Flood Analysis: On the Automation of the Geomorphological-Historical Method

Elena Fernández, Miguel Fernández, Soledad Anadón,
Gil González-Rodríguez, and Ana Colubi

Abstract. Different methods to assess the flood return period are available in the literature. The hydrological-hydraulic approaches, among the best-known quantitative methods, oversimplify the complex characteristics of the fluvial systems. Additionally, they rely on data that are usually criticized because of their low quality and representativity. In contrast, the semi-quantitative approach based on geomorphological and historical information has lead to more realistic and promising results in pilot studies. This approach is based on highly informative field data providing valuable knowledge which can be used to test the aforementioned quantitative approaches. The aim of this work is to analyze the kind of information that is required to apply the latter method and to explore the possibilities of its automation.

Keywords: Flood frequency, Geomorphological-historical information, Imprecise data, Supervised classification.

1 Introduction

Floods are one of the most common hazards in Europe. They are causing nowadays large losses. To reduce such losses, it is essential to improve the assessment of the return period of these events. The approach that has traditionally been employed to estimate the return period is based on hydrological-hydraulic models. Nevertheless, this approach is being more and more criticized due to the unrealistic assumptions that it requires and the poor

Elena Fernández, Miguel Fernández, Soledad Anadón, and Ana Colubi
INDUROT, University of Oviedo, 33600 Mieres, Spain
e-mail: elena,soledad,miguelfa,colubi@indurot.uniovi.es

Gil González-Rodríguez
European Center for Soft Computing, 33600 Mieres, Spain
e-mail: gil.gonzalez@softcomputing.es

C. Borgelt et al. (Eds.): Combining Soft Computing & Stats. Methods, AISC 77, pp. 239–246.
springerlink.com © Springer-Verlag Berlin Heidelberg 2010

results that are obtained when the available data are scarce [7, 9]. These limitations make the hydrological-hydraulic models not always suitable for scarcely populated mountain zones where not enough reliable data have historically been recorded [6]. An increasing number of authors are suggesting the need of considering complementary information [3, 10]. In this sense, the Spanish Authorities have approved a program to elaborate a National Cartography System of Flooding Areas combining different methodologies.

The geomorphological-historical method [2] has shown to lead to more realistic results in recent studies developed in North Spain [4, 7]. Nevertheless, the employed information is very heterogeneous, has different degrees of reliability and precision and the final combination to assess the return period is made by expert criteria. In order to guarantee the objectivity of the approach a systematic analysis of the information and an automation of the final assessment is required. The usual flooding categories are those established by the EU Flood Directive, namely, low, medium and high flooding probability (respectively associated with return periods of about 500, 100 and 10 year). The final aim will be to obtain an automatic classification rule from a supervised experiment which has to be properly designed. In this work, some results obtained by pilot studies are discussed.

2 Analyzing the Flood Frequency with Geomorphological-Historical Information

Geomorphological and historical information is collected both in field and office work. In [4] historical data obtained from documentary sources and riverside inhabitants interviews are used to define an index of the flood magnitude. The index is based on 5 partial indicators. Namely, the discharge measure, the event magnitude according to the interviews, the proportion of interviewees mentioning the event, the flooded area percentage and the proportion of other documentary sources mentioning the event.

Given that the effect of some of the considered indicators cannot be precisely evaluated, intervals and fuzzy sets reflecting the imprecision are employed. The imprecision varies depending on different factors, so different ways of obtaining the intervals as a function of those factors are introduced. On the other hand, the importance and/or reliability of the indicators is different according to the expert criteria. Thus, the synthetic index gathering the information of all the indicators is computed as a weighted (interval-valued) mean of the valid data.

