ELSEVIER

Journal of Hydrology 239 (2000) 132–147

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Comparison of short-term rainfall prediction models for real-time flood forecasting

E. Toth*, A. Brath, A. Montanari

Dipartimento di Ingegneria delle Strutture, dei Trasporti, delle Acque, del Rilevamento e del Territorio, University of Bologna, Bologna, Italy

Received 19 November 1999; revised 15 August 2000; accepted 21 September 2000

Abstract

This study compares the accuracy of the short-term rainfall forecasts obtained with time-series analysis techniques, using past rainfall depths as the only input information. The techniques proposed here are linear stochastic auto-regressive moving-average (ARMA) models, artificial neural networks (ANN) and the non-parametric nearest-neighbours method. The rainfall forecasts obtained using the considered methods are then routed through a lumped, conceptual, rainfall–runoff model, thus implementing a coupled rainfall–runoff forecasting procedure for a case study on the Apennines mountains, Italy. The study analyses and compares the relative advantages and limitations of each time-series analysis technique, used for issuing rainfall forecasts for lead-times varying from 1 to 6 h. The results also indicate how the considered time-series analysis techniques, and especially those based on the use of ANN, provide a significant improvement in the flood forecasting accuracy in comparison to the use of simple rainfall prediction approaches of heuristic type, which are often applied in hydrological practice. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Rainfall forecasting; Flood warning; Stochastic processes; Artificial neural networks; Non-parametric predictors

1. Introduction

The incorporation of quantitative precipitation forecasting (QPF) in flood warning systems has been acknowledged to play a key role, allowing for an extension of the lead-time of the river flow forecast, which may enable a more timely implementation of flood control (Brath et al., 1988). The QPF integration is particularly needed in small and medium-sized mountainous basins where, given the short response time of the watershed, a precipitation forecast is necessary for an extension of the lead-time of the flood warning. It is widely recognised that obtaining a reliable QPF is not an easy task, rainfall being one of the most difficult elements of the hydrological cycle to forecast (e.g. French et al., 1992), and great uncertainties still affect the performances of both stochastic and deterministic rainfall prediction models.

Interesting perspectives for the future are opened by numerical weather prediction models, but, up to now, they unfortunately do not seem able to provide accurate rainfall forecasts at the temporal and spatial resolution required by many hydrologic applications (Brath, 1999).

The timely use of remote sensing observations (radar data and satellite images) allows the issue of very short-term forecasts (nowcasting), based on the extrapolation of current weather conditions. Unfortunately, the outputs from satellite and radar images, although providing useful information on the precipitation pattern, do not allow a satisfactory assessment

* Corresponding author.
E-mail address: elena.toth@mail.ing.unibo.it (E. Toth).

0022-1694/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0022-1694(00)00344-9

of rain intensities yet (Krzysztofowicz, 1995). The radar detection is, in addition, particularly difficult in mountainous regions, because of ground occultation and altitude effects.

The QPF may be obtained also by means of time-series analysis techniques (Brath et al., 1998; Burlando et al., 1993). One should be aware that the predictive ability of these methods is limited owing to the low persistence in time that usually characterises rainfall time-series, even if sampled at a fine temporal scale. However, the moderate computer time and data availability required by such techniques make their application very attractive in the context of real-time flood forecasting. Thus, the benefits potentially achievable from their application with real precipitation data, in terms of efficiency of the flood forecast, are worth analysing.

The following time-series analysis techniques have been applied: (1) Linear stochastic auto-regressive moving-average models (ARMA), which express the future rainfall as a linear function of past data. The approach is thus linear, model-driven and parametric, i.e. it first requires the identification of the type of relationship among the variables (model identification) and then the estimation of model parameters. (2) Artificial neural network architectures (ANN), belonging to the non-linear, data-driven approaches: the resulting model depends on the available data to be "learned", without any a priori hypothesis about the kind of relationship, which is allowed to be complex and non-linear. (3) *K*-nearest-neighbour method (*K*-NN), a non-parametric regression methodology, not implying any structured interaction, but exploiting the closeness ("neighbourhood") between the most recent observations and *K* "similar" sets of observations chosen in an adequately large training sample.

The aim of our analysis is a comparison of the above methods, not only in terms of obtained rainfall depths, but also from an hydrological point of view, considering an integrated rainfall and runoff forecasting system operated on a real-world case study. The issued rainfall forecasts, therefore, will be routed through a lumped conceptual rainfall–runoff transformation model and the performances of the flow forecast will be analysed and compared.

Section 2 describes the study area and the data sets, along with the use of the data in the analysis. Sections 3–5 describe the examined methodologies and some relevant aspects regarding the application to short-term rainfall forecasting. We have gone into more details about artificial neural networks and nearest-neighbour methods because their application in hydrology is relatively new and less well established with respect to ARMA models. Section 6 presents and compares the obtained results in terms of predicted rainfall depths. Section 7 describes the hydrologic rainfall–runoff transformation model and introduces heuristic rainfall forecasting approaches used as benchmarks for the comparison of the time-series rainfall forecasting methods. The performances of the coupled rainfall–runoff forecasting systems corresponding to all the rainfall predictive schemes, both time-series methods and heuristic approaches, are then compared, evaluating the efficiency of the rainfall predictions in terms of the transformed river discharges. Section 8 offers conclusions and a perspective on future developments.

2. Case study

2.1. Study area and data set description

The case study considered herein is referred to the Sieve River basin, a first-order tributary of the Arno River, located in the Apennines Mountains in Central Italy. Since the Sieve joins the right bank of the Arno just a few kilometres upstream of the city of Florence, the prediction of the corresponding flood hydrographs plays a key role in determining the flood risk for the city. The main stream length is 58 km, the areal extension of the watershed is around 830 km^2 and the time of concentration is approximately 10 h. The closure section is Fornacina, where hourly discharge observations were collected between 1 January 1992 and 31 December 1996.

For the same observation period, hourly temperatures (for estimating the potential evapo-transpiration) were measured at four stations and hourly rainfall depths at 12 raingauges located inside the study basin, thus allowing the computation of the average areal precipitation over the watershed. The rainfall is spatially averaged (with the Thiessen polygons method) so that it is consistent with the use in the lumped rainfall–runoff transformation model. The

averaging also improves the efficiency of the forecast, compared to the efficiency of making forecasts for each one of the gauges separately before averaging the results over the basin. In fact, the averaging produces a smoothing of non-stationarities and fluctuations recorded at each gauge, while preserving the general pattern of the storm event over the watershed (Burlando et al., 1993).

Given the predominance of the presence of null or very low values in the rainfall series and our interest in flood forecasting, we limited our analysis, both in the calibration and validation phases, to the rainfall values belonging to storm events, so as to identify the temporal pattern characterising the storms, whose persistence properties are different from those of dry or low rainfall sequences. A storm event was conventionally defined as an interval containing at least one hourly rainfall depth exceeding 1 mm ("wet" observation), preceded by at least 5 and followed by at least 20 h of rainfall lower than 1 mm, provided that in the same time interval the corresponding observed hourly discharge increases by at least 20 m^3/h. In this way, the wet observations are followed by a time interval not shorter than the concentration time of the basin, so that the observed river discharges include the total contribution of the surface runoff generated by each storm event. In addition, the requirement on the change in discharge allows the exclusion of the events that do not produce hydrologic response (dry periods). In the observation period a total of 84 storm events were identified, including a total number of 4580 hourly rainfall observations, for an average storm length of 55 h.

