

Water level forecasting through fuzzy logic and artificial neural network approaches

S. Alvisi¹, G. Mascellani¹, M. Franchini¹, and A. Bárdossy²

¹Dipartimento di Ingegneria, Università degli Studi di Ferrara, Italia

²Institut für Wasserbau, Universität Stuttgart, Germany

Received: 29 May 2005 – Published in Hydrology and Earth System Sciences Discussions: 22 June 2005

Revised: 16 September 2005 – Accepted: 18 December 2005 – Published: 8 February 2006

Abstract. In this study three data-driven water level forecasting models are presented and discussed. One is based on the artificial neural networks approach, while the other two are based on the Mamdani and the Takagi-Sugeno fuzzy logic approaches, respectively.

All of them are parameterised with reference to flood events alone, where water levels are higher than a selected threshold. The analysis of the three models is performed by using the *same input and output variables*. However, in order to evaluate their capability to deal with different levels of information, two different input sets are considered. The former is characterized by significant spatial and time aggregated rainfall information, while the latter considers rainfall information more distributed in space and time.

The analysis is made with great attention to the reliability and accuracy of each model, with reference to the Reno river at Casalecchio di Reno (Bologna, Italy). It is shown that the two models based on the fuzzy logic approaches perform better when the physical phenomena considered are synthesised by both a limited number of variables and IF-THEN logic statements, while the ANN approach increases its performance when more detailed information is used. As regards the reliability aspect, it is shown that the models based on the fuzzy logic approaches may fail unexpectedly to forecast the water levels, in the sense that in the testing phase, some input combinations are not recognised by the rule system and thus no forecasting is performed. This problem does not occur in the ANN approach.

liable water level forecasts enable the use both of early warning systems to alert the population, and real time control of hydraulic structures, like diversion, gates, etc., to mitigate the flood effects. Information on the flood evolution must be provided with a reasonable lead time to be effective, but this is not an easy task, particularly when only rainfall data observed up to the forecasting time instant are available, without any assumption of future rainfall behaviour.

As is well known, the rainfall-runoff transformation processes leading to the formation of a flood wave and to the consequent water level evolution in the river, is extremely complex being non-linear, time varying and spatially distributed (Singh, 1964; Pilgrim, 1976). A plethora of rainfall-runoff models have been proposed and used in the past (see, for example, Todini, 1988; Franchini and Pacciani, 1991; Singh and Woolisher, 2002, for a general classification and analysis). Among them, the most widely used for flood forecasting purposes have a conceptual structure with different levels of physical information (e.g. Stanford Watershed Model IV (Crawford and Linsey, 1966), Sacramento (Burnash et al., 1973), TOPMODEL (Beven and Kirby, 1979; Beven et al., 1984; Sivapalan et al., 1987), ARNO (Todini, 1996), TOPKAPI (Liu and Todini, 2002)), or a stochastic structure (typical examples are those based on ARMA and/or ARIMA structures – see, for example, Montanari et al., 2000; Toth et al., 2000). There are also examples where the two approaches are combined to improve the forecasting performance (Ferraresi et al., 1990; Todini, 1988).

In the last decade a new type of data-driven models, based on artificial intelligence and soft computing technique have been applied. In particular, Artificial Neural Network (ANN) is one of the most widely used technique in the forecasting field (e.g. Hsu et al., 1995; Shamseldin, 1997; Thimuralaiah and Deo, 2000). More recently, Fuzzy Logic (FL) (e.g. Hundecha et al., 2001; Özelkan and Duckstein, 2001; Chang et al., 2005), Model Trees (e.g. Solomatine and Dulal, 2003) and hybrid systems based on both ANN and FL have also

1 Introduction

Water level forecasting is important for environmental protection and flood control since, when flood events occur, re-

Correspondence to: S. Alvisi
(salvisi@ing.unife.it)

been used (e.g. See and Openshaw, 1999, 2000; Abrahart and See, 2002).

Most applications based on these models consider the discharge as forecasting variable (e.g. Imrie et al., 2000; Dawson et al., 2002; Moradkhany et al., 2004), probably because of a historical contiguity with the classes of conceptual and physically based rainfall-runoff models. Such an approach requires the knowledge of the rating curve in the cross section of interest (i.e. the basin outlet) to parameterize the model. However, the knowledge of the water level is required within the framework of a flood warning system and thus the rating curve has to be used also to transform the forecasted flows into water levels.

Models based on ANN, FL, etc., and, more in general, all data-driven models, can be designed to forecast water levels directly, given their very nature. For example, Campolo et al. (1999), See and Openshaw (1999), See and Openshaw (2000), Thirumalaiah and Deo (2000), Campolo et al. (2003), Young (2001, 2002), develop models based mainly on applications of the ANN techniques, while Krzysztofowicz (1999, 2001), Krzysztofowicz and Kelly (2000), Krzysztofowicz and Herr (2001) develop a Bayesian forecasting system which produces a short-term probabilistic river level forecast based on a probabilistic quantitative precipitation forecast.

The possibility of operating directly on the levels does not apply to the conceptual rainfall runoff models, which are based on the respect of the mass conservation and even less to the physically based rainfall runoff models, which, beyond the respect of the mass conservation at each step of the rainfall runoff transformation, include further equations reflecting details on the energy balance. As a consequence, these latter models can only deal with the flow.

As previously mentioned, the water level is the quantity of interest within the framework of a flood warning system. Furthermore, it frequently happens that a basin is closed by a cross section where a long series of registered levels is available but, at the same time, the rating curve is unknown. In this latter situation, the only models that can be used to perform a forecast are those of data-driven type directly applied to the water levels.

Within this framework, an analysis of two data-driven water level forecasting approaches is developed: one approach is based on artificial neural networks, whereas the other is based on fuzzy logic. These two models are selected since they are similar, that is, both of them create a quantitative *inner* chain of links between input and output quantities without explaining/justifying the physical reason of those links. They mimic, in fact, the intuitive human way of relating *causes* with their *effects*.

Several other papers perform comparisons between data-driven/black box models (see, for example, Goswami et al., 2005), usually focusing on model precision both in calibration and validation phase. In this paper the analysis of the selected models is made with great attention to precision, re-

liability and capability to deal with different levels of information.

Reliability is considered since the abandoning of any physical constraint such as the mass conservation (typical of any data-driven model) can represent a potential risk of unexpected failures in the forecasts, outside the calibration phase.

The capability to deal with different levels of information is also considered since the FL approach links input and output through a decomposition based on categories (low, high, etc.) while the ANN approach links input and output through a system of numerical weights. Indeed the two approaches can behave differently when facing an increasing amount of detail and for this reason two different input sets are considered in the numerical test: the former is characterized by significant spatial and time aggregated rainfall information, while the latter considers rainfall information more distributed in space and time.

The structure of the paper is as follows. The architecture of the ANN and FL models is first presented. Subsequently, with reference to the Reno river at Casalecchio di Reno (Bologna, Italy), two different experiments, characterised by different levels of information in the input sets, are set up and discussed, highlighting the behaviour and reliability of the two approaches.

2 Architecture of the ANN and FL models

2.1 The Artificial Neural Network model

Artificial neural networks reproduce the behaviour of the brain and nervous systems in a simplified computational form. They are constituted by highly interconnected simple elements, called artificial neurons, which receive information, elaborate them through mathematical functions and pass them to other artificial neurons. In particular, in multi-layer perceptron feed-forward networks (Rosenblatt, 1958; Hagan et al., 1996), the artificial neurons are organized in layers: an input layer, one or more hidden layers and an output layer. In this study, one hidden layer is considered, since it is shown that this type of network can approximate any function (Hornik et al., 1989; Maier and Dandy, 2000).

With reference to a generic three layer feed-forward network with n_p input neurons, n_h hidden neurons and n_o output neurons, the input vector \mathbf{p} , consisting of n_p elements, is multiplied by a weight matrix \mathbf{WP} ($n_h \times n_p$) generating a vector that is summed with a bias vector \mathbf{bp} (n_h). In the hidden layer neurons, each element of the vector obtained is transformed using a nonlinear transfer function f_h , thus generating the vector \mathbf{h} (n_h):

$$\mathbf{h} = f_h (\mathbf{WP}\mathbf{p} + \mathbf{bp}) \quad (1)$$

The same procedure is repeated in the output layer. Thus, the vector \mathbf{h} is multiplied by a weight matrix \mathbf{WO} ($n_o \times n_h$) generating a vector that is summed with a bias vector \mathbf{bh} (n_o).

In the neurons of the output layer each element of the vector obtained is transformed using a nonlinear transfer function f_o generating the output vector \mathbf{o} (n_o):

$$\mathbf{o} = f_o(\mathbf{WHh} + \mathbf{bh}) \quad (2)$$

In particular, the transfer functions f_h and f_o used in this study in the hidden and output layers, respectively, are a hyperbolic tangent sigmoid transfer function,

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

and a logsigmoid transfer function,

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

where x is the generic element of the vectors $\mathbf{WPp} + \mathbf{bp}$ and $\mathbf{WHh} + \mathbf{bh}$. Other different transfer functions were tested, but the attention was focused on functions (3) and (4) since they produce better results and show high flexibility without increasing model parameterization. In order to avoid the problem of output signal saturation (Smith, 1993; Hsu et al., 1995;) the input datum p_i is standardized in the range [0.05:0.95] through:

$$p_i^{\text{norm}} = 0.05 + 0.9 \frac{p_i - p_{\min}}{p_{\max} - p_{\min}} \quad (5)$$

where $[p_{\min}, p_{\max}]$ is the variation range of the input variable considered.

In summary, the full definition of an ANN model implies the quantification of the number of neurons in the hidden layer and the weight values, since the neuron numbers in the input and output layers are fixed by the numbers of input and output variables, respectively.

As regards the neuron number in the hidden layer, it is usually defined by a trial and error procedure, searching for the lowest number of neurons without penalizing model efficiency (Hsu et al., 1995; Zealand et al., 1999; Chiang et al., 2004).

