Application of fuzzy sets to estimate uncertainty of hydraulic conductivity

Bayat!, H., M.R. Neyshabouri, A. A. Mahboubi, M.A. Hajabbasi, and M.R. Mosaddeghi
1-Department of Soil Science, College of Agriculture, Tabriz University, Tabriz, Iran. Email:
h.bayat@tabrizu.ac.ir

Introduction

The output of mathematical models is often approximate and lying within an imprecision range. This imprecision may result from measuring physical magnitudes, parameter estimation procedures, temporal and spatial variability of the parameters, subjective interpretations and expert judgment of available information (Freissinet et al., 1998). One way of analyzing imprecision which is arisen during the last decades is the fuzzy theory (Zadeh,1965), which abandons the Aristotelian logic to two values and defines fuzzy sets as those whose confinements are not precise but rather vague. This theory has the ability to represent the concepts of vagueness that natural language uses for pointing out qualitative variables, such as warm, cold, lukewarm, etc. (Enea and Salemi, 2001). At first it is necessary to briefly explain the “fuzzy set” theory.

Fuzzy set

Fuzzy sets were first introduced by Zadeh (1965) as a possible way to handle uncertainty in processing uncertain or imprecise data and vague expert knowledge. This theory allows the notion of gradation to express whether or not an element belongs to a set. Contrary to the ordinary set theory, in which each element either belongs to or does not belong to a set, a fuzzy set is described by a membership function, representing numerically the degree to which an element belongs to a set. Let us consider a classical set $U$, usually called “universe of discourse” or “reference” set. $F$ is called a fuzzy set of $U$ if $F$ is completely defined by the set of pairs:

$$F=\{(x,m_F(x); x \in U, m_F(x) \in [0,1]\}$$

in which $m_F$ is called the membership function of $F$. The closer $m_F(x)$ is to 1, the more $x$ belongs to $F$ and inversely the closer $m_F(x)$ is to 0, the less $x$ belongs to $F$. The upper boundary of $m_F(x)$ is called the height of $F$ and a fuzzy set with height =1 is called a normalized fuzzy set (Freissinet et al., 1998). The “possibility theory” was coined by L A.
Zadeh in 1978 as an approach that describes uncertainty induced by pieces of vague linguistic information and it can be defined by fuzzy sets (Zadeh, 1978). In using the possibility theory, knowledge about a variable x is represented by a possibility distribution $\Pi x$ (Zadeh, 1978; Dubois et al., 1988) mapping the domain of x into the $[0,1]$ interval. The distribution $\Pi x$ can be viewed as the membership function of the fuzzy set of possible values of x. For any value $u$, $\Pi x u$ is the degree of possibility that $x=u$, with the convention $\Pi x(u)=0$ indicating that x can not take the value u, and $\Pi x(u)=1$ means that x would take the value of u. In the transition area with possibilities strictly between 0 and 1, $\Pi x(u_1) > \Pi x(u_2)$ expresses that $u_1$ is a more plausible value than $u_2$.

Saturated hydraulic conductivity is one of the soil physical properties with relatively high coefficient of variation. Concerning soil heterogeneity, literature values had to be adopted. Chappell et al. (1998) computed 514% and reported 86-190 and 358%, and Devis et al. (1999) reported 135-740% as coefficient of variation using the geometric mean for K$_{sat}$ in their data set. Because of high coefficient of variation for this parameter, researchers often have not been able to clarify significant differences for K$_{sat}$ due to applying different treatments affecting K$_{sat}$. Although in current decade the fuzzy approaches are being applied to soil data analysis, but the problem mentioned above with K$_{sat}$ and interpreting its measured values, has not still gotten away. How this problem can be solved? How can we rely on our discussion and result using classical analysis, and how far we continue to represent our knowledge with the uncertainty attached to it? The most difficulty encountered in K$_{sat}$ analysis by classical statistics is its imprecision. By fuzzification of K$_{sat}$, its analysis may be simplified and the complication can be avoided.

In this paper we introduce the fuzzy set theory as a new approach to analyze measured data of K$_{sat}$. The method presented herein is based on the idea of using fuzzy sets to describe imprecision in saturated hydraulic conductivity. Given a fuzzy set F defined on U and any number $\alpha \in [0,1]$, the $\alpha - cut$ represents the set of elements with membership function $\geq \alpha$ (Ross, 1997). The estimate of imprecision depends on the choice of the $\alpha - cut$ value used in the fuzzy calculations. It represents the degree of confidence of the results required by the user (Freissinet et al., 1998). The non-classical methodology to describe K$_{sat}$ based on the fuzzy set and fuzzy logic theory leads to the result of a calculation being expressed by a triplet $(Y, \Delta \alpha Y, \alpha)$, which represents the main value Y of the measured data such as K$_{sat}$ given by the mathematical calculation, the imprecision range $\Delta \alpha Y$ of the main value and $\alpha$ the degree of confidence of each value belonging to this imprecision range (Freissinet et al.,
1998). At first we will define a “universe of discourse” or “reference” set for \( K_{\text{sat}} \). Then fuzzify each value of \( K_{\text{sat}} \) using bell shape fuzzification method.

