Using Fuzzy Logic as a Complement to Probabilistic Radioactive Waste Disposal Facilities Safety Assessment -8450

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

ABSTRACT

Probabilistic calculation assumes randomness in events and processes, i.e., it assumes that all the observations in a probability distribution have the same possibility of occurrence. The only difference from one event to another is their respective frequency.

To avoid the situation that a very low probability event drives the decision making process, there must be a criterion for the predicted dose as a function of probability to be considered acceptable by regulatory authorities. For example, it can be required that the predicted dose at 2 standard deviations above the mean value for dose to be no more than three times the regulatory standard.

This paper proposes the use of possibility analysis, as a complement to the probability analysis. In this approach two separate performance analyses, probabilistic and possibilistic, are performed and the results are used to complement each other. A case example is provided to illustrate the methodology.

INTRODUCTION

According to the 1989 International Atomic Energy Agency (IAEA) report, [1], the uncertainties can be classified as type A, aleatoric, and type B, epistemic. Aleatoric uncertainty is generated by occurrence of random and independent events, while epistemic uncertainty is generated by factors such as lack of data, ignorance and high complexity of the system.

Probability theory is used to model aleatoric uncertainty. A number of methodologies exist to deal with the epistemic uncertainty and they are called non-probabilistic methods.

The two types of uncertainty, aleatoric and epistemic, are sometimes mixed and modeled as probabilistic. This happens because the input distributions of data are also used to represent lack of understanding of some processes. For example, in high-level waste performance assessment, the fuel dissolution rate or input data related to canister failure contain both types of uncertainty [2]. This made it very difficult to evaluate the impact of the epistemic uncertainty in the Total System Performance

Assessment (TSPA). This is even more evident in the cases where some conditions are characterized by linguistic expressions.

Because TSPA is comprised of very complex studies where data with different formats have to be aggregated in the same framework, it is almost always practically impossible to distinguish, and treat, the uncertainties according to their different sources.
The method presented in this paper addresses this difficult situation and consists of using possibility as a complement to the traditional probabilistic approach.

MONOTONE MEASURE

The monotone measure describes the assignment of an element, x_i, to two or more crisp sets. Where a crisp set is a set with well defined boundaries, as opposed to a fuzzy set which is a set with flexible boundaries. Please refer to the literature for more information. [3, 4]

Monotone measures are an umbrella theory that is comprised of some different forms theories such as, belief, plausibility, possibility, necessity, etc. A more complete study of all of them is beyond the scope of this text. Additional information is available from [4].

In a universe *U*, an element, x_i, is assigned to a crisp set, *Qj*, according to the available evidence. There is no uncertainty about the definition of the crisp set, but uncertainty exists about the evidence to establish an assignment of an element to the set. This is not a random notion.. The evidence can be completely lacking – the case of total ignorance – or evidence can be complete – the case of probability assignment. [3]

Example of assignment

Figure 1 shows a schematic representation of assignment of elements to a crisp set. As an example, suppose that the set Q_1 is a collection of all boxes that are colored with shades of green, and Q_2 is a collection of shades of yellow.
Then suppose that each element, B_i, is a box that is assigned to the crisp sets Q_1 or Q_2 according to their color.

Depending on the mechanism of analysis, it may be difficult to distinguish some colors from each other so each box is correctly assigned to the right crisp set. For example, the box B_5 can have some shades of yellow mixed with green, so it has membership in $Q_1$and Q_2.

