BIOSYSTEMS ENGINEERING 97 (2007) 283–293

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper: SW—Soil and Water

Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems

Inmaculada Pulido-Calvo[a,*], *Pilar Montesinos*[b], *José Roldán*[b], *Francisco Ruiz-Navarro*[b]

[a]*Dpto. Ciencias Agroforestales, EPS, Campus de La Rábida, Universidad de Huelva, 21819 Palos de la Frontera (Huelva), Spain*
[b]*Dpto. Agronomía, ETSIAM, Universidad de Córdoba, Apdo. 3048, 14080 Córdoba, Spain*

ARTICLE INFO

Article history:
Received 2 June 2006
Accepted 6 March 2007
Available online 21 May 2007

Information regarding water demand is key to managing consumption in irrigation districts. Forecasting water demand is one of the main problems for designers and managers of water delivery systems. This paper evaluates the performance of linear multiple regressions and feed forward computational neural networks (CNNs) trained with the Levenberg–Marquardt algorithm for the purpose of irrigation demand modelling. The models are established using data recorded from an irrigation water distribution system located in Andalusia, southern Spain, during two irrigation seasons (2001/2002, 2002/2003). A commercial telemetry system was installed on 28 farms of the irrigation network to record water volumes in real time. The input or independent variables used in various CNN and multiple regression models are: (a) water demands from previous days; (b) water demands and climatic data (rainfall, maximum, minimum and average temperatures, relative humidity and wind speed) from previous days. Good predictions were obtained when water demand original data were modified in the calibration period by a smoothing process to reduce the noise in the data acquisition during the start-up of the research project. The best predictions were obtained when water demand recorded during the two previous days was used as input data.

1. Introduction

Demands for water are increasing in both quantity and quality; a phenomenon that is conditioned by social, political and environmental factors. The pressures to meet growing demands have led to greater competition for available water resources among traditional water consumers, namely agriculture, industry and cities. These competing interests are already limiting social, industrial and rural development actions of many countries. Furthermore, the fact that this growing demand for water is not coupled or sychronised with increased resources is giving rise to greater competition between regions or countries for access to water (FAO, 1993; Ohlsson, 1995; Sumpsi *et al.*, 1998).

The current concern for environmental protection has given rise to a new factor affecting competition for water. Certain non-consumptive uses for recreational, ecological or landscaping purposes are now being considered when assigning water for consumptive uses. Thus, not only has competition increased in terms of the amount demanded, but also the quality.

As a result of these different factors affecting competition for water resources, water is increasingly considered a scarce and valuable resource requiring rigorous management and extreme caution to prevent its depletion. One of the keys to solving these problems lies in the agricultural sector given that irrigated agriculture is the largest user of water throughout the world, accounting for 87% of consumptive uses (ONU, 1997; Sumpsi *et al.*, 1998).

*Corresponding author.
E-mail addresses: ipulido@uhu.es (I. Pulido-Calvo), pmontesinos@uco.es (P. Montesinos), jroldan@uco.es (J. Roldán).
1537-5110/$ - see front matter
doi:10.1016/j.biosystemseng.2007.03.003

The improvement of water management in an irrigation district requires the analysis of water demand in order to determine ways in which it may be modified and rationalised with a view to making water management policy more efficient. Information regarding water demand in irrigated areas is key to the development of policies on irrigation water consumption. These data can provide us with information regarding the marginal value of water and the response level to different irrigation water rates. This also provides reference data for the design, modernisation and exploitation of water-delivery systems (Kadra & Lamaddalena, 2006).

Daily water requirements for crop irrigation can be estimated by the rates of percolation and evapotranspiration and used for irrigation planning. Many models have been used to simulate these water requirements, from empirical or functional (Doorenbos & Pruit, 1977; Doorenbos & Kassam, 1979; Allen *et al*., 1998) to mechanistic (Van Aelst *et al*., 1988). However, water requirements calculated for irrigation planning are not always suitable for predicting actual use (i.e., consumer demand) due to changes in the field environment such as weather conditions and farm management practices, which can influence the actual amounts of water needed.

Additionally, to facilitate data acquisition and irrigation system management and operation, recently developed tools, such as remote sensing and geographic information systems GIS (Hartkamp *et al*., 1999; Kite, 2000; Kite & Droogers, 2000; Lorite *et al*., 2004), and monitoring and controlling systems (Leib *et al*., 2003; Mareels *et al*., 2005; Miranda *et al*., 2005), have been combined with hydrologic models to assess and to improve the behaviour of irrigation schemes.

Despite these advances water management in some irrigation districts is carried out using only the experience and knowledge of the administrator although there is always a need to forecast daily water demand. Significant progress in the field of forecasting has recently been made possible through advances in a branch of nonlinear system theory modelling called artificial or computational neural networks (ANNs or CNNs). The neural approaches are increasingly being applied in many fields of science and engineering and usually providing highly satisfactory results.

Some of the applications of CNNs for the management of water resources include modelling the rainfall-runoff process (Hsu *et al*., 1995; Lorrai & Sechi, 1995; Mason *et al*., 1996; Abrahart *et al*., 1999; Tokar & Johnson, 1999; Thirumalaiah & Deo, 2000; Tokar & Markus, 2000; Chiang *et al*., 2004; Moradkhani *et al*., 2004; Anctil & Rat, 2005; Agarwal *et al*., 2006), short-term river stage forecasting (Thirumalaiah & Deo, 1998, 2000; Abrahart & See, 2000, 2002; See & Openshaw, 2000; Cameron *et al*., 2002; Nayebi *et al*., 2006; Pulido-Calvo & Portela, 2007), rainfall forecasting (French *et al*., 1992; Zhang *et al*., 1997; Kuligowski & Barros, 1998), groundwater modelling (Roger & Dowla, 1994; Yang *et al*., 1997), predicting the soil water contents (Givi *et al*., 2004) and nitrate-nitrogen in drainage water (Sharma *et al*., 2003), and drought analysis (Shin & Salas, 2000), among others. Previous works on water demand forecasting both in urban supply systems and irrigation districts (Griñó, 1992; Pulido-Calvo *et al*., 2002, 2003) show that the use of CNNs provide very satisfactory results.

The objective of this paper was to forecast consumer demands of an irrigation area using on-farm water-use information from supervisory control system and approaches based on linear regression, traditional forecasting methods, and on computational neural networks; heuristic models included in the knowledge field known as soft-computing. The purpose of forecasting is the real-time control of the daily water uses at the farm-scale for the various crops, as proposal of improvement water supply management in on-demand irrigation districts.

