Uncertainty and Risk: Mathematical Concepts and some Geological Applications

János Fodor[1] and György Bárdossy[2]

[1]Department of Biomathematics and Informatics, Szent István University, Budapest, Hungary. E-mail: Fodor.Janos@aotk.szie.hu

[2]Hungarian Academy of Sciences, Budapest, Hungary
E-mail: h4750bar@helka.iif.hu

Abstract: The aim of this paper is to review traditional and recent approaches for handling uncertainty in geological investigations. First we classify main types of uncertainties occuring in this framework. Then we outline traditional mathematical methods frequently applied in geology. We point out their limitations as well. In the second part of the paper we present new methods that are more suitable for handling some types of uncertainty such as vagueness and incomplete information. Among these methods, fuzzy set theory seems to be one of the most efficient for geological purposes. The problems of uncertainty in risk analysis are shortly discussed. The authors present their experiences in the geological application of the reviewed methods. Finally, the main test calculations, preformed by the authors are listed.

1 Introduction

Uncertainty is involved in many real applications. Apart from the definition, it is an important matter of decision whether one would like to model uncertainty in an explicit way. If not then it is typical to use a deterministic model as an approximation of the uncertain phenomenon. In addition, the worst case analysis is an approach that notices the presence of uncertainty without modeling it explicitly. It works with upper (or lower) bounds, trying to ensure that no larger (or smaller) value of an uncertain parameter may occur in the given system. In this paper we shall focus on those cases in which the decision is to model uncertainty explicitly.

Before the mid-sixties, probability theory and statistics were the only tools to model uncertainty. Since that time, additional alternative theories have been suggested for uncertainty modeling. However, we share the surprise of Zimmermann (2000) who could not find any general definition of uncertainty. In this paper we also follow him in the sense that „we shall focus on the human-related, subjective interpretation of uncertainty which depends on the quantity and quality of information which is available to a human being about a system or its behavior that the human being wants to describe, predict or prescribe" (Zimmermann, 2000). The choice of an appropriate uncertainty calculus should depend on the causes of uncertainty, quantity and quality of information available, type of information processing required by the respective uncertainty calculus.

The paper is organized as follows. In the next section we summarize some of the fundamental concepts, especially those ones related to uncertainty. Then we briefly recall traditional methods for handling uncertainty, together with their shortcomings. After that a summary of some promising alternative uncertainty calculi is presented. Before the concluding remarks we touch upon our experiences in geological applications of these new methods, performed by us in diverse test calculations. The interested reader should consider the recent book by Bárdossy and Fodor (2004) for more details both on the theory and on the applications.

2 Basic Concepts

Geological investigations are characterized by particularly high uncertainties. It is therefore of paramount importance to understand the concept of uncertainty in general and its application to geology. In our opinion, uncertainty is a general term expressing lack of certainty and of precision in describing a geological object, a feature or a process. Highly competent scientists distinguish the following types of uncertainties (Dubois and Prade 2000, Zimmermann 2000):

1. *Imprecision* or inaccuracy, expressing the deviation of measurements from the true value.
2. *Vagueness* or *ambiguity*, the uncertainty of non-measurable objects or properties.
3. *Incompleteness,* the uncertainty due to incomplete information.
4. *Conflicting evidence,* the uncertainty arising from contradicting evidences present in the studied system.
5. *Presumption* or *belief,* when all available information is subjective. The well known „expert's opinion" belongs to this group.

The above classification is valid for geological investigations as well. However, the complexity of most geological problems requires a more detailed and specific classification. The one elaborated by us distinguishes two main groups (Bárdossy and Fodor 2004):

A. Uncertainties due to natural variability (called also *aleatory* uncertainties).
B. Uncertainties due to human shortcomings and incomplete knowledge (called also *epistemic* uncertainties).

