

An adaptive predictive approach for river level forecasting

José V. Aguilar, Pedro Langarita, Lorenzo Linares, Manuel Gómez
and José Rodellar

ABSTRACT

Efficient flood management requires accurate real-time forecasts to allow early warnings, real-time control of hydraulics structures, or other actions. Commercially available computing tools typically use hydraulic models derived from the numerical approximation of Saint-Venant equations. These tools need powerful computers, accurate knowledge of the riverbed topography, and skilled operators with a not insignificant hydraulic background. This paper presents an alternative approach in which the river basin is modeled as a network of cascade interconnected input–output systems. Each system is modeled by an adaptive predictive expert model, which provides real-time fast and accurate forecasts over a moving prediction horizon. The approach is evaluated using real data from the Ebro river basin in Spain. The main concluded advantages of the new approach are: (1) the formulation is simple with low computational burden; (2) it does not require topographic information on the river waterbeds; (3) the forecast may be performed autonomously.

Key words | adaptive predictive models, data-driven modeling, floods, river level forecasting

José V. Aguilar

Pedro Langarita

Automation and Control Group,
Confederación Hidrográfica del Ebro,
50071 Zaragoza,
Spain

Lorenzo Linares

Automation and Control Group, SICE,
50009 Zaragoza,
Spain

Manuel Gómez

FLUMEN Group, Department of Hydraulics,
Marine and Environmental Engineering,
Universitat Politècnica de Catalunya-
BARCELONATECH, 08034 Barcelona,
Spain

José Rodellar (corresponding author)

Control, Dynamics and Applications Group,
Department of Applied Mathematics III,
Universitat Politècnica de Catalunya –
BARCELONATECH, 08034 Barcelona,
Spain
E-mail: jose.rodellar@upc.edu

INTRODUCTION

The problem of water in Spain is mainly caused by the unequal precipitation distribution, both in space and time, which reduces its availability. As a result, severe drought and flood episodes have occurred over the years, which has motivated the development of major hydraulic infrastructures whose safety it is vital to ensure during flood episodes. In view of this situation, and after tragic floods in the north and east of Spain at the beginning of the 1980s, the need arose to deploy Automatic Hydrological Information Systems (AHIS) in the main river basins. With the real-time hydrologic–hydraulic data provided by AHIS and other complementary information (weather forecasts, snow storage, etc.), a decision support system (DSS) is in operation for flood forecasting (Romeo *et al.* 2004). This system integrates several mathematical models and computer tools to estimate the time evolution of levels and flows along the different watershed rivers. One of these models is in charge of describing and calculating the discharge evolution along the rivers in the basin.

The most popular discharge routing models, currently in use and commercially distributed, are based on the laws of physics (HEC-RAS, GISPLANA, QUABIS, MIKE 11). Their physical background makes these models very well suited to describe the water flow process with accuracy. When used for operational forecast, their main drawbacks are twofold: (1) the prediction accuracy may significantly depend on the estimation of the parameters, the geometry of the waterbeds and the adopted numerical scheme, so that the user has to be skilled enough to ensure a meaningful implementation; (2) they may require high computation times, so that they are normally used for medium- to long-term forecasts with daily updating.

The use of black-box (data-driven) models has been proposed as an alternative way to overcome some of the above drawbacks. These models identify a direct relation between inputs and outputs with reduced insight on the physical processes. Although the ability of these models to describe the physical phenomena is reduced, the interest is to exploit

their simplicity. Their main role is in real-time applications with short operating times, such as short-term forecasting or automatic control. Artificial neural networks and soft computing have been among the most commonly used in recent years to build data-driven models (See & Opeshaw 1999; Young 2002; Moradkany *et al.* 2004; Bruen & Yang 2005; Alvisi *et al.* 2006; Pulido-Calvo & Portela 2007; Tsanis *et al.* 2008; Fernández *et al.* 2010; Islam 2010; Shamseldin 2010; Seckin 2011).

Another approach, motivated by the aim of deriving simplified forecasting models, is the so-called data-based mechanistic modeling (Romanowicz *et al.* 2006, 2008). In this approach, input–output discrete-time linear models are identified to describe river reaches. Upstream water levels are the inputs for these models, which are estimated by means of a nonlinear rainfall–water level relation. At the end, the approach gives real-time updated water level forecasting at specified desired locations over a specific future time horizon.

The aim of this work is to propose a new, simple but effective forecasting method exhibiting the following features:

- It is based on the formulation of a novel prediction scheme with simple input/output models with time varying parameters and time delays, which incorporate real-time parameter adaptation and fuzzy logic rules.
- It is designed to operate autonomously to supply online hourly forecasts over a long-range prediction horizon.
- It is formulated without the need to know the geometry of the waterbed.

In this paper, a large river basin is modeled as a set of cascade interconnected discrete-time linear adaptive models representing the different reaches. All the inputs and outputs are water levels, which are measured by a network of gauging stations with a short (hourly) sampling time. The model parameters are updated at each sampling time by an adaptation algorithm (Martín-Sánchez & Rodelar 1996) and used to perform forecasts of the downstream level of each reach. These forecasts are performed in real time at each sampling instant over a future prediction horizon, which usually has the same length as the observed time delay in the reach. In order to improve the predictions, the updated model parameters are weighted by coefficients

supplied by a fuzzy set with if-then rules. This fuzzy set is built using available data of recorded past floods.

The paper presents the methodology and validates its effectiveness by using real data of a major flood recorded by the AHIS of the Ebro river in Spain. The results are compared with those obtained by means of a MIKE 11 based DSS system.

This paper is an extended version of a previous paper first presented at the 2011 IEEE International Conference on Networking, Sensing and Control in Delft, The Netherlands (Aguilar *et al.* 2011).

PREDICTION ARCHITECTURE

Adaptive models

A river basin can be modeled as a set of cascade interconnected systems, where each system output is the input to the subsequent system. Each system receives additional inputs from a number of gauging stations located along the main river and their tributary rivers and supplies the output through a gauging station located at the downstream end of the main river reach. Water levels are the physical variables measured at each station.

It is clear that the behavior of these systems is complex and physically based nonlinear models should be adopted if the purpose is to perform accurate simulations. But, under certain operation conditions and for online purposes (as predictions or automatic control), the dynamics can be locally described by input–output discrete-time linear equations with appropriate parameters. Different operation conditions for the same physical process may require different local linear representations. However, a single model can be adopted for the overall set of conditions with a linear structure but with time varying parameters.

For a generic system, the following input–output model is considered:

$$y(k) = \sum_{i=1}^n a_i(k)y(k-i) + \sum_{i=1}^{m_1} b_{1i}(k)u_1(k-r^i(k)-i) + \dots + \sum_{i=1}^{m_p} b_{pi}(k)u_p(k-r_p(k)-i) + \Delta(k), \quad (1)$$

where u_i and y denote the inputs and the output, respectively, at each sampling time k ; $r_i(k)$ denotes the time delays between the output and the corresponding inputs and $a(k)$ and $b(k)$ are parameters. In general, delays and parameters can be time dependent, trying to capture the system nonlinearities in a reasonably practical way. Variable $\Delta(k)$ may represent some unknown disturbances.