2. Material and methods

2.1. *Study area*

The methods developed in this paper were applied to the Genil-Cabra irrigation district for purposes of comparison. The district currently comprises about 15,000 irrigated hectares belonging to the Guadalquivir River basin in the provinces of Córdoba and Seville (southern Spain) (Fig. 1). The water used to irrigate the district is supplied from the Iznájar reservoir, which has a maximum storage capacity of $0.978\,km^3$. The water is then transported by the Genil River to the Cordobilla dam, with a capacity of $0.034\,km^3$. This is where the main pumping station that lifts the water to the principal canal is located. This canal conveys water to several pumping stations. The main pipe networks (pressurised or gravity-fed according to hydrant elevations) distribute water from these pumping stations to the different sectors where it is delivered to the farms by secondary pipe networks. This water distribution system permits farmers to use irrigation water throughout the year 24 h a day, as long as they do not exceed the preset maximum instantaneous flow value.

The predominant crops in the district include olive, cotton and winter cereals. Although crops such as sunflower predominated in the past, accounting for 45% of the surface area during the 1992/1993 irrigation season, during the 2002/2003 season sunflower crops accounted for less than 5% of the total surface area due to market fluctuations and European Agricultural Policy. Maize growing is rising accounting for almost 10% of the cropping area during the 2002/2003 season as compared to less than 0.5% in the 1992/1993 season. During this same season, olive crops accounted for almost 16% of the cropping area and currently occupy the largest percentage—some 34%.

One of the main problems when managing water resources in irrigated districts is the lack of reliable and comparable information on consumption. For this reason, an observation network was established in this study as an indispensable condition for the global and integrated management of water resources in irrigated areas. The network selected is a commercial telemetry system known as SIGA (Sistema Integral de Gestión del Agua—Integral Water Management System in English). This network consists of a telemetric system with the corresponding software for its control and the management and treatment of the data obtained, thus permitting irrigation water distribution systems to be monitored. Specifically, this system was installed on 28 farms located in sectors II–VII of the irrigation network to record

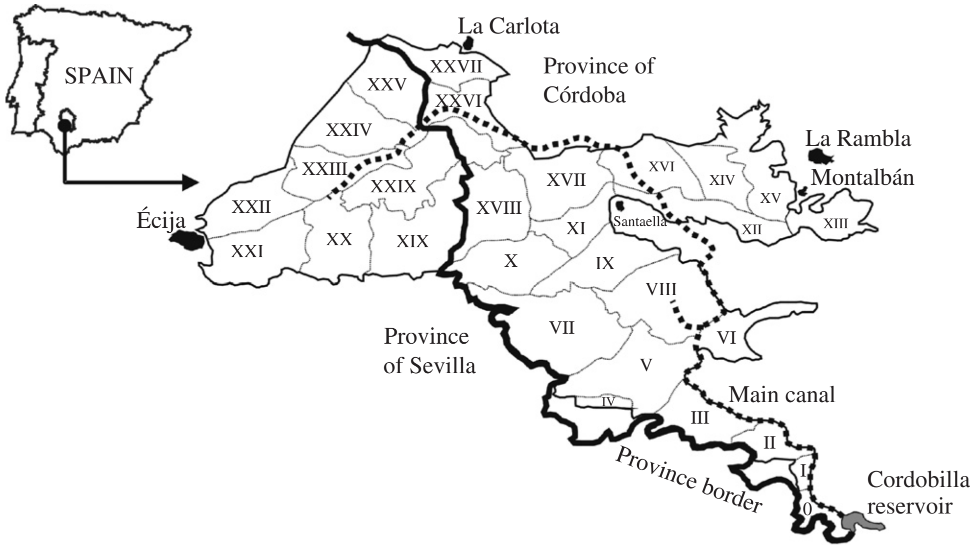

Fig. 1 – Sectors of the Genil–Cabra irrigation district.

water volumes and pressures in real time and transmit data via cable and radio in order to monitor the hydrants. The system was installed in two stages: on 16 farms during the 2001/2002 irrigation season and on 12 additional farms during the 2002/2003 irrigation season. These particular farms were chosen as they represented a wide range of crops (14 olive crops and 14 dedicated to herbaceous crops), soil types (vertisol and a local type known as villar) and irrigation methods (sprinkler and drip).

Equations recommended by the FAO (Allen *et al.*, 1998; Pirmoradian & Sepaskhah, 2006) were used to determine irrigation water requirements. Daily reference evapotranspiration E_{To} has been estimated using the Penman–Monteith equation. Crop and climatic data from 2001/2002 and 2002/2003 irrigation seasons were used. The crop data (crop coefficients and duration of the development stages) were obtained from other study in the area that adjusted these values to local conditions (Lorite, 2003). The climatic data were collected at the Santaella (Córdoba) weather station located in the irrigation district. Mean annual rainfall in the area was 606 mm, with a standard deviation of 216 mm. Because of the area's Mediterranean climate, there is a wet season in winter and a very dry season in summer. Consequently, monthly rainfall distribution is very irregular. The mean air temperature ranges from 10 °C in winter to over 27 °C in summer. Fig. 2 shows accumulated rainfall and E_{To} at the 2002/2003 irrigation season. There is rainfall deficit versus evaporative demand from May month.

The mean irrigation water requirements of the selected farms were higher than measured values of consumer water demands. So, the average value for the index Annual Relative Water Supply ARWS (Malano & Burton, 2001; Lorite *et al.*, 2004; Malano *et al.*, 2004; Rodríguez-Díaz *et al.*, 2004), which relates the total volume of water applied (irrigation plus rainfall) to the volume of water required by the crop (computed as gross irrigation requirements plus rainfall), was 0.81 ± 0.12. This indicates a deficit-irrigation situation. The relatively high water productivity found in the Genil–Cabra irrigation district ($1.42\,€\,m^{-3}$) is due to a combination of deficit irrigation and the widespread use of herbaceous winter crops and olive which efficiently use a substantial proportion of the annual rainfall in Mediterranean climates, thus lowering their irrigation requirements (Lorite *et al.*, 2004).

2.2. *Telemetry system*

To determine the water consumption in the irrigation district, the SIGA remote control system was installed in the 28 hydrants selected. There is an automatic system in SIGA that allows the data captured by the sensors installed on the farm hydrants to be recorded, stored, sent, processed and displayed in a sequential manner. The information recorded for our study included time, water volumes and instantaneous pressures. Due to the tree-shaped structure of the data capturing system, points of measurement distributed irregularly throughout the area of over 8000 ha were monitored. The information was sent to a central control site where it was received, stored and displayed by means of SCADA (Supervisory Control and Data Acquisition) commercial software (Fig. 3).

