Journal of Hydrology (2007) **332**, 1–15

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds

Inmaculada Pulido-Calvo [a,*], Maria Manuela Portela [b]

[a] *Departamento Ciencias Agroforestales, EPS, Campus Universitario de La Rábida, Universidad de Huelva, 21819 Palos de la Frontera, Huelva, Spain*
[b] *Departamento Engenharia Civil, SHRH, Instituto Superior Técnico (IST), Avda. Rovisco Pais, 1049-001 Lisboa, Portugal*

Received 1 June 2005; received in revised form 20 June 2006; accepted 21 June 2006

KEYWORDS
Daily flow forecasting;
Neural networks;
Intervention series;
Convolution process;
Hybrid model;
Portuguese watersheds

Summary Since a few years ago, the computational neural networks (CNNs) had become one of the most promising tools for stream flow forecasting. However, most of the works presented in the literature is focused in watersheds that, besides stream flow records, generally utilizes as inputs records of other hydro-meteorological variables.

In this study the performance of feed forward CNNs to forecast one-day ahead daily flows at large Portuguese watersheds considering that only flows in previous days are available for the calibration of the models were analyzed. For that purpose several CNN approaches were implemented and compared. Besides the CNN having as inputs the flows in previous days or those flows plus differenced flow data also in previous days, auxiliary inputs were used to apply intervention series to the CNN predictor model; a convolution process of the input variables was carried out; and a hybrid methodology combining CNN and ARIMA models was applied.

The CNN having inputs of flows in the three previous days combined with a convolution process of the input sequence proved to be able to provide very accurate estimates of the daily flows. A preliminary analysis of the capability of that approach to forecast daily flows at a watershed different from the one considered in the calibration of the parameters of the model was also carried out. The results showed that it is also possible to get daily flows forecasts at watersheds with insufficient flow data.

Introduction: general scope of the work

The development and the implementation of successful water resource management tools to assess engineering and environmental problems, such as flood control, on-line reservoir operation, hydropower generation, water quality

* Corresponding author. Tel./fax: +34959217533.
E-mail addresses: ipulido@uhu.es (I. Pulido-Calvo), mps@civil.ist.utl.pt (M.M. Portela).

0022-1694/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2006.06.015

control or river ecosystem constraints, among several others, often require the analysis of stream flow data. Whenever such data sets are insufficient or unavailable, models providing stream flow estimates must be applied.

For many years, the transformation of rainfall into runoff has been studied in order to extend stream flow series or to establish stream flows at ungauged watersheds. Two of the major approaches presented in the literature to assess the rainfall-runoff process involve physical based models and 'black box' approaches.

A physical based model generally aims to formulate the physical process in terms of each of its relevant components, such as interception, depression storage, infiltration, overland flow, interflow, percolation, evaporation, and transpiration. Thus, it is often too complex and too demanding in terms of data and parameters requirements, for widespread practical application (Jakeman and Hornberger, 1993; Abrahart and See, 2000). Also, the accuracy of the model predictions is generally affected by subjective components as the establishment of its parameters it is often highly dependent on the user's expertise and experience in what concerns the watershed behavior (Duan et al., 1992, 1993, 1994; Sorooshian et al., 1993; Hsu et al., 1995; Yapo et al., 1996; Tokar and Johnson, 1999; Chang and Chen, 2001).

In a 'black box' approach a model is applied to identify a direct mapping between inputs and outputs without detailed consideration about the internal structure of the physical processes. Compared with a physical based model, the 'black box' approach has much less data requirements. There are many practical situations such as stream flow real-time forecasting where the main goal is to make accurate predictions at specific watershed locations, and therefore, a 'black box' approach may be preferred to not expend the time and effort required to develop, validate and implement a physical based model whose importance is in the understanding of the rainfall-runoff process (Hsu et al., 1995; Lorrai and Sechi, 1995).

It should be pointed out that the data availability often determines the model choice. In fact, continuous measurements of precipitation and stream discharge can be easily and cost effectively obtained (at least, in Portugal) when compared with continuous measurements of soil characteristics, initial soil moisture, infiltration, and groundwater characteristics. Therefore, a 'black box' approach that operates based only on the first set measurements can be much more suitable for operational forecasting purposes than a physical based model that also requires the latter set of measurements (Tokar and Johnson, 1999).

The artificial or computational neural networks (ANNs or CNNs) can be classified as 'black box' type models. A CNN is a non-linear mathematical structure capable of representing the complex non-linear processes that relate the inputs to the outputs of a system. CNNs models are increasingly being applied in many fields of science and engineering and usually provide highly satisfactory results. Some specific applications of CNN to water resource management and planning include the modeling of monthly, daily and hourly rainfall-runoff process (Hsu et al., 1995; Lorrai and Sechi, 1995; Mason et al., 1996; Abrahart et al., 1999; Tokar and Johnson, 1999; Thirumalaiah and Deo, 2000; Tokar and Markus, 2000; Chiang et al., 2004; Moradkhani et al., 2004; Anctil and Rat, 2005), real-time river stage forecasting (Thirumalaiah and Deo, 1998, 2000; Abrahart and See, 2000, 2002; See and Openshaw, 2000; Cameron et al., 2002), rainfall forecasting (French et al., 1992; Zhang et al., 1997; Kuligowski and Barros, 1998), groundwater modeling (Roger and Dowla, 1994; Yang et al., 1997), assessment of stream's hydrologic and ecological response to climate change (Poff et al., 1996), water demand prediction in urban and irrigation delivery systems (Griñó, 1992; Pulido-Calvo et al., 2002, 2003), and drought analysis (Shin and Salas, 2000), etc. A comprehensive review of the applications of CNNs in hydrology was presented by the ASCE task committee (ASCE, 2000a,b).

In the time series forecasting issues past observations of one or more variables are collected and introduced as input data in a model that describes the underlying relationships among those variables and allows estimating future realizations of one of the same (Zhang, 2003). Recently, CNNs have been extensively applied to time series forecasting (Griñó, 1992; Al-Saba and El-Amin, 1999; Abrahart and See, 2000; Zhang et al., 2001; Zhang, 2003; Gutiérrez-Estrada et al., 2004) and data infilling techniques (Elshorbagy et al., 2000; Khalil et al., 2001).

In this paper, the CNN approach was applied to the short-term forecasting of daily stream flows in Portuguese rivers. However, and as much as the literature about the subject reveals, the constraints assumed in the analysis as well as and some of the procedures applied were different from those typically considered.

In fact, in its typical applications, the calibration and validation of a CNN based model are accomplished for the same watershed where flow forecasts are required. For that purpose long enough stream flows series at the river section where flow forecasts are required must be available. This constraint often inhibits the application of CNNs. Hence, the study performed included the analysis of the applicability of a CNN with parameters established for a given watershed having long enough flow series to daily flow forecasting both at the same watershed and at another watershed having insufficient flow data (Elshorbagy et al., 2000).

Two more aspects distinguish the analysis performed from the typical CNN applications. One of those aspects is related with the watershed area: while many of the CNN results presented in the literature are restricted to small watersheds, the study accomplished was focused on very large watersheds. This way, larger implies more averaging effects and consequently more stability in modeling and forecasting results.

The other aspect is related with the input data. Typical applications of CNNs to short-term forecasting of stream flows consider as input data flows and precipitations in previous time steps. Also, potential evapotranspiration data and/or of air temperature data are considered in some of the studies. In the applications carried out for Portuguese rivers only daily stream flows (stream flow time series) were considered as input data. Other measurements, such as those of meteorological variables or of watershed features were not included because, on one hand, they are not as easy to obtain as stream flow data and, on the other hand, the forecasted flows achieved based only on stream flow data were accurate.

This paper evaluates the performance of feed forward CNN models trained with the Levenberg–Marquardt algorithm (Shepherd, 1997) for the purpose of stream flow time series forecasting (one-step daily flow forecast model). To verify if more accurate CNN solutions could be achieved, three variations of the typical CNN models or additions to those models were tested, namely: (a) auxiliary inputs were used to apply intervention series to the CNN predictor model (Griñó, 1992; Abrahart and See, 2000); (b) a convolution process of the input variables to the CNN model was carried out (De Vries and Principe, 1991); and (c) a hybrid methodology combining CNN and ARIMA models was applied to take advantage of the unique strength of both models in nonlinear and linear modeling, respectively (Wedding and Cios, 1996; Hansen and Nelson, 1997; Zhang, 2003).

Theoretical background, data, and approaches

Computational neural network models: general concepts

Computational neural networks (CNNs) are mathematical models inspired by the neural architecture of the human brain. The CNNs can recognize patterns and learn from their interactions with the 'environment'. The most widely studied and used structures are multilayer feed forward networks (Rumelhart et al., 1986). A typical four-layer feed forward CNN has g, n, m and s nodes or neurons in the input, first hidden, second hidden and output layers, respectively [the notation of the neural network is (g, n, m, s)]. The parameters associated with each of the connections between nodes are called weights. All connections are 'feed forward'; that is, they allow information transfer only from an earlier layer to the next consecutive layers.

Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal (x_i) is a weight (W_{ji}). The effective incoming signal (I_j) to node j is the weighted sum of all the incoming signals, according to

$$I_j = \sum_{i=1}^{g} x_i W_{ji}. \tag{1}$$

The effective incoming signal, I_j, is passed through an activation function (sometimes called a transfer function) to produce the outgoing signal (y_j) of the node j. In this study, the linear function ($y_j = I_j$) was used in the output layer and the sigmoid non-linear function, in the hidden layers:

$$y_j = f(I_j) = \frac{1}{1 + \exp(-I_j)} \tag{2}$$

in which I_j can vary on the range $(-\infty, \infty)$, and y_j is bounded between zero and one. Because of the use of sigmoid functions in the CNN model, the values of the data variables must be normalized onto range $[0, 1]$ before applying the CNN methodology.

