2012

Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series

JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT-ASCE, RESTON, v. 138, n. 3, supl. 1, Part 3, pp. 245-256, MAY-JUN, 2012
http://www.producao.usp.br/handle/BDPI/41096

Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series

Frederico Keizo Odan[1] and Luisa Fernanda Ribeiro Reis[2]

Abstract: This paper addressed the problem of water-demand forecasting for real-time operation of water supply systems. The present study was conducted to identify the best fit model using hourly consumption data from the water supply system of Araraquara, São Paulo, Brazil. Artificial neural networks (ANNs) were used in view of their enhanced capability to match or even improve on the regression model forecasts. The ANNs used were the multilayer perceptron with the back-propagation algorithm (MLP-BP), the dynamic neural network (DAN2), and two hybrid ANNs. The hybrid models used the error produced by the Fourier series forecasting as input to the MLP-BP and DAN2, called ANN-H and DAN2-H, respectively. The tested inputs for the neural network were selected literature and correlation analysis. The results from the hybrid models were promising, DAN2 performing better than the tested MLP-BP models. DAN2-H, identified as the best model, produced a mean absolute error (MAE) of 3.3 L/s and 2.8 L/s for training and test set, respectively, for the prediction of the next hour, which represented about 12% of the average consumption. The best forecasting model for the next 24 hours was again DAN2-H, which outperformed other compared models, and produced a MAE of 3.1 L/s and 3.0 L/s for training and test set respectively, which represented about 12% of average consumption. **DOI:** 10.1061/(ASCE)WR.1943-5452.0000177.

CE Database subject headings: Forecasting; Artificial intelligence; Fourier series; Hybrid methods; Water demand; Water supply.

Author keywords: Forecasting; Artificial intelligence; Fourier series; Hybrid methods; Water demand; Water supply.

Introduction

Water supply systems endeavor to meet the daily needs of various consumers, ranging from households to industrial consumers. Moreover, the main objective of water companies is to supply quality water at a reasonable price. Water demand, however, varies throughout the day, weekdays, months, and the year. It is also influenced by socioeconomic factors such as economic level, population characteristics, the size of homes, commercial, and industrial establishments; climatic and meteorological factors such as air temperature, precipitation; and the cost of water supply.

Aside from the variability of water demand, it has a tendency to increase over the years. However, water is a finite resource and, just like energy, the options for new sources are increasingly more limited and expensive. In this context, rationalizing water use has become necessary—not only to minimize waste and loss of water, but also to minimize the energy spent on water supply, through awareness campaigns, rehabilitation of water networks, and streamlining the operation of water supply systems.

The advent of computer use and the supervisory control and data acquisition (SCADA) has made it possible to remotely monitor and control water supply systems. Consequently, the online rationalization operation can be performed with the support of a hydraulic simulator, to reproduce the behavior of the water supply network, which—combined with optimization models and water-demand forecasting—can provide the optimized operation to satisfy the forecasted demand.

According to Jain et al. (2001), there are two types of water-demand forecasting: "long-term forecasting and short-term forecasting. Long-term forecasting is useful in planning and design, and making extension plans for existing water systems, while the short-term forecasting is useful in the efficient operation and management of an existing water system." According to Jain et al. (2001), short-term water-demand forecasts can be applied to settings such as the management of an urban water distribution system, water conservation programs during drought conditions, and water pricing policy assessment. They emphasize the need for accurate short-term water demand forecasts for improving the performance of an urban water system.

Zhang (2001) pointed out that there are basically two types of approaches to modeling time-series: linear and nonlinear. The linear approach employs: time-series regression, exponential smoothing, and autoregressive integrated moving average (ARIMA). In the nonlinear one, there are nonlinear regression models, bilinear models, and the threshold autoregressive model (AR).

The linear statistical forecasting methods were widely used because they are easy to develop and implement, in addition to being simple to understand and interpret. However, in the real world, data have varying degrees of nonlinearity, which may not be adequately handled by linear applications.

To improve the forecasting performance, many nonlinear methods were proposed. However, forecasting by nonlinear methods was a complex task "since there are many possible nonlinear patterns and one specific form may not capture all the nonlinearities in the data" (Zhang 2001).