The main components of the telemetry system included: (a) the irrigation node or remote station which captured the signals from the volumetric water meters and the pressure gauges installed in the irrigation hydrants; (b) the concentration station which received and stored the field signals sent from the irrigation nodes; and (c) the central control site which displayed and stored the data recorded in the concentrating stations and their associated nodes (Fig. 3).

Data was read hourly in order to obtain 24 daily records from each of the hydrants. This reading frequency, which could be adapted accordingly to user needs, allowed initiation time, duration of irrigation, volume consumed and used by the farmer to be determined. In this way, hourly water demand data (which was then transformed into daily demand) was available for each of the farms under study.

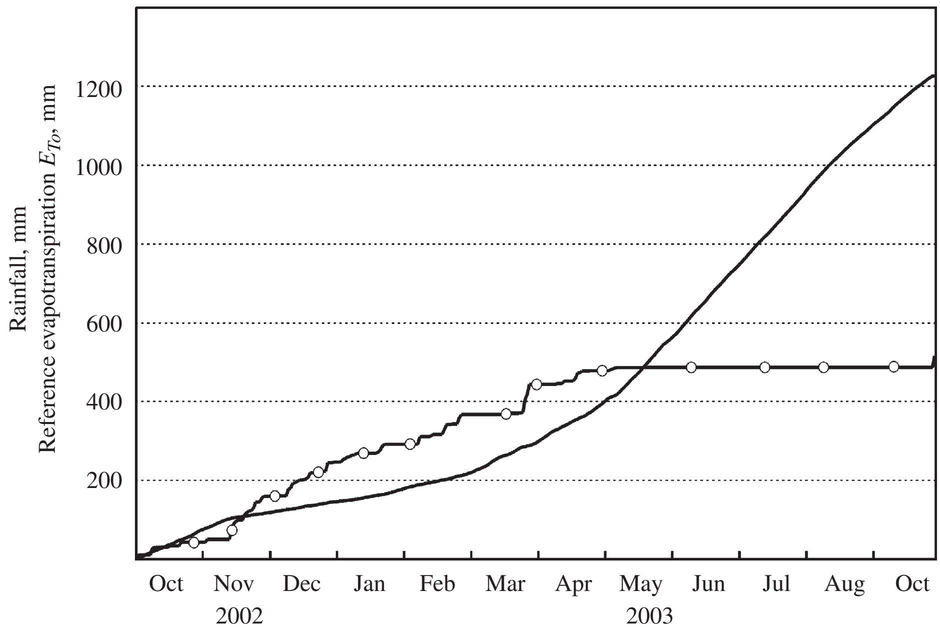

Fig. 2 – Accumulated rainfall (—○—) and reference evapotranspiration E_{To} (—) at the 2002/2003 irrigation season in the Genil–Cabra irrigation district.

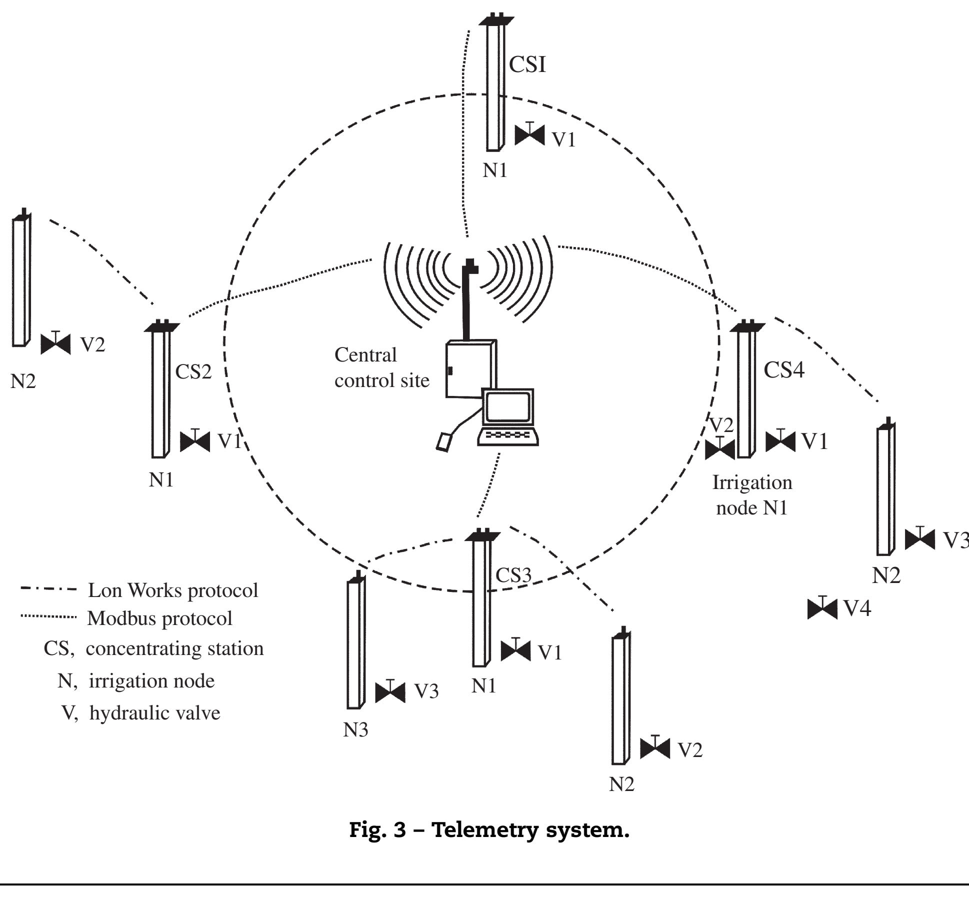

Fig. 3 – Telemetry system.

2.3. *Daily water demand forecasting by computational neural networks*

A computational neural network (CNN) is a heuristic model that simulates the structure and functioning of a biological neural system. The basic processor or neuron is a simple calculation device which uses an external input vector or intermediate results from neurons to provide a single response or output. In general, a neural network consists of a set of nodes or neurons which are grouped together in several layers and interconnected (input layer, intermediate or hidden layers and output layer). The most widely studied and used structures are multilayer feed forward networks (Rumelhart *et al.*, 1986). The feed forward connections transfer information only from earlier layer to the next consecutive layers. A typical four-layer feed forward CNN has g, n, m and r nodes or neurons in the input, first hidden, second hidden and output layers, respectively. The notation of this neural network is (g:n:m:r).

The connections that join the neurons are assigned a numeric weight W. To determine this set of weights a corrective-repetitive process called learning or training or calibration of the CNN is performed. This training helps to define the interconnections among neurons (weights W), and it is accomplished by using both 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.

