

Complexity selection of a neural network model for karst flood forecasting: The case of the *Lez* Basin (southern France)

Line Kong A Siou^{1,2}, Anne Johannet¹, Valérie Borrell², Séverin Pistre²

¹Ecole des Mines d'Alès, CMGD, 6 avenue de Clavières, 30319 Alès Cedex, France

²Université Montpellier II, Hydrosiences Montpellier, Place E. Bataillon, 34095 Montpellier Cedex 5, France

Abstract

A neural network model is applied to simulate the rainfall-runoff relation of a karst spring. The input selection for such a model becomes a major issue when deriving a parsimonious and efficient model. The present study is focused on these input selection methods; it begins by proposing two such methods and combines them in a subsequent step. The methods introduced are assessed for both simulation and forecasting purposes. Since rainfall is very difficult to forecast, especially in the study area, we have chosen a forecasting mode that does not require any rainfall forecast assumptions. This application has been implemented on the *Lez* karst aquifer, a highly complex basin due to its structure and operating conditions. Our models yield very good results, and the forecasted discharge values at the *Lez* spring are acceptable up to a 1-day forecasting horizon. The combined input selection method ultimately proves to be promising, by reducing input selection time while taking into account: i) the model's ability to accommodate nonlinearity, and ii) the forecasting horizon.

1. Introduction

The need to develop accurate prediction tools for flood events has been underscored by the recent occurrences of catastrophic floods in the northern Mediterranean region, where such events caused at least 4,566 fatalities and €29,136 million in property damage over the period 1990-2006 (Llasat *et al.*, 2010). Between July and September 2002, southern France, Italy and Spain were all affected by fast, or "flash", floods. Flooding is France's leading natural hazard and responsible for major damage; the flooding risk encompasses 20,000 km² of land area, with some 4.5 million residents (8% of the French population) living in flood zones (Source: *France's Ministry of Ecology, Energy and Sustainable Development*). In this country, a number of major flood events have taken place, including *Nîmes* (1988), *Vaison-la-Romaine* (1991), *Aude* (1999), *Gardons* (2002), *Arles* (2003) and *Var* (2010).

In seeking to mitigate flood risks, the Ministry of Ecology, Energy and Sustainable Development created in 2003 the National Center for Flood Forecasting and Warning: SCHAPI (*Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*), along with 22 regional forecasting services (*Services de Prévision des Crues*), which were placed in charge of the "Vigicrue" flood monitoring service. This system is based on a round-the-clock watch of the main rivers and generates warning messages broadcast to Civil Security authorities and the local citizenry in the form of an Internet map (<http://www.vigicrues.ecologie.gouv.fr>). Four levels of risk are evaluated by tools ranging from simple hydrological analyses to detailed and spatialized hydrological / hydraulic forecasting models spanning a 20,000-km network of river streams. At present, SCHAPI identifies the four following types of floods (for which forecasting techniques still require significant improvements): slow plain floods, fast floods, estuarine floods and water table floods.

Floods in karst aquifers have been investigated more recently; the pilot basin selected was the *Lez* aquifer, which is the topic for the present article. The research work displayed herein has been performed in collaboration with the SCHAPI Unit within the framework of the ongoing project focusing on karsts.

This paper proposes a methodology devoted to the input selection of a neural network model applied to the simulation and forecasting of daily discharges from a karst spring. The first part will discuss the major issues involved in studying karst springs, in order to better understand their behavior, particularly during floods. The case of Mediterranean flash floods, which have caused numerous casualties, will be developed primarily for karst models and flood models. Section 4 will describe neural networks, as an interesting method for karst and flood forecasting by taking advantage of their potential for universal and parsimonious approximations. The key issue of variable selection will be addressed in order to avoid overfitting; moreover, two input selection methods will be presented, with the aim of choosing a rainfall width that explains spring discharge. Based on both these methods, a procedure offering the nonlinear capability of neural networks, while reducing network design time, will be proposed. The paper's final section concerns application of this method to the complex karst aquifer at *Lez* spring. Both simulations and forecasts will be discussed to demonstrate that the method can accurately forecast floods.

2. Context and larger societal issues

2.1 Karst basins

A karst basin is composed of two distinct elements, each of which behaves differently: the surface watershed, and the aquifer. These two components are not necessarily vertically aligned from a geographical standpoint. Moreover, the occurrence of an intense rainfall event is capable of causing surface flooding over a portion of the non-karstified basin, in addition to a rise in the groundwater level over the portion containing karst. Should these two sudden rises in water level coincide, the underground action could exacerbate the surface flood. It is necessary therefore to fully understand the hydrodynamic behavior of both basin components, i.e. karst and non-karst, in order to determine the total extent of flooding. Each karst aquifer is structured differently, as a result of the dissolution of carbonate rock (Bakalowicz, 2005), with different types of underground flows able to coexist. Water flowing within larger drains can reach velocities of several hundreds of meters per hour, while at the other extreme, water flowing through the crack matrix travels much more slowly (with a permeability on the order of 10^{-5} or 10^{-7}). In light of this discrepancy, karst aquifers exhibit nonlinear hydrodynamic behavior that might also include threshold effects, which are naturally difficult to quantify and forecast.

Water infiltration deep into the ground may be either diffused or localized in swallow holes, thus requiring knowledge of the precise precipitation location in order to characterize the karst recharge. Given the interactions between groundwater and surface water flows, several authors (Jourde *et al.*, 2007; Maréchal *et al.*, 2008; Bailly-Comte *et al.*, 2009) have underscored the importance, along with the difficulties, encountered in modeling and forecasting the flooding that occurs in karst aquifers. Specifically, the karstic portion of the basin is capable of either intensifying the surface flood (via swallow holes found in border springs) or mitigating it, in which case the karst acts like a flood control reservoir. A greater understanding of the hydrodynamic behavior of karst is therefore critical to improving the quality of flood forecasts and helping protect the local population.

2.2 Mediterranean floods

The northern Mediterranean zone is especially concerned by flooding due to its unique climate, which is subject to intense rainfall during the autumn: daily precipitation can reach as high as 650 mm, as demonstrated during the 2002 event in southeastern France. Such high-volume rainfall events, referred to as *Cévenol episodes* in the south of France, have had catastrophic consequences (over 100 fatalities during the past two decades in

southeastern France, and €1.2 billion in property damage for the sole event on September 9th, 2002 (Gaume and Bouvier, 2004; Le Lay and Saulnier, 2007). Moreover, the recent event that took place in June 2010 in the *Var* (France) caused 25 deaths and €600 million in direct damage (www.keraunos.org: *French Observatory of Tornadoes and Severe Thunderstorms*).

Rainfall forecasts for the area under investigation lack sufficient accuracy, given that the rain generation process itself has not yet been clearly understood. It remains a difficult task to accurately forecast when or where the next storm will stabilize, even for the most recent meteorological models (such as AROME (Bouttier, 2003)), due to the "backward regenerative system" that feeds the storm with water from the Mediterranean Sea (Delrieu *et al.*, 2005). Flood forecasting models are therefore implemented without accurate rainfall forecasts.

In addition, the issue of karstic basin flood forecasts is not limited to the *Lez* aquifer since 35% of the European territory, primarily the Mediterranean zone, is covered by limestone outcrops.

2.3 Karst modeling and flood forecasting

2.3.1 Karst modeling

Several approaches have been proposed to model karsts and simulate spring hydrographs (see reviews by White, 2002; Bakalowicz, 2008). Reservoir models have been used effectively to simulate spring discharge and flow rate in combination with rainfall data. Several authors (Fleury *et al.*, 2007; Fleury *et al.*, 2008; Jukić and Denić-Jukić, 2009) applied reservoir models to karst-based systems. From knowledge of the overall behavior of karst aquifers, the composition of such models includes several connected reservoirs representing contributions of the saturated zone, the low-infiltration zone and the rapid-infiltration zone, respectively.

Physical models are intended to depict the physical characteristics of karst aquifers. This type of modeling requires extensive system knowledge, which may not always be available due to system heterogeneity. More specific models have thus been developed for karsts. Double-permeability models featuring fracture permeability and conduit permeability have been introduced, for example, by Mohrlök and Teutsch (1997), Cornaton and Perrochet (2002) and Long (2009). These models consider an average distribution for both the matrix and conduits. More recently, discrete models have taken into account an explicit conduit/matrix system identified through exploration and/or the physical process of conduit development (Mohrlök and Sauter, 1997; Jaquet *et al.*, 2004; Kaufmann and Romanov, 2008). The models developed are thus specific to each aquifer and not immediately applicable to a karst aquifer.

A third type of model is based on a systemic approach, as initiated by (Mangin, 1984). The karst is considered as a system able to both transform an input (rainfall) into an output (discharge) and evaluate the input-output

relation using mathematical functions. The systemic approach yields "black box" models based on deconvolutions, functions or machine learning techniques (Thiery *et al.*, 1983; Karam, 1989; Fleury, 2005; Fleury *et al.*, 2007). Such models offer satisfactory results when approximating overall karst system behavior (Larocque *et al.*, 1998; Labat *et al.*, 1999); furthermore, this approach may be conducted on any aquifer without specific system knowledge. Subsequent to this approach, neural networks have been used for karst modeling by Johannet *et al.* (1994) and Kurtulus and Razack (2007) due to their nonlinear capabilities. The present paper has adopted this particular approach.

2.3.2 Flood forecasting

In the context of real-time flood forecasting, a distinction must be drawn between *simulations*, in which the outflow at time t is estimated from rainfall at the same time t and earlier, and *forecasts*, for which outflow is computed ahead of time. In general, forecasts are obtained by running simulating models using forecasted rainfall values. In this study context, most models studied are distributed physical models like SHE (Beven *et al.*, 1980; Abbott *et al.*, 1986), TOPMODEL (Beven *et al.*, 1984), TOPKAPI (Liu and Todini, 2002) or TOPODYN (Bartholmes and Todini, 2005; Vincendon *et al.*, 2010), which is a flash flood-dedicated version of TOPMODEL. In France, SCHAPI uses distributed models available on the ATHYS platform (Marchandise, 2007). Global conceptual models, such as GR, have also been used for forecasting purposes (Formigué and Lavabre, 2005; Javelle *et al.*, 2008). These models however do not take into account the karst contribution. Recent conceptual models, developed specifically for karst aquifers, have been introduced (Maréchal *et al.*, 2008).