To determine the set of weights a corrective–repetitive process called 'learning' or 'training' of the CNN is performed. This training helps to define the interconnections among neurons (weights), and it is accomplished by using both known inputs and outputs (training sets or training patterns), and presenting these to the CNN in some ordered manner, adjusting the interconnection weights until the desired outputs are reached. The strength of these interconnections is adjusted using an error convergence technique so that a desired output will be produced for a given input. There are many training methods. In this work, a variation of back-propagation algorithm (Rumelhart et al., 1986), known as the Levenberg–Marquardt algorithm (Shepherd, 1997) was applied. This is a second-order non-linear optimization algorithm with very fast convergence and is recommended by several authors (Tan and van Cauwenberghe, 1999; Anctil and Rat, 2005).

Levenberg–Marquardt algorithm operates by making the assumption that the underlying function being modeled by the neural network is linear. Based on this assumption, the minimum of the objective function can be exactly determined in a single step. The computed minimum is tested, and if the error is lower than in the previous step, the algorithm moves the weights to the new point. This process is repeated iteratively on each generation. Since the linear assumption is ill-founded, it can easily lead Levenberg–Marquardt to test a point that is lower (perhaps substantial lower) than the current one. The most ingenious aspect of Levenberg–Marquardt is that the computation of the new point is actually a compromise between a step in the direction of steepest descent and the above-mentioned leap. Successful steps are accepted and lead to a strengthening of the linearity assumption (which is approximately true near a minimum). Unsuccessful steps are rejected and lead to a more cautious 'downhill' step. Thus, Levenberg–Marquardt continuously switches its approach and can make very rapid progress.

Levenberg–Marquardt algorithm uses the following formula that is continuously updated:

$$\Delta W = -(Z^{\mathrm{T}} Z + \lambda I)^{-1} Z^{\mathrm{T}} \varepsilon, \tag{3}$$

where ε is the vector of errors, Z is the matrix of the partial derivatives of these errors with respect to the weights W, and I is the identity matrix. The first term of the second member of the Levenberg–Marquardt formula represents the linear assumption and the second, the gradient-descent step. The control parameter λ governs the relative influence of these two approaches. Each time Levenberg–Marquardt succeeds in lowering the error, it decreases the control parameter by a factor of 10, thus strengthening the linear assumption and attempting to jump directly to the minimum. Each time it fails to lower the error, it increases the control parameter by a factor of 10, giving more influence to the gradient descent step, and also making the step size smaller.

Let epoch denote the time period that encompasses all the iterations performed after all the patterns are displayed. In the study presented in this paper, the learning process was controlled by the method of internal validation (about 20% of calibration data to test the error at the end of each epoch) (Tsoukalas and Uhrig, 1997; Gutiérrez-Estrada et al., 2004). The weights are updated at the end of each epoch. The number of epochs with the smallest error of the internal validation indicates the weights to select.

The numbers hidden layers and nodes in the hidden layers were determined by trial and error. CNNs with ranges of 1–2 hidden layers and 2–14 hidden nodes were successively trained based on the calibration data set. The CNN having the best performance when applied to the validation set, within a pool of 30 repetitions, was selected (Anctil and Rat, 2005). The CNN models were implemented using STATISTICA 6.0 (Statsoft, Inc., 1984–2002).

Data and general procedures (cases 1 and 2)

The CNN technique was used to develop a one-step daily flow forecast model for the watershed of the Tua River in the Douro Basin (north Portugal). Only daily flows were considered to be available for the model calibration and validation. Additionally, to evaluate the generalization capacity of the best CNN model calibrated in watershed of the Tua River, prediction of the dependent variable (daily streamflow) using data from other contrasting Portuguese watershed was carried out. This way, the extension of the validation process to other watershed were carried out.

Table 1 shows the general information of the two watersheds utilized in the study. Both watersheds of the Tua and Côa Rivers are tributaries of Douro River (north of Portugal).

Despite the two watersheds of Table 1 coincide with stream gauging stations (and therefore they have flow measurements) only the daily flows at Castanheiro were utilized to calibrate and to validate the CNN parameters, as well as to identify the daily flow forecasting procedure with the best performance. The model thus identified was then applied to the prediction of daily flows at Cidadelhe. The main goal of this 'transposition' procedure was to analyze if a CNN model with parameters established for a given watershed having long stream flow data series could also provide accurate estimates of daily flows in another watershed where the available data is insufficient to establish the CNN parameters.

The previous transposition procedure is supported by the studies extensively developed for the Portuguese watersheds (Quintela, 1967; Portela and Quintela, 2000, 2002a,b, 2005a,b, 2006). These studies showed that Portuguese watersheds having similar mean annual flow depths (the mean annual flow depth is equal to the mean annual flow expressed as water depth over the watershed) exhibit annual, monthly and daily flow regimes with similar relative temporal variability when expressed in a non-dimensional form, obtained by diving each flow by the mean annual flow. Based on this circumstance, the non-dimensional annual, monthly and daily flow series at a given watershed can be adopted at another watershed provided that the two watersheds have close mean annual flow depths. The proximity of the mean annual flows depths is the only restriction. The watersheds of Castanheiro and Cidadelhe are under these conditions circumstances because no significant differences were found between the respective mean annual flow depths (Portela and Quintela, 2000, 2002a,b, 2005a,b, 2006).

The transposition procedure from Castanheiro watershed to Cidadelhe watershed and the forecasting procedure applied is schematized in Fig. 1. Let μ denote the CNN based model having the best performance when applied to Castanheiro flow records and $[\mu]_{\text{Castanheiro}}$ the corresponding set of parameters.

To identify the model μ as well as to estimate the corresponding parameters $[\mu]_{\text{Castanheiro}}$, 38 years of daily flow records (from 1st October 1958 to 30th September 1996 – Table 1) at Castanheiro were utilized. This data was split into two sets, one for the CNN calibration or training – 30 years from 1st October of 1958 to 30th September of 1988 – and another for the CNN validation – 8 years from 1st October 1988 to 30th September 1996. To appreciate the

Table 1 General information for watersheds used in case studies

Parameter (1)	Tua River (2)	Côa River (3)
Stream gauging station	Castanheiro	Cidadelhe
Main watershed	Douro	Douro
Latitude	41°14′28″	40°54′49″
Longitude	7°23′25″	7°5′57″
Drainage area (km^2)	3718	1685
Mean discharge (m^3/s)	44.80	15.82
$\overline{H}$ (mm)[a]	366 ± 216	360 ± 177
Recording period	October 1958 to September 1996	October 1958 to September 1997

Stream flow records in Portuguese watersheds are available via Internet in http://snirh.inag.pt/ (*Sistema Nacional de Informação de Recursos Hídricos, SNIRH*). This public database was created in 1995 by the Portuguese water authority (*Instituto da Água*) and it provides information related with flows in more than 400 Portuguese stream gauging stations.

[a] Average of the annual flow expressed as water depth over the watershed (mean annual flow depth) and corresponding standard deviation (Portela and Quintela, 2000, 2002a,b, 2005a,b).

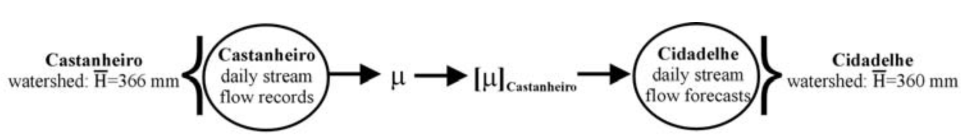

Figure 1 Case studies: (a) Calibration and validation with Castanheiro stream flow series; and (b) Calibration with Castanheiro stream flow series and validation with Cidadelhe stream flow series.

adequacy of model μ at Cidadelhe 39 years of daily flows records were utilized (from 1st October 1958 to 30th September 1997 – Table 1).

For those approaches based on a typical application of CNNs only daily flows in the previous days were considered as input variables. Under these circumstances, the output variable in each day t was the flow in that day and the input variables were the flows in the previous days ($t-1$, $t-2$, etc.) (case 1). The number of previous days to be considered in the CNN model was fixed based on the analysis of the autocorrelation and partial autocorrelation of the daily flows at Castanheiro. CNNs having as inputs flows in previous δ days plus differenced data also in previous days ($\mathrm{DIFF}_{t-1} = Q_{t-1} - Q_{t-2}, \ldots, \mathrm{DIFF}_{t-\delta-1} = Q_{t-\delta-1} - Q_{t-\delta-2}$) were also tested (case 2).

The partial autocorrelation showed three significant peaks related with the three previous days. This denotes a third-order autoregressive process, according to which the flow in each day may be considered a function of its own past values in the three previous days (Wilson and Keating, 1996). However, delays until fifteen antecedent days were tested to see if a partial autocorrelation of higher order could improve the performance of the model during the validation phase.

Beside CNNs having as input data flows in previous days (case 1) or those flows plus differenced data (case 2) others approaches were tested and compared, as presented in the following sections.

Measures of accuracy (validation phase)

To assess the performance of the CNN during the validation phase and therefore to identify the model μ of Fig. 1, several measures of accuracy were applied, as there is not a unique and more suitable performance evaluation test (Yapo et al., 1998; Legates and McCabe, 1999; Abrahart and See, 2000). The correlation between observed and predicted flows was expressed by means of the correlation coefficient R. The coefficient of determination (R^2) describes the proportion of the total variance in the observed data that can be explained by the model. Others measures of variances applied were the percent standard error of prediction (%SEP) (Ventura et al., 1995), the coefficient of efficiency (E_j) (Nash and Sutcliffe, 1970; Kitanidis and Bras, 1980) and the average relative variance (ARV) (Griñó, 1992). These four estimators are unbiased estimators that are employed to see how far the model is able to explain the total variance of the data.

