Fuzzy alpha-cut vs. Monte Carlo techniques in assessing uncertainty in model parameters

A. J. Abebe
IHE Delft, P.O.Box 3015, 2601 DA Delft, the Netherlands, Email: abebe@ihe.nl

V. Guinot
IHE Delft, Delft, the Netherlands

D. P. Solomatine
IHE Delft, Delft, the Netherlands

ABSTRACT: This paper presents a comparison of two methods of analysis of uncertainty arising from uncertain model parameters. The first method is Monte Carlo simulation that treats parameters as random variables bound to a given probabilistic distribution and evaluates the distribution of the resulting output. The second one is fuzzy logic-based alpha-cut analysis in which uncertain parameters are treated as fuzzy numbers with given membership functions. Both techniques are tested on a model of ground water contaminant transport where the decay rate of the contaminant is considered to be uncertain. In order to provide a basis for comparison between these two approaches, the shapes of the membership function used in the fuzzy alpha-cut method is the same as the shape of the probability density function used in the Monte Carlo simulations. The analysis indicates that both methods give similar results provided that the correlation distance of the decay rate is assumed to be infinite. However, particular details of the analysis steps, computation time and representation of uncertainty are different, which may lead to the choice of one method or another depending on the nature of the problem.

1 INTRODUCTION

1.1 Problem position

Parameters of physically-based models bare some meaning and can be determined using in situ measurements, calibration, expert judgement, etc. However, the values of these parameters may be subject to uncertainty, due to the lack of measurement points, over–calibration or imprecise expert judgement. Uncertainty in model parameters is one of the main causes of uncertainty in model outputs.

Model uncertainty arising from parameters can be analyzed using several techniques. In the present paper, methods based on (1) Monte-Carlo simulation (MCS) and (2) fuzzy logic analysis (alpha-cuts) are considered.

The MCS technique treats any uncertain parameter as random variable that obeys a given probabilistic distribution. Any model output is then a random variable. This technique is widely used for analyzing probabilistic uncertainty.

The fuzzy alpha-cut analysis is based on fuzzy logic and fuzzy set theory (introduced by Zadeh, 1965) which is widely used in representing uncertain knowledge. Uncertain model parameters can be treated as fuzzy numbers that can be manipulated by specially designed operators.

Fuzzy set approach has been applied recently in various fields, including decision-making, control and modelling. In hydrology, it has been used for example for solving the regression problems (Bardossy et al., 1990) and for restoration of the missing rainfall data at neighbouring measuring stations (Abebe and Solomatine 2000). An example of its use in modelling with imprecise parameters can be found in Dou et al. (1995).

The present paper aims to compare the features, advantages and drawbacks of these two techniques when applied to the analysis of uncertainty in ground water contaminant transport modelling. In this study, the decay rate of the contaminant was considered to be the uncertain parameter.

1.2 Transport Model

In presence of advection, hydrodynamic dispersion and first–order kinetics degradation in the liquid phase only, the two-–dimensional equation that describes contaminant transport in an aquifer is:

$$\begin{aligned}&\frac{\partial}{\partial t}(h\theta C)+\nabla\mathbf{F}=RC_s-\lambda h\theta C\\&\mathbf{F}=h\mathbf{v}C+h\boldsymbol{\delta}\mathbf{v}\nabla\mathbf{C}\end{aligned}\qquad(1)$$

where C (kg.m^{-3}) is the contaminant concentration in the aquifer, C_s (kg.m^{-3}) is the contaminant concentration of the recharge water, h (m) is the aquifer thickness, $\mathbf{v}$ (m.s^{-1}) is the Darcy velocity vector, R (m.s^{-1}) is the recharge rate, $\boldsymbol{\delta}$ (m) is the dispersivity tensor, λ (s^{-1}) is the degradation rate, θ (dimensionless) is the aquifer porosity and ∇ represents the divergence operator when applied to a vector and the gradient when applied to a scalar. Usually, the typical order of magnitude for the dispersivity in non–fractured aquifers is 10 metres or smaller (Gelhar et al., 1992). Since the distances D (m) dealt with in the present paper is one or several kilometres, the Peclet number Pe = D/δ is much greater than unity, which indicates that the predominant phenomenon in this type of problem is advection. The dispersive terms in Eq. (1) were therefore neglected.