2.2. Description of the calibration approaches

Two alternative approaches were followed for estimating the parameters of the models: split-sample calibration and adaptive calibration.

In the *split-sample calibration*, the storm events were divided into two sets: a calibration (or training) set and a validation set, where the former contains twice the number of events as the latter. The sets were chosen so as to have approximately the same proportion of low duration–high intensity and high duration–low intensity rainfall events. It follows that the training set should be representative of the characteristics of different kinds of event, thus allowing the calibration phase to "learn" an accurate representation of the problem domain. These first two sets of events will be referred to as split-sample pair A. In order to verify the variability of the forecasts results if different split-samples are used, three further pairs (B–D) of calibration and validation sets were selected, by slightly changing the definition of storm event described in Section 2.1, so as to include also events not considered before. In pair B, the validation set remains the same, so that a fair comparison with the results described above is allowed. For the calibration set, instead, the events (again twice the number of the validation events) containing the highest rainfall values (above 3 mm/h) in the remaining data were chosen, lowering at the same time the threshold imposed on the change in the corresponding discharge from 20 to 10 m^3/s. In pairs C and D, new events were chosen for both the calibration and validation sets, so as to force the presence of the most extreme rainfall events in the calibration set still more (even if at the expense of the corresponding increase in discharges), with the aim of improving the "learning" of rainfall peaks. The new events in these last two experiments were chosen requiring a peak of at least 4 and 5 mm/h, respectively. In addition, the number of observations following the wet observations was reduced in the last case from 20 to 5; in this way, there is a further reduction of the presence of low rainfall values in the sets. We here anticipate that the forecast performances on the split-samples B–D in the analysis were found to be very close to those obtained when using the A split-sample pairs. For this reason and for brevity and clarity of presentation, only the results referring to case A will be described in what follows.

In the *adaptive calibration* no database of past significant observed events is supposed available, but only the most recently observed values, immediately preceding the forecast instant. Thus, the set of data is extremely poor for generalisation purposes, but, on the other hand, the limited amount of data to be processed allows the re-calibration of the model on-line, at each time step, as soon as new observations become available. This adaptability enables the model parameters to adjust to the properties of an ongoing event, by capturing the characteristics of the current meteorological situation.

For both calibration approaches, in correspondence

of each hourly time step belonging to the validation set, a rainfall forecast was issued for the subsequent 1–6 h, using the most recent observations as input. The resulting forecasted rainfall values were processed as inputs to the rainfall–runoff transformation model, thus providing the discharge forecast.

3. ARMA models

Most of the time-series techniques traditionally used for modelling water resources series fall within the framework of the ARMA class of linear stochastic processes. They are usually denoted as ARMA(p,q) models, where p and q are the auto-regressive and moving-average orders, respectively (Box and Jenkins, 1976; Brockwell and Davis, 1987; Bras and Rodriguez-Iturbe, 1994). They describe each observation of the time series as a weighted sum of p previous data, and the current as well as q previous values of a white noise process:

$$x_t = \phi_1(x_{t-1} - \mu_x) + \phi_2(x_{t-2} - \mu_x) + \cdots + \phi_p(x_{t-p} - \mu_x) + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2} + \cdots + \theta_q\eta_{t-q} + \mu_x, \quad (1)$$

where x_t is the investigated time series; η_t, a white noise, i.e. a non-correlated, zero-mean random variable that is also not correlated with the past values of x_t; $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_p$, the auto-regressive and moving-average parameters, respectively; and μ_x, the mean of the time series.

Parameter estimation for ARMA models can be performed in several ways. We applied here an approximation in the spectral domain of the Gaussian maximum likelihood function, which was first proposed by Whittle (1953) for short-memory models.

3.1. ARMA model application

The application of low-order ARMA processes to model short-term precipitation values is considered here, following the modelling framework proposed by Brath et al. (1988) and Burlando et al. (1993). The application of ARMA models requires the data to be stationary and this is often not the case for hourly rainfall observations, whose statistical properties may vary with the season. Nonetheless, the limited number of rainfall events in the observation period prevented us, in the split-sample calibration, from grouping the events in monthly periods, as it is usually done in hydrology to circumvent non-stationarity. In the adaptive calibration application, non-stationarity is accounted for by allowing the model parameters to vary with time since the calibration is performed solely on the progress of the current event. We preferred not to perform any preliminary transformation of the data in order to make them as close to Gaussian as possible. In fact, Gaussian data are not required for the forecast application of ARMA models, since they provide the best linear prediction even in the non-Gaussian case (Brockwell and Davis, 1987).

The selection of the model orders, p and q, was driven by some results available in literature. Obeysekera et al. (1987) determined an equivalence between the correlation structure of an ARMA(1,1) model and some point process models, like the Poisson rectangular pulses and the Neyman–Scott white noise models (see Rodriguez-Iturbe et al., 1984). On the other hand, the Neyman–Scott rectangular pulses model, which has proved to represent the stochastic structure of rainfall better (Rodriguez-Iturbe et al., 1987), has a correlation structure equivalent to that of an ARMA(2,2) process.

In the adaptive calibration, the parameters are estimated in correspondence with each forecast instant, on the basis of the last values measured in real-time. The number of past observations to be used for each calibration was chosen on the basis of the results of a previous study (Brath et al. 1998). The estimation of the parameters was performed there with a number w of observations x_t immediately preceding each forecast instant, with w varying from 7 to 100, aiming at identifying the value of w that provides the best forecasting performances. The results showed that for increasing w, the efficiency of the forecast improved moderately for short lead-times (1–3 h), but a longer set of past data (more than 3 days of previous hourly observations) provided a much better performance for lead-times longer than 4 h. Thus, we set the moving window of past rainfall observations to be used in each adaptive calibration equal to the 100 last measured hourly observations (that is, $w = 100$).

4. Artificial neural networks

As mentioned above, the rainfall generation mechanism is extremely difficult to identify and model: there is still much that is not understood regarding the small-scale behaviour of precipitation and the underlying physical laws. It is not easy to recognise all the existing complex, and typically non-linear, relationships between the various aspects of the dynamical process (French et al., 1992; Kuligowski and Barros, 1998).

To accommodate the inability of an ARMA-type model to account for these effects, non-linear statistical models have been proposed, such as the threshold model and the bilinear model, but their complexity often makes them unsuitable for operational applications (Chakraborty et al., 1992). In addition, such models still belong, like the ARMA type, to the model-driven approaches, requiring an identification of the kind of relation between the variables (model selection) and an estimation of the selected model parameters. From these considerations stems the interest in alternative non-linear forecasters, belonging to the data-driven approaches, where no a priori relationship between known parameters and observed values has to be hypothesised and no knowledge of the underlying process is needed. Both artificial neural network architectures and nearest-neighbour methods present these appealing characteristics.