In addition, it is advisable to quantify the error in the same units of the variables. These measures, or absolute error measures, included the square root of the mean square error (RMSE) and the mean absolute error (MAE), given by

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Q_t - \widehat{Q}_t)^2}{N}} = \sqrt{\mathrm{MSE}}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}\left|Q_t - \hat{Q}_t\right|}{N}, \tag{4}$$

where Q_t is the observed stream flow at the time step t; $\widehat{Q}_t$ is the estimated stream flow at the same time step t; and N is the total number of observations of the validation set.

The percent standard error of prediction, %SEP, is defined by

$$\%\mathrm{SEP} = \frac{100}{\overline{Q}}\ \mathrm{RMSE}, \tag{5}$$

where $\bar{Q}$ is the average of the observed stream flows of the validation set. The principal advantage of %SEP is its non-dimensionality that allows comparing in a same basis forecasts given by different models. The coefficient of efficiency E_j and the average relative variance ARV are used to see how the model explains the total variance of the data and represent the 'proportion' of the variation of the observed data considered by the model. E_j and ARV are given by

$$E_j = 1.0 - \frac{\sum_{i=1}^{N}\left|Q_t - \widehat{Q}_t\right|^j}{\sum_{i=1}^{N}\left|Q_t - \overline{Q}\right|^j}$$

$$\mathrm{ARV} = \frac{\sum_{i=1}^{N}(Q_t - \widehat{Q}_t)^2}{\sum_{i=1}^{N}(Q_t - \overline{Q})^2} = 1.0 - E_2. \tag{6}$$

The sensitivity to outliers due to the squaring of the difference terms is associated with E_2 or, equivalently, with ARV. E_1 (named here the modified coefficient of efficiency) reduces the effect of the squared terms. A value of zero for E_2 indicates that the observed average $\overline{Q}$ is as good as predictor as the model, while negative values indicate that the observed average is a better predictor than the model (Legates and McCabe, 1999).

For a perfect match, the values of R^2 and of E_j should be close to one and those of %SEP and ARV close to zero.

Also the persistence index, PI, was used for the model performance evaluation (Kitanidis and Bras, 1980):

$$\mathrm{PI} = 1 - \frac{\sum_{i=1}^{N}(Q_t - \widehat{Q}_t)^2}{\sum_{i=1}^{N}(Q_t - Q_{t-L})^2}, \tag{7}$$

where Q_{t-L} is the observed flow at the time step $t-L$ and L is the lead-time. In the applications carried out L was set equal to one, since only one-day ahead forecasts were performed. A PI value of one reflects a perfect adjustment between predicted and observed values, and a value of zero is equivalent to say that the model is no better than a 'naïve' model, which always gives as prediction the previous observation. The PI is well designed for assessing forecasts, as the stream flows in previous days are the main CNN inputs. A negative PI value would mean that the model is degrading the original information, thus denoting a performance worse than the one of the 'naïve' model (Anctil and Rat, 2005).

Other indexes used to identify the best linear and nonlinear models were the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Qi and Zhang, 2001) given by

$$\mathrm{AIC} = \log(\mathrm{MSE}) + \frac{2m}{N} \quad \mathrm{BIC} = \log(\mathrm{MSE}) + \frac{m\log(N)}{N}. \tag{8}$$

In the previous equations m is the number of parameters of the model. In those equations both first terms of the second members measure the goodness-of-fit of the model to the data while the second terms set a penalty for the model over-parameterization. The optimal model is selected when AIC and BIC are minimized.

Computational neural network with intervention series (case 3)

In a CNN with intervention series the input layer comprehends two classes of input variables: the stream flows in previous days ($t-1$, $t-2$, etc.) and auxiliary inputs. The first class of inputs is responsible for modeling the autoregressive structure of the series. Auxiliary inputs are used to apply intervention series to the CNN model. The intervention series contain additional information about the characteristics of the day in which the stream flow is predicted. That information may consist of four binary [0,1] intervention series each series corresponding to a season of the year (spring, summer, autumn or winter) (case 3 – see Section *Data and general procedures (cases 1 and 2)*). So that these input variables intended to allow one variation in CNN output according to a hypothetical seasonal influences. In this approach, the best CNN is also determined by trial and error. The tested ranges of hidden layers and nodes in the hidden layers were the same of cases 1 and 2.

Convolution process of the input variables to computational neural network (case 4)

The multilayer feed forward network is a static CNN architecture. To assess the temporal patterns, the CNN must have an appropriate memory that stores past information. The simplest form of memory consists of a buffer that contains the δ most recent inputs, that is to say, that contains multiple copies of the input data at various time delays δ (the stream flows in previous days $t-1$, $t-2$, etc., as defined in Section *Data and general procedures (cases 1 and 2)* as case 1). Other common methodology is to represent the memory as a convolution of the input sequence of daily flows x_i with a kernel or smooth function (De Vries and Principe, 1991). In the analysis carried out, both methodologies – buffer (case 1)/smooth function (case 4) – were implemented.

Taking into account that the correlation coefficient between consecutive daily flows measures the dependency between those flows, being a sort of memory of the systems, the following triangular smooth function was developed and tested (case 4):

$$\bar{x}_i = \frac{\sum_{\tau=-\delta}^{\delta} R_{t/t-\tau} x_{i-\tau}}{\sum_{\tau=-\delta}^{\delta} R_{t/t-\tau}} \qquad \forall i = \delta \text{ to } M-\delta, \tag{9}$$

where M is the total number of observations of the training data set, and $R_{t/t-\tau}$ is the coefficient of correlation between data series at time t and time $t-\tau$ ($\tau = -\delta, \ldots, 1, \ldots, \delta$). The $\bar{x}_i$ series thus achieved (Fig. 2) are supplied to the model as input data. This approach was combined with the best CNN determined with the previous trainings.

Hybrid neural network and ARIMA model (case 5)

ARIMA models and CNNs are often compared with no clear conclusions in terms of the relative forecasting performance superiority (Zhang, 2003). The ARIMA models assume that a time series is a linear combination of its own past values and of current and past values of an error term (Box and Jenkins, 1976). In this paper, a hybrid approach to time series forecasting using both CNN and ARIMA models was analyzed. The main idea of the combination of those two models was to use each model feature to capture different patterns in the data. The methodology implemented consists of two steps: (a) in the first step, a CNN is developed to model the stream flows (the best CNN determined with the training of the cases 1, 2 or 3); and (b) in the second step, an ARIMA model is used to describe the residuals from the CNN model (Fig. 3). The ARIMA model helps to interpret the unexplained variance by the CNN model.

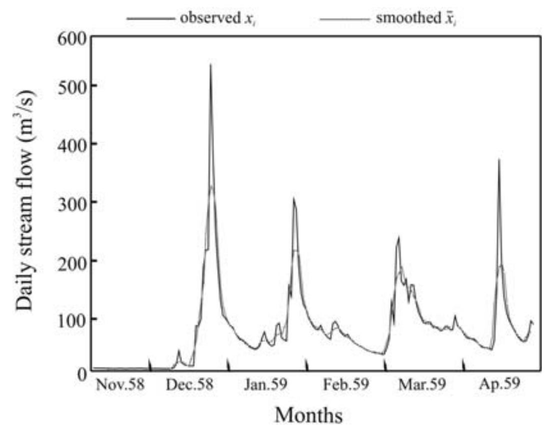

Figure 2 Convolution of the input sequence of daily flows x_i with a smooth function $\bar{x}_i$ from November 58 to April 59.

The CNN and ARIMA combined model can be formulated as follows:

$$Q_t = f(Q_{t-1}, Q_{t-2}, \ldots, Q_{t-\delta}) + \phi^{-1}(B)\theta(B)\eta_t, \tag{10}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \tag{11}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \tag{12}$$

where f is a function determined by the neural network structure and connection weights; ϕ_j ($j = 1, \ldots, p$) are the weights of the ARIMA model associated with each previous observation ε_t (in the applications carried out, the residuals from the CNN model); θ_j ($j = 1, \ldots, q$) are the weights of the ARIMA model associated with each previous noise terms; B is the backshift operator that assigns a value to a variable in the previous instant ($B\varepsilon_t = \varepsilon_{t-1}$ and $B^p\varepsilon_t = \varepsilon_{t-p}$); and η_t is the noise term in instant t. Differencing transformation

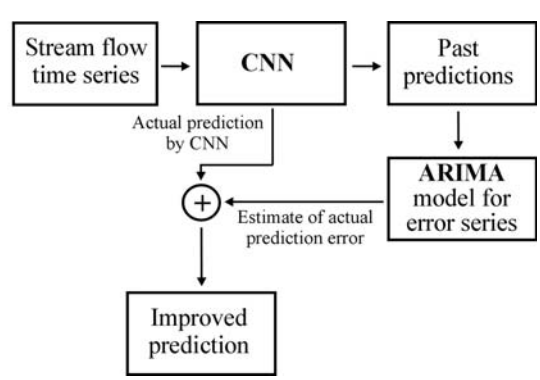

Figure 3 Schematic structure of hybrid neural network with modelling ARIMA of prediction residuals.

(d) is often applied to the data to remove the trend and to stabilize the variance before an ARIMA model can be fitted.

In the ARIMA (p,d,q) model, values of p and q varying from zero to five (with a unitary step), and values of d varying from zero to two (also with a unitary step) were tested. The values of p, d, and q that proved to be more appropriate according to the accuracy measures presented in Section *Measures of accuracy (validation phase)* were then used (case 5). The parameters ϕ_j and θ_j were fixed by using function minimization procedures, so that the sum of squared residuals was minimized. The level of significance of these parameters should be evaluated (acceptable if $P_\alpha < 0.05$). This approach was combined with the best CNN determined with the previous trainings (cases 1, 2 and 3).

