

Using fuzzy sets for data interpretation in natural analogue studies

F. L. De Lemos

CNEN- National Nuclear Energy Commission; Rua Prof. Mario Werneck, s/n, BH –
CEP: 30123-970, Brazil

T. Sullivan

BNL-Brookhaven National Laboratory, USA

K-H Hellmuth

STUK- Radiation and Nuclear Safety Authority, Finland

ABSTRACT

Natural analogue studies can play a key role in deep geological radioactive disposal systems safety assessment. These studies can help develop a better understanding of complex natural processes and, therefore, provide valuable means of confidence building in the safety assessment.

In evaluation of natural analogues, there are, however, several sources of uncertainties that stem from factors such as complexity; lack of data; and ignorance. Often, analysts have to simplify the mathematical models in order to cope with the various sources of complexity and this adds uncertainty to the model results. The uncertainties reflected in model predictions must be addressed to understand their impact on safety assessment and therefore, the utility of natural analogues.

Fuzzy sets can be used to represent the information regarding the natural processes and their mutual connections. With this methodology we are able to quantify and propagate the epistemic uncertainties in both processes and, thereby, assign degrees of truth to the similarities between them. An example calculation with literature data is provided.

INTRODUCTION

Some of the natural processes that comprise a radioactive waste disposal system are too complex to be properly represented in performance assessment mathematical models. Besides complexity another problem faced by analysts is that the models are applied to evaluate the behavior of the system over a long period of time, while input data are based on short term laboratory experiments.

Natural analogue studies can provide important information to help improve models either during the development of conceptual models or in the validation of the performance assessment results.

Due to the complexity of natural systems, analogue information is primarily qualitative or semi-quantitative, because it is impossible, in most cases, to quantify all relevant parameters in natural systems. [1]

The following paragraph illustrates perfectly the need for mathematical tools that are able to interpret the obtained data and properly represent the uncertainty that stems from complexity, and lack of knowledge, about the processes under

observation.

According to [2]: ..“we must accept that much of the information stemming from analogue studies is inherently “fuzzy”. Where this is the case, it should not necessarily be seen as a drawback, as it simply reflects the complexity of the natural environment. Rather, we should learn from it. If unraveling past events produce fuzzy results, than it is clear that making predictions into the future will give equally, if not more fuzzy results. We are, perhaps, learning to be more realistic about this predictive side of the coin, and appreciating more the unavoidably fuzzy context in which numerical performance assessment results are produced. This will eventually begin to temper the way in which we evaluate and act on these results.”

Fuzzy sets theory was developed to deal with fuzziness and vagueness in information. The objective of using fuzzy logic in natural analogues studies is to quantify the degree of similarity between natural systems and a repository system. This is because, although it is impossible to have a natural system that is exactly equal to a repository system, it is always possible to find degrees of chemical, physical and morphological similarity between both systems. This can help transfer information from natural systems more effectively.

FUZZY SETS AND FUZZY LOGIC

In this section some of the principles used for data analysis and interpretation are presented.

A fuzzy set is a subset \tilde{A} of the universe of discourse X, where the transition between full membership and no membership is gradual rather than crisp, [0, 1]. In a classical set, on the other hand, the transition between full membership and no membership is abrupt (either 0 or 1).

A fuzzy set \tilde{A} in X can be represented by a set of ordered pairs such that

$$\tilde{A} = \{x, \mu_{\tilde{A}}(x)\}, \quad x \in X, \quad (1)$$

Where $\mu_{\tilde{A}}(x)$ is the degree of membership of x to \tilde{A} and is a number in the interval [0, 1].

Fuzzy Logic and fuzzy rule based systems

Fuzzy logic is an extension of classical logic when partial truth is included to extend bi-valued logic (true or false) to a multi-valued logic (degrees of truth between true and not true). [3]

The use of mathematical equations as a representation of linguistic expressions makes it possible to keep record of important information and analysis of uncertainty propagation. This will be the basis for development of a rule based system. One of the most common ways of knowledge representation is through the natural language expressions

IF premise (antecedent)...THEN conclusion (consequent). [3]

The above rule can be represented as relations between input and output (mappings).

$$y = x \circ \underset{\sim}{R} \quad (2)$$

Where $\underset{\sim}{R}$ is a result of union or intersection depending to the type of aggregation “AND” or “OR”.

Figure 1 depicts an example of a simulation of a non linear relation between input (x) and output (y) in a hypothetical process. In this figure it is shown that, instead of points in the input universe (X) and output (Y), there are fuzzy sets that are represented by linguistic descriptions.

