

# Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of “Cévenol flash floods”

Mohamed Toukourou · Anne Johannet ·  
Gérard Dreyfus · Pierre-Alain Ayrat

© Springer Science+Business Media, LLC 2010

**Abstract** “Cévenol flash floods” are famous in the field of hydrology, because they are archetypical of flash floods that occur in populated areas, thereby causing heavy damages and casualties. As a consequence, their prediction has become a stimulating challenge to designers of mathematical models, whether physics based or machine learning based. Because current, state-of-the-art hydrological models have difficulty performing forecasts in the absence of rainfall provisions, new approaches are necessary. In the present paper, we show that an appropriate model selection methodology, applied to neural network models, provides reliable two-hour ahead flood forecasts.

**Keywords** Flood forecasting · Rainfall-runoff relation · Neural networks · Dynamic system · Generalization

## 1 Introduction

The prediction of flash floods in populated areas is a major challenge for early flood warning systems. For the watershed under consideration in the present paper, current physics-based models have difficulties performing accurate real-time forecasting in the absence of accurate rainfall forecasts. Therefore, machine-learning models based on past flash flood measurements in the same watershed are attractive alternatives.

After a presentation of the “*Cévenol flash floods*” whose prediction is investigated in the present paper, the current physics-based methods are described cursorily, and a short presentation of past attempts at forecasting flash floods with neural networks is provided. In the subsequent section, two candidate models for nonlinear dynamic process forecasting are presented: recurrent neural networks and feedforward neural networks with time delays; it is shown that, given the specific features of flash flood forecasting in the absence of rainfall predictions, the latter models are more appropriate. The contents and structure of the database is described, and a useful typology of the flood events present therein is provided.

The subsequent section describes the model design methodology, with emphasis on variable and model selection by leave-one-sequence-out cross-validation, training and regularization, and independent testing. In the final section, the results are described, and we show that satisfactory two-hour ahead predictions are feasible, thereby opening the way to issuing reliable population warnings in real time.

## 2 Flood forecasting

Flood forecasting involves essentially finding an appropriate relation between rainfalls (past and present) and future river runoff. For heterogeneous watersheds, this is a very difficult task, for various reasons. First, the rainfall-runoff relation depends on the season due to evaporation and evapotranspiration (consumption and evaporation of water by vegetation) [1], which may lead to significant spatial variations of the initial water storage [2]. Moreover, the watershed response also depends on soil moisture and land use. As a consequence, the rainfall-runoff relation is nonlinear, and it depends on variables that are difficult to measure and/or to estimate [2–4].

---

M. Toukourou · A. Johannet (✉) · P.-A. Ayrat  
Ecole des Mines d’Alès, CMGD/LGEI, 6 av. de Clavières, 30319  
Alès Cedex, France  
e-mail: [anne.johannet@mines-ales.fr](mailto:anne.johannet@mines-ales.fr)

G. Dreyfus  
ESPCI-Paristech, Laboratoire d’Electronique, 10 rue Vauquelin,  
75005 Paris, France

## 2.1 Flash floods of Gardon d'Anduze

The Experimental Digital Watershed (EDW) program<sup>1</sup> was initiated by the French Ministry of Environment (MEEDADT) in order to compare the real-time performances of different forecasting models on a few test watersheds. Three EDW test sites were proposed: the *Meuse* and the *Marne* watersheds for slow floods and “water table” floods respectively, and the *Gardon* watershed at *Anduze* for flash floods. Flash floods are floods that exhibit a sharp flow peak occurring on a very short time scale (less than 4–6 hours in the case of the *Gardon* watershed) after a very intense rainfall. The prediction of the latter flood type is still an open issue, because of its complexity, which is due mainly to the spatial heterogeneity of the rainfall and of the soil moisture [2, 4]. In the absence of accurate measurements of these important variables, it is very difficult for physical models to predict accurately the evolution of the water levels or flows.

In view of the damages caused by flash floods in populated areas (over 100 fatalities over the past two decades in the southeast of France, and 1.2 billion € of economical damages for the sole event of 9th September 2002 [7, 8]), predictive methods that do not require the accurate knowledge of the above factors are potentially very useful.

The *Gardon* catchment exhibits the typical behavior that the present article aims at modeling: floods are very irregular and may rise up to several meters in a few hours. In addition, it is vulnerable: a population lives and works in this basin, which explains the huge damage costs and loss of human lives.

The *Gardon d'Anduze* catchment, sub-catchment of the *Gardon* catchment (Rhône river tributary) is located in the southeast of France as shown in Fig. 1, in the *Cévennes* mountainous area. The basin area is 546 km<sup>2</sup>, the catchment is mountainous with large mean slopes (40%) and high soil infiltration capacity, which explains the velocity of the floods. The basin contains three main geological units: schists (60%), granite (30%) and limestone (10%), which accounts for the heterogeneity of soil moisture and permeability.

The *Anduze* catchment is subjected to very intense storms delivering huge amounts of water: for example, a 600 mm rainfall was recorded in the *Anduze* rain gauge in less than 12 hours in 2002.<sup>2</sup> These storms occur most frequently during autumn when the Mediterranean Sea is almost warm.

Rainfalls and flows were measured both for very intense and for “moderate” flood events, over a period of thirteen

years. Spatially distributed rainfall measurements were provided by six rain gauges. The list of recorded flash flood events for the *Gardon* watershed at *Anduze* is reported in Table 1.

Usually the flow is not measured directly: it is derived from the water level by means of a “rating curve”. The rating curve must be derived carefully; it depends on the section of the river, on the roughness of the waterbed, on its slope, and it may vary in time due to local changes in the waterbed during floods. In the present study, water levels were forecasted and subsequently converted to discharges using the rating curve. The discharge is the physical value characterizing the river state, while the level is location-specific.

As an additional difficulty, various sensor failures may affect the accuracy of the measurements.

At present, measurements of the *Gardon d'Anduze* catchment are sampled with a five-minute sampling rate and transmitted to the local Flood Forecasting Service which analyzes the incoming data.

## 2.2 Hydrological methods for flood forecasting

One can distinguish between *simulations*, where the flow at time  $t$  is computed from rainfalls at the same time  $t$  or earlier, and *forecasts*, where the flow is computed ahead of time. Since water flow is due solely to rainfalls, flow forecasting requires either rain forecasts, or *ad hoc* assumptions on future rains, such as null or constant rainfall. The last two assumptions are not satisfactory in the case of flash flood as investigated here, because rainfalls are highly variable in time. Furthermore, these assumptions cannot take into account the spatial variability of rainfalls. Rainfall predictions in the area under investigation are not accurate enough, because the rain generation process itself is not yet clearly understood: it is still a difficult task to forecast accurately when or where the storm will be stabilized, even for recent meteorological models (such as AROME [5]), due to the “backward regenerative system” that feeds the storm with the water of the Mediterranean sea [6]. Hydrological models may belong to one of three categories: (i) physical models, (ii) conceptual models and (iii) statistical models. Some models implement a combination of these approaches. The first two families of models will be presented briefly, before describing neural network based statistical models in detail.