To calibrate and validate CNN models, STATISTICA 6·0 (Statsoft, Inc., 1984–2002) software module was used. The linear function l was used as an information transfer function between neurons of the output layer, while the logistic transfer function, also known as sigmoid s was used for each of the neurons in the intermediate layers. The learning algorithm chosen for calibration and later validation of the models was the Levenberg–Marquardt (LM) second-order supervised algorithm (Shepherd, 1997).

The standard backpropagation learning algorithm (Rumelhart *et al.*, 1986) is the most widely used supervised algorithm in neural systems. This algorithm is based on modifying the values of the weights proportionally to the gradient of the error function in order to obtain a local minimum (gradient-descent algorithm). Given that the convergence velocity of this algorithm is slow, additional steps were used to accelerate convergence making use of second-order information of the error function, that is, from its second derivatives, or in matrix form, of its Hessian derivative (second-order algorithms). In the particular case of minimising a squared error function, the Hessian derivative can be approximated by using only the first derivatives of the neural system outputs, in the same manner as the Gauss–Newton algorithm. Given that this algorithm can be unstable when the approximation is not positive definite, the LM algorithm resolves this problem by introducing an adaptive term. For this reason, it is recommended as a first option by many authors (Tan & van Cauwenberghe, 1999; Martín-del-Brío & Sanz-Molina, 2001; Anctil & Rat, 2005; Nayebi *et al.*, 2006).

The key feature of the LM algorithm is that the determination of the new search direction of the minimum of the error function is obtained by a compromise between the step λ and the error variation with respect to the weights W. Hence, the change in the weight coefficients W is expressed as:

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

where ε is the vector of errors; $\boldsymbol{Z}$ is the matrix of partial derivatives of these errors with respect to the weights W; $\boldsymbol{I}$ is the identity matrix; and T is the transpose operator. The first term of the second member of the LM formula represents the linear assumption and the second, the gradient-descent step. The control parameter λ governs the relative influence of these two approaches.

Epoch refers to the period of time that encompasses all the iterations performed when all the patterns are presented once. A fundamental feature of neural networks is their ability to generalise by examples. Generalisation is understood as the ability of the network to respond correctly to behaviour patterns that have not been used during training or calibration. Thus, a given network architecture is trained until it reaches an optimal point at which the generalisation error is minimal. In this paper, controlled learning was performed by means of an internal validation method (20% of the calibration data to test the error at the end of each epoch) (Tsoukalas & Uhrig, 1997; Gutiérrez-Estrada *et al.*, 2004; Pulido-Calvo & Portela, 2007). Weights W are updated at the end of each epoch. The weights that are selected are those which correspond to the number of epochs that present the least error in the internal validation.

Although a neural network with an intermediate or hidden layer can represent any continuous function (Cybenco, 1989), networks with two intermediate layers were used (universal approximator of any function: Cybenco, 1989) since the relationship among variables was not known a priori. Furthermore, when there is only one intermediate layer, the number of hidden neurons needed to achieve a certain level of error can be so high that the application was not feasible in practice. After testing from 2 to 18 nodes in each of the hidden layers, the one that obtained the best results in the validation period was selected.

2.4. *Comparison with multiple regression analysis*

The aim of multiple regression analysis is to obtain a linear equation that allows the dependent variable or criterion e, to be estimated when the values of the q independent or predictive variables $x_1,\ldots, x_q$ are known:

$$e = b_0 + b_1\, x_l + \ldots + b_q x_q, \tag{2}$$

where the parameters $b_0, b_1,\ldots, b_q$ represent the contributions of each independent variable to the estimation of the dependent variable.

2.5. *Data and identification of models*

Two types of multiple regression and neural network models were developed according to the selection of variables used as input variables to the neural network or as independent variables of the multiple regression: (a) water demands from several previous days; (b) water demands and climatic data (rainfall, maximum, minimum and average temperatures, relative humidity and wind speed) from several previous days. Daily climatic data were recorded during the 2001/2002 and 2002/2003 irrigation seasons at the Santaella (Córdoba) weather station located in the irrigation district. The same cases were developed to estimate daily water demand with multiple regression analysis and the CNNs. The inputs to the neural network were the same as the independent variables of the multiple regression. In all of the cases, the output or dependent variable was the estimated water demand for the day following those which were used to forecast demand, expressed as $m^3 ha^{-1} day^{-1}$.

With multiple regression and neural network analyses, data were processed as follows: (a) the model was calibrated with data from the 2001/2002 irrigation season; and (b) the dependent variable (daily water demand) was forecast using data from the 2002/2003 season (model validation) to ensure that the results of the model could be generalised to the population and were not specific to the sample used in the calibration.

The number of previous days was selected as input data by means of the partial autocorrelation function analysis of the daily water demand series. The partial autocorrelation showed two significant peaks related with the two previous days. This denotes a two-order autoregressive process, according to which the water demand in each day may be considered a function of its own past values in the two previous days (Wilson & Keating, 1996).

The reliability of the forecast largely depends on the quality and quantity of the available data. In the case under study, inaccuracies and noise were detected in the daily water consumption data series from the 2001/2002 irrigation season as a result of errors in the data acquisition during the start-up of the research project. To correct this problem, the water demand information contained in the training data set (the 2001/2002 irrigation season) was pre-treated by means of filtering, triangular smoothing or weighted moving average (Coulbeck, 1988; De Vries & Principe, 1991; Gutiérrez-Estrada *et al.*, 2004):

$$S_j = \frac{x_{j-2} + 2x_{j-1} + 3x_j + 2x_{j+1} + x_{j+2}}{9} \quad \forall j = 3 \text{ to } n-2, \tag{3}$$

where S_j is the j^{th} observation of the series of smoothed data, x_j is the j^{th} observation in the training data set and n is the total number of observations in the training data set. The smoothing process of the water demand was only carried out during the calibration or training period of models (the 2001/2002 irrigation season). With the trained neural network or the calibrated multiple regression, a weight W configuration or regression parameters b_0, $b_1, \ldots, b_q$, respectively, used for the validation period (the 2002/2003 irrigation season) considering as variables the original data without smoothing.

The models were also calibrated using a 2002/2003 irrigation season data set (60% of the original data of 2002/2003 irrigation season) while the remaining data of 2002/2003 irrigation season were used for model validation. In this case the results were not as good.