Eq. (1) was solved numerically using a finite volume, Godunov–type method (Guinot, 2000), adapted for the two–dimensional computation of passive contaminant transport in two dimensions.

2 METHODOLOGY

2.1 Monte-Carlo Simulation (MCS)

The principle of MCS is illustrated by Figure 1. As mentioned previously, any input parameter P subject to uncertainty is considered as a random variable $\mathbf{P}$. A number of realisations P_i of $\mathbf{P}$ is generated and the deterministic, physically–based model is run for each of them, hence producing an output R_i. The set of outputs R_i represents the set of realisations of the random variable $\mathbf{R}$. The statistical properties of $\mathbf{R}$ are therefore computed from the realisations R_i .

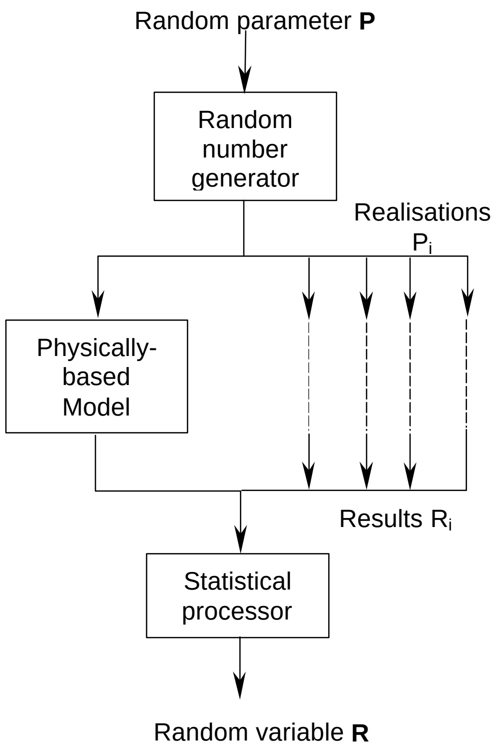

Figure 1. Sketch of the Monte-Carlo Simulation (MCS) method

2.2 Fuzzy alpha-cut (FAC) technique

This technique uses fuzzy set theory to represent uncertainty or imprecision in the parameter(s). Uncertain parameters are considered to be fuzzy numbers with some membership functions. Figure 2 shows a parameter P represented as a triangular fuzzy number with support of A_0. The wider the support of the membership function, the higher the uncertainty. The fuzzy set that contains all elements with a membership of $\alpha \in [0,1]$ and above is called the a-*cut* of the membership function. At a resolution level of α, it will have support of A_α. The higher the value of α, the higher the confidence in the parameter (Li & Vincent, 1995).

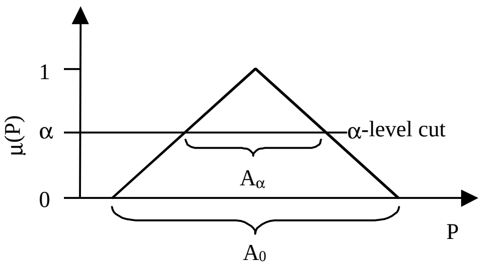

Figure 2. Fuzzy number, its support and α-cut

The method is based on the *extension principle,* which implies that functional relationships can be extended to involve fuzzy arguments and can be used to map the dependent variable as a fuzzy set. In simple arithmetic operations, this principle can be used analytically. However, in most practical modeling applications, relationships involve partial differential equations and other complex structures that make analytical application of the principle difficult. Therefore, interval arithmetic is used to carry out the analysis.

The membership function is cut horizontally at a finite number of α-levels between 0 and 1. For each α-level of the parameter, the model is run to determine the minimum and maximum possible values of the output. This information is then directly used to construct the corresponding fuzziness (membership function) of the output which is used as a measure of uncertainty.

If the output is monotonic with respect to the dependent fuzzy variable/s, the process is rather simple since only two simulations will be enough for each α-level (one for each boundary). Otherwise, optimization routines have to be carried out to determine the minimum and maximum values of the output for each α-level.