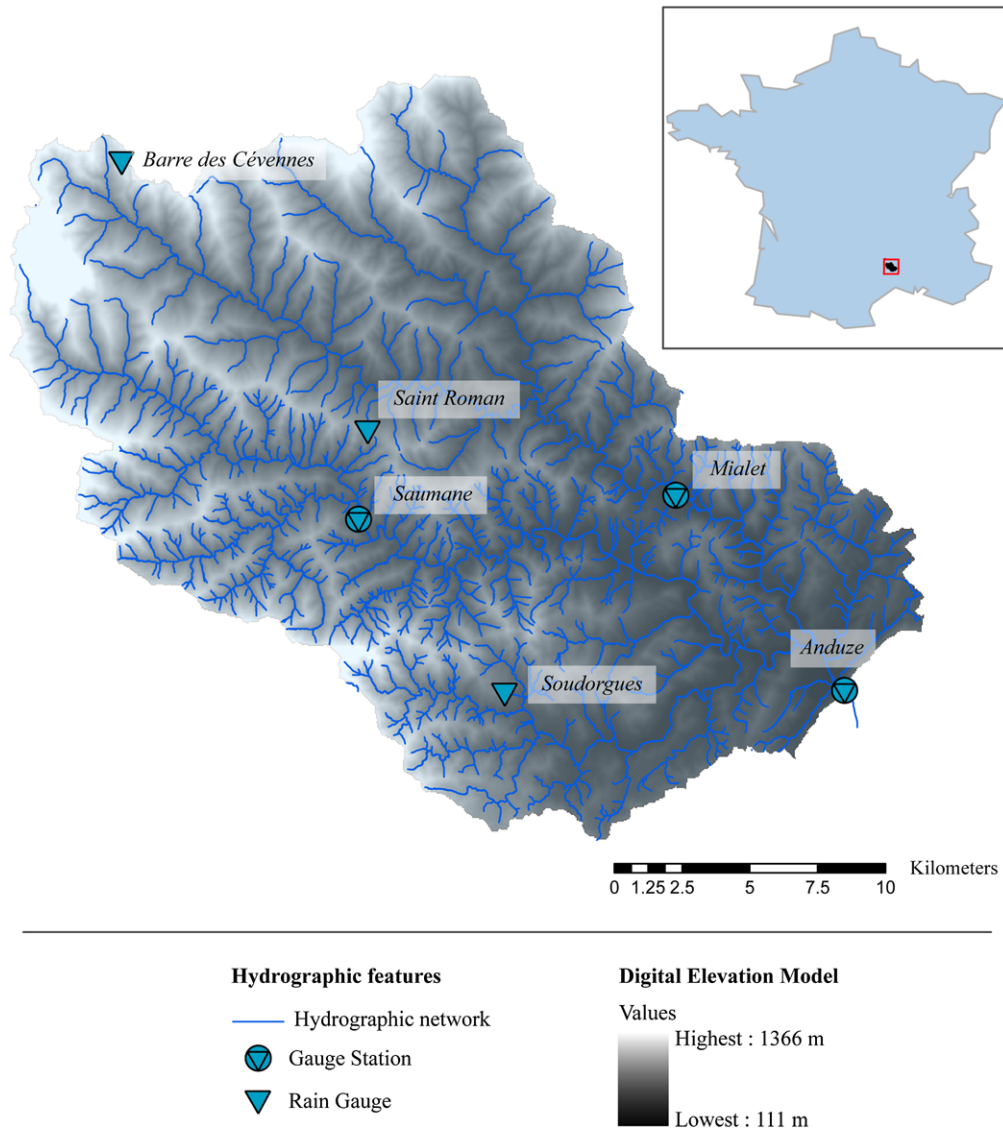
### 2.2.1 Physical modeling

Physical modeling aims at modeling the physical phenomena that result in flash floods. In the case of the catchment under investigation, the heterogeneity of rainfalls and the behavior of the rainfall-soil couple are the main factors that account for the phenomena: physics-based models must take

<sup>1</sup>French acronym: BVNE for Bassin Versant Numérique Expérimental.

<sup>2</sup>For comparison purposes, 600 mm is approximately the mean yearly rainfall in Paris (France).

*Gardon d'Anduze watershed*



**Fig. 1** The *Gardon d'Anduze* watershed and its location in France. Discharges are measured in *Anduze* and cumulated rainfalls are measured by six rain gauges distributed over the watershed in order to obtain information about the spatial variability of rain

them into account. For example, the model that is used in the “Experimental Digital Watershed” project, and applied to the *Anduze* catchment [2], estimates, at each time step (hourly), the spatial distribution of the soil water content (with three parameters). In a subsequent step, for six appropriately defined hydrological meshes, two elements of the global discharge are estimated at the pixel resolution of the “Digital Terrain Model”: the subsurface lateral flow (water flowing underground laterally in the first few meters of soil and drained by the river network) and the saturated area runoff (water that flows on the surface because the soil is already saturated due to previous rainfall, thus unable to absorb additional water). Then a geomorpholog-

ical approach allows an estimation of the time necessary for underground or surface water to reach the river at any location. Six sub-catchments are thus defined, where discharge measurements are performed (one measurement station per catchment), three physical parameters are required to compute the total runoff for each sub-catchment, so that finally eighteen parameters are defined; two additional parameters, accounting for evapotranspiration and for water transfer, are required. The accurate estimation of those physical quantities is a major difficulty of that approach, because the inverse problem does not have a unique solution.