2.6. *Measures to evaluate the models (validation phase)*

To assess the performance of the neural networks and the multiple regressions during the validation phase and therefore to identify the best model, several measures of accuracy were applied. This was because there is not a unique performance evaluation test (Legates & McCabe Jr., 1999; Abrahart & See, 2000). The determination coefficient R^2 describes the proportion of total variance in the observed data that can be explained by the model. Others measures of variances included the percent of standard error of prediction E_S, % (Ventura *et al.*, 1995) and the efficiency coefficient E (Nash & Sutcliffe, 1970; Kitanidis & Bras, 1980).

Likewise, it is appropriate to quantify the error in the same units of the variables. These absolute error measures, included the square root of the mean square error (E_{RMS}) given by

$$E_{RMS} = \sqrt{\frac{\sum_{i=1}^{N} (Q_t - \hat{Q}_t)^2}{N}}, \tag{4}$$

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

The percent of standard error of prediction E_S is defined as

$$E_S = \frac{100}{\overline{Q}} E_{RMS}, \tag{5}$$

where $\overline{Q}$ is the average of the flows observed in the validation set. The principal advantage of E_S is its non-dimensionality since it allows forecasts given by different models to be compared on the same basis.

The efficiency coefficient E is calculated as

$$E = 1.0 - \frac{\sum_{i=l}^{N} (Q_t - \hat{Q}_t)^2}{\sum_{i=l}^{N} (Q_t - \overline{Q}_t)^2}. \tag{6}$$

To obtain an acceptable goodness of fit, the values of R^2 and E must be close to one, while the values of E_S must be close to zero. A value of zero for E indicates that the average observed value $\overline{Q}$ is as good predictor as the model, while negative values indicate that the observed average is a better predictor than the model (Legates & McCabe Jr., 1999).

3. Results

When the calibration data series for recorded water usage was not pre-filtered, the measures of prediction accuracy were not good (Table 1). For both the neural models and multiple regressions the best values for the evaluation magnitudes were obtained for the farms dedicated to olive monoculture when water demand for the 2 days prior to forecasting was used as the input or for the independent variables. In all of the models considered, the neural networks provide better forecasts than the multiple regressions. Fig. 4 shows the results of water demand forecasting on olive farms during the validation period of the best neural network (2:9s:9s:1 l) (two neurons in the input layer, nine neurons in each of the intermediate layers and one neuron in the output layer) and of the best regression with the water demand of the two previous days as independent variables. The predictive equation of this regression analysis obtained in calibration period is

$$\hat{Q}_t = 4.01 + 0.91\, Q_{t-1} - 0.18\, Q_{t-2}, \tag{7}$$

where $\hat{Q}_t$ is the estimated water demand in day t on olive farms; Q_{t-1} is the observed water demand in day $t-1$ on olive farms; and Q_{t-2} is the observed water demand in day $t-2$ on olive farms. The determination coefficient was $R^2 = 0.61$, the Snedecor statistic was $F = 295.09$, the statistical level of significance is $p < 0.001$, and degrees of freedom were 120.

When the calibration data series was subject to a triangular smoothing, the results were greatly improved. In almost all

Table 1 – Goodness of fit of the multiple regression and neural network models using original calibration data recorded by the control system (SIGA)

Crop	Independent variables	Model	R^2	E_{RMS}, $m^3\ ha^{-1}\ day^{-1}$	E_S, %	E
Olive	Demand of the two previous days	CNN(2:9s:9s:1 l)	0.40	22.76	51.69	0.38
		Regression	0.28	35.33	88.32	0.18
	Demand and climatic data of the two previous days	CNN(14:11s:11s:1 l)	0.36	28.38	60.09	0.30
		Regression	0.25	37.22	89.24	0.16
Cotton	Demand of the two previous days	CNN(2:7s:7s:1 l)	0.38	24.96	54.09	0.28
		Regression	0.25	36.22	90.02	0.15
	Demand and climatic data of the two previous days	CNN(14:12s:12s:1 l)	0.33	30.09	63.19	0.29
		Regression	0.22	40.61	91.14	0.13
Maize	Demand of the two previous days	CNN(2:7s:7s:1 l)	0.36	25.23	55.34	0.22
		Regression	0.27	36.10	89.44	0.16
	Demand and climatic data of the two previous days	CNN(14:13s:13s:1 l)	0.31	32.39	65.90	0.32
		Regression	0.23	42.11	91.89	0.14

R^2, determination coefficient; E_{RMS}, square root of mean square error; E_S, standard error of prediction; E, efficiency coefficient; CNN(2:9s:9s:1 l), neural network with 2 neurons in the input layer, 9 neurons in each of the intermediate layers with sigmoid information transfer function s and 1 neuron in the output layer with linear transfer function l.

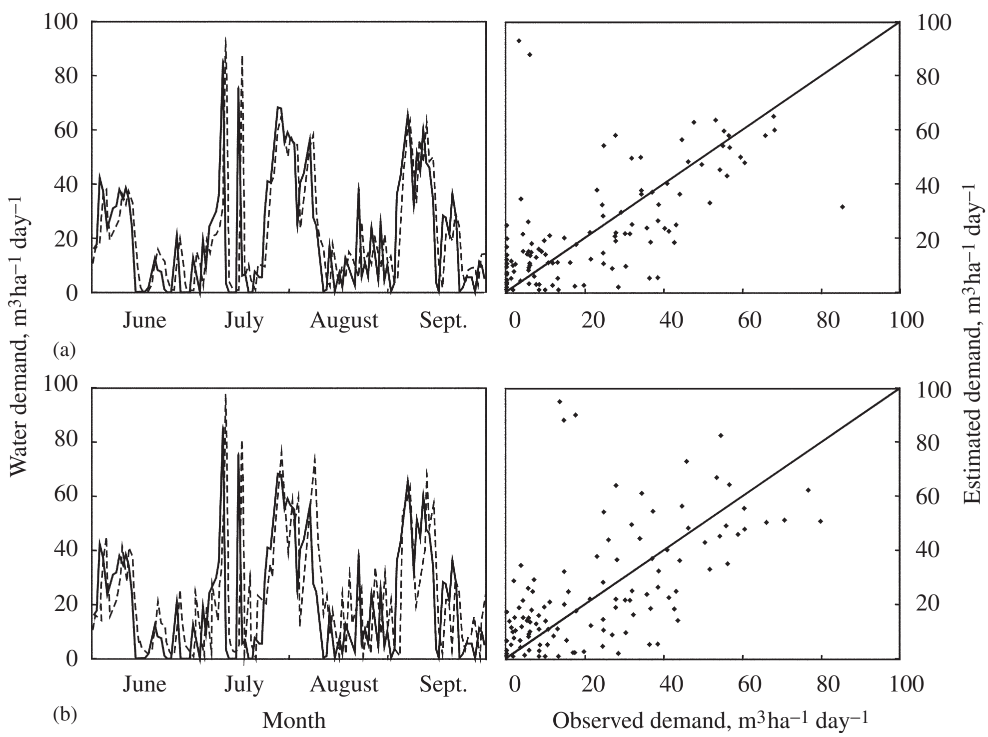

Fig. 4 – One-step-ahead prediction of water demands on the olive farms considered the demand of the two previous days as independent variables and scatterplots comparing observed (—) and estimated (- - -) daily water demands (validation data): (a) neural network with two neurons in the input layer, nine neurons in the intermediate layers and one neuron in the output layer CNN(2:9 s:9 s:1 l); and (b) multiple regression.