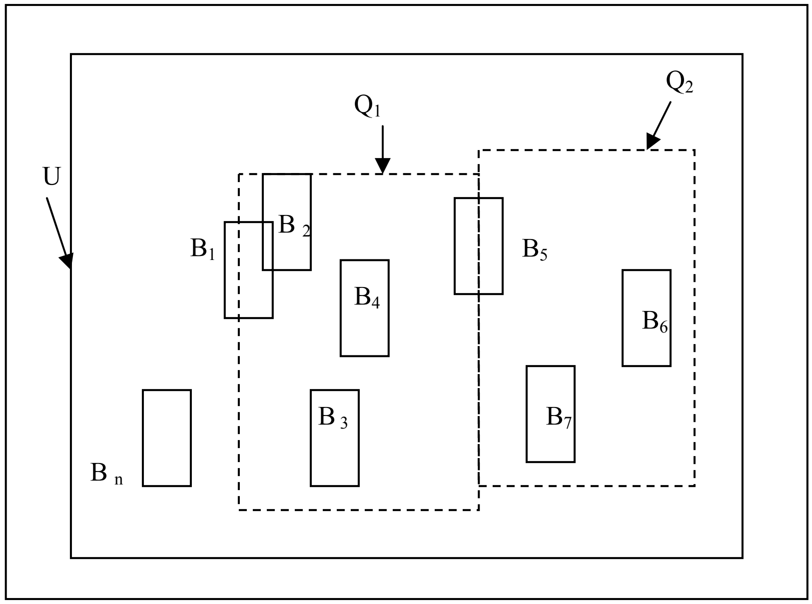

Figure 1: Schematic representation of elements to crisp sets in a universe U

This can be compared to a classification process where the elements are assigned to crisp set according to the available evidence. As the quality of evidence can change, the assignment of elements, or subsets, to the crisp set, *Q*, could be viewed as a fuzzy classification. This means that one ball, B_i, can belong to one or more crisp sets, Q_i.

In this example, as more information is made available, or a new mechanism is used for color identification, the composition of the sets Q_1 and Q_2 can change.
Within this context, the fraction of elements that have degree of membership =1 to the set *Q* is called belief (*Q*), and the fraction of elements that have degrees of membership higher than 0 to the set Q is called plausibility, (*Q*)

Possibility theory

It can be shown that, under certain conditions, the belief and plausibility measures are defined as necessity and possibility. The reader is encouraged to find more explanations in the literature. [3], [4]

A possibility distribution function π is a mapping of the singleton elements, *x*, in the universe, *X*, to the unit interval:

$$\pi(X) = \max_{x \in A} r(x) \qquad \text{Eq. (1)}$$

Where:

$$r : X \rightarrow [0,1] \qquad \text{Eq. (2)}$$

Because of the similarity of axioms of fuzzy logic to those of possibility theory, the possibility distributions can be represented as a membership function. [4, 5]

Therefore the mathematical developments for propagation of fuzzy sets, fuzzy arithmetic and interval analysis, are applied for possibility distribution as well.

It is in this context that fuzzy logic is used in this paper.

Fuzzy Sets

Fuzzy sets can be interpreted as being a collection of elements which respective degrees of membership represent the degrees of compatibility of the elements with the others members of that collection or class. The fuzzy set is mathematically represented by a membership function. A fuzzy set can then be viewed as a representation of a classification of the elements and their respective degrees of membership are assigned according to the support, or evidence, for the classification.

In a calculation, if the input data are given in the form of fuzzy sets, then the result is also given in the form of a fuzzy set. This result, in the form of a fuzzy set, can then be viewed as a classification of the several results, R_n, to a class, $\underset{\sim}{R}$ (fuzzy result).

Similar to the Monte Carlo method, a fuzzy set can be viewed as a collection of results of realizations, each one with a specific combination of parameters that will provide support for the assignment of degrees of membership to a set of results.

In other words, a fuzzy result $\underset{\sim}{R}$ is a picture of the system under analysis showing its compatibility to a collection of possible results.

A fuzzy set, $\underset{\sim}{A}$, is represented as a vector as follows:

$$\underset{\sim}{A}=\left\{\frac{\mu_1}{x_1},\frac{\mu_2}{x_2},.....\frac{\mu_n}{x_n}\right\} \qquad \text{Eq. (2)}$$

Where:
μ = degree of membership

x_i = element of the set with n elements.

Probability measure

In a probability measure all the elements are assumed to have complete membership to the set *Qi*. There is no doubt about the evidence to the assignment, i.e., their membership can be described as a binary relation, zero or one.

If the assignment is thought of as a membership function, then only the elements with degree of membership =1 can be described probabilistically. This concept is shown in Figure 2. The top figure shows the degree of membership of all values x to a set. The bottom of figure 2 shows the frequency (probability distribution function) of the values of *x* in the set. Note, frequency is not defined when the membership is less than 1.