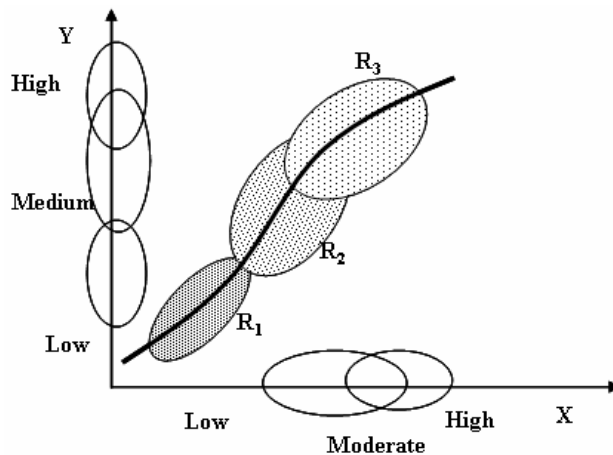


Figure 1: A fuzzy nonlinear relation between input and output. (Adapted from [3])

RADIONUCLIDE MIGRATION

One of the best examples of application of natural analogues is the study of the retardation of radionuclides in the geosphere. The term “retardation” includes a number of mechanisms (diffusion into the matrix, molecular filtration, ion exclusion, physical sorption, ion-exchange, mineralization and precipitation, etc.) but, in practice, they tend to be not well-differentiated, and the net of the individual mechanisms is often (imprecisely) represented by a distribution coefficient, K_d . [1]

This coefficient only refers to an empirical representation of kinetically fast and reversible sorption/desorption process. As sorption is a result of several processes, that are site specific, the meaning of the distribution coefficient needs to be evaluated on a case-by-case basis.

For the above mentioned reasons there will be degrees of similarity between the site conditions, obtained data, and the repository site. In terms of fuzzy sets nomenclature degrees of membership are assigned to parameter values allowing for site specific

differences to be taken into account in the analogy process between the natural environment and repository system.

In the next section a methodology to build a fuzzy set from empirical data is presented. In the following section an example of how fuzzy sets can be applied for data interpretation is provided.

DERIVING FUZZY SETS FROM OVERLAPPING INTERVALS

For field data to be represented by fuzzy sets some mathematical developments are necessary. The first step is to show the equivalence of fuzzy sets and possibility distribution. A detailed discussion of this subject is not in the scope of this work and can be found in the literature [3]. However, a short discussion of this subject is presented.

Under special conditions a possibility distribution can be interpreted as a fuzzy set. Zadeh, [4], defined a possibility distribution as a fuzzy restriction that acts as an elastic constraint on the values that may be assigned to a variable. For example, let \tilde{A} be a fuzzy set on a universe X , and let the membership value, μ , be a variable on X that assigns a “possibility” that an element of x is in \tilde{A} . So we get

$$\Pi(x) = T(x)$$

Where:

$\Pi(x)$ is the induced possibility distribution over the set X

$T(x)$ is the grade of membership x in the fuzzy subset \tilde{A}

Or,

$$\Pi(x) = \mu_{\tilde{A}}(x) \quad (3)$$

In order to develop a possibility distribution, from empirical data, the intervals should be consonant, i.e. they should be nested within each other. However, empirical data are seldom nested and special methods exist to transform nonconsonant intervals into consonant intervals in accordance with available evidence. [3]

Considering a system that can be described within a domain $X = \{x_1, x_2, \dots, x_n\}$, the behavior of which is described by evidence obtained as observations over a collection of sets; F . The set X is called the domain of the system. For example, the domain of the system can be all the possible K_d values within a specific region.

Now, let $f = \{[A_j, w_j]\}$ represent the original intervals along with their weights. If M measurements are observed, then the weights w_j of each observation A_j is calculated by frequency analysis as

$$w_j = n(A_j) / M \quad (4)$$

Where $n(A_j)$ is the frequency count of interval A_j , $\sum_j w_j = 1$ and $\sum_j n(A_j) = M$.

The most possible interval is the one that occurs most frequently, and in which experts have the most confidence. After selecting the consonant e intervals the weights from nonconsonant intervals are then redistributed in such a way that the total information from underlying evidence is preserved.

In next section, empirical data analysis based on this methodology is presented. With this analysis it was possible to assign degrees of membership to K_d values. For more information on possibility theory the reader is encouraged to refer to the literature. [3, 4]

DERIVING A FUZZY K_D FOR URANIUM

This study is based on the work by Alexander and McKinley (CEC, 1990). That study attempted to validate a transport mathematical model by comparing calculated concentrations of radionuclide with observed ones. Values of K_d are tested to find the best fit between predicted and measured values.

The observed concentrations can not be exactly equal to the calculated ones because mathematical models always have simplifications regarding processes involved in retardation. Also, some natural processes may have not been considered in the model. Therefore, as mentioned earlier, degrees of fuzziness are observed when empirical data are compared with the predicted ones.