Summary of the implemented approaches

As result of the procedures presented in the previous sections, the following five approaches were implemented based on Castanheiro daily flow series and compared in terms of their relative accuracy in order to identify the model with the best performance (model μ of Fig. 1):

case 1 flows in the previous days as input variables (typical approach — Section *Data and general procedures (cases 1 and 2)*);
case 2 flow in the previous days and differenced data in previous days as input variables (typical approach plus differenced series — Section *Data and general procedures (cases 1 and 2)*);
case 3 flow in the previous days and auxiliary data as input variables (binary seasonal intervention series — Section *Computational neural network with intervention series (case 3)*);
case 4 convolution process of the input variables (smooth function, Eq. (9), Section *Convolution process of the input variables to computational neural network (case 4)*); and
case 5 hybrid neural network and ARIMA model (Section *Hybrid neural network and ARIMA model (case 5)*).

Results

Castanheiro stream gauging station. selection of the best model μ

Cases 1—5 were tested based on Castanheiro daily flow records in order to identify the CNN model with the best performance (model μ of Fig. 1).

For cases 1—3 the one-day ahead daily flow forecasts thus achieved during the validation period (8 years, from 1st October 1988 to 30th September 1996) and within a pool of 30 repetitions of each CNN model showed that:

case 1 had the best estimates when the flows in the three previous days were used as input data in a CNN with 4 nodes in the first and second hidden layers [CNN (3,4,4,1)] and 200 epochs in the training process.
case 2 the best results were obtained when the flows of the three previous days and the differenced data in previous days ($\mathrm{DIFF}_{t-1} = Q_{t-1} - Q_{t-2}$, $\mathrm{DIFF}_{t-2} = Q_{t-2} - Q_{t-3}$) were used as inputs in a CNN with 4 nodes in the two hidden layers [CNN (5,4,4,1)] and 250 epochs in the training process.
case 3 the CNN (7,5,5,1) having as input data the flows in the three previous days and four binary intervention series (each series corresponding to a season of the year) led to the best daily forecasts with 300 epochs in the training process.

The previous conclusions were supported by the values provided by the accuracy measures presented in Section *Measures of accuracy (validation phase)*. For the best CNN model of each of the cases 1—3, Fig. 4 shows those values in terms of their averages and of their standard deviations, after 30 repetitions of the CNN model (validation period).

Fig. 4 shows for the cases 1—3 two sets of contradictory results: some of the accuracy measures (namely R^2, E_1 and E_2) point towards the good performance of most of the CNNs while other measures (MAE, RMSE, %SEP and PI) suggest the opposite (poor performance of most of those models). A MAE of more than a third of the average of the observed flows in validation period (that is to say, higher than $33.28/3 \approx 11\ \mathrm{m^3/s}$) indicates mediocre models.

A detailed analysis of the results showed that the poor performance of the cases 1—3 was mainly due to the occurrence of a kind of systematic 'displacement' between estimated and observed flows, as exemplified on Fig. 4 (values of PI around zero or even negatives and high values of MAE, RMSE and %SEP) and Fig. 5 (case 1). In fact, cases 1—3 systematically led to predictions in day t very close to the observed flows in day $t-1$, as is shown in the schematic representation of the forecasted flows at time t as a function of the observed flows at time $t-1$ (Fig. 6).

In terms of accuracy measures, the differences among cases 1—3 prove not to be significant (one-way analysis of variance or ANOVA — R^2: $F = 0.229$, $P_\alpha = 0.797$; RMSE and %SEP: $F = 0.205$, $P_\alpha = 0.816$; MAE and E_1: $F = 0.142$, $P_\alpha = 0.869$; E_2, ARV and PI: $F = 0.224$, $P_\alpha = 0.800$; AIC: $F = 0.646$, $P_\alpha = 0.532$; BIC: $F = 1.143$, $P_\alpha = 0.334$). Case 1 presented the best averages and the smallest standard deviations for all accuracy measures. Cases 2 and 3 had averages very similar except for AIC and BIC measures. This was due to the higher number, m, of parameters of case 3 (case 1: $m = 32$; case 2: $m = 40$; case 3: $m = 65$). However, case 2 had the highest standard deviations for all accuracy measures.

When a convolution process of the input variables (case 4) and a hybrid neural network and ARIMA model (case 5) were combined with the best CNN according to case 1, better performances were achieved during the validation phase, especially for case 4 (Fig. 4). In fact, no significant differences were found between accuracy measures of cases, on one hand, 1, 2 and 3 and, on the other hand, 5, except for R^2 (ANOVA — R^2: $F = 8.206$, $P_\alpha < 0.001$; RMSE and %SEP: $F = 0.538$, $P_\alpha = 0.660$; MAE and E_1: $F = 0.303$, $P_\alpha = 0.823$; E_2, ARV and PI: $F = 0.511$, $P_\alpha = 0.677$; AIC: $F = 1.056$, $P_\alpha = 0.380$; BIC: $F = 1.545$, $P_\alpha = 0.220$).

However, case 4 showed a better performance when compared with other cases, with significant improvements

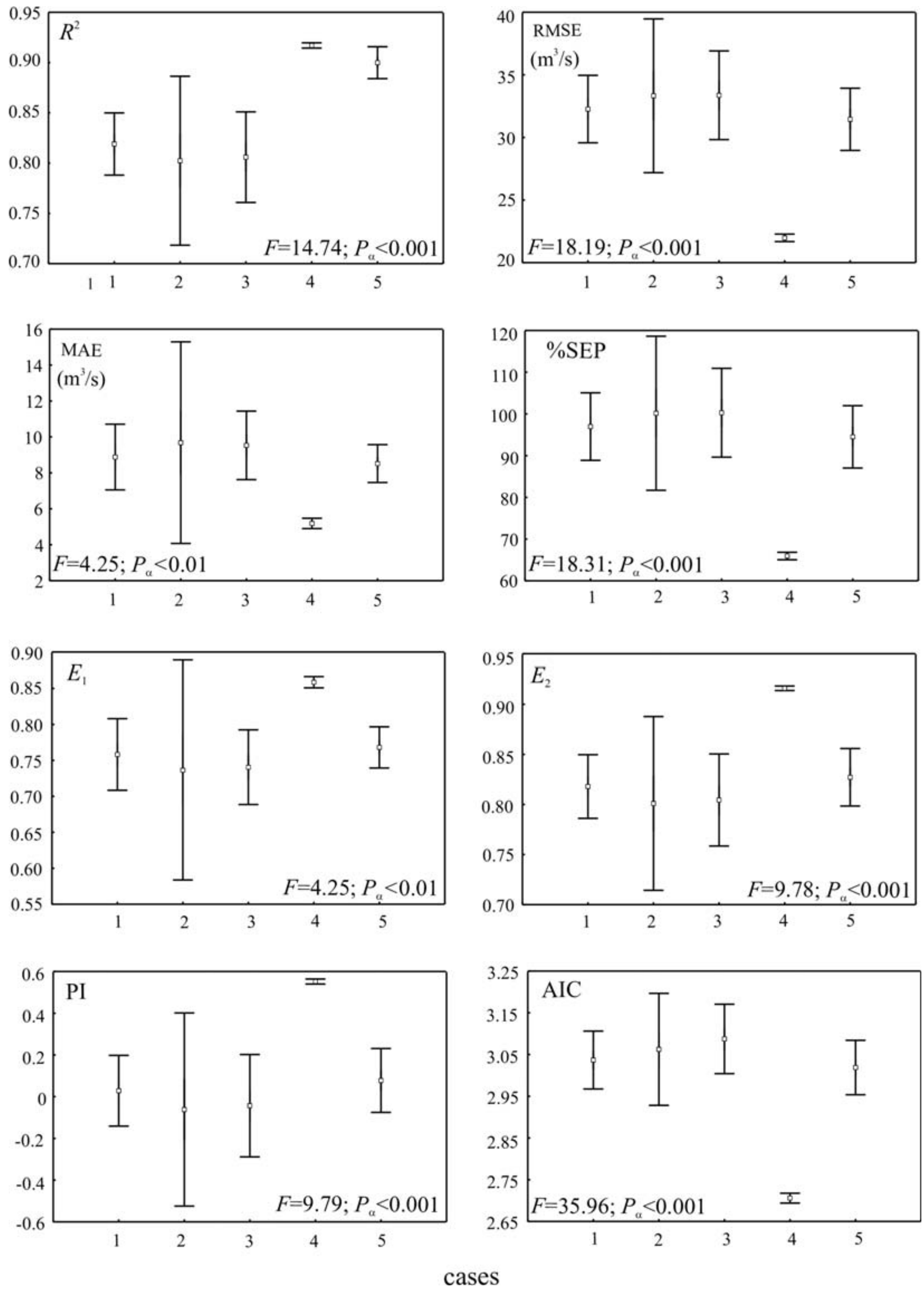

Figure 4 Means and standard deviations of results of accuracy measures (validation period, Castanheiro station).

in terms of accuracy measures (Fig. 4). In all instances, case 4 led to PI values greater than 0.5 and MAE values smaller than 6 m^3/s. These measures mean that the flow in each day resulting from case 4 did not necessarily approach the flow in the previous day, as happened in the other cases (cases 1—3). That is to say, case 4 proved to be able to eliminate the systematic 'displacement' between observed and forecasted flows. It should be pointed that the best repetitions of the pools of cases 4 and 5 did not coincide with those of case 1.