Maidment et al. (1985) developed a time-series model forecast for daily municipal water use as a function of rainfall and air

[1]Doctoral student, Dept. of Hydraulics and Sanitary Engineering, Univ. of São Paulo, Av. Trabalhador Sãocarlense, 400. CEP 13566-590, São Carlos, São Paulo, Brazil (corresponding author). E-mail: frederico_keizo@yahoo.com.br

[2]Professor, Dept. of Hydraulics and Sanitary Engineering, Univ. of São Paulo, Av. Trabalhador Sãocarlense, 400. CEP 13566-590, São Carlos, São Paulo, Brazil. E-mail: fernanda@sc.usp.br

Note. This manuscript was submitted on June 7, 2010; approved on August 4, 2011; published online on April 16, 2012. Discussion period open until October 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 138, No. 3, May 1, 2012. ©ASCE, ISSN 0733-9496/2012/3-245–256/$25.00.

temperature. The model did not consider the effect of weekdays, which is usually a determining factor to represent a constant average process. Conversely, Smith (1988) considered the weekday effect, proposing a conditional autoregressive model, to forecast the municipal water demand.

Shvartser et al. (1993) combined the pattern recognition technique with the time-series analysis to forecast the water demand. The model performed the pattern recognition for three segments of the daily cycle, ascending, oscillating, and descending phase, as "states" of the water demand curve, which can be described as successive states of a Markov process. Then the transition probabilities between the states were calculated and each segment fitted by an ARIMA model. The model obtained was applied to Israel's water supply, which, according to the author, produced satisfactory results.

Other examples of short-term water-demand forecasting models on the basis of time-series analysis include Valdes and Sastri (1989); Miaou (1990); and Jowitt and Xu (1992). However, the online forecast research which considered the consumption data as a dynamic sequence of values with parameters constantly updated in time represents a real breakthrough. In this regard, the models developed by Perry (1981) and Zahed (1990) stand out among those presented in the literature. Perry (1981) decomposed the consumption sequence in a series of harmonics and residual flows, which were fitted by an ARMA model. Both the harmonic coefficients and the ARMA coefficients, which were dynamically updated, were obtained from a time series and used for a 24-h forecast. Zahed (1990), on the basis of Perry's model (Perry 1981), proposed and applied two-hour water demand forecasting in the operation of part of the metropolitan pipeline system of the city of São Paulo, Brazil. One was on the basis of the Taylor series and the other on the Fourier series. As important premises of the former model, he stated the elimination of the weekday effect by adjusting the harmonic coefficients to the consumption of a week and by an hourly update of them. Additionally, recent data were used to refine the forecast instead of auxiliary models.

Odan et al. (2009) performed comparative studies between the Fourier series model proposed by Zahed (1990) and the *k*-nearest neighbors to forecast the water demand for the water supply sector in the city of São Carlos, Brazil. The first model consisted of fitting a harmonic equation to a time-series data and extrapolating it for forecasting future values, with the harmonic coefficients updated on a daily basis within 7-day cycles. The second model stored the sequences to be used as the basis from which the closest ones to the sequence to be forecasted were selected to obtain the future value. Several configurations of both models were studied, and then a comparison was made between the best of each. The models produced good results, with a mean absolute error (MAE) of around 13.5% for predicting the next hour consumption.

Recently, the artificial neural networks (ANNs) were being proposed as efficient tools for modeling and forecasting water demand. Crommelynck et al. (1992) were some of the pioneers to use the multilayer ANN and the back-propagation training algorithm for modeling hourly and daily water demand. They observed that the ANN presented results at least comparable with those from statistical methods.

Jain et al. (2001); Jain and Ormsbee (2002); and Bougadis et al. (2005) concluded that the ANN performs better than regression and time-series analysis. Jain et al. (2001) observed during the investigation of the inputs for the ANN that meteorological factors have significant influence on the short-term forecast, especially the air temperature and the occurrence of rainfall, the latter being more significant than the amount of rain itself. The ANN model developed by Jain and Ormsbee (2002) for daily water demand was a function of the daily water consumption from the previous day and the daily maximum air temperature of the current day. Bougadis et al. (2005), unlike Jain et al. (2001), found that water demand on a weekly basis was better correlated with the rainfall amount than the occurrence of rainfall.