Natural variability is a property of nature, existing independently of us. It can be described and quantified by mathematical methods, but not diminished by further empirical studies (although it may be better characterized). That's why it is also called irreducible uncertainty. The term aleatory uncertainty intends to emphasize its relation to the randomness in gambling and games of chance.

On the other hand, the second kind of uncertainty is the incertitude that comes from scientific ignorance, measurement uncertainty, inobservability, censoring, or other lack of knowledge. In contrast with aleatory uncertainty, epistemic uncertainty is sometimes called reducible uncertainty since, at least in principle, it can generally be reduced by additional empirical effort. This fundamental difference has been often neglected in geological investigations.

The second group has been divided by us for both computational and geological reasons into the following subgroups:

B1. Uncertainties in the phase of preparing the input data.
B2. Uncertainties in the phase of geological evaluation.

They are generally evaluated separately for each variable but in some cases the calculation of the total uncertainty is required.

The sources and characteristics of these uncertainties have been discussed in detail in the book of Bárdossy and Fodor (2004).

3 Traditional Approaches for Handling Uncertainties

As we already mentioned in the Introduction, there exist deterministic methods that notice the presence of uncertainty without modeling it explicitly. On the other hand, probability theory and statistics (may be called stochastic methods) allow for a certain type of uncertainty in the given system. In this theory probability is considered as a measure of uncertainty. (We suppose that the reader is familiar with the main concepts and methods of the probability theory.) Two main approaches can be distinguished. The frequentist approach generally applied in geology requires repeated identical experiments to calculate the main characteristics (such as expected value, standard deviation, and/or the probability distribution) of the studied variable. However this can be fulfilled only in few cases because of technical and financial reasons. For example, how to repeat borehole sets? This represents a significant limitation for the application of the frequentist approach in geological investigations. Neglecting this requirement leads to more or less important bias of the results. Within the stochastic framework, there exists a second approach as well, the Bayesian one. This approach works with subjective probabilities, depending not only on the phenomenon itself, but also on the state of knowledge and on the belief of the researcher. Bayesian probabilities can be applied also to unrepeatable phenomena.

They change with time as new pieces of information are acquired by the investigations. This concept is based on the famous Bayes' Theorem.

In this context prior probabilities refer to the event prior to updating by new pieces of information and posterior probabilities to the same event after updating by new pieces of information acquired by new investigations. Geological investigations are characterized by new pieces of information obtained step-by-step during the investigation. This is why they are particularly suitable for the application of the Bayesian approach. Unfortunately, relatively little has been done so far for the geological application of this concept.

For several decades, the methods based on probability theory have been applied successfully for the evaluation of geological investigations. However, some important limitations and difficulties appeared particularly for the evaluation of geological uncertainties, as listed below:

1. The input data of the deterministic and stochastic approaches are single valued (real numbers), called also „crisp" numbers. Only natural variability can be expressed by this type of data and not the uncertainties due to human shortcomings and incomplete knowledge.
2. The basic axioms of the probability theory, called Kolmogorov axioms, deal only with mutually exclusive subsets. However in geology disjoint subsets are rare, transitions are much more frequent.
3. Uncertain propositions and statements cannot be evaluated in terms of repeated experiments.
4. Semi quantitative and qualitative (linguistic) variables are not suitable for most statistical evaluations.

For the reasons listed above, we consider that methods based on probability theory are mathematically correct, but they do not offer optimal solutions for several geological problems, particularly for the evaluation of uncertainties.