In case 5, the best estimates were obtained by applying an ARIMA(5,0,1) to the residuals of the CNN model. By this way the smallest error magnitudes, parameters ϕ_j and θ_j with a level of acceptable statistical significance ($P_\alpha < 0.05$), and autocorrelation coefficients of the prediction errors statistically acceptable (all values of residual

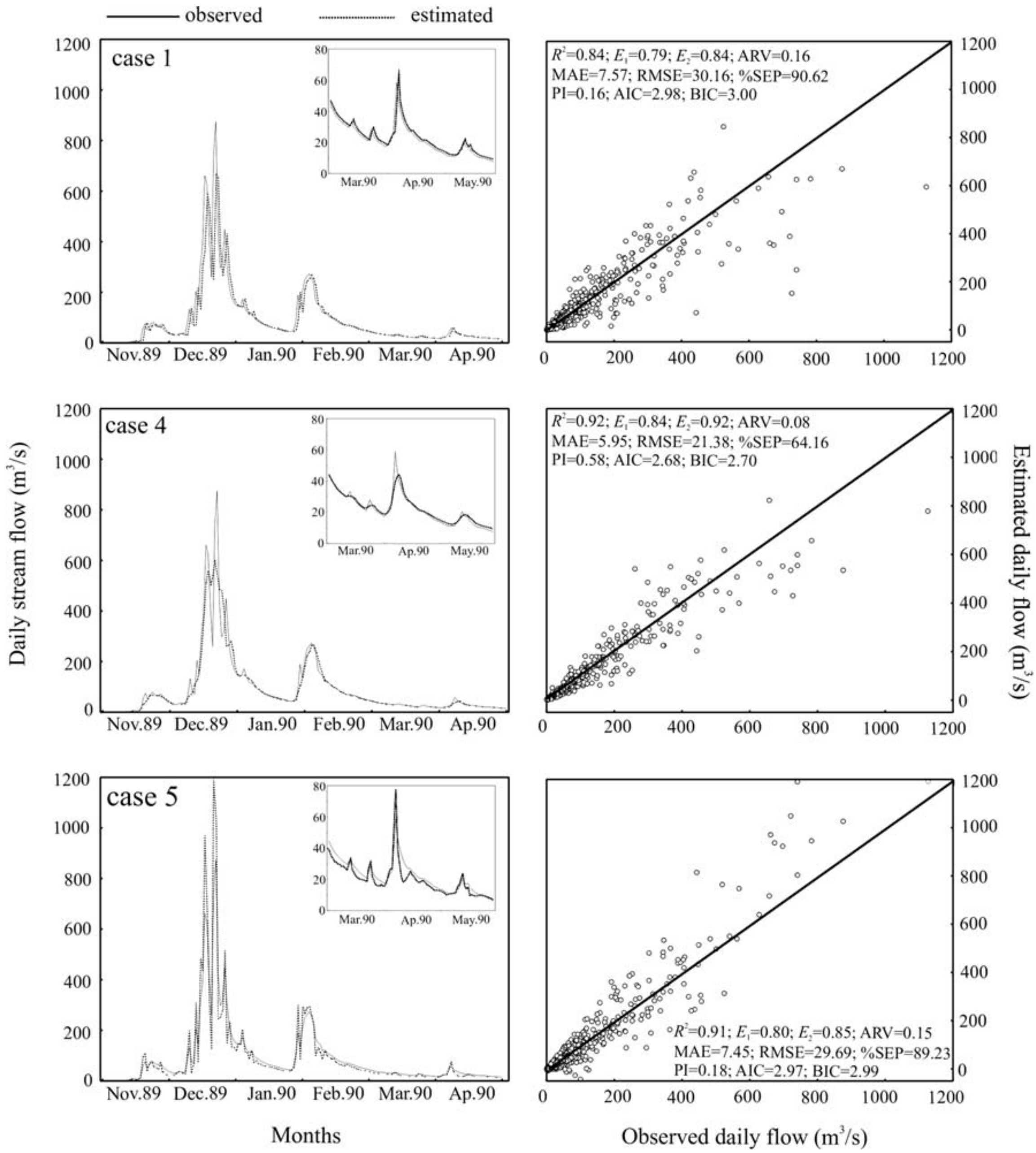

Figure 5 Exemplification of one-step-ahead prediction of daily stream flows for the period from November 89 to April 90, and scatterplot between observed and forecasted flows, for best repetition of cases 1, 4 and 5 (validation period, Castanheiro station).

correlations were close to zero and within the confidence limit to 95%; also there was not serial dependency between residuals, Wilson and Keating, 1996) were achieved. The parameters of the ARIMA model of the repetition with the best results were $\phi_1 = 0.4783^a$; $\phi_2 = -0.1516^a$; $\phi_3 = 0.0746^a$; $\phi_4 = 0.0495^a$; $\phi_5 = 0.0789^a$; $\theta_1 = 0.4014^a$; a = parameters ϕ_j and θ_j with a level of acceptable statistical significance $P_\alpha < 0.05$ (Eqs. (10)—(12)).

The schematic representation of the forecasted flows as a function of the observed flows (scatterplots between observed and forecasted flows) showed that cases 1—3 presented the highest dispersion along the line 1:1 (that would correspond to the perfect adjustment between the two previous kinds of flows), while cases 4 and 5 denoted a higher approximation of observed and estimated values. Scatterplots for the best repetition of the pools for cases 1, 4 and 5 are presented in Fig. 5 that also includes, as examples, diagrams of the observed and forecasted flows from November 89 to April 90. This figure shows that case 4 presents the closest match between forecasted and observed flows over the entire flow range (points with smaller dispersion along line 1:1). Notice the tendency of cases 1 and 4 to underestimate the high flows and case 5 to overestimate those flows.

The averages (in the validation period) of the observed daily flows (33.28 m^3/s) and of the forecasted daily flows are very close, that is to say, without statistically significant differences (Fig. 7). The comparison of the standard

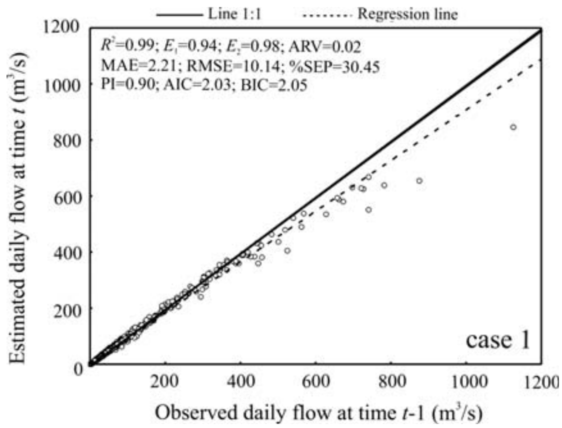

Figure 6 Scatterplot between observed flows at time $t-1$ and estimated flows at time t for best repetition case 1 (validation period, Castanheiro station).

deviation of the observed daily flows ('observed' standard deviation with the value of 78.86 m^3/s) with the standard deviations of the forecasted daily flows showed that, due to the previously mentioned overestimation and underestimation of stream flows, case 5 led to standard deviations that were almost 20 m^3/s higher than the 'observed' one while the standard deviations achieved in cases 1—4 were almost 20 m^3/s smaller than the same (Fig. 7).

The performance of the different cases is also shown in Table 2 that contains the accuracy measures for the entire range of daily flows as well as those measures depicted by successive ranges of daily flows. The averages of the observed and forecasted daily flows during the validation period and standard deviations of the former daily flows were also included in Table 2.

Table 2 shows that the best adjustment between observed and forecasted daily flows (best results of the accuracy measures) mainly occur in the lowest range from 0 to 50 m^3/s (in which about 80% of the observed flow are included). In general the error magnitudes are quite acceptable for the entire range of daily flows. In relative terms, those magnitudes become worse as the flow range is higher, being much worse for the highest flow range. The 'deterioration' of the quality of the different models as the flow ranges become higher is due to the decrease of the number of flows belonging to the successive ranges. By other words, due to the progressive smaller number of flows utilized during the calibration phase, that number becoming insufficient to allow the model to 'learn' from the available observations. All cases showed significance differences among accuracy measures in different flow ranges (for example, Kruskal—Wallis test in case 4 gave the following results for R^2, RMSE, MAE, AIC and BIC: $H(3,120) = 36.585$, $P_\alpha < 0.001$; %SEP: $H(3,120) = 32.933$, $P_\alpha < 0.001$; E_1: $H(3,120) = 35.633$, $P_\alpha < 0.001$; E_2 and ARV: $H(3,120) = 30.786$, $P_\alpha < 0.001$; and PI: $H(3,120) = 29.706$, $P_\alpha < 0.001$).

With exception of case 4, the values of PI are most of the time around zero or negative denoting model inadequacy, mainly due to the occurrence of systematic 'displacements' between observed and forecasted flows (Figs. 5 and 6, case 1). In case 4, those values are quite good, though denoting a less good quality in the range from 0 to 50 m^3/s also due to some 'displacement' between observed and estimated flows. Case 4 estimated less extreme values than case 5, which implies better fittings.

Globally, the results achieved, summarized in Figs. 4 and 7 and in Table 2, clearly indicate case 4 as the one, not only with the far best performance when compared with the other cases, but also with an excellent performance by itself, as a tool to forecast daily flows at Castanheiro. Therefore, case 4 was adopted as the model $\boldsymbol{\mu}$ of Fig. 1.

Validation of case 4 at Cidadelhe watershed

To check the capability of case 4 to estimate one-day ahead daily flows at Cidadalhe, the approach with the set of parameters previously established for Castanheiro ($[\boldsymbol{\mu}]_{\text{Castanheiro}}$) was applied to daily flow forecasting at Cidadelhe, based on the 39-year period, from 1st October 1958 to 30th September 1997. As pointed earlier (Section *Data and general procedures (cases 1 and 2)*), this transposition/regionalization procedure is justified by the similarity of the average annual flow depths at the two stream gauging stations under consideration.

