Flash Flood Forecasting by Statistical Learning in the Absence of Rainfall Forecast: a Case Study

Mohamed Samir Toukourou[1], Anne Johannet[1], Gérard Dreyfus[2]

[1] Ecole des Mines d'Alès, CMGD, 6 av. de Clavières
30319 Alès cedex, France
[2] ESPCI-Paristech, Laboratoire d'Electronique, 10, rue Vauquelin
75005 Paris, France

Abstract. The feasibility of flash flood forecasting without making use of rainfall predictions is investigated. After a presentation of the "cevenol flash floods", which caused 1.2 billion Euros of economical damages and 22 fatalities in 2002, the difficulties incurred in the forecasting of such events are analyzed, with emphasis on the nature of the database and the origins of measurement noise. The high level of noise in water level measurements raises a real challenge. For this reason, two regularization methods have been investigated and compared: early stopping and weight decay. It appears that regularization by early stopping provides networks with lower complexity and more accurate predicted hydrographs than regularization by weight decay. Satisfactory results can thus be obtained up to a forecasting horizon of three hours, thereby allowing an early warning of the populations.

Keywords: Forecasting, identification, neural network, machine learning, generalization, weight decay, early stopping.

1 Introduction

The need for accurate predictions of flash floods has been highlighted by the recent occurrences of catastrophic floods such as in *Vaison-la-Romaine* (1991), *Nîmes* (1988), *Gardons* (2002), *Arles* (2003), to name only a few, located in the south of France. These disasters result from intense rainfalls on small (some hundreds of km^2), high-slope watersheds, resulting in flows of thousands of m^3/s with times of concentration of a few hours only. The death toll (over 100) in these circumstances in the southeast of France, and the cost of more than 1.2 billion Euros in 2002, showed that the design of a reliable tool to forecast such phenomena is mandatory.

Faced with this major risk, the French Ministry in charge of Sustainable Development (currently MEEDADT) created in 2003 the national center for flood forecasting and warning SCHAPI (*Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*), which is in charge of the "*vigicrue*" surveillance service. The *Gardon d'Anduze*, in the South-East of France, has been chosen by this Center as a pilot site to compare the flash floods (concentration time of 2h-4h) forecasting models. In this context, this paper describes the study of neural network models to build an efficient real time flash flood forecaster.

Real time flash flood forecasting is usually addressed by coupling complex atmospheric and hydrologic models. The complexity generated by this coupling is huge, and the performances of the present models are limited by several factors: the observations may not be accurate enough for these models to produce useful predictions, the models may be biased by a lack of observations on the ground at an appropriate scale, and the models themselves do not take into account the whole complexity of the phenomena.

An alternative approach consists of capitalizing on the available data in order to build models by statistical machine learning. This will reduce the computational burden and free the model designers from the limitations of physical modeling when the phenomena are too complex, or when the estimation of physical parameters is difficult.

Due (i) to the lack of accurate estimations of rainfalls, and (ii) to the high noise level in water level measurements, and in order to guarantee the best possible generalization capabilities, complexity control is a particularly critical issue. Two traditional regularization methods have been investigated: early stopping and weight decay. After careful variable and model selection, the ability of models, obtained by either regularization method, to predict the most dramatic event of the database (September 2002) is assessed. Hydrographs are displayed and comparisons between the results of both methods are performed. Finally, we conclude that, in the present case, training with early stopping provides networks with lower complexity, longer training but more satisfactory predictions.