Artificial neural networks have been widely studied and applied to a variety of problems, including hydrometeorological simulation and forecasting. Several studies have been dedicated to the prediction of river flows (at a time scale ranging from one year to one day) with no exogenous inputs, that is with only the use of past flow observations (e.g. Karunanithi et al., 1994). However, the large majority of the ANN hydrologic applications predict future flows based on the knowledge of previous rainfall depths (and other meteorological variables) along with past observed flows. The appeal of the use of ANN as black-box rainfall–runoff models lies mainly in their ability to reproduce the non-linear nature of the rainfall–runoff transformation, and encouraging results have been obtained in literature on both real and synthetic hydrologic data (e.g. Hsu et al., 1995; Minns and Hall, 1996; Shamseldin, 1997; Zealand et al., 1999).

The use of ANN for rainfall forecasting has not been fully explored, yet. A pioneer work is the study by French et al. (1992), who applied a neural network to forecast 1 h ahead, two-dimensional rainfall fields on a regular grid. The forecast was based on synthetically generated rainfall grid values corresponding to the previous hourly period. Kuligowski and Barros (1998) generated a QPF of point precipitation accumulated over the following 6-h period using as inputs the antecedent rainfall depths measured in adjacent gauges and the radiosonde-based wind direction. Our interest is mainly focused on the hydrological use of rainfall forecasts: the implemented scheme provides a spatially averaged rainfall forecast for all the lead-times from 1 to 6 h, directly usable as input to the rainfall–runoff lumped model, and the only available information used are past rainfall observations.

Neural networks distribute computations to relatively simple processing units called *neurons*, grouped in layers and densely interconnected. Three different layer types can be distinguished: *input layer*, connecting the input information (and not carrying out any computation), one or more *hidden layers*, acting as intermediate computational layers between input and output, and an *output layer*, producing the final output. In correspondence to each computational node J, each entering value (II_i) is multiplied by a connection weight (w_{ij}). Such products are then all summed with a neuron-specific parameter, called bias (b_j), used to scale the sum of products into a useful range. The computational node finally applies a non-linear *activation function* (f), often a sigmoid, to the above sum producing the node output (OJ).

Neural networks are *trained* (calibrated) with a set of observed input and output (called *target* to be distinguished from the network final output) data pairs, the *training* data set, which is processed repeatedly, changing the values of the parameters until they converge to values such that each input vector produces output values as close as possible to the desired target vectors. Several variants of the popular and extensively tested BackPropagation (BP) training algorithm were tested in the present work. The Levenberg–Marquardt algorithm, a quasi-Newton method that proved to be the quickest and less easily trapped in local minima among all the tested training techniques, was finally chosen.

The comparison of the performance of networks

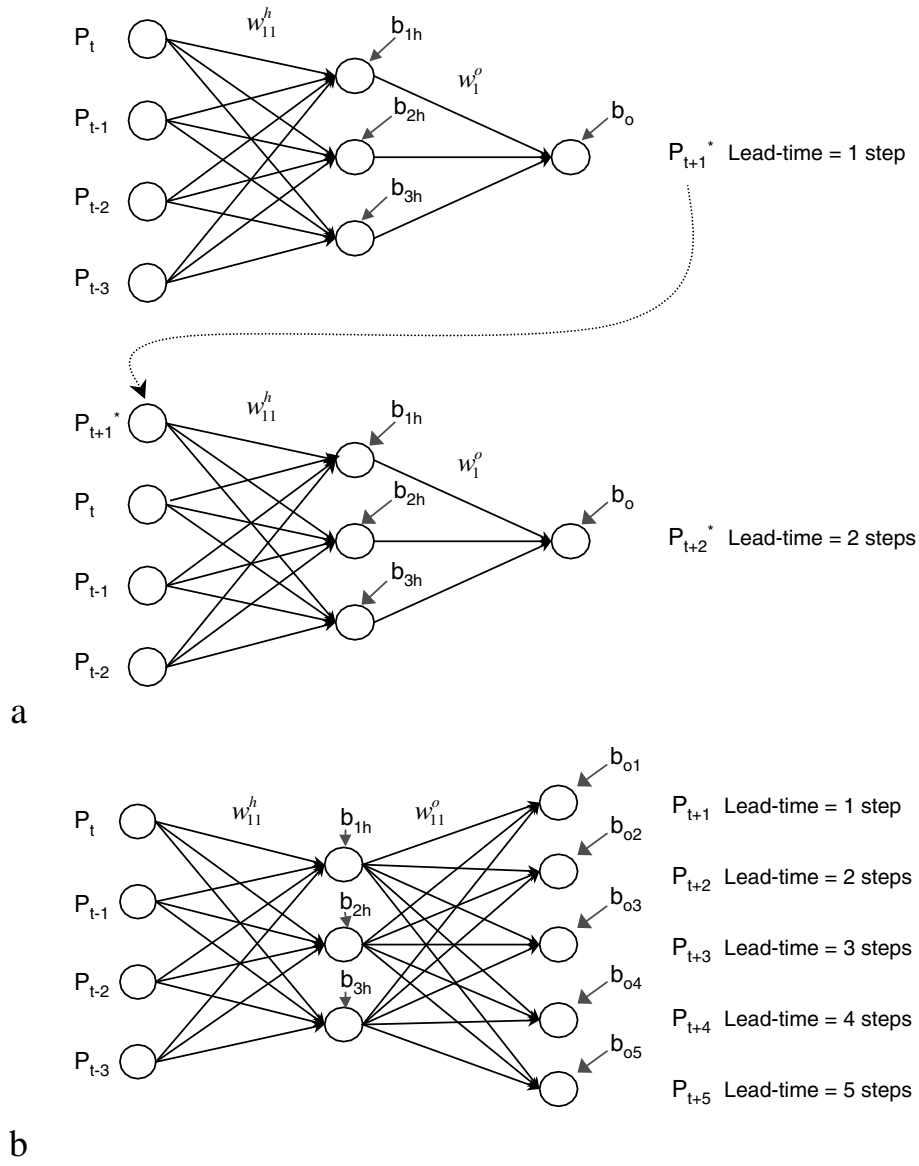

Fig. 1. (a) Feed-forward recursive network. P_t is the precipitation process; w_{ij}^h, w_i^o are the connection weights towards the hidden and output layers, respectively, and b_j are the node biases. (b) Feed-forward direct multi-step network. P_t is the precipitation process; w_{ij}^h, w_{ij}^o are the connection weights towards the hidden and output layers, respectively, and b_j are the node biases.

allowing different feed-forward and feedback *connections between the nodes* led to the choice of a classic multi-layer feed-forward network, where the information flows only in one direction, from the input through the hidden up to the output layer. Networks allowing other types of connections between the layers (recurrent and cascade-forward networks) were also tested: they provided comparable results, but longer time and more random initialisations were needed for training.

As far as the *number of hidden layers* is concerned, there is no theory yet to tell how many hidden layers are needed. It has been proved that only one layer of hidden units "suffices to approximate any function

with finitely many discontinuities to arbitrary precision", provided the activation functions of the hidden units are non-linear (the "Universal Approximation Theorem", see Hornik et al., 1989). The hidden nodes also allow taking into account the presence of non-stationarities in the data, such as trends and seasonal variations (Maier and Dandy, 1996). Obviously, the introduction of additional hidden layers allows the fit of a larger variety of target functions. On the other hand, the use of more than one hidden layer substantially increases the number of parameters to be estimated and the training time, and it exacerbates the problem of local minima, increasing the need for several random initialisations. Moreover, the addition of hidden layers often fails to provide noticeable improvement in the out-of-training forecasting application (see, for example, Zealand et al., 1999), as confirmed also in our case study applications.