Equation (1) is usually written in the following compact form:

$$y(k) = \theta(k)^T X(k) + \Delta(k) \quad (2)$$

where $\theta(k)^T = [a_1(k), \dots, b_{11}(k), \dots, b_{p1}(k), \dots]$ is the parameter vector and $X(k)^T = [y(k-1), \dots, u_p(k-r_p(k)-1) \dots]$ is the regression vector including all the inputs and outputs previous at time instant k . In this work, all these variables are water levels, which are measured at the gauging stations and transmitted to the control center.

Time delays

It is known that the time required for a flood to propagate along the river is not constant. Indeed, it varies depending on several reasons, particularly on the flow magnitude. Accurate estimation of time delays is really difficult and requires hydraulic and topographical data of the river basin. Since the objective of this work is not to rely on such types of information, the time delays $r_i(k)$ used in the models (1) are calculated as follows:

$$r_i(k) = \begin{cases} r_{i1}, & \text{if } y(k) \leq \gamma_{i1} \\ f_i(y(k)), & \text{if } \gamma_{i1} \leq y(k) \leq \gamma_{i2} \\ r_{i2}, & \text{if } y(k) \geq \gamma_{i2} \end{cases} \quad (3)$$

where r_{i1} , r_{i2} , γ_{i1} , and γ_{i2} are constant parameters and f_i are linear functions, which are selected through the analysis of historical data of past flood events.

Adaptation mechanism

In order to perform the parameter adaptation, the first step is to obtain an estimation of the real measured output $y(k)$. An appropriate mathematical model for the estimation

may have the structure of the model in Equation (1), but without the unknown disturbance variable. The designer has to choose the model orders for the input and output variables. In this work, a first order model has been chosen for all variables. Further studies could consider second or higher orders to implement the proposed approach.

Therefore, the following estimation model is considered:

$$\hat{y}(k|k-1) = \hat{\theta}(k-1)^T X(k), \quad (4)$$

where $\hat{\theta}(k-1)^T = [\hat{a}(k-1), \hat{b}_1(k-1), \dots, \hat{b}_p(k-1)]$ is the vector containing the parameters estimated at time instant $k-1$. This is referred to as *a priori* estimation, since $\hat{y}(k|k-1)$ denotes the estimation of the current output at instant k but it is performed using the information available at previous instant $k-1$, which is included in the regression vector $X(k)^T = [y(k-1), \dots, u_p(k-r_p(k)-1) \dots]$.

The *a priori* estimation error is defined as follows:

$$e(k|k-1) = y(k) - \hat{y}(k|k-1) \quad (5)$$

Using this error and the input-output vector, the parameters are updated by the following adaptation law:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{e(k|k-1)}{1 + X^T(k)BX(k)} BX(k) \quad (6)$$

where B is a positive definite weighting matrix. Once the parameters are updated, the following *a posteriori* estimation error is defined and may be used to validate the effectiveness of the adaptation mechanism (4)–(6):

$$e(k|k) = y(k) - \hat{y}(k|k) = y(k) - \hat{\theta}(k)^T X(k) \quad (7)$$

Such a validation will be presented for the Ebro river basin under study.

This adaptation and estimation scheme is described in detail in [Martín-Sánchez & Rodellar \(1996\)](#), including a mathematical study on convergence and stability.

It is worth mentioning that the above estimation approach is not based on an optimization perspective, where parameters are updated to minimize some criterion involving the estimation error. This is a quite natural objective in a modeling context, where the objective is the

identification of the parameters to achieve the best possible model in terms of capturing the system dynamics for simulation purposes. In this case, *a priori* correlation or mutual information analysis among variables may be relevant to ensure a statistically consistent estimation (Bowden et al. 2005a, 2005b).

The methodology used in this work is based on stability, and the emphasis is on ensuring that the estimated parameters converge to some values, while keeping the estimation error bounded in the presence of unknown disturbances. This convergence is efficient in the sense that iteratively reduces the square of the estimation error in the gradient direction, which means in the direction of the best actual system output tracking. Within this perspective, the problem of correlation of variables is not relevant since the objective is not to identify the best possible model. This scheme allows working satisfactorily with reduced order models. The objective is to have a simplified model able to track the system output and to have adaptive capability to system changes for control or prediction purposes.

Recent applications of this methodology to automatic control of irrigation canals have been presented in Aguilar et al. (2009, 2012). The motivation of the present study is to explore their potential for online prediction.

Prediction

The final objective of this work is to develop and validate a prediction scheme to forecast the output level of each system. To be more precise, at each real sampling time k , the aim is to perform an online prediction over a future time interval $[k, k + N]$, where N is a prescribed time horizon. This prediction will use information on inputs and outputs that has been captured prior to the prediction instant k , along with the adaptive models described above.

As previously discussed, linear representation of non-linear processes can be reasonably acceptable if the parameters are updated as the operation point varies. Once the parameters $\hat{a}(k)$, $\hat{b}_1(k)$, \dots , $\hat{b}_p(k)$ are adapted at time k by a mechanism such as the one presented in the previous section, the models are well suited to estimate the outputs for a short future time. But if we consider a flood

prediction over a longer time interval $[k, k + N]$, it could be expected that the operation conditions will change, particularly as the time horizon and the flood magnitude increase. Consequently, it does not seem advisable to model the dynamics of the reaches over the whole prediction interval $[k, k + N]$, just by keeping constant the updated parameters $\hat{a}(k)$, $\hat{b}_1(k)$, \dots , $\hat{b}_p(k)$.

The way proposed to deal with this issue is to assign time-varying weights to the model parameters aiming to accommodate varying conditions along the prediction horizon. Altogether, the following prediction scheme is proposed in this work based on weighted first order adaptive models:

$$\begin{aligned} \hat{y}(k+j|k) = & \hat{a}(k)A(j)\hat{y}(k+j-1|k) \\ & + \hat{b}_1(k)B_1(j)\hat{u}_1(k-r_1(k)+j-1|k) \\ & + \dots + \hat{b}_p(k)B_p(j)\hat{u}_p(k-r_p(k)+j-1|k). \end{aligned} \quad (8)$$

In this model the notation $\hat{y}(k+j|k)$ represents the output predicted for the future instant $k+j$ ($j = 1, \dots, N$) made at the present time k . The model is redefined at each time k starting from the available measured inputs and outputs, i.e.,

$$\hat{y}(k+1-1|k) = y(k) \quad \text{for } j = 1;$$

$$\begin{aligned} \hat{u}_n(k-r_n(k)+j-1|k) = & u_n(k-r_n(k)+j-1) \\ \text{for } j = 1, \dots, r_n(k)+1; \end{aligned} \quad (9)$$

$$\begin{aligned} \hat{u}_n(k-r_n(k)+j-1|k) = & \hat{y}_{nc}(k-r_n(k)+j-1|k) \\ \text{for } j = r_n(k)+2, \dots, N; \end{aligned}$$

$$n = 1, \dots, p$$

In this model, $A(j)$, $B_1(j)$, \dots , $B_p(j)$ are the weighting parameters, which are calculated at each future prediction time $k+j$. A fuzzy logic scheme has been designed for this purpose and is described in the next subsection.

