and permuted. The result is M individual realizations that are consistent with the physics of the field.

It is clear that the error between the most possible values of the model parameters and the calculated means of the sampled realizations is minimal after a small number of realizations (Figure 7.4). The convergence of the model concentration estimates was confirmed by sampling *10000* realizations for the simplified Woburn example; the contour plot of the most possible nodal concentration values was identical to Figure 7.9.

Even though concentration estimates are easily reduced to means and variances, this is inconsistent with the manner in which input model uncertainty is characterized. As such, each node's concentrations estimates are interpreted in a possibilistic manner. Since concentration estimates defined by possibility distributions may not be ideal, the most possible value at each node is taken, in order to plot the resulting concentration field. In stochastic modeling, this is analogous to considering only the most probable value at each node, as determined from a histogram. The analogy to mean values in possibility distribution amounts to defuzzifying (Klir and Yuan, 1995) the possibility distributions.

Aleatoric and Epistemic Uncertainty in Groundwater Simulation

Because thorough hydrogeological investigations cost significant amounts of money and time, an efficient means of data acquisition is valuable. One such means is expert knowledge extraction for uncertainty characterization. Uncertainty has two forms, aleatory and epistemic, and expert insight has generally applied to the characterization of aleatory, or irreducible, uncertainty. However, inherent in human thought, is a certain degree of epistemic, or reducible, uncertainty. Thus, appropriate characterizations of expert knowledge should represent both uncertainty types.

Probability theory requires extraordinary precision of an expert and is not well suited for uncertainty characterization. The hazard associated with applying traditional probability theory is that the opining expert may provide inaccurate and artificially precise characterizations of the random variable that best captures the naturally stochastic nature of hydraulic conductivity.

Random sets were introduced in this paper as a more appropriate means to the characterization of both aleatory (the random variable) and epistemic (expert-characterized measurement error) uncertainty. This method of uncertainty characterization provides a means of bounding this true, yet unknown random variable, accurately capturing the imprecise nature of expert knowledge. It is important to reiterate, as well, that uncertainty characterization via random sets eliminates the need for any model definition or assumption.

Computationally, the application of the vertex method provides a direct approach to operating upon fuzzy-based transformations of random set data with standard modeling equations. Because the vertex method is based upon interval arithmetic, little knowledge of fuzzy set theory is required for its implementation. Though the extension principle provides greater accuracy, it is more computationally intensive than the vertex method, and the latter approach is a fair approximation to the extension principle given the perceived monotonicity of the concentration values with α-cuts of the input hydraulic conductivity values.

While the representation of model concentration estimates as upper and lower probabilities provides a transparent comparison between the random variable and random set approaches defined above, the utility of data in such a form may not be immediately obvious. Nevertheless, the representation of concentration estimates as possibility distributions, a form equivalent to the probability bounds, user-friendly. Inspection of a possibility distribution reveals not only the most possible concentration value, but also a range of concentrations that are also possible to lesser and varying degrees. In fact, algorithms have been developed to complement and update possibilistic model estimates with updated information (Fruhwirth-Schnatter, 1993; Pan and Klir, 1996; Yang, 1997; Ross et al, 2007).

COMPLETE BIBLIOGRAPHY

Aitken, M., Butler, G., Lemmon, D., Saindon, E., Peters, D., and Williams, G. (2004). "The Lord of the Rings: the visual effects that brought middle earth to the screen." *International Conference on Computer Graphics and Interactive Techniques*, *ACM SIGGRRAPH 2004 Course Notes*, 11.

Allen, D.M., Schuurman, N., and Zhang, Q. (2007). „Using fuzzy logic for modeling aquifer architecture." *Journal of Geographical Systems*, 9(3), 289-310.

Altman, D. (1994). "Fuzzy set theoretic approaches for handling imprecision in spatial analysis." *Journal of Geographical Information Systems*, 8(3), 271-289.

Alyamani, M.S., and Sen, Z. (1993). "Determination of hydraulic conductivity from complete grain-size distribution curves." *Ground Water*, 31(4), 551-555.

Amaral, T.G.B., Crisostomo, M.M. (2002). "Automatic helicopter motion control using fuzzy logic." *IEEE International Conference on Fuzzy Systems*, 2, 860-863.

Apostolakis, G. (1990). "The concept of probability in safety assessments of technological systems." *Science*, 250(4986), 1359-1364.

Arya, L.M., Leij, F.J., Shouse, P.J., and van Genuchten, M.Th. (1999). "Relationship between the hydraulic conductivity function and the particle-size distribution." *Soil Science Society of America Journal*, 63, 1063-1070.

Asli, M., and Marcotte, D. (1995). "Comparison of approaches to spatial estimation in a bivariate context." *Mathematical Geology*, 27(5), 641-658.

Bagtzoglou, A.C., Nedungadi, A., and Sagar, B. (1996). "A fuzzy rule-based model for flow simulation in heterogeneous media." *Proceedings of the 11st International Conference on Computational Methods in Water Resources*, 1, 629-637.

Bandemer, H., and Gebhardt, A. (2000). “Bayesian fuzzy kriging.” *Fuzzy Sets and Systems*, 112, 405-418.

Bardossy, A. (1996). “The use of fuzzy rules for the description of elements of the hydrological cycle.” *Ecological Modelling*, 85, 59-65.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1988). “Imprecise (fuzzy) information in geostatistics.” *Mathematical Geology*, 20(4), 287-311.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1989). “Geostatistics utilizing imprecise (fuzzy) information.” *Fuzzy Sets and Systems*, 31, 311-328.