Since water outflow is due to rainfall, outflow forecasting methodologies require either rainfall forecasts or *ad hoc* assumptions regarding future rain events, such as null or constant rainfall. Cloke and Pappenberger (2009) indicated that many forecasting systems were moving towards the use of climatic forecasts (Numerical Prediction Model), thus coupling a climatic model with a forecasting model (Bartholmes and Todini, 2005). Cloke and Pappenberger (2009) pointed out that a major consideration focuses on improving the Numerical Prediction Model.

In addition, the null or constant rainfall hypothesis is unsatisfactory in the case of the Mediterranean rainfalls studied herein, due to their greater temporal and spatial heterogeneity. For this reason, black box-inspired models provide an effective means of investigation. Following this approach, Toukourou *et al.* (2009) showed that neural networks could be successfully applied to forecasting floods without the availability of rainfall forecasts for a fast, but predominantly non-karst, watershed in the Mediterranean zone.

3. Neural networks for flood forecasting

3.1 State-of-the-art assessment

One alternative to physical models consists of utilizing available data in order to build models by machine learning, which for this study involves neural networks. This step will alleviate computational burdens while freeing model designers from the limitations imposed by physical modeling whenever the physical phenomena are too complex or their parametric estimation too difficult.

Studies on rainfall-runoff identification by neural networks were first carried out by Halff *et al.* (1993) and Johannet *et al.* (1994) using a multilayer perceptron. Many studies have used similar models and compared them with the classical rainfall-runoff model (Hsu *et al.*, 1995; Zealand *et al.*, 1997; Sajikumar and Thandaveswara, 1999; Govindaraju, 2000). The most frequently cited advantage of neural networks in hydrology is their ability to model the complexity of the rainfall-runoff relation thanks to their property of universal approximation (Hornik *et al.*, 1989). According to the authors, neural network models provide at the very least similar results to other models. In addition, they noted that such models are capable of managing missing data and offering real-time training.

In most cases, neural networks are used as black box models, yet Johannet *et al.* (1994) proposed transforming the black box into a transparent box, a step that includes *a priori* knowledge in the model architecture. Johannet *et al.* (2008b) observed hidden neuron outputs as the evolution in physical variables, e.g. an estimation of potential evapotranspiration.

Before 2000, research work was primarily performed using static networks and the back-propagation learning rule (Rumelhart *et al.*, 1986); currently however, several authors have presented the advantages associated with dynamic modeling (Coulibaly *et al.*, 2000a; Chiang *et al.*, 2004) as well as the Levenberg-Marquardt learning rule (Coulibaly *et al.*, 2000b). Hydrological data contain special statistical properties with little resemblance to a normal distribution, i.e. high skewness and heteroscedasticity. Extra care had to be taken when normalizing the input and output datasets, e.g. by choosing a specific transformation to reduce these irregularities (Sudheer *et al.*, 2003). Inputs must also be chosen with precaution: data on rainfall, outflow at previous times and sometimes temperature and evapotranspiration are typically available. Several authors have analyzed the relative importance of inputs using correlations (Coulibaly *et al.*, 2000b; Sudheer *et al.*, 2002), while others have posited that the nonlinearity of these relations justifies adding other, less correlated inputs (Dae-II and Young-Oh, 2005).

A major component of this body of work focuses on presenting how neural models are more effective than others (either physical or conceptual) without necessarily investigating in depth the issues surrounding the neural method. Yet the generalization capacity of the neural network is often cited as the main issue. While this has been clearly grasped by Dae-II and Young-Oh (2005), it remains a difficult problem due to: i) the dynamic properties of the rainfall-runoff relation; ii) the specific statistical properties of the outflow and rainfall signals; and iii) the significant level of noise inherent in the data. To overcome these difficulties, regularization methods, such as "early stopping", need to be introduced. The application of early stopping entails evaluating model quality on the basis of a dataset independent of the calibration set: this "stopping set" serves to prevent overtraining (Coulibaly *et al.*, 2000b).

In several cases, the database is too small to build two sets having the same statistical properties (one for training, the other for stopping and validation); it can thus be observed that stopping specializes the model on the stopping set. Model quality is therefore likely to be unduly optimistic. In order to achieve "early stopping", a third independent set must then be introduced (called the "test set", Kurtulus and Razack, 2007).

Concerning flash floods in particular, forecasting was performed by means of neural networks, radial basis function networks and nearest-neighbors, all within the framework of nonlinear, autoregressive models, which take into account past rainfall measurements and past runoff measurements (Piotrowski *et al.*, 2006). The authors note that "all investigated multilayer perceptron networks perform significantly worse when applied to the test data sets". In Sahoo *et al.* (2006) and Sahoo and Ray (2006), flash floods are forecasted by various methods including neural networks. The reported results are satisfactory, due most likely to the very small size of the basins (2 and 20 km²), which simplifies the rainfall-runoff relation; moreover, results have not been obtained on an independent test set. In Schmitz and Cullmann (2008), the generalization step is assessed on an independent test set, but the most intense event of the database is always included in the training set. The importance of such issues was studied by Toukourou (2009), who employed cross-validation (Dreyfus *et al.*, 2004) in order to select model complexity and variables.

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 forecasts. A method is proposed herein that, by capitalizing on the correlation analysis proposed by Mangin (1981) and cross-validation, is less time-consuming than cross-validation applied on its own.

3.2 Universal approximation

The interested reader is referred to Dreyfus *et al.* (2004) for the fundamentals on neural networks. Only the main properties are presented herein. The multilayer perceptron is a feedforward neural network with one hidden layer of N_c sigmoid neurons and a linear output neuron, as shown in Figure 1. This set-up acts as a universal approximator (Hornik *et al.*, 1989) and is nonlinear relative to both its input and parameters. This model is also parsimonious in comparison with other statistical, nonlinear models (Barron, 1993).

The proposed model is intended, at discrete time kT ($k \in \mathbb{N}^+$, where T is the sampling period), to forecast or simulate the outflow at time $(k+f)T$, where f is the forecasting horizon ($f \in \mathbb{N}^+$).

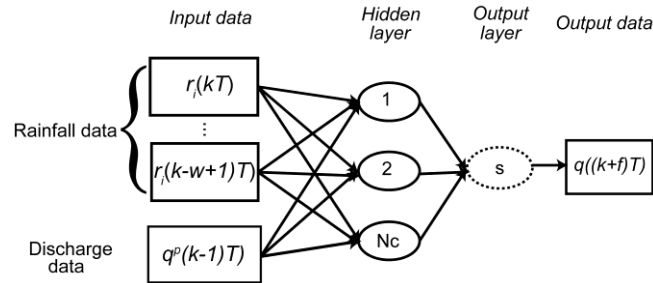


Figure 1: Rainfall-runoff model using a neural network

By denoting the forecasted outflow value as $q(k)$, the observed outflow (i.e. process output) as $q^p(k)$, the rainfall vector as $\mathbf{r}(k)$ and the nonlinear function implemented by a feedforward neural network as g_{NN} , the input-output "neural" model has been designed, based on Nerrand *et al.* (1993), as follows:

$$(a) \quad q(k) = g_{NN}[q^p(k-1), q^p(k-2), \dots, q^p(k-n), r(k), r(k-1), \dots, r(k-w+1)]$$

Rainfall information was conveyed to the network as a sliding window of width w , whose optimal value was chosen as described in Section 3.4. The information on past outflows was provided by the outflow for the previous discrete time $q^p(k-1)$. Even though this architecture is static (i.e. if all input variables are constant, then the output is also constant), it still allows simulating a dynamic process by the presence of measured discharge as an input variable. This architecture is thus known as "directed" since the model is being directed by measurements of previous outflows. Nerrand *et al.* (1993) demonstrated that this model was the best predictor of data whose noise is mainly added to the output.

Moreover, introduction of the measured discharge as an input variable at previous times contributes information on the process state to the network (Johannet *et al.*, 2008a).

Training and generalization

The training procedure entails minimizing an error function, which is usually the least squares error. Several training rules exist, although second-order gradient, steepest-descent methods like the Levenberg-Marquardt algorithm are the most efficient (Zhang *et al.*, 1998).

3.2.1 Levenberg-Marquardt algorithm

In using the training dataset, N input-output couples are presented to the network and enable calculating the error function J for each time k of the learning set, i.e.:

$$(b) \quad J^k(\mathbf{c}) = \frac{1}{2} (q^p - q(\mathbf{c}))^2,$$

with \mathbf{c} corresponding to the matrix of parameters.

Based on the classical second-order gradient, steepest-descent methods, the Levenberg-Marquardt algorithm then calculates parameter increments after each presentation of all examples:

$$(c) \quad \mathbf{c}^{i+1} = \mathbf{c}^i - [\tilde{\mathbf{H}}\mathbf{c}^i + \lambda_{i+1}\mathbf{I}]^{-1} \nabla J \mathbf{c}^i$$

where $\tilde{\mathbf{H}}\mathbf{c}^i$ is the Hessian approximation of the error function at iteration i . The matrix is composed of the first derivative of the J function. \mathbf{I} corresponds to the identity matrix. λ_{i+1} is a parameter introduced to transform $[\tilde{\mathbf{H}}\mathbf{c}^i + \lambda_{i+1}\mathbf{I}]$ into a diagonally dominant matrix at the beginning of training.

3.2.2 Accuracy and generalization

Neural networks are able to approximate any differentiable and continuous function (Hornik *et al.*, 1989). As raised in the state-of-the-art section above however, one major issue with machine learning is overfitting and complexity control. An overly complex model will be highly-fitted with learning data and therefore excessively sensitive to data noise. As a consequence, the model will possess poor generalization capability. In contrast, a low-complexity model will not be able to generate enough of a fit with learning data. This problem is known as the bias-variance dilemma and has been mathematically defined in Geman *et al.* (1992). In sum, training error is not a good estimator of the generalization error, hence the generalization error must be estimated over a set independent of the training procedure.