As regards the quantification of the weight values, two different algorithms are frequently used to train the model: the Levenberg Marquardt algorithm (Hagan and Menhaj, 1994) and the scaled conjugate gradient algorithm (Moller, 1993). The former algorithm seems to perform better with ANN models characterized by few neurons, and thus few weights, while the latter with ANN models characterized by many neurons, and thus many weights, (Demuth and Beale, 2000).

In order to avoid overfitting and to improve the ANN model robustness, an early stopping procedure is used (ASCE, 2000; Demuth and Beale, 2000). In this procedure three data sets are considered: a training, a validation and a testing set. The first and the second subsets are used to set up the model, the third subset to test it. More in detail, the first subset is used for training the model. At each training step, the calibrated model is validated using the second subset. While the first training steps are performed, the error

decreases, as it does in the corresponding validation phase. As the model begins to overfit the data, the error in the validation phase begins to rise and thus the training procedure is stopped. The artificial neural network model was implemented in Matlab environment where both Levenberg Marquardt and scaled conjugate gradient training techniques are available in the Neural Network Toolbox.

2.2 Fuzzy logic model

A fuzzy logic model (Zadeh, 1973) is a logical-mathematical procedure based on a “IF-THEN” rule system that allows for the reproduction of the human way of thinking in computational form. In general, a fuzzy rule system has four components:

- fuzzification of the input: process that transforms the “crisp” (traditional) input into a fuzzy input;
- fuzzy rules: IF-THEN logic system that links the input to the output variables;
- fuzzy inference: process that elaborates and combines rule outputs;
- defuzzification of the output: process that transforms the fuzzy output into a crisp output.

The most widespread methodologies for developing fuzzy rule systems are those proposed by Mamdani (1974) and Takagi-Sugeno (1985). The Mamdani method (FL-M) follows exactly the above mentioned scheme, whereas the Takagi-Sugeno method (FL-TS) uses a composite procedure for fuzzy inference and output defuzzification. In this study, both methods are used for developing two different forecasting models.

With reference to the Mamdani method, being the k th crisp input variable defined as a_k , $\hat{A}_{i,j,k}$ its corresponding j th fuzzy input number considered in the i th rule and $\hat{B}_{i,l}$ the l th fuzzy output number relevant to the i th rule, the generic Mamdani rule $(R_i)_M$ is:

$$(R_i)_M : \begin{array}{l} \text{IF } a_1 \text{ is } \hat{A}_{i,j,1} \text{ AND } a_2 \text{ is } \hat{A}_{i,j,2} \text{ AND } \dots \text{ AND } a_K \text{ is } \hat{A}_{i,j,K} \\ \text{THEN } \hat{B}_{i,l} \end{array} \quad (6)$$

In the algorithm developed in this study, the degree of fulfillment v_i of the i th rule is obtained with the product inference procedure (Larsen, 1980), then the weighted sum combination is used to define the final output membership function μ_B generated by the fuzzy rule system for the crisp input vector (a_1, \dots, a_K) (Bardossy and Duckstein, 1995). Finally the crisp output number b is obtained by applying the centroid defuzzification method to μ_B .

With reference to the Takagi-Sugeno method, being the k th crisp input variable defined as a_k , $\hat{A}_{i,j,k}$ its corresponding

Table 1. Hydrometeorological and morphological information about the Reno river basin at Casalecchio di Reno (Bologna, Italy). \bar{Q}_{peak} is the mean value of the annual maximum peak discharges, H_m is the mean altitude of the basin respect to the basin outlet, L is the length of the Reno river upstream from Casalecchio di Reno, T_c is the time of concentration, S is the mean bed slope and P_{year} is the yearly average areal rainfall depth.

Basin	\bar{Q}_{peak} (m ³ /s)	Area (km ²)	H_m (m)	L (km)	T_c (h)	S (%)	P_{year} (mm)
Reno at Casalecchio	767.71	1051	581	84.2	12	1.18	1336.00

Table 2. Main characteristics of training, validation and testing sets. ΔW is the hourly water level variation, P_{max} is the maximum hourly areal rainfall and $P_{c_{\text{max}}}$ is the maximum 12 h cumulated areal rainfall.

Phase	n° of events	n° of patterns	Peak Flood (cm)		ΔW (cm)		P_{max} (mm)	$P_{c_{\text{max}}}$ (mm)
			min	max	min	max		
Training (10/1994–12/1999)	36	1549	54	357	−46	119	15.20	66.09
Validation (09/1993–09/1994)	9	386	79	250	−33	151	12.15	63.25
Testing (03/2000–12/2000)	7	343	80	317	−42	126	14.43	73.25

j th fuzzy input number considered in the i th rule, the generic Takagi-Sugeno i th rule $(R_i)_{TS}$ is:

$$(R_i)_{TS} : \begin{array}{l} \text{IF } a_1 \text{ is } \hat{A}_{i,j,1} \text{ AND } a_2 \text{ is } \hat{A}_{i,j,2} \text{ AND...AND } a_K \text{ is } \hat{A}_{i,j,K} \\ \text{THEN } b_i = c_{i,0} + c_{i,1}a_1 + \dots + c_{i,K}a_K \end{array} \quad (7)$$

where b_i is the crisp output and $c_{i,k}$ are the parameters of the linear equation relevant to the i th rule.

Eventually, the final crisp output b is given by a linear combination of the outputs of the activated rules:

$$b = \frac{\sum_{i=1}^I v_i \left[c_{i,0} + \sum_{k=1}^K c_{i,k}a_k \right]}{\sum_{i=1}^I v_i} = \frac{\sum_{i=1}^I v_i b_i}{\sum_{i=1}^I v_i} \quad (8)$$

where v_i is the degree of fulfilment of the i th rule obtained through the product inference procedure (Larsen, 1980).

In summary, having fixed the inference, combination and defuzzification procedures, the full definition of a FL model, when triangular membership functions are considered, requires, in the case of the Mamdani method, a decision on (a) how many fuzzy numbers to define for each input variable, (b) their supports and peaks, (c) how many fuzzy numbers to define for each output variable, (d) their supports and peaks, and (e) the number of rules and the corresponding IF-THEN relationships. In the case of the Takagi-Sugeno method, a decision is necessary about the linear equations defining the crisp values shown in Eqs. (7) and (8) since the output fuzzy numbers, the rules combinations and the defuzzification process are substituted by a composite procedure.

The input (and output, for the Mamdani approach) membership functions can be generated on the basis of the frequency of the observed values in the crisp data set used for parameterization of the model (Bardossy and Samaniego, 2002), whereas the IF-THEN rule system (and the coefficients of the linear equation producing the crisp output numbers, in the case of the Takagi-Sugeno approach) can be obtained through an optimization process based on the simulated annealing algorithm (Kirkpatrick et al., 1983).

As for the ANN model, an early stopping procedure is applied for the parameterization of both the FL models used in this study. The fuzzy logic models were implemented by using the FORTRAN programming language.

3 The set up of the numerical experiment

The data used in this study refer to the Reno river basin closed at Casalecchio di Reno (Bologna, Italy). This basin is located in the central part of the Emilia Romagna region and drains an area of 1051 km². The main river flows for 84.4 km through the basin (south-north direction), and the time of concentration of the basin is about 12 h. Further hydrometeorological and morphological information is reported in Table 1. A schematic representation of the basin and the location of the 24 rain-gauges considered in this study are shown in Fig. 1.

The available hourly water level data cover the period 1993–2000. Hourly rainfall data are available in the 24 rain-gauges as well. A set of 45 flood events observed in the years 1993–1999 is selected and used for setting up both types of

models; in particular, since an early stopping procedure is used, this set is split into 2 subsets: 36 flood events (observed in the period 10/1994–12/1999) are used for training the model and 9 flood events (observed in the period 09/1993–09/1994) for its validation. A set of 7 flood events observed in the year 2000 is selected and used for testing the models. The main characteristics of training, validation and testing sets are summarized in Table 2. The events selected are all characterised by a water level peak equal to or greater than 50 cm above the reference zero of the measuring station at Casalecchio di Reno. The initial and the final points of the flood events were marked in the water level time series where the periodic oscillations, due to anthropic activities, terminate and begin again, respectively.

As previously written, the aim of this study is to compare and analyse fuzzy logic and neural network approaches for setting up data driven water level forecasting models. To perform this experiment, the same input and output variables are considered for both the approaches.

The input variables are represented by rainfall volume over the basin estimated through the rain-gauges and water level measurements, both evaluated at and before the forecasting time t . More precisely, both approaches are set up to link the rainfall and the water levels, registered until the time t , to the water level variations $\Delta W(t, t+m\Delta t)$ at the basin outlet for 5 selected time horizons $m\Delta t$ ($m=1, 3, 6, 9, 12$; $\Delta t=1$ h), at least in the initial part of the study. It is worth stressing immediately, that one forecasting model has been set up for each time horizon, thus producing 5 ANN models, 5 FL-M models and 5 FL-TS models.

The use of water level variation $\Delta W(t, t+m\Delta t)$ with respect to the current value, instead of the water level itself $W(t, t+m\Delta t)$, is selected for the following two reasons. The first reason is related to the consequence on the final result (i.e. the water level at $m\Delta t$ time steps ahead) of the same relative error on the output variable. When the output variable is the water level variation, the forecasted level is $W(t) + \Delta W(t, t+m\Delta t)$ and the relative error acts only on the second component which is usually a fraction of the first one. On the contrary, when the output variable of the model is directly the level $W(t+m\Delta t)$, the same relative error produces a higher absolute error on the forecasted water level. The second reason is related to the calibration of the model, independently whether it is based on the ANN or FL approach. While a specific value of the water level can be present both in the rising and depletion phase, the corresponding water level variation gives more precise information about the phase, which is important information for “training” both types of models.