The use of an ANN for forecasting time series implies that the input nodes are connected to a number of past observed values supposed sufficient to identify the process at future time steps. For forecasting several time steps ahead (*multi-step ahead prediction*), two methods have been considered. The first is the *recursive multi-step* method: the network has only one output node, forecasting a single step ahead, and the network is applied recursively, using the previous predictions as inputs for the subsequent forecasts (see Fig. 1a). This forecasting technique is similar to the one used by the ARMA-type models. The second method (*direct multi-step*) exploits the capability of a neural network to provide a multiple output, when several nodes are included in the output layer, and each output node represents one time step to be forecasted (see Fig. 1b). In our preliminary tests, the *recursive multi-step* method provided very good results for lead-time equal to one time step, but there was a drastic deterioration when the lead-time increased, as could be expected, because in the recursive methodology the forecast errors are propagated into subsequent forecasts. Given the importance of good performance in correspondence with the longer lead-times, the *direct multi-step* method was chosen.

4.1. ANN application

The most crucial disadvantage of ANN models is that the optimal network architecture and properties are highly problem-dependent and no definitive established methodology exists to deal with the neural network modelling problem. As described above, preliminary forecast analyses were performed on a few storm events, for choosing, on the analysed time series, different architectures and properties of the networks.

From all the above considerations it was decided to extensively test ANN architectures consisting of a multi-layer feed-forward network with only one hidden layer. The network is trained with the Levenberg–Marquardt algorithm and the multi-step approach is the *direct multi-step* method. The optimal complexity of the model, that is the number of input and hidden nodes, will be determined, as it is usually done, by a trial-and-error approach.

5. *K*-NN method

The *K*-nearest-neighbour method has its origins as a non-parametric statistical pattern recognition procedure, aiming at distinguishing between different patterns according to chosen criteria. Among the various non-parametric techniques, in the sense that no theoretical or analytical relation is known or assumed between the inputs and the outputs, it is the most intuitive, but nevertheless possesses powerful statistical properties. Yakowitz (1987) and Karlsson and Yakowitz (1987a,b) did considerable work in extending the *K*-NN method to time-series and forecasting problems, obtaining satisfactory results and constructing a robust theoretical base for the *K*-NN method. The intuitiveness of the approach and the powerful theoretical basis have made the method attractive to forecasters, especially in the hydrologic field, where the method found successful applications (Karlsson and Yakowitz, 1987a,b; Galeati, 1990; Kember and Flower, 1993; Todini, 1999).

The prediction of a time series is based on a local approximation, making use of only the nearby observations. For each forecast instant t, let $\bar{\mathbf{x}}^d(t) = (x_t, \ldots, x_{t-d+1})$ be a *feature vector* of past records. A *feature vector* is a vector that summarises the whole past history in a smaller-dimension vector of observations supposed to contain most of the information relevant to the forecast. The method assumes that

the probability distribution of the random variable conditioned on the entire past $(x_{t+1}/x_t, x_{t-1}, \ldots)$, is the same as that of the random variable conditioned on only the d past observations $(x_{t+1}/\bar{\mathbf{x}}^d(t))$.

It was proved that, even if $\bar{\mathbf{x}}^d(t)$ does not satisfy the above "history summarisation" properties, the K-NN forecaster will be asymptotically optimal among all the forecasters defined on the feature vector $\bar{\mathbf{x}}^d(t)$. That is, under fairly general circumstances, convergence to the optimal forecaster is assured as the historical data set increases (Karlsson and Yakowitz, 1987b). Let us indicate the expectation of the next value as $\hat{x}_{t+1}$, conditioned on the current feature vector $\bar{\mathbf{x}}^d(t)$, that is,

$$\hat{x}_{t+1} = E[x_{t+1}|\bar{\mathbf{x}}^d(t)]. \tag{2}$$

To estimate $\hat{x}_{t+1}$, the K-NN method imposes a metric, denoted by $\|\cdot\|$, on the feature vector $\bar{\mathbf{x}}^d(t)$ to find the set of K past *nearest neighbours* of $\bar{\mathbf{x}}^d(t)$, i.e. the K d-dimensional vectors of past observations: $\bar{\mathbf{x}}^d(t_j)$, $J = 1, \ldots, K$, which minimise $\|\bar{\mathbf{x}}^d(t) - \bar{\mathbf{x}}^d(t_j)\|$.

The most intuitive and widely used metric to identify neighbours is the *Euclidean norm*, which, for a d-dimensional vector $\bar{\mathbf{Z}}^d = (z_1, z_2, \ldots, z_d)$, is

$$\|\overline{Z}^d\| = \left(\sum_{i=1}^{d} z_i^2\right)^{1/2}. \tag{3}$$

The forecast is then obtained by averaging the temporal evolution of the nearest neighbours, assumed to be similar to the evolution of the current situation, that is,

$$\hat{x}_{t+1} = \frac{1}{K}\sum_{j=1}^{K} x_{t_j+1}. \tag{4}$$

The generalisation to higher lead-times L is straightforward:

$$\hat{x}_{t+L} = \frac{1}{K}\sum_{j=1}^{K} x_{t_j+L}. \tag{5}$$

Thus, in our case, the K-NN algorithm looks through all consecutive d-dimensional vectors in the entire historical rainfall depths database and locates K of these d-ples, which are closest to the vector of the d most recent rainfalls. The prediction of the next rainfall is then taken to be the average of the rainfall subsequent to these K historical nearest neighbours. It may be noticed that the K-NN approach does not require the selection of a class of models and the estimation of the model parameters, so that the identification of a specific form of the input/output relationship is not needed.

5.1. K-NN method application

The nature of the nearest-neighbour method makes an adaptive calibration approach completely meaningless, because the approach is based on the presence of an extended database. In fact, in addition to being data-driven, it is a method that does not detect any input/output mapping function, not even a posteriori (whereas the ANNs do), and it has, therefore, no extrapolation ability when presented with an unfamiliar input vector. As a consequence, only the split-sample application was performed.

A trial-and-error procedure was implemented for finding the number of nearest neighbours, K, and the dimension of the feature vector d (corresponding to the number of past rainfall data considered representative for the forecast), providing the best performing forecasts. As Karlsson and Yakowitz (1987a) emphasise, K and d seem, and indeed are, parameters but the method itself is nevertheless non-parametric. In fact, K and d do not imply a model for $\hat{x}_{t+L}$: the purpose of their search "is pragmatic; it is to make the Nearest-Neighbour forecaster as accurate as possible for a given database".

6. Analysis and comparison of rainfall forecasting results

The performances of the considered time-series methods (ARMA models, ANN and K-NN method) are first of all analysed and compared assessing the respective ability to forecast the spatially averaged rainfall depths belonging to the validation set storms. The issued rainfall forecasts will be successively routed through the conceptual rainfall–runoff transformation model and the performances of the resulting river flow forecasts will be analysed and compared in the next session.