2 Problem Statement

2.1 Flash Flood Forecasting

Real time flood forecasting is currently addressed on a disciplinary basis by coupling atmospheric and hydrologic models, hydrologic and geographic models, sometimes in conjunction with a risk management system or an expert system. The difficult issue is the modeling of flash floods in mountainous areas. That problem is usually considered from the hydraulic or hydrologic viewpoints, and, in sharp contrast to the present study, the forecast of rainfall or climate is an important and necessary part of current projects (European projects Flood Forecasting System [1] and PREVIEW [2]). In PREVIEW, Le Lay [3] derives a space-dependent rainfall-runoff relation, whose inputs are the rainfall radar observations and/or the rainfall forecasts provided by the weather models, and the output is the discharge for the *Gardon d'Anduze* watershed. This work shows a fundamental limitation of hydrologic models: they need rainfall forecasts because they can only propagate the rainfall over the watershed, in agreement with the physical behavior. However, in the case of small basins subject to intense storms, no rainfall forecast is available with suitable time scale and accuracy. Traditionally, two hypotheses are postulated: either null rainfall, or persistency of past observed rainfalls. Obviously these assumptions are inappropriate, and, as a consequence, the forecasts of the model are not satisfactory. A probabilistic approach is possible in order to downscale the rainfall predictions [4]; nevertheless the time scale is not appropriate given the rising time of the *Gardon d'Anduze* flood. The coupling of atmospheric and geographical data can also be performed with remote sensing data [5]. As a consequence, the huge quantity of data to be processed leads also to investigate parallel simulation as in the CrossGrid project, where a prototype of flood forecasting operates on Grid technologies [6]. From the viewpoint of the end users, the very short computation times involved in the execution of neural network algorithms once training has been performed makes them very attractive as components of a warning system, without having to resort to grid computing. Another advantage is that any nonlinear, dynamical behavior may be modeled by neural networks, particularly the relation between the rainfall up to time t and the discharge at time $t+f$. Forecast is thus possible without estimating future rainfalls. Although neural networks were applied previously to the forecasting of outflows at several forecasting horizons [7] [8], or for water supply management in mountainous areas [9], they were never applied to events of such speed and intensity.

2.2 *Gardon d'Anduze* Flash Floods

The *Gardon* catchment is emblematic of flash flood behavior: first, its floods are very irregular and may rise up to several meters in a few hours; in addition, the basin is populated, which explains the huge damage costs and loss of human lives.

In a few words, the *Gardon d'Anduze* catchment, sub-catchment of the *Gardon* catchment (*Rhône* river tributary) is located in the southeast of France, in the *Cévennes* mountainous area. The basin area is 546 km^2, the catchment is mountainous and has a large mean slope of 40%, which explains the velocity of the floods. The basin contains three main geological units: schist (60 %), granite (30%) and limestone (10%), which results in heterogeneous soil moisture and permeability.

The *Anduze* catchment is subject to very intense storms delivering huge amounts of water: for example, a 500 mm rainfall was recorded in the *Anduze* rain gauge in less than nine hours in 2002. These storms occur most often in autumn, when the Mediterranean Sea is almost warm. They are called "*épisodes cévenols*".

Fifteen flash flood events are available in the database, whose characteristics are shown in Table 1 (1700 records sampled every 30 min). Five very intense events are indicated.

In the present paper, we show that, despite the difficulty of the task, the evolution of the water level at *Anduze* can be forecast up to 3 hours ahead of time, without any assumption about the evolution of future rainfall, for the catastrophic, most intense event of the database, namely the event of September 2002.

Table 1. Characteristics of events of the database

Date	Duration (hours)	Maximum level (m)	Discharge Peak (m^3/s)	Very intense
September, 21-24, 1994	35	3,71	181	
October, 4-5, 1995	54	5.34	975	y
October, 13-14, 1995	92	5	864	
November, 10-12, 1996	82	2,71	268	
December, 18-19, 1997	104	5.37	985	y
October, 20-21 1997	34	3,64	473	
November, 5-7 1997	74	4,20	624	
November, 26-27 1997	66	2,58	244	
December, 18-19 1997	104	5,37	985	y
September, 28-29, 2000	46	4.80	800	
September, 8-9, 2002	29	9.71	2742	y
September, 24-25 2006	23	2,24	186	
October, 19-20, 2006	*55*	6.61	1436	y
November, 17 2006	34	2,75	275	
November, 20-23 2007	70	2,69	264	

2.3 Noise and Accuracy

In the present section, we focus on the nature and quality of the information available in the database.