It is worth noting that no rainfall data are considered known or forecasted within the time interval $m\Delta t$. This assumption limits the time horizon extension since long term water level forecasts are significantly influenced by the amount of rainfall fallen within the time interval $m\Delta t$ (Toth et al., 2000). Nonetheless, the target of this study is the com-

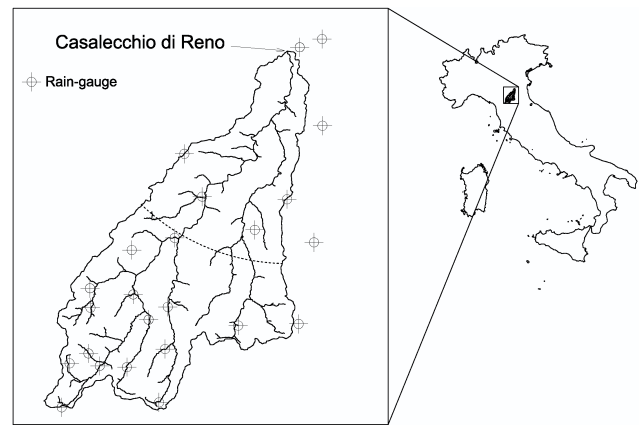


Fig. 1. The Reno river basin at Casalecchio di Reno outlet and position of the rain-gauges. The dashed line subdivides the basin into two sub-regions used to define the DRI input data set.

parison of the two approaches and the evaluation of their reliability and not the implementation of a sophisticated water level forecasting model which can obviously gain higher efficiency and robustness by using this further information.

In order to evaluate the capability of the two approaches to deal with different levels of information, two different input sets are considered in this study: the former is characterized by a significant spatial and time aggregated rainfall information (ARI), while the latter considers more distributed rainfall information both in space and time (DRI).

3.1 Input variables

As regards the ARI input, the following variables aggregating spatial and time rainfall information are considered:

- $P_c(t-12\Delta t, t)$: the cumulated areal rainfall in the 12 h before the forecasting time t , Δt being equal to 1 h (the time spell of 12 h is selected because this is the time of concentration of the basin; see, for example, Anctil and Rat, 2005);
- $S(t)$: the “temporal dissymmetry coefficient” indicating the position of the rainfall peak into the time interval $[t-12\Delta t, t]$ and defined by:

$$S(t) = \frac{\sum_{n=1}^N \left[\left(-\frac{N+1}{2} + n \right)^3 \cdot P(t-n\Delta t, t-(n-1)\Delta t) \right]}{\sum_{n=1}^N P(t-n\Delta t, t-(n-1)\Delta t)} \quad (9)$$

where $N=12$ is the number of the time steps into the time interval $[t-12\Delta t, t]$ and $P(t-n\Delta t, t-(n-1)\Delta t)$ is the areal rainfall registered in the time interval $[t-n\Delta t, t-(n-1)\Delta t]$.

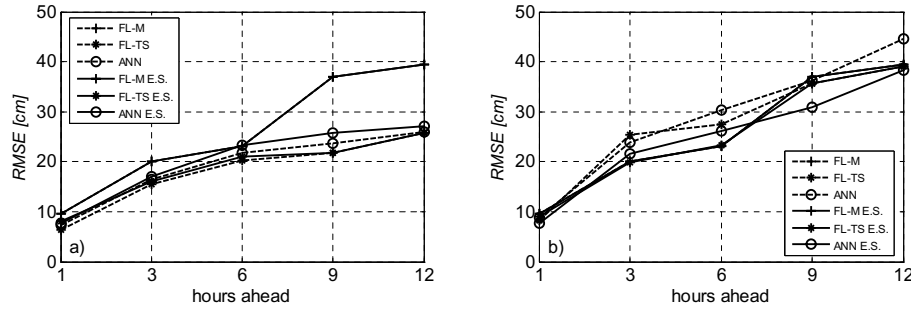


Fig. 2. ARI input data set. *RMSE* relevant to the FL-M, FL-TS and ANN models. (a) training phase, (b) testing phase. Solid and dashed lines are relative to the models trained with and without early stopping procedure, respectively.

As previously mentioned, the comparison between the two different data-driven approaches for water level forecasting is made by using the same input sets. To do this, the input sets are necessarily selected a priori and the criterion used for their selection is “physical” representativity.

With this in mind, the variables $P_c(t-12\Delta t, t)$ and $S(t)$ are used to summarize the total amount and the time distribution of the rainfall within a time interval equal to the time of concentration of the basin. Obviously, their combination in some way affects the time evolution and position of the peak of the water levels at the basin outlet.

As regards the DRI input, it is characterized by more distributed information both in space and time. The basin is divided into two sub-regions (see Fig. 1) and the following 7 input variables are defined for each sub-basin:

- $P(t-n\Delta t, t-(n-1)\Delta t)$ with $n=1, \dots, 6$: the hourly areal rainfall registered in the time interval $[t-6\Delta t, t]$;
- $P_c(t-12\Delta t, t-6\Delta t)$: the cumulated areal rainfall in the time interval $[t-12\Delta t, t-6\Delta t]$.

Both the ARI and DRI inputs consider these 2 further input variables describing the water level status:

- $W(t)$: the current water level at time t .
- $\Delta W(t-\Delta t, t)$: the water level variation with respect to the previous hour.

The current water level gives the reference point to the forecasting. The water level variation in the previous hour gives two important indications: the module indicates the flood variation rate, the sign determines the flood phase (rising or depletion).

To summarise, the ARI input data sets is characterised by 4 variables, while the DRI input is characterised by 16 variables. In the following, results relevant to the ARI input are discussed and a sensitivity analysis with respect to the rate of positive and negative water level variations in the training input data set is performed. Subsequently, the results of the two approaches, i.e. ANN and FL approaches, are compared and discussed with reference to the DRI input.

4 Analysis of results

4.1 Aggregated Rainfall Information (ARI) input data set

The 5 ANN models, one for each of the 5 time horizons, are parameterized by using the 45 flood events previously selected. In particular, the ANN architecture of each single forecasting model is characterized by 4 neurons in the input layer (according to the ARI input), 12 in the hidden layer (number obtained by trial and error procedure, ranging between 5 and 20 neurons) and 1 in the output layer. The Levenberg Marquardt algorithm is used for training the model since it is characterised by a low number of inputs.

Similarly, the 5 + 5 FL models are set up on the same flood events. As regards the FL-M, the input and output membership functions are generated on the basis of the frequency of the observed values in the crisp input data set (Bardossy and Samaniego, 2002), resulting in 10 input triangular fuzzy numbers and 7 output triangular fuzzy numbers. The number of rules is set to 20: a smaller number produces worse results, a larger number increases the computational time enormously and does not improve the performance of the model. A similar parameterisation procedure is applied to the 5 FL-TS models: 10 input triangular fuzzy numbers are generated on the basis of the frequency of the observed values in the crisp input data set (Bardossy and Samaniego, 2002). In the same as the FL-M model, the number of rules for the FL-TS has been selected through a trial-and-error procedure. The optimal number for this method is set to 10: both smaller and larger number of rules produce worse results. It can be observed that the optimal number of rules in the FL-TS formulation is lower than that of the FL-M formulation. This difference can be explained according to their different architectures (see Sect. 2.2): the “consequent” of the FL-M rules is represented by a single fuzzy number, and thus just 1 parameter is calibrated (i.e., the integer number which identifies the output fuzzy number to be related to the generic rule, see Eq. 6), whereas the “consequent” of the FL-TS rules is represented by a linear combination of the input variables, and thus $K+1$ parameters are calibrated for each rule (in this

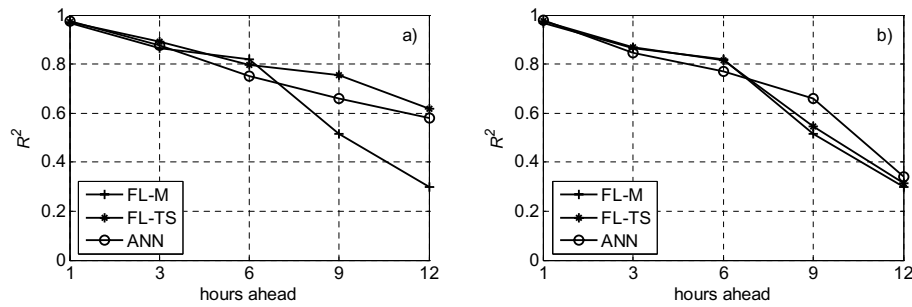


Fig. 3. ARI input data set. R^2 relevant to FL-M, FL-TS and ANN models. (a) training phase, (b) testing phase.

case, $K+1$ is equal to 5, since 4 is the number of the input variables, see Eq. 7). In other words, the performances of the two models are mainly controlled by the number of rules (each of them connected to a single parameter) in the case of the FL-M formulation and by the combined effect given by the number of rules and parameters in the case of the FL-TS formulation since, in this last case, each rule is connected to a higher number of parameters.

The forecasting accuracy of all the models considered in this study is compared through the root mean square error (*RMSE*) and the coefficient of determination (R^2). In Fig. 2, the statistics relevant to the results in training and testing phases are compared. It can be observed that in the training phase (Fig. 2a) the *RMSEs* of the FL-TS and ANN models (trained with the early stopping procedure – solid line), are similar even if the FL-TS model shows smaller *RMSEs*, ranging from 8 to 26 cm for 1 to 12 h ahead, respectively. Still with reference to the training phase, the FL-M model shows higher *RMSEs* than the FL-TS and ANN models, particularly at 9 and 12 h ahead. On the other hand, in the testing phase, (Fig. 2b) the forecasting accuracy of the FL-M and FL-TS models is almost the same at all the forecasting lead times and up to 6 h ahead these two models perform slightly better than the ANN model.

