An approach for integrating fuzzy rule system in hydrological modeling
F. Delclaux
UMR Géofluides-Bassin-Eau (CNRS-ORSTOM-UM2), ORSTOM Hydrologie, BP5045, 34032 Montpellier Cedex, France
Email: Francois.Delclaux@mpl.orstom.fr

Abstract

When uncertainty is important, the fuzzy set theory offers a framework for processing qualitative information. In such a model, variables are treated as fuzzy numbers and calculations are replaced by inference rules. Therefore, we built some tools in order to introduce qualitative knowledge in the field of hydrology.

Firstly, a C library has been developed, including basic fuzzy tools. Secondly, we built a fuzzy rule model software in which several algorithms are taken into account: defuzzification, aggregation rule, logical operator definition. The case-study concerns the determination of the ten year runoff coefficient in the Sahel area for which an evaluation method is based on three variables: the area and the slope of the basin, and an infiltrability index. In the classical ORSTOM method, the runoff coefficient is determined by interpolation between these variables on predefined curves. In the fuzzy approach, the runoff coefficient is computed by combination of fuzzy rules defined by the expert knowledge. The comparison shows a good agreement between the two methods.

Lastly, a specific fuzzy library has been implemented in a GIS in order to apply fuzzification and inference rules operations to raster maps.

1 Introduction

When considering environmental modeling, hypotheses often impose strict classifications, when qualitative knowledge would be valuable. Receiving an exact value from an imprecise variable is often resolved using a statistical approach where the variable is replaced by a
probability function. The fuzzy sets theory, based on the concept of gradual membership, provides a different non-probabilistic framework about this problem. From this point, many mathematical developments have been undertaken in various fields: decision making, process control, pattern recognition, image processing.

Hydrological modeling is faced with the information quality issue. Most often, the specialist’s knowledge is good, the observed data are accurate enough and the small-scale models provide good results. But the problem of transposing or extending results is not easy to solve. Then the lack of accuracy for extensive or ungauged areas can be balanced by the expert’s qualitative knowledge: it is the case for the determination of the runoff coefficient of Sahelian small basins in which, apart from some well-equipped basins, the data are scarce and the expertise is important. This case-study was a good opportunity for testing fuzzy concepts.

The first part of this paper presents the basic concepts of the fuzzy logic and the fuzzy rule system. In a second part, we briefly describe the fuzzy software components. Next, we present the ORSTOM method for the runoff coefficient determination and compare this method with the fuzzy one. The fourth part deals with the spatial fuzzification.

2 Fuzzy Sets and Fuzzy Logic

2.1 Fuzzy sets

Fuzzy sets, introduced by Zadeh [1] are based on a generalization of the classical set theory. Contrary to the ordinary set, the border of a fuzzy set is vague and is described by a gradual membership function which takes its values between 0 and 1: the membership value of an element x in a fuzzy set A expresses the degree to which x belongs to A. We define a fuzzy set as follows: given a set X, so-called universe set, a fuzzy subset A of X is a function \( \mu_A(x) : X \rightarrow [0,1] \). Some examples are reported on Figure 1.

For ordinary or crisp sets, \( \mu_A(x) \) is identical to the characteristic function which yields value 0 or 1 (see Figure 1-a). As for crisp sets, intersection and union are defined on fuzzy sets (cf. Zadeh [1]):

\[
\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \quad \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))
\] (1)

Some other functions called \( t \)-norm and \( t \)-conorm functions can be applied, as listed in Tong-Tong [2].
2.2 Fuzzy numbers

Fuzzy numbers are a particular case the fuzzy sets and are defined in the following way (see Dubois and Prade [3] for a complete definition): X is the real set \( \mathbb{R} \), and A is normal and convex (cf. Figure 1-d).

![Figure 1: Representation of membership functions. (a) ordinary set (b and c) fuzzy sets (d) fuzzy number](image)

Then, arithmetic operations can be induced on fuzzy numbers, all based on interval arithmetic according to the convexity assumption. It may be noted that the results of theses operations are generally fuzzy numbers while union or intersection generate fuzzy sets. The simplest way for describing fuzzy numbers usually is triangular (TFN) or trapezoidal (TRFN) functions. At last, we can define a connection between a fuzzy set and an equivalent crisp number thanks to the defuzzication operation for which several algorithms are available (Bardossy and Duckstein [4]): maximum value, center of gravity and median.