Notice that, in the prediction scheme stated in Equations (8) and (9), the information can be divided into two groups: (1) the inputs and outputs that are physically available at the present time k where the prediction is carried out; and (2) the inputs that correspond to future instants, which have not been measured. Clearly all the outputs at k and previous time instants are known. Thus, the

predicted output is redefined with the current value $y(k)$ as stated in the first equation in Equation (9). The values of the inputs $u_n(n = 1, \dots, p)$ are known for the prediction instants $j = 1, \dots, r_n(k) + 1$, since they correspond to instants prior to k . Thus, these values may be used in the prediction model (8) as stated in the second equation in Equation (9). However, when extending the prediction to instants beyond the time delay $j = r_n(k) + 2, \dots, N$, we have inputs that correspond to future instants and the corresponding values are not known at the current prediction time k . Most of the systems in the network are connected in such a way that the inputs for a system are outputs resulting from previous systems. This is clearly observed in the section devoted to the case study. Since the predictions for all the systems are carried out in a sequential cascade order, the inputs $u_n(n = 1, \dots, p)$ for a specific system for the future instants $j = r_n(k) + 2, \dots, N$ are not exactly known, but their values have already been predicted by the prediction scheme. Thus, their predicted values can be used in Equation (8) as inputs. This strategy is summarized in the third equation in Equation (9), where $\hat{y}_{nc}(k - r_n(k) + j - 1|k)$ denotes these predicted values. In the case that the system under prediction has some input that is not the output of a connected system (like those at the head of the network), the above future values are unavailable either from measurements or from the prediction scheme. The prediction scheme allows the inclusion of these future values from other alternative sources. In this study, for those specific cases, estimations from a rainfall-runoff model have been used, which is in operation in the management center of the Ebro river basin.

Fuzzy logic scheme

Fuzzy logic is a mathematical scheme with capacity to operate with ambiguous concepts and information, which are typical in the qualitative human reasoning. It allows drawing quantitative conclusions from a collection of observations and a set of qualitative rules.

The structure of a fuzzy system has four main components:

Fuzzyfication interface: it performs scaling and conversion of the input variable data into appropriate linguistic values.

Knowledge base: it is composed by the database for the membership functions of the input/output variables and a set of rules that characterizes the policy and objectives of the human expert knowledge.

Inference engine: it combines the rules and gives the fuzzy output variable.

Defuzzyfication interface: it transforms the fuzzy output variable into a numerical variable.

With an intuitive view, fuzzy logic may be interpreted as a black box, which gives an output variable from several input variables. As an example, without loss of generality, consider the following single input-single output first order predictive model:

$$\hat{y}(k + j|k) = \hat{a}(k)A(j)\hat{y}(k + j - 1|k) + \hat{b}(k)B(j)\hat{u}(k - r(k) + j - 1|k). \tag{10}$$

Figure 1 shows a block diagram of the fuzzy structure to calculate the parameter $B(j)$ in Equation (10). It has two input variables: (1) Input_1, which represents the increment in the predicted output system at prediction time instant j ; and (2) Input_2, which represents the increment in the system input also at time j . The output given by the fuzzy block is the value of parameter $B(j)$.

The incremental variables are defined as follows:

$$Input_1(j) = \hat{y}(k + j - 1|k) - \hat{y}(k + j - 2|k)$$

$$\begin{aligned} \hat{y}(k + 1 - 1|k) &= y(k) && \text{for } j = 1 \\ \hat{y}(k + 1 - 2|k) &= y(k - 1) && \text{for } j = 1 \\ \hat{y}(k + 2 - 2|k) &= y(k) && \text{for } j = 2 \end{aligned} \tag{11}$$

$$Input_2(j) = \hat{u}(k - r(k) + j - 1) - \hat{u}(k - r(k) + j - 2)$$

The use of incremental variables in the fuzzy blocks has clear physical meaning: the objective is that the calculation

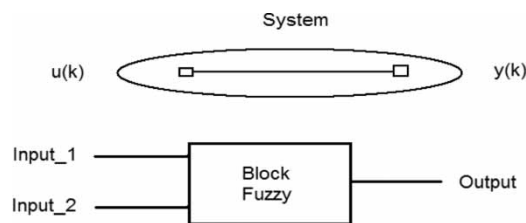


Figure 1 | Fuzzy block to compute parameter $B(j)$.

of the parameter $B(j)$ takes into account whether the flood at prediction time $k + j$ is increasing or decreasing, both at the system input and output, and uses the corresponding time variations.

As a practical example, let us consider the fuzzy structure for a particular system. It is part of the case study fully described in the next section, specifically System 9 in Table 1, which relates two gauge stations, named A002 and A011, respectively. The elements of the fuzzy structure are summarized as follows and illustrated by Figures 2–4.

Table 1 | Information on systems and models

System	Variables	Station	Initial model parameters	Input-output distances (km)	Initial delays (h)
1	Input_1	A161	0.06	52.8	7
	Input_2	A188	0.08	17.3	3
	Output	A001	0.8		
2	Input_1	A001	0.2	123.9	16
	Input_2	A165	0.034	124.8	17
	Input_3	A074	0.104	119.7	16
	Input_4	A281	0.034	90.4	8
3	Output	A120	0.8		
	Input_1	A071	0.12	48.0	5
4	Output	A003	0.8		
	Input_1	A084	0.12	65.6	7
5	Input_2	A069	0.2	69.7	7
	Output	A004	0.8		
6	Input_1	A264	0.04	23.9	4
	Input_2	A064	0.24	25.0	4
	Output	A065	0.8		
7	Input_1	A065	0.072	75.1	10
	Input_2	A101	0.1	73.6	10
	Output	A005	0.8		
8	Input_1	A290	0.16	40.0	6
	Output	A260	0.8		
9	Input_1	A120	0.15	74.9	14
	Input_2	A003	0.03	44.9	8
	Input_3	A004	0.08	32.9	7
	Input_4	A005	0.1	33.8	7
	Output	A002	0.8		
9	Input_1	A002	0.13	137.4	20
	Input_2	A260	0.02	79.4	12
	Output	A011	0.8		

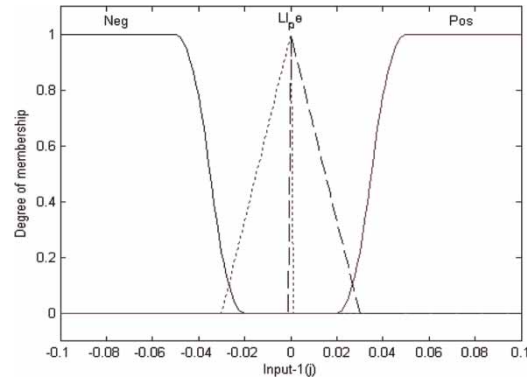


Figure 2 | Fuzzy sets for $Input_1(j)$.