Bardossy, A., Bogardi, I., and Duckstein, L. (1990a). “Fuzzy regression in hydrology.” *Water Resources Research*, 26(7), 1497-1508.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990b). “Kriging with imprecise (fuzzy) variograms. I: Theory.” *Mathematical Geology*, 22(1), 63-79.

Bardossy, A., Bogardi, I., and Kelly, W.E. (1990c). “Kriging with imprecise (fuzzy) variograms. II: Application.” *Mathematical Geology*, 22(1), 81-94.

Bardossy, A., Bronstert, A., and Merz, B. (1995). “1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach.” *Advances in Water Resources*, 18(4), 237-251.

Bardossy, A., and Disse, M. (1993). “Fuzzy rule-based models for infiltration.” *Water Resources Research*, 29(2), 373-383.

Bauer, M. (1996). “Approximation algorithms and decision making in the Dempster-Shafer theory of evidence: An empirical study.” *International Journal of Approximate Reasoning*, 17, 217-237.

Bear, J. (1979). *Hydraulics of Groundwater*, McGraw-Hill, New York.

Beasley, D., Bull, D.R., and Martin, R.R., (1993). "An overview of genetic algorithms: Part 1, fundamentals." *University Computing*, 15(2), 58-69.

Begemann, H.K.S. (1953). "Improved method of determining resistance to adhesion by sounding through a loose sleeve placed behind the cone." *Proceedings of the 3rd International Conference on Soil Mechanics and Foundation Engineering (ICSMFE)*, 213-217.

Begemann, H.K.S.P. (1963). "Use of static soil penetrometer in Holland." *New Zealand Engineering*, 18(2), 41-49.

Berardinis, L.A. (1991). "It washes! It rinses! It talks!" *Machine Design*, 63(18), 43-47.

Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.

Bezdek, J.C., and Pal, S.K. (1992). *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*, Institute of Electrical and Electronics Engineers, New York.

Boadu, F.K. (2000). "Hydraulic conductivity of soils from grain-size distribution: New models." *Journal of Geotechnical and Geoenvironmental Engineering*, 126(8), 739-746.

Bordogna, G., and Chiesa, S. (2003). "A fuzzy object-based data model for imperfect spatial information integrating exact objects and fields." I*nternational Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 11(1), 23-41.

Bragato, G. (2004). "Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain." *Geoderma*, 118, 1-16.

Brown, W., Groves, D., and Gedeon, T. (2003). "Use of fuzzy membership input layers to combine subjective geological knowledge and empirical data in a neural network method for mineral-potential mapping." *Natural Resources Research*, 12(3), 183-200.

Burrough, P.A., van Gaans, P.F.M., and Hootsmans, R. (1997). "Continuous classification in soil survey: spatial correlation, confusion and boundaries." *Geoderma*, 77, 115-135.

Caers, J., and Zhang, T (2002). "Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models." *Integration of Outcrop and Modern Analog Data in Reservoir Models*, Grammer, G. M., Harris, P.M., and Eberli, G. P., eds., American Association of Petroleum Geologists.

Carman, P.C. (1937). "Fluid flow through granular beds." *Transactions, Institution of Chemical Engineers*, 15, 150-166.

Carman, P.C. (1939). "Permeability of saturated sands, soils and clays." Journal of Agricultural Science, 29, 263-273.

Carranza, E.J.M., and Hale, M. (2001). "Geologically constrained fuzzy mapping of gold mineralization potential, Baguio District, Philippines." *Natural Resources Research*, 10(2), 125-136.

Castro, P.A.D., and Camargo, H.A. (2005). "Focusing on interpretability and accuracy of a genetic fuzzy system." *14th International Conference on Fuzzy Systems*, 696-701.

Chanas, S., and Nowakowski, M. (1988). "Single value simulations of fuzzy variable." *Fuzzy Sets and Systems*, 25, 43-57.

Chapellier, D. (1992). *Well logging in hydrogeology*, A.A. Balkema Publishers, Brookfield, MA.

Chapuis, R.P., and Aubertin, M. (2003). "On the used of the Kozeny-Carman equation to predict the hydraulic conductivity of soils." *Canadian Geotechnical Journal*, 40, 616-628.

Chen, G., Xie, Q., and Shieh, L.S. (1998). “Fuzzy Kalman filtering.” *Information Sciences*, 109(1-4), 197-209.

Cheng, Q., and Agterberg, F.P. (1999). “Fuzzy weights of evidence method and its application in mineral potential mapping.” *Natural Resources Research*, 8(1), 27-35.

Chiang, H. H., Wu, S.J., and Lee, T.T. (2003). “Application of fuzzy sets theory in process data filtering.” *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems*, 325-330.

Choquet, G. (1954). “Theory of capacities.” *Annales de L’Institut Fourier*, 5, 131-295.

Cooper, N. (1978). “The law of excluded middle.” *Mind*, 87(2), 161-180.

Coppola, E.A., Duckstein, L., and Davis, D. (2002). “Fuzzy rule-based methodology for estimating monthly groundwater recharge in a temperate watershed.” Journal of Hydrologic Engineering, 7(4), 326-335.