the models, values for R^2 were higher than 0.7 and for E_S were lower than 30% (Table 2). The best results were achieved when the water demand for the two days prior to forecasting were considered as input or independent variables. In this case, the regression models provided somewhat better evaluation magnitudes than the CNNs, although the differences were negligible. However, the neural networks were better than the multiple regressions when water demand and climatic data from the two previous days were used as input variables. Fig. 5 shows the results of demand forecasting on olive farms during the validation period of the best neural network (2:6s:6s:1 l) (two neurons in the input layer, six neurons in each of the intermediate layers and one neuron in the output layer) and of the best regression with the water demand of the

Table 2 – Goodness of fit of the multiple regression and neural network models using filtered calibration data recorded by the control system (SIGA)

Crop	Independent variables	Model	R^2	E_{RMS}, $m^3 ha^{-1} day^{-1}$	E_S, %	E
Olive	Demand of the two previous days	CNN(2:6s:6s:1 l)	0.95	3.17	19.19	0.94
		Regression	0.96	3.03	17.10	0.95
	Demand and climatic data of the two previous days	CNN(14:7s:7s:1 l)	0.78	7.29	24.11	0.67
		Regression	0.75	8.33	47.10	0.45
Cotton	Demand of the two previous days	CNN(2:8s:8s:1 l)	0.93	4.36	19.40	0.92
		Regression	0.95	3.71	15.49	0.95
	Demand and climatic data of the two previous days	CNN(14:7s:7s:1 l)	0.78	7.32	25.00	0.67
		Regression	0.73	9.07	48.89	0.47
Maize	Demand of the two previous days	CNN(2:7s:7s:1 l)	0.92	4.89	19.87	0.91
		Regression	0.93	4.78	19.02	0.92
	Demand and climatic data of the two previous days	CNN(14:11s:11s:1 l)	0.79	6.76	24.03	0.69
		Regression	0.73	9.33	48.09	0.46

R^2, determination coefficient; E_{RMS}, square root of mean square error; E_S, standard error of prediction; E, efficiency coefficient; CNN(2:6s:6s:1 l), neural network with 2 neurons in the input layer, 6 neurons in each of the intermediate layers with sigmoid information transfer function s and 1 neuron in the output layer with linear transfer function 1.

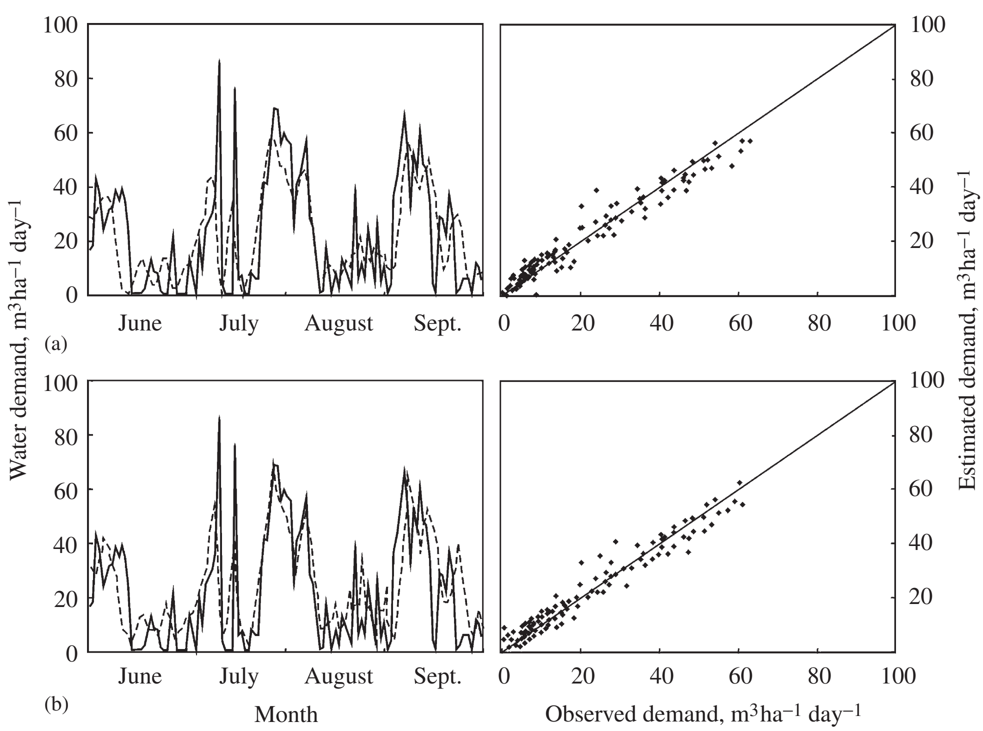

Fig. 5 – One-step-ahead prediction of water demands on the olive farms considered the demand of the two previous days as independent variables and scatterplots comparing observed (—) and estimated (- - -) daily water demands (validation data) (filtered calibration data series): (a) neural network with two neurons in the input layer, six neurons in the intermediate layers and one neuron in the output layer CNN(2:6 s:6 s:1 l); and (b) multiple regression.

two previous days as independent variables and filtered calibration data series. The predictive equation of this regression analysis obtained in calibration period is

$$\hat{Q}_t = 3.98 + 0.81\,Q_{t-1} - 0.28\,Q_{t-2}, \quad (8)$$

where the determination coefficient is $R^2 = 0.98$, the Snedecor statistic was $F = 307.8$, the statistical level of significance was $p<0.001$, and degrees of freedom was 120.

4. Discussion

Irrigation water demand forecasting based on original data recorded by the SIGA system in calibration period for both the neural network and multiple regression models lead to unacceptably high error levels. This lack of goodness of fit was due inaccuracies, errors and noise in the data series which was recorded when the control system was being installed.