Rainfall measurements are performed with rain gauges. These are very accurate sensors, which broadcast the water level every five minutes; however, they provide very local information, so that the heterogeneity of the rainfalls is an important source of inaccuracy: for example, for the event of September 2002, the cumulated rainfalls were 3 times as large in *Anduze* as in *Saint-Jean-du-Gard,* which is only ten kilometers away. Therefore, the most important rainfall may be located between rain gauges, thereby causing inaccurate estimates due to the too large mesh of the rain gauge network. For this reason, radar acquisition of rainfalls with a definition of 1 km^2 has been performed since 2002, but complete, homogeneous sequences are not yet available for all events.

Water level measurements are available with several sampling periods: 1 hour from 1994 to 2002, and 5 minutes after this date. However, because of real time constraints, the sampling period used in this work is T = 30 mn, although variance analysis has shown that 15 minutes would be more appropriate. Thus for events recorded before 2002, the peak value is probably underestimated, possibly by 10% to 30%. For the event of 2002, the error results from an accident: instrumentation was damaged during the event, and the water level was estimated *a posteriori*.

Therefore, the unreliability of the available data makes the forecasting of such catastrophic events a challenging task.

3. Model Design

3.1 Definition of the Model

Given a forecasting horizon f, the model is intended to forecast, at discrete time kT, ($k \in N^+$) the water level at *Anduze* at time $(k + f)\,T$ ($f \in N^+$).

The available information for the *Anduze* catchment is the water level at the *Anduze* station, the rainfalls at 6 rain gauges delivering the cumulated rainfalls over the sampling period (30 min), and the soil moisture (Soil Water Index, given by the ISBA (Interactions between Soil, Biosphere, and Atmosphere) model [10]).

The 6 rain gauges: *Barre-des-Cévennes*, *Saint-Roman de Tousques*, *Saumane*, *Mialet*, *Soudorgues* and *Anduze* are spatially well distributed and one can consider that each of them is important. The information about rainfalls is conveyed to the network as sliding windows. All sliding windows have equal width w, whose optimal value is chosen as described in section 3.2. Different values were found, depending on the forecasting horizon (Table 2). Similarly, the information about past water levels is conveyed as sliding windows, whose optimal width was found to be $r = 2$, irrespective of the forecasting horizon.

Table 2. Sliding window width for rainfalls

Forecasting horizon (f)	0.5 hour	1 hour	2 hours	3 hours	4 hours	5 hours
w	2.5	3	3	2	0.5	0.5

Because of the non-linearity of the physical process, we take advantage of the universal approximation property of a neural network with one hidden layer of sigmoid neurons and a linear output neuron [11]. The water level at time $t+f$ is forecast from (i) the measured rainfalls in a sliding window of width w, and (ii) from the measured water levels in a sliding window of width r. The training data is chosen (see section 3.2) in the set of flood sequences recorded over several years (1994-2007), described in Table 1.

Since the model takes into account measured past values of the water level, during the same flood, the available information about soil moisture is not explicitly conveyed to the model since it is implicitly present in the input data.

3.2 Model Selection

One of the events was set apart for use as a test set (see section 3.3); another event was selected for use either as an early stopping set when the latter regularization technique was used, or as an additional test set when regularization was performed by weight decay (see section 4 for more details on regularization). In the latter case, it was also set apart and used neither for training nor for model selection.