Zhang et al. (2006) forecasted the short-term demand for the city of Louisville, Kentucky, by using ANN. The work differed from others by separating the year into winter and summer season, wherein the winter uses only the daily time series as input, because the weather factors had little influence in that season. For the summer, six ANN models were developed, which used the following as input: the daily time-series demand; difference between weekday and weekend; and the weather factors, such as temperature, rainfall, relative humidity, dew point, and wind. The best summer forecast model used maximum temperature, relative humidity, rainfall, and historical demand to forecast the demand for the next two days.

Adamowski (2008) forecasted the daily maximum temperature on the basis of the historical demand of Canada, confirming the findings of Jain et al. (2001), which indicated the demand is better correlated with rainfall occurrence than the amount of rainfall itself. The author compared 39 multiple linear regression models, 9 ARIMA models, and 39 ANN models, the last one having produced better results than the other models. The best ANN model was obtained considering previous daily maximum demand, temperature of the current day and the day before, and rainfall occurrence of the last 5 days.

Ghiassi et al. (2008) developed a dynamic artificial neural network (DAN2) to forecast the hourly water demand of a city in California that surpassed the ARIMA models in its performance. The authors observed that modeling weather variables enhanced the performance of DAN2 model, but even without them, the results were considered excellent.

It was observed that among online forecast models, those which constantly update their parameters proved to be especially useful to incorporate the latest consumption data, such as the ones developed by Perry (1981) and Zahed (1990).

To improve performance of the forecasting model, some authors (Zhang 2003; Athanasiadis et al. 2005; Jain and Kumar 2007; Pulido-Calvo and Gutierrez 2009) developed hybrid forecasting models on the basis of two or more different models.

Zhang (2003) combined ANN and ARIMA models to take advantage of the unique strength of each model, nonlinear and linear modeling, respectively. The methodology of the hybrid model consisted of two steps: first, the ARIMA model was used to model the linear part of the data, then the ANN was used to model the residuals between the ARIMA model and the observed data. The hybrid model was applied to three benchmark data series (Wolf's sunspot data, the Canadian lynx data, and the British pound/U. S. dollar exchange rate data). The overall forecasting performance of the hybrid model was improved by the combination of both models.

Athanasiadis et al. (2005) developed a hybrid model to forecast residential water demand called DAWN, which integrated an agent-based social model for the consumer with a conventional econometric model, to simulate the residential water supply chain; in other words, it forecasted water demand through the iterative interaction between the consumer, represented by the econometric model; the water supply manager, which determined the water's price on the basis of total demand calculation; and meteorological data.

Jain and Kumar (2007) combined ANN and ARIMA models with a filtered data series to forecast monthly streamflow. Three categories of data were tested: raw data and two sets of filtered data, one with detrended data and the other with detrended deseasonalized data. The detrended data have their long-term trend removed; the deseasonalized data do not have their seasonal-variation

J. Water Resour. Plann. Manage. 2012.138:245-256.

component. The ANN with the detrended deseasonalized data produced much more accurate results than its ARIMA counterpart.

Pulido-Calvo and Gutierrez (2009) combined the ANN and fuzzy logic to forecast daily water demand at irrigations districts located in Andalucía, Spain. Univariate and multivariate models were compared, in which the former performed better considering the demands and the maximum temperatures of the two previous days. It was found that the multivariate hybrid models performed significantly better than ANN alone.

Data

The consumption data used in this study was from the water supply subsector of the city of Araraquara, Brazil, called Iguatemi High Zone (Iguatemi HZ), supplied by the Autonomous Department of Water and Sewage (DAAE). The subsector consumed about 27 m^3/h. The hourly data of temperature and relative humidity was obtained from the São Paulo State Environmental Company (CETESB). The time frame of the data used was from September 27th, 2008, to July 14th, 2009, totaling almost 7,000 h of record.

Before the use of the observed values, a preprocessing procedure was performed on missing and faulty data, resulting from failures in the records or presence of values greater than or less than twice the absolute value of standard deviation (i.e., those outside the 95% confidence interval), considering a normal distribution of data. It was observed that missing and faulty data did not exceed 5.5% of the observed values. After identifying the missing and faulty data, the data preprocessing was performed using a tool developed by Oba et al. (2003), called Bayesian principal component analysis (BPCA), freely distributed on MATLAB or Java language (JBPCA 2004).