In this study, we have used several regularization methods to prevent overfitting. Some are referred to as active (examples include early stopping), while other methods are called passive (these include cross-validation

(Dreyfus *et al.*, 2004)). Early stopping consists of observing generalization performance on an independent dataset (i.e. the stop set) during the training period and then stopping training when the error increases. This increasing error indicates that the model can no longer generalize to unseen data. Using early stopping implies that three distinct sets must be defined, one for training, one for stopping and one for testing. More specifically, for this study, the Nash criterion (see Equation (e)) has been calculated for the stop set, and training was stopped one iteration before the Nash stop criterion starts to decrease.

3.3 Model selection

Model selection by cross-validation

Model complexity depends on its number of free parameters. For the multilayer perceptron used in the present study, this number is equal to $(w+3)N_c+1$ ¹. Minimizing complexity thus means minimizing N_c and w . Cross-validation offers a method for simultaneously choosing the appropriate number of hidden neurons N_c and the rainfall sliding window w . The complexity is then selected in order to minimize the squared error function. For each model, the cross-validation score was calculated as follows.

Using the learning set of K subsets, each subset one at a time was reserved as the *validation set*. Training was then performed K times on K subsets (Dreyfus *et al.*, 2004), with the mean squared error also being calculated K times on each validation set. The validation score could be calculated afterwards. To assess the model's generalization capability through the use of several criteria, the score was calculated with respect to a given "criterion". The cross-validation score thus equals:

$$(d) \quad S = \sqrt{\frac{1}{K} \sum_{i=1}^K \text{Criterion}_i}$$

In this study, Criterion_i was the Nash or persistence criterion (see Section 3.5) of the forecasted outflow for year i of the validation set (extending one complete year):

$$(e) \quad \text{Nash}_i = 1 - \frac{\sum_{k=1}^n (q^p(k) - q(k))^2}{\sum_{k=1}^n (q^p(k) - \bar{q}^p)^2}$$

Model complexity was selected by retrieving the configuration that maximized the Nash cross-validation score. For each model, 50 different parameter initializations were performed prior to training.

¹ Input variables are: w rainfalls, 1 bias, 1 previous outflow; i.e. $w+2$ variables correlated with N_c hidden neurons. A total of $(w+2) N_c$ free parameters were thus present in the first layer of the neural network. In the second layer, N_c hidden neurons were present along with 1 bias connected to the output neuron, hence $N_c + 1$ free parameters. The number of free neural model parameters is the sum of both first- and second-layer parameters.

The cross-validation method therefore selects the best complexity corresponding to the specific training set, with each subset used like during validation. The model is not specialized on a given validation set; it can be noted however that this procedure implies investigating the entire field of possibilities, a step that regrettably is very time consuming.

Rainfall width selection by cross-correlation

Rainfall width represents the rainfall history that influences discharge at discrete time k . Mangin (1975) showed that a rainfall-discharge cross-correlogram provides information on basin response to a given rainfall event. By plotting the rainfall-discharge cross-correlation, the linear relation between rainfall and discharge can be observed versus the lag between the two time series. Jenkins & Watts (1968) and Mangin (1975) both determined that when the cross-correlation between two time series is less than 0.2, the series can be considered uncorrelated (in which case the cross-correlation becomes like a white noise). The lag time at which the correlation drops below 0.2 is called the "memory effect" in hydrology (Fig. 2), and this lag corresponds to the rainfall width that can reasonably be applied as a network input variable. From a rigorous standpoint, the 0.2 value needs to be adjusted to the number of time series samples under the hypothesis (for a series less than 30 samples long) of a student distribution. Nevertheless, since the student distribution for the very short series in this research cannot be indicated, the 0.2 threshold is maintained without modification.

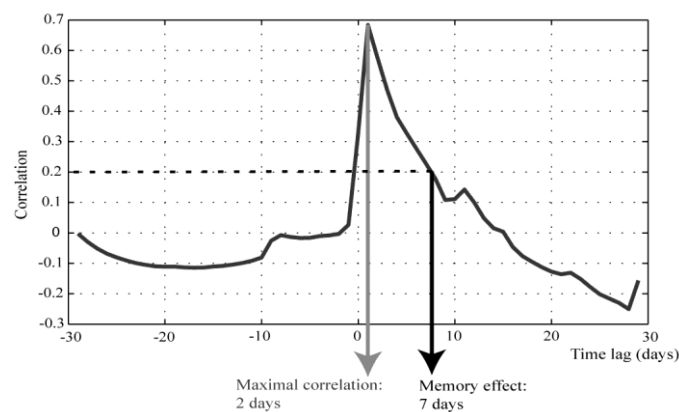


Figure 2: Example of a rainfall-discharge cross-correlogram

The cross-correlation method can only be used for the variable selection and takes into account the linear dependence between both time series, while cross-validation enables choosing the best complexity that also includes the number of hidden neurons and thus considers the nonlinearity implemented by the neural network.

Both selection methods can be used either separately or combined for the input variable selection. Cross-correlation offers a possible initial approach to determine rainfall width, with the aim of diminishing the number

of combinatorial experiments; then, cross-validation can accurately adjust the width length. This two-step variable selection process was applied in the present study to the *Lez* karst aquifer.

3.4 Forecasting

In considering the model presented in Figure 1, forecasting can be performed over several time horizons. When $f=0$, the model calculates discharge at time k , using the discharge from time $k-1$ to $k-n$, along with a rainfall history from k to $k-w+1$. This architecture corresponds to a simulation mode (as defined in Section 2.3.2). If f is positive, the model calculates discharge at time $k+f$ using the same input dataset as for the simulation mode, yielding a forecast over an f -day horizon without any rainfall forecasting involved.

In order to evaluate more specifically the model's forecasting performance, a persistence criterion (Kitanidis and Bras, 1980) was introduced to compare the model forecast to a so-called "naive" forecast. This naive value proposes as a forecast at time $k+f$ the same signal as that observed at time k : in other words, the discharge is not varying between k and $k+f$:

$$(f) \quad C_p = 1 - \frac{\sum_{k=1}^s [q^p(k+f) - q(k+f)]^2}{\sum_{k=1}^s [q^p(k+f) - q^p(k)]^2}$$

where s is the number of samples and f the forecasting horizon. For the naive prediction, $q^p(k) = q^p(k+f)$, which implies that $C_p=0$. When the model forecast is better than the naive prediction, C_p is positive; if it is lower, C_p is negative.

In addition, the naive prediction lies close to the observed signal if the forecasting horizon is short (e.g. see Table 8, in which the Nash criteria for several naive predictions have been reported for various forecasting horizons). In contrast, the naive prediction has no correlation with the observed signal if the forecasted horizon is long. Beating the naive forecast thus becomes an easy task.

Persistence cross-validation scores as well as the Nash cross-validation scores can be calculated; these results in turn enable selecting the model by means of an optimally-adapted criterion.

4. Application to the *Lez* Basin

4.1 The *Lez* karst aquifer

The *Lez* spring is the major outlet of the karst aquifer located north of *Montpellier* (see Fig. 3). Other outlets (including *Lirou*, *Restinclières* and *Fleurette*) are seasonal springs, functioning during periods of high flow rates. The boundaries of the *Lez* aquifer are the *Vidourle* River Basin to the north and east, the *Hérault* River Basin to

the west, and the presumed southern boundary is the *Montpellier* "fold", an east-west directed thrust fault. In spite of the extensive research devoted to the *Lez* aquifer, the hydrogeological basin (as will be detailed hereafter) is not defined accurately at present, especially for its northern and northeastern boundaries. Nevertheless, its area has been estimated at 380 km² (Thiery *et al.*, 1983).

The karst system was developed in late Jurassic and early Cretaceous limestone, with a thickness of approx. 650 m to 1,100 m (Marjolet and Salado, 1976). The lower limit is marl and marly limestone of the middle Jurassic, while the upper limit is marl and marly limestone of the early Cretaceous. The main karst formation outcrops are situated southwest and northeast of the basin (Fig. 3). Including swallow holes, the recharge area is estimated to encompass 130 km² (Fleury *et al.*, 2008; Dörfliger *et al.*, 2008). Most of the aquifer is thus confined, with several perched aquifers existing above its upper limit: Cretaceous limestone (late Valanginian) constituting the *Causse de l'Hortus*, late Eocene and late Cretaceous limestone constituting the other perched aquifers. Pyrenean and Oligocene tectonics have led to large faults, in a NE-SW direction, along with folds in the E-W direction. At several points of the basin, these faults join karstified limestone and impermeable formations in leading to border springs (*Lirou*, *Lez*). As a consequence, the aquifer has been divided into compartments (Lacas and Avias, 1976; Drogue and Grillot, 1976; Bérard, 1983). Of special attention is the *Corconne* fault, which divides the basin into western and eastern parts (Fig. 3) and whose exchanges are poorly known. In addition, the *Corconne* fault could function as both a drain (due to the presence of swallow holes and preferential runoff towards the spring) and a "dam" (as a result of permeable/impermeable contact) (Bérard, 1983). A Messinian salinity crisis (from roughly 5 million years ago) has likely affected the karst structure and its southern part in particular; during this crisis, the Mediterranean Sea, which was partially dried, caused deep karstification in karst systems throughout the surrounding area (Gèze, 1979; Clauzon, 1982; Avias, 1995). For this reason, a connection with the southern part of the *Montpellier* fold might be possible beneath sea level (Touet, 1987).