Still referring to Fig. 2, it can also be observed that the early stopping procedure is useful to avoid over-fitting and to improve robustness for both the ANN and FL-TS models, whereas it seems to have no effect in the case of the FL-M model. In fact, with reference to the *RMSE* observed when the models are trained without the early stopping procedure (dashed line), its values are slightly smaller in the training phase in the case of the FL-TS and ANN models, but, in testing phase, the performance of these models decreases significantly (the *RMSEs* increase by 5–10 cm) at all the lead times. These results show that these two models, when trained without the early stopping procedure, tend to work better in the training phase and worse in testing phase. On the other hand, the performance of the FL-M is independent of the early stopping procedure: in fact it has been observed that the trend of the objective function in the validation phase is always concordant with that observed in the

training phase. This may be due to its very structure which does not present the many coefficients which are instead included in the composite inference and defuzzification phase of the FL-TS rule system.

With this point clarified, from now on, only the results produced by the ANN, FL-M and FL-TS models parameterized with the early stopping procedure will be discussed to maintain homogeneity among the training techniques.

The considerations developed with reference to the *RMSE* are confirmed by the analysis of the coefficient of determination R^2 relevant to the training (Fig. 3a) and testing (Fig. 3b) phase. In particular, it can be observed that all the models produce an accurate 1 h ahead forecast, both in training and testing phase ($R^2 \simeq 0.97 \div 0.98$); the performances decrease as the lead time increases, showing, in testing phase, $R^2 \simeq 0.87 \div 0.78$ for the 3 and 6 h ahead forecasts where the higher values of R^2 are relevant to the FL models. The R^2 definitely decreases for the 9 and 12 h ahead forecasts. It seems then that the information available before the forecasting time instant is not sufficient to perform reliable forecasts for a time spell greater than 6 h, at least in the basin considered here. Certainly, more efficient forecasts might be done if information on “future” rainfall were available. However, as already written, this case is not considered in this paper and the subsequent discussion will be restricted to a time spell of 6 h ahead, since the results for greater time horizons are too poor. It is worth noting that these limits in the forecasting time spells are reported in several other studies, where similar methodologies and basin dimensions are considered (see, for example, Campolo et al., 1999, 2003; See and Openshaw, 1999; Solomatine and Dulal, 2003; Kim and Barros, 2001).

The analysis of the scatter plots of Fig. 4 (in which, referring to the testing phase, the water levels forecasted at 1, 3 and 6 h ahead by FL-M, FL-TS and ANN models are plotted versus the observed water levels) shows that the models behave in slightly different ways. With reference to the one hour lead time, the forecasted water levels lie close to the 45° line, which represents the exact fit. As the lead time increases to 3 and 6 h, the dispersion increases, and particularly for the ANN model, the highest points tend to lie below the 45° line: so this model tends to underestimate the higher water levels.

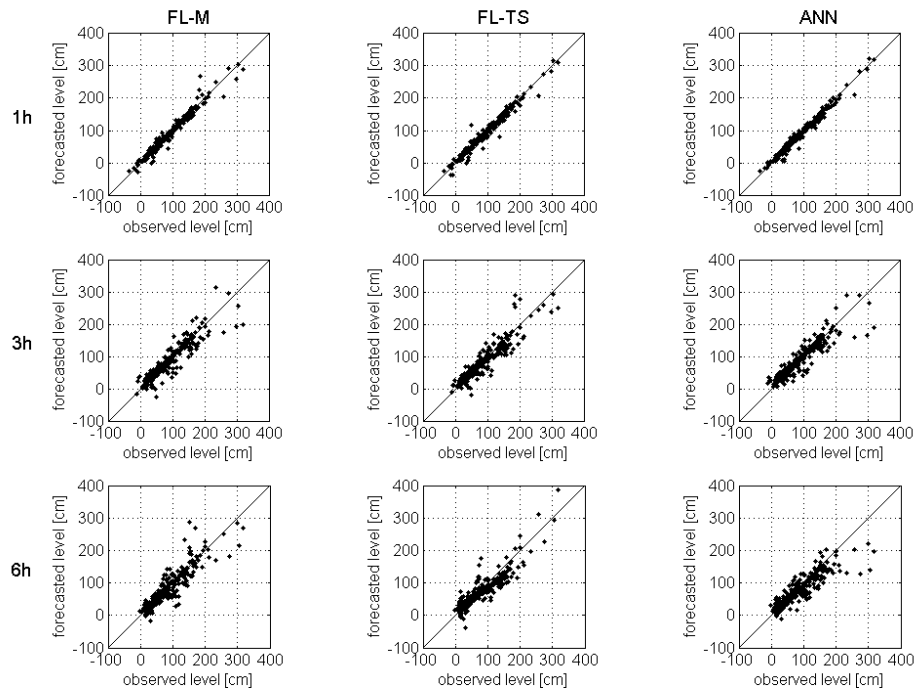


Fig. 4. ARI input data set – testing phase. Water levels forecasted at 1, 3 and 6 h ahead (rows 1, 2 and 3, respectively) by the FL-M, FL-TS and ANN models (columns 1, 2 and 3, respectively) versus the observed data.

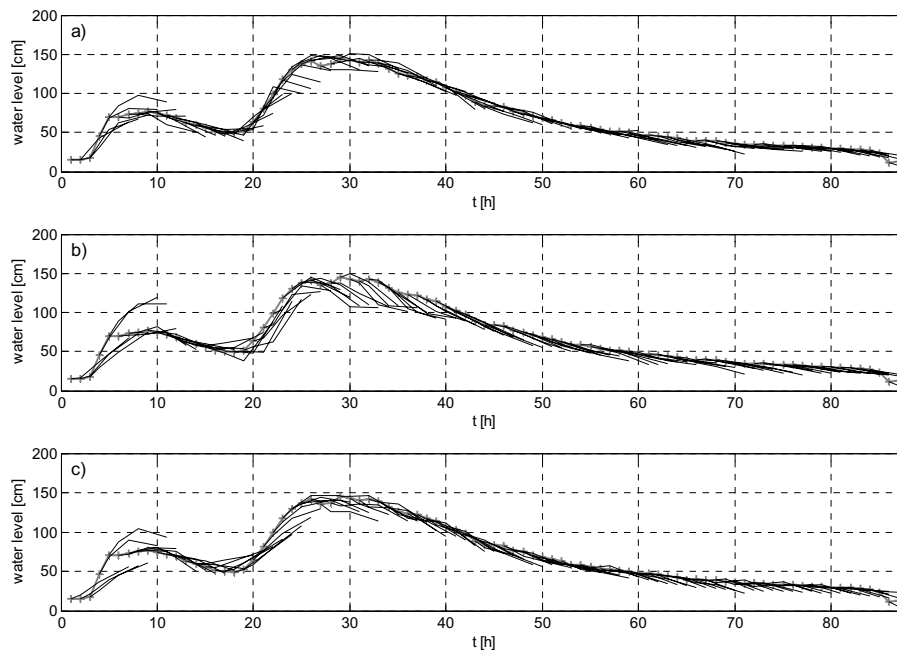


Fig. 5. ARI input data set – testing phase. Ensemble of the water levels forecasted at 1, 3 and 6 h ahead starting from each hour of the flood event considered. (a) FL-M model, (b) FL-TS model, (c) ANN model.

To compare the models in more detail, an event of the testing set is shown in Fig. 5 where the 1, 3 and 6 h ahead forecasts, calculated at each time step, are plotted. In this specific event, the FL-M, FL-TS and ANN models show a

similar behaviour: in all the cases the greatest differences between observed and forecasted water levels occur in the rising part of the wave, where a marked underestimation is observed. The similar behaviour of the FL-M, FL-TS and

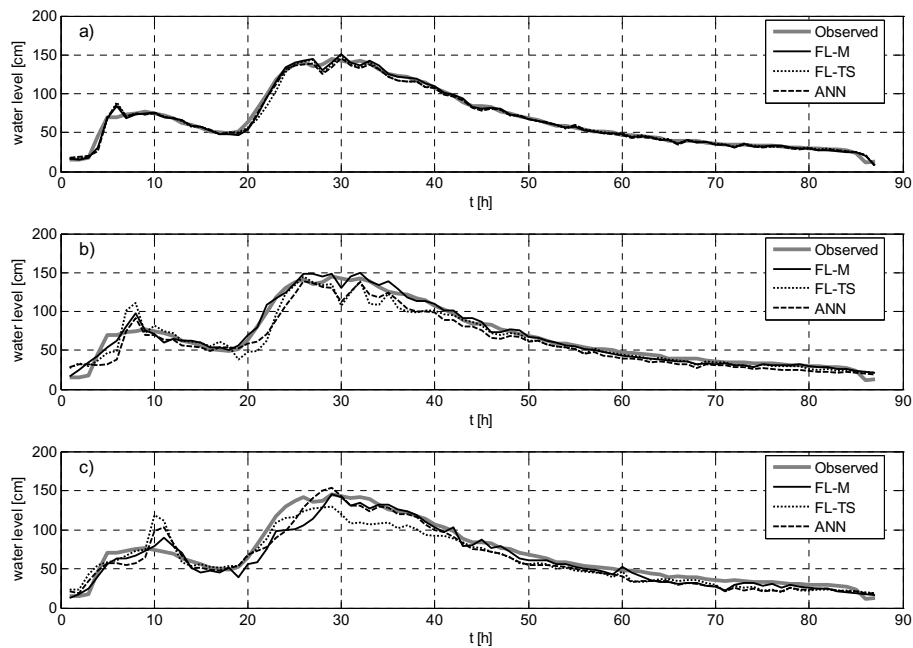


Fig. 6. ARI input data set – testing phase. Water levels forecasted (a) 1 h, (b) 3 h and (c) 6 h ahead by the three models.

ANN models is also confirmed by Fig. 6 in which all the levels forecasted at 1, 3 and 6 h ahead are connected among them. It may be worth noting that all three models produce similar false peaks, particularly between the first and the second rising phase of the flood event, whereas the forecast of the depletion phase of the flood event is good for all the three models.