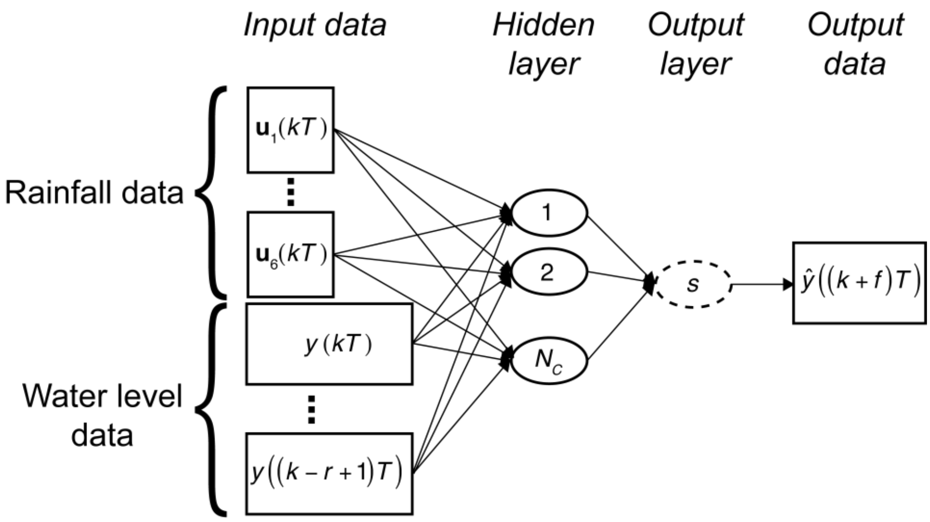

Fig. 1. The model is fed by cumulated rainfall measurements provided by the 6 rain gauges over a temporal window of width w: $\mathbf{u}_i(kT)$ is the vector of the w rainfall level measurements provided by rain gauge i (i = 1... 6) at times kT, $(k–1)T$, ..., $(k–w+1)T$. Water level measurements, over a sliding window of width r, are also input to the model. The output is the forecast water level, f sampling periods ahead.

Model complexity selection was performed by partial K-fold cross-validation on the remaining N–2 events of the database. K models were trained from the N–3 remaining floods; therefore, N–K–2 events (the least intense ones) were present in the training set of all K models.

The generalization ability of the model was assessed by the cross-validation score (see e.g. [12]):

$$S = \sqrt{\frac{1}{K}.\sum_{i=1}^{K} MSE_i} \qquad \textbf{(1)}$$

where MSE_i is the mean squared forecasting error of the model, for the time sequence recorded during event i of the validation set.

In the present case, K = 4 events were chosen: the 1995, 1997, and 2006 very intense events reported in Table 1. The 2002 event was selected as a test set because it is typical of events whose forecasting is crucial for early warning. In addition, this event is also more intense than those used for training and validation: it is a difficult test for the models. The other nine floods were always present in the training set.

The above procedure was used for complexity selection, spanning the space of rainfall window width w, water level window width r and number of hidden neurons N_C. For each model, 100 different parameter initializations were performed. Complexity selection was performed separately for each forecasting horizon.

After selecting the appropriate complexity (i.e. after selecting the appropriate values of w, r and N_C) for a given forecasting horizon f, a final model was trained for that horizon, from thirteen sequences: all floods except the test sequence and the early stopping sequence, or all floods except the two test sequences when weight decay regularization was performed. Its performance was assessed on the test sequence(s).

3.3 Training

The usual squared error cost function was minimized by the Levenberg-Marquardt algorithm [13] during training, after computation of the gradient of the cost function by backpropagation.

4 Regularization

In addition to performing input and model selection by cross-validation, regularization methods were applied during training. Two different methods were assessed in this study: weight decay and early stopping.

4.1 Weight Decay

Weight decay prevents parameters from taking excessive values (resulting in overfitting), by introducing a term in the cost function that penalizes large parameter values; this idea is implemented in a systematic fashion in Support Vector Machines. In the present case, the new cost function is expressed as:

$$J = \gamma MSE + (1-\gamma)\|\theta\|^2 . \quad \textbf{(2)}$$

where MSE is the usual mean squared prediction error, θ is the vector of parameters and γ is the hyperparameter that controls the balance between the terms of the cost function.

Similarly to model selection, the hyperparameter γ was selected by cross validation, for each forecasting horizon, for γ varying from 0.5 to 0.95 with an increment step of 0.05. Table 3 shows the optimal value of γ obtained for each forecasting horizon.

Table 3. Optimal hyperparameter values for each forecasting horizon

Forecasting horizon (f)	0.5 hour	1 hour	2 hours	3 hours	4 hours	5 hours
γ	0.6	0.9	0.7	0.55	0.75	0.75

4.2 Early Stopping

As an alternative regularization technique, early stopping was used in the present investigation. Early stopping consists of terminating training when the prediction error, assessed on a stopping set, different from the training set, starts increasing. It has been shown [14] that it is theoretically equivalent to weight decay. However, due to finite sample size, the results may vary widely depending on the choice of the stopping set. In the present investigation, the stopping set was the event of September 2000 of the database, which was well learnt when it was in the training set, and well predicted when it was in the test set. Therefore, it appeared as a "prototype" of the behavior of the flood process.