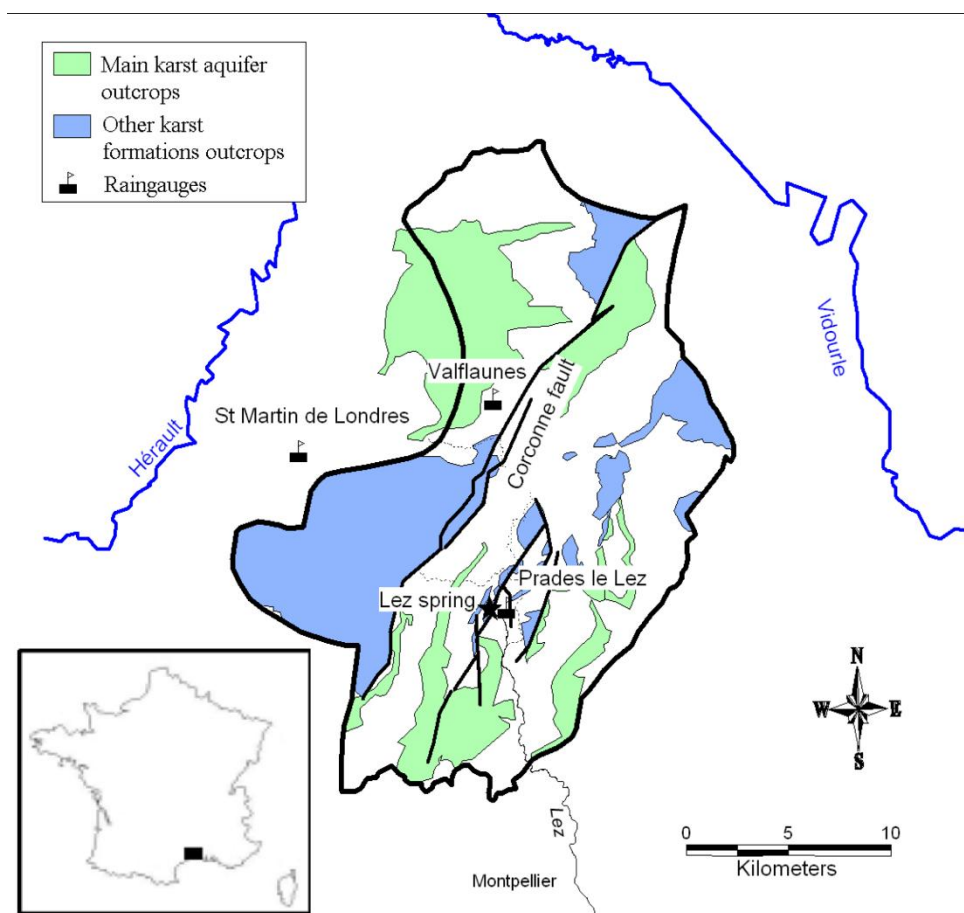


Figure 3: Presumed feeder basin of the *Lez* spring and available rain gauges - the black line is the boundary of the presumed hydrogeological basin drawn by Conroux (2007).

The *Lez* spring is tapped in order to provide drinking water for the city of *Montpellier*. In 1982, boreholes were drilled to enable pumping inside the main drain of the spring, which often dried up the spring and emptied part of the aquifer. Nevertheless, in order to maintain minimal outflow at the spring, a Public Utility Decree imposes releasing 160 L/s into the river when the spring is dry. Measured discharges at the *Lez* spring include both pumped and recovered discharges. The measurements are therefore a sum of natural and artificial discharges. Moreover, during the autumn, the *Lez* River experiences high flows, which induce a flood hazard in the city of *Montpellier*. Major events like those occurring in December 2002, December 2003 and September 2005 caused widespread damage by flooding residences and destroying bridges, roads and dykes.

4.2 *Lez* spring models

Due to its position and use as a potable water source, the *Lez* Basin has been extensively studied and has given rise to a considerable set of available data. However, the complexity of both the basin and its incoming rainfall, along with the artificial discharge, poses a very difficult modeling problem.

Modeling studies first began during the 1970's with rainfall-runoff models (Chemin, 1974; Guilbot, 1975). These early versions corresponded to a former behavior of the system, at a time when pumping was performed in the pool. They do however provide valuable knowledge of these complicated systems. Thierry *et al.* (1983) proposed a 3-reservoir model, which is capable of simulating the water level in the spring pool as well as the overflow discharge. Rainfall and pumping data were processed in order to simulate the drawdown. A linear black box model was proposed by Karam (1989) to forecast the water level in the spring pool at a 5- and 10-day horizon during low-flow periods using piezometric data. Karam (1989) also contributed an interesting discussion relative to delimitation of the supply or feeder basin.

More recently, by proceeding with this system through drain pumping, Conroux (2007) and Dörfliger *et al.* (2008) took into account the relations between: i) water level in the pool, ii) overflow discharge, iii) water table in the boreholes, and iv) pumped discharge, all in order to build a model capable of simulating "natural" outflows at the spring. A "natural" discharge chronicle could thus be produced spanning from 1970 through 2005. Based on this chronicle and in an effort to characterize behavior of the *Lez* basin, Fleury *et al.*, 2008 developed a reservoir model to simulate spring discharge and water level in the main drain. The inputs to this model were a combination of effective rainfall measurements from three rain gauges: *Prades le Lez*, *St Martin de Londres* and *Valflaunès* (Fig. 3). The Nash criterion value of 0.80 was obtained for the discharge simulation, with most errors occurring during flood events.

From a synthesis viewpoint, the *Lez* Basin raises a very difficult modeling issue. On the one hand, the tectonically-disturbed karst aquifer is complex and obviously highly heterogeneous, while on the other, rainfall is highly variable in both time and space and very complicated to forecast. Moreover, the measured discharge is, at least in part, artificial due to tapping the spring. As a last consideration, tapping causes the spring to run dry during the summer, thus preventing any direct outflow measurement.

As a result of these considerations, machine learning inspired by the systemic approach seems to be a well-adapted tool for identifying the behavior of such a nonlinear and partially unknown system.

4.3 Application details

The *Lez* Basin modeling exercise focuses on the most recent period of tapping the spring, after the beginning of pumping in the main drain (1982). The discharge time series for the *Lez* spring and the rainfall measurements on three rain gauges presented in Figure 3 are available daily from 1988 to 2006; these values are distributed as indicated in Table 1. Table 2 lists the altitudes of all three rain gauges and of the *Lez* spring. Distances between rain gauges and the spring are also shown.

As indicated in Section 3.3, the database must be divided into three sets: a training set as complete as possible, a test set to evaluate model performance, and a stop set to apply the early stopping method. This last set is to be composed of cycles that are statistically close to the training set (Toukourou, 2009). For this reason, the stop set is composed of two years, i.e. two complete hydrological cycles, exhibiting nearly the same mean cumulative rainfall and discharge with respect to the total database contents. The test set was chosen so as to contain the leading flood events, i.e. in this case the two most intense events: December 2002 and December 2003 (see Table 3). Model generalization was thus assessed on a significant set. Table 3 lists the major flood events included in the database, along with their maximum daily discharge, rainfall, cumulative rainfall during the event, and the corresponding rain gauge that recorded the maximum precipitation. It is interesting to note that the maximum rainfall was often recorded by the *St Martin* gauge, which is furthest from the spring (Table 2) yet closer to the basin's primary aquifer. At an hourly sampling rate, the rainfall data are highly heterogeneous: correlations equal 0.6 and 0.4 respectively for the most intense events in December 2002 and December 2003, although the daily samplings as used in this study have provided well-correlated signals (Table 4).

Set	Number of cycles	Date
Training	10	01/09/1988 to 31/08/1993, 01/09/1996 to 31/08/2001, 01/09/2004 to 31/08/2006
Stop	2	01/09/1993 to 31/08/1995
Test	2	01/09/2002 to 31/08/2004

Table 1: Distribution of database contents into training, stopping and testing sets

	Altitude	Distance to the <i>Lez</i> spring
<i>Lez</i> spring	65 m	
Valflaunès	133 m	9.8 km
Prades le Lez	85 m	800 m
St Martin de Londres	194 m	12 km

Table 2: Altitudes of the rain gauges and *Lez* gauge station; distances between rain gauges and *Lez* spring

Date	Maximal discharge (m ³ /s)	Maximal daily rainfall (mm/24h)	Cumulated rainfall (mm)	Corresponding rain gauge	Set
November 1988	6,7	104	208	St Martin	Training
November 1993	7,1	47	227	St Martin	Stop
February 1994	6,5	95	148	Prades	Stop
October 1994	10,6	205	297	St Martin	Stop
November 1994	9,4	151	230	St Martin	Stop
December 1996	10,6	186	110	St Martin	Training
January 1997	8,5	105	170	St Martin/Prades	Training
December 1997	11,5	117	295	St Martin	Training
November 1999	6,9	88	156	St Martin	Training
November 2000	7,4	87	143	St Martin	Training
January 2001	9,9	88	151	St Martin	Training
December 2002	15,4	161	366	St Martin	Test
December 2003	12,6	149	267	Prades	Test
September 2005	8,8	233	284	Valflaunès	Training
January 2006	7,9	102	162	St Martin	Training

Table 3: Major flood events included in the database

	Prades	St Martin	Valflaunès
Prades	1	0.85	0.90
St Martin		1	0.87
Valflaunès			1

Table 4: Rain gauge cross-correlations. The cross-correlogram was computed for each hydrological cycle (September to August, between 1988 and 2006). The correlations presented correspond to the mean of the primary correlation peaks of each cycle.

The first matter therefore is to evaluate the advantage gained by using just one rain gauge within the neural network, as a means of diminishing complexity and thus increasing the generalization capability. From another standpoint, the linear correlation does not presume any advantage that the nonlinear neural network can derive from the small differences between signals. To proceed with this evaluation, an initial step entailed designing a model for each rain gauge and then comparing results. As a second step, the three rain gauges were applied to the network, with a subsequent discussion of results.

4.3.1 Rainfall-runoff modeling using a single rain gauge

The number N_c of hidden neurons was determined by cross-validation. Alternatively, input selection, namely rainfall width selection, was determined using both cross-correlation and cross-validation.

The various results are reported in Table 5. A first observation to be shared is that due to the architecture (directed mode), these results are excellent. The *Lez* spring was simulated to a high level of accuracy (Fig. 5). The Nash criteria calculated on the test set (covering two complete years, including the intense flood event of December 2003) all exceed 0.95. The number of hidden neurons had always been low, which means that the complexity selection was indeed efficient.

In addition, both input selection methods have produced roughly the same results with no great differences from one rain gauge to the next. Except for *Prades'* model, cross-correlation yielded a less complex network; this

method however is based on a linear cross-correlation, whereas neural networks are able to accommodate nonlinearity.