**Table 1** List of available events between 1994 and 2007 on the *Gardon d'Anduze* catchment (from EDW database). Events are identified by a number, and each row of the table reports the date of the event, its duration, the cumulated rainfalls over the duration of the event (averaged over the 6 rain gauges), the maximum peak discharge (averaged over half an hour), the maximum water level at the *Anduze* gauge station and the mean discharge over the duration of the event. The most intense event is number 19, whose prediction is one of the objectives of the present work

No.	Date	Duration (hours)	Cumulated rainfall (cm)	Maximum discharge (m <sup>3</sup> /s)	Maximum water level (m)	Mean flow (m <sup>3</sup> /s)
1	11/21–24/1994	35	18.9	491	3.71	181
3	10/4–5/1995	54	32	975	5.34	175
4	10/13–14/1995	92	19	864	5	108
8	11/10–12/1996	82	16	268	2.71	33
13	9/28–29/2000	46	18	800	4.80	74
19	9/9/2002	29	41	2742	9.71	511
22	9/24–25/2006	23	6	186	2.24	21
23	10/19–20/2006	55	20	1436	6.61	181
24	11/17/2006	34	6	275	2.75	65
25	11/20–23/2007	70	12	264	2.69	139
107	11/5–7/1997	74	27	624	4.20	225
108	11/26–27/1997	66	11	244	2.58	117
109	12/18–19/1997	104	38	985	5.37	327
117	10/20–21/1999	34	12	473	3.64	212
120	11/12–14/2000	71	13	279	2.77	110

### 2.2.2 Conceptual modeling

Conceptual modeling assumes that the rainfall-runoff relation may be described by global concepts, such as the reservoir concept. This approach has been applied to the *Gardon d'Anduze* catchment using the SCS model (Soil Conservation Service) [3]. It is considered that rainfalls fill a “soil reservoir” whose level is computed using a balance between infiltration and discharge. The reservoir is modeled as a first-order linear filter. Therefore, two parameters only are required: the capacity of the reservoir and its time constant. The soil reservoir depends also on the soil characteristics and is thus spatially distributed. The results of the simulations performed with the model depend strongly on the postulated initial level of the soil reservoir.

For all hydrological models addressing flash floods, the prominent variable is the spatially distributed rainfall. In the case of TOPMODEL, variable ranking shows [2] the major relevance of (i) distributed rainfalls, and (ii) soil moistures. This expert knowledge will be useful for the design of our statistical models, as shown in the “Model design” section.

Moreover, in both conceptual and physical models, rainfall information must be available; therefore, it is very difficult to forecast the future water level in the absence of rainfall forecast. Neural networks, which can learn functions based on rainfall and runoff time series, are thus good candidates for forecasting runoff at a predefined horizon.

### 2.3 Neural networks for flash flood forecasting

Although neural networks have been widely publicized for flood forecasting, the prediction of flash floods, in the absence of rainfall predictions, has been seldom investigated

by machine learning techniques. In [9], flash flood forecasting is performed by neural networks, radial basis functions networks and nearest-neighbors, in the framework of nonlinear, autoregressive models, taking into account past rainfall measurements, and past runoff measurements; the authors note that “all investigated MLP networks perform significantly worse when applied to the test data sets”. This overfitting phenomenon can probably be traced to the fact that no regularization mechanism was implemented during training. By contrast, early stopping is used for regularization in [10] and [11], but the scope of that research is long-term predictions. In [12, 13], flash floods are predicted by various methods including neural networks. The reported results are satisfactory probably due to the very small size of the basins (2 and 20 km<sup>2</sup>) which simplify the rainfall runoff relation; moreover they are not obtained on an independent test set: therefore, the results are likely to be unduly optimistic. In [14], even though generalization is assessed on an independent test set, the most intense event of the database is always included in the training set, so that the ability of the model to forecast reliably a more intense event is not demonstrated. The importance of such issues as complexity control, variable selection, independent testing, are not always acknowledged when machine learning techniques are used in hydrology. Some well-known techniques are summarized in [15]. In addition, many publications in the field are plagued by various problems (inefficient optimization techniques, overparameterized models, inappropriate performance indices, and failure to realize that neural networks are just another nonlinear regression technique); as a consequence, results on difficult problems such as the forecasting of extreme floods are often disappointing, although these are events that most urgently need fast, accurate predictions.

In the present work, we show that a rigorous model design procedure, involving model complexity control, variable selection, regularization, and independent testing, leads to satisfactory predictions of extreme events.

### 3 Problem statement and data processing

#### 3.1 Neural network models for flood forecasting

Time plays a functional role in the rainfall-runoff relation: physically, the flow at time  $t$  depends on previous rainfalls. Therefore, discrete-time dynamic models must be designed. In the following, we denote by  $T$  the sampling period of the physical quantity of interest  $y^p(kT)$ , hereinafter abbreviated as  $y^p(k)$ .

Denoting by  $y(k)$  the predicted value of the quantity of interest at discrete time  $kT$ , by  $\mathbf{u}(k)$  the vector of exogenous variables at time  $kT$ , and by  $g_{NN}$  the nonlinear function implemented by a feedforward neural network, the following simple input-output “neural” predictors can be designed [16]:

(i) recurrent neural networks

$$y(k) = g_{NN}(y(k-1), y(k-2), \dots, y(k-n), \mathbf{u}(k-1), \mathbf{u}(k-2), \dots, \mathbf{u}(k-m)))$$

where  $n$  and  $m$  are positive integers; recurrent neural networks predict the quantity of interest at time  $k$  from its past *predicted* values and from past measured values of the exogenous variables,

(ii) feedforward neural networks with time delays

$$y(k) = g_{NN}(y^p(k-1), y^p(k-2), \dots, y^p(k-n), \mathbf{u}(k-1), \mathbf{u}(k-2), \dots, \mathbf{u}(k-m))),$$

which predict the quantity of interest at time  $k$  from its past *measured* values and from past values of the exogenous variables.

It has been shown [16] that the first category of predictors is optimal if the modeled process is subjected to output noise (typically measurement noise), while the second category is optimal if the modeled process is subjected to state noise (typically unknown disturbances acting on the process itself). If both types of noise are present, recurrent models fed both with past predicted variables and past measured variables must be designed.

In the present study, the quantity of interest  $y^p$  is the water flow (or equivalently the water level), and the exogenous variables are the rainfalls and possibly additional factors as described below.

In addition, the purpose here is to forecast, at time  $kT$ , the water flow or water level at time  $(k+f)T$ , where

$f$  is a given integer. The prediction horizon is typically a few hours, while the sampling period is typically a few minutes. In order to perform such a forecast, a recurrent network might be used by running the predictor  $f$  times at time  $kT$ : that would require the availability of  $\mathbf{u}(k+1)$ ,  $\mathbf{u}(k+2)$ ,  $\dots$ ,  $\mathbf{u}(k+f)$ , i.e. the availability of rainfall forecasts on that horizon. This requirement precludes the use of recurrent predictors.

Therefore, the predictors designed in the present study are of the form

$$y(k+f) = g_{NN}^f(y^p(k), y^p(k-1), \dots, y^p(k-n), \mathbf{u}(k), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m)))$$

where  $g_{NN}^f$  is implemented as a feedforward neural network whose number of variables and number of hidden neurons depend on the forecasting horizon  $f$ .