Figure 2 shows the calculated and measured uranium activity ratios with distance from an outcrop. The decrease in the activity ratios with distance in the aquifer were taken as a measure of the rate of U movement in the aquifer, which is related to the rate of groundwater movement via the R (or K_d) factors of the transport equation. (CEC, 1990)

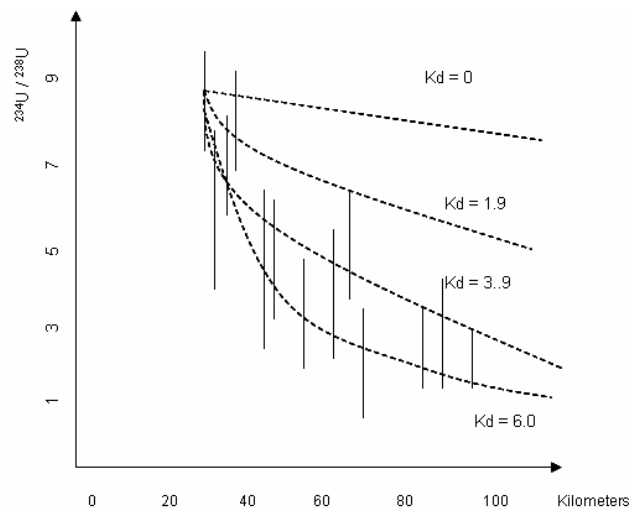


Figure 2: Calculated and observed activity ratios of uranium with distance.(Adapted from CEC, 1990)

We are interested in the comparison of assumed K_d values for the calculation and the observed data. Four K_d values were tested: 0; 1.9; 3.9 and 6.0. It was found that a

$K_d = 6 \text{ ml g}^{-1}$ was the best fit. In the case of deterministic calculation this value would be chosen. But according to Figure 2, the different K_d values have different degrees of fit to the data depending on the analyzed region.

In the fuzzy sets perspective, all of those K_d values can be considered as members of a set of values that compose a fuzzy set for K_d . Our purpose is to derive degrees of membership for these K_d values based on the classification in order to build a fuzzy K_d .

To this end, a first step will be the classification of the data according to the level of activity ratios and distance from the source. By observing Figure 3a it can be seen that the activity ratios can be classified into High, Medium and Low levels of activity, which are represented by the respective circles.

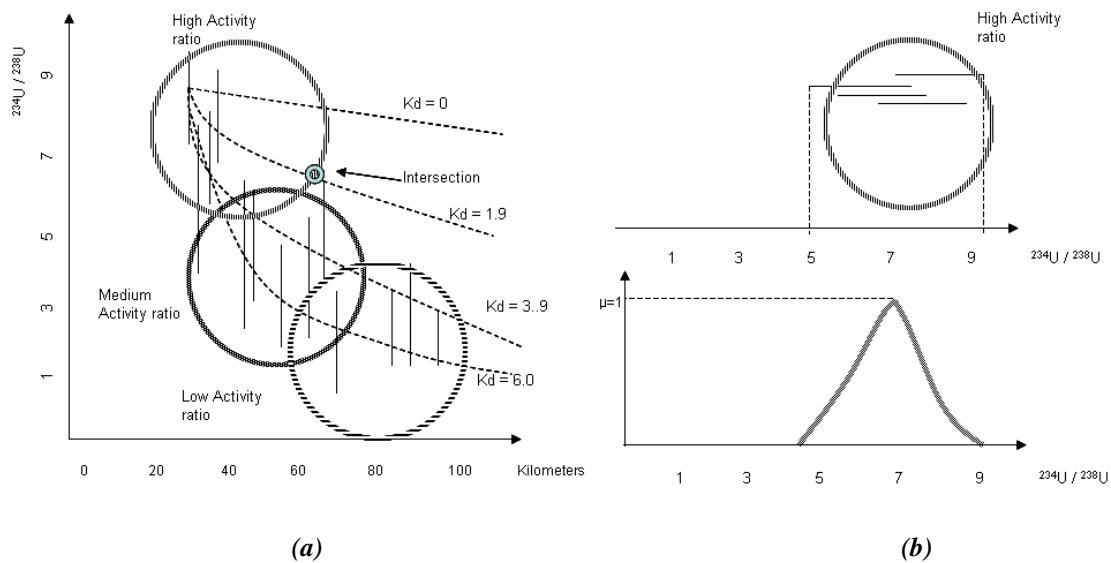


Figure 3a and 3 b: Fuzzy classification of the activity ratios. Adapted from (CEC, 1990)

Figure 3b shows an example of translation of ranges of data of the class “High Activity” ratio into a fuzzy set. The top circle in Figure 3b represents the region of high activity ratio, in Figure 3a and the activity ration ranges from about 5.5 to 9.5. The bottom graph in Figure 3b shows the membership function, which peaks around 7 and is contained in the activity ratio region of 5.5 to 9.5. According to the process, described in [3], weights are assigned to intervals that are more frequent in all ranges.