This approach was used, for example by Schulz & Huwe (1997, 1999) to model soil water pressure in the unsaturated zone subject to imprecise boundary conditions and hydraulic properties.

3 CASE STUDY

3.1 Test Site

The two approaches described above were applied to contaminant transport modelling in groundwater over the Vannetin basin in France. This catchment is a sub-catchment of the Grand Morin basin, which contributes to the water supply of the urban area of Paris (Fig. 3). Water is mostly taken from the main rivers that flow east of Paris. Since these rivers drain a wide aquifer system, the quality of drinking water depends directly on that of water in the aquifer system.

This basin is used extensively for agricultural practices. The regular application of pesticides over the past 35 years has lead to an increase of their concentrations in the aquifers and consequently in the river systems. Therefore, the modelling of contaminants in groundwater is of outmost importance for the management of surface water quality. The Vannetin basin was chosen as a test site (see Figure 3).

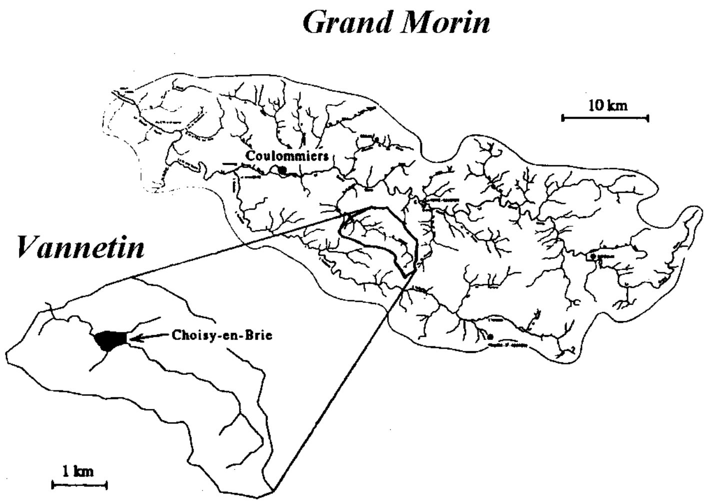

Figure 3. Geography of the test site

The behaviour of the aquifer flow in the neighbourhood of the aquifer wells in the town of Choisy-en-Brie has already been investigated in a previous study (Guinot 1995). This investigation was carried out with the aim of defining protection perimeters for the groundwater wells against contamination by pesticides. A physically-based model of the Vannetin catchment had been built using the MIKE SHE (Abbott et al. 1986a,b) modelling system. The model predicted a quasi-steady flow regime in the Champigny aquifer, where the pumping wells are located. This is due to the fact that the Champigny aquifer is separated from the ground surface by a top aquifer and by a semi-permeable marn layer. Figure 4 shows the computed water table.

The study consisted of studying the effect of a pointwise contamination of the aquifer by Atrazine. Atrazine is a pesticide used for protecting maize. The injection point is indicated by a square on the figure.

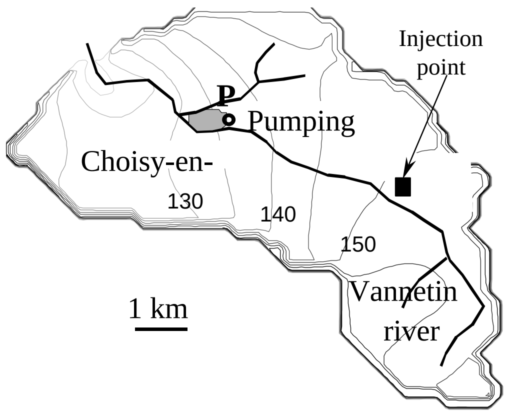

Figure 4. Computed aquifer water table. Heads are given in metres.

3.2 *Model Input Data*

Although the degradation rates of most pesticides in the unsaturated zone are quite well–known, very little is known about their behaviour in the saturated zone. Their half–life is generally estimated to be two or three years. On the basis of previous studies carried out on uncertainty assessment in the unsaturated zone (Carsel et al., 1988), the assumption was made of a triangular probability density function for the degradation rate. The mean value for the degradation rate was taken equal to 2.2×10^{-8} s^{1}, which corresponds to a half–life of 2 years. The lower bound of the probability density function is 1.64×10^{-8} s^{-1} and the upper bound is 2.7×10^{-8} s^{-1} (see figure 5).