Cordon, O., Gomide, F., Herrera, F., Hoffmann, F., and Magdalena, L. (2004). “Ten years of genetic fuzzy systems: Current framework and new trends.” *Fuzzy Sets and Systems*, 141, 5-31.

Cronican, A.E., and Gribb, M.M. (2004). “Hydraulic conductivity prediction for sandy soils.” *Ground Water*, 42(3), 459-464.

de Marsily, G., Delay, G., Teles, V., Schafmeister, M.T. (1998). “Some current methods to represent the heterogeneity of natural media in hydrogeology.” *Hydrogeology Journal*, 6, 115-130.

De Gruijter, J.J., and McBratney, A.B. (1988). “A modified fuzzy k means for predictive classification.” *Classification and Related Methods of Data Analysis*, Bock, H.H., ed., Elsevier, Amsterdam.

De Gruijter, J.J., Walvoort, D.J.J, and van Gaans, P.F.M. (1997). “Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models.” *Geoderma*, 77, 169-195.

Delgado, M., and Moral, S. (1987). “On the concept of possibility-probability consistency, *Fuzzy Sets and Systems*, 21, 311-318.

Demmico, R.V. (2004). “Fuzzy logic and earth science: An overview.” *Fuzzy Logic in Geology*, Demmico, R.V., and Klir, G.J., eds., Elsevier, Amsterdam.

Demmico, R.V., and Klir, G. (2004). *Fuzzy Logic in Geology*, Elsevier, Amsterdam.

Dempster, A.P. (1968). “A generalization of Bayesian inference.” *Journal of the Royal Statistical Society,Series B,* 30, 205-247.

de Silva, C.J.S., and Attikiouzel, Y. (1994). “A better path to duplicating human reasoning.” IEEE Expert 9(4), 14.

Dhiman, S.D., and Keshari, A.K. (2003). “Quantifying uncertainties using fuzzy logic for groundwater driven contaminant exposure assessment.” Probabilistic Approaches to Groundwater Modeling Symposium, 236-247.

Diamond, P. (1989). “Fuzzy kriging.” *Fuzzy Sets and Systems*, 33, 315-332.

Dokou, Z. (2008). *Optimal Search Strategy for the Definition of a DNAPL Source*, Ph.D. Dissertation, University of Vermont.

Domenico, P.A., and Schwartz, F.W. (1990). *Physical and Chemical Hydrogeology*, John Wiley and Sons, New York.

Dong, W., and Shah, H.C. (1987). “Vertex method for computing functions of fuzzy variables.” *Fuzzy Sets and Systems*, 24, 65-78.

Dou, C., Woldt, W., and Bogardi, I. (1999). "Fuzzy rule-based approach to describe solute transport in the unsaturated zone." *Journal of Hydrology*, 220, 74-85.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1995). "Steady state groundwater flow simulation with imprecise parameters." *Water Resources Research*, 31(11), 2709-2719.

Dou, C., Woldt, W., Bogardi, I., Dahab, M. (1997a). "Numerical solute transport simulation using fuzzy sets approach." *Journal of Contaminant Hydrology*, 27, 107-126.

Dou, C., Woldt, W., Dahab, M., and Bogardi, I. (1997b). "Transient ground-water flow simulation using a fuzzy set approach." *Ground Water*, 35(2), 205-215.

Douglas, B.J., and Olsen, R.S. (1981). "Soil classification using electric cone penetrometer." *Proceedings of a Session at the ASCE National Convention: Cone Penetration Testing and Experience. ASCE*, 209-227.

Druschel, B., Ozbek, M.M., and Pinder, G.F. (2006). "An application of Dempster-Shafer theory to hydraulic conductivity." *Proceedings of the 16th International Conference on Computational Methods in Water Resources*, Copenhagen, Denmark.

Dubois, D., Foulloy, L., Mauris, G., and Prade, H. (2004). "Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities." *Reliable Computation*, 10, 273-297.

Dubois, D., and Prade, H. (1988). *Possibility theory: An approach to computerized processing of uncertainty*, Plenum Press, New York.

Dubois, D., and Prade, H. (1990). "Consonant approximations of belief functions." *International Journal of Approximate Reas*oning, 4, 419-449.

Dubois, D., and Prade, H. (1991a). "Basic issues on fuzzy rules and their application to fuzzy control, *Proceedings IJCAI-91 Workshop on Fuzzy Control*, 5-17.

Dubois, D., and Prade, H. (1991b). "Random sets and fuzzy interval analysis." *Fuzzy Sets and Systems*, 42, 87-101.

Dubois, D., Prade, H., and Sandri, S. (1991). "On possibility/probability transformations." *Proceedings of the Fourth International Fuzzy Systems Association World Congress* (*IFSA '91*), 50-53.

El-Kadi, A. (1985). "On estimating the hydraulic properties of soil, Part 2. A new empirical equation for estimating hydraulic conductivity for sands." *Advances in Water Resources*, 8, 148-153.

Elkan, C. (1994). "The paradoxical success of fuzzy logic." IEEE Expert, 9(4), 3-8.

Esterline, A.C., Dozier, G., and Homaifar, A. (1997). "Fuzzy spatial reasoning." *NASA URC Technical Conference*.

Fang, J.H., and Chen, H.C. (1997). "Fuzzy modeling and the prediction of porosity and permeability from the compositional and textural attributes of sandstone." *Journal of Petroleum Geology*, 20(2), 185-204.