	Prades		St Martin		Valflaunès	
Input selection method	Cross-validation	Cross-correlation	Cross-validation	Cross-correlation	Cross-validation	Cross-correlation
Rainfall width w (days)	8	9	11	9	10	8
Number of hidden neurons N_c	3	3	3	3	3	5
Nash Test	0,96	0,95	0,95	0,95	0,95	0,95

Table 5: Architecture and performance of simulation models using each rain gauge one at a time

As suggested in Section 3.4.1, the next step combined these two input selection methods. The cross-correlation analysis was performed, yielding rainfall widths around which cross-validation could be applied by calculating cross-validation scores. If the selected width were to correspond to a border of the tested width, then another trial would be required in order to select the better model in terms of the Nash cross-validation score.

The results for the *St Martin* rain gauge are shown in Table 6.

Rain gauge	Memory effect	Tested width	Selected Width	Nash Cross-validation score	Nash test
St Martin	8 days	8 to 10 days	10 days	0.81	(non optimum)
		11 and 12 days	11 days	0.92	0.95

Table 6: Width selection using a combination of cross-validation and cross-correlation methods

It can be noted that the combination of both methods leads to the same architecture as cross-validation on its own. However, combining methods allows choosing the number of hidden neurons and considerably reduces network design time, as compared to an exhaustive application of the cross-validation method.

4.3.2 Rainfall-runoff modeling using 3 rain gauges

As previously mentioned, the rain gauges are well correlated with a daily sampling rate, yet building a network receiving information from all three rain gauges offers some interesting prospects. Unfortunately, including information stemming from the three rain gauges in the network increases complexity and thus could lead to degrading the generalization. To avoid this drawback, a slightly rearranged network that reduces the number of parameters has been proposed. This revised network includes one more linear neuron for each rain gauge, and this extra neuron can be viewed as an additional hidden layer.

The rainfall width of each rain gauge is thus connected to a single linear hidden neuron, which itself is connected to the classical nonlinear hidden layer (Fig. 4). As an example, a 10-day rainfall width for each of the three rain gauges with four hidden neurons requires 124 parameters with the classical two-layer network (as shown in Fig. 1) and just 46 parameters using the 3-layer network (Fig. 4).

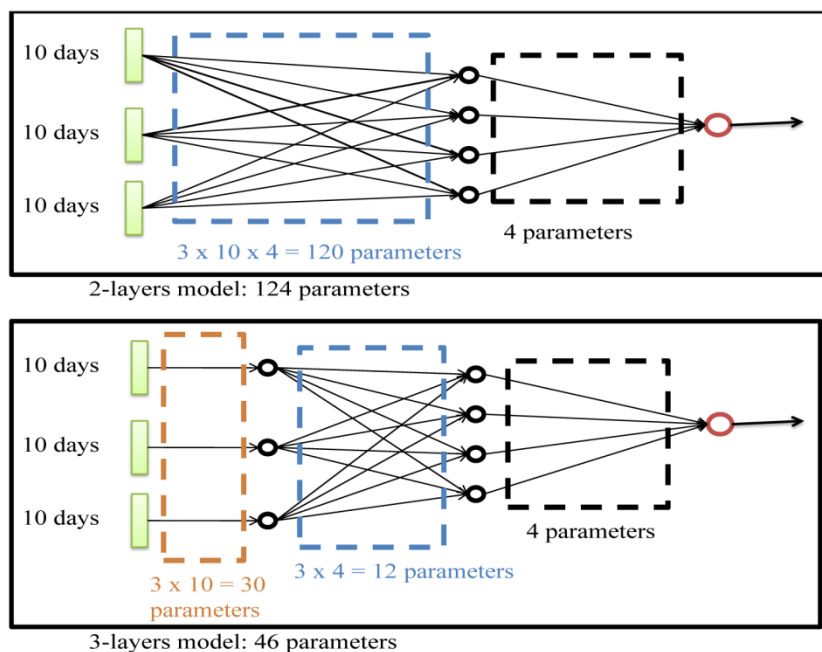


Figure 4: Presentation of the 2-layer and 3-layer networks

Since use of all three rain gauges implies three rainfall widths and in order to reduce model design time, both input selection methods were combined, like in Section 4.3.1. The resulting model complexity is presented in Table 7.

Let's note that the selected rainfall widths for *Prades* and *St Martin* correspond to the upper boundary of tested widths; the largest widths (10 and 11 days) were also tested, though they produced a lower cross-validation score than the 9-day width.

The 3-rain gauge model has led to a similar accuracy as the 1-rain gauge model, with a Nash criterion on the test set equal to 0.95.

	Prades	St Martin	Valflaunès
Memory effect	8 days	8 days	7 days
Runned width	8 to 10 days	8 to 10 days	7 to 9 days
Width giving best cross-validation score	9	9	7

Table 7: Complexity selection for the three-rain gauge model

The simulated hydrograph is displayed in Figure 5. Since results are very similar for all three gauges, we have only presented hydrographs from the *St Martin* model ($w=11$ and $N_c=3$) and the 3-gauge model (w as given in Table 7 and $N_c=4$). Both simulated hydrographs slightly underestimate the December 2002 peak. On the other

hand, the December 2003 event is underestimated by the 3-gauge model while it is overestimated by the *St Martin* model.

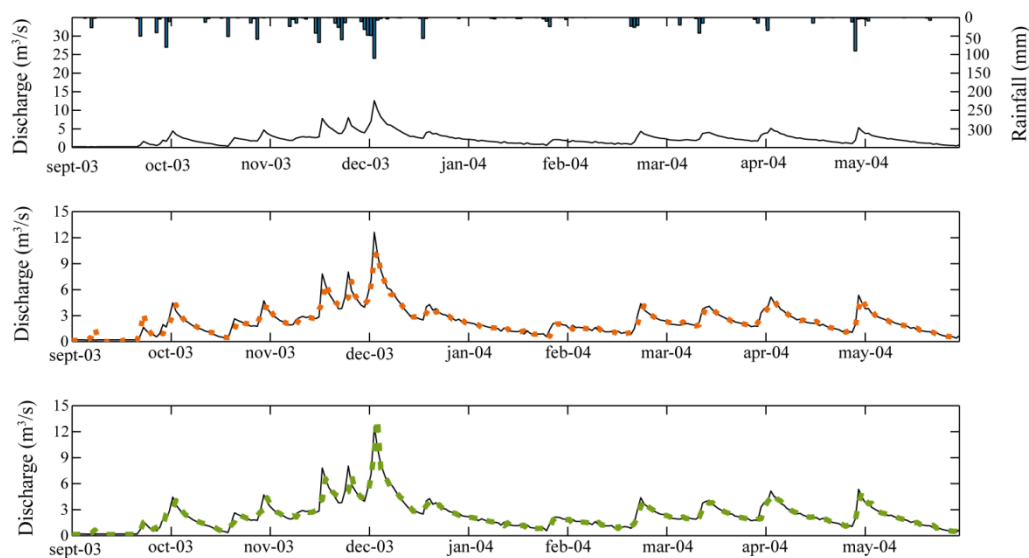


Figure 5: Simulated discharges at the Lez spring. The bar graph corresponds to measured daily cumulative rainfall at the *St Martin de Londres* station. The black line shows the measured discharge at the *Lez* spring; the orange dotted line on the second graph is the simulated discharge using the three-rain gauge model; the green dotted line on the third graph is the simulated discharge using the *St Martin* model with cross-correlation chosen as the rainfall width selection method.

4.3.3 Forecasting

When forecasting, for each horizon, the model selection using both Nash and the persistence cross-validation score have provided the same complexity for most cases. For other cases, results are very close in terms of model complexity and the values of Nash and persistence criteria on the test set.

Both the one-gauge and three-gauge models were used for this forecasting exercise. For one-gauge models, results from the cross-correlation and cross-validation input selection methods are similar. The results listed correspond to the cross-correlation rainfall width selection method.

Moreover, the Nash criterion found by the naive forecast is included; this score is very high for the short forecasting horizon (0.9 at *St Martin*) and decreases as the forecasting horizon increases. It should also be mentioned that the Nash and persistence criteria do not systematically evolve in a proportional manner with respect to one another; it is indeed possible to obtain negative persistence for a Nash criterion greater than that of the naive forecast (see Table 8), which depends on the signal shape. The Nash criterion of the naive forecast nevertheless offers an interesting guideline.

Horizon	1 day	2 days	3 days
Naive forecast (measured signal)	0.85	0.67	0.53
Prades	0.76	0.81	0.68
St Martin	0.90	0.81	0.70
Valflaunès	0.76	0.78	0.68
3 rain gauges	0.88	0.80	0.68

Table 8: Nash criteria for various models and forecasting horizons (models were selected by Nash cross-validation score). The width selection method for the one-gauge model (*Prades*, *St Martin*, *Valflaunès*) was cross-correlation. For the three-gauge model, both width selection methods were used in combination.

Horizon	1 day	2 days	3 days
Prades	-0.68	0.40	0.30
St Martin	0.33	0.39	0.33
Valflaunès	-0.74	0.31	0.30
3 rain gauges	-0.13	0.10	0.28

Table 9: Persistency criteria for various models and forecasting horizons (models were selected by the persistence cross-validation score). The width selection method for the one-gauge model (*Prades*, *St Martin*, *Valflaunès*) was cross-correlation. For the three-gauge model, both width selection methods were combined.

Better results in terms of both Nash and persistence criteria were obtained with the *St Martin* model (w , as given in Table 10 and $N_c=4$) and the three-rain gauge model (w , as given in Table 10 and $N_c=4$). The *St Martin* model probably offered the better one-gauge model by virtue of recording greater quantities of rainfall than the others (also note that it has been placed at a higher elevation, see Table 2). By being the furthest from the *Lez* spring (Table 2), the *St Martin* rain gauge allows this model to produce the better forecasts (the *St Martin* flood energy arrives after the others at the *Lez* spring). As shown in Tables 8 and 9, these results are excellent when considering that no information regarding future rainfall is available. In examining Tables 5 and 9, which report the performances of the various selected models, it appears that the Nash criterion was roughly constant for the simulation models, while persistency differed for the forecasting models (for the 1-day and 2-day forecasting horizons).