On the whole, the results obtained by using the ARI input data sets (which is characterised by a low number of input variables) highlight that the FL-M and FL-TS models present, with reference to forecasts up to 6 h ahead, a slightly higher accuracy in the testing phase than that shown by the ANN model (see Fig. 3). Nonetheless, similar underestimation in forecasting the rising part of the flood event is observed in all the models.

4.2 Sensitivity analysis to the data set

As shown in Fig. 5, all the models tend to underestimate the water level in the rising part of the flood waves, particularly in the case of the 6 h ahead forecasts. Analysis of the training water level variation data set $\Delta W(t, t+m\Delta t)$ shows that the negative variations, mainly relevant to the depletion phase of the floods, are 70% of the whole data set. This is due to the typical shape of the flood events: relatively short rising phase and a long depletion phase. Since the models here considered are data driven, this percentage of negative variations could be the reason for the general underestimation observed in the prediction of the rising part of the flood waves. For this reason, a sensitivity analysis to the rate of positive and negative water level variation data in the training set is developed.

The analysis is structured in the following way: the amount of negative water level variation in training and validation set is progressively reduced from 70% to 0%, so that the percentages of positive water level variations can vary from 30% to 100%, i.e., up to a calibration procedure based only on the rising part of the floods.

These new input data sets are used for training both FL and ANN models again for each forecasting time horizon (1, 3 and 6 h), then the corresponding calibrated models are applied to the whole original testing data set. The results, in terms of scatter plot obtained from the testing set for FL-M, FL-TS and ANN models, are shown in Figs. 7, 8 and 9, respectively.

It can be observed that the forecasts, for any model and any lead time, are not influenced by the percentage of positive-negative water level variations of the training data set, except for the case of 100% positive water level variations, where a general (and obvious) deterioration can be observed. Some deeper insight can be gained by observing Fig. 10 where the flood event shown in Fig. 5 is considered. When the percentage of the positive water level variations in the training input data set passes from 30% (see Fig. 5) to 50% a slight reduction of the underestimation on the rising part of the flood wave is observed in all the three models but, at the same time, an increment of the underestimation is observed in the depletion phase. Similar observation applies to the case where the percentage of the positive variation is equal to 80%. This explains the general behaviour shown in Figs. 7, 8 and 9, which, in fact, summarise the global patterns both in rising and depletion phases. Finally, with reference to the percentage of

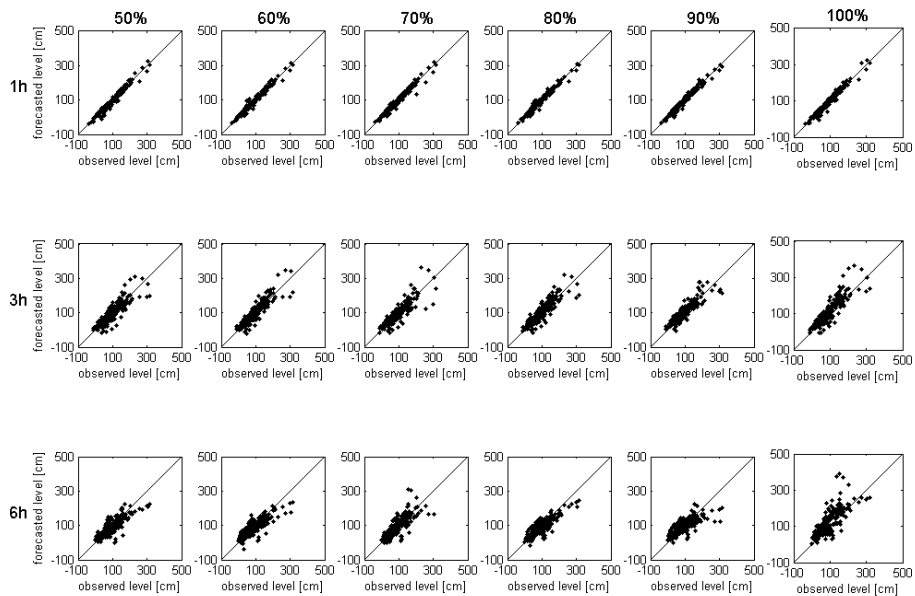


Fig. 7. ARI input data set – testing phase. Scatter plots of the water levels forecasted at 1, 3 and 6 h ahead (rows 1, 2 and 3, respectively) by the FL-M model trained with data set characterized by different percentage of positive water level variations (columns 1–6).

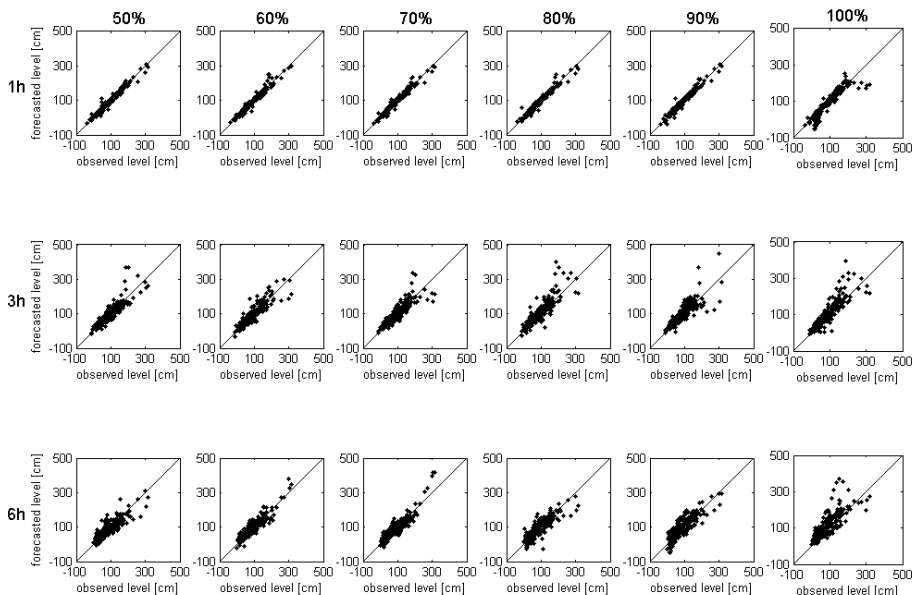


Fig. 8. ARI input data set – testing phase. Scatter plots of the water levels forecasted at 1, 3 and 6 h ahead (rows 1, 2 and 3, respectively) by the FL-TS model trained with data set characterized by different percentage of positive water level variations (columns 1–6).

100%, it is easy to realize that this fraction of positive water level variations would be a bad choice in selecting the input data set.

These results can be read as an indication that the fraction of the positive water level variations in the training input data set is not the key point in improving the forecasts. Other aspects, such as a more detailed information on the rainfall dynamic, are fundamental to improve the general quality of

the forecasts. This aspect will be discussed in the next section making reference to the DRI input data set.

It is now worth stressing that while performing all the previous numerical experiments, mainly dedicated to the analysis of the model efficiency, some problems related to the reliability of the two FL models have been observed. It has been noted that when these latter models are applied, the input training data set can lead to a rule system which is not able

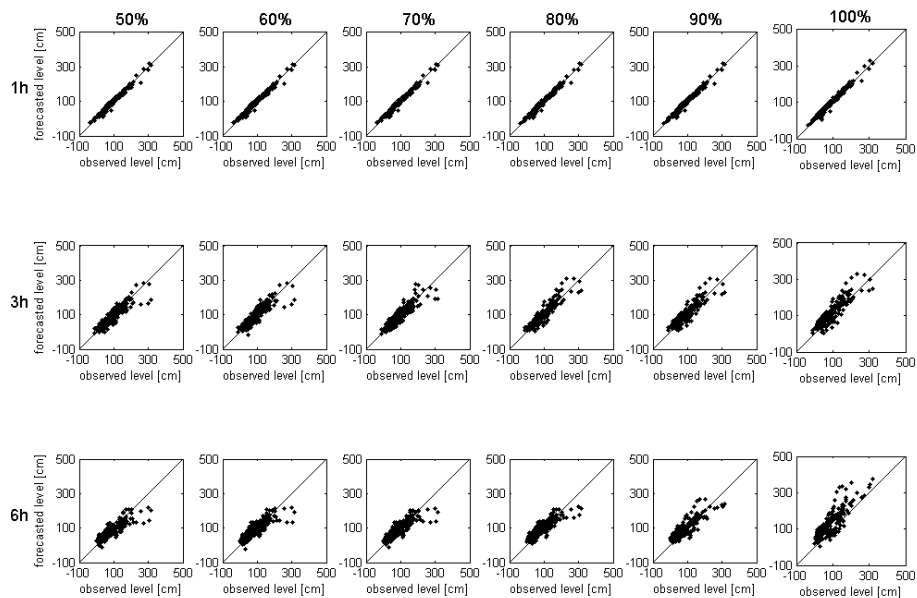


Fig. 9. ARI input data set – testing phase. Scatter plots of the water levels forecasted at 1, 3 and 6 h ahead (rows 1, 2 and 3, respectively) by the ANN model trained with data set characterized by different percentage of positive water level variations (columns 1–6).