Faybishenko, B.A. (2004). "Introduction to modeling of hydrogeologic systems using fuzzy differential equation." *Fuzziness and Soft Computing*, 142, 267-284.

Ferson, S., and Ginzburg, L.R. (1996). "Different methods are needed to propagate ignorance and variability." Reliability Engineering and System Safety, 54, 133-144.

Finol, J., and Jing, X.-D.D. (2002). "Permeability prediction in shaly formations: The fuzzy modeling approach." *Geophysics*, 67(3), 817-829.

Fogg, G.E., Noyes, C.D., and Carle, S.F. (1998). "Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting." *Hydrogeology Journal*, 6(1), 131-143.

Freeze, R.A. (1975). "A stochastic-conceptual analysis of one dimensional groundwater flow in nonuniform homogeneous media, *Water Resources Research*, 11(5), 725-741.

Fruhwirth-Schnatter, S. (1993). "On fuzzy bayesian inference." *Fuzzy Sets and Systems*, 60(1), 41-58.

Ganoulis, J. (1996). "Modeling hydrologic phenomena." *Revue des Sciences de L'Eau*, 9(4), 421-434.

Garcia, O.N. (1994). "Known concerns about fuzzy logic." IEEE Expert, 9(4), 22-24.

Gelhar, L.W. (1986). "Stochastic subsurface hydrology from theory to applications." *Water Resources Research*, 22(9), 135S-145S.

Gelhar, L.W. (1993). *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, NJ.

Goodman, I. R., and Nguyen, H. T. (1985). *Uncertainty Models for Knowledge-based Systems: A Unified Approach to the Measurement of Uncertainty*, Elsevier, Amsterdam.

Gotway, C.A., and Hartford, A.H. (1996). "Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables." *Journal of Agricultural, Biological, and Environmental Statistics*, 1(1), 17-39.

de Gruijter, J.J., Walvoort, D.J.J., and van Gaans, P.F.M. (1997). "Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models." *Geoderma*, 77(2), 169-195.

Guan, J., and Aral, M. M. (2004). "Optimal design of groundwater remediation systems using fuzzy set theory." *Water Resources Research*, 40, W01518, doi: 10.1029/2003WR002121.

Helton, J.C., and Burmaster, D.E. (1996). "Guest Editorial: Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems." *Reliability Engineering and System Safety*, 54(2-3), 91-94.

Helton, J.C., Davis, F.J., and Johnson, J.D. (2000a). "Characterization of stochastic uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant." *Reliability Engineering and System Safety*, 69, 167-189.

Helton, J. C., Johnson, J. D., and Oberkampf, W. L. (2004). "An exploration of alternative approaches to the representation of uncertainty in model predictions." *Reliability Engineering and System Safety*, 85(1-3), 39-71.

Helton, J.C., Martell, M.-A., and Tierney, M.S. (2000b). "Characterization of subjective uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant." *Reliability Engineering and System Safety*, 69, 191-204.

Helton, J.C., and Oberkampf, W.L. (2004). "Alternative representations of epistemic uncertainty." *Reliability Engineering and System Safety*, 85, 1-10.

Herrera, F., Lozano, M., and Verdegay, J.L. (1993). "Genetic algorithms applications to fuzzy logic based systems." *Proceedings of the 9th Polish-Italian and 5th Polish-Finnish Symposium on Systems Analysis and Decision Support in Economics and Technology*, 125-134.

Herrera, F., Lozano, M., and Verdegay, J.L. (1995). "Tuning fuzzy logic controllers by genetic algorithms." *International Journal of Approximate Reasoning*, 12, 299-315.

Ho, S.Y., Cheng, T.K., and Ho, S.J. (2000). "Designing an efficient fuzzy classifier using an intelligent genetic algorithm." *Proceedings of the IEEE Computer Society's International Computer Software and Applications Conference*, 293-298.

Hofer, E., Kloos, M., Krzykacz-Hausmann, B., Peschke, J., and Woltereck, M. (2002). "An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties." *Reliability Engineering and System Safety*, 77(3), 229-238.

Højberg, A.L., and Refsgaard, J.C. (2005). "Model uncertainty – parameter uncertainty versus conceptual models." *Water Science Technology*, 52(6), 177-186.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Michigan.

Hong, T.P., Chen, C.H., Wu, Y.L. and Lee, Y.C. (2006). "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions." *Soft Computing*, 10, 1091-1101.

Hsieh, B., Lewis, C., and Lin, Z. (2005). "Lithology identification of aquifers from geophysical well logs and fuzzy logic analysis: Shui-Lin Area, Taiwan." *Computers and Geosciences*, 31, 263-275.

Hsiao, C.-Y. (1997). "Analysis and design of a fuzzy Kalman filtering algorithm." *Journal of Control Systems and Technology*, 5(2), 75-86.

Hsiao, C.-Y. (1999). "Modified gain fuzzy Kalman filtering algorithm." *IEEE Transactions on Systems, Man and Cybernetics, Series C*, 42(2), 363-368.

Hurley, P. (2006). *A Concise Introduction to Logic*, Thomson Wadsworth Publishing, California.

Hwang, S.I., Powers, S.E. (2003). "Using particle-size distribution models to estimate soil hydraulic properties." *Soil Science Society of America Journal*, 67, 1103-1112.

Issaks E.H., and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Joslyn, C., and J. M. Booker (2004), Generalized information theory for engineering modeling and simulation, in *Engineering Design and Reliability Handbook*, edited by E. Nikolaidis, D. Ghiocel, and S. Singhal, pp. 9-1 – 9-40, CRC Press, New York.