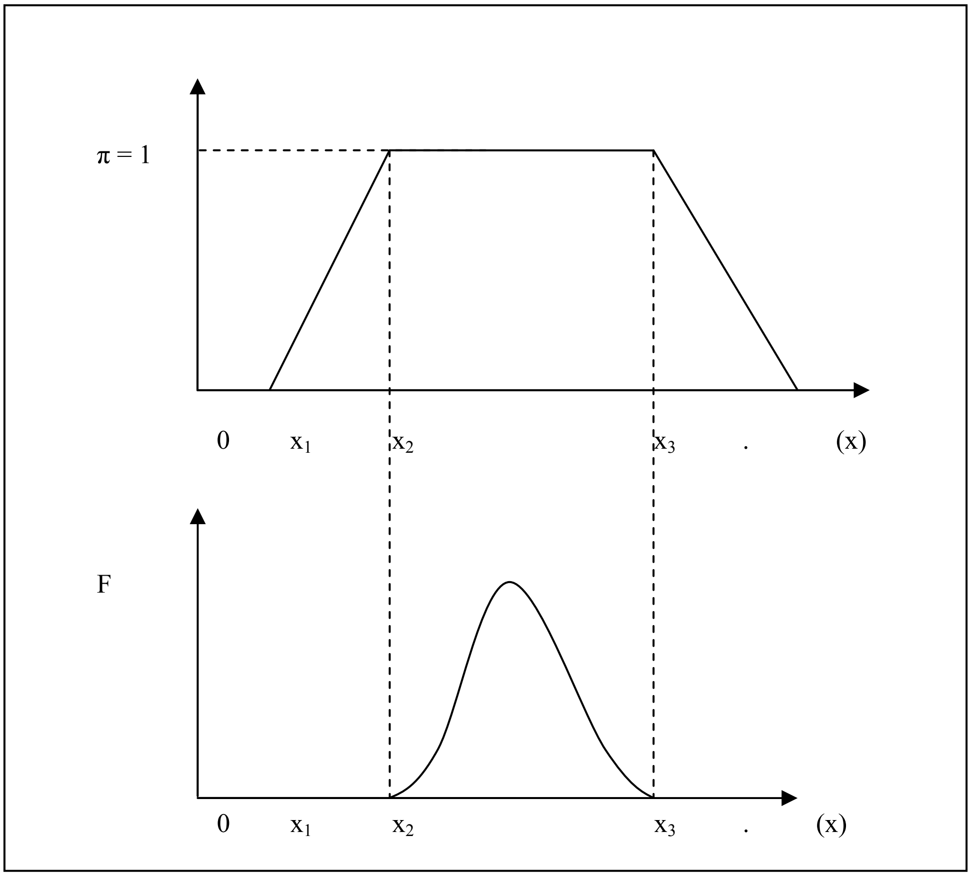

Figure 2: *Elements defined probabilistically have the same degree of membership to a crisp set Q_i.*

Having a degree of membership of 1 in the set leads to the law of the excluded middle. From this law, if one knows the probability of *A, P (A)*, then the complement of the probability of A is also known:

$$P(A) = 1 - P(\overline{A}) \qquad \text{Eq. (3)}$$

In the case of a possibility distribution, the elements can have different degrees of membership and, therefore, can have membership to different sets.

POSSIBILITY VERSUS PROBABILITY DISTRIBUTIONS

As mentioned above, a probability measure assumes that all the elements have complete evidence in their assignment to a crisp set, Q_i. Then, if in a collection of elements, $x_i \ldots x_n$, with various degrees of evidence, only those elements, $x_k \ldots x_m$, which have complete evidence, will be described probabilistically.

In TSPA, due to lack of enough data, the probability distribution functions are often based on expert's judgment. In other words, experts use the collection of discrete points and extend it to a continuous distribution. The new values thus created do not necessarily have the same degree of support (i.e. membership less than 1), which means that the assignment of an element, x_i, to a crisp set Q_i, may not be based on complete evidence. Consequently the probability distribution functions may not be based on the assumption of full membership.