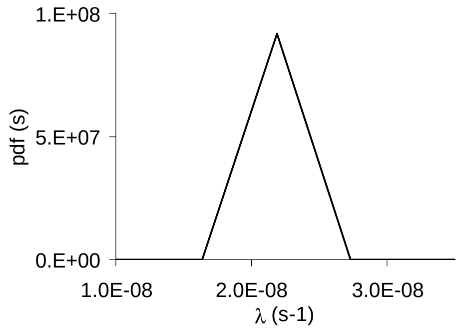

Figure 5. Probability density function for the degradation rate.

From this theoretical distribution, 500 values were generated using a random number generator. The histogram of the generated sample is shown on Figure 6.

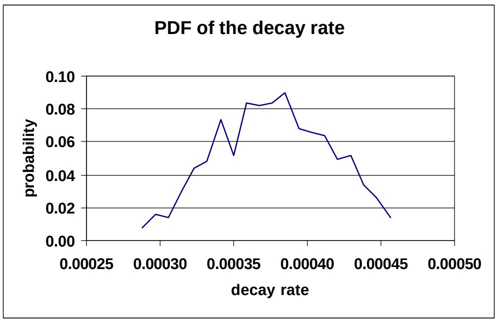

Figure 6: Histogram for the degradation rate derived from the sample obtained with a random number generator.

Each of these 500 values was taken as an input for the transport model described by Eq. (1). In these simulations, the correlation distance of the degradation rate was assumed to be infinite. This means that the value of l was assumed to be the same everywhere. This assumption was made for the purpose of unbiased comparison with the fuzzy approach, that assumes the fuzzy variable (i.e. its membership function) to be the same at every point in space. As shown by previous studies, the assumption of infinite correlation leads to overestimating the effects of uncertainty (Guinot, 1995).

The membership function for the degradation rate that was used for the fuzzy technique is shown on figure 7.

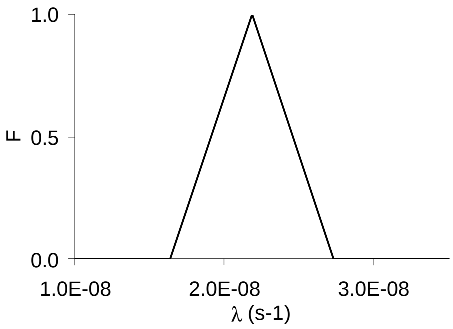

Figure 7: Membership function of the degradation rate.

3.3 Results

Two analyses were carried out: a spatial analysis, for which a measure of uncertainty was devised, and a pointwise analysis, where the density probability function (for the MCS technique) and the membership function (for the FAC technique) of the concentration were analysed at a given point.

3.3.1 Spatial analysis

In order to evaluate the spatial distribution of uncertainty, it is necessary to establish a measure of uncertainty for the two methods. For the MCS technique, such a measure is given by the ratio of the standard deviation to the mean concentration of the solute at each grid cell. The measure of uncertainty used for the FAC technique is the ratio of the 0.1-level support to the value of the concentration for which the membership function is equal to 1 (see figure 8).

Figure 9 shows the measure of uncertainty obtained with the MCS, whereas Fig. 10 shows the results obtained with the FAC technique. The results indicate that the relative width of the fuzzy membership function of the output and the ratio of the standard deviation to the mean of the concentration is an increasing function of the distance to the injection point.

The magnitudes of the measures of uncertainty in both cases are different but they convey the same information about the distribution of uncertainty over the physical domain.