Joslyn, C., and S. Ferson (2004), Approximate representations of random intervals for hybrid uncertainty quantification in engineering modeling, in *Sensitivity Analysis of Model Output (SAM2004)*, edited by K. M. Hanson, and F. M. Hemez, pp. 453-469, Los Alamos National Laboratory.

Joslyn, C., and V. Kreinovich (2005). "Convergence properties of an interval probabilistic approach to system reliability estimation." *International. Journal of General Systems*, 34(4), 465-482.

Journel, A.G. (1983). "Nonparametric-estimation of spatial distributions." *Journal of the International Association for Mathematical Geology*, 15(3), 445-468.

Jumarie, G. (1995). "Further results on possibility-probability conversion via relative information and informational invariance." *Cybernetics and Systems*, 26(1), 111-128.

Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems." *Transactions of the ASME - Journal of Basic Engineering* (*Series D*), 82, 35-45.

Keller, J.M., and Wang, X. (2000). "A fuzzy rule-based approach to scene description involving spatial relationships." *Computer Vision and Image Understanding*, 80, 21-41.

Khanbagi, M., and Malhame, R.P. (1994). "Reducing travel energy costs for a subway train via fuzzy logic controls." *Proceedings of the 1994 IEEE International Symposium on Intelligent Control*, 99-104.

Klir, G.J. (1989). "Is there more to uncertainty than some probability theorists might have us believe?" *International Journal of General Systems*, 15, 347-378.

Klir, G.J. (2006). *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley & Sons, Hoboken, NJ.

Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River.

Koltermann, C.E., and Gorelick, S.M. (1995). “Fractional packing model for hydraulic conductivity derived from sediment mixtures.” *Water Resources Research*, 31(12), 3283-3297.

Kosko, B. (1993). *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York.

Kravchenko, A.N. (2003). “Influence of spatial structure on accuracy of interpolation methods.” *Soil Science Society of America Journal*, 67, 1564-1571.

Krumbein, W.C., and Monk, G.D. (1943). “Permeability as a function of the size parameters of sedimentary particles.” *American Institute of Mining and Metallurgical Engineers Technical Publication*, 1492, 153-163.

Kulkarni, R.B. (1984). “Bayesian kriging in geotechnical problems.” *Geostatistics for Natural Resources Characterizations*, Verly, G., David, M., Journel, A.G., and Marechal, A., eds., Reidel, Dordrecht, The Netherlands.

Kurup, P.U., and Griffin, E.P. (2006). “Prediction of soil composition from CPT data using general regression neural network.” *Journal of Computing in Civil Engineering*, 20(4), 281-289.

Larue, J., and Tyagi, A.K. (1997). “Groundwater simulation using fuzzy numbers.” *Groundwater: An Endangered Resource*, Findikakis, A.N., and Stauffer, F., eds., ASCE, New York.

Larue, J., and Tyagi, A.K. (1998). “Fate/transport modeling using fuzzy numbers.” *After the Rain Has Fallen: Ground Water Management Symposium*, Smith, J.W., ed., ASCE, Reston, VA.

Leung, Y. (1983). “Fuzzy-sets approach to spatial-analysis and planning – A nontechnical evaluation.” *Geografiska Annaler Series B – Human Geography*, 65(2), 65-75.

Leung, Y. (1984). “Basic issues of fuzzy sets theoretic spatial analysis.” *24th European Congress of the Regional Science Association*, Milan, Italy.

Li, Y., and Li, S. (2004). “A fuzzy set theoretic approach to approximate spatial reasoning.” *IEEE Transactions on Fuzzy Systems*, 12(6), 745-754.

de Lima, V., and Olimpio, J.C. (1989). *Hydrogeology and Simulation of Ground-Water Flow at Superfund-Site Wells G and H, Woburn, Massachusetts*, U.S. Geological Survey: Water Resources Investigations Report 89-4059, Boston, Massachusetts.

Malinowski, G. (2001). “Many-Valued Logics.” *The Blackwell Guide to Philosophical Logic*, Goble, L., ed.,. Blackwell.

Mamdani, E.H., and Assilian, S. (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." *International Journal of Man-Machine Studies*, 7(1), 1-13.

de Marsily, G., Delay, F., Teles, V., and Schafmeister, M.T. (1998). “Some current methods to represent the heterogeneity of natural media in hydrogeology.” *Hydrogeology Journal*, 6, 115-130.

Masch, F.D., and Denny, K.J. (1966). “Grain size distribution and its effect on the permeability of unconsolidated sands.” *Water Resources Research*, 2(4), 665-677.
Masson, M., and Denœux, T. (2006). “Inferring a possibility distribution from empirical data.” *Fuzzy Sets and Systems*, 157, 319-340.

Matheron, G. (1971). “The theory of regionalized variables and its applications.” *Les Cahiers du Centre de Morphologie Mathematique de Fontainbleau*, 5, Ecole Nationale Supérieure des Mines de Paris.

Mathon, B., Ozbek, M.M., and Pinder, G.F. (2008) “Dempster-Shafer theory applied to uncertainty surrounding hydraulic conductivity.” *Mathematical Geology*, *in press*.