4 Review of Uncertainty Oriented Mathematical Methods

New mathematical methods have been developed since the sixties with the aim to efficiently handle the uncertainties; especially, to represent what is known about real valued but uncertain quantities. For situations in which the uncertainty is purely aleatory, probabilistic and statistical methods are usually preferred. When the gaps in our knowledge involve both aleatory and epistemic uncertainty, several competing approaches have been suggested. Let us stress that the frequently used term of „uncertainty analysis" has been applied so far only for probabilistic evaluations. Thus it could not offer the entire evaluation of all types of uncertainties. The most important new methods consist of interval analysis

(Moore 1979), fuzzy set theory (Zadeh 1965), neuro-fuzzy systems (Fullér 2000), possibility theory (Zadeh 1978, Dubois and Prade 1988), probability bounds theory (Walley and Fine 1982, Williamson and Downs 1990), and different hybrid methods (Cooper, Ferson, and Ginzburg 1996, Guyonnet et al. 2003). The common feature of these methods is that instead of single valued crisp input data they apply different new types of data expressing the amount of uncertainty related to the given input data. The basic properties of these methods and their advantages as well as their limitations are shortly outlined below.

Interval analysis (Moore 1979) replaces crisp numbers by uncertainty intervals (Figure 1). The topic has become even more important with the advent of computers (Dubois et al., 2000): the motivation is “the quest for rigor in numerical computation on machines”. It is assumed that the true value is somewhere within the interval. Interval analysis lacks gradations and is the simplest method to express uncertainty through arithmetic calculations. The method guarantees that the true value will always remain within the interval, but this goal is achieved at cost of precision. During the calculations the intervals become wider and wider and the final results become too conservative.

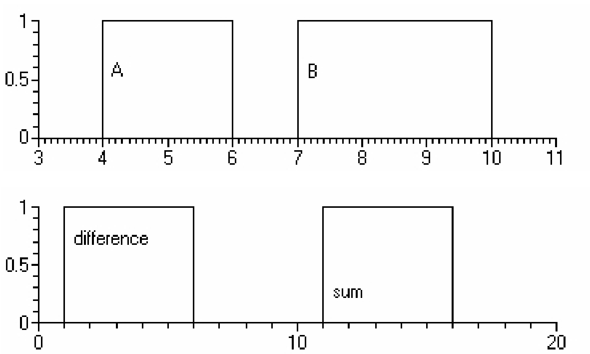

Figure 1: Two intervals and their sum (A+B) and difference (B-A)

Possibility theory, a generalization of interval analysis, provides a suitable model for the quantification of uncertainty by means of the possibility of an event (Zadeh 1978, Dubois and Prade, 1988). The theory acknowledges that not all types of uncertainty can be handled by probability distributions. Instead, it uses membership functions to represent non quantified uncertainty. The membership value of a number, varying between zero and one, expresses the plausibility of the occurrence of that number. The theory has been applied successfully in biology, health and medicine (Ferson and Ginzburg 1996, Ferson et al. 1999) and in different branches of industry and economy (Bardossy A. and Duckstein 1995, Fodor and Roubens 1994).

The related fuzzy set theory expresses uncertainty very often by the use of fuzzy numbers. They represent estimates of uncertainty at different levels of possibility (or membership degree). Membership functions of fuzzy numbers are by definition unimodal and they have to reach at least in one point the possibility level one, that is, the full possibility. In geology mainly trapezoidal and triangular fuzzy numbers are applied. They can be both symmetrical and asymmetrical. The smallest and the largest possible values of the given variable represent the lower and the upper bounds of the support of the fuzzy number. All values of the variable must be within these boundaries. The values reaching the possibility level one are considered as the most possible estimates of the given variable, and this interval is called the core of the fuzzy number. Fuzzy numbers are generalizations of traditional real numbers, as the latter ones can be regarded as fuzzy numbers with a single point support.

All usual arithmetic calculations can be carried out with fuzzy numbers. In contrast with probability, one of their great advantages is that they do not require the knowledge of the correlations among the variables and the type of their probability distribution (Takács and Várkonyi-Kóczy 1999a, 1999b). For the sake of numerical comparisons and ranking, fuzzy numbers can be reconverted into crisp numbers. This calculation is called defuzzification. But the main advantage of the fuzzy method is that prior geological experience can be incorporated into the construction of fuzzy numbers. This goal can be achieved by joint constructing of the fuzzy numbers by geologists and mathematicians. The method allows the appropriate evaluation of semi-quantitative and qualitative input data as well. The frequent transitions of the geological populations, as mentioned before, can be also represented by fuzzy numbers (Figure 2). Cagnoli (1998) showed the application of the fuzzy set theory in the study of volcanic rocks. In the last years it found a broad application in the geographical information systems as well (Altman 1994, Macmillan 1995, and Unwin 1995).