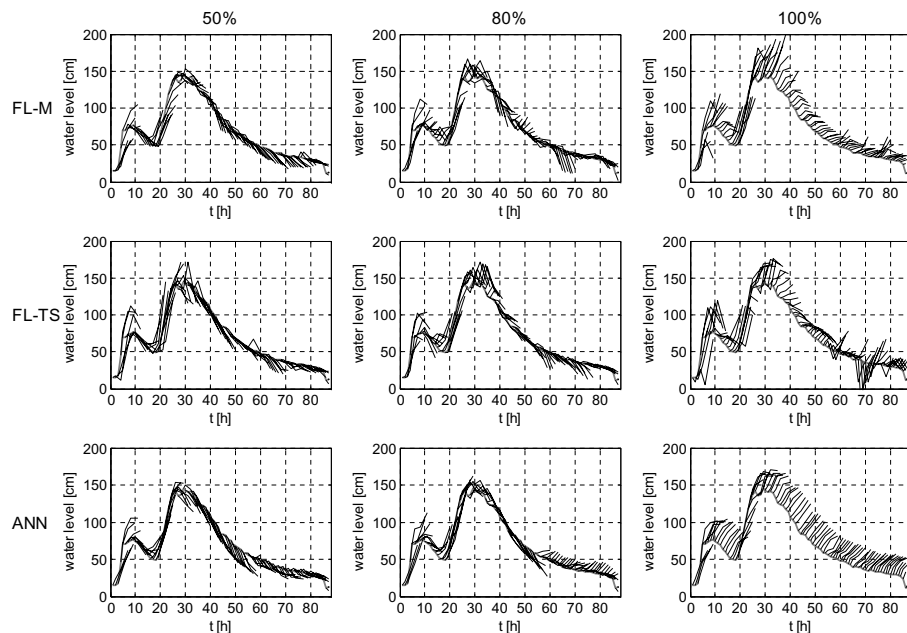


Fig. 10. Ensemble of the water levels forecasted at 1, 3 and 6 h ahead starting from each hour of the flood event considered by FL-M, FL-TS and ANN models (rows 1, 2 and 3, respectively) trained with data set characterized by different percentage of positive water level variations (columns 1–3).

to furnish a result for each testing input data. In other words, it happens that some input vectors do not satisfy any rules during the testing phase, and thus it is not possible to execute the forecast. In particular, considering a total amount of about 6900 cases (number obtained by applying the 985 input vectors of the testing data set to the models calibrated with

the 7 different percentages of data relevant to flood rising and depletion phases), the FL-M rule systems are not satisfied 18 times and the FL-TS 2 times, that is the 0.3% and 0.03% of the total cases, respectively. From a statistical viewpoint these percentages are very low, but, independently of the specific number of failures, what is more important is that these

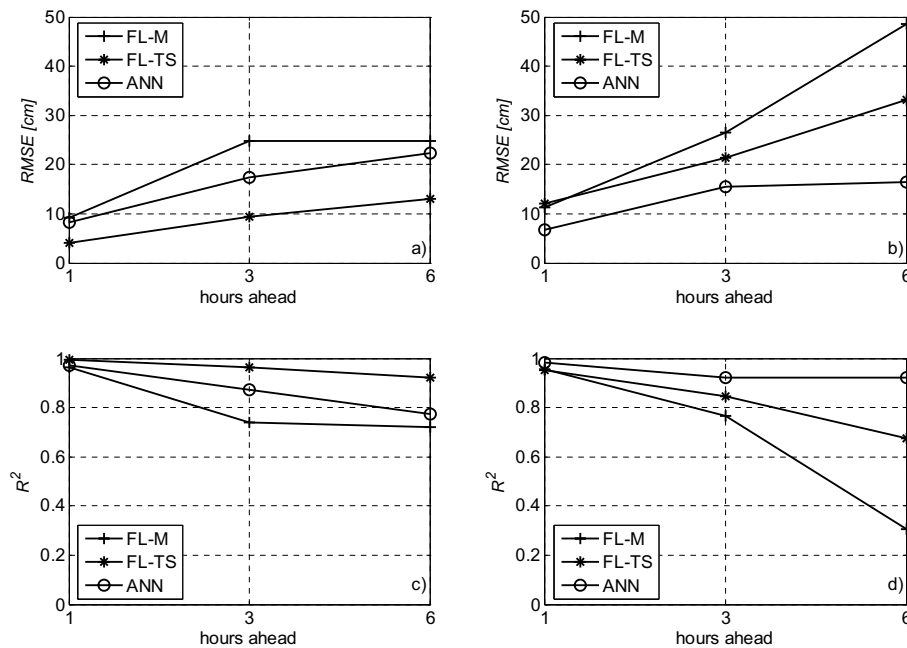


Fig. 11. DRI input data set. $RMSE$ and R^2 relevant to the FL-M, FL-TS and ANN models. (a) training phase, (b) testing phase.

models can fail in an unpredictable way. In fact, it has been observed that the “forecasting failure” does not follow any predictable scheme: these failures are not due to “abnormal” input vectors and no correlation has been detected with the forecasting time horizon, the number of rules and the training data set applied. This (un)reliability of the FL approach could represent a limit in their application in a flood warning system. The ANN model does not present this problem since, given its own architecture, for each input vector \mathbf{p} , an output vector \mathbf{o} is always obtained through Eqs. (1) and (2). Moreover, it is worth stressing that when the input vectors which cause the failure of FL-M and/or FL-TS models are applied to the ANN model, the forecasting accuracy is coherent with the $RMSEs$ presented in Sect. 4.1.

4.3 Distributed Rainfall Information (DRI) input data set

As previously written in Sect. 3.1, the DRI input data set is composed of 16 inputs: the current water level $W(t)$, the water level variation in the previous hour $\Delta W(t-\Delta t, t)$ and, for each of the two sub-regions, the hourly areal rainfall registered in each of the 6 antecedent hours $P(t-n\Delta t, t-(n-1)\Delta t)$, $n=1, \dots, 6$ and the cumulated areal rainfall $Pc(t-12\Delta t, t-6\Delta t)$ registered in the time interval $[t-12\Delta t, t-6\Delta t]$.

The 45 complete flood events previously considered are now used for training and validating all the models with the DRI input data sets.

The ANN models, each of them relevant to a selected time horizon, are characterized by 16 neurons in the input layer, 12 in the hidden layer (number obtained by trial and error

procedure, ranging between 5 and 20 neurons) and 1 in the output layer; the scaled conjugate gradient algorithm is used for training the model, since it is now characterised by a high number of inputs. The FL-M and FL-TS models remain equal to those previously described with reference to the ARI input case.

In the same way as the ARI input data set, the performances of the models are analysed through the root mean square error ($RMSE$) and the coefficient of determination (R^2). In Fig. 11, these statistics relative to the training and testing phases are presented.

With respect to the ARI input data set, the FL-M model presents similar $RMSEs$ in training phase, but the forecasting accuracy decreases in testing phase, particularly with reference to the 6 h ahead forecasts (see Fig. 11a and b). Similarly, the FL-TS method shows better results in the training phase but this improvement is not confirmed in the testing phase. In fact, the $RMSEs$ for the 1 and 3 h ahead forecasts are similar to those obtained with the ARI input set, while the 6 h ahead forecasts are characterised by a $RMSE$ which is worse than that in the case of the ARI input data set. These results are confirmed by analysis of the scatter plots presented in Fig. 12, where both the FL models show a significant increment of the dispersion when the forecasting horizon increases, and, furthermore, that dispersion is greater than that shown in the ARI input case. This is particularly true for the FL-M model.

On the other hand, the ANN model improves the performances significantly, especially in the testing phase, when the DRI input data set is used. In fact, it produces the smallest

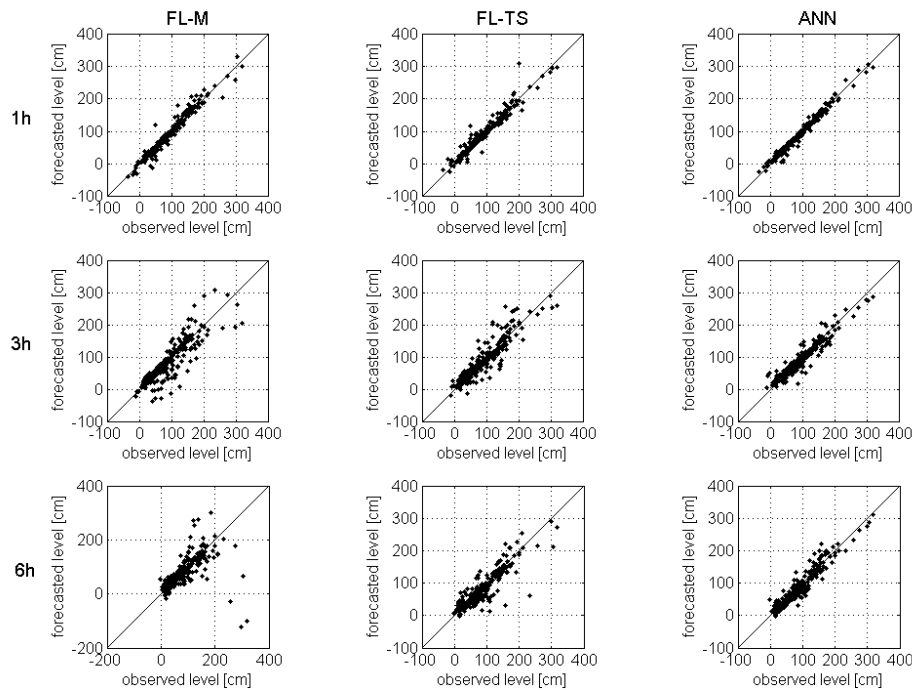


Fig. 12. DRI input data set – testing phase. Water levels forecasted at 1, 3 and 6 h ahead (rows 1, 2 and 3, respectively) by the FL-M, FL-TS and ANN models (columns 1, 2 and 3, respectively) versus the observed data.

RMSEs for all the time horizons, ranging from 6 to 16 cm for 1 to 6 h ahead forecasts, respectively (see Fig. 11b), and a R^2 always greater than 0.9 (see Fig. 11d). The analysis of the scatter plot presented in Fig. 12 shows that all the forecasted water levels lie close to the 45° line for each time horizon. Moreover, when the DRI set is used, the ANN model does not underestimate the water levels. This is probably due to the more detailed rainfall information now used, which thus seems extremely important for forecasting the rising part of the flood event.

Figure 13 shows the same event of the testing set, already used and discussed for the ARI input data set case, where the 1, 3 and 6 h ahead forecasts, calculated at each time step, are plotted. Similarly, in Fig. 14 the same water levels forecasted at 1, 3 and 6 h ahead are shown but this time connecting the values relevant to the same time horizon. The performances of the models observed in terms of *RMSE*, R^2 and scatter plot are confirmed.