Matia, F., Jimenez, A., Al-Hadithi, B.M., Rodriguez-Losada, D., and Galan, R. (2006). “The fuzzy Kalman filter: State estimation using possibilistic techniques.” *Fuzzy Sets and Systems*, 157, 2145-2170.

Mazaheri, S.A., Koppi, A.J., and McBratney, A.B. (1995). “A fuzzy allocation scheme for the Australian Great Soil Groups classification system.” *European Journal of Soil Science*, 46, 601-612.

Mazaheri, S.A., McBratney, A.B., and Koppi, A.J. (1997). “Sensitivity of memberships to attribute variation around selected centroiods and intergrades in the continuous Australian Great Soil Groups classification system.” *Geoderma*, 77, 155-168.

McBratney, A.B., and de Gruijter, J.J. (1992). “A continuum approach to soil classification by modified fuzzy *k*-means with extragrades, *Journal of Soil Science*, 43, 159-175.

McBratney, A.B., De Gruijter, J.J, Brus, D.J. (1992). “Spacial prediction and mapping of continuous soil classes.” *Geoderma*, 54, 39-64.

McBratney, A.B., and Odeh, I.O.A. (1997). “Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions.” *Geoderma*, 77(2), 85-113.

McKay, M.D., Conover, W.J., and Breckman, R.J. (1979). “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code.” *Technometrics*, 21, 239-245.

Metropolis, N., and Ulam, S. (1949). "The Monte Carlo method." *Journal of the American Statistical Association*, 44(247), 335–341.

Minasny, B., Homans, J.W., Harter, T., Eching, S.O., Tuli, A., and Denton, M.A. (2004). “Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data.” *Soil Science Society of America Journal*, 68, 417-429.

Mishra, S., Parker, J.C., and Singhal, N. (1989). "Estimation of soil hydraulic properties and their uncertainty from particle size distribution data." *Journal of Hydrology,* 108, 1-18.

Moore, R. (1966). *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Mullarky, P.W., and Fenves, S.J. (1986). "Fuzzy logic in a geotechnical knowledge based system: CONE." *Civil Engineering Systems*, 3(2), 58-81.

Nedungadi, A., Bagtzoglou, A.C., and Sagar, B. (1994). "A fuzzy methodology to solve the groundwater flow problem in the presence of uncertain hydrogeological data." Applications of Artificial Intelligence in Engineering IX, Adey, R.A., Rzevski, G., Russell, D.W., eds., Computational Mechanics Publications, Southampton, UK.

Nilsson, B., Højberg, A.L., Refsgaard, J.C., and Troldborg, L. (2007). "Uncertainty in geological and hydrogeological data." *Hydrology and Earth System Sciences*, 11, 1551-1561.

Nobre, M.M., and Sykes, J.F. (1992). "Application of Bayesian kriging to subsurface characterization." *Canadian Geotechnical Journal*, 29, 589-598.

Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S. (2004). "Challenge problems: uncertainty in system response given uncertain parameters." *Reliability Engineering and System Safety*, 85(1-3), 11-19.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J. (1992). "Fuzzy-*c*-means and kriging for mapping soil as a continuous system." *Soil Science Society of America Journal*, 56, 1848-1852.

Odeh, I.O.A, McBratney, A.B., and Slater, B.K. (1999). "Predicting soil properties from ancillary information: Non-spatial models compared with geostatistical and combined methods." *Geostatistics Wollongong*, 2, 1008-1019.

Odeh, I.O.A., Todd, A.J., and Triantafilis, J. (2003). "Spatial prediction of particle size fractions as compositional data." *Soil Science*, 168(7), 501-515.

O'Hagan, A., and Oakley, J.E. (2004). "Probability is perfect, but we can't elicit it perfectly." Reliability Engineering and Risk Assessment, 85, 239-248.

Omre, H. (1987). "Bayesian kriging—Merging observations and qualified guesses in kriging." Mathematical Geology, 19(1), 25-39.

Onisawa, T., and Anzai, T. (1999) "Acquisition of intelligible fuzzy rules." *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 5, 268-273.

Oussalah, M. (2000). "On the probability/possibility transformations: A comparative analysis." *International Journal of General Systems*, 29(5), 671-718.

Oussalah, M., and Schutter, J.D. (2000). "Possibilistic kalman filtering for radar 2-D tracking." *Information Sciences*, 130, 85-107.

Ozbek, M.M., and Pinder, G.F. (2006). "Non-probabilistic uncertainty in subsurface hydrology and its applications: An overview." *Water, Air, and Soil Pollution: Focus*, 6, 35-46.

Pan, Y., and Klir, G. J. (1996). "Bayesian inference of fuzzy probabilities." *International Journal of General Systems*, 26(1-2), 73-90.

Parasuraman, K., Elshorbagy, A., and Si, B.C. (2006). "Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles." *Soil Science Society of America Journal*, 70, 1851-1859.

Passarella, G., Vurro, M., D'Agostino, V., and Barcelona, M.J. (2003). "Cokriging optimization of monitoring network configuration based on fuzzy and non-fuzzy variogram evaluation." *Environmental Monitoring and Assessment*, 82, 1-21.

Pelletier, F.J. (1994). "Fuzzy logic: A misplaced appeal." IEEE Expert, 9(4), 29-31.

Pham, T.D. (2005). “Integration of fuzzy and geostatistical models for estimating

Piotrowski, J.A. (1996). “Estimation of hydrogeological parameters for groundwater modeling with fuzzy geostatistics: Closer to nature?” *Calibration and Reliability in Groundwater Modelling*, 237, 511-520.