Therefore, given a forecasting horizon  $f$ , one has to find the appropriate time window  $m$  for rainfalls, the appropriate time window  $n$  for past water levels, the relevant exogenous variables, together with the appropriate complexity in terms of number of hidden neurons, given the available data. We perform this task as described below in the “model and variable selection” subsection of the “model design” section.

#### 3.2 Data acquisition

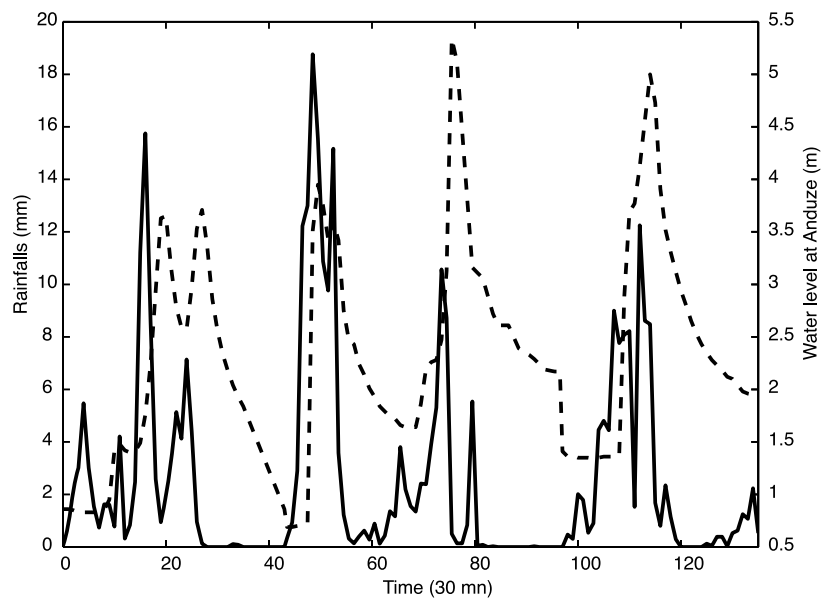
Rainfalls are measured by rain gauges, which are the most accurate sensors available at present. However, they provide local information, so that the heterogeneity of rainfalls is a major problem for flash floods. For instance, in the case of event 19, which is of particular interest in the present study, the cumulated rainfall was three times as large in *Anduze* as in *Soudorgues*, which is only fifteen kilometers away. In order to obtain a more global picture, radar acquisitions of rainfalls with a definition of 1 km<sup>2</sup> have been performed since 2002, but the number of flood events thus monitored in the catchment under investigation is still too small for reliable use in a machine learning approach.

The sampling period for water level measurements during flood events was 1 h before 2002 and has been 5 minutes since 2002. In the present work, a sampling period of 30 min was chosen, which is appropriate considering the 2–4 hours rise time of this catchment. For events that occurred before 2002, re-sampling was performed by linear interpolation.

#### 3.3 Database structure

The water level of the catchment under investigation is sampled all year long with a sampling period of one hour, but the sampling period is shortened to 5 min when flash floods occur. Therefore, the database was organized as follows: at

**Fig. 2** Concatenation of events 1, 3 and 4. *Solid line*: water level; *dashed line*: rainfall averaged over the six rain gauges



the end of each flood event, a non-flood (zero or weak rainfall) time period of 15 hours (more than twice the concentration time of the basin, i.e. the time that is necessary for water to flow from the most distal point of the basin to the catchment where the measurement is performed) was kept in the database after each event. The flood events were subsequently concatenated into a single time series (Fig. 2). Between two concatenated events, the non-flood zone is used simultaneously as rainfall for the end of the present event and as the rainfalls for the future event. A transition zone is also introduced between two adjacent events in order to decrease the outflow discontinuity between the end of an event (high discharge) and the beginning of the adjacent event (low discharge). The resulting database contains 1743 measurements of rainfalls and water levels at *Anduze*.

### 3.4 Typology of flood events

In order to perform (i) training, (ii) model and variable selection, and (iii) testing, three data sets are necessary: a training set, a validation set and a test set. They must be chosen appropriately. To this end, a typology of the available events was created as follows.

The 15 events listed in Table 1 were clustered manually into three categories depending on their intensity, time length and number of peaks: an event was classified as “very intense” if the maximum water level exceeded 4.5 m; it was classified as “long” if its duration exceeded 50 h; finally, single-peak events were separated from events with multiple water level peaks. This resulted in 8 categories, which are listed in Table 2.

As might be expected, classes D and E are empty: an intense and short event cannot have more than one peak, and nonintense but long events necessarily have several peaks.

**Table 2** List of classes and of their number of elements

Very intense	Long	Single-peak	Class and number of elements
Y	Y	Y	A (3)
Y	Y	N	B (1)
Y	N	Y	C (2)
Y	N	N	D (0)
N	Y	Y	E (0)
N	Y	N	F (5)
N	N	Y	G (3)
N	N	N	H (1)

## 4 Model design

As indicated in the previous section, the constraints of the problem lead naturally to choosing a model of the form

$$y(k+f) = g_{NN}^f(y^p(k), y^p(k-1), \dots, y^p(k-n), \mathbf{u}(k), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m))$$

where  $y$  is the estimated water level,  $y^p$  is the measured water level,  $\mathbf{u}$  is the vector of exogenous variables  $g_{NN}^f$  is the function implemented by a feedforward neural network with one layer of hidden neurons with tanh nonlinearities, and a linear output neuron.

As pointed out in section “hydrological method for flash flood modeling”, the vector of exogenous candidate variables includes the rainfalls from 6 rain gauges (spatially distributed rainfalls), and the soil moisture (Soil Water Index, given by the ISBA model [17]). The actual relevance of the candidate variables must be assessed by a variable selection procedure.  $n$  and  $m$  are positive integers that must also be

chosen by that procedure. The latter is described below, in the section entitled “model and variable selection”.

#### 4.1 Training

The usual least squares cost function was optimized by the Levenberg-Marquardt algorithm; the gradient of the cost function was computed by backpropagation.

#### 4.2 Regularization

Regularization was performed by early stopping on event 13, which appears to be prototypical of the available events. For more details on training and regularization, see e.g. [18].