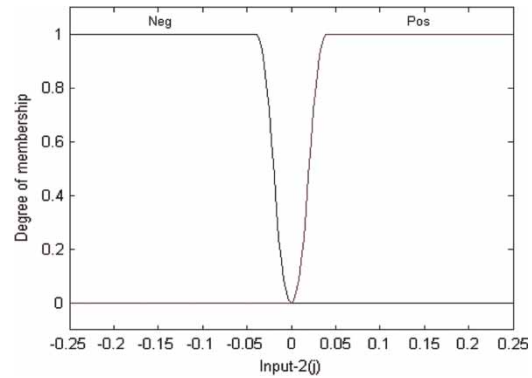


Figure 3 | Fuzzy sets for $Input_2(j)$.

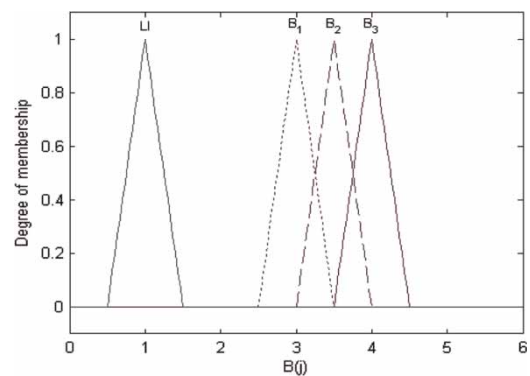


Figure 4 | Fuzzy sets for output $B(j)$.

The two input variables are $Input_1(j)$ for the increments in A011 and $Input_2(j)$ for the increments in A002. Four fuzzy sets are associated with the first input and two sets for the second one. They are shown in Figures 2 and 3, respectively.

Four fuzzy sets are associated with the parameter $B(j)$ (output), which are represented in Figure 4.

The knowledge base is designed with eight fuzzy rules:

1. If $Input_1(j)$ Neg and $Input_2(j)$ Neg then $B(j)$ L1
2. If $Input_1(j)$ Neg and $Input_2(j)$ Pos then $B(j)$ B2
3. If $Input_1(j)$ Ll_ne and $Input_2(j)$ Neg then $B(j)$ B2
4. If $Input_1(j)$ Ll_ne and $Input_2(j)$ Pos then $B(j)$ B3
5. If $Input_1(j)$ Ll_po and $Input_2(j)$ Neg then $B(j)$ L1
6. If $Input_1(j)$ Ll_po and $Input_2(j)$ Pos then $B(j)$ B1
7. If $Input_1(j)$ Pos and $Input_2(j)$ Neg then $B(j)$ L1
8. If $Input_1(j)$ Pos and $Input_2(j)$ Pos then $B(j)$ L1

An inference mechanism is used for systems of Mandani type with the 'and' connectivity operator (product method) and maximum rule aggregation. The centroid method is used for defuzzification.

The selection of the intervals for each fuzzy set is done when the prediction scheme is validated with the historical data of past floods.

Fuzzy logic has been used in a number of forecasting approaches, mainly combined with artificial neural networks (Alvisi *et al.* 2006; Fernández *et al.* 2010; Seckin 2011). The way fuzzy logic enters into the approach presented in this paper is different. The idea here is to help to modify the model parameters to compensate for possible

unknown variations in the system dynamics during the future prediction intervals.

CASE STUDY: LEVEL PREDICTION IN THE EBRO RIVER

The adaptive predictive expert (APE) prediction methodology proposed in this paper is validated using the real data of a flood that occurred in the Ebro river basin in February 2009. The surface of the basin is 32,700 km² and is modeled decomposed into nine dynamic systems with a total of 21 gauging stations. In these stations, the levels of the basin rivers are measured in real time and the data are transmitted to a control center each 15 minutes. In this study, the prediction scheme is used with a time period of 1 h to forecast all the station levels that are selected as the outputs of the systems. Figure 5 shows the modeling of the whole basin indicating the 21 gauging stations. Table 1 gives the most relevant information to characterize all the systems, including the input–output stations, the initial parameters of the estimation models (4), and the corresponding initial time delays.

Two classes of systems coexist in the basin: those with fast response (Systems 1, 3, 4, 5, 6, 7) and those with slow

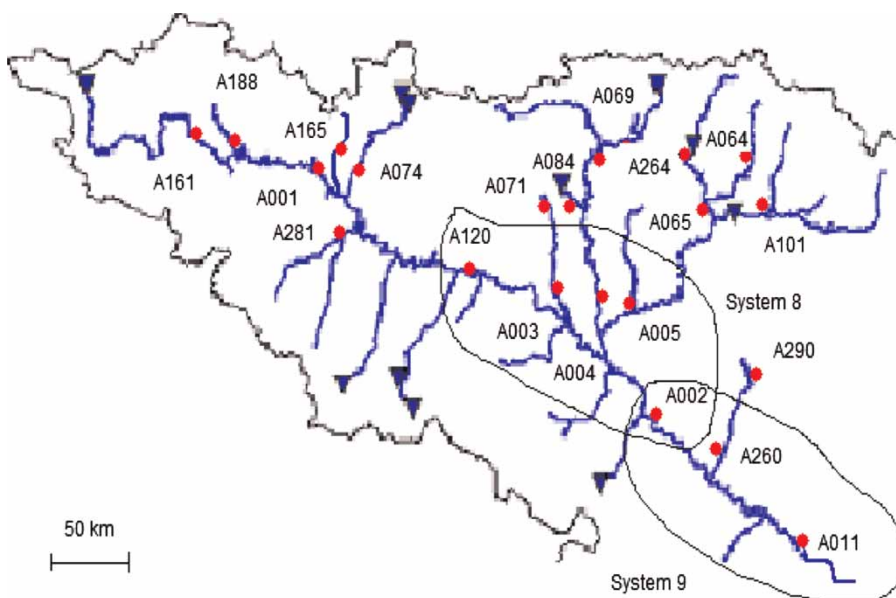


Figure 5 | Map with the river gauge network.

response (2, 8, 9). Fast systems are characterized by short lengths and big slopes, which make the time delays between upstream inputs and downstream outputs relatively small and with small variations. On the other hand, slow systems have significant dimensions and relatively small slopes. This corresponds to normal small time delays but, when the flood magnitude increases, the riverbeds become significantly wider and the time delays become longer. To have an idea of the flood magnitudes, the maximum flows in the event of February 2009 ranged between 150 and 2,600 m³/s. A detailed description of the river basin is freely available through the official web site of the Confederación Hidrográfica del Ebro: www.sai-hebro.com.