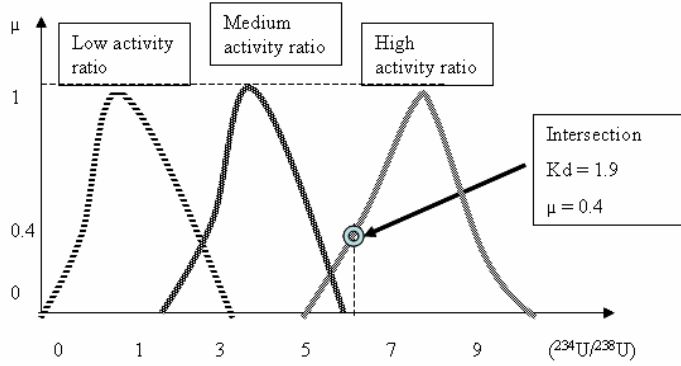


Figure 4: Fuzzy sets representing three intervals of activity ratios for Uranium.

The method is then applied to each class, generating three possibility distributions, or fuzzy sets shown in Figure 4. Note in Figure 3a, the intersection of the $K_d = 1.9$ curve with the High Activity Region occurs just below 7. From the membership function, this intersection leads to a degree of membership of 0.4 as shown in Figure 4. This is only possible because the generated possibility distributions will have the same properties of a fuzzy set. Now we use these fuzzy sets for the activity ratios to generate the degrees of membership of the K_d values to the fuzzy set for K_d 's.

We assume then that this degree of membership represents the degree of truth for that specific K_d value. In other words, a degree of membership is assigned to a K_d value according to the degree of membership of it's intersection point at the activity ratio's class. This approach then generates a collection of K_d values that comprises a fuzzy K_d for each class of U activity ratio, Table I.

Table I: Fuzzy K_d 's derivation according to intersection to the three classes of activity ratio

K_d Value	Degrees of Membership		
	K_{d1} (Low Activity ratio)	K_{d2} (Medium Activity ratio)	K_{d3} (High Activity ratio)
0	0	0	0.6
1.9	0	0	0.4
3.9	0.5	1	0.3
6.0	1	0.6	0.2

As, in our example there are three fuzzy activity ratios then three fuzzy K_d 's (K_{d1} , K_{d2} , K_{d3}) are derived and can be carried on to and propagated throughout the performance assessment calculation. In other words, all the information is conveyed throughout the calculation. Then, instead of one value, the K_d is represented as a fuzzy number, or fuzzy K_d , while the safety assessment calculation is performed accordingly to the appropriate mathematics.

With this approach the $K_d = 6$, which was considered the best fit, now has a degree of membership, $\mu = 1.0$ in K_{d1} (Low activity ratio), $\mu = 0.6$ in K_{d2} (medium activity ratio), and $\mu = 0.2$ in the K_{d3} (high activity ratio), Table I. Thus we can use each value of K_d in the traditional performance assessment calculations and interpret the results according to the degrees of truth or confidence. This allows us to convey more information attached to each value.

CONCLUSION AND DISCUSSION OF RESULTS

Fuzzy sets are an effective way of quantifying semi-quantitative information such as natural analogues data. Epistemic uncertainty that stems from complexity and lack of knowledge regarding natural processes are represented by the degrees of membership. It also facilitates the propagation of this uncertainty throughout the performance assessment by the extension principle. This principle allows calculation with fuzzy numbers, where fuzzy input results in fuzzy output.

This may be one of the main applications of fuzzy sets theory to radioactive waste disposal facility performance assessment. Through the translation of natural data into fuzzy numbers, the effect of parameters in important processes in one site can be quantified and compared to processes in other sites with different conditions. The approach presented in this paper can be extended to facilitate a comparison between sites. For example, the degrees of membership can be a measure of similarities between sites and, consequently, confidence in validation tests of models can be enhanced.

REFERENCES

1. IAEA - International Atomic Energy Agency: *Use of natural analogues to support radionuclide transport models for deep geological repositories for long lived radioactive wastes*; IAEA, Vienna 1999.
2. CEC - Commission of the European Communities; *Fourth natural analogue working group meeting and Poços de Caldas project final workshop* – EUR-13014 EN; Scotland, 1990.
3. Ross T, (2004). *Fuzzy Logic with Engineering Applications*; McGraw-Hill, New York, 627 pp
4. Zadeh, L. *Fuzzy sets as a basis for a theory of possibility*; Fuzzy Sets and Systems 100 Supplement (1999); 9-34; North Holland.