Piotrowski, J.A., Bartels, F., Salski, A., and Schmidt, G. (1996). “Geostatistical regionalization of glacial aquitard thickness in northwestern Germany, based on fuzzy kriging.” *Mathematical Geology*, 28(4), 437-452.

Porwal, A., Carranza, E.J.M., and Hale, M. (2003). “Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping.” *Natural Resources Research*, 12(1), 1-25.

Prasad, R. K., and Mathur, S. (2007). “Groundwater flow and contaminant transport simulation with imprecise parameters.” *Journal of Irrigation and Drainage Engineering,* 133(1), 61-70.

Reyes, C.A.P. (2004). *Coevolutionary fuzzy modeling*, Springer, Berlin.

Ricciardi, K., Pinder, G.F., and Karatzas, G.P. (1998). “A new probability density function for hydraulic conductivity in optimal design.” *Eos*, *Transactions*, AGU. 79(45), Suppl., Page F280.

Robertson, P.K. (1990). “Soil classification using the cone penetration test.” *Canadian Geotechnical Journal*, 27(1), 151-158.

Robertson, P.K., Campanella, R.G., Gillespie, D., and Grieg, J. (1986). “Use of piezometer cone data.” *Proceedings*, *In-situ ’86*, ASCE Specialty Conference, Blacksburg, Virginia.

Robinson, V.B. (1990). “Interactive machine acquisition of a fuzzy spatial relation.” *Computers and Geosciences*, 16(6), 857-872.

Romo, M.P., and Garcia, S.R. (2003). "Neurofuzzy mapping of CPT values into soil dynamic properties." *Soil Dynamics and Earthquake Engineering*, 23, 473-482.

Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, John Wiley & Sons, West Sussex.

Ross, J., Ozbek, M., and Pinder, G.F. (2006). "Fuzzy kalman filtering of hydraulic conductivity." *Computations Methods in Water Resources Conference 26th International Conference (CMWRXVI)*.

Ross, J., Ozbek, M. and Pinder, G.F. (2007). "Hydraulic conductivity estimation via fuzzy analysis of grain size data." *Mathematical Geology*. 39(8), 765-780.

Schaap, M.G., and Leij, F.J. (1998). "Database-related accuracy and uncertainty of pedotransfer functions." *Soil Science*, 163(10), 765-779.

Schaap, M.G., Leij, F.J., van Genuchten, M.Th. (2001). "ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions." *Journal of Hydrology*, 251, 163-176.

Schmertmann, J.H. (1978). "Guidelines for cone test, performance, and design." *Federal Highway Administration*, *Report FHWA-TS-78209*.

Schockaert, S., De Cock, M., and Kerre, E.E. (2006). "Towards fuzzy spatial reasoning in geographic IR systems." *Proceedings of GIR06 (Workshop on Geographic Information Retrieval), Workshop at SIGIR2006 (29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval),* 34-36.

Schultz, K., and Huwe, B. (1997). "Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach." Journal of Hydrology, 201, 211-229.

Selley, R.C. (1985). *Elements of Petroleum Geology*, W. H. Freeman and Company, New York.

Sen, Z. (1992). "New statistical grain-size method for evaluating the hydraulic conductivity of sand aquifers – Comment." *Journal of Hydrology*, 130, 399-403.

Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University, Princeton.

Shepherd, R.G. (1989). "Correlations of permeability and grain size." *Ground Water*, 27(5), 633-638.

Shokir, E.M.E.-M. (2003). "A novel model identification for the prediction of heterogeneous reservoir permeability." *Journal of Engineering and Applied Science*, 50(6), 1197-1214.

So, A.T.P., Li, K.C., Chan, W.L. (1993). "Implementation of fuzzy logic based automatic camera focuser." *Proceedings of the 1993 IEEE Region 10 Conference on Computer, Communication, Control and Power Engineering*, 4, 292-295.

Sudicky, E.A. (1986). "Natural gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process." *Water Resources Research*, 22(13), 2069-2082.

Strebelle, S. (2000). *Sequential simulation drawing structure from training images*, Ph.D. Thesis, Stanford University.

Sunila, R., and Horttanainen, P. (2003). "Fuzzy model of soil polygons for managing the imprecision." *Proceeding of Interfacing Geostatistics, GIS and Spatial Databases* (*STATGIS2003*).

Tamari, S., Wösten, J.H.M, and Ruiz-Suárez, J.C. (1996). "Testing an artificial neural network for predicting soil hydraulic conductivity." *Soil Science Society of America Journal*, 60, 1732-1741.

Taylor, R.B., and Steven, T.A. (1983). "Definition of mineral resource potential." *Economic Geology*, 78, 1268-1270.

Temples, T.J., and Waddell, M.G. (1996). "Application of petroleum geophysical well logging and sampling techniques for evaluating aquifer characteristics." *Ground Water*, 34(3), 523-531.

Thore, P. (2006). "Accuracy and limitations in seismic modeling of reservoir." *SEG Technical Program Expanded Abstracts*, 1674-1677.

Titi, H.H., and Tumay, M.T. (2000). "Cone penetration tests at the NGES-Texas A&M University: Clay site." *ASCE Geotechnical Special Publication*, 93, 186-205.

Triantifilis, J., Odeh, I.O.A., Minasny, B., and McBratney, A.B. (2003). "Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy k-means classification of EM34 data." *Environmental Modelling and Software*, 18(7), 667-680.