Once the information of the different indicators is computed and merged into the synthetic interval-valued index, a representation as the one in Fig. 1 is obtained. This index allows us to complete the flood chronology by deduction. For instance, although no historical information was obtained for 1993 in a given unit, it can be deduced because it is known that the unit was flooded

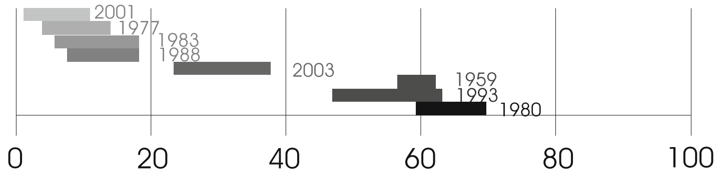

Fig. 1 Interval values for the synthetic index measuring the magnitude of each event.

in 2003: since it is known that a smaller event flooded the area, the larger event had to flood that unit too.

After reconstructing the series of historical floods, a lower bound for the return period can be obtained. The flood probability can be estimated through the flood frequency in the considered period. Since such a period is just a sample, the underlying stochastic variability can be considered. Specifically, confidence intervals based on the score method are proposed to be computed in [4], due to the performance of this method for small sample sizes. Time non-stationarity could also be considered to improve the results in this approach.

It is clear that having documentary references to all the historical floods is not feasible, even in the current information society. This is especially critical in sparsely populated areas. For this reason, it is proposed to complement this information with geomorphological data.

If a given unit is frequently flooded, visible geomorphic evidences can be found by the experts [8, 9]. Thus, observing the presence/absence of morphologies such as those in Fig. 2 provides us with highly valuable information about the flood frequency. Specifically, they allow to identify the high frequently flooded plains, which is essential for the flood hazard management. The shortcoming of this kind of data is that they do not allow to determine high return periods. Additionally, a certain degree of expert knowledge is required in order to identify the morphologies.

Nevertheless, there are other indicators, as the height of the river bank which may also supply information about the flooding frequency. For instance, in Fig. 3 a flooding plain with low river bank is shown. In contrast, Figure 4 displays the opposite situation. If they refer to the same stretch of the river, it is clear that the first unit may be easily flooded. According to the experts, the corresponding water cross-section to reach the flooding plain, and the surface of the drainage basin of each flooding plain are other quantitative variables to be considered. To quantify all these indicators, Digital Elevation Models (DEMs) are frequently used when available. In this particular study, DEMs were available with a 1-meter pixel resolution and up to milimetric precision for the height values.

Pilot studies have shown that some of those variables (height and cross-section) are statistically related to the probability of belonging to the class of high/medium/low frequency flood determined by expert criteria. However,

Fig. 2 Geomorphological evidences.

Fig. 3 Flood plain with low height.

considering only the height leads to results almost as good as those obtained by taking into account more variables (see Table 1).

From Table 1, we can conclude that DEM height measurements are very valuable for classifying between Medium and Low frequent flooding plains. Unfortunately, this accuracy is partially lost when the critical classification problem between High and Medium classes is considered. In this case, it seems that the combination with the cross-section information improves the results. This lack of accuracy is probably connected with the lower reliability of DEM measurements due to the usual abundance of vegetation in the High and Medium frequency flooding plains. Consequently, we consider that obtaining more reliable measures of the height is an essential task.

Table 1 Leave-One-Out percentage of right classification for discriminant analysis based on DEM measurements.

	High/Medium	Medium/Low	High/Medium/Low
Height	78.9%	93.3%	84.6%
Cross-section	78.9%	63.3%	64.1%
Surface	00.0%	68.8%	37.2%
Height and cross-section	84.2%	90.0%	87.2%

Fig. 4 Flood plain with high height.

Measuring exactly the height in field work would be too expensive. Nevertheless, obtaining a subjective valuation from the field researchers by a simple visual inspection is very easy. On the contrary, the cross-section and surface measurements require more expensive tools, as DEM or 1:2000-scale topographic maps.