The tool was originally developed to complete missing DNA sequences through the estimation of those missing values, and was recommended when there was a large amount of data. The BPCA was on the basis of principal component regression, Bayesian estimation, and a repetitive expectation-maximization (EM) algorithm. The method fits a probabilistic model on the basis of the principal component analysis (PCA) through Bayesian estimation. An iterative variational Bayesian algorithm was used to estimate the posterior distribution of the model parameters and faulty data, until convergence was reached. The BPCA technique used in that study (i.e., Oba et al. 2003) for missing data preprocessing was found to be quite good even when 40% of data were missing, surpassing the performance of the models on the basis of k-nearest neighbors (kNN) and singular value decomposition (SVD). The tool was also used by Lim and Zainuddin (2008), which compared the BPCA with the local least square imputation (LLS) and the radial basis function (RBF) network. They concluded that BPCA had the best results according to the normalized root mean squared error (NRMSE) criterion, and the LLS produced competitive results when compared with BPCA, followed by RBF, which had the lowest performance.

After data preprocessing, the correlation analysis was performed to identify the input variables to forecast the water demand. The autocorrelation between the lagged consumption series of 1, 2, 3, 4, 24, and 168 h (7 days x 24 h/day = 168h) and the last 24 h mean consumption were analyzed as shown in Table 1. Correlation between consumption from Iguatemi HZ with temperature, relative humidity, the hour and reservoir level are presented in Table 2.

Table 1 presents the autocorrelation calculated up to a lag of 4 h, beside the lag of one day (lag of 24 h) and 1 week(lag of 168 h). The correlation for 1 day and 1 week lags were quite significant, given the cyclical behavior of water consumption.

Table 1. Autocorrelation between Lagged Consumption of 1, 2, 3, 4, 24, 168 h, and Last 24-h Mean Consumption for Subsector Iguatemi HZ

Variable						
C						Cmed (24 hour-mean)
Lag						lag
1	2	3	4	24	168	0
0.89	0.73	0.53	0.32	0.87	0.89	0.19

Table 2. Correlation between Consumption from Iguatemi HZ with Temperature, Relative Humidity, Hour, and Reservoir Level

Variable			
Temperature	Relative Humidity	Hour	Reservoir Level
0.91	0.99	−0.15	−0.23

The weather variables (temperature and relative humidity) were selected on the basis of previous studies by Jain et al. (2001), Adamowski (2008), and Ghiassi et al. (2008), who concluded that such variables improve the water-demand forecasting.

The hour of consumption was considered in the correlation analysis given the cyclic behavior of water-demand during the day. The level of the reservoir that supplies each subsector was chosen because the consumption may be influenced by pressure of the network, which was a function of the head of the reservoir in gravity feed network.

Model Development

Implementing demand forecasting involved selecting the forecasting model, defining the best structures for the models and the input variables, and defining the evaluation criteria. This study chose to apply the ANNs primarily on the basis of their capability to match or even surpass the other forecasting models as reported in the specialized literature. Hence, the ANN multilayer perceptron with back-propagation training algorithm (MLP-BP), the DAN2, and the hybrid neural networks were used for this research. The ANN MLP-BP was used to serve as a basis of comparison, given its wide use and successful testing. DAN2 is a relatively simple network to implement with the potential for excellent results in predicting urban water demand (Ghiassi et al. 2008). The hybrid neural network proposed in this paper shall be tested against the other two reasonably good models to establish its predictive capabilities.

The developed water demand forecasting model was built to help with problems that require prediction. In this paper, estimates were made 24 h ahead at each time step (1 h), whose values were used to perform the operation of a water network for the next 24 h.

The demand forecasting 24 h ahead was performed in multiple steps. For example, the 24 h prediction for the fourth model on the basis of MLP-BP (ANN 4) was presented schematically in Fig. 1. It shows that the predicted consumption value for the first hour ahead ($C(t+1)$), P1, was a set of 5 consumptions lagged by 4 hours ($C(t-4)$), 3 hours ($C(t-3)$), 2 hours ($C(t-2)$), 1 hour ($C(t-1)$) and (0) hour ($C(t)$). Similarly, the lagged consumption set $[C(t-3), C(t-2)), C(t-1), (C(t)]$ together with P1 from the previous step were used to estimate the consumption second hour ahead ($C(t+2)$), P2. Thus, it was in this stepwise manner that the forecast progressed: the estimates of 2 hours ahead was the forecast of 1 h ahead; the forecast of 3 h ahead was on the basis of forecasts of 1 and 2 ahead, from previous two steps, and so on.