Uddameri, V. (2004). "Relationships of longitudinal dispersivity and scale developed from fuzzy least-squares regressions." *Environmental Geology*, 45(8), 1172-1178.

Uma, K.O., Egboka, B.C.E, and Onuoha, K.M. (1989). "New statistical grain-size method for evaluating the hydraulic conductivity of sandy aquifers." *Journal of Hydrology,* 108, 343-366.

van Genuchten, M. Th. (1980). "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils." *Soil Science Society of America Journal*, 44, 892-898.

Vernieuwe, H., De Baets, B., Verhoest, N.E.C. (2002a). "Simulation of one-dimensional water movement in the unsaturated zone by means of a first order Takagi-Sugeno model." IEEE International Conference on Fuzzy Systems, 2, 1069-1074.

Vernieuwe, H., Verhoest, N.E.C., and De Baets, B. (2002b). "A Takagi-Sugeno model of flow in the unsaturated zone." *Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality (Proceedings of ModelCARE 2002)*, 277, 79-85.

Vernieuwe, H., Verhoest, N.E.C., de Baets, B., Hoeben, R., and de Troch, F.P. (2007). "Cluster-based fuzzy models for groundwater flow in the unsaturated zone." *Advances in Water Resources*, 30, 701-714.

Vos, J.D. (1982). "The practical use of CPT in soil profiling." *Proceedings of the 2nd European Symposium on Penetration Testing*, 933-939.

Wang, X., and Keller, J.M. (1997). "Fuzzy surroundedness." *IEEE International Conference on Fuzzy Systems*, 2, 1173-1178.

Wang, X., and Keller, J.M. (1999). "Human-based spatial relationship generalization through neural/fuzzy approaches." *Fuzzy Sets and Systems*, 101(1), 5-20.

Wang, X.J., and Qi, F. (1998). "The effects of sampling design on spatial structure analysis of contaminated soils." *The Science of the Total Environment*, 224, 29-41.

Wang, Z., and Klir, G.J. (1992). Fuzzy Measure Theory, Springer, New York.

Webster, R., and Oliver, M.A. (1992). "Sample adequately to estimate variograms of soil properties." *Journal of Soil Science*, 43, 177-192.

Welch, G., Bishop, G. (2001). "An introduction to the kalman filter." *SIGGRAPH 2001 course 8*, *In Computer Graphics*, *Annual Conference on Computer Graphics & Interactive Techniques*. ACM Press, Addison-Wesley, Los Angeles, CA, USA.

Winter, C.L. (2004). "Stochastic hydrology: practical alternatives exist." *Stochastic Environmental Research and Risk Assessment*, 18(4), 271-273.

Yager, R. (2004). "Cumulative distribution functions from Dempster-Shafer belief structures." *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 34(5), 2080-2087.

Yang, C.C. (1997). "Fuzzy bayesian inference." *IEEE Transactions of Systems, Man, and Cybernetics*, 3, 2707-2712.

Zadeh, L. (1965). "Fuzzy sets", *Information Control*, 8, 338-353.

Zadeh, L. (1973). "Outline of a new approach to the analysis of complex systems and decision processes." *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(1), 28-44.

Zadeh, L. (1975). "The concept of a linguistic variable and its application to approximate reasoning – I." *Information Sciences*, 8, 199-249.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 1(1), 3-28.

Zadeh, L. (1992). "The calculus of fuzzy if-then rules." *Artificial Intelligence Expert*, 7(3), 23-27.

Zhang, D.Q. and Chen, S.C. (2003). "Clustering incomplete data using kernel-based fuzzy C-means algorithm." *Neural Processing Letters*, 18(3), 155-162.

Zhang, Y., and Pinder, G. (2003). "Latin hypercube sample selection strategy for correlated random hydraulic conductivity fields." *Water Resources Research*, 39(8), doi:10.1029/2002WR001822.

Zhang, Z., and Tumay, M. (1999). "Statistical to fuzzy approach toward CPT soil classification." *Journal of Geotechnical and Geoenvironmental Engineering*, 125(3), 179-186.

Zhang, Z., and Tumay, M.T. (2003). "Non-traditional approaches in soil classification derived from the cone penetration test." *ASCE Geotechnical Special Publication*, 121, 101-149.

Zhang, N., and Wunsch, D.C. (2003). "An extended Kalman filter (EFK) approach on fuzzy system optimization problem." *The 12th IEEE Internation Conference on Fuzzy Systems, FUZZ '03*, 2, 1465-1470.

Zhu, A.-X. (1997). “A similarity model for representing soil spatial information.” *Geoderma*, 77, 217-242.

Zhu, A.-X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. (2001). “Soil mapping using GIS, expert knowledge, and fuzzy logic.” *Soil Science Society of America Journal*, 65, 1463-1472.

Zhu, A.-X., Band, L.E., Dutton, B., and Nimlos, T.J. (1996). “Automated soil inference under fuzzy logic.” *Ecological Modelling*, 90, 123-145.
missing multivariate observations.” *WSEAS Transactions on Systems*, 4(4), 233-237.

Zong, Q., Wang, Z., Yue, Y., and Lei, X. (2000). “The design of multi-objective fuzzy control algorithm in elevator group control system.” *Proceedings of the World Congress on Intelligent Control and Automation (WCICA),* 3, 2009-2012.