The performances of the various forecasting techniques are investigated through a trial-and-error process. The rainfall forecasting results are extremely difficult to

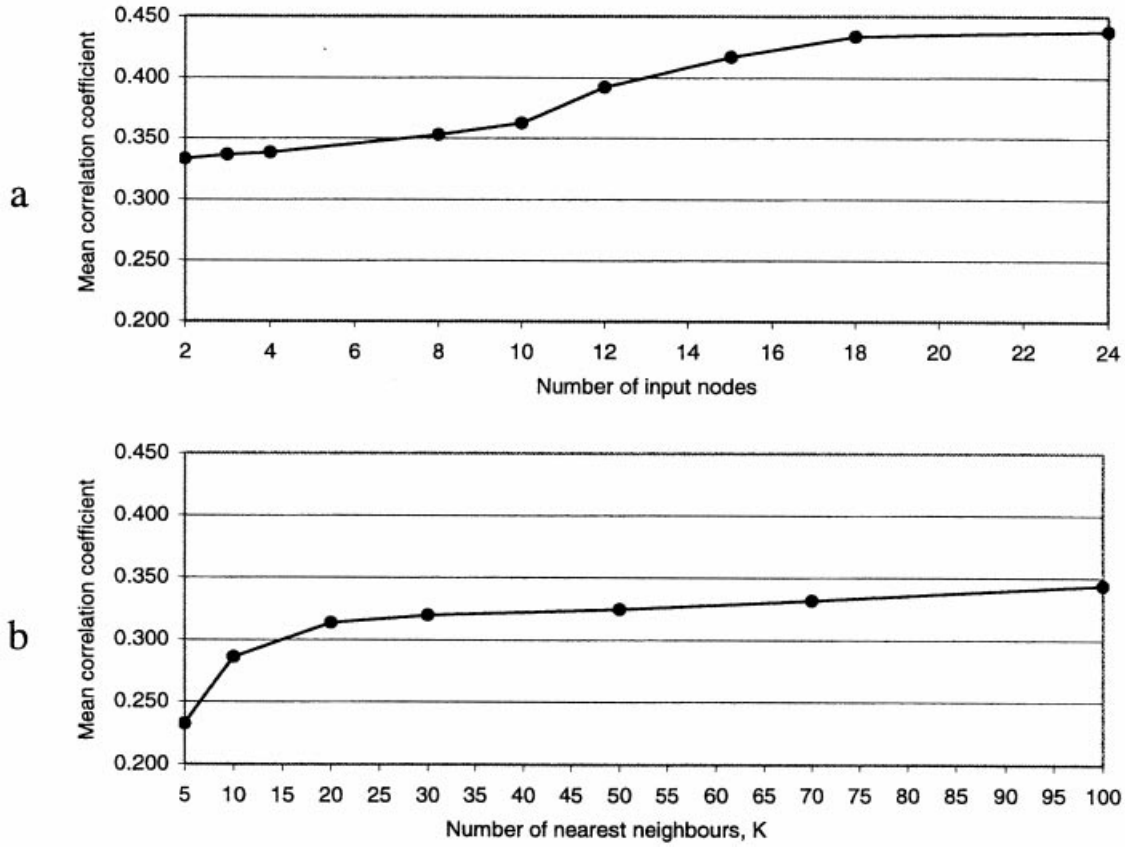

Fig. 2. Mean correlation coefficients of the one to six steps ahead rainfall forecasts issued: (a) by ANN architectures with varying number of input nodes, and (b) by nearest-neighbour implementations with varying number of neighbours.

analyse, and a variety of quantitative measures (root mean square error, mean absolute error, coefficient of persistence, efficiency coefficient, correlation coefficient, index of agreement) were considered for synthesising the effectiveness of the performances over all the lead-times, so to make a comparison possible. Different methods provided the best results for different forecast performance measures and for different lead-times, so that a performance classification could not be assessed unequivocally.

Among the considered measures, the correlation coefficient (CC) was chosen, following other rainfall forecasting studies (French et al., 1992; Kuligowski and Barros, 1998), as the most representative for assessing rainfall forecast performance. The correlation coefficient is given by the covariance of the forecasts and observations divided by the product of the respective standard deviations. It ranges from -1.0 to 1.0, with higher values indicating better agreement.

6.1. Description of the split-sample and adaptive applications

6.1.1. ARMA models: trial-and-error processes

In the *split-sample calibration*, low-order ARMA models were implemented, both purely auto-regressive and with a moving-average component, provided that the sum of the auto-regressive and moving-average orders, $p + q$, is equal to or less than 6. All the tested ARMA models provided almost analogous performances (mean CC = 0.291–0.293).

In the *adaptive calibration*, the tested ARMA models had auto-regressive orders of 1 and 2 and moving-average orders less or equal to 3. The trend

Table 1
Correlation coefficients of rainfall predictions

Time-series analysis technique	Optimal model configuration	Lead-time (h)						Mean CC
		1	2	3	4	5	6	
ANN split-sample	$NI = 18$, $NH = 2$	0.689	0.511	0.407	0.358	0.331	0.327	0.437
ARMA split-sample	Almost all equivalent	0.686	0.430	0.276	0.173	0.114	0.077	0.293
Nearest Neighbours	$K = 70$, $d = 2$	0.709	0.493	0.336	0.239	0.174	0.110	0.344
ANN adaptive	$NI = 3$, $NH = 3$	0.527	0.307	0.196	0.171	0.169	0.162	0.255
ARMA adaptive	$p = 1$, $q = 0$	0.744	0.472	0.283	0.134	0.060	0.003	0.281

of the obtained performances, for all the ARMA models, is characterised by relatively good performance for lead-time of 1 h, followed by a collapse in correspondence of longer time horizons. The overall best results are provided by ARMA models with parsimonious configurations, indicating ARMA(1,0) as the best performing model (mean CC = 0.281).

6.1.2. ANN: trial-and-error processes

In the *split-sample application*, architectures with a number of input nodes *NI* ranging from 2 to 24 were tested. For each input layer dimension, the number of hidden nodes (*NH*) was progressively increased from 2 to 8 nodes. A deterioration of the forecasting performance on the validation set, indicating over-fitting, was always shown for moderate dimensions of the hidden layer, the best results corresponding to *NH* between 2 and 6. The performance of ANN architectures, considering all the lead-times, improved as the number of input nodes (*NI*) increased, with modest additional gain for more than 15–18 nodes (see Fig. 2a).

The networks tested in the *adaptive calibration* of were extremely parsimonious (both *NI* and *NH* ranging from 2 to 5), because the limited number of training samples (here chosen corresponding to 100 past observations for a fair comparison with the ARMA adaptive approach) would make a complex network easily subject to over-fitting. The optimal network complexity for adaptively calibrated neural networks seems to correspond to numbers of input and hidden nodes, *NI* and *NH*, equal to 3.

6.1.3. Nearest-neighbours: trial-and-error process

A trial-and-error procedure was implemented for a number of nearest neighbours, *K*, ranging from 5 to 100 and a dimension of the feature vector, *d*, ranging from 2 to 12. The improvement of the performance with an increasing number of nearest neighbours is less noticeable for more than 20 neighbours and there is no marginal improvement in the overall performance when increasing *K* beyond 70 (see Fig. 2b). Small values (from 2 to 4) of the feature dimension *d* gave the most satisfactory results for each given number of neighbour vectors *K*.

6.2. Overall comparison of rainfall forecasts

A direct comparison of all the implemented methodologies is presented here, so as to highlight the relative strengths and limitations. Table 1 shows the correlation coefficients of the rainfall predictions provided by each time-series analysis technique (with the modelling configurations identified in the trial-and-error processes), both for different lead-times and for the mean over all the lead-times. The table ranks ANNs above the nearest-neighbour method and this latter above the ARMA models, on our precipitation data. Such results seem to indicate the appropriateness of non-linear approaches when modelling rainfall time-series.