J. Water Resour. Plann. Manage. 2012.138:245-256.

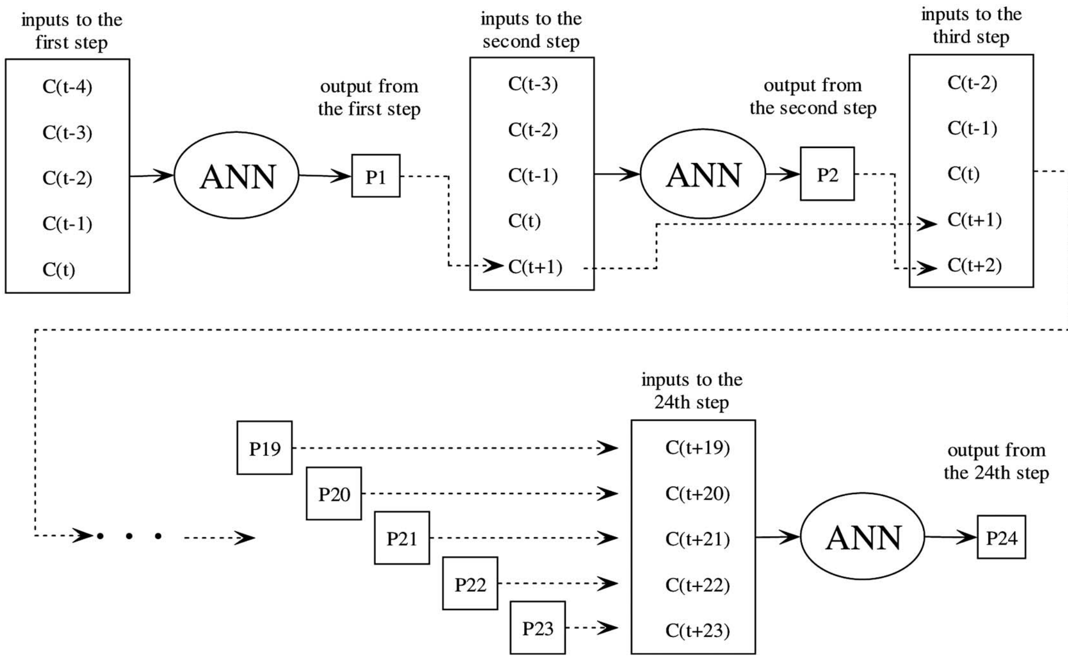

Fig. 1. Scheme of the 24-h demand forecasting

Initially, evaluations of the first hour forecasting demand were done to select models for the prediction of the next 24 hour, which was performed in a multi-step mode, so that repeated one-step ahead predictions were made up to the desired horizon.

The evaluation of the ANNs was done using the k-fold cross validation, to make results more independent from the training and test set choice. The method was comprised of dividing the data set in k subsets for this purpose. The evaluation method was applied k times. Each time, one of k sets was used as a test set, and the remaining subsets were used for training. The final evaluation (error or correlation criteria) is the average of the k trials performed. In this study, k was set to ten (10), which is a commonly used value.

ANN MLP-BP

ANN models are massive parallel-distributed information processing systems with certain performance characteristics resembling biological neural networks of the human brain (Haykin 1999). According to Zhang et al. (2006), an ANN is basically composed of several layers of nodes, and that structure is denominated architecture, as the three-layer ANN illustrated in Fig. 2, where **X** is a system input vector composed of a number of input variables, and **Y** is system output vector consisting of a number of output variables. The weight matrix **W**, composed of individual weights for each neuron, connects input, hidden and output neurons. A schematic diagram of a generic jth node is illustrated in Fig. 3.