5 Results and Discussion

The quality of the forecast can be estimated by various criteria, each of which focuses on a particular desired feature of the model: accuracy of the prediction of the water level at the peak of the flood, accuracy of the prediction of the time of occurrence of the peak, absence of spurious water level peak, etc. The most widely used criterion is the coefficient of determination of the regression

$$R^2 = 1 - \frac{MSE}{\sigma^2}$$

where σ^2 is the variance of the observations. If the model simply predicts the mean of the observations, $R^2 = 0$; conversely, if the model predicts the observations with perfect accuracy, $R^2 = 1$. In the hydrology literature, the coefficient of determination is known as the "Nash-Sutcliffe criterion".

The "persistency criterion" is more specific to forecast models [15]; it is defined as:

$$P = 1 - \frac{\sum\limits_{test\ sequence} (y(t+f) - \hat{y}(t+f))^2}{\sum\limits_{test\ sequence} (y(t) - y(t+f))^2}. \quad (3)$$

where y is the observed water level and $\hat{y}$ is the estimated water level. P is equal to 0 if the predictor is perfectly dumb, i.e. it always predicts that the future value is equal to the present one, and it is equal to 1 if the predictor provides perfectly accurate forecasts.

Tables 4 and 5 describe the models and the accuracy of their predictions on the 2002 flood.

Table 4. Models obtained with regularization by weight decay.

Forecasting horizon (f)	0.5h	1 h	2 h	3 h	4 h	5 h	mean
N_C	7	7	5	7	5	3	5.6
Persistency criterion	0.62	0.75	0.62	0.63	0.70	0.68	0.66
R^2 (Nash-Sutcliffe criterion)	0.96	0.91	0.81	0.78	0.72	0.49	0.78
Estimated/Observed peak values	0.86	0.81	0.77	0.69	0.69	0.57	0.73

Table 5. Models obtained with regularization by early stopping.

Forecasting horizon (f)	0.5 h	1 h	2 h	3 h	4 h	5 h	mean
N_C	2	2	5	3	3	3	3
Persistency criterion	0.45	0.65	0.32	0.28	0.23	0.59	0.42
R^2 (Nash-Sutcliffe criterion)	0.98	0.93	0.87	0.93	0.84	0.58	0.85
Estimated/Observed peak values	0.90	0.84	0.73	0.82	0.79	0.60	0.78

In the present case, early stopping provides consistently more parsimonious models than weight decay, with consistently higher values of the determination criterion. However, the persistency criterion is larger for models obtained with weight decay.

Figures 2 and 3 show the predicted and observed curves for the test sequence (2002 event). Given the difficulty of the task, these results are extremely encouraging, since they show that the model would have allowed the public services to issue early warnings to the population if it had been available during that event.

6 Conclusion

Flash flood forecasting is a very challenging task due to high variability and noise in the data, especially when no rainfall forecast is available. In the present study, we have shown the feasibility of forecasting the catastrophic event of September 2002 in *Anduze* with an accuracy and forecasting horizon that are compatible with an early warning of the populations.

This requires a careful methodology for model selection and regularization; it is shown that early stopping and weight decay result in different generalization capabilities, and that, in this specific case, early stopping provides more satisfactory results on the test set. This is not claimed to be a general result, but it shows that a variety of methods must be used in order to solve such difficult problems satisfactorily.

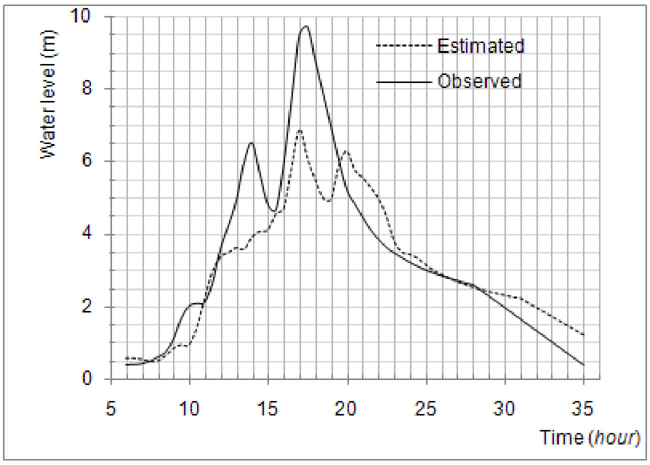

Fig. 2. Hydrograph of the 2002 event for $f = 3$ h – regularization by weight decay