For the sake of space limitations, this paper presents results for the station A011 only. This is the output of System 9, the most important in the whole network since it is located in a big city with a population over 500,000. Besides presenting the results, practical insights are also given about the implementation of the methodology, particularly on the calculation of the time delays and the selection of the initial values of the model parameters.

Calculation of time delays

When describing the prediction architecture in previous sections, it has been stated that the time delay between the inputs and the output for each system in the network is a function of the output level at each time k , as given in Equation (3). The constant parameters in Equation (3) have been obtained in practice by analyzing the data from four past flood events that occurred in 2001, 2003, 2007, and 2008. Figure 6 displays the time histories of the water level at stations A002 and A011 (System 9) for the flood of 2007. By observing these data and the corresponding time histories of the other three floods, the time delay between the input (A002) and the output (A011) are given for several values of the output in Table 2. The time delays for intermediate output levels are calculated by means of linear interpolation.

The calculation of the time delays for all the stations that supply inputs to the system in the network is performed as for the case described for station A002.

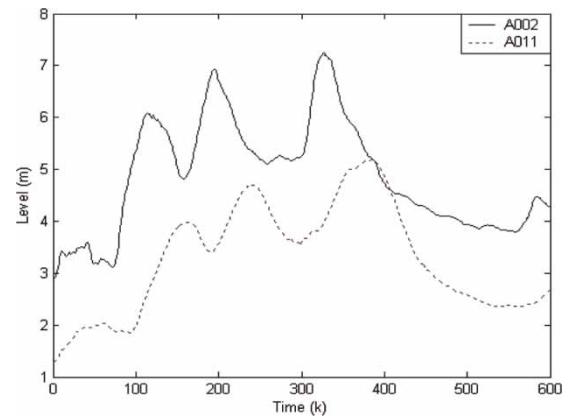


Figure 6 | Time history of flood that occurred in 2007.

Table 2 | Output levels in station A011. Time delays for the input at station A002

1.5 m	2.5 m	4 m	4.75 m	5.75 m
20 h	32 h	40 h	50 h	65 h

Selection of the initial model parameters

A good selection of the initial parameters is important and has to be done with some physical insight. The initial parameters for the outputs are selected as $\hat{a}(0) = 0.8$ for all the systems. The reason is that all the systems are stable (a bounded input produces a bounded output). This imposes the condition that the adaptation mechanism must keep this parameter always under the condition $\hat{a}(k) < 1$, since the models in this study in Equation (4) have a first order structure and $\hat{a}(k)$ is the single pole. The initial value 0.8 is coherent with this condition.

In order to understand the systematic procedure that has been developed in this work to select the initial input parameters, let us consider again System 9 as a prototype case. It has two inputs supplied by stations A002 and A260 and has the single output at station A011.

A first observation is done to the gauging plots (flow-level measurements) at the stations A002 and A260 shown in Figure 7. They are very relevant for the selection of the initial model parameters, since these plots give information on the relative influence of both inputs over the output.

By taking data from these two plots, a relation close to 6.5 is detected between the flows at the stations A002 and

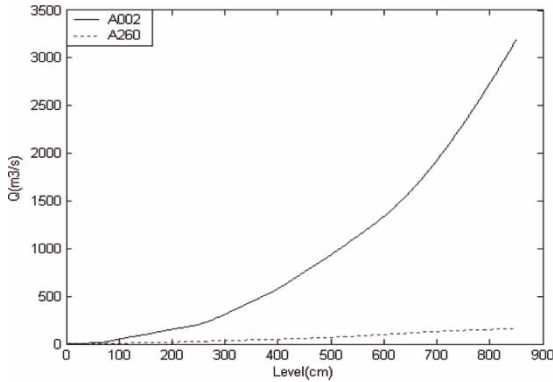


Figure 7 | Gauging plots.

A260 for a range of levels between 2 and 3 m. This means that the initial parameter for the input A002 should be around 6.5 times bigger than the parameter for the input A260 if the adaptation–prediction scheme starts operating when the A002 level is around 2–3 m range. The experience shows that this range is normal for long time periods.

To obtain the specific initial values for the inputs A002 and A260, data of a past flood episode are used when the system is in a steady state and the A002 level is within the 2–3 m range. Figure 8 displays the flood which occurred in 2001; note that these conditions hold at the beginning of the time history. The relation between input and output variables at this steady-state scenario can be written in the following form:

$$y_{A011}(k) = \hat{y}_{A011}(k) = 0.8_{A011}(k - 1) + 6.5 \hat{b}_{A260} u_{A002}(k - 20 - 1) + \hat{b}_{A260} u_{A260}(k - 12 - 1) \tag{12}$$

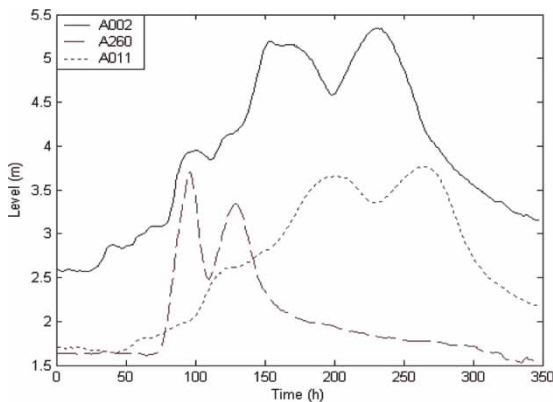


Figure 8 | Levels in System 9 for the 2001 flood.

where \hat{b}_{A260} is the remaining parameter value to be chosen. By taking the input and output variables from the plots in Figure 8 at time $k = 25$ and substituting them into Equation (12), we may solve for \hat{b}_{A260} and thus obtain all the initial model parameters for System 9. These values and those corresponding to all the systems in the network are given in Table 1.

It can be noticed that this procedure chooses the initial model parameters in normal steady-state conditions. Then, if the adaptation–prediction scheme is turned on in such a condition, the models are well prepared for the operation of the adaptation mechanism in flood conditions.

RESULTS

After having the models prepared as described in the previous sections, the prediction scheme is validated using a flood event that occurred in 2009. Figure 9 shows the time histories for the variables of System 9. It is recognized that the initial output level A011 is lower than 1.5 m, so that the initial time delay in station A002 is 20 h. It is also seen that the A002 level is 2 m, which allows the initial model parameters given in Table 1 to be taken.

The effectiveness of the prediction methodology is tested in two steps:

1. The ability of the estimation model (4) and the adaptation mechanism (6) is evaluated.
2. The precision of the predictive expert model (8) is checked.

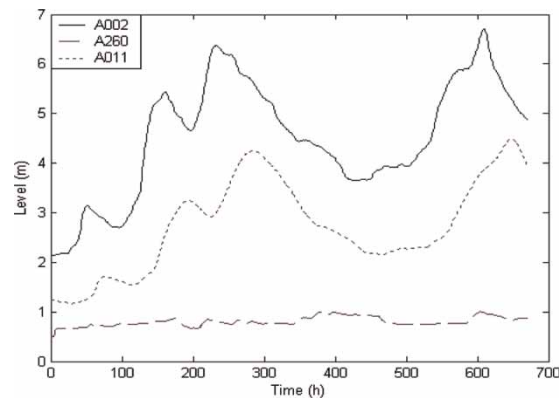


Figure 9 | Input-output levels in System 9 for the 2009 flood.