The output value y_j was obtained from the transfer function f, with respect to the sum of the inner product of vector X and W_j minus bias b_j, as represented by the following equation:

$$y_j = f(X \cdot W_j - b_j)$$

The sigmoid function was used as transfer (or activation) f for neurons in hidden layer and output layer as well

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

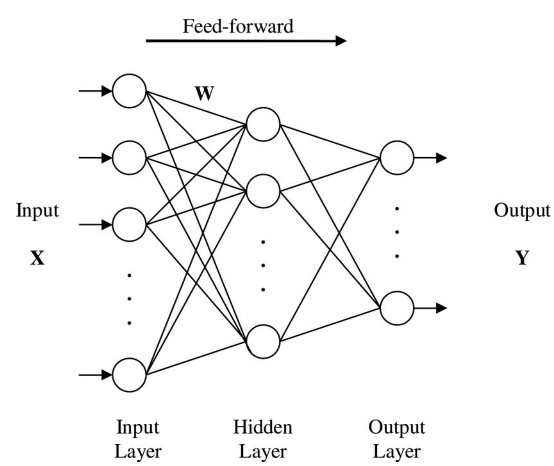

Fig. 2. Schematic architecture of a three-layered feed-forward ANN

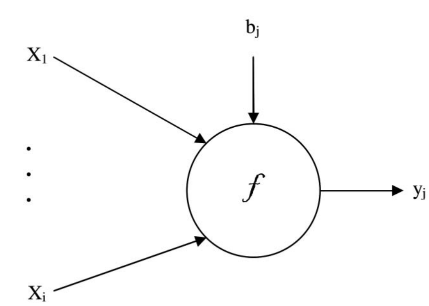

Fig. 3. A generic jth node

To generate the output vector **Y**, which approaches a desired vector Y_d, a training process was applied to modify the weight matrix **W** and bias vector **B** such that one minimizes an objective function, for example, the sum of the squared error (Min E)

$$\mathrm{Min}\, E = \sum e^2 = \frac{1}{2}\sum (y - y_d)^2$$

The training process consisted of solving the weight adjusting problem by optimization. The most used training algorithm was the back-propagation (BP), on account of its simplicity and easy implementation. For the BP algorithm, supposing $g(w) = \nabla E(w)$ is the gradient vector, then the steepest descent move is described as

$$w_{k+1} = w_k - \lambda g(w_k)$$

where λ is the user-specified constant step size, known as the learning rate. Once the network is trained, the weights are optimized and the ANN model can be used for prediction.

The learning rate (λ) can take on any value between 0 and 1, and it plays an important role in ANN training. To select λ, the first architecture of MLP-BP (ANN 1 from Table 3) was used. Through the variation of λ (0.1, 0.2, …, 1.0), it was found that $\lambda = 0.2$ produced the best results in terms of error criteria used in this paper.

For most ANNs, it was necessary to normalize the data X to the interval [0, 1] for sigmoid transfer functions, or to the interval $[-1, 1]$, when using hyperbolic transfer function. However, it was generally recommended that data should not be normalized to the limits of activation function, because these nonlinear functions have asymptotic limits, reaching the limits only for infinite values (Tang and Fishwick, 1993), and also in the case the future values are slightly outside the range reached by the observed data so far. A linear normalization for the interval [0.25, 0.75] was used in this study

$$x_{\mathrm{nor}} = (b - a)\frac{x_{\mathrm{obs}} - x_{\mathrm{min}}}{x_{\mathrm{max}} - x_{\mathrm{min}}} + a$$

where x_{nor} = the normalized value; x_{obs} = the original value; x_{min} and x_{max} = respectively, the minimum and maximum values of the variable; and a and b = limits of the activation function, 0.25 and 0.75, respectively.

The ANN Toolbox 0.4.2.2, developed by Hristev (2000) and updated by Alan Cornet, was used to model the ANN MLP-BP. It is written in computer language Scilab 5.3, a free alternative version of MATLAB. The choice of the architecture of the ANN was done by a trial and error process and involved choosing the number of layers and neurons on each layer. In this study, only one hidden layer was used following Cybenko (1989) and Funahashi (1989) who showed that any continuous function can be reasonably approximated by an ANN MLP-BP with a single hidden layer using a logistic activation function. The number of neurons at the hidden layer was varied from 1 to 8 neurons. The training will be stopped when the maximum number of epochs (cycles) previously established is reached.