The selected rainfall widths for various forecasting horizons are presented in Table 10. It can be noted that the *St Martin* model for a 3-day forecasting horizon, with a rainfall width of 6 days, was particularly parsimonious. In fact, the vanishing role of rainfall for such an extended horizon was due to the fact that the network, for a forecast at $k+3$, was able to glean more significant information from the previous observed outflows than from the previous rainfalls. Even though the forecasting function without a rainfall forecast differs from the simulation, it is still interesting to observe that 6 days in the rainfall window plus 3 days of forecasting horizon provides 9 days, which is the window length selected in the simulation mode for the *St Martin* model (using the

linear cross-correlation method, see Table 5). For the three-rain gauge model, this phenomenon did not appear, possibly because of the more complex architecture using three layers of neurons, thus making better use of the nonlinear capabilities of neural models thanks to the cross-validation rainfall selection method.

Model/Horizon	1 day	2 days	3 days
St Martin rain gauge	14 days	13 days	6 days
3 rain gauges	Prades: 8 days St Martin: 9 days Valflaunès: 7 days	Prades: 9 days St Martin: 10 days Valflaunès: 8 days	Prades: 8 days St Martin: 8 days Valflaunès: 8 days

Table 10: Selected rainfall widths for various forecasting horizons, in combining both width selection methods.

The hydrographs calculated for the 2003 test cycle are shown for both these models as well as for each forecasting horizon (Figs. 6-8). The low-flow period has not been shown since the discharge during this period corresponds to a recovery of the minimum prescribed outflow, i.e. 160 L/s (see Section 4.1).

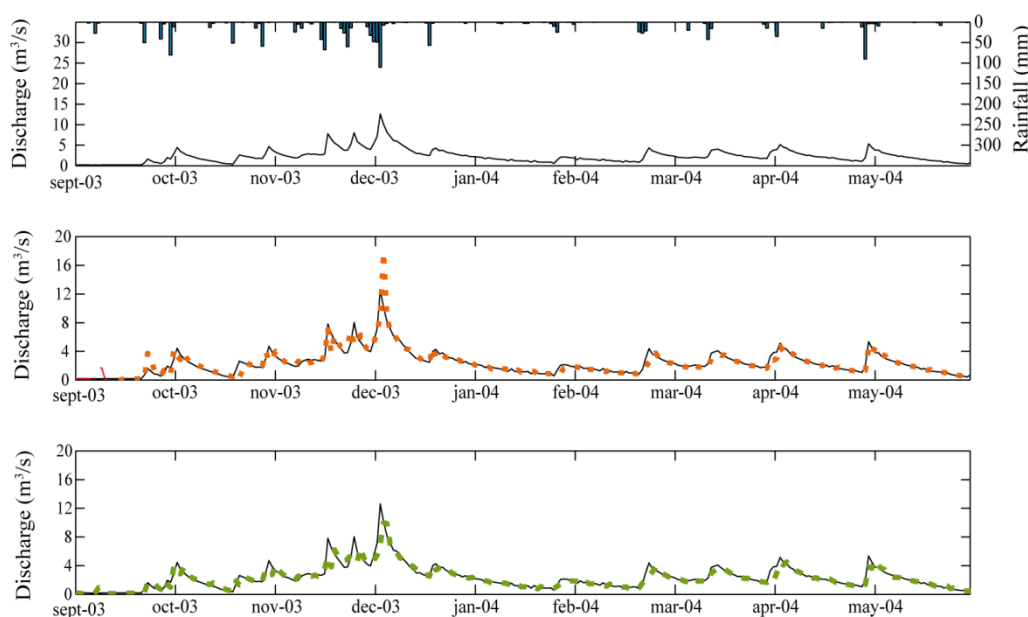


Figure 6: One-day forecasted discharges at the Lez spring. The bar graph corresponds to measured daily cumulative rainfall at the St Martin de Londres station. The black line indicates the measured discharge at Lez spring; the orange dotted line on the second graph is the forecasted discharge using the three-gauge model (in combining both width selection methods); the green dotted line on the third graph is the forecasted discharge using the St Martin model with cross-correlation as the rainfall width selection method.

The three-gauge model overestimates the major flood peaks, as opposed to the St Martin model underestimations. For other less intense flood peaks, the forecasted and observed values were very similar and synchronous over time. For the St Martin model, the persistency index (Table 9) and the hydrograph both tend to suggest that the model has provided a significant forecast one day ahead of time. For the three-gauge model, the

persistence index was negative, even though it was the better model in terms of flood forecasting (good synchronism and extrapolation).

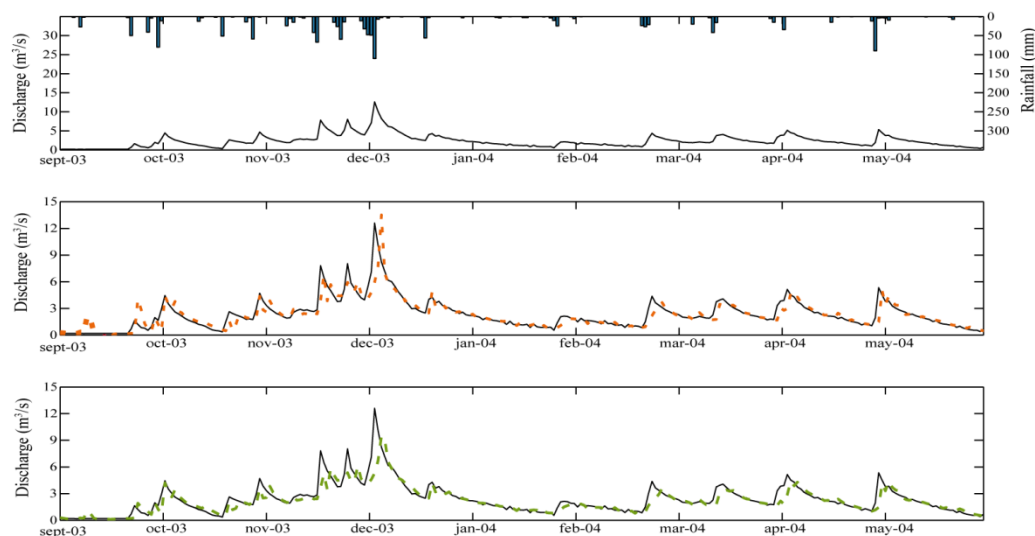


Figure 7: Two-day forecasted discharges at the Lez spring. The bar graph corresponds to measured daily cumulative rainfall at the *St Martin de Londres* station. The black line depicts the measured discharge at *Lez* spring; the orange dotted line on the second graph is the forecasted discharge using the three-gauge model; the green dotted line on the third graph is the forecasted discharge using the *St Martin* model with cross-correlation as the rainfall width selection method.

For the two-day forecast, a one-day lag time appears at the major flood peak, demonstrating the one-day anticipation capability. Interestingly, the three-gauge model still yields greater peaks than the *St Martin* model; nevertheless, forecasted outflow at the date of the observed peak revealed less than 30% error. The forecast hydrographs however are still acceptable for moderately-intense floods with an actual anticipation of two days for autumn floods (both models) and a shorter anticipation for spring floods. Considering the persistency criteria (Table 9), it appears that the model offers good anticipation capability, which is remarkable for a model devoid of rainfall forecasts.

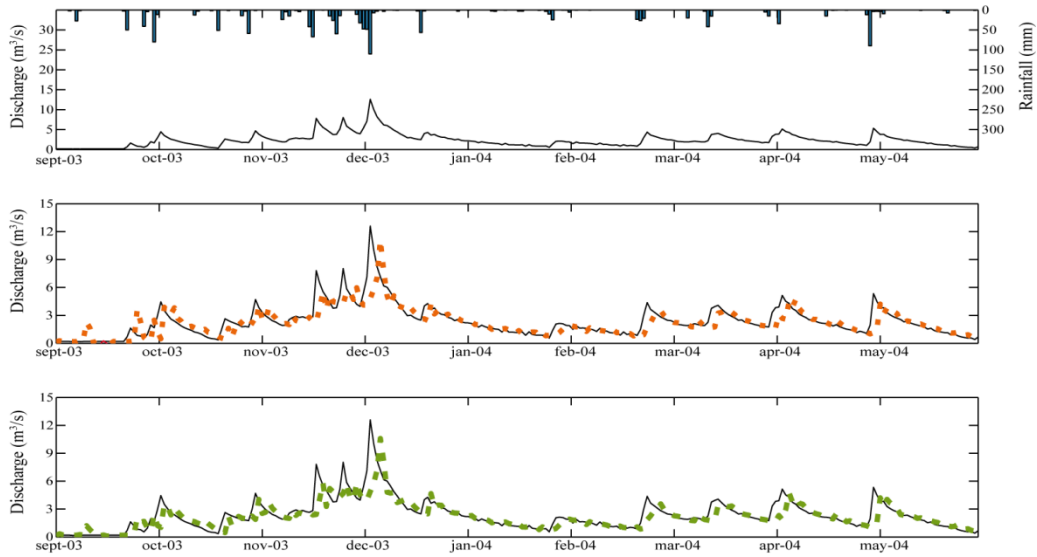


Figure 8: Three-day forecasted discharges at the *Lez* spring. The bar graph corresponds to measured daily cumulative rainfall at the *St Martin de Londres* station. The black line reveals the measured discharge at *Lez* spring; the orange dotted line on the second graph is the forecasted discharge using the three-gauge model; the green dotted line on the third graph is the forecasted discharge using the *St Martin* model with cross-correlation as the rainfall width selection method.

Over the three-day horizon, regarding the persistency criteria (Table 9) which are all positive and high, both models would be deemed acceptable; nevertheless, the graphical analysis contradicts this first-level analysis. A one- or two-day lag time appears for all peak floods, yielding an actual anticipation capability of 1 or 2 days, as suggested by the two previous figures. Both models underestimated these peaks, which illustrate the limitations of models without any rainfall forecasts.