With regard to the ARMA models, the adaptive calibration application provides very good performance for very short lead-times, while the split-sample approach achieves better results for lead-times longer than 3 h. This is consistent with the results obtained by Burlando et al. (1993), who found a superiority of the adaptive calibration method tested for predictions 1 and 2 h ahead.

The ANN adaptive calibration application proves to be unreliable for short lead-times and especially inadequate for reproducing low rainfall, even if it is satisfactorily stable for lead-times longer than 3 h. For a better result on short lead-times, we should probably

Table 2
Mean correlation coefficients of predictions for different rainfall ranges

	Rainfall intensity range		
	Low rainfall (<0.1 mm)	Medium rainfall (0.1–1 mm)	High rainfall (>1 mm)
ANN split-sample	0.203	0.257	0.178
ARMA split-sample	0.216	0.180	−0.001
Nearest neighbours	0.268	0.121	0.119
ANN adaptive	0.075	0.115	0.028
ARMA adaptive	0.132	0.205	0.008

have resorted to the *recursive multi-step* method, but at the expense of the performance for several steps ahead. The ANN corresponding to the split-sample calibration gives the overall best results for lead-times longer than 2 h, even if this kind of structure slightly penalises the 1-h ahead forecast.

Table 2 presents the correlation coefficients obtained for different rainfall intensity ranges, to get an insight into the type of rainfall for which each method performs better. Both the nearest-neighbour and ARMA split-sample models provide a good fit of low-intensity precipitation, but the ANN with split-sample calibration allows the best results for medium and high rainfall values.

The forecasts issued when the calibration is of the split-sample type present, overall, smaller deviations from the observed values. Nevertheless, Table 2 highlights the difficulties experienced by all the methods in reproducing high-intensity rainfall. The analysed techniques do not seem to properly reproduce high precipitation values. This is probably due to the fact that they are influenced by the majority of low-rainfall observations in the calibration sets. Analysing the issued forecasts, it may be noticed that the adaptive calibrations often issue unreliable forecasts in correspondence with abrupt changes in the rainfall intensities; on the other hand, the forecasts obtained with the split-sample approaches tend, when the lead-time increases, towards the average of the calibration data, often underestimating the values that follow the high-intensity occurrences. It may be inferred that this is the same reason that causes the presence of an apparent upper limit in the split-sample forecasts.

If a different goodness-of-fit criterion had been chosen in the trial-and-error processes, for instance maximising the fit on the highest rainfall values alone, better results probably would have been obtained when predicting extreme rainfall. However, such a choice would provide poor performance in the prediction of events that, although not extreme, are relevant in terms of vulnerability of flood-prone areas.

Considering all the lead-times and all the rainfall categories, the ANN with split-sample calibration seems the most adequate among all the considered approaches, at least in reference to our case study. In addition, it should be underlined that almost all the computational effort spent for the implementation of the ANN split-sample application is concentrated in the training phase, while the issue of the forecasts with the trained network is practically instantaneous, thus making this approach very appealing in a real-time forecasting framework.

7. Rainfall–runoff transformation: analysis and comparison of discharge forecasts

7.1. Hydrologic model description

The deterministic model used for simulating the rainfall–runoff transformation is a conceptual continuous simulation model called ADM (Franchini, 1996), which is based on the concept of probability distributed soil moisture storage capacity. The catchment is assumed to be composed of an infinite number of elementary areas (each one with a different soil moisture content and a different soil moisture capacity) and the proportion of elementary areas that are saturated is described by a distribution function: the total surface runoff is the spatial integral of the infinitesimal contribution deriving from the saturated elementary areas. The model is divided into two main blocks: the first represents the water balance at soil level, while the second represents the transfer of runoff production at

the basin outlet. The soil, in turn, is divided into two zones: the upper zone produces surface and subsurface runoff, having as inputs precipitation and potential evapo-transpiration, while the lower zone (whose input from the first one is the percolation flow) produces base runoff. The transfer of these components to the outlet section takes place in two distinct stages: the first represents the flow along the hillslopes towards the channel network, while the second, the flow along the channel network towards the basin outlet. The 11 parameters of the ADM model were accurately calibrated with the Shuffled Complex Evolution global optimisation algorithm (Duan et al., 1992). The conceptual model has been separately parameterised off-line (fixed calibration) so that its parameters do not change during the forecasting period. Joint optimisation of the rainfall–runoff model coupled with the rainfall forecasters would, of course, be possible, but this might cause, owing to compensation effects, undesirable biases in the calibration of the parameters of the rainfall–runoff model (and possibly a departure from their physical meaning).

7.2. Standards of reference: heuristic rainfall predictive approaches

To evaluate the performances of the analysed time-series forecasting methods when used for providing the inputs to the rainfall–runoff transformation model, the resulting discharges will be compared with those obtained with some predictive benchmarks, consisting of rainfall forecasting approaches of a purely heuristic nature. Three predictive procedures have been considered among the possible alternatives that a hydrological practitioner may envisage in case sophisticated modelling tools are not available and in case the only information at his/her disposal are the most recent rainfall observations.

Probably, the most widespread approach when using a rainfall–runoff transformation model in a real-time framework is to assume that the future rainfall will be null (*null rainfall*). It is an optimistic hypothesis, assuming that the forecast is issued at the end of the event, whereas, especially in watersheds with short response time, a forecast is needed earlier in the storm progress.

A second term of comparison, widely used in forecasting theory, is the *persistent* method, which equals the future rainfall intensity, over all the investigated lead-times, to the last measured value,

$$\hat{x}_{t+L} = x_t, \qquad \forall L. \tag{6}$$

The last investigated heuristic approach, somehow similar to the *persistent* method, consists in extrapolating future values setting the intensity for each given lead-time L equal to the mean intensity measured over the last L observations, that is,

$$\hat{x}_{t+L} = \frac{\sum_{i=1}^{L} x_{t-i+1}}{L}. \tag{7}$$

This last predictive scheme will be denoted as the *modified persistent* method.

7.3. Analysis and comparison of flow forecasting performances

The performance of the discharge forecasts attainable using the QPF provided by the different rainfall predictive models was evaluated by computing the corresponding coefficient of efficiency, which is widely recognised as one of the most suitable goodness-of-fit measures for runoff. For the analysis of discharge performance results, the discharge series chosen as a reference was not the series of observed discharges, but the hourly discharges simulated by the conceptual model when using as inputs the observed precipitation ("true" discharges). This scenario was considered in order to be able to evaluate the improvement obtainable by the rainfall forecasting module alone, independently of the effects of the simulation errors induced by possible residual inadequacies of the hydrologic model.

The coefficient of efficiency for each lead-time L is given by:

$$E_L = 1 - \frac{\sum(Q_{t+L} - \hat{Q}_{t+L})^2}{\sum(Q_{t+L} - \bar{Q})^2}, \qquad L = 1,\ldots,6 \tag{8}$$

where $\hat{Q}_t$ is the discharge at time t forecasted using as input the predicted rainfall values; Q_t, the value of the corresponding "true" discharge (known rainfall); and $\bar{Q}$, the mean of the Q_t series. The summations are extended to all the issued forecasts, that is, to all the forecast instants t belonging to all the validation events.