Here is where the application of possibility theory would be useful as a complement to the probability approach to TSPA. It may happen that the elements of the probability distribution function have different degrees of compatibility with the set being analyzed. The degree of compatibility of an element could be used as a complement to its probability.

As it will be shown in the next section, soil samples are assigned to a class of K_d values (e.g. Low, Medium, High) according to each sample composition.

The analyzed samples are used as a basis for a fuzzy classification [7]. In the example, the samples are classified according to their features, or composition. As each sample has a specific K_d value, these values will be classified following the same membership function, in the same classes.

Within a class, the degrees of membership of each element are also a measure of their respective degrees of compatibility with that class. When new data are available, it is possible to calculate its degree of membership to the already built classes. This process is called pattern recognition.

Through this process a membership for the data is calculated, which is translated into degree of compatibility. This process takes into account the characteristics, or features, of the sample, and not just the frequency of occurrence. Therefore, the probability distribution function can show different results from a membership function.

A CASE EXAMPLE: Kd VALUES AS POSSIBILITY DISTRIBUTION

For a certain radionuclide, the distribution coefficient, K_d, depends on the soil features. In analogy to the example of colored boxes, where the color of each is the evidence for its assignment to a set; we can consider the soil characteristics as the evidence to assign a certain sample to a set, or class, of K_d values. This class of K_d values can be defined, for example, as Low, Medium and high K_d's.

Because of the complexity of factors that influence the distribution coefficient, this assignment is not a straightforward task. Also, even though the laboratories measurements are precise, laboratory conditions are different from field conditions and, therefore, the actual values of K_d, for a certain region, can only be roughly evaluated.

In other words, having a universe of possible K_d values, there is not complete evidence for the assignment of a soil sample to a class of K_d value, or more importantly for TSPA, to a deterministic value. In this example, the evidence is based on soil composition, for example, the percentage of clay, sand, organic material and silt.

Following the nomenclature provided by evidence theory, in this case example we consider soil samples as the elements, B_i, to be assigned to a crisp set; and the single Kd value as the crisp set Q. Each element (soil sample) can have membership to one or more crisp sets,

As mentioned before, the process of assignment can be compared to a fuzzy classification which will generate a membership function that defines the degree of membership, or compatibility, of individual elements to a class. The definition of a class can be done in reference to a specific K_d value as shown in Table 1.

Table 1: K_d values as center of a class

K_d	**Class**
0	Low
200	Medium
1000	High

In the following example, soil samples are classified according to their degree of compatibility to the K_d value 200, or "Medium K_d".

Another explanation is that the degree of support for the assignment of a soil sample to a particular value of K_d can be viewed as the degree of compatibility of that soil sample to that particular K_d. Geochemical conditions have a strong spatial and temporal variability, and the use of a deterministic value (e.g., geometric mean) in TSPA, or even a probabilistic distribution may not capture the actual situation. By using this methodology each sample will have different degrees of assignment to a class of values and the region can be defined by a membership function, or a fuzzy number. For example, membership functions could be developed based on soil properties to suggest the Kd class that is most compatible with the soil properties.

This methodology is not meant to be a substitute to a probabilistic representation of the data. The probabilistic distribution assumes that all points have the same degree of membership to a class, or μ =1. However due to several reasons this is not always true. Then the representation of the data also in terms of a fuzzy set can have an important impact on the confidence one can pose in the TSPA results.

As a practical example, let's take the case presented in [7]. Data from 25 soil samples were collected from the literature and classified according to their respective composition, i.e., percentage of clay, sand, organic and silt. Each sample is also associated with a K_d Value which was determined through laboratory tests. Therefore, the classification of a sample, according to its composition, was also was a measure of the compatibility of each sample's K_d value, to a particular value.

Table 2 shows the composition of the elements (samples) of one of the classes. The compositions of the samples were used for the fuzzy classification. [3]. As each sample corresponds to a specific K_d, the result, shown in the Figure 3, is given in terms of a membership function of K_d values.