4.3.4 Discussion

Regarding the selection method

As underscored in the literature, complexity selection is the major issue in neural network design, given that it determines the generalization capability. Along these lines, cross-validation is a general method that can be used for selecting not only the number of hidden neurons but also the input variables. Unfortunately, this method may be very expensive, from a computational standpoint, due to the combinatorial complexity of the trials to be performed. Concerning determination of the width of the temporal rainfall window application, i.e. the memory effect, we have proposed limiting the field of possible trials by proceeding with two steps: i) preselecting the width around the value provided by the cross-correlation; and ii) adjusting this width accurately in a nonlinear mode (using the neural network) by means of cross-validation.

The proposed method has proven its efficiency in the selection of complexity (i.e. w and N_c), for simulations as well as for forecasting. It can be noticed that the adjustments provided by the cross-validation step to the

memory effect were not always in the same direction. Length was reduced for the *Valflaunès* simulation model and increased for the *St Martin* and three-gauge models. These slight differences can be explained by the ability of the neural network to manage the nonlinearity and/or by the fact that the memory effect is not an exact estimation (student distribution hypothesis). The next section will focus on the quality of models provided by one method versus the other.

Regarding the comparison of input selection methods

A comparison of input selection methods was conducted using the one-gauge model. The simulation results are similar regardless of the rain gauge or input selection method used (Table 5). On the other hand, forecasting results are equivalent in terms of the Nash test criteria, yet not identical. Table 10 shows that the models selected by the combined method vary the rainfall width as a function of forecasting horizon, in particular for the one-gauge model. Moreover, a model selected by cross-correlation alone is unable to adapt to the given forecasting horizon because such a window is predetermined and not allowed to vary. On this application, results have revealed that the performances of both methods were roughly equivalent.

Regarding the simulation mode

The proposed selection method was applied to both one-gauge and three-gauge models. The results provided in simulation mode showed that the three-gauge model was not more efficient than any of the single-gauge models. This finding is consistent with the observation that rain gauges exhibit a high cross-correlation (Table 4) at the daily sampling frequency. At the available time scale, no further information is being provided by the additional rain gauges.

Regarding the forecasting mode

As opposed to the observations in simulation mode, the *St Martin* model yields better forecasts than the other models, and slightly better than the three-gauge model forecasts (Tables 8 and 9), especially for one-day forecasts. This outcome is understandable since the highest rainfall values were mainly recorded by the *St Martin* rain gauge (Table 3). Moreover, this analysis can be refined by examining the hydrographs. For one-day and two-day forecasts, the *St Martin* model underestimated the peaks for principal flood events (December 2002 and December 2003), whereas the three-gauge model overestimated these peaks. The three-gauge model, which overestimated the highest peak of the entire database, while estimating very well the less intense peaks until two days ahead of time, is thus better suited for flood forecasting.

It can also be noted that the rainfall widths (Table 10) narrowed as the forecasting horizon was extended: this behavior is logical since independent of the forecasting horizon, the same rainfall length is required to estimate outflow. This behavior was not observed however for the three-gauge model.

With this as a satisfactory outcome, with a Nash criterion equal to 0.9 (respectively 0.8 and 0.7 for the 1-day and 2-day forecasts), the hydrographs were simulated with good accuracy, primarily due to the network architecture supplied by previous outflows. The ability of neural networks to take advantage of previous outflow observations and not require rainfall forecasts thus offers a very attractive feature.

Lastly, the forecasted hydrographs demonstrate that neural network models are able to extrapolate relative to training data. The highest event in the training set (December 1997) actually reaches a maximum discharge of 11.5 m³/s (Table 3), while the calculated discharge for the major test set event (December 2002) with a 2-day forecasting horizon ranged between 9 m³/s and 18 m³/s (Fig. 6) when the measured value was 15 m³/s.

5. Conclusion

Forecasting floods in karst and populated basins is a difficult yet critical task. The *Lez* Basin (in southern France) was chosen as a pilot site to develop and compare forecasting methods for the purpose of setting up an early warning system for the local population. When combining the problems of karst nonlinearity, rainfall unpredictability and an artificial runoff (via pumping for water supply), the *Lez* Basin has been a preferred study target over the past three decades. In the present paper, we have demonstrated that the systemic approach can be performed by neural networks on the *Lez* Basin for discharge forecasting until one day ahead of time. The anticipation capability was confirmed until two days ahead of time. Inspired from the *memory effect*, which is well known within the framework of correlation analysis, an efficient methodology has been proposed to determine neural model complexity. This methodology was then applied to several models based on either a single rain gauge or several gauges. A blind test was conducted on the most intense database event, hence proving that the method, thanks to complexity selection, is able to accurately forecast floods in real time with satisfactory accuracy, in the absence of actual rainfall forecasts. The follow-up to this study will be intended to introduce *a priori* knowledge on geology and the karst structure. This additional information should constrain the model to physically output interpretable values for its parameters and/or hidden neurons, including for example the relative contributions of various geological zones of this basin.

Acknowledgments

The authors would like to thank the METEO-France weather agency for providing the rainfall datasets, as well as Gilles Le Gac from the DIREN Montpellier Office for contributing discharge data.

We are also very grateful to Caroline Wittwer, Bruno Janet and Arthur Marchandise for the exciting collaboration shared with the SCHAPI Unit, and to Nathalie Dörfliger, Perrine Fleury, Henri Paloc, Daniel Diep, Pierre Roussel-Ragot, Bernard Vayssade and Marc Vinches for the helpful discussions they helped organize.

Lastly, our thanks are extended to Dominique Bertin for his extremely fruitful collaboration in the design and implementation of the Neural Network simulation tool: RnfPro.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A. O'Connell, P.E., Rasmussen, J., 1986. Introduction to the European Hydrological System- Systeme Hydrologique Europeen, 'SHE', 2: Structure of a Physically-based, Distributed Modelling System. *Journal of Hydrology JHYDA* 7, 87(1-2).

Avias, J., 1995. Active management of Lez karstic spring, Herault, France, 1957-1994. *Hydrogeologie Editions du BRGM*, 113–128.

Bailly-Comte, V., Jourde, H. & Pistre, S., 2009. Conceptualization and classification of groundwater-surface water hydrodynamic interactions in karst watersheds: Case of the karst watershed of the Coulazou River (Southern France). *Journal of Hydrology*, 376(3-4), 456-462.

Bakalowicz, M., 2005. Karst groundwater: a challenge for new resources. *Hydrogeology Journal*, 13(1), 148-160.

Bakalowicz, M., 2008. Le milieu karstique : études et perspectives, identification et caractérisation de la ressource. In *CFH - Colloque Hydrogéologie et karst au travers des travaux de Michel Lepiller*. 17 mai 2008.

Barron, A.R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, 39(3).

Bartholmes, J. & Todini, E., 2005. Coupling meteorological and hydrological models for flood forecasting. *Hydrol. Earth Syst. Sci.*, 9(4), 333-346.

Bérard, P., 1983. Alimentation en eau de la ville de Montpellier. Captage de la source du Lez. Etude des relations entre la source et son réservoir aquifère. Rapport n°2 Définition des unités hydrogéologiques, *BRGM Montpellier*.

Beven, K.J., Warren, R. & Zaoui, J., 1980. SHE: towards a methodology for physically-based distributed forecasting in hydrology. *IAHS Publ*, 129.

Beven, K., Kirkby M.J., Schofield, N., Tagg A.N. 1984. Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. *Journal of Hydrology*, 69(1-4), 119-143.

Bouttier, F., 2003. The AROME mesoscale project. In *Proceedings of a seminar on Recent developments in data assimilation for atmosphere and ocean*. p. 8–12.

Chemin, J., 1974. Essai d'application d'un modèle mathématique conceptuel au calcul du bilan hydrique de l'aquifère karstique de la source du Lez (Région Nord de Montpellier). PhD. Université des Sciences et Techniques du Languedoc (Montpellier).

Chiang, Y.M., Chang, L.C. & Chang, F.J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *Journal of Hydrology*, 290(3-4), 297–311.

Clauzon, G., 1982. Le canyon messinien du Rhône: une preuve décisive du "desiccated deep-basin model [Hsü, Cita et Ryan, 1973]". *Bulletin de la Société géologique de France, Océans-paléocéans, Paris*, 24(3), 597-610.

Cloke, H. & Pappenberger, F., 2009. Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3-4), 613-626.

Conroux, Y., 2007. Caractérisation du fonctionnement hydrodynamique de l'aquifère karstique du Lez (Hérault) à l'état naturel. Master Université Avignon.

Cornaton, F. & Perrochet, P., 2002. Analytical 1D dual-porosity equivalent solutions to 3D discrete single-continuum models. Application to karstic spring hydrograph modelling. *Journal of Hydrology*, 262(1-4), 165–176.

Coulibaly, P., Anctil, F., Rasmussen P., Bobée B., 2000a. A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrological Processes*, 14(15), 2755-2777.

Coulibaly, P., Anctil, F. & Bobée, B., 2000b. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230(3-4), 244-257.

Dae-II, J. & Young-Oh, K., 2005. Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological processes*, 19(19), 3819–3835.

Delrieu, G., Nicol J., Yates, E., Kirstetter P-E., Creutin J-D., Anquetin, S., Obled, C., Sauliner G.M., Ducrocq, V., Gaume E., Payrastra, O., Andrieu, H., Ayrat, P-A., Bouvier, C., Neppel, L., Livet, M., Lang, M.,

Parent-duChâtelet, J., Walpersdorf, A., Wobrock, W., 2005. The Catastrophic Flash-Flood Event of 8–9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes–Vivarais Mediterranean Hydrometeorological Observatory. *American Meteorological Society Journal*, 6(1).

Dörfliger, N., Jourde, H., Ladouche, B., Fleury, P., Lachassagne, P., Conroux, Y., Pistre, S., Vestier, A., 2008. Active water management resources of karstic water catchment : the example of Le Lez spring (Montpellier, South France). In *World Water Congress*. Montpellier.

Dreyfus, G., 2004. *Neural Networks Methodology and Applications*, Springer, Berlin.

Droque, C. & Grillot, J.C., 1976. Structure géologique et premières observations piézométriques à la limite du sous-système karstique de Terrieu (Périmètre expérimental). *Ann. Sc. Univ. Besançon*, 25, 195-210.