The validation of the transposition of case 4 was accomplished in a way equivalent to the one adopted for Castanheiro, based on the accuracy measures presented in Section *Measures of accuracy (validation phase)*. It should be pointed out that the main target of the procedure was

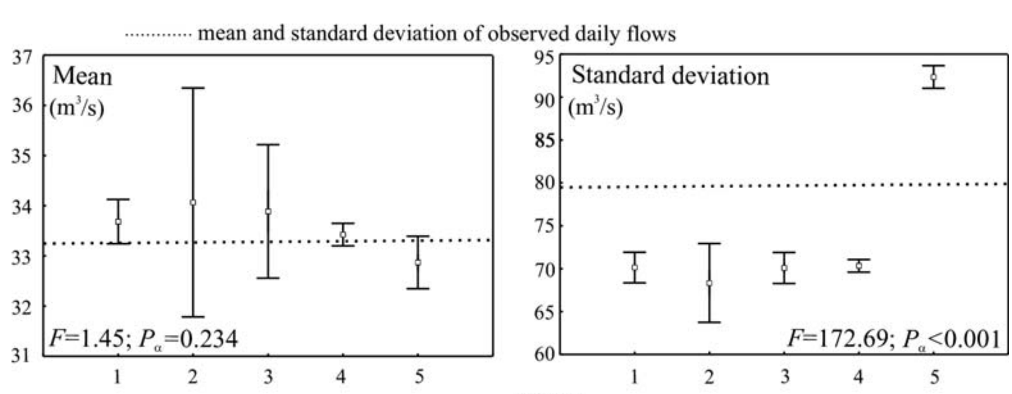

Figure 7 Comparison of observed and forecasted means and standard deviations of flows (validation period, Castanheiro station).

Table 2 Accuracy measures and averages of the forecasted flows for the complete range of flows and depicted by continuous ranges of flows (validation period)

Case (1)	Range[a] (m^3/s) (2)	R^2 (3)	RMSE (m^3/s) (4)	MAE (m^3/s) (5)	%SEP (6)	E_1 (7)	E_2 (8)	ARV (9)	PI (10)	AIC (11)	BIC (12)	Average of the forecasted flows (m^3/s) (13)
1	All data	0.82 ± 0.03	32.26 ± 2.69	8.88 ± 1.83	96.96 ± **8.10**	0.76 ± 0.05	0.82 ± 0.03	0.18 ± 0.03	0.03 ± 0.17	3.04 ± 0.07	3.05 ± 0.07	33.68 ± 0.44
	[0,50)	0.90 ± 0.08	4.48 ± 1.33	3.00 ± 1.40	37.73 ± 11.18	0.87 ± 0.06	0.96 ± 0.02	0.04 ± 0.02	−0.94 ± 1.37	1.30 ± 0.21	1.32 ± 0.21	13.25 ± 0.40
	[50,150)	0.52 ± 0.02	26.73 ± 4.69	17.57 ± 4.39	35.05 ± 6.15	0.59 ± 0.10	0.69 ± 0.13	0.31 ± 0.13	− 0.28 ± 0.53	3.01 ± 0.13	3.06 ± 0.13	81.67 ± 2.64
	[150,450)	0.41 ± 0.05	96.43 ± 4.45	72.17 ± 3.63	36.87 ± 1.70	0.68 ± 0.02	0.84 ± 0.01	0.16 ± 0.01	0.20 ± 0.07	4.61 ± 0.04	4.61 ± 0.04	252.20 ± 6.44
	⩾450	0.004 ± 0.005	290.69 ± 35.56	230.48 ± 30.96	44.81 ± 5.48	0.63 ± 0.05	0.79 ± 0.06	0.21 ± 0.06	−0.04 ± 0.27	7.97 ± 0.10	6.94 ± 0.10	493.35 ± 20.57
2	All data	0.80 ± 0.08	33.33 ± 6.14	9.68 ± 5.61	100.16 ± 18.46	0.74 ± 0.15	0.80 ± 0.09	0.20 ± 0.09	−0.06 ± 0.46	3.06 ± 0.13	3.08 ± 0.13	34.06 ± 2.28
	[0,50)	0.89 ± 0.14	5.35 ± 4.89	3.75 ± 4.90	45.08 ± 41.18	0.84 ± 0.21	0.92 ± 0.18	0.08 ± 0.18	−3.49 ± 10.07	1.33 ± 0.45	1.36 ± 0.45	14.48 ± 4.98
	[50,150)	0.47 ± 0.13	29.06 ± 9.08	17.64 ± 6.42	38.11 ± 11.90	0.59 ± 0.15	0.62 ± 0.30	0.38 ± 0.30	−0.60 ± 1.26	3.11 ± 0.21	3.17 ± 0.21	78.89 ± 7.26
	[150,450)	0.38 ± 0.07	100.32 ± 14.11	76.18 ± 14.47	38.36 ± 5.39	0.67 ± 0.06	0.83 ± 0.05	0.17 ± 0.05	0.12 ± 0.27	4.80 ± 0.11	4.80 ± 0.11	246.87 ± 22.58
	⩾450	0.003 ± 0.004	294.14 ± 42.64	235.56 ± 39.62	45.35 ± 6.57	0.62 ± 0.06	0.78 ± 0.07	0.22 ± 0.07	−0.07 ± 0.35	8.74 ± 0.11	7.45 ± 0.11	479.31 ± 39.54
3	All data	0.81 ± 0.05	33.37 ± 3.54	9.53 ± 1.91	100.29 ± 10.64	0.74 ± 0.05	0.80 ± 0.05	0.20 ± 0.05	−0.04 ± 0.25	3.09 ± 0.08	3.12 ± 0.08	33.89 ± 1.33
	[0,50)	0.86 ± 0.14	5.65 ± 2.44	3.73 ± 1.60	47.62 ± 20.54	0.84 ± 0.07	0.94 ± 0.07	0.06 ± 0.07	−2.34 ± 3.64	1.51 ± 0.28	1.55 ± 0.28	13.75 ± 2.17
	[50,150)	0.48 ± 0.05	26.41 ± 1.48	16.53 ± 1.37	34.63 ± 1.95	0.62 ± 0.03	0.71 ± 0.03	0.29 ± 0.03	−0.22 ± 0.14	3.19 ± 0.05	3.29 ± 0.05	80.75 ± 3.21
	[150,450)	0.38 ± 0.08	104.30 ± 15.21	75.32 ± 8.75	39.88 ± 5.82	0.67 ± 0.04	0.81 ± 0.06	0.19 ± 0.06	0.05 ± 0.31	5.33 ± 0.11	5.33 ± 0.11	249.90 ± 10.31
	⩾450	0.004 ± 0.006	293.83 ± 21.99	240.49 ± 21.32	45.30 ± 3.39	0.61 ± 0.03	0.78 ± 0.03	0.22 ± 0.03	−0.05 ± 0.17	11.12 ± 0.06	9.03 ± 0.06	491.81 ± 24.95
4	All data	0.92 ± 0.003	21.98 ± 0.30	5.19 ± 0.29	65.95 ± 0.90	0.86 ± 0.01	0.92 ± 0.002	0.08 ± 0.002	0.55 ± 0.01	2.71 ± 0.01	2.72 ± 0.01	33.42 ± 0.22
	[0,50)	0.96 ± 0.01	2.86 ± 0.11	1.27 ± 0.33	24.05 ± 0.95	0.94 ± 0.01	0.99 ± 0.001	0.01 ± 0.001	0.27 ± 0.06	0.94 ± 0.03	0.96 ± 0.03	12.58 ± 0.34
	[50,150)	0.71 ± 0.01	15.66 ± 0.57	9.01 ± 0.34	20.54 ± 0.75	0.79 ± 0.01	0.90 ± 0.01	0.10 ± 0.01	0.57 ± 0.03	2.56 ± 0.03	2.61 ± 0.03	78.63 ± 1.24
	[150,450)	0.54 ± 0.01	75.14 ± 2.64	53.91 ± 2.24	28.73 ± 1.01	0.76 ± 0.01	0.90 ± 0.01	0.10 ± 0.01	0.52 ± 0.03	4.39 ± 0.03	4.39 ± 0.03	264.58 ± 1.84
	⩾450	0.24 ± 0.03	186.79 ± 5.86	157.41 ± 5.39	28.80 ± 0.90	0.74 ± 0.01	0.91 ± 0.01	0.09 ± 0.01	0.58 ± 0.03	7.59 ± 0.03	6.56 ± 0.03	524.97 ± 9.06
5	All data	0.90 ± 0.02	31.45 ± 2.48	8.52 ± 1.05	94.50 ± 7.47	0.77 ± 0.03	0.83 ± 0.03	0.17 ± 0.03	0.08 ± 0.15	3.02 ± 0.07	3.04 ± 0.07	32.87 ± 0.52
	[0,50)	0.92 ± 0.02	4.26 ± 0.64	2.72 ± 0.62	35.92 ± 5.40	0.88 ± 0.03	0.97 ± 0.01	0.03 ± 0.01	−0.66 ± 0.53	1.28 ± 0.12	1.30 ± 0.12	10.60 ± 0.55
	[50,150)	0.40 ± 0.12	26.71 ± 4.34	17.75 ± 3.86	35.03 ± 5.69	0.59 ± 0.09	0.70 ± 0.11	0.30 ± 0.11	−0.27 ± 0.48	3.05 ± 0.12	3.11 ± 0.12	71.31 ± 2.13
	[150,450)	0.49 ± 0.06	92.85 ± 3.65	69.56 ± 2.51	35.51 ± 1.40	0.70 ± 0.01	0.85 ± 0.01	0.15 ± 0.01	0.26 ± 0.06	4.70 ± 0.03	4.70 ± 0.03	266.86 ± 5.46
	⩾450	0.74 ± 0.08	284.46 ± 33.49	221.56 ± 24.20	43.85 ± 5.16	0.64 ± 0.04	0.80 ± 0.05	0.20 ± 0.05	0.01 ± 0.26	8.52 ± 0.09	7.30 ± 0.09	801.03 ± 10.42

[a] For each flow range, number of records, N, average (m^3/s) ± standard deviation (m^3/s) of the observed daily flow (validation period): [0,50): $N = 2423, \overline{Q} = 11.87 \pm 12.51$; [50,150): $N = 375$, $\overline{Q} = 76.25 \pm 23.65$; [150,450): $N = 100$, $\overline{Q} = 261.51 \pm 85.88$; ⩾450: $N = 21$, $\overline{Q} = 648.67 \pm 159.67$.