For the first step, Figure 10 shows the *a posteriori* estimation $\hat{y}(k|k)$ defined below in Equation (13) for the whole flood episode for the output A011, together with the real measured output:

$$\hat{y}_{A011}(k|k) = \hat{a}(k)y_{A011}(k-1) + \hat{b}_1(k)u_{A002}(k-r_1(k)-1) + \hat{b}_2(k)u_{A260}(k-r_2(k)-1) \tag{13}$$

The estimation error (7) is also plotted in Figure 10. The absolute mean value of this error for the whole episode is really small (0.003 m), thus emphasizing the effectiveness of the adaptive model to represent the actual input-output behavior.

In the second step, the effectiveness of the predictive expert model is tested, which takes the following form:

$$\hat{y}_{A011}(k+j|k) = \hat{a}(k)A(j)\hat{y}_{A011}(k+j-1|k) + \hat{b}_1(k)B_1(j)\hat{u}_{A002}(k-r_1(k)+j-1|k) + \hat{b}_2(k)B_2(j)\hat{u}_{A260}(k-r_2(k)+j-1|k) \tag{14}$$

where the values of the input and outputs for each j are obtained as indicated in Equation (9).

Two criteria are used for the assessment:

1. An analysis is carried out to study how the level peaks on the flood episode are predicted in terms of precision and anticipation. To do this, tests are performed where

forecasts are calculated within a time interval prior to the occurrence of significant peaks in the registered 2009 episode under study.

2. At each real-time instant k over the whole flood episode, the prediction for a future instant $k+j$ is performed and the absolute mean value of the prediction error over the whole episode is obtained. By doing this second analysis for different prediction instants j , further significant insights on the precision and anticipation ability are obtained.

In Figure 9, it is observed that two significant peaks are produced in station A011 for hourly time instants 285 and 648. Figures 11 and 12 show the predictions (14) made at time instants $k=239$ and $k=602$ over a future time horizon of 48 h. The time axis represents the prediction instants $j=1, \dots, 48$. The lower plots give the corresponding prediction errors.

In the Ebro river basin processing center, the system operators perform forecasts (usually once a day) of river levels in flood episodes with the help of MIKE 11-based DSS system. They solve the one-dimensional Saint-Venant equations with discretization time of 90 seconds and spatial discretization of 8 km.

In Figures 13 and 14, two forecasts (previous to maximum levels) calculated with the DSS are shown and are compared with forecasts (corresponding to the same time instant) calculated with APE models. The lower part of each figure plots the prediction errors resulting for each

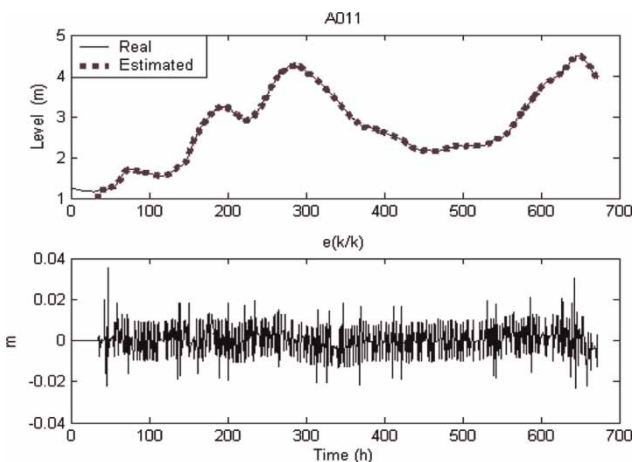


Figure 10 | Estimated and real output A011 plus estimation error.

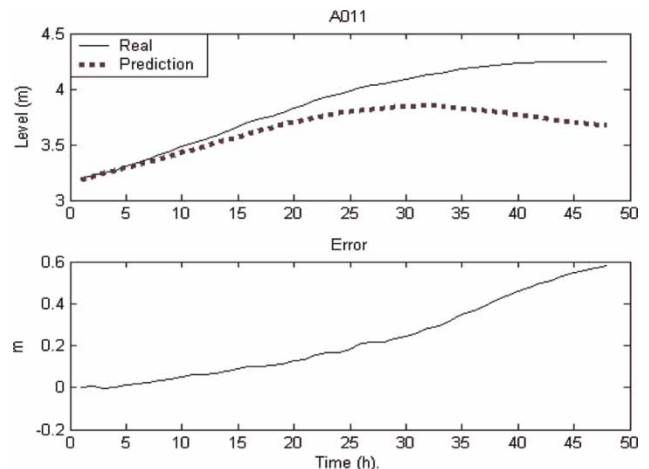


Figure 11 | Prediction at time 239 with a horizon of 48 h.

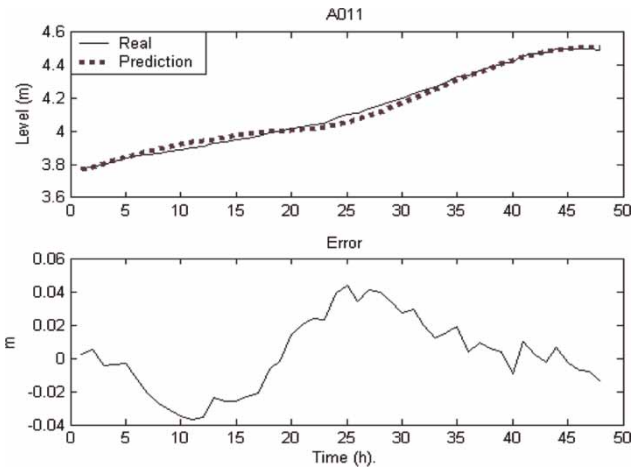


Figure 12 | Prediction at time 602 with a horizon of 48 h.

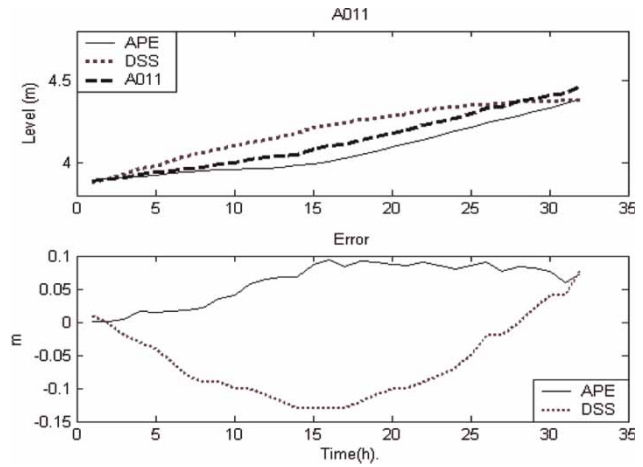


Figure 14 | Prediction at time 611 with a horizon of 32 h.