Following the approach in [5], a fuzzy scale allowing to capture, not only the perception but also the uncertainty of the field researcher is proposed. Specifically, in the pilot study carried out, the field researchers were asked to collect their valuation of the height by means of trapezoidal fuzzy sets. A trapezoidal fuzzy set T is characterized by a $[0,1]-$valued function defined on $S \subset \mathbb{R}$ assuming positive values over an interval $[a,b]$, called 0-level, the value 1 over an interval $[c,d]$, called $1-$level, linearly increasing between a and c and linearly decreasing between b and d (see an example in Fig. 5). For each $x \in S$, $T(x)$ represents the degree of compatibility of the perception of the expert with the assertion "the height is x".

Thus, the experts are asked to choose the $0-$level of the fuzzy set as the smaller interval that they would not completely discard as containing the "true" Height value, whereas the $1-$cut is to be chosen as the interval of

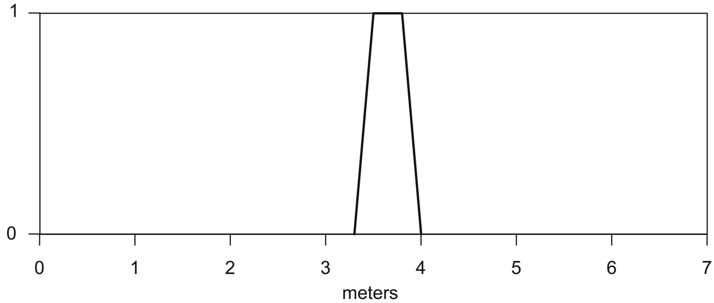

Fig. 5 Example of fuzzy perception of the height.

Table 2 Leave-One-Out percentage of right classification for discriminant analysis based on fuzzy field valuations.

	High/Medium	Medium/Low	High/Medium/Low
X_1 =Inf T_0	89.5%	83.3%	79.5%
X_2 =Inf T_0	89.5%	83.3%	82.1%
X_3 =Sup T_0	94.7%	83.3%	84.6%
X_4 =Sup T_0	89.5%	83.3%	82.1%
X_1, X_2, X_3, X_4	84.2%	90.0%	84.6%
$D = (X_1 + X_2 + X_3 + X_4)/4$	89.5%	83.3%	82.1%

values that they indeed consider completely compatible with the height that they are observing. In other words, the 1−level would contain their personal opinion and the 0−level the range that they could admit to a greater or lesser extent. In this way, fuzzy perceptions of the length as that in Fig. 5 are available.

In order to verify if the collected fuzzy information is useful for the considered classification problem, several approaches can be considered. On the one hand, a classical discriminant analysis based on the 4 variables recorded for each trapezoidal fuzzy perception T (infima and suprema of the 0− and the 1− level set), as well as an average of all of them as a defuzzifier can be applied (see Table 2).

According to the results in Tables 1 and 2 it seems that the classification results between High and Medium are better when the field valuation is considered. On the contrary, the classification results between Medium and Low are better when the DEM measures are employed.

However, the analysis in Table 2 is not taking into account the structure of fuzzy data in the classification problem. To consider the fuzzy sets as structured data, the Proximity-based Classification Criteria for Fuzzy data

Table 3 Leave-One-Out percentage of right classification for Proximity-based classification with fuzzy field valuations.

	High/Medium	Medium/Low	High/Medium/Low
heigh valuation	89.5%	90.0%	84.6%

(PCCF) in [1] can be employed. The simplified idea of PCCF is to consider the fuzzy data as observations of a fuzzy random variable X and to proceed as follows:

- the 'center' C_i of each group G_i is computed by averaging of the fuzzy data in this group.
- To classify a new fuzzy data T, the conditional probability
$$P(d(X,C_i) > d(T,C_i)/G_i)$$
is estimated. This probability is a kind of measure of the affinity of T to each one of the groups.
- T is assigned to the group G_i with highest estimated probability.