**Method and Materials**

Data discussed in this paper were taken from a field experiment on a sandy loam soil at research center in Hamadan. Undisturbed samples were taken and their \( K_{\text{sat}} \) were measured with constant head method. To define a reference set \( X = \{x_1, \ldots, x_n\} \), where \( x_i \) is each member of \( X \), we first sorted the data presuming each value as a member of universe set, and then added some members to the set with magnitudes less than the minimum and more than the maximum value of the data. We chose an arbitrary value and fuzzified it by the bell shape method \( \pi_i = \pi(x_i), \quad i = 1, \ldots, n \), where \( \pi_i \) is the membership function for \( i \)th member and can use membership functions as a possibility distribution (Zadeh, 1978).

**Fuzzification**

Fuzzification is the process of making a crisp quantity fuzzy. We do this by simply recognizing that many of the quantities that we consider to be crisp and deterministic are actually not deterministic at all: they carry considerable uncertainty. If the form of uncertainty happens to arise because of imprecision, ambiguity or vagueness, then the variable is probably fuzzy and can be represented by a membership function (Ross, 1997). There are possibly more ways to assign membership value. We used bell shape fuzzification method that follows:

\[
\mu_a(x) = \frac{1}{1 + d(a - x)^2} \tag{2}
\]

where \( \mu_a(x) \) is the membership function for the value \( x \) from the set \( X \), \( x \) is each member of the set \( X \), \( a \) is the value that has to be fuzzified and \( d \) is wideness coefficient (Ross, 1997). Then by differentiating the \( \mu_a(x) \) to several partitions \( X = \{A_1, \ldots, A_m\} \) and \( \pi_j = \pi(A_j), \quad j = 1, 2, \ldots, m \) in such a way that \( \pi_1 \geq \pi_2 \geq \pi_m \) (where \( \pi_j \) is the possibility that \( A_j \) could take), a probability distribution can be derived by the following rule (Dubois and Prade, 1988):

\[
\forall x \in A_j, \quad PA_j(x) = \pi_j - \pi_{j+1} \tag{3}
\]

in which \( PA_j(x) \) is the probability level for the portion \( A_j \). MATLAB software was used to fuzzify \( K_{\text{sat}} \) and the EXCEL program was used to draw the graphs.
Results and Discussion

Figure 1 illustrates the $K_{sat}$ fuzzy number equal to 0.001 cm/h and also uncertainty of $K_{sat}$. Although in the classic method we were unable to assign any other values to the sample that its $K_{sat}$ was measured, but in the fuzzy method it is developed to take a lot of values in such a way that the possibility of measured value is equal to 1 and the possibility of other values (more than measured value and less than it) is less than 1. Using this diagram (Figure 1) it is possible to estimate the degree of confidence for each value of $K_{sat}$ that could be taken. We use $\alpha$-cut method to determine the range of imprecision in which the possibility of the values was more than a specific membership function.

By transforming diagram in Figure 1 to probability distribution, we could discuss about the probability level of each value and then classify $K_{sat}$ fuzzy number into nine classes (Figure 2). By using equation 3, Figure 2 was transformed to probability distribution diagram (Figure 3). In this way we expressed a triplet ($Y, \Delta \alpha Y, \alpha$), similar to $\alpha$-cut expression, where $Y$ is the main measured value, $\alpha$ is probability level that values occurred in $\Delta \alpha Y$ ($\Delta \alpha Y=Y_2-Y_1$), $Y_1$ and $Y_2$ are the beginning and end of the chosen range, respectively. By increasing $\Delta \alpha Y$ more values were added to the range between the two values and thus the related $\alpha$ was increased, meaning that larger $\Delta \alpha Y$ resulted in higher $\alpha$ values. In this way all the probable values to $K_{sat}$ for describing the variability coefficient was considered. But in the classic method these values of $K_{sat}$ were ignored and only the measured values are considered.

The probability distribution curves maybe used to show the probability levels of measured $K_{sat}$ between the two specified values. But fuzzy method illustrates a measured value with its variability and interprets it much better than measured value per se. By this process a measured value can be expressed as a curve instead of just a single measured value that is not precise. By using the probability distribution, the probability level for each value can be determined. The highest probability belongs to the measured value, and the other values which are less probable are also considered. The more the difference between the measured and probable values, the less the probability level would be possessed. On the other hand in the classic method only one measured value is considered and other values are ignored and thus in the case of occurrence and probability, may cause a big misinterpretation in the discussions. Although the classic method may seem simple but the fuzzy method has the following advantages: a) it enables one to avoid from the invalidity of the statistical data, b) by this method one can illustrate the variability of the $K_{sat}$, c) the fuzzy method might be
better than the conventional method, because the $K_{sat}$ is an imprecise parameter, and the fuzzy number for $K_{sat}$ may be more reliable, d) in contrast to the classic method, fuzzy can discuss about the variability of $K_{sat}$ only by using a single specified or measured value, as it is demonstrated for the $K_{sat}$ value of 0.001 cm/h. The fuzzy method adds an imprecision range and a probability level to the measured value and could be a complementary method.

References
Figure 1: Fuzzy numbers for a measured $K_{sat}$ of 0.001 cm/h

Figure 2: Possibility distribution of $K_{sat}$

Figure 3: Probability distribution of $K_{sat}$