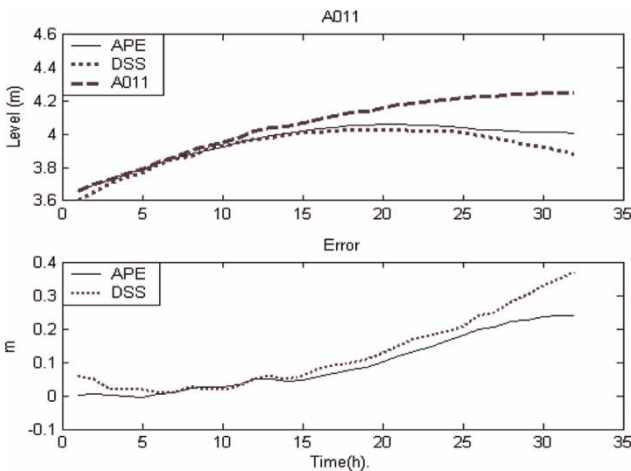


Figure 13 | Prediction at time 253 with a horizon of 32 h.

method. The absolute mean value of this error for the two forecasts is shown in Table 3.

From an overall view of Figures 11–14 and Table 2, it can be observed that the APE scheme supplies a satisfactory forecast of peak levels, with significant anticipation and reduced errors, being competitive with the system under current operation.

The second analysis is quantified by the following criterion:

$$E_{A011}(j) = \frac{1}{672} \sum_{k=1}^{672} |y_{A011}(k+j) - \hat{y}_{A011}(k+j|k)|. \quad (15)$$

Table 3 | Absolute mean prediction errors for water level

Prediction at time	Absolute mean prediction error (m)
253	APE 0.13 DSS 0.17
611	APE 0.06 DSS 0.07

It gives the absolute mean error, over the whole flood episode, between the predicted level output and the real one when a prediction for a prescribed future time instant $k + j$ is done at each real-time instant k . Table 4 gives this error for several future times.

Figures 15–19 show the full time histories of the predictions (and the errors) made at each real time instant k for the future time instants $k + j$.

The predictions observed in Figures 15–19, along with the absolute mean errors, confirm the ability of the proposed

Table 4 | Absolute mean prediction errors for level A011

Prediction instant j (h)	Absolute mean prediction error (m)
6	0.024
15	0.051
24	0.071
33	0.097
41	0.122

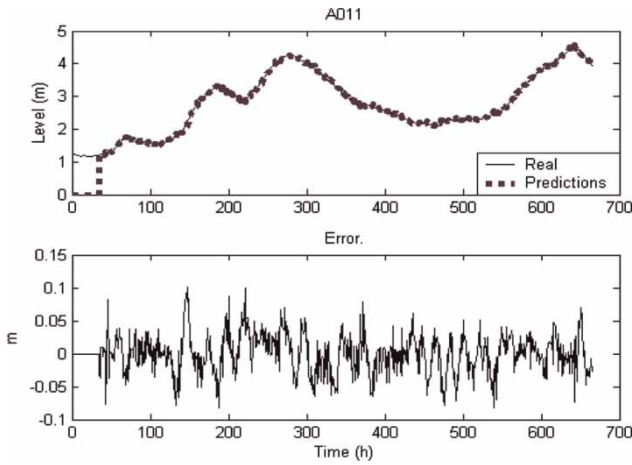


Figure 15 | Predictions made at each time k for future instants $k + 6$.

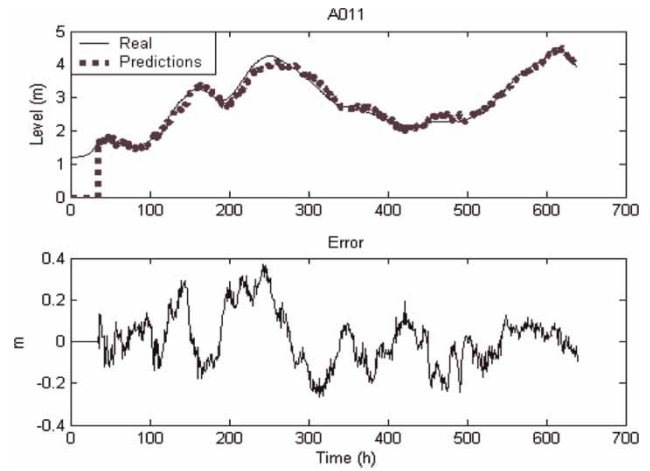


Figure 18 | Predictions made at each time k for future instants $k + 33$.

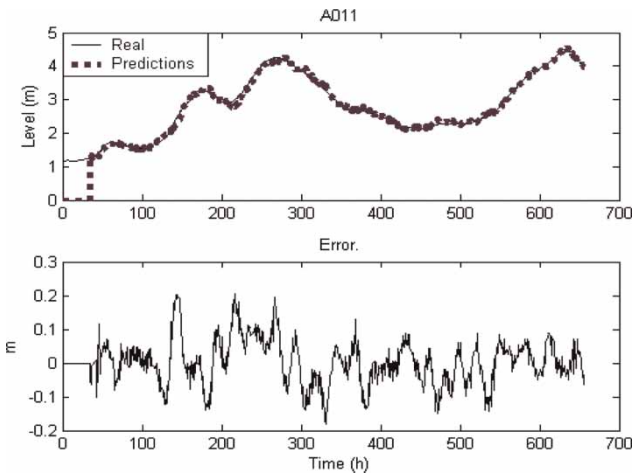


Figure 16 | Predictions made at each time k for future instants $k + 15$.

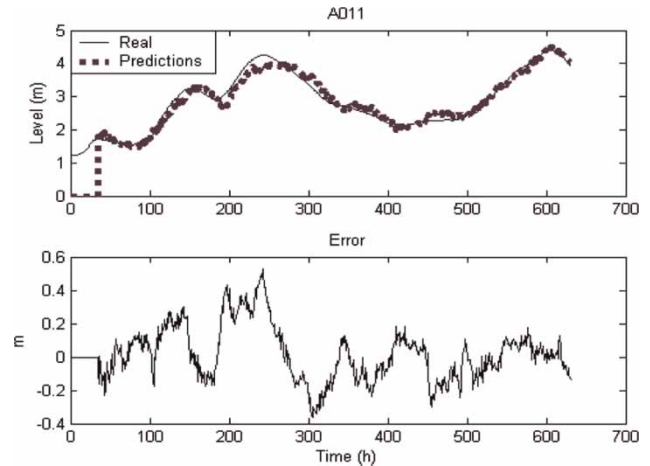


Figure 19 | Predictions made at each time k for future instants $k + 41$.