Water demand forecasting based on the original filtered series in calibration period provided correct estimations and was acceptable in error terms. With the defined filtering process, a change in the scale of the original water usage series occurred, increasing amplitudes and decreasing frequencies. In this way, sudden events decreased in the parameter to be estimated, while the smoothing effect eliminated the noise and inaccuracies of the original series in the calibration period (Bloomfield, 1976). With this a filtering process in the calibration period the multiple regressions linear models and the neural networks provided similar results. However, the CNNs performed better than the regressions when water demand and climatic variables were considered as input data. Therefore, the CNNs performed better than the linear regressions with input variables that have a high correlation.

The best results were obtained with the models that used water demand from the 2 days prior to forecasting as input or independent variables, when a priori it would seem highly probable that rainfall would condition irrigation water use. This could be due to the low rainfall during the irrigation seasons studied. These seasons were very dry in the Mediterranean area climate. Thus, it will be of interest in future studies to analyse irrigation seasons or other irrigation districts with periods of more abundant rainfall in order to evaluate the sensitivity of the rainfall variable. On the other hand, the results obtained could indicate that rainfall factor, and other climatic variables, are implicitly taken into account in water demand observations (Saporta & Muñoz, 1994).

With the most accurate models proposed (using the demand of the two previous days as independent variables) the estimated curves were displaced with respect to the observed curves. This displacement was because the water demand variable of the previous days had the highest significance in the analysis of variance. In consequence, very good results were obtained in the validation period with the filtered series in the calibration period according to the evaluation magnitudes R^2 and E, as indicators of goodness of fit. However, the E_S magnitude was greater as it performed a point to point evaluation between the observed and estimated values. This characteristic has also been described in papers on the forecasting of other parameters using multiple regressions; for example ARIMA (autoregressive integrated moving average) models and/or neural networks (Park, 1998; Abrahart & See, 2000; Gutiérrez-Estrada *et al.*, 2005; Pulido-Calvo & Portela, 2007). These papers indicate that a time delay could be due to factors not included in the models. In the case of water demand forecasting, this last hypothesis could be valid given that in our study the abovementioned displacement was not observed in the models that used climatic data from previous days.

The training method proposed with internal validation has been shown to be very efficient during generalisation or validation, given that the neural networks responded correctly to patterns that were not employed in the training process, that is, there is no overadjustment of the examples during the calibration phase.

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 multiple regressions. Given that the neural networks are heuristic models, specific rules cannot be given regarding the controlling of convergence, network design or the initialisation and change of weights to resolve a concrete problem. In the specialist literature, only some general guidelines taken from the experience of numerous authors were found. For this reason it was necessary to determine the nature of the problem beforehand; a condition which is not necessary in many other statistical forecasting methods. Thus, in this paper the analyses were begun by calculating the maximum delay of the input variables using the Box–Jenkins methodology following Griñó (1992). In other papers, principal components analyses were 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).

5. Conclusions

In this paper, consumer water demand forecasting systems that can support decision making of the irrigation district administrators are proposed using multiple regressions and computational neural networks (CNNs). Determination coefficients higher to 92%, error magnitudes of E_S lower to 20% and efficiency coefficient E higher to 0.91 have been obtained in the validation period, when water demand of the 2 days prior to forecasting is used as input or independent variables to the neural network or multiple regression, respectively. In this situation, the multiple regressions linear models and the neural approaches provide similar results. However, the CNNs performed better than the regressions when water demand and climatic variables were considered as input data.

Short-term demand modelling can be used as input in methods and/or programs for the management of water-delivery systems in real time. Furthermore, this approach achieves a better fit of the pumped volumes and the real demand of the distribution network, thereby leading to a more rational use of water and energy resources.

It would be of interest to broaden the methodology developed in this paper and implement it in other areas in order to establish a general model for the efficient management of water in irrigated areas. This model constitutes a first step in the analysis and forecasting of water demands and should be of aid in decision-making processes to develop efficient water management policy.

REFERENCES

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. Hydrological Processes, **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. Hydrology and Earth System Sciences, **6**(4), 655–670

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

Agarwal A; Mishra S K; Ram S; Singh J K (2006). Simulation of runoff and sediment yield using artificial neural networks. Biosystems Engineering, **94**(4), 597–613

Anctil F; Rat A (2005). Evaluation of neural network streamflow forecasting on 47 watersheds. Journal of Hydrologic Engineering, **10**(1), 85–88

Allen R G; Pereira L S; Raes D; Smith M (1998). Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56, FAO, Rome

Bloomfield P (1976). Fourier Analysis of Time Series: An Introduction. John Wiley & Sons, New York

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. Hydrological Processes, **16,** 1033–1046

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

Coulbeck B (1988). Computer Control of Water Supply. Leicester Polytechnic, Leicester, UK

Cybenco G (1989). Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems, **2**(4), 303–314

De Vries B; Principe J C (1991). A theory for neural networks with time delays. In: Advances in Neural Information Processing Systems, Vol. 3. Morgan Kaufmann Publishers, California

Doorenbos J; Kassam A H (1979). Yield Response to Water. FAO Irrigation and Drainage Paper 33, FAO, Rome

Doorenbos J; Pruitt W O (1977). Guidelines for Predicting Crop Water Requirements. FAO Irrigation and Drainage Paper 24, FAO, Rome

FAO (1993). Las políticas de recursos hídricos y la agricultura. In: El Estado Mundial de la Agricultura y la Alimentación. [The politics of water resources and the agriculture. In: The World State of the Agriculture and the Feeding.] Organización de las Naciones Unidas para la Agricultura, Rome

French M N; Krajewski W F; Cuykendall R R (1992). Rainfall forecasting in space and time using a neural network. Journal of Hydrology, **137,** 1–31

Givi J; Prasher S O; Patel R M (2004). Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point. Agricultural Water Management, **70**(2), 83–96

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. Aquacultural Engineering, **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. [Water temperature short-term forecasting. Application in aquaculture systems.]. Ingeniería del Agua, **12**(1), 77–92

Hartkamp A D; White J W; Hoogenboom G (1999). Interfacing geographic information systems with agronomic modeling: a review. Agronomy Journal, **91,** 761–772

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

Kadra R; Lamaddalena N (2006). A simulation model to generate the demand hydrographs in large-scale irrigation systems. Biosystems Engineering, **93**(3), 335–346

Kitanidis P K; Bras R L (1980). Real time forecasting with a conceptual hydrological model. 2: Applications and results. Water Resources Research, **16**(6), 1034–1044

Kite G (2000). Using a basin-scale hydrological model to estimate crop transpiration and soil evaporation. Journal of Hydrology, **229,** 59–69

Kite G; Droogers P (2000). Comparing evapotranspiration estimates from satellites, hydrological models and field data. Journal of Hydrology, **229,** 3–18