The results in Table 3 are obtained by applying PCCF to the fuzzy perceptions of the height collected in the pilot study. These results indicate that the consideration of fuzzy field valuations of the height is very valuable in comparison with the consideration of the DEM measures for High/Medium classes. Additionally, for Medium/Low and the overall classification, the results are comparable. Thus, taking into account these results, the cost of both kinds of data, and since the experts have to visit anyway the flooding plains to look for geomorphological evidences, we recommend to consider the systematic collection of fuzzy valuations.

3 Concluding Remarks

In this paper we have surveyed some of the most useful geomorphological and historical information that can be used in order to assess the flooding frequency. The different sources show various degrees of imprecision and reliability. None of them is uniformly the best to classify the flooding plains according to the expert criterion, nevertheless they supply complementary information that can be merged to build a quantitative model in the future.

The documentary sources, the historical records and the interviews can be used to determine lower bounds that may avoid critical underestimates provided by other methods.

The geomorphological evidences are, of course, very useful to classify high frequency flooding. However, these dichotomic variables are not enough to accurately determine low return periods. Thus, although the expert criterion is highly linked to these variables, they require more information to distinguish between medium and low frequency classes. One of the characteristics that the experts use for that purpose is the river bank height, as well as other quantitative variables involved in the hydrological-hydraulic models. The pilot studies have shown that considering the height is essential. Nevertheless, reliable measures are required. One of the ways of measuring this indicator is to use DEM, but there are problems in presence of lush vegetation, which is related to high frequency classes. The pilot studies have indicated that incorporating fuzzy valuations of height provided by the field researchers gives, in general, better results than using DEM measures. Additionally, it is a not expensive source of information which does not require expert knowledge on geomorphology. Nevertheless, the necessity of developing pilot studies for each new basin should be underlined, because the heights determining the cuts between classes are specific of each basin, whence the sample to train the supervised classification has to be updated.

Acknowledgements. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grants MTM2009-09440-C02-01 and MTM2009- 09440-C02-02 and by the Principality of Asturias Projects IB09-042C1 and IB09-042C2. Their financial support is gratefully acknowledged.

References

1. Colubi, A., González-Rodríguez, G., Gil, M.A., Trutchnig, W.: Discriminant Analysis for fuzzy random variables based on nonparametric regression (submitted for publication, 2010)
2. Baker, V.R., Kochel, R.C., Patton, P.: Flood geomorphology. Wiley, New York (1988)
3. Baker, V.R.: Paleoflood hydrology: Origin, progress, prospects. Geomorphology 101, 1–13 (2008)
4. Fernández, E., Colubi, A., González-Rodríguez, G., Anadón, S.: Integrating statistical information concerning historical floods: ranking and interval return period estimation (submitted for publication, 2010)
5. González-Rodríguez, G., Colubi, A., Gil, M.A.: Fuzzy data treated as functional data. A one-way ANOVA test approach (submitted for publication, 2010)
6. Jarret, R.D.: Hydrologic and hydraulic research in mountain rivers. Hydrology of Mountainous Areas 190, 107–117 (1990)
7. Lastra, J., Fernández, E., Díez-Herrero, A., Marquínez, J.: Flood hazard delineation combining geomorphological and hydrological methods: an example in the Northern Iberian Peninsula. Natural Hazards 45, 277–293 (2008)
8. Magilligan, F.J., Phillips, J.D., James, L.A., Gómez, B.: Geomorphic and sedimentological controls on the effectiveness of an extreme Flood. J. Geology 106, 87–96 (1998)
9. Ortega, J.A., Garzón, G.: Interpretación de los depósitos de avenida como clave para establecer la dinámica de la llanura de inundación. In: Pérez Alberti, A., López Bedoya, J. (eds.) Geomorfología y territorio. Cursos e Congresos da Universidade de Santiago de Compostela, pp. 629–644 (2006)
10. Thorndycraft, V.R., Benito, G., Gregory, K.J.: Fluvial geomorphology: A perspective on current status and methods. Geomorphology 98, 2–12 (2008)