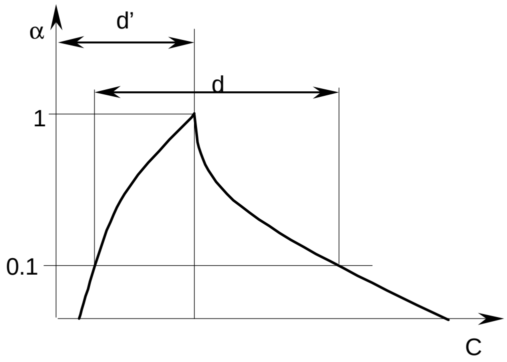

Figure 8 Measure of uncertainty for the FAC technique : the measure of uncertainty is given by the ratio d/d'.

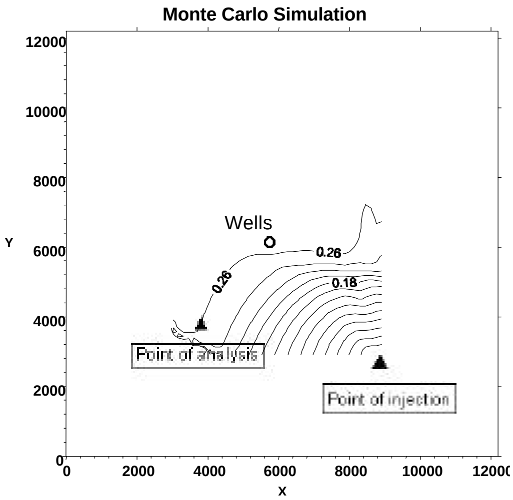

Figure 9. Measure of uncertainty for the MCS technique. Distances are in metres.

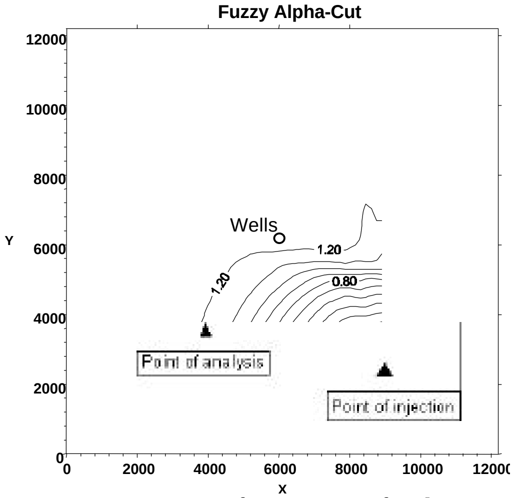

Figure 10: Measure of uncertainty for the FAC technique. Distances are in metres.

3.3.2 *Pointwise analysis*

Figure 11 shows that the uncertainty at the selected point of analysis. The normalized frequency distribution of the concentration obtained from the MCS is plotted in the same set of axes as the fuzzy number representing the concentration obtained from the FAC method.

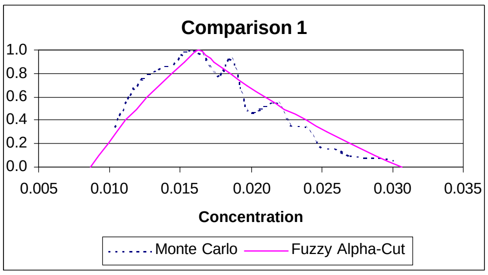

Figure 11: Normalized PDF and fuzzy membership function of the output at the selected point of analysis.

In Figure 12, the distribution function and the normalized-integrated fuzzy number are plotted.

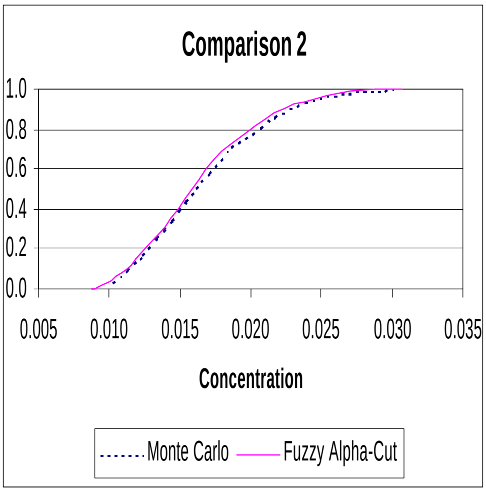

Figure 12: Cumulative distribution of the PDF and normalized and integrated membership function of the output at the selected point of analysis.