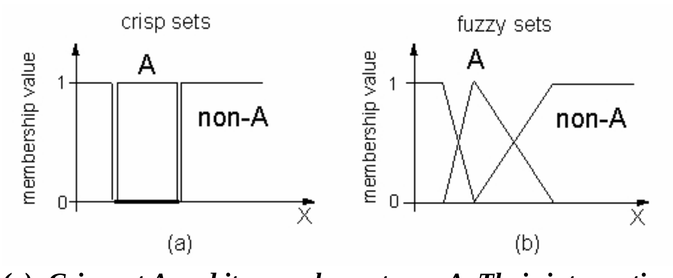

Figure 2 (a): Crisp set A and its complement non-A. Their intersection is empty, and their union is the set of all elements of the universe.
(b): Fuzzy set A and its complement non-A. They overlap.

The way of constructing fuzzy numbers raises the problem of their robustness. Imagine that several well trained and experienced experts are asked to construct fuzzy numbers, based on the same crisp data. It is certain that the resulting fuzzy numbers will not be exactly identical. However, the differences are expected to be rather small. Luckily all the mathematical operations one has to carry out with fuzzy numbers are stable, that is, small changes in the input data yield only small changes in the results. As a consequence, the final results are not sensitive to small differences in the initial fuzzy numbers.

The probability bounds theory (Walley and Fine 1982, Williamson and Downs 1990) is a combination of probability theory with interval analysis, by representing epistemic uncertainty within the context of probability theory. The final idea is that one can work with bounds on probability for this purpose. It expresses uncertainty by two cumulative probability distributions, called "probability boxes", or "p-boxes" for short. The area between the two curves represents the extent of uncertainty of the given variable, as it can be seen in Figure 3. Probability bounds are considered as a generalization of crisp numbers, intervals and probability distributions. Williamson and Downs 1990 also described algorithms to compute arithmetic operations (addition, subtraction, multiplication and division) on pairs of p-boxes. These operations generalize the notion of convolution between probability distributions. The great advantage of this method is that it can apply different probability distributions, e.g. normal, lognormal, exponential etc. and correlations for the variables to be studied. But the method works also without making any prior assumptions. The probability bounds get narrower with more empirical information about the given geological object. Its disadvantages are the more complicated calculations to be carried out. Nevertheless it seems for us to be a highly efficient approach in the case of safety assessments, when prior information is abundant. A detailed overview of p-boxes and their links to other representation of uncertainties can be found in Ferson et al. 2002.

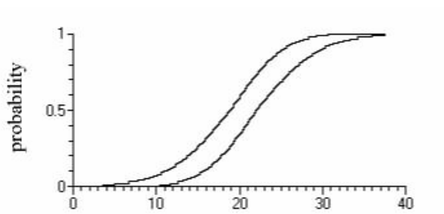

Figure 3: Probability bounds

The method of hybrid arithmetic (Cooper et al. 1996, Guyonnet et al. 2003) combines probability distributions with intervals, fuzzy numbers and probability bounds in different manner. These methods allow the use of all kinds of numbers, this being their greatest advantage. These are the newest among the methods of uncertainty analysis and there are very few publications on its application.