#### 4.3 Model and variable selection

Model and variable selection aim at finding the appropriate number of hidden neurons, the appropriate values of  $m$  and  $n$ , and the relevance of the soil moisture variable. The procedure was a “leave-one-sequence-out” procedure, inspired by the usual leave-one-out cross-validation method. First, sequence 13, which is used for early stopping (see section “Regularization”), and sequence 19, which is used for testing (see section “Test”) are set apart. From the set of the remaining 13 sequences, each sequence in turn is extracted, a predictor is trained on the other 12 sequences, and its performance on that sequence is computed. The leave-one-sequence-out score is the average of the sum of squared prediction errors, over all left-out sequences. However, since the accuracy of the predictions of intense sequences is more critical than the accuracy of the predictions of moderately intense ones, the following variant was found more successful: instead of leaving out each sequence in turn, the least intense nine sequences were always kept in the training set, and one sequence among the most intense four sequences (3, 4, 109 and 23) was left out in turn.

At the end of the procedure, the combination of  $n$ ,  $m$ , of the number of hidden neurons, and of the presence or absence of the soil moisture variable, that provides the best leave-one-sequence-out score was selected. Such a model was subsequently trained with all available sequences except the early stopping sequence and the test sequence.

#### 4.4 Test

The quality of the resulting model was assessed on the test sequence, i.e. event 19 (class C), whose intensity made history in 2002. Several quality criteria can be considered.

The coefficient of determination of the regression, known to hydrologists as Nash’s criterion, is defined as

$$R^2 = 1 - \frac{\sum_{k \in \text{test sequence}} (y^p(k) - y(k))^2}{\sum_{k \in \text{test sequence}} (y^p(k) - \bar{y}^p)^2}$$

where  $\bar{y}^p$  is the mean of the observations.  $R^2 = 1$  if the model predicts the test data with perfect accuracy, and  $R^2 = 0$  if the model simply predicts the mean of the observations.

For forecasting, the persistence coefficient  $C_p$  is of special interest. For a forecasting horizon  $f$ , it is defined as [19]:

$$C_p = 1 - \frac{\sum_{k \in \text{test set}} (y^p(k + f) - y(k + f))^2}{\sum_{k \in \text{test set}} (y^p(k) - y^p(k + f))^2}$$

$C_p$  ranges from  $-\infty$  to 1.  $C_p = 1$  if the model predicts the test data with perfect accuracy, and  $C_p = 0$  if the predictor is perfectly dumb, i.e. if it predicts that the future value of the quantity of interest will be equal to the present one. A predictor with negative  $C_p$  performs worse than a dumb predictor.

### 5 Results

In order to obtain baseline results, static predictors of the form

$$y(k + f) = g_{NN}^f(\mathbf{u}(k), \mathbf{u}(k - 1), \dots, \mathbf{u}(k - m)),$$

where  $\mathbf{u}(k)$  is the vector of exogenous variables, will first be described. Dynamic predictors will be considered in the subsequent section.

#### 5.1 Static predictors

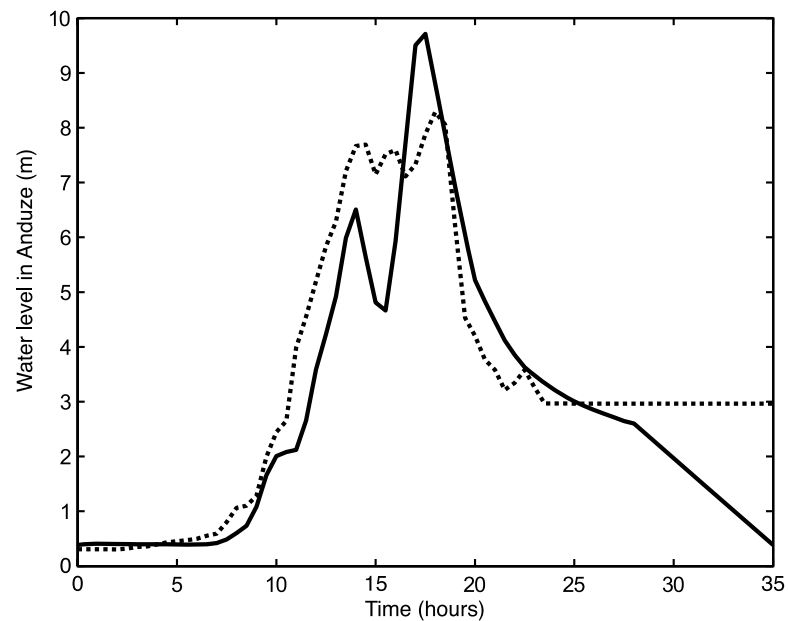
The procedure described in section “Model and variable selection” showed that the best leave-one-sequence-out scores were obtained by networks having 2 hidden neurons, with  $m = 11$  (12 rainfall values, sampled at 30-min intervals) for thirty-minute ahead forecasts,  $m = 7$  for one-hour ahead forecasts, and  $m = 3$  for two-hour ahead forecasts. Soil moisture was not found to be relevant: it was not included in vector  $\mathbf{u}$ . An additional input variable was found to be relevant, and was therefore included in vector  $\mathbf{u}$ : the rainfall accumulation from the beginning of the event.

For two forecasting horizons out of three, the persistence coefficient was negative, showing that the predictor is less efficient than a dumb predictor. Figure 3 shows the two-hour ahead forecasts of event 19 and Table 3 the values of the quality criteria.

**Table 3** Quality criteria for static network (event 19)

Forecasting horizon	Thirty-minutes	One hour	Two hours
$R^2$ (Nash criterion)	0.83	0.79	0.76
$C_p$	-2.68	-0.32	0.45

**Fig. 3** Limnograph of event 19. Solid line: observation; dotted line: two-hour ahead forecast by a static network.  $R^2$  and  $C_p$  are 0.76 and  $-0.45$  respectively



## 5.2 Dynamic predictors: feedforward neural networks with time delays

In the present section, we consider predictors of the form

$$y(k+f) = g_{NN}^f(y^p(k), y^p(k-1), \dots, y^p(k-n), \mathbf{u}(k), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m)).$$

As in the case of static predictors, the procedure described in section “Model and variable selection” resulted in the selection of  $m$  and of the number of hidden neurons, and in the assessment of the relevance of the candidate exogenous variables. In addition, the appropriate value of  $n$  was selected by the procedure.

The results are summarized as follows:

- $n = 1$  for all values of  $f$  (present value and one past value of water level),
- 2 hidden neurons for thirty-minute and one-hour ahead forecasts, 5 hidden neurons for two-hour ahead forecasts,
- $m = 4$  for thirty-minute ahead forecasts,  $m = 5$  for one-hour and two-hour ahead forecasts,
- soil moisture and cumulated rainfalls from the beginning of the event were found irrelevant.