The analysis at the selected grid point indicates that a symmetrical uncertainty in the decay rate results in unsymmetrical uncertainty in the solute concentration. This is a reflection of the effect of the equations used by the model. The width of the output membership function (fuzzy number) is the indication of the sensitivity of the model to this parameter.

Both methods have shown comparable results when integrated membership function and cumulative frequency distribution are used for comparison. However, when probabilistic density functions are compared to membership functions, there is a clear indication of the variability in the case of MCS and consistency in FAC approach.

4 DISCUSSION AND CONCLUSION

In The MCS approach for sampling the input variables the number of 500 model runs has been chosen. This number could have been reduced, but only at the expense of the smoothness of the generated frequency distribution and the accuracy of the resulting probability density function. The FAC approach needed 20 model runs only. This indicates an obvious advantage of the FAC over the MCS approach for this particular case study.

However, it should be noted that the finite-difference solution of the advection–degradation equation is monotonic with respect to the degradation rate. If the results are not monotonic with respect to the uncertain parameter (e.g. in the case of diffusion problems), it would be necessary to use an optimisation technique to find the support of the fuzzy variable so as to reconstruct the fuzziness of the output. The approach would then become more time–consuming than it was in the present case. In the framework of the present study such problems have not been investigated yet.

The drawback of the Monte Carlo approach for the present application is its time–consuming character, due to the large number of simulations needed to achieve a satisfactory smoothness and accuracy of the results. On the other hand, it is possible to generate random fields for the uncertain parameter that take into account the spatial structure of these parameters (e.g. shape of the variogram, correlation distance, etc.).

The Fuzzy alpha–cut technique presents a strong alternative to the Monte-Carlo approach. It is faster for applications where the output is a monotonic function of the uncertain parameters, but the effect of non-monotonicity of outputs with respect to parameters still have to be investigated in terms of computational effort. Its drawback is that, so far, it is applicable only under the assumption of infinite correlation distances.

REFERENCES

Abbott, M.B., J.C. Bathurst, J.A. Cunge, P.E. O’Connell and J.Rasmussen. “An introduction to the European Hydrological System – Systeme Hydrologique Europeen, “SHE”, 1: History and philosophy of a physically-based, distributed modelling system”. *J. Hydrol.* 17: 45-59, 1986a.

Abbott, M.B., J.C. Bathurst, J.A. Cunge, P.E. O’Connell and J.Rasmussen. “An introduction to the European Hydrological System – Systeme Hydrologique Europeen, “SHE”, 2: Structure of a physically-based, distributed modelling system”. *J. Hydrol.* 17: 61-77, 1986b.

Abebe A.J., Solomatine D.P., Venneker R. Fuzzy logic in restoration of missing rainfall records. Hydrological Science J., 2000 (accepted for publication).

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

Carsel, R.F., R.S. Parrish, R.L. Jones, J.L. Hansen and R.L. Lamb. “Characterising the uncertainty of pesticide leaching in agricultural soils”. *Journal of Contaminant Hydrology*, 2 ; 111–124, 1988.

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

Gelhar, L.W., C. Welty and K.R. Rehfeldt, “A critical review of data on field–scale dispersion in aquifers”, *Water Resources Research*, 28 (**7**) 1955–1974, 1992.

Guinot, V. “Modélisation mécaniste du devenir des produits phytosanitaires dans l’environnement souterrain. Application à la protection des captages en aquifère”. *PhD Thesis*, University of Grenoble I, 1995.

Guinot, V. “The Discontinuous Profile Method (DPM) for simulating two–phase flow in pipes”, Submitted to *Int. J. Num. Meth. Fluids* (2000).

Li, H.X. and V.C. Yen, Fuzzy Sets and Fuzzy Decision-Making. CRC Press, 1995.

Schulz, K. and B. Huwe, “Uncertainty and sensitivity analysis of water transport modeling in a layered soil profile using fuzzy set theory,” *Journal of Hydroinformatics*, 1 (2) 127-138, IWA Publishing, 1999.

Schulz, K. and B. Huwe, “Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach,” *Journal of Hydrology,* 201 pp. 211-229. Elsevier, 1997.