5 Uncertainty of Risk Analysis in Geology

Risk is a common term in science, economy and industry. According to the definition of the Society of Risk Analysis, *risk is the potential for realization of unwanted consequences of a decision or an action*. *Risk analysis* is defined by the same society as „the process of quantification of the probabilities and expected consequences of risks" (2001). Risk analysis has been applied to several problems in geology, such as mineral exploration, mining projects, landslides, floods, volcanic and earthquake hazards. The safety assessments of toxic and radioactive waste repositories represent particularly important applications of risk analysis. All these calculations have been carried out so far by traditional deterministic and probabi-listic methods (Bonano and Cranwell 1988, Craig 1988, Hunter and Mann 1992). At our knowledge, no uncertainty oriented methods have been applied for risks of geological problems so far.

The basic requirement of risk analysis is to exclude the possibility of under-estimation of risk at the given conditions. With the traditional methods measures of central tendency (mean, median etc.) are produced. However, experience showed that not these measures, but the tail of the distributions are of paramount importance, as they represent risks of low proba-bility, but of severe consequences. *Dependency bounds analysis*, suggested by Ferson (1996), seems to assure sufficiently secure estimates of these tail-probabailities.

The methods of interval analysis and fuzzy arithmetic have been first applied to risk analysis by Ferson and Kuhn (1992) for ecological problems. Our aim is to apply these methods for the calculation of geological risks as well.

6 Possible Applications in Geology

As outlined in the foregoing sections, membership functions can express the degree of uncertainty of any input data collected in the course of a geologic investigation. But Dubois and Prade (2000) pointed out that they can express the degree of similarity and the degree of preference as well. In our opinion these two latter meanings can be applied also to geological problems. In mathematical sense the degree of similarity expresses the proximity of x to the chosen prototype elements of A. In geology a prototype can be a rock, a lithologic unit, a mineral deposit, a fossil species etc. Note that probability theory, as defined by the axioms of Kolmogorov, does not admit transitions, only full or no membership (see the characteristic functions), that is, only disjoint sets of data. However, in geology the transitions are very frequent. When we apply a probabilistic evaluation, we are obliged to draw sharp boundaries and to cut transitional zones into different pieces. Obviously this is a distortion of the natural reality. Membership functions can be applied with success to solve this problem, as presented in the next section.

The degree of preference refers to a set of more or less preferred objects and $\mu_A(x)$ represents the degree of preference in favor of the object x. This case may occur in the exploration of mineral deposits, or of groundwater resources, when we have to choose among several potential regions. The degree of our preference for a given region can be represented by a membership function. The choice between suitable locations for commercial, toxic or radioactive waste disposal can also be represented by membership functions, instead of the traditional ranking of crisp numbers. Note that when using the degree of preference, the choice of the preferred object is ours.

7 Some Real Applications of Fuzzy Arithmetic

In the last four years a number of test calculations were carried out by us – applying the fuzzy set theory – on different geological problems. Parts of the results were published in separate articles or in the book Bárdossy and Fodor, 2004. The main topics of our calculations have been:

1. Estimation of the resources of solid mineral deposits.
2. Quantitative mineralogic phase analysis of rocks and ores by X-ray diffractometry and by thermal analysis (separately).
3. Safety assessment of radioactive waste disposal.
4. Application to paleontological biometry.
5. Measurements of ground-water transmissivity in boreholes.
6. Evaluation of geochemical transitions in bauxite deposits.

8 Conclusions

In geology, uncertainty has long been considered a removable adverse circumstance that should gradually disappear with the overall development of the Earth-Sciences. However, one must recognize that a part of this uncertainty is an inherent feature of Nature. Therefore, understanding and appropriate handling of uncertainties should be part of all future geological investigations.

Traditional mathematical methods – deterministic and probabilistic – applied so far in geological investigations are mathematically correct, but by far not optimal for the treatment of all kinds of uncertainties.

New mathematical methods summarized in this paper are suitable to evaluate in a mathematically correct way semi-quantitative and qualitative („linguistic") input data and to determine the uncertainties and errors connected with them.

It is important to emphasize that the traditional and the new uncertainty oriented methods are context-dependent, and complete each other in geological investigations.

A thorough study of the geological objects and processes is indispensable for any mathematical evaluation in geology. Without that even the most sophisticated method becomes an empty formalism.