These predictors are much more efficient than the static predictors. Table 4 summarizes the values of the coefficient of determination  $R^2$  and of the persistence coefficient  $C_p$  for various forecasting horizons, for the most intense, single-peak event (19) and for a more moderate, multiple-peak event (107, class F). The results on event 107 were obtained by exactly the same procedure as for event 19, except that event 19 was in the training set and that event 107 was used

**Table 4** Summary of test results obtained on two different events at three different forecasting horizons

Event	Forecasting horizon	$R^2$	$C_p$
19	1/2 h	0.98	0.60
	1 h	0.92	0.51
	2 h	0.86	0.68
107	1/2 h	1.00	0.72
	1 h	0.99	0.77
	2 h	0.98	0.76

neither for training nor for validation. In both cases the results are very satisfactory. For event 19, the prediction of the peak is performed two hours before its occurrence, but the rainfall peak occurs two hours before the peak of the discharge; this explains in part why that particular event is very difficult to predict accurately. To the best of our knowledge, no forecasting method has ever achieved a comparable accuracy on such an extremely intense flash flood as event 19.

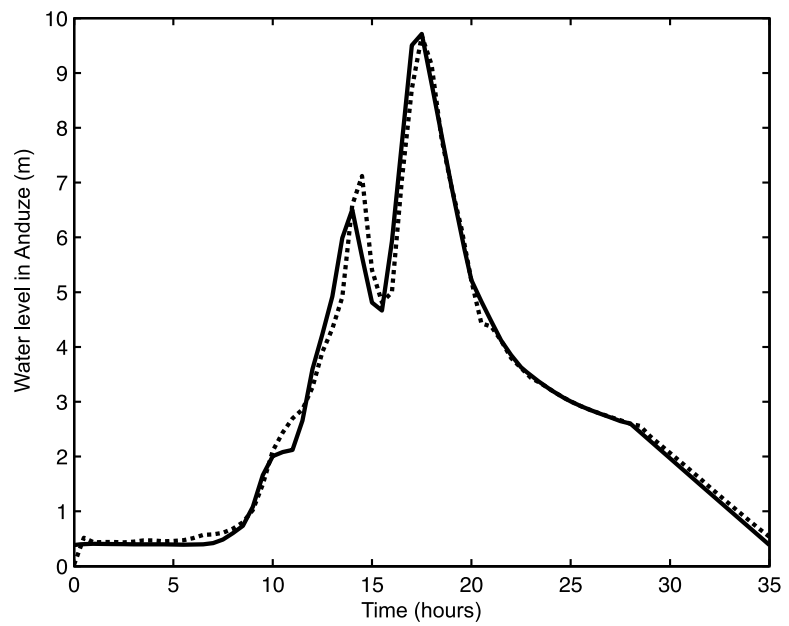
Figures 4a, b, c, and 5a, b, c show the observed and forecast limnographs for the above events.

## 5.3 Assessment of the relevance of the predictor with respect to present vigilance thresholds

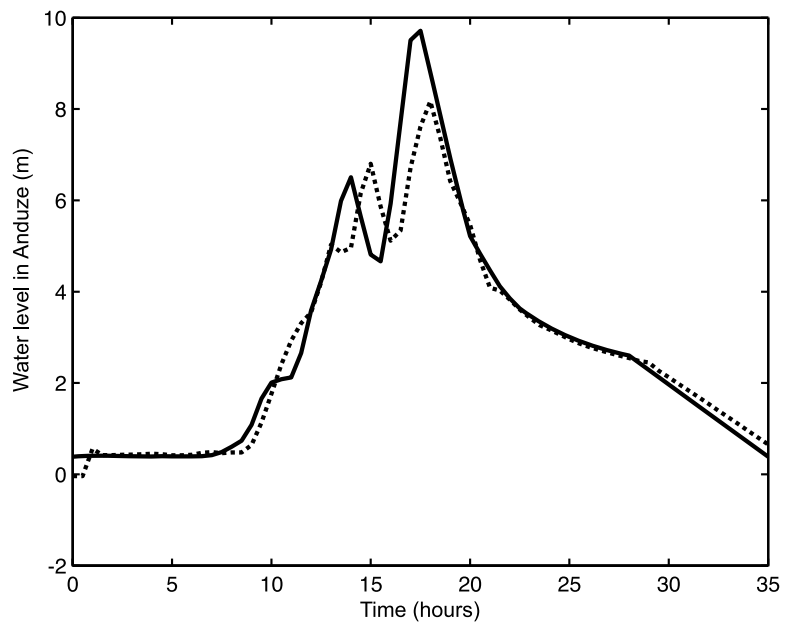
From the end user’s point of view, it is important to assess the behavior of the predictor in the framework of the present flood warning system. Four vigilance levels, defined by the



**Fig. 4** **a.** Limnograph of event 19. *Solid line:* observation; *dotted line:* thirty-minute ahead forecast. **b.** Limnograph of event 19. *Solid line:* observation; *dotted line:* one-hour ahead forecast. **c.** Limnograph of event 19. *Solid line:* observation; *dotted line:* two-hour ahead forecast



(a)



(b)

SCHAPI, are currently in use:

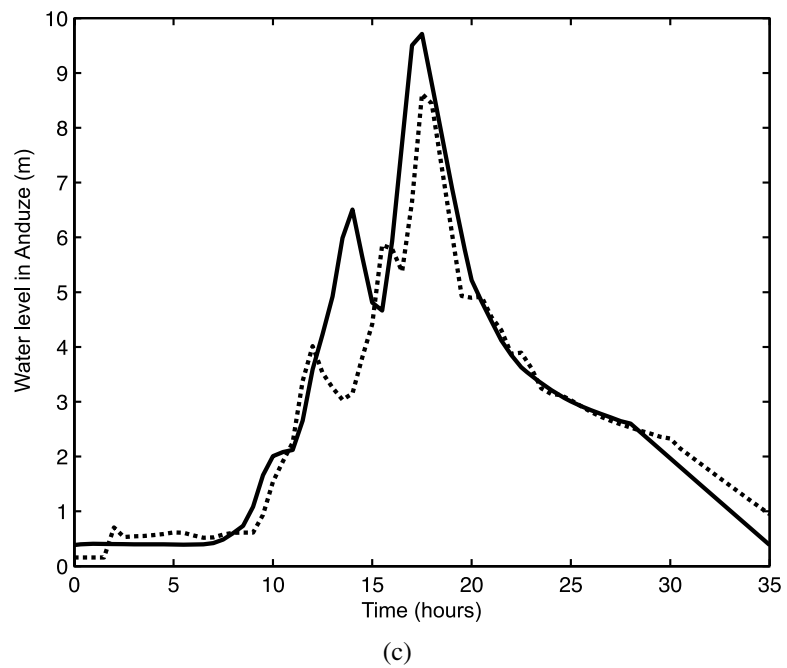
- Red: major risk of high water level with serious threats to people and property,
- Orange: risk of high water level with considerable overflow liable to have a significant effect on daily life and on the safety of people and property,
- Yellow: risk of high or rapidly rising water not involving significant damage but requiring particular vigilance in the case of seasonal and/or outside activities,

- Green: no particular vigilance required.