Table 3 exhibits the input and output for ANN MLP-BP, for which the input data was selected the correlation analysis previously performed in the previous section, for Iguatemi HZ. The ANN 1 to 9 were spatially univariate models, and ANN 10 is a spatially multivariate model, where the Iguatemi HZ input is combined 2×2 with the other subsectors, Iguatemi Low Zone, Martinez and Eliana, and also combined with all inputs from these subsectors. For the multivariate models, the inputs and outputs Iguatemi HZ, Iguatemi Low Zone, Eliana and Martinez are denoted as IHZ, ILZ, E, and M, respectively.

The nomenclature used for the input of the forecasting models corresponds to the variables indicated in the correlation analysis (Tables 1 and 2), where C denotes consumption data, T temperature, and RH relative humidity. The term in parenthesis refers to time before forecast, [i.e., (t) is the present time and $(t-1)$, $(t-2), \ldots, (t-n)$ are respectively 1 h, 2 h, …], n hours before the present time.

DAN2

The basic idea of DAN2 was to learn and accumulate knowledge in each layer, propagating and adjusting the knowledge in each added layer, until a stopping criterion is reached. Before continuing the study, the input matrix was defined as

$$X = \{X_i, i = 1, \ldots, n\}$$

composed of n independent patterns of m attributes, in other words,

$$Xi = \{x_{ij}, j = 1, \ldots, m\}$$

as shown in Fig. 4 (input matrix X).

Following Ghiassi and Saidane (2005), the model used normalization to avoid any problem with different data metrics. The normalization process defined a reference vector

$$R = \{r_j, j = 1, \ldots, m\}$$

Table 3. ANN MLP-BP Models and their Respective Inputs and Outputs for the Univariate Model of Iguatemi HZ (ANN 1 to 9) and Spatially Multivariate Model (ANN 10)

Model	Input	Output
ANN 1	$C(t)$, $C(t-1)$, $C(t-2)$	$C(t+1)$
ANN 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$	$C(t+1)$
ANN 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1)$
ANN 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-4)$	$C(t+1)$
ANN 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$	$C(t+1)$
ANN 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1)$
ANN 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $T(t)$	$C(t+1)$
ANN 8	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $RH(t)$	$C(t+1)$
ANN 9	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$
ANN 10-1	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{ILZ}}$, $C(t-1)_{\mathrm{ILZ}}$, $C(t-2)_{\mathrm{ILZ}}$	$C(t+1)_{\mathrm{IHZ}}$
ANN 10-2	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{E}}$, $C(t-1)_{\mathrm{E}}$, $C(t-2)_{\mathrm{E}}$	$C(t+1)_{\mathrm{IHZ}}$
ANN 10-3	$C(t)_{\mathrm{IHZ}}$, F, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{M}}$, $C(t-1)_{\mathrm{M}}$, $C(t-2)_{\mathrm{M}}$	$C(t+1)_{\mathrm{IHZ}}$
ANN 10-4	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{ILZ}}$, $C(t-1)_{\mathrm{ILZ}}$, $C(t-2)_{\mathrm{ILZ}}$, $C(t)_{\mathrm{E}}$, $C(t-1)_{\mathrm{E}}$, $C(t-2)_{\mathrm{E}}$, $C(t)_{\mathrm{M}}$, $C(t-1)_{\mathrm{M}}$, $C(t-2)_{\mathrm{M}}$	$C(t+1)_{\mathrm{IHZ}}$