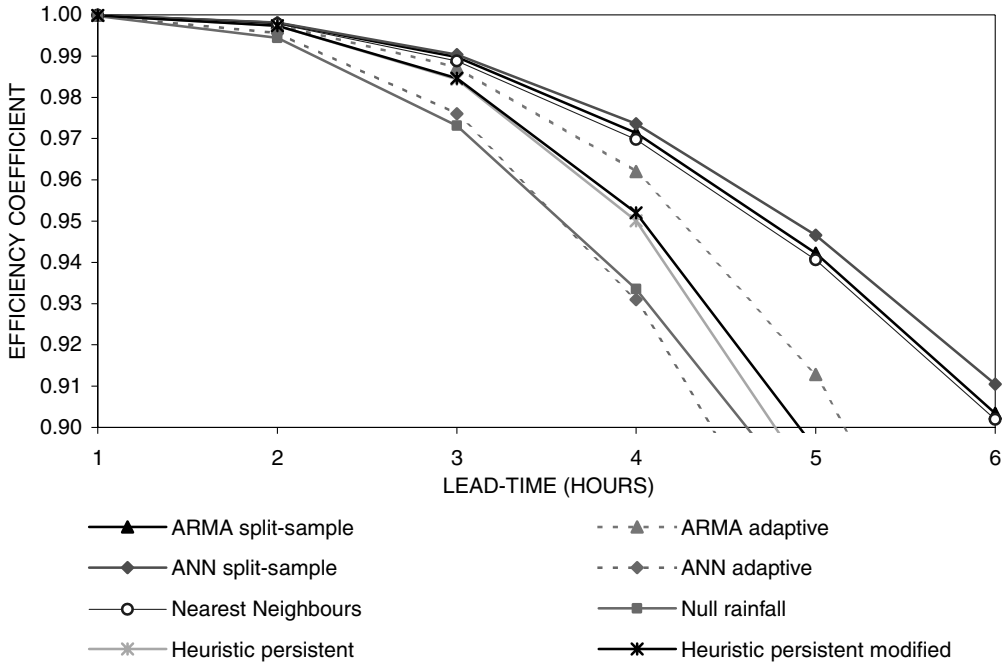

Fig. 3. Efficiency coefficients of the river flows corresponding to the different rainfall forecasting procedures: ARMA models with split-sample and adaptive calibration; ANN with split-sample and adaptive calibrations; and nearest-neighbour method. The heuristic approaches: null rainfall, *persistent* rainfall method and *persistent modified* method.

Fig. 3 shows the performances, in terms of efficiency coefficient, of the coupled rainfall–runoff forecasting schemes obtained with all the considered rainfall forecasting procedures. It may be observed that the hydrological processes taking place in the rainfall–runoff transformation tend to level out all the rainfall forecasts corresponding to very short lead-times and it dampens out most of the variability between the different methods highlighted in the analysis of the rainfall depths described in the previous section. As a consequence, the good performance of the ARMA adaptive calibration approach for 1-h lead-time becomes unnoticeable, whereas the relevant deterioration with increasing lead-time confirms how the method is less appropriate than the three split-sample calibration methods.

Moreover, the response time of the watershed shifts the poor results provided by the adaptive calibration of ANN from the low to the large lead-times. The unsatisfactory results of the ANN with adaptive calibration is probably due to the strong limitation of the forecasting ability of ANN when trained on inadequate data sets, as the records of low-intensity rainfall immediately preceding the arrival of the highest precipitation intensities certainly are. As was expected, the null rainfall hypothesis proves to be unrealistic, since it may strongly underestimate the rainfall volumes, whereas the *persistent* methods (both in the traditional formulation and *modified*) provide an improvement with respect to the null rainfall approach.

The ARMA model with adaptive calibration shows a relevant deterioration with increasing lead-time, as could be expected, given the poor performance of the rainfall forecasts obtained using this approach in correspondence to longer lead-times. The discharges simulated with the ARMA split-sample rainfall predictions appear slightly closer to the reference discharges in comparison with the results obtained with the nearest-neighbours scheme, whilst the performance of the rainfall forecasts appeared superior for the nearest-neighbour method. It may be hypothesised that this is due to the non-linear and threshold effects characterising the rainfall–runoff transformation modelling.

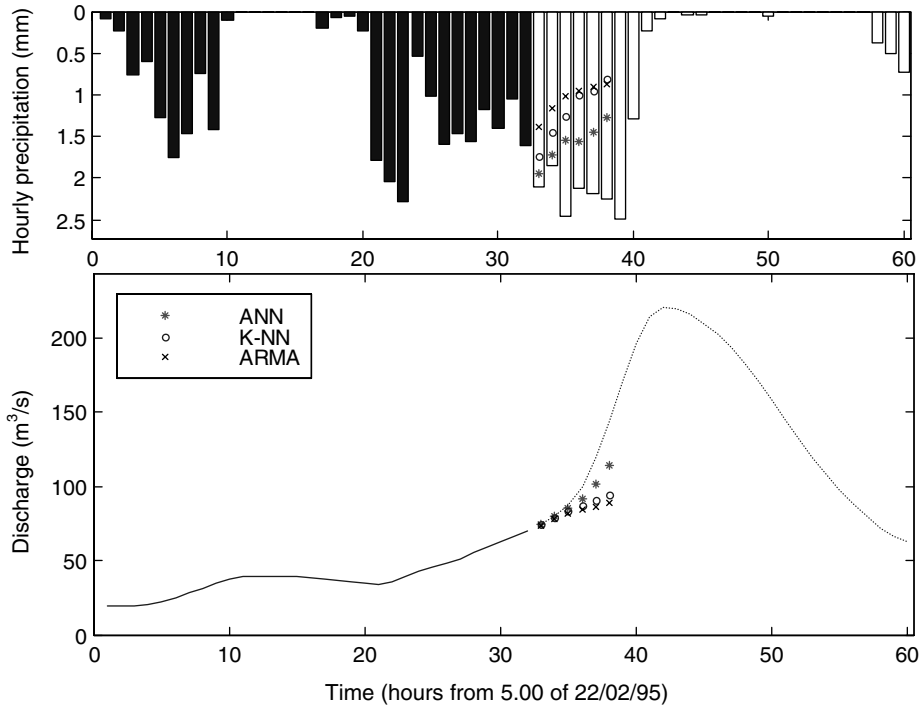

Fig. 4. Example of rainfall and runoff observations and forecasts during the event of 22 February 1995: observed precipitation (past, black bars; future, white bars), observed discharge (past, solid line; future, dotted line) and 1 to 6 h ahead predictions issued by ANN, nearest-neighbour and ARMA methods with split-sample calibration.

Overall, the split-sample calibration techniques seem to be preferable with respect to the adaptive calibrations. The reason lies probably in the "experience" they learned from past samples, which allows them to better reproduce the rainfall evolution mechanism for all the lead-times. An example of the rainfall and runoff forecasts obtained with the split-sample calibrations in the course of a typical event, is shown in Fig. 4.

The split-sample calibrated ANNs produced overall the highest efficiency values. Therefore, the coupled rainfall–runoff forecasting comparison confirms, with regard to our case study, the superiority of ANN already shown in the analysis of the performances of rainfall forecasts. In order to improve the performance when forecasting the rainfall peaks, a larger number of heavy precipitation events (as might be found in longer observation periods) would be needed in the training set, thus exposing the neural network to a larger number of extreme events in the calibration of the model.