For the *Gardon* at *Anduze*, the thresholds are defined as follows by the local Flood Forecasting Service,

- discharge < 800 m<sup>3</sup>/s: green
- 800 m<sup>3</sup>/s < discharge < 1600 m<sup>3</sup>/s: yellow
- 1600 m<sup>3</sup>/s < discharge < 2800 m<sup>3</sup>/s: orange
- discharge > 2800 m<sup>3</sup>/s: red.

Fig. 4 (Continued)



**Table 5** Forecasted vigilance level for all events and all forecasting horizons. Boldface characters highlight misclassifications

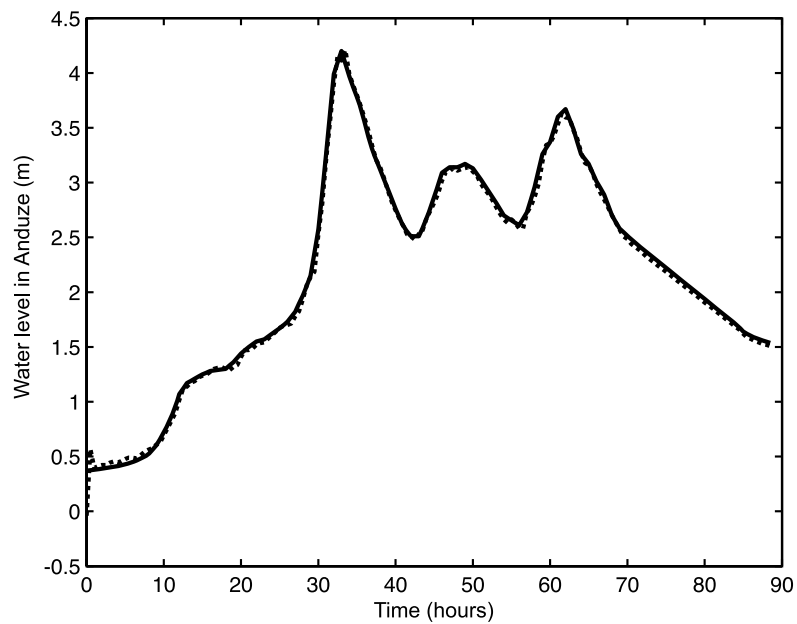
N° event	Observed Vigilance level	$f = 30$ min Vigilance level	$f = 1$ h Vigilance level	$f = 2$ h Vigilance level	$f = 3$ h Vigilance level	$f = 4$ h Vigilance level	$f = 5$ h Vigilance level
01	Green	Green	Green	Green	<b>Yellow</b>	Green	Green
03	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
04	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
08	Green	Green	Green	Green	Green	Green	Green
19	Orange	Orange	Orange	Orange	Orange	Orange	<b>Yellow</b>
22	Green	Green	Green	Green	Green	Green	Green
23	Yellow	Yellow	Yellow	Yellow	<b>Green</b>	<b>Green</b>	<b>Green</b>
24	Green	Green	Green	Green	Green	Green	Green
25	Green	Green	Green	Green	Green	Green	Green
107	Green	Green	Green	Green	Green	Green	Green
108	Green	Green	Green	Green	Green	Green	Green
109	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
117	Green	Green	Green	Green	Green	Green	<b>Yellow</b>
120	Green	Green	Green	Green	Green	Green	Green

Therefore, predicting whether a given vigilance threshold will be reached may be viewed as a classification problem. This task is much less demanding than the prediction problem addressed in the present paper: the classification may be correct although the predicted water level, or the timing of the peaks, may be quite erroneous.

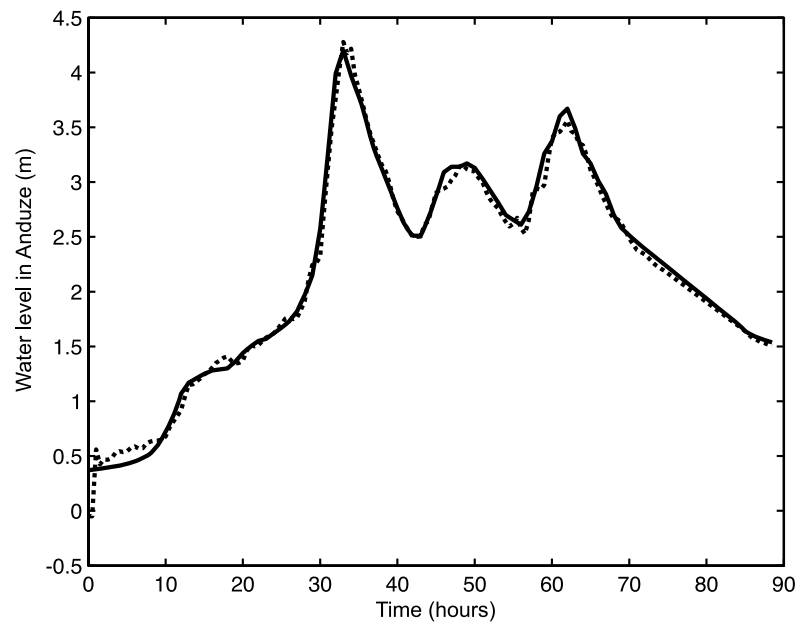
For the different prediction horizons  $f$ , the predicted and the observed vigilance levels can be compared (*a posteriori* because the flood warning system was not available in 2002), so that the ability of the model to forecast the risk of

the event, in the framework of the present warning system, can be assessed. Table 5 presents the results of this analysis. The agreement between the vigilance forecasts and the issued vigilance warning is very good: for 10 events out of 14, the risk is correctly classified for all prediction horizons. When misclassifications occur, it is only for high forecasting horizons ( $\geq 3$  h), with two false negatives. The 2002 event (event 19, with a discharge peak value of 2742 m<sup>3</sup>/s) is classified in orange vigilance but it is very close to the red level.