Both the FL models present an evident “instability” in the water level forecasts relevant to the rising part and peak of the flood event (see Fig. 14). Overall, the event is forecasted with less accuracy than that observed when the ARI input data sets are used (compare Fig. 13a with Fig. 5a and Fig. 13b with Fig. 5b).

The ANN model fits the observed values very well, with a moderate underestimation in the higher part of the flood (see Figs. 13c and 14c).

Summing up, the results obtained with the DRI input data sets highlight that the increment of input information improves the forecasting accuracy of the ANN model, while this is not true for both the FL models, which, on the contrary, present worse results than those produced when the ARI input data set is used.

Finally, as regards the reliability aspect, none of the FL models present any failure when the DRI input data sets are used. In fact, differently from the ARI input data set, both fuzzy rule systems always furnish a result for each testing input vector. However, it is worth stressing that a lower number (985 Vs 6900) of testing cases is analysed with the DRI input data set, since the testing data set was applied only once, as we did not develop any sensitivity analysis to the percentage of the positive water level variations, which instead had been done with reference to the ARI input data set. Thus, considering the low failure percentage and their unpredictability previously highlighted, it is not possible to exclude, a priori, the possibility of a FL forecasting failure if other or further testing data were available.

5 Discussion and conclusions

This paper addresses the problem of comparing two data-driven approaches (the artificial neural network approach and the fuzzy logic approach) in terms of accuracy, reliability

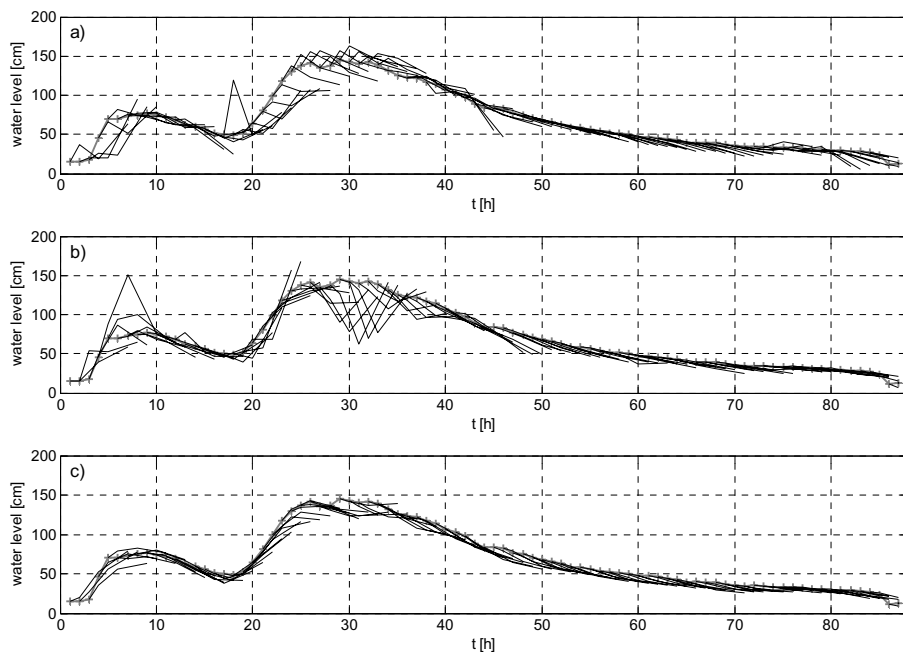


Fig. 13. DRI input data set – testing phase. Ensemble of the water levels forecasted at 1, 3 and 6 h ahead starting from each hour of the flood event considered. (a) FL-M model, (b) FL-TS model, (c) ANN model.

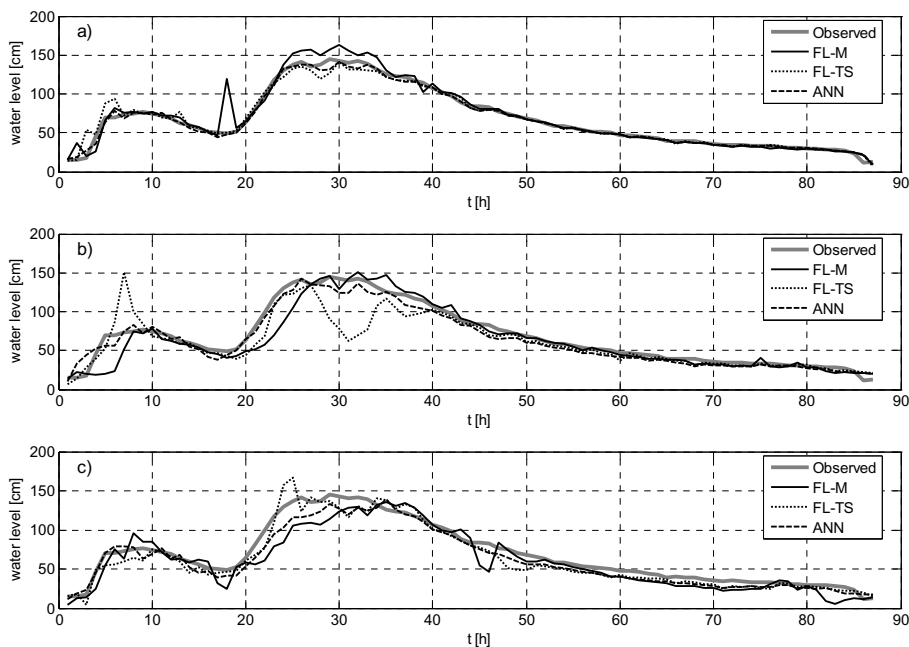


Fig. 14. DRI input data set – testing phase. Water levels forecasted (a) 1 h, (b) 3 h and (c) 6 h ahead by the three models.

and capability of dealing with different levels of information, within the framework of a water level forecasting system.

As regards the accuracy, all the models provide good accuracy for short time horizon forecasts which however decreases when longer time horizons are considered and this is particularly true for the rising phase of the flood wave where

a systematic underestimation is observed when the models are trained with the ARI input data set. A lead time up to 6 h ahead can be however considered acceptable for both FL and ANN approaches. This temporal limit is coherent with that detected by other authors using similar data-driven models applied to basins with similar extension to that considered in

this study (e.g. Campolo et al., 1999, 2003; See and Openshaw, 1999; Solomatine and Dulal, 2003), and this limit is certainly due to the fact that no information or forecast of rainfall is considered available within the time spell ahead with respect to the time instant when the forecast is performed.

The analysis of the model accuracy, when the ARI input data set is used, shows that, overall, the FL-M and FL-TS models perform slightly better than the ANN model in terms of *RMSEs* and R^2 . However, as previously recalled, all the models underestimate the water levels in the rising part of the flood waves and this can be related to the little information on rainfall dynamic available in the ARI input data set.

As regards the reliability, it has been shown that the models considered present a different level of reliability. Both the FL models produce some unexpected failure. In particular, they are not able to execute a forecast in some testing cases since the input vectors do not satisfy any “IF” condition of the trained/calibrated rules. This lack of response can suggest that the FL approach is certainly appropriate when the enumeration of all the possibilities can be done a priori, as in the case of some mechanical or electronic tool, but it may not be totally reliable when dealing with natural phenomenon where the number of possible combinations may be extremely large and where input combinations not considered in the training phase can produce no results. The ANN model does not present this problem since, given its own architecture, for each input vector an output vector is always obtained through the transfer functions of the hidden and output layers.

As regards the capability to deal with different level of information, the use of the DRI input data set highlights further differences between the FL and ANN approaches. The ANN accuracy increases significantly and, at the same time, the tendency to underestimate the future water levels decreases significantly with respect to that observed in the case of the ARI input data set. This behaviour indicates that the greater detail in the rainfall pattern is useful to forecast the water levels more accurately, especially in the rising part of the flood events.

As regards the FL-M and FL-TS models, the forecasting accuracy in the testing phase does not increase, or becomes even worse when the DRI input data set are used. This indicates that the FL approach, independently of its formulation, has a limited capability of dealing with too detailed information, and this result is in line with other hydrological studies based on fuzzy rules system which, in fact, are generally characterized by a low number (from 2 to 5) of input variables (see, for example, Abebe et al., 2000; Hundecha et al., 2001; Xiong et al., 2001; Han et al., 2002; Chang et al., 2005). A similar limit exists also for the number of rules that can be implemented within the framework of a FL model. It has in fact been shown that the accuracy of the model initially increases with the number of rules, but beyond a certain number, the accuracy of the model starts

to decrease again. All these latter considerations show that, given their very structure, the FL approaches perform better when the physical phenomena considered are synthesised by both a limited number of variables and IF-THEN logic statements, while the ANN approach increases its performance when more detailed information is used.

Overall, both approaches may be used within the framework of a real time forecasting system, though with different levels of reliability. Both of them however show performances which remain satisfactory up to 6 h ahead for the basin considered, which indeed is too short for operational flood forecasting. While it is worth stressing that the aim of this paper was the comparison between the two approaches within the terms recalled at the beginning of this section, this latter observation highlights the need to feed these models with further information, with respect to the time of forecasting, relevant to the future rainfall amount and its spatial and time distribution. This further information should arrive mainly from radar data, possibly elaborated by another model dedicated to the rainfall forecast, which would allow the time spell of prediction to be extended to useful operational time intervals.

Finally, we can observe that the models here considered, for their very nature, are “blind” with respect to specific physical information: they only deal with data (e.g. rainfall and levels) which indirectly contain the “integral” behaviour of the system to be modelled. As a consequence, since the comparison presented was performed with reference to a basin of a humid region, it is expected that similar results (in terms of comparison between the two approaches) can be obtained when other basins located in similar regions are considered. On the contrary, if basins located in dry regions were considered, different inputs and structures may be selected and identified, so the relative performances of the two models may change. Further analyses are currently being developed with reference to catchments in dry regions to analyse these latter aspects.