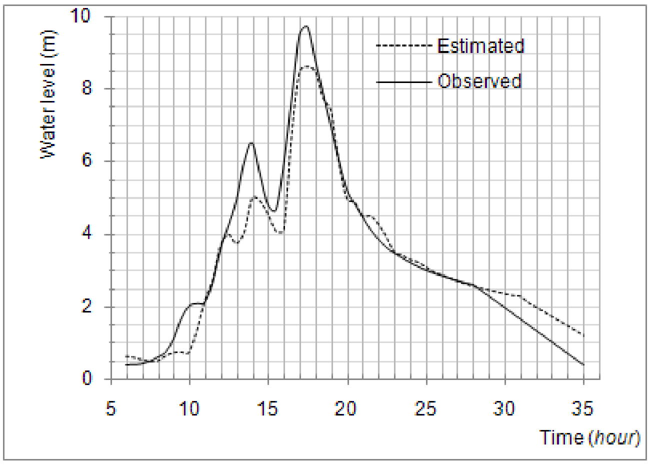

Fig. 3. Hydrograph of the 2002 event for $f = 3$ h – regularization by early stopping.

From the viewpoint of hydrology, this methodology should easily be applied to small (less than 1000 km^2), fast (concentration time less than 10 h) basins providing only rainfalls and water level. Because no exogenous data is necessary, the method should be applicable to many European mountainous watersheds.

Acknowledgments. The authors acknowledge the SCHAPI and its regional service SPC of Gard for their collaboration on this exciting application.

References

1. European Flood Forecasting System, http://effs.wldelft.nl/ (2003)
2. PREVIEW, http://www.preview-risk.com (2008)
3. Le Lay, M., Saulnier, G.-M.: Exploring the Signature of Climate and Landscape Spatial Variabilities in Flash Flood Events: Case of the 8–9 September 2002 Cévennes-Vivarais Catastrophic Event. Geophysical Research Letters 34, 5p (2007)
4. Taramasso, A. C., Gabellani, S., Marsigli, C., Montani, A., Paccagnella, T., Parodi, A.: Operational flash-flood forecasting chain: an application to the Hydroptimet test cases. Geophysical Research Abstracts 7 09--014 (2005)
5. Jasper, K., Gurtz, J., Lang, H.: Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. Journal of Hydrology 267, 40--52 (2002)
6. CrossGrid, http://www.eu-crossgrid.org (2005)
7. Zealand, C M., Burn, D. H., Simonovic, S. P.: Short term streamflow forecasting using artificial neural networks. Journal of Hydrology 214, 32--48. (1999)
8. Schmitz, G.H., Cullmann, J.: PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments. Journal of Hydrology 360, 1--14. (2008)
9. Iliadis, S.L., Maris, F.: An artificial neural networks model for mountainous water-resources management: the case of Cyprus mountainous watersheds. Environmental Modelling & Software 22, 1066--1072 (2007)
10. Noilhan, J., Mahfouf, J. F.: The ISBA land surface parameterization scheme. Global and Planetary Change 13 145--159. (1996)
11. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks Are Universal Approximators. Neural Networks 2, 359--366 (1989)
12. Dreyfus, G.: Neural Networks, Methodology and Applications. Springer, Heidelberg (2005)
13. Hagan, M.-T., Menhaj, M.-B.: Training feedforward networks with the Marquardt Algorithm. IEEE Transaction on Neural Networks 5, n°6, 989--993 (1994)
14. Sjöberg, J., Ljung, L.: Overtraining, regularization, and searching for a minimum, with application to neural networks. International Journal of Control, 62 (6), 1391--1407 (1995)
15. Kitadinis, P.K., Bras, R.L.: Real-time forecasting with a conceptual hydrologic model: 2 applications and results. Water Resour. Res. 16, 1034--1044 (1980)