2.3 Fuzzy logic

Multivariate logic has been introduced in order to address complicated logical problems describing the human reasoning: therefore the number of the truth values is increased, but these values still are numerical in \([0,1]\), and their meaning depends on the author. Another generalized logic is the linguistic fuzzy logic in which the truth values are words or expressions of the natural language described by fuzzy sets. Thus a linguistic variable is defined by a triple \((X, U, T_x)\) where \(X\) is the variable name, \(U\) the universe the discourse of \(X\) and \(T_x\) the linguistic values of \(X\), each one modeled by a membership function. Figure 2 shows an example of such a variable. Then the extension of the modus ponens logical rule can be expressed in term of linguistic variable combination. The most often used fuzzy rule system is expressed as a set of rules: IF A \(\otimes\) IF B THEN C where A, B, C are expressions including fuzzy variable and predicate value, and \(\otimes\) is one of the operators AND, OR or XOR.
Solving a fuzzy rule system is described in the following steps:
- definition of the linguistic variables dedicated to each system variable;
- definition of the fuzzy rules according to the knowledge of the system.
Then, for each uplet of the input crisp values, the fuzzy inference is applied in the following way (Bardossy and Duckstein [4]):

- valid rule evaluation and calculation of the membership function values;
- for each valid rule, calculation of a degree of fulfillment (DOF) according to a pair (t-norm, t-conorm) functions;
- aggregation of DOF weighted rules
- generation of the final consequence fuzzy set and crisp output calculation (defuzzification).

3 Computer Framework

The software structure is based on a C library whose components handle basic fuzzy functions. The developments have been made according three axes (cf. Figure 3).
Firstly, a software, MODFLOU -MODélisation à base de règles FLOUes- has been developed, including a Tcl/Tk user interface. As shown on Figure 4, MODFLOU can simulate dynamic systems taking into account internal variables. Therefore, an extension of the rule system has been introduced according to the following semantic:

$$\text{IF } (X_i \text{ is } x_i) \text{ AND } (Z_k \text{ is } z_k) \text{ THEN } (Z_k \text{ is } z'_k), (Y_j \text{ is } y_j)$$ \quad (2)

where $X_i, Y_j$ are external and $Z_k$ internal variables, and $x_i, y_j, z_k, z'_k$, fuzzy predicate values. In the case where $Z$ is time-dependent, $Z(t+dt)$ can be calculated according to $Z(t)$ with eq (2).
Secondly, we implemented fuzzy spatial modules in GRASS, a Public Domain GIS developed at the USA-CERL. These commands allow the user to assign to fuzzify a raster map, to achieve basic fuzzy operations and to activate MODFLOU for fuzzy modeling between raster maps.

The third part is a FORTRAN library which can be linked with the C library and called from a FORTRAN program.

4 Orstom’s Method to Determine Ten-Year Floods

Since 1965, ORSTOM carried many studies to determine the ten-year runoff coefficient of Sahelian water basins. From rainfall-runoff lumped concept, Rodier [5] proposed a deterministic approach which is based on the hydrogram unit theory where the peak discharge $Q_{10}$ of a ten-year flood is given by the relation:

$$Q_{10} = K_{10} P_{10} K_{r10} \alpha_{10} S / T_{b10}$$  \hspace{1cm} (3)

with $K_{10}$: spatial reduction coefficient, $P_{10}$: ten-year daily precipitation, $K_{r10}$: runoff coefficient, $\alpha_{10}$:peak to volume ratio, $S$: catchment area, $T_{b10}$: base time. Subsequently, we focused on the $K_{r10}$ evaluation.

4.1 $K_{r10}$ determination with ORSTOM method

For a given basin and a precipitation $P_{10}$, $K_{r10}$ is interpolated between two coefficients ($K_{r70}$, $K_{r100}$) corresponding to (70, 100)mm
precipitation. These coefficients are determined according to the basin surface (S) for five infiltration classes (Ci) and for different global slope indices (Ig), all based on observations. Figure 5 shows an example of the curves providing $K_{r70}$ values for these different parameters.

Figure 5: Fitting curves of $K_{r70}$ vs Surface for different slope indices and permeability classes. The data come from observation of Sahelian basins. Similar curves exist for $K_{r100}$ coefficient.

The global slope index Ig is the ratio between the global altitude difference and the length of the watershed equivalent rectangle.

The quantification of the infiltration classes is the most delicate point in the proposed method. Casenave and Valentin [6] determined the surface features factors influencing the infiltration and adopted a classification based on five groups characterized by a theoretical water depth $Pr$ corresponding to 50 mm net rainfall, as shown in Table 1.

<table>
<thead>
<tr>
<th>Infiltration classes</th>
<th>$Pr$ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>very impermeable basin</td>
</tr>
<tr>
<td>I</td>
<td>impermeable basin</td>
</tr>
<tr>
<td>RI</td>
<td>relatively impermeable basin</td>
</tr>
<tr>
<td>P</td>
<td>permeable basin</td>
</tr>
<tr>
<td>TP</td>
<td>very permeable basin</td>
</tr>
</tbody>
</table>

Table 1: Infiltration capacity classification.

Thus $K_{r10}$ is written as following:

$$K_{r10} = K_{r70}(S,Ci,Ig) + [K_{r100}(S,Ci,Ig) - K_{r70}(S,Ci,Ig)](P_{10-70})/(30)$$  (4)
Actually a fuzzy rule system is particularly suitable to this approach which obviously handles uncertain data and expert knowledge.