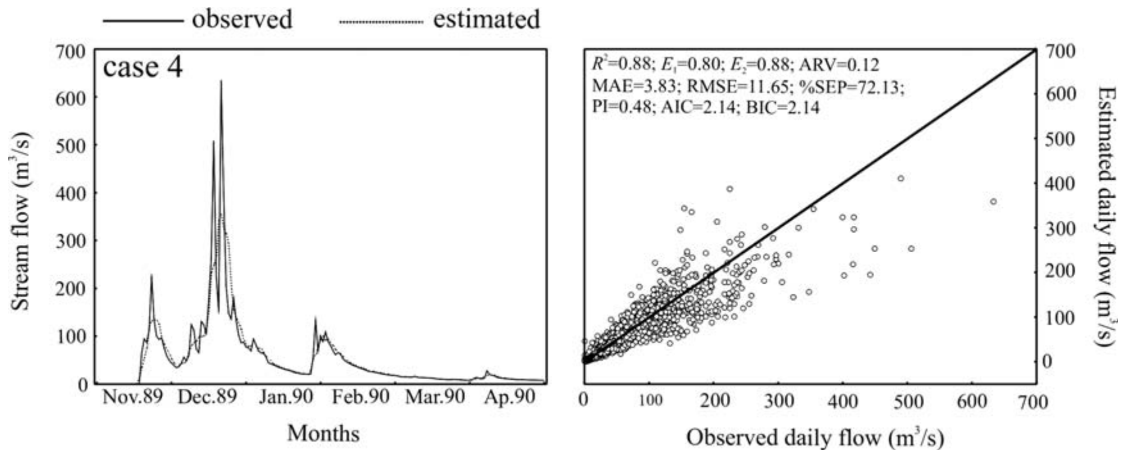

Figure 8 Exemplification of one-step-ahead prediction of daily stream flows for the period from November 89 to April 90, and scatterplot between observed and forecasted flows, for validation period of Cidadelhe station for best repetition using $[\mu]_{\text{Castanheiro}}$.

to obtain one-day ahead flow forecasts at Cidadelhe. Hence, once a daily flow was estimated and compared with the observed one the application of the model pursuit by replacing the forecasted flow by the observed one.

The accuracy measures thus achieved for Cidadelhe were the following ones: $R^2 = 0.86 \pm 0.03$; RMSE = 12.90 ± 1.51 m^3/s; MAE = 5.08 ± 1.84 m^3/s; %SEP = 79.85 ± 9.32; E_1 = 0.73 ± 0.10; E_2 = 0.85 ± 0.04; ARV = 0.15 ± 0.04; PI = 0.35 ± 0.16; AIC = 2.22 ± 0.09; BIC = 2.23 ± 0.09. The averages ± standard deviations of observed and of the estimated daily flows were 16.15 ± 33.89 and 17.74 ± 31.37 m^3/s, respectively.

In Fig. 8 forecasted daily flows versus observed daily flows (scatterplot) for the CNN with the best performance of the pool of 30 repetitions are shown. The figure also includes, as example, a diagram of the observed and forecasted flows from November 89 to April 90.

Discussion

The adequacy of computational neural network models for daily stream flow forecasting at large Portuguese watersheds was analyzed in this paper. The best correlation and error statistics were obtained based on a CNN having as input data the flows in the three previous days combined with a convolution process of the input sequence by means of a triangular smooth function — case 4, Eq. (9). The results thus achieved were better than those obtained with typical autoregressive training based only on flow-input data at various time delays (case 1).

In general, the consideration of differenced stream flows (case 2) or of intervention time series (case 3) in the inputs of the neural model did not improve the quality of the predictions. This fact is, some how, in disagreement with the conclusions of Griñó (1992) who showed that the additional consideration of seasonal factors improve the forecasts of the daily demands in an urban water system by removing many of the peak errors. Also Abrahart and See (2000) used seasonal factors as inputs of neural networks for the prevision of river flows. Their results were better than those obtained in this paper. A possible explanation is the displacement of the estimated curves with regard to the observed curves in cases 1, 2 and 3. Moreover, it is obvious that this divergence of results depends of a wide spectrum of conditions: different sites with heterogeneity in precipitation, evaporation, infiltration, interception, soil moisture, overland flow, land use and geomorphology.

A trained feed forward neural network has a static architecture (fixed weights) and the output is solely determined by the present input state to the network and not by the initial and past states of the neurons in the network. Despite the memory added to the network by means of a buffer containing recent inputs (the flows in the three previous days), the forecasts provided by cases 1—3 in each day were systematically very close to the flows observed in the previous day. This circumstance is probably due to the fact that the correlation between observed flows in any two consecutive days were most of the time very high (R = 90.37%; $P_\alpha < 0.001$) and so, each daily 'occurrence' is highly responsible for the next daily 'realization'. Park (1998), Abrahart and See (2000), Pulido-Calvo et al. (2002, 2003) and Gutiérrez-Estrada et al. (2005) also reported this 'lag-one' difference between actual values and values resulting from multiple regression, ARIMA and/or neural network models applied to forecasting of different kinds of time variables. One of the possible ways to solve the systematic 'displacement' is to improve the learning capacity of the network by providing it additional inputs, such as those of other relevant hydrological or meteorological variables.

The convolution process of the input variables (case 4) resulted in less extreme values, in the underestimation of the high flows, and in a slight 'displacement' in the low flow range. The accuracy measures showed that case 4 was the one with the best capacity to implicitly access the non-linearity inherent to the stream flows, but there is clearly still matter for research. Consideration of hybrid neural network and ARIMA model (case 5) was in general favorable as it avoided the previously mentioned systematic 'displacement', though resulting in worse estimates of the maximum and minimum flows.

The poor performance of all CNNs in the range of the highest daily flows was a consequence of the much smaller number of observed flows belonging to those ranges. This circumstance is due to the highly irregular hydrologic regime of the Portuguese rivers characterized by low flows most of the time and just once a while by high flows, that however, can be very exceptional. The comparatively small number of high flows in the training set naturally compromises the 'learning' process of the CNNs during the calibration phase and, therefore, the capability of the models to forecast those flows. To overcome this problem, the use of distributed input data, that is to say, of flows from more than one stream gauging station, could be equated. By this way, a higher number of peak flows would be available for the model training. Other alternative approach could be based on dynamic-feedback neural networks. Chiang et al. (2004) proved the better performance of those models when data with insufficient length or content is involved in the training phase.

In cases 1, 2 and 3, the accuracy measures R^2, E_j and ARV based on the complete range of flows pointed towards the good performance of the models, with means values of R^2 and E_j higher than 0.7 and of ARV lower than 0.2. However, such measures may not provide information about the performance of the models as feasible as desired. In fact, the poor values of the error statistics RMSE, MAE %SEP and PI suggested that is necessary to carry out a multicriteria performance assessment based on different accuracy evaluation procedures applied to different ranges of flows. In general, the results of the accuracy measures for the complete range of flows are not good indicators of the capability of the models to predict peak flows, as Abrahart and See (2000) reported, owing to the overwhelming presence of low flow occurrences. One possible approach to overcome this situation is to group the original data into distinct hydrological event types, each individual cluster type being assessed by an independent model, Abrahart and See (2000).

The main problem associated with the CNNs developed in this paper was to identify the architecture that involved the least error of validation. This resulted in a longer data processing time compared to the others traditional forecasting techniques. Given that the neural networks are heuristic models, specific rules cannot be given regarding the controlling of convergence, network design or the initialization and change of weights to resolve a concrete problem. In the specialized literature, we only found some general guidelines taken from the experience of numerous authors. For this reason it is necessary to determine the nature of the problem beforehand; a condition which is not necessary in many other statistical forecasting methods. Hence, in this paper the analyses are begun by calculating the maximum delay of the input variables using the Box-Jenkins methodology following Griñó (1992). In other papers, principal components analyses are used to preprocess the data (Ventura et al., 1997) or multiple regressions to identify the independent variables that have a significant influence on the dependent variable (Pulido-Calvo et al., 2002, 2003).

In summary, it should be stressed the capability revealed by case 4 to forecast daily flows in a watershed different from the one utilized in the model calibration. That capability reinforces the results from the hydrological data transposition studies performed for Portuguese rivers. It can also play an important role, especially in the management of water resources systems. For instances, if the CNN model works on-line integrated in a centralized remote-control system, adaptive answers can be quickly implemented at some of the locations based on the changing conditions detected in different locations.

Summary and conclusions

The work presented focused the application of computational neural networks (CNNs) to short-term daily flow forecasting at large Portuguese watersheds. For the calibration of the CNNs only stream flow records were considered to be available. To identify the best approach five CNN based models (cases 1–5) were developed and compared in terms of the accuracy of the daily flows thus forecasted.

A CNN having as input data the flows in the three previous days combined with a convolution process of the input data – case 4 – proved to be the best approach. The adequacy of that approach to daily flow forecasting at a watershed different from the one utilized in the calibration of the model parameters was also preliminary analyzed.