Edited by: E. Todini

References

- Abebe, A. J., Solomatine, D. P., and Venneker, R. G. W.: Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events, *Hydrol. Sci. J.*, 45(3), 425–436, 2000.
- Abrahart, R. J. and See, L.: Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, *Hydrol. Earth Syst. Sci.*, 6, 655–670, 2002, **SRef-ID: 1607-7938/hess/2002-6-655**.
- Anctil, F. and Rat, A.: Evaluation of Neural Network Streamflow forecasting on 47 Watershed, *J. Hydrol. Eng.*, 10(1), 85–88, 2005.

- ASCE Task Committee on the application of ANN in Hydrology: Artificial Neural Networks in Hydrology. I: Preliminary Concepts, *J. Hydrol. Eng.*, 5(2), 115–123, 2000.
- Bardossy, A. and Duckstein, L.: Fuzzy rule-based modelling with applications to geophysical, Biological and Engineering Systems, CRC Press, Boca Raton, Florida, 1995.
- Bardossy, A. and Samaniego, L.: Fuzzy Rule-Based Classification of Remotely Sensed Imagery, *IEEE Transactions on Geosciences and Remote Sensing*, 40, 362–374, 2002.
- Beven, K. J. and Kirby, M. J.: A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 1–3, 1979.
- Beven, K. J., Kirby, M. J., Schofield, N., and Tagg, A. F.: Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments, *J. Hydrol.*, 69, 119–143, 1984.
- Burnash, R. J. C., Ferrel, R. L., and Mc Guire, R. A.: A general Streamflow Simulation System – Conceptual Modelling for Digital Computers, Report by the Joint Federal State River Forecast Center, Sacramento, 1973.
- Campolo, M., Andreussi, P., and Soldati, A.: River flood forecasting with neural network model, *Water Resour. Res.*, 35(4), 1191–1197, 1999.
- Campolo, M., Andreussi, P., and Soldati, A.: Artificial neural network approach to flood forecasting in the river Arno, *Hydrol. Sci. J.*, 48(3), 381–398, 2003.
- Chang, L. C., Chang, F. J., and Tsai, Y. H.: Fuzzy exemplar-based inference system for flood forecasting, *Water Resour. Res.*, 41, W02005, doi:10.1029/2004WR003037, 2005.
- Chiang, Y. M., Chang, L. C., and Chang, F. J.: Comparison of static-feedforward and dynamic-feedback neural network for rainfall-runoff modelling, *J. Hydrol.*, 290, 297–311, 2004.
- Crawford, N. H. and Linsey, R. K.: Digital simulation in hydrology, Stanford Watershed Model IV. Tech. Rep. No. 39, Dep. Civ. Eng., Stanford University, 1966.
- Dawson, C. W., Harpham, C., Wilby, R. L., and Chen, Y.: Evaluation of artificial neural network technique in the River Yangtze, China, *Hydrol. Earth Syst. Sci.*, 6, 619–626, 2002, **SRef-ID: 1607-7938/hess/2002-6-619**.
- Demuth, H. B. and Beale, M.: *Neural Network Toolbox for use with Matlab*, The Math Works, Inc., Natick, 2000.
- Ferraresi, M., Pacciani, M., and Todini, E.: Sull'applicazione di alcuni schemi di previsione di piena fluviale in tempo reale, *Atti del XXII convegno di Idraulica e Costruzioni Idrauliche*, 3, 307–320, 1990.
- Franchini, M. and Pacciani, M.: Comparative analysis of several conceptual rainfall-runoff models, *J. Hydrol.*, 122, 161–219, 1991.
- Goswami, M., O'Connor, K. M., Bhattarai, K. P., and Shamseldin, A. Y.: Assessing the performance of eight real-time updating models and procedures for the Brosna River, *Hydrol. Earth Syst. Sci.*, 9, 394–411, 2005, **SRef-ID: 1607-7938/hess/2005-9-394**.
- Hagan, M. T., Demuth, H. B., and Beale, M.: *Neural Network Design*, PWS Publishing Company, Boston, 1996.
- Hagan, M. T. and Menhaj, M.: Training feedforward networks with the Marquardt algorithm, *IEEE Transaction on Neural Networks*, 5(6), 989–993, 1994.
- Han, D., Cluckie, I. D., Karbassioun, D., Lawry, J., and Krauskopf, B.: River Flow Modelling using Fuzzy Decision Trees, *Water Resour. Manag.*, 16, 431–445, 2002.
- Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 359–366, 1989.
- Hsu, K. L., Gupta, H. V., and Sorooshian, S.: Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, 31(10), 2517–2530, 1995.
- Hundecha, Y., Bardossy, A., and Theisen, H. W.: Development of a fuzzy logic-based rainfall-runoff model, *Hydrol. Sci. J.*, 46(3), 363–376, 2001.
- Imrie, C. E., Durucan, S., and Korre, A.: River flow prediction using artificial neural networks: generalizations beyond the calibration range, *J. Hydrol.*, 233, 138–153, 2000.
- Kim, G. and Barros, A. P.: Quantitative flood forecasting using multi-sensor data and neural networks, *J. Hydrol.*, 246, 45–62, 2001.
- Kirkpatrick, S., Gelat Jr., C. D., and Vecchi, M. P.: Optimization by simulated annealing, *Science*, 220(4598), 671–680, 1983.
- Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, 35(9), 2739–2750, 1999.
- Krzysztofowicz, R. and Kelly, K. S.: Hydrologic uncertainty processor for probabilistic river stage Forecasting, *Water Resour. Res.*, 36(11), 3265–3277, 2000.
- Krzysztofowicz, R. and Herr, H. D.: Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model, *J. Hydrol.*, 249, 46–68, 2001.
- Krzysztofowicz, R.: Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model, *J. Hydrol.*, 249, 69–85, 2001.
- Larsen, P. M.: Industrial applications of Fuzzy Logic Control, *Int. J. Man Mach. Stud.*, 12(1), 3–10, 1980.
- Liu, Z. and Todini, E.: Towards a comprehensive physically based rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 6, 859–881, 2002, **SRef-ID: 1607-7938/hess/2002-6-859**.
- Maier, H. R. and Dandy, G. C.: Neural Networks for prediction and forecasting of water resources variables: a review of modelling issue and applications, *Environmental Modelling & Software*, 15, 101–124, 2000.
- Mamdani, E. H.: Application of Fuzzy Algorithm for Control of Simple Dynamic Plant, *Proc. IEEE*, 121, 1585–1888, 1974.
- Moller, M. F.: A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, 6, 525–533, 1993.
- Montanari, A., Rosso, R., and Taqqu, M. S.: A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, *Water Resour. Res.*, 36(5), 1249–1259, 2000.
- Moradkhany, H., Hsu, K. L., Gupta, H. V., and Sorooshian, S.: Improved streamflow forecasting using self-organizing radial-basis function artificial neural networks, *J. Hydrol.*, 295, 246–262, 2004.
- Özelkan, E. C. and Duckstein, L.: Fuzzy conceptual rainfall-runoff models, *J. Hydrol.*, 253, 41–68, 2001.
- Pilgrim, D. H.: Travel times and non linearity of flood runoff from tracer measurements on a small watershed, *Water Resour. Res.*, 12(3), 487–496, 1976.
- Rosenblatt, F.: The Perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65, 386–408, 1958.
- See, L. and Openshaw, S.: Applying soft computing approaches to river level forecasting, *Hydrol. Sci. J.*, 44(5), 763–778, 1999.

- See, L. and Openshaw, S.: A hybrid multi-model approach to river level forecasting, *Hydrol. Sci. J.*, 45(4), 523–536, 2000.
- Shamseldin, A. Y.: Application of a neural network technique to rainfall-runoff modelling, *J. Hydrol.*, 199, 272–294, 1997.
- Singh, V. P.: Non linear instantaneous unit hydrograph theory, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, 90(HY2), 313–347, 1964.
- Singh, V. P. and Woolhiser, D. A.: Mathematical Modeling of Watershed Hydrology, *J. Hydrol. Eng.*, 7(4), 270–292, 2002.
- Sivapalan, M., Beven, K. J., and Wood, K. F.: On Hydrological similarity 2. A scaled model of storm runoff production, *Water Resour. Res.*, 23, 2266–2278, 1987.
- Smith, M.: *Neural Network for statistical modelling*, Van Nostrand Reinhold, New York, 1993.
- Solomatine, D. P. and Dulal, K. N.: Model trees as an alternative to neural networks in rainfall runoff modelling, *Hydrol. Sci. J.*, 48(3), 399–411, 2003.
- Takagi, T. and Sugeno, M.: Fuzzy identification of systems and its application to modelling and control, *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1), 116–132, 1985.
- Thimuralaiah, K. and Deo, M. C.: Hydrological forecasting Using Neural Networks, *J. Hydrol. Eng.*, 5(2), 180–189, 2000.
- Toth, E., Brath, A., and Montanari, A.: Comparison of short-term rainfall prediction models for real-time flood forecasting, *J. Hydrol.*, 239, 132–147, 2000.
- Todini, E.: Rainfall-Runoff modeling – Past, present and future, *J. Hydrol.*, 100, 341–352, 1988.
- Todini, E.: The ARNO rainfall-runoff model, *J. Hydrol.*, 175, 339–382, 1996.
- Xiong, L., Shamseldin, A. Y., and O'Connor, K. M.: A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system, *J. Hydrol.*, 245, 196–217, 2001.
- Young, P. C.: Data-based mechanistic modelling and validation of rainfall-flow processes, in: *Model Validation: Perspectives in Hydrological Science*, edited by: Anderson, M. G. and Bates, P. D., Wiley, Chichester, pp. 117–161, 2001.
- Young, P. C.: *Advances in real-time flood forecasting*, Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, 2002.
- Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning, *Memorandum ERL-M 411*, Berkeley, 1973.
- Zealand, C. M., Burn, D. H., and Simonovic, S. P.: Short term streamflow forecasting using artificial neural networks, *J. Hydrol.*, 214, 32–48, 1999.