Table 2: Samples used for building the regression equations *

Sample (n)	Sand (%)	Silt (%)	Clay (%)	Organic (%)	Kd (ml/g)
0	74	3	23	0	405
4	100	0	0	.03	119
6	96	4	0	.51	74
11	95	2	2	.3	510

* Adapted from [7]

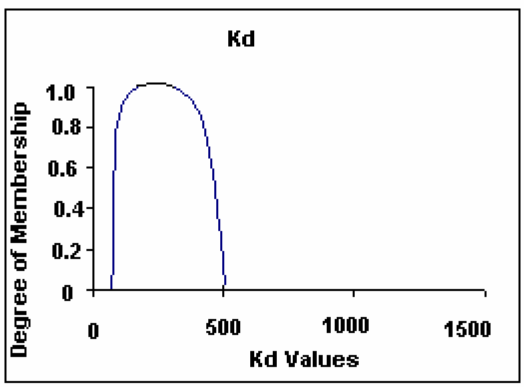

Figure 3: *Membership distribution function of Kd values. Adapted from [7]*

The degrees of membership of the K_d s can be viewed as degrees of compatibility of that value with the region from where the samples were taken. This is different information than a probabilistic distribution, or even a deterministic value.

One advantage of this methodology is for analysis of regions where we do not have data but may have other information that suggests certain values are compatible or not compatible with this information (i.e. soil type for Kd).

Experts could build a probability distribution function based on their experience. This methodology may improve over an expert redefining the Probability Distribution Function (PDF) to account for these soil types.

In this sense the two approaches, probabilistic and possibilistic, can be complementary. In a probabilistic approach, all the elements have the same likelihood to occur, therefore they should have the same degree of possibility, or membership $\mu = 1$.

However, traditionally, due to lack of data, experts build the PDF's based on their professional experience. Therefore, the concept in Figure 2 may not apply to this case. In this case the possibilistic distribution could be used as a complement to the PDF, acting as a measure of the "reasonable assurance" that the element under consideration has a high degree of compatibility with the situation it represents.

In other words, if a high probability event has a low degree of membership, or a very low probability event has a high compatibility, then further studies should be carried out in order to clarify the results. An ideal situation for building confidence in model predictions would be a match between probability and compatibility. For example, in the example, a $K_d = 499$ has a very low degree of membership to the set, "Low k_d" in Table 1, therefore, it should also have a very low probability of occurrence.

CONCLUSIONS

The probabilistic approach assumes independent and random events, which means the events are equally possible to occur. However, this is not always the case for empirical data. Traditionally, because of lack of adequate information, experts build the probabilistic distribution functions based on their professional experiences.

In other words, there is no complete evidence for characterization of data and their assignment to a class. Consequently, the data may not be interpreted correctly. i.e., a high probable event may have a low degree of possibility, or vice versa.

This paper suggests that the possibility analysis be used as a complement to probabilistic analysis. The degree of membership can be interpreted as a degree of compatibility, or assurance, of a particular data to the situation being studied. In this sense, a reasonable degree of compatibility, or membership, could be interpreted as reasonable assurance that the results are being correctly interpreted.

REFERENCES

1. IAEA – International Atomic Energy Agency. *Evaluating the Reliability of Predictions Made using Environmental Transfer Models*. Vienna: 1989 (a). 106p. (Safety Series, 100)

2. SKB_ Swedish Nuclear Fuel and Waste Management Co (1999); *Long-term safety for KBS-3repositories at Forsmark and Laxemar – a first evaluation Main Report of the SR-Can project*; TR-06-09, Stockholm

3. Ross T, (2005). *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.

5. Donald S. *Development of Empirical Possibility Distributions in Risk Analysis*; Dissertation; The University of New Mexico; Albuquerque, 2003

6. Martz H. F. and & Waller R. A. (1982). *Bayesian Reliability Analysis*, John Willey & Sins, New York, 745 pp.

7. Lemos, F. L. and Sullivan, T. M., *Kd Determination Using Fuzzy Sets Theory*, International Conference on Probabilistic Safety Assessment and Management, New York, 1998