8. Conclusions

This paper reports the results of a comparison of time-series analysis techniques for short-term rainfall forecasting to be used as input in a deterministic rainfall–runoff model for real-time flash-flood forecasting.

Different structures of ARMA models, ANNs and nearest-neighbour approaches were applied for forecasting storm rainfalls that occurred in the Sieve River basin, Italy, in the period 1992–1996. The forecast performances of each technique were evaluated by comparing observed and predicted rainfall data and also by comparing the river discharge predictions provided by a conceptual rainfall–runoff model using observed and predicted rainfall as input.

Different approaches for calibrating the rainfall predictors were applied. The calibration procedures making use of an extended training set of past rainfall data provided the best performances, especially for longer forecast time horizons, with a superiority of

the predictors based on ANN architectures, within the analysed range of parameters and model structures. The dampening effect induced by the rainfall–runoff transformation processes tends to level the performances of the different methods, so that, when considering the predicted runoff, the satisfactory fit that some of the methods allow for very short-term rainfall forecasts is of little worth.

Overall, the study indicates that the considered time-series analysis techniques provide an improvement in the flood forecasting accuracy with respect to the use of intuitive, heuristic rainfall prediction approaches, even if the rainfall forecasting performance measures indicate only a weak to moderate relationship between forecasted and observed values. This is due to the fact that past rainfall observations alone are not sufficient to predict future precipitation accurately, not even for short time periods.

The results shows that the use of time-series analysis techniques for precipitation forecasting may allow an extension of the lead-time up to which a reliable flood forecast may be issued, providing a quick prediction based on past values solely and directly in the format required by the rainfall–runoff transformation model. On the other hand, strong limitations to a time-series analysis approach are due to the lack of information needed for a reliable prediction. More substantial improvement should certainly be pursued through numerical weather prediction models, once they are able to provide timely rainfall forecasts at a temporal and spatial scale compatible with the requirements of flood forecasting in small and medium-sized basins.

Acknowledgements

We are grateful to the three anonymous reviewers whose suggestions allowed us to improve the presentation of our results. The work presented here has been partially supported by the European Community through the grant ENV4-CT97-0529 (FRAMEWORK project), by the National Research Council of Italy — Group for the Prevention of Hydrogeological Disasters, contract no. 98.00587.PF42 and by the Ministry of University and Research in Science and Technology (MURST) of Italy through its 40% national grant to the program on “Climatic and anthropogenic effects on hydrological processes”, contract no. 9908032871_001.

References

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, CA.

Bras, R.L., Rodriguez-Iturbe, I., 1994. Random Functions and Hydrology. Dover, New York.

Brath, A., 1999. On the role of numerical weather prediction models in real-time flood forecasting. In: Proceedings of the International Workshop on River Basin Modeling: Management and Flood Mitigation, September 25–26, 1997, Monselice (Italy), pp. 249–259.

Brath, A., Burlando, P., Rosso, R., 1988. Sensitivity analysis of real-time flood forecasting to on-line rainfall predictions. In: Siccardi, F., Bras, R.L. (Eds.), Selected Papers from the Workshop on Natural Disasters in European-Mediterranean Countries, Perugia, Italy, pp. 469–488.

Brath, A., Montanari, A., Toth, E., 1998. Su alcune tecniche stocastiche per il miglioramento delle prestazioni del preannuncio di piena (in Italian, with English Abstr.). Proc. XXVI Convegno di idraulica e costruzioni idrauliche, Catania, Italy, vol. II, pp. 159–172.

Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer, New York.

Burlando, P., Rosso, R., Cadavid, L.G., Salas, J.D., 1993. Forecasting of short-term rainfall using ARMA models. J. Hydrol. 144, 193–211.

Chakraborty, K., Mehootra, K., Mohan, C.K., Ranka, S., 1992. Forecasting the behavior of multivariate series using neural networks. Neural Netw. 5, 961–970.

Duan, Q., Sorooshian, S., Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resour. Res. 28 (4), 1015–1031.

Franchini, M., 1996. Use of a genetic algorithm combined with a local search method for automatic calibration of conceptual rainfall runoff models. Hydrol. Sci. J. 41 (1), 21–39.

French, M.N., Krajewski, W.F., Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network. J. Hydrol. 137, 1–31.

Galeati, G., 1990. A comparison of parametric and non-parametric methods for runoff forecasting. Hydrol. Sci. J. 35 (1), 79–94.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.

Hsu, K., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall–runoff process. Water Resour. Res. 31 (10), 2517–2530.

Karlsson, M., Yakowitz, S., 1987a. Nearest-neighbor methods for nonparametric rainfall–runoff forecasting. Water Resour. Res. 23 (7), 1300–1308.

Karlsson, M., Yakowitz, S., 1987b. Rainfall–runoff forecasting methods, old and new. Stochastic Hydrol. Hydraul. 1, 303–318.

Karunanithi, N., Grenney, W.J., Whitley, D., Bovee, K., 1994.

Neural networks for river flow prediction. J. Comput. Civil Engng 8 (2), 201–220.

Kember, G., Flower, A.C., 1993. Forecasting river flow using nonlinear dynamics. Stochastic Hydrol. Hydraul. 7, 205–212.

Krzysztofowicz, R., 1995. Recent advances associated with flood forecast and warning systems. US National Report to IUGG, 1991–1994, Rev. Geophys. 33 (Part 2, Suppl. S), 1139–1147.

Kuligowski, R.J., Barros, A.P., 1998. Experiments in short-term precipitation forecasting using artificial neural networks. Mon. Weather Rev. 126, 470–482.

Maier, H.R., Dandy, G.C., 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. 32 (4), 1013–1022.

Minns, A.W., Hall, M.J., 1996. Artificial neural networks as rainfall–runoff models. Hydrol. Sci. J. 41 (3), 399–417.

Obeysekera, J.T.B., Tabios III, G.Q., Salas, J.D., 1987. On parameter estimation of temporal rainfall models. Water Resour. Res. 23 (10), 1837–1850.

Rodriguez-Iturbe, I., Gupta, V.K., Waymire, E.C., 1984. Scale considerations in the modeling of temporal rainfall. Water Resour. Res. 20 (11), 1611–1619.

Rodriguez-Iturbe, I., Febres de Power, B., Valdes, J.B., 1987. Rectangular pulses point process models for rainfall: analysis of empirical data. J. Geophys. Res. 92 (D8), 9645–9656.

Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall–runoff modelling. J. Hydrol. 199, 272–294.

Todini, E., 2000. Real-time flood forecasting: operational experience and recent advances. In: Marsalek, J. et al. (Eds.), Flood issues in contemporary water management, Kluwer Academic Publisher, The Netherlands, pp. 261–270.

Whittle, P., 1953. Estimation and information in stationary time series. Ark. Mat. 2, 423–434.

Yakowitz, S., 1987. Nearest neighbor method for time series analysis. J. Time Ser. Anal. 8 (2), 235–247.

Zealand, C.M., Burn, D.H., Simonovic, S.P., 1999. Short term streamflow forecasting using artificial neural networks. J. Hydrol. 214, 32–48.