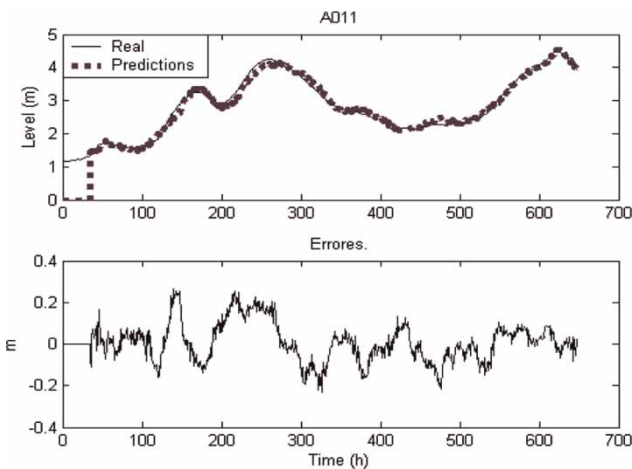


Figure 17 | Predictions made at each time k for future instants $k + 24$.

scheme to perform online forecasts with several future time horizons. These time horizons go from 6 h ahead to 41 h ahead. The error increases as the prediction horizon is longer, but the error is small in all cases. To interpret the main reason why the prediction error increases with the prediction horizon, we may come back to Equations (8) and (9) defining the prediction scheme and the subsequent discussion on which is the available information for the predictive models in the prediction intervals $[k, k + N]$. When the time horizon is small (in comparison with the time delays in the different reaches), all the required inputs may be available from measurements at the time k when the predictions are performed, since they may be

outputs from upstream reaches. In this case, the prediction uses exact information without uncertainty. For longer time horizons, some predictions may demand inputs that have not yet been measured at time k , which are predicted and thus subject to errors.

CONCLUSIONS

This paper has presented a new approach for river level forecasting with the following main components: (1) the river basin is modeled as a set of cascade interconnected input–output system; (2) each system is described by a discrete-time linear model with time delays, having water levels as inputs and outputs; (3) the model has time-varying parameters, which are updated by means of a stability-based adaptation mechanism to track the measured output; (4) the forecast scheme is based on the adapted models, which are redefined at each prediction time to incorporate the measured available information and complemented with expert knowledge through a fuzzy logic scheme.

The combination of the aforementioned four elements represents a novel approach to river level forecasting. The results shown in the paper confirm the potential for its application, which could be considered through an independent system or as a complement to be integrated into more complete operational DSSs. The main practical advantages added by the proposed approach can be summarized as follows:

- The APE models, once the initial parameters are adjusted, do not need any additional further parameterization performed by the flood manager technician.
- The computational needs are low.
- Knowledge on the riverbed topography is not necessary.
- In combination with a real-time data acquisition system, the models allow an autonomous forecast with a 1-h period.
- Since calculations are carried out only with levels and not with discharges, predictions are not affected by changes that may occur in river level-flow tables of gauge stations. These changes could be produced as a consequence of significant floods that could alter the geometry of the riverbed.

- During floods of major intensity, river overflows and floods happen. Such flows coming out of the river between two river gauge stations are always unknown. Therefore, they are a source of errors in the predictions made by hydraulic model-based systems. In these flood scenarios, APE models may be particularly well suited since they may be able to identify in real time the new flood dynamics.

REFERENCES

- Aguilar, J. V., Langarita, P., Linares, L., Gómez, M. & Rodellar, J. 2011 An adaptive predictive approach for river level forecasting. *Proceedings of IEEE Conference on Networking, Sensing and Control*, Delft, The Netherlands, pp. 92–97.
- Aguilar, J. V., Langarita, P., Linares, L. & Rodellar, J. 2009 Automatic control of flows and levels in an irrigation canal. *IEEE Trans. Ind. Appl.* **45** (6), 2198–2208.
- Aguilar, J. V., Langarita, P., Linares, L., Rodellar, J. & Soler, J. 2012 Adaptive predictive expert control of levels in large canals for irrigation water distribution. *Int. J. Adapt. Control.* **26**, 945–960.
- Alvisi, S., Mascellani, G., Franchini, M. & Bárdossy, A. 2006 Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrol. Earth Syst. Sci.* **10**, 1–17.
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005a Input determination for neural network models in water resources applications. Part 1. Background and methodology. *J. Hydrol.* **301** (1–4), 75–92.
- Bowden, G. J., Maier, H. R. & Dandy, G. C. 2005b Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *J. Hydrol.* **301** (1–4), 93–107.
- Bruen, R. & Yang, J. 2005 Functional networks in real-time flood forecasting – a novel approach. *Adv. Water Res.* **28**, 899–909.
- Fernández, N., Jaimes, W. & Altamiranda, E. 2010 Neuro-fuzzy modeling for level prediction for the navigation sector of the Magdalena river (Colombia). *J. Hydroinform.* **12** (1), 36–50.
- Islam, A. S. 2010 Improving flood forecasting in Bangladesh using an artificial neural network. *J. Hydroinform.* **12** (3), 351–364.
- Martín-Sánchez, J. M. & Rodellar, J. 1996 *Adaptive Predictive Control, from the Concepts to Plant Optimization*. Prentice Hall, Hemel Hempstead, UK.
- Moradkany, H., Hsu, K. L., Gupta, H. V. & Sorooshian, S. 2004 Improved streamflow forecasting using self-organizing radial-basis function artificial neural networks. *J. Hydrol.* **295**, 246–262.
- Pulido-Calvo, I. & Portela, M. M. 2007 Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. *J. Hydrol.* **332**, 1–15.
- Romanowicz, R. J., Young, P. C. & Beven, K. J. 2006 Data assimilation and adaptive forecasting of water levels in the

- river Severn catchment, United Kingdom. *Water Resour. Res.* **42** (6), W06407.
- Romanowicz, R. J., Young, P. C., Beven, K. J. & Pappenberger, F. 2008 A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting. *Adv. Water Res.* **31**, 1048–1056.
- Romeo, R., Linares, A. & García, E. 2004 Flood control in Ebro river basin. *Proceedings of the II Days on Decision Support Systems for Real Time Hydraulic and Hydrological Problems*, CEDEX, Zaragoza, Spain, pp. 259–278.
- Seckin, N. 2011 Modeling flood discharge at ungauged sites across Turkey using neuro-fuzzy and neural networks. *J. Hydroinformatic.* **13** (4), 842–849.
- See, L. & Opeshaw, S. 1999 Applying soft computing approaches to river level forecasting. *Hydrol. Sci.* **44** (5), 763–778.
- Shamseldin, A. Y. 2010 Artificial neural network model for river flow forecasting in a developing country. *J. Hydroinformatic.* **12** (1), 22–35.
- Tsanis, I. K., Coulibaly, P. & Daliakopoulos, N. 2008 Improving groundwater level forecasting with a feedforward neural network and linearly regressed projected precipitation. *J. Hydroinformatic.* **10** (4), 317–330.
- Young, P. C. 2002 Advances in real-time flood forecasting. *Phil. Trans. Royal Soc.: Math. Phys. Eng. Sci.* **362**, 1433–1450.

First received 31 October 2011; accepted in revised form 23 October 2012. Available online 19 December 2012