References

[1] Altman, D., 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis: International Journal of Geographical Information Systems, v. 8, p. 271-289.

[2] Bardossy, A., and Duckstein, L., 1995, Fuzzy rule based modeling with applications geophysical, biological and engineering systems: New York, CRC Press, 232 p.

[3] Bárdossy, Gy., and Fodor, J., 2004, Evaluation of uncertainties and risks in Geology: Berlin, Heidelberg, London, New York, Springer Verlag, 221 p.

[4] Cagnoli, B., 1998, Fuzzy logic in vulcanology: Episodes, v. 21, p. 94-96. Cooper, J. A., Ferson, S., and Ginzburg, L. R., 1996, Hybrid processing of stochastic and subjective uncertainty data: Risk Analysis, v. 16, p. 785-791.

[5] Dubois, D. and Prade, H., 1988, Possibility theory: An approach to computerized processing of uncertainty: New York, Plenum Press, 263 p.

[6] Dubois, D., Kerre, E., Mesiar, R., and Prade, H., 2000, Fuzzy interval analysis, in Dubois, D., and Prade, H., eds., Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series 7: Boston, London, Dordrecht, Kluwer Academic Publishers, p. 483–581.

[7] Ferson, S., Root, W., and Kuhn, R., 1999, RAMAS Risk Calc: Risk assessment with uncertain Numbers: Setauket, New York, Applied Biomathematics, 184 p.

[8] Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., and Sentz, K., 2002, Constructing probability boxes and Dempster-Shafer structures: SAND2002-

0835 Technical Report, Sandia National Laboratories, Albuquerque, New Mexico, p. 143.

[9] Fodor, J., and Roubens, M., 1994, Fuzzy preference modeling and multicriteria decision support: Dordrecht, Kluwer Academic Publishers, 272 p.

[10] Fullér, R., 2000, Introduction to neuro-fuzzy systems: Heidelberg, Physica Verlag, 289 p.

[11] Guyonnet, D., Bourgine, B., Dubois, D., Fargier, H., Côme, B., and Chilès, J.-P., 2003, A hybrid approach for addressing uncertainty in risk assessments: Journal of Environmental Engineering, v. 129, p. 68-78.

[12] Macmillan, W., 1995, Modeling: fuzziness revisited: Progress in Human Geography, v. 19, p. 404-413.

[13] Moore, R. E., 1979, Methods and applications of interval analysis: Philadelphia, SIAM Studies on Applied Mathematics, Vol.2.

[14] Takács, O., and Várkonyi-Kóczy, A.R., 1999a, Fuzzy handling of uncertainty in nonlinear systems, in De Baets, B., Fodor, J. and Kóczy, L.T., eds., Proceedings of EUROFUSE – SIC ’99, Budapest, p. 22-27.

[15] Takács, O., and Várkonyi-Kóczy, A.R., 1999b, Information processing based on mixed-classical and fuzzy-data models, in IEEE International Workshop on Intelligent Signal Processing, Budapest, p. 23-27.

[16] Unwin, D., 1995, Geographical information systems and the problem of error and uncertainty: Progress in Human Geography, v. 19, p. 549-558.

[17] Walley, P., and Fine, T.L., 1982, Towards a frequentist theory of upper and lower probability: Annals of Statistics, v. 10, p. 741-761.

[18] Williamson, R.C., and Downs, T., 1990, Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds: International Journal of Approximate Reasoning, v. 4, p. 89–158.

[19] Zimmermann, H.-J., 2000, An application-oriented view of modeling uncertainty: European Journal of Operational Research v. 122, p. 190–198.

[20] Zadeh, L. A., 1965, Fuzzy sets: Information and Control v. 8, p. 338–353.

[21] Zadeh, L. A., 1978, Fuzzy sets as a basis for a theory of possibility: Fuzzy Sets and Systems, v. 1, p. 3-28.