As main conclusion of the study it can be stated that the case 4 proved to be a powerful tool that, with very modest data requirements, enable accurate daily flow forecasts at Portuguese watersheds. By combining hydrologic regionalization criteria with a CNN approach it is also possible to get daily flows forecasts at watersheds with insufficient flow data.

References

Abrahart, R.J., See, L., Kneale, P.E., 1999. Using pruning algorithms and genetic algorithms to optimize network architectures and forecasting inputs in a neural network rainfall-runoff model. J. Hydroinformatics 1 (2), 103–114.

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol. Process. 14, 2157–2172.

Abrahart, R.J., See, L., 2002. Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. Hydrol. Earth Syst. Sci. 6 (4), 655–670.

Al-Saba, T., El-Amin, I., 1999. Artificial neural networks as applied to long-term demand forecasting. Artif. Intell. Eng. 13 (2), 189–197.

Anctil, F., Rat, A., 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. 10 (1), 85–88.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I. Preliminary concepts. J. Hydrol. Eng. 5 (2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology. II. Hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.

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

Cameron, D., Kneale, P., See, L., 2002. An evaluation of a traditional and a neural net modelling approach to flood forecasting for an upland catchment. Hydrol. Process. 16, 1033–1046.

Chang, F.-J., Chen, Y.-C., 2001. A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. J. Hydrol. 245, 153–164.

Chiang, Y.-M., Chang, L.-C., Chang, F.-J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. J. Hydrol. 290, 297–311.

De Vries, B., Principe, J.C., 1991. A Theory for Neural Networks with Time DelaysAdvances in Neural Information Processing Systems, 3. Morgan Kaufmann Publishers, California.
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.
Duan, Q., Gupta, V.K., Sorooshian, S., 1993. A shuffled complex evolution approach for effective and efficient global minimization. J. Optim. Theory Appl. 73 (3), 501—521.
Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265—284.
Elshorbagy, A., Simonovic, S.P., Panu, U.S., 2000. Performance evaluation of artificial neural networks for runoff prediction. J. Hydrol. Eng. 5 (4), 424—427.
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.
Griñó, R., 1992. Neural networks for univariate time series forecasting and their application to water demand prediction. Neural Network World 2 (5), 437—450.
Gutiérrez-Estrada, J.C., de Pedro-Sanz, E., López-Luque, R., Pulido-Calvo, I., 2004. Comparison between traditional methods and artificial neural networks for ammonia concentration forescasting in an eel (*Anguilla anguilla* L.) intensive rearing system. Aquacult. Eng. 31, 183—203.
Gutiérrez-Estrada, J.C., de Pedro-Sanz, E., López-Luque, R., Pulido-Calvo, I., 2005. Estimación a corto plazo de la temperatura del agua. Aplicación en sistemas de producción en medio acuático. Ing. Agua 12 (1), 77—92 (in Spanish).
Hansen, J.V., Nelson, R.D., 1997. Neural networks and traditional time series methods: a synergistic combination in state economic forecasts. IEEE Trans. Neural Networks 8 (4), 863—873.
Hsu, K., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res. 31 (10), 2517—2530.
Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? Water Resour. Res. 29 (8), 2637—2649.
Khalil, M., Panu, U.S., Lennox, W.C., 2001. Groups and neural networks based streamflow data infilling procedures. J. Hydrol. 241, 153—176.
Kitanidis, P.K., Bras, R.L., 1980. Real time forecasting with a conceptual hydrological model. 2. Applications and results. Water Resour. Res. 16 (6), 1034—1044.
Kuligowski, R.J., Barros, A.P., 1998. Experiments in short-term precipitation forecasting using artificial neural networks. Mon. Weather Rev. 126 (2), 470—482.
Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233—241.
Lorrai, M., Sechi, G.M., 1995. Neural nets for modelling rainfall-runoff transformations. Water Resour. Manage. 9, 299—313.
Mason, J.C., Tem'me, A., Price, R.K., 1996. A neural network model of rainfall-runoff using radial basis functions. J. Recherches Hydrauliques 34 (4), 537—548.
Moradkhani, H., Hsu, K., Gupta, H.V., Sorooshian, S., 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. J. Hydrol. 295, 246—262.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I. A discussion of principles. J. Hydrol. 10, 282—290.
Park, H.-H., 1998. Analysis and prediction of walleye pollock (*Theragra chalcogramma*) landings in Korea by time series analysis. Fish. Res. 38, 1—7.
Poff, L.N., Tokar, A.S., Johnson, P.A., 1996. Stream hydrological and ecological responses to climate change assessed with an artificial neural network. Limnol. Oceanogr. 41 (5), 857—863.
Portela, M.M., Quintela, A.C., 2000. A altura do escoamento anual médio numa bacia hidrográfica como parâmetro de regionalização de informação hidrométrica (Mean annual flow height as a regionalization parameter of hydrologic information). 1° Congresso sobre Aproveitamentos e Gestão de Recursos Hídricos em Países de Idioma Português, Rio de Janeiro, Brasil, pp. 218—227 (in Portuguese).
Portela, M.M., Quintela, A.C., 2002a. Assessment of the streamflow characteristics under unavailability of discharge data: the mean annual flow depth over the watershed as a regionalization parameter. In: The Portuguese case, 2002 EGS Conference, European Geophysical Society, Nice, France.
Portela, M.M., Quintela, A.C., 2002b. Evaluation of the water resources in Portuguese watersheds without streamflow data. In: International Conference of Basin Organizations, Madrid, Spain.
Portela, M.M., Quintela, A.C., 2005a. Regionalization of hydrologic information: establishment of flow series at ungauged watersheds. In: Cunha, M.C., Brebbia, C.A. (Eds.), Water Resources Management III. WIT Press, UK.
Portela, M.M., Quintela, A.C., 2005b. Estimação de séries de caudais médios diários na ausência de informação hidrométrica (Evaluation of daily stream flow series under unavailability of stream flow data). VII Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa (7° SILUSBA), APRH, Associação dos Recursos Hídricos, Portugal (in Portuguese).
Portela, M.M., Quintela, A.C., 2006. Preliminary design of the storage capacity of reservoirs based on a flow regionalization parameter. BALWOIS 2006, International Conference on 'Water Observation and Information System for Decision Support', IAHS, IRD (France), HMS (Republic of Macedonia), HIO (Republic of Macedonia), Republic of Macedonia.
Pulido-Calvo, I., Roldán, J., López-Luque, R., Gutiérrez-Estrada, J.C., 2002. Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola. Ing. Agua 9 (3), 319—331 (in Spanish).
Pulido-Calvo, I., Roldán, J., López-Luque, R., Gutiérrez-Estrada, J.C., 2003. Demand forecasting for irrigation water distribution system. J. Irrig. Drain. Eng. 129 (6), 422—431.
Qi, M., Zhang, G.P., 2001. An investigation of model selection criteria for neural network time series forecasting. Eur. J. Oper. Res. 132, 666—680.
Quintela, A.C., 1967. Recursos de águas superficiais em Portugal Continental. Ph.D. Thesis, IST, Technical University of Lisbon, Portugal (in Portuguese).
Roger, L.L., Dowla, F.U., 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Resour. Res. 30 (2), 457—481.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. 'Learning' representations by backpropagation errors. Nature 323, 533—536.
See, L., Openshaw, S., 2000. A hybrid multi-model approach to river level forecasting. Hydrol. Sci. J. 45 (4), 523—536.
Shepherd, A.J., 1997. Second-order Methods for Neural Networks. Springer, New York.
Shin, H-S., Salas, J.D., 2000. Regional drought analysis based on neural networks. J. Hydrol. Eng. 5 (2), 145—155.
Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model. Water Resour. Res. 29 (4), 1185—1194.
Tan, Y., van Cauwenberghe, A., 1999. Neural-network-based *d*-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intell. 12 (1), 21—25.
Thirumalaiah, K., Deo, M.C., 1998. River stage forecasting using artificial neural networks. J. Hydrol. Eng. 3 (1), 26—32.
Thirumalaiah, K., Deo, M.C., 2000. Hydrological forecasting using neural networks. J. Hydrol. Eng. 5 (2), 180—189.

Tokar, A.S., Johnson, P.A., 1999. Rainfall-runoff modeling using artificial neural networks. J. Hydrol. Eng. 4 (3), 232—239.

Tokar, A.S., Markus, M., 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models. J. Hydrol. Eng. 5 (2), 156—161.

Tsoukalas, L.H., Uhrig, R.E., 1997. Fuzzy and Neural Approaches in Engineering. Wiley Interscience, New York.

Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, C., 1995. Artificial neural networks for estimation of kinetic analytical parameters. Anal. Chem. 67 (9), 1521—1525.

Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, C., 1997. Computational neural networks in conjunction with principal component analysis for resolving highly nonlinear kinetics. J. Chem. Inf. Comput. Sci. 37 (2), 287—291.

Wedding II, D.K., Cios, K.J., 1996. Time series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. Neurocomputing 10, 149—168.

Wilson, J.H., Keating, B., 1996. Business Forecasting. Irwin, London.

Yang, C.C., Prasher, S.O., Lacroix, R., Sreekanth, S., Patni, N.K., Masse, L., 1997. Artificial neural network model for subsurface-drained farmland. J. Irrig. Drain. Eng. 123 (4), 285—292.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J. Hydrol. 181, 23—48.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204, 83—97.

Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159—175.

Zhang, M., Fulcher, J., Scofield, R.A., 1997. Rainfall estimation using artificial neural network group. Neurocomputing 16, 97—115.

Zhang, G.P., Patuwo, B.E., Hu, M.Y., 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting. Comp. Oper. Res. 28 (4), 381—396.