Fleury, P., 2005. Sources sous-marines et aquifères karstiques côtiers méditerranéens. Fonctionnement et caractérisation. PhD, Université Pierre et Marie Curie, Paris, 286 p.

Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., Dörfliger, N., 2008. Modelling the hydrologic functions of a karst aquifer under active water management - The Lez spring. *Journal of Hydrology*, 365 (3-4), 235-243.

Fleury, P., Plagnes, V. & Bakalowicz, M., 2007. Modelling of the functioning of karst aquifers with a reservoir model: Application to Fontaine de Vaucluse (South of France). *Journal of Hydrology*, 345(1-2), 38-49.

Formigué, P. & Lavabre, J., 2005. Flood forecasting with the GR3J conceptual rainfall-runoff model. Adaptability to rainfall uncertainties. *Journal of Water Science*, 18(1), 87-102.

Gaume, E. & Bouvier, C., 2004. Analyse hydro-pluviométrique des crues du Gard et du Vidourle des 8 et 9 septembre 2002 : Crues Méditerranéennes : l'aléa et la gestion de crise = Hydrological analysis of the Gard and Vidourle river floods on the 8th and 9th September 2002. *La Houille Blanche*, (6).

Geman, S., Bienenstock, E. & Doursat, R., 1992. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1–58.

Gèze, B., 1979. Languedoc méditerranéen. Montagne Noire. Masson.

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2), 115-123.

Guilbot, A., 1975. Modélisation des écoulements d'un aquifère karstique (Liaison pluie-débit). Application aux bassins de Saugras et du Lez. PhD. Université des Sciences et Techniques du Languedoc (Montpellier).

Half, A.H., Half, H.M. & Azmoodeh, M., 1993. Predicting runoff from rainfall using neural networks. In *Proc. Engng. Hydrol., ASCE, New York*, 768-775.

Hornik, K., Stinchcombe, M. & White, H., 1989. Multilayer Feedforward networks are universal approximator. *Neural Networks*, (2), 359-366.

Hsu, K.L., Gupta, H.V. & Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31(10), 2517-2530.

Jaquet, O., Siegel, P., Klubertanz, G., Benabderrhamane, H., 2004. Stochastic discrete model of karstic networks. *Advances in Water Resources*, 27(7), 751–760.

Javelle, P., Berthet, L., Arnaud, P., Lavabre, J., Perrin, C., 2008. Comparaison de deux versions du modèle GR pour la prévision des crues sur un grand échantillon de bassins versants français. In *Colloque SHF-191e CST «Prévisions hydrométéorologiques»*, Lyon, 18-19 novembre 2008.

Jenkins, G.M. & Watts, D.G., 1968. *Spectral Analysis and its Applications*, Holden-Day, Boca Raton, Fla.

Johannet, A., Mangin, A. & D'Hulst, D., 1994. Subterranean water infiltration modelling by neural networks : use of water source flow. In *Proceedings of the International Conference on Artificial Neural Networks*. International Conference on Artificial Neural Networks. p. 1033 - 1036.

Johannet, A., Mangin, A. & Vayssade, B., 2008a. Modélisation d'un système karstique par réseaux de neurones: simulation des débits du karst du Baget-France. *Collection EDYTEM Cahiers de Géographie*, (7).

Johannet, A., Vayssade, B. & Bertin, D., 2008b. Neural Networks: From Black Box towards Transparent Box Application to Evapotranspiration Modeling. *International Journal of Computational Intelligence*, 4(3).

Jourde, H., Roesch A., Guinot, V., Bailly-Comte V., 2007. Dynamics and contribution of karst groundwater to surface flow during Mediterranean flood. *Environmental Geology*, 51(5), 725-730.

Jukić, D. & Denić-Jukić, V., 2009. Groundwater balance estimation in karst by using a conceptual rainfall–runoff model. *Journal of Hydrology*, 373(3-4), 302-315.

Karam, Y., 1989. Essais de modélisation des écoulements dans un aquifère karstique. Exemple de la source du Lez. (Hérault, France). PhD. Université des Sciences et Techniques du Languedoc (Montpellier).

Kaufmann, G. & Romanov, D., 2008. Cave development in the Swabian Alb, south-west Germany: A numerical perspective. *Journal of Hydrology*, 349(3-4), 302–317.

Kitanidis, P. & Bras, R., 1980. Real-time forecasting with a conceptual hydrologic model. 2. Application and results. *Water Resources Research*, 16(6), 1034-1044.

Kurtulus, B. & Razack, M., 2007. Evaluation of the ability of an artificial neural network model to simulate the input-output responses of a large karstic aquifer: the La Rochefoucauld aquifer (Charente, France). *Hydrogeology Journal*, 15(2), 241–254.

Labat, D., Ababou, R., Mangin, A., 1999. Linear and nonlinear input/output models for karstic springflow and flood prediction at different time scales. *Stochastic Environmental Research and Risk Assessment*, 13(5), 337-364.

Lacas, J.L. & Avias, J., 1976. Introduction à la méthodologie d'étude et d'utilisation des champs hydrothermiques des aquifères karstiques: d'après l'exemple du site de l'exurgence de la source du Lez (Hérault, France), Université des sciences et techniques du Languedoc (Montpellier), CERGA.

Larocque, M., Mangin, A., Razack, M., Bandon, O., 1998. Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *Journal of Hydrology*, 205(3-4), 217-231.

Le Lay, M. & Saulnier, G., 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. *Geophysical Research Letters*, 34(13).

Liu, Z. & Todini, E., 2002. Towards a comprehensive physically-based rainfall-runoff model. *Hydrology and Earth System Sciences*, 6(5), 859-881.

Llasat, M.C., Llasat-Botija, M., Prat, M.A., Porcu, F., Price, C., Mugnai, Lagouvardos, K., Ktroni, V., Katsanos, D., Michaelides, S., Yair, Y., Savvidou., Nicolaidis, K., 2010. High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database. *Advances in Geosciences*, 23, 47-55.

Long, A.J., 2009. Hydrograph separation for karst watersheds using a two-domain rainfall-discharge model. *Journal of Hydrology*, 364(3-4), 249-256.

Mangin, A., 1975. Contribution à l'étude hydrodynamique des aquifères karstiques. PhD. Laboratoire souterrain du Centre national de la recherche scientifique. Université de Dijon.

Mangin, A., 1981. Utilisation des analyses corrélatoire et spectrale dans l'approche des systèmes hydrologiques. *CR Acad. Sci. Paris*, 293, 401-404.

Mangin, A., 1984. Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *Journal of Hydrology*, 67(1-4), 25-43.

Marchandise, A., 2007. Modélisation hydrologique distribuée sur le Gardon d'Anduze: étude comparative de différents modèles pluie-débit, extrapolation de la normale à l'extrême et tests d'hypothèses sur le processus hydrologiques. PhD. Université des Sciences et Techniques du Languedoc (Montpellier).

Maréchal, J., Ladouche, B. & Dörfliger, N., 2008. Karst flash flooding in a Mediterranean karst, the example of Fontaine de Nîmes. *Engineering Geology*, 99(3-4), 138-146.

Marjolet, G. & Salado, J., 1976. *Contribution à l'étude de l'aquifère karstique de la source du Lez (Hérault)*. PhD. Université des Sciences et Techniques du Languedoc (Montpellier).

Mohrlok, U. & Sauter, M., 1997. Modelling groundwater flow in a karst terrane using discrete and double-continuum approaches; importance of spatial and temporal distribution of recharge. In *Proceedings of the 12th International Speleological Congress*. p. 167–170.

Mohrlok, U. & Teutsch, G., 1997. Double continuum porous equivalent (DCPE) versus discrete modelling in karst terranes. In *Proceedings of the International Symposium and Field Seminar on Karst*. p. 319–326.

Nerrand, O., Roussel-Ragot, P., Personnaz, L., Dreyfus, G., Marcos, S., 1993. Neural networks and nonlinear adaptive filtering: Unifying concepts and new algorithms. *Neural Computation*, 5(2), 165-199.

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 Processes in Geophysics*, 13(4), 443–448.

Rumelhart, D.E., Hintont, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

Sahoo, G. & Ray, C., 2006. Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology*, 317(1-2), 63-80.

Sahoo, G., Ray, C. & De Carlo, E., 2006. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *Journal of Hydrology*, 327(3-4), 525-538.

Sajikumar, N. & Thandaveswara, B.S., 1999. A non-linear rainfall-runoff model using an artificial neural network. *Journal of Hydrology*, 216(1-2), 32-55.

Schmitz, G.H. & Cullmann, J., 2008. PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments. *Journal of Hydrology*, 360(1-4), 1–14.

Sudheer, K.P., Gosain, A.K. & Ramasastri, K.S., 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes*, 16(6).

Sudheer, K.P., Nayak, P.C. & Ramasastri, K.S., 2003. Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, 17(3), 677–686.

Thiery, D., Bérard, P. & Camus, A., 1983. Captage de la source du Lez. Etude de relation entre la source et son réservoir aquifère - rapport n°1 : recueil des données et établissement d'un modèle de cohérence. *BRGM*

Touet, F., 1987. Détermination de l' origine des ressources en eau captées en bordure Sud du pli de Montpellier entre Montpellier et Vendargues (Hérault). PhD. Université Paris-Sud 11.

Toukourou, M.S., 2009. Application de l'apprentissage artificiel à la prévision des crues éclair. PhD. Ecole des Mines de Paris.

Toukourou, M.S., Johannet, A. & Dreyfus, G., 2009. Flash Flood Forecasting by Statistical Learning in the Absence of Rainfall Forecast: A Case Study. *Engineering Applications of Neural Networks*, 98-107.

Vincendon, B., Ducrocq, V., Saulnier, G.M., Bouilloud, L., Chancibault, K., Habets, F., Noilhan, J., 2010. Benefit of coupling the ISBA land surface model with a TOPMODEL hydrological model version dedicated to Mediterranean flash-floods. *Journal of Hydrology*, 394(1-2), 256-266.

Zealand, C.M., Burn, D.H. & Simonovic, S.P., 1997. Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*, 214(1-4), 32-48.

Zhang, G., Patuwo, E.B. & Hu, M., 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.