Kuligowski R J; Barros A P (1998). Experiments in short-term precipitation forecasting using artificial neural networks. Monthly Weather Review, **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 Resources Research, **35**(1), 233–241

Leib B G; Matthews G; Kroeger M (2003). Development of an on-time logger for irrigation systems. Agricultural Water Management, **62,** 67–77

Lorite I J (2003). Evaluación del Riego de la Zona Regable del Genil-Cabra (Santaella, Córdoba): Índices de Calidad del Riego. [Irrigation Evaluation of Genil-Cabra District (Santaella, Córdoba): Indexes of Irrigation Quality.] Analistas Económicos de Andalucía, Málaga, Spain

Lorite I J; Mateos L; Fereres E (2004). Evaluating irrigation performance in a Mediterranean environment. I: Model and general assessment of an irrigation scheme. Irrigation Science, **23,** 77–84

Lorrai M; Sechi G M (1995). Neural nets for modelling rainfall-runoff transformations. Water Resources Management, **9,** 299–313

Malano H; Burton M (2001). Guidelines for Benchmarking Performance in the Irrigation and Drainage sector. International Programme for Technology and Research in Irrigation and Drainage, FAO, Rome

Malano H; Burton M; Makin I (2004). Benchmarking performance in the irrigation and drainage sector: a tool for change. Irrigation and Drainage, **53,** 119–133

Mareels I; Weker E; Ooi S K; Cantoni M; Li Y; Nair G (2005). Systems engineering for irrigation systems: successes and challenges. Annual Reviews in Control, **29,** 191–204

Martín-del-Brío B; Sanz-Molina A 2001. Redes Neuronales y Sistemas Borrosos. [Neural Networks and Fuzzy Systems.] Ra-Ma, Madrid

Mason J C; Tem'me A; Price R K (1996). A neural network model of rainfall-runoff using radial basis functions. Journal of Hydraulic Research, **34**(4), 537–548

Miranda F R; Yoder R E; Wilkerson J B; Odhiambo L O (2005). An autonomous controller for site-specific management of fixed irrigation systems. Computers and Electronics in Agriculture, **48,** 183–197

Moradkhani H; Hsu K; Gupta H V; Sorooshian S (2004). Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. Journal of Hydrology, **295,** 246–262

Nash J E; Sutcliffe J V (1970). River flow forecasting through conceptual models. I: A discussion of principles. Journal of Hydrology, **10,** 282–290

Nayebi M; Khalili D; Amin S; Zand-Parsa Sh (2006). Daily stream flow prediction capability of artificial neural networks as influenced by minimum air temperature data. Biosystems Engineering, **95**(4), 557–567

Ohlsson L (1995). Hydropolitics-Conflicts over Water as a Development Constraint. Zed Books&University Press, London

ONU (1997). Comprehensive Assessment of the Freshwater Resources of the World. United Nations Department for

Policy Coordination and Sustainable Development (DPCSD), Commission on Sustainable Development, New York

Park H H (1998). Analysis and prediction of walleye pollock (Theragra chalcogramma) landings in Korea by time series analysis. Fisheries Research, **38,** 1–7

Pirmoradian N; Sepaskhah A R (2006). A very simple model for yield prediction of rice under different water and nitrogen applications. Biosystems Engineering, **93**(1), 25–34

Pulido-Calvo I; Portela M M (2007). Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. Journal of Hydrology, **332,** 1–15

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. [Short-term forecasting techniques of water demand. Application to agricultural use.]. Ingeniería del Agua, **9**(3), 319–331

Pulido-Calvo I; Roldán J; López-Luque R; Gutiérrez-Estrada J C (2003). Demand forecasting for irrigation water distribution systems. Journal of Irrigation and Drainage Engineering, **129**(6), 422–431

Rodríguez-Díaz J A; Camacho-Poyato E; López-Luque R (2004). Applying benchmarking and data envelopment analysis (DEA) techniques to irrigation districts in Spain. Irrigation and Drainage, **53,** 135–143

Roger L L; Dowla F U (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Resources Research, **30**(2), 457–481

Rumelhart D E; Hinton G E; Williams R J (1986). Learning representations by backpropagation errors. Nature, **323**(9), 533–536

Saporta D; Muñoz M (1994). El consumo en redes de distribución. Predicción diaria de la demanda. In: Mejora del Rendimiento y de la Fiabilidad en Sistemas de Distribución de Agua (Cabrera E; Vela A F eds). [The consumption in water distribution networks. Demand daily prediction. In: Yield and Reliability Improvement in Water Distribution Systems.] Universidad Politécnica de Valencia, Valencia, Spain

See L; Openshaw S (2000). A hybrid multi-model approach to river level forecasting. Hydrological Sciences Journal, **45**(4), 523–536

Sharma V; Negi S C; Rudra R P; Yang S (2003). Neural networks for predicting nitrate-nitrogen in drainage water. Agricultural Water Management, **63**(3), 169–183

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. Journal of Hydrologic Engineering, **5**(2), 145–155

Sumpsi J M; Garrido A; Blanco M; Varela C; Iglesias E (1998). Economía y política de gestión del agua en la agricultura. [Water Management Economy and Politics in the Agriculture.] Ministerio de Agricultura, Pesca y Alimentación&Ed. Mundi-Prensa, Madrid

Tan Y; van Cauwenberghe A (1999). Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Engineering Applications of Artificial Intelligence, **12**(1), 21–25

Thirumalaiah K; Deo M C (1998). River stage forecasting using artificial neural networks. Journal of Hydrologic Engineering, **3**(1), 26–32

Thirumalaiah K; Deo M C (2000). Hydrological forecasting using neural networks. Journal of Hydrologic Engineering, **5**(2), 180–189

Tokar A S; Johnson P A (1999). Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering, **4**(3), 232–239

Tokar A S; Markus M (2000). Precipitation-runoff modeling using artificial neural networks and conceptual models. Journal of Hydrologic Engineering, **5**(2), 156–161

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

Van Aelst P V; Ragab R A; Feyen J; Raes D (1988). Improving irrigation management by modelling the irrigation schedule. Agricultural Water Management, **13,** 113–125

Ventura S; Silva M; Pérez-Bendito D; Hervás C (1995). Artificial neural networks for estimation of kinetic analytical parameters. Analytical Chemistry, **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. Journal of Chemical Information and Computer Sciences, **37**(2), 287–291

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. Journal of Irrigation and Drainage Engineering, **123**(4), 285–292

Zhang M; Fulcher J; Scofield R A (1997). Rainfall estimation using artificial neural network group. Neurocomputing, **16,** 97–115