**Fig. 5** **a.** Limnograph of event 107. *Solid line:* observation; *dotted line:* thirty-minute ahead forecast. **b.** Limnograph of event 107. *Solid line:* observation; *dotted line:* one-hour ahead forecast. **c.** Limnograph of event 107. *Solid line:* observation; *dotted line:* two-hour ahead forecast



(a)



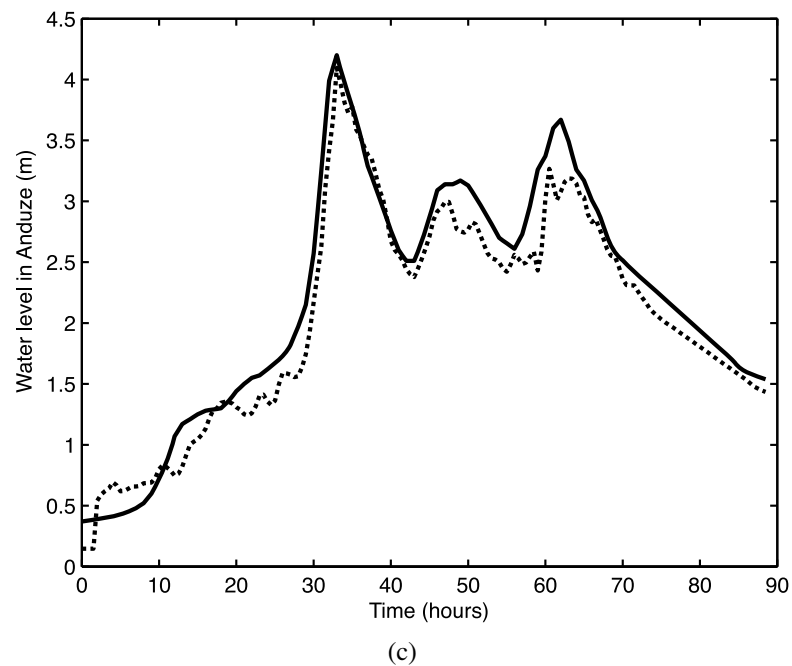
(b)

## 6 Conclusion

Forecasting flash floods in populated areas is an important and difficult task, which is traditionally approached by physics-based models or by hydrology-based conceptual models. In the present paper, we have shown that such forecasts can be made with very satisfactory accuracy by machine-learning methods, in the absence of rainfall predictions, provided an appropriate methodology is used. Leave-one-sequence-out cross-validation was used for model and variable selection in the framework of feedforward mod-

els with time delays. Regularization was performed by early stopping. Blind testing was performed on very intense events, which caused large damages. The accuracy of the forecasts, for the water level as well as for the vigilance level, opens the way to early warnings of the population in the area under investigation. Future research will involve the application of the methodology to other flash-flood areas in France. In addition, support vector regression and dynamic modeling with kernel methods [20], on the same data, will be investigated.

Fig. 5 (Continued)



**Acknowledgements** This work was funded in the framework of Experimental Digital Watershed (BVNE), a program initiated by the SCHAPI of the French Ministry of Environment (MEEDADT).

We are very grateful to MM. Pierre Roussel-Ragot, Georges-Marie Saulnier, Bernard Vayssade and Marc Vinches for helpful discussions.

We also wish to thank M. Dominique Bertin for his highly effective collaboration in the design and implementation of the Neural Network simulation tool: RnfPro.

## References

- Oudin L, Hervieu F, Michel C, Perrin C, Andréassian V, Anctil F, Loumagne C (2005) Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2: toward a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *J Hydrol* 303:290–306
- Le Lay M, Saulnier G-M (2007) Exploring the signature of climate and landscape spatial variabilities in flash flood events: case of the 8–9 September 2002 Cévennes-Vivarais catastrophic event. *Geophys Res Lett* 34:L13401
- Bouvier C, Ayrat P-A, Brunet P, Crespy A, Marchandise A Martin C (2006) Recent advances in rainfall-runoff modelling: extrapolation to extreme floods in Southern France. In: International workshop on hydrological extremes. Observing and modelling exceptional floods and rainfalls rende (CS), 3–4 May 2006
- Estupina-Borrell V, Chorda J, Dartus D (2005) Flash-flood anticipation. *CR Geosci* 337:1109–1119
- Bouttier F (2003) The Arome mesoscale project. In: Proceedings of the ECMWF, September 2003
- Delrieu G, Ducrocq V, Gaume E, Nicol J, Payrastré O, Yates E, Kirstetter P, Andrieu H, Ayrat P, Bouvier C, Creutin J, Livet M, Anquetin S, Lang M, Neppel L, Obled C, Parent-du-Chatelet J, Saulnier G-M, Walpersdorf A, Wobrock W (2004) The catastrophic flash-flood of 8–9 September 2002 in the Gard region, France: a first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory. *J Hydrometeorol* 6:34–52
- Gaume E, Livet M, Desbordes M, Villeneuve J-P (2004) Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999. *J Hydrol* 286:135–154
- Sauvagnargues-Lesage S, Ayrat P-A (2007) Using GIS for emergency management—a case study during the 2002 and 2003 South East flooding in France. *Int J Emerg Manag* 4(4):682–703
- Piotrowski A, Napiorkowski J-J, Rowinski P-M (2006) Flash flood forecasting by means of neural networks and nearest neighbour approach—a comparative study. *Nonlinear Process Geophys* 13:443–448
- Coulibaly P, Anctil F, Rasmussen P, Bobée B (2000) A recurrent neural networks approach using indices of low frequency climatic variability to forecast regional annual runoff. *Hydrol Process* 14:2755–2777
- Coulibaly P, Anctil F, Bobée B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J Hydrol* 230:244–257
- Sahoo G-B, Ray C (2006) Flow forecasting for a Hawaii stream using rating curves and neural networks. *J Hydrol* 317:63–80
- Sahoo GB, Ray C, de Carlo E-H (2006) Use of neural networks to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *J Hydrol* 327:525–538
- Schmitz G-H, Cullmann J (2008) PAI-OFF: a new proposal for online flood forecasting in flash flood prone catchments. *J Hydrol* 360:1–14
- Giustolisi O, Laucelli D (2005) Improving generalization of artificial neural networks in rainfall-runoff modelling. *Hydrol Sci J* 50:439–457
- Nerrand O, Urbani D, Roussel-Ragot P, Personnaz L, Dreyfus G (1994) Training recurrent neural networks: why and how? An illustration in process modeling. *IEEE Trans Neural Netw* 5:178–184
- Noilhan J, Mahfouf J-F (1996) The ISBA land surface parameterisation scheme. *Glob Planet Change* 13:145–159
- Dreyfus G (2005) Neural networks, methodology and applications. Springer, Berlin
- Kitadinis P-K, Bras R-L (1980) Real-time forecasting with a conceptual hydrologic model: 2 applications and results. *Water Resour Res* 16:1034–1044
- Qu H-n, Oussar Y, Dreyfus G, W Xu (2009) Regularized recurrent least squares support vector machines. In: International joint conference on bioinformatics, systems biology and intelligent computing, Shanghai