4.2 Modeling $K_{r10}$ with fuzzy rules

Using experimental data obtained by Rodier for 74 water basins, we establish a calibration and validation datasets. Then according to the methodology described in §2.3, $K_{r70}$ and $K_{r100}$ are calculated, and $K_{r10}$ determined thanks to eq (4). The calibration was manual and based on a Nash index defined as following:

$$\text{In} = 1 - \frac{\sum (K_r^c - K_r^o)^2}{\sum (K_r^o - K_r^m)^2}$$

where $K_r^c, K_r^o, K_r^m$ are the calculated, observed and average $K_r$ values.

Many computations have been carried out by Delclaux and Lamachère [7] to test the sensitivity of the model to the algorithms. Thus the best results were obtained with the gravity center defuzzification method, the (min., max.) t-norm and t-conorm functions and the cresting maximum combination of responses for rule aggregation. TFN functions have been assigned to all variables except for the Surface (TRFN) for which uncertainty is weak because of the acceptable accuracy in area calculation.

Figure 6 contains the fuzzy $K_{r10}$ results obtained on the validation dataset and shows a good agreement with Rodier observed data. It may be noted that the ORSTOM method provided little better results than the fuzzy model in term of Nash Index. This is mainly due to an information loss resulting from the rule definition as they have been determined with the help of the abacus.

Figure 6: Calculated $K_{r10}$ vs Rodier observed $K_{r10}$ for validation dataset.

Nash index: fuzzy $K_{r10}$ = 0.69, curve $K_{r10}$ = 0.78

Moreover, simulations have been carried out concerning the influence of the variable number. On one hand, the model efficiency decreases as the
variable number increases, as it can be seen on Figure 7-c where output variables and rules have been generated by automatic algorithms from the calibration dataset. On the other hand, if the variable number is divided, for example by a factor 2, the efficiency decrease is more severe (cf. Figure 2-a). In this case, the knowledge loss is too important.

![Figure 7](chart.png)

**Figure 7**: Evolution of model efficiency vs number of variables.
(a) simplified knowledge  (b) expert knowledge  
(c) rules and Kr variable derived from calibration dataset

## 5 Spatial Fuzzy Modeling

A part of the ORSTOM CATCH project - Couplage Atmosphère Tropicale et Cycle Hydrologique- is dedicated to the hydrological environment modeling in West Africa. One of the steps presently in progress concerns the human knowledge integration in spatial analysis to determine the more consistent parameters as regards to hydrology-GCM coupling.

As stated by Burrough and Heuvelink [8], modeling human reasoning with binary logical approach can be deficient due to the limitation of operations on Boolean classified maps and the pixel value uncertainty. But fuzzy approach explicitly addresses this point.

Consequently fuzzy treatments has been implemented in GRASS for improving the spatial operations according to the ORSTOM knowledge in this part of Africa. Presently the spatial fuzzy model is not yet operational due to its incapacity to deal with large raster maps. We just point on the fuzzification effect on classification. The data are slope maps derived from GTOPO30 DEM over Benin country. The first map (cf. Figure 8-left) is the result of a standard classification and shows that medium slopes (greater than 2 %, in black) are located on small parts of
the study area. On the contrary, the fuzzy classified map of Figure 8-right shows how large is the boundary (in black) between the low and medium slope zones as we consider slope as a fuzzy variable. So, the question of the data quality cannot be avoided: in this case, fuzzy analysis not only evaluates the sensitivity to the uncertainty, what can be achieved with classical tools, but also proposes a method for controlling error propagation in further treatments.

Figure 8: Boolean and fuzzy classification applied on Benin slope maps. The selected slope classes are low and medium.
- left map: Boolean classification (black=medium, white=low)
- right map: fuzzy classification (black=medium and low, white=medium or (exclusive) low)

6 Conclusion

Some fuzzy computer tools have been developed, from a basic library up to a ready-to-use software, MODFLOU. Then, the fuzzy set theory has been applied towards two hydrological problems.

In the first one, we calculate fuzzy ten-year runoff coefficients in Sahelian area and we compare them with the Rodier observed data. This comparison gives a good agreement with the observed data, as well as with the ORSTOM method based on curve interpolating. A particular point concerning fuzzy approach stays in the rule and output variables determination: the closer to the expert knowledge the rules are set, the more efficient the model will be. In other words, too many rules do not really improve the system, and too few rules decrease its efficiency.
In the second one, we introduce fuzzy spatial analysis in the context of a large scale hydrological environment modeling in West Africa for which data are not always accurate enough. The first treatments applied on Benin slope maps show the limits of a Boolean classification approach and is an encouragement to carry on this fuzzy spatial inference work in order to integrate the ORSTOM knowledge in the project.

References