J. Water Resour. Plann. Manage. 2012.138:245-256.

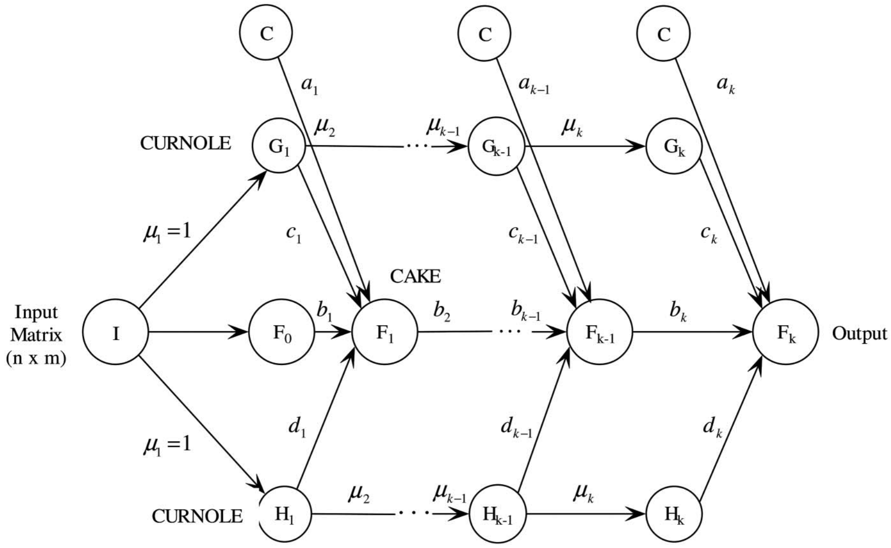

Fig. 4. Generic architecture of DAN2

and projected each observation i over R. The normalization defined and angle α_i, between observation data vector i, X_i, and the reference vector R.

Unlike traditional ANNs, DAN2 had a different architecture, in which the number of nodes in the hidden layer was fixed at four, as shown in Fig. 4. The first node is a constant C. The second node, which accumulated acquired knowledge, is called current accumulated knowledge element (CAKE), represented by F_k. Finally, the third and fourth nodes are called current residual nonlinear element (CURNOLE), responsible for capturing nonlinear residual elements, and represented by G_k and H_k.

G_k and H_k nodes represented a trigonometric transfer function using the extraction of nonlinear components in a similar manner to the Fourier series. The DAN2 differed from other neural networks by using the training set simultaneously and repetitively at each layer. Furthermore, the connections were many-to-one, reducing the network complexity. The greatest advantage of this neural network was the fact that it was not necessary to define the number of neurons in the hidden layers. Therefore, this network can be classified as constructive. The hidden layers were added iteratively, until it satisfied a stopping criterion. As proposed by the authors, three stopping criteria were used: maximum number of iterations, underfitting criterion, and overfitting criterion. The second criterion avoided the model to be insufficiently trained, and the third avoided excessive training, (i.e., it provides a good fit for the training set and for the test set). The limits should be determined experimentally for each problem.

Underfitting was avoided following the criterion expressed as

$$\varepsilon_1 = (SSE_k - SSE_{k+1})/SSE_k \leq \varepsilon_1^*$$

where SSE = the sum of squared error; index k and $k+1$ = the iteration; ε_1 and ε_1* = respectively, the calculated error and error limit.

Similar criterion for overfitting is

$$\varepsilon_2 = |MSE_T - MSE_V|/MSE_T \leq \varepsilon_2^*$$

where MSE = the mean square error, index T and V = respectively for training and test set; ε_2 and ε_2^* = the overfitting errors. In this study, the number of iterations was limited to 15 and errors limited as: $\varepsilon_1^* = 0.0001$ and $\varepsilon_2^* = 0.25$.

The training began with a special CAKE F_0, responsible for extracting the linear components through a linear combination of input data, as

$$F_0(X_i) = a_0 + \sum_{j=1}^{m} b_{0j} x_{ij}$$

where a_0 and b_{0j} = coefficients determined by linear regression.

In case of an unacceptable calculated error, a new CAKE F_k was added

$$F_k(X_i) = a_k + b_k F_{k-1}(X_i) + c_k \cos(\mu_k \cdot \alpha_i) + d_k \sin(\mu_k \cdot \alpha_i)$$

where a_k, b_k, c_k, d_k and μ_k = constants, and

$$G_k = \cos(\mu \cdot \alpha_i) \qquad H_k = \sin(\mu_k \cdot \alpha_i)$$

Angles α_i between vector R and X_i were obtained according to the following equations:

$$\alpha_i = Arc\cos(R * X_i)_N \quad for\ i = 1, \ldots, n$$

where

$$(R * X_i)_N = \frac{(R * X_i)}{\|R\| * \|X_i\|} \qquad (R * X_i) = \sum_{j=1}^{m} r_j x_{ij}$$

By default, $r_j = 1$, for $j = 1, \ldots, m$, and

$$\|R\| = \sqrt{\sum_{j=1}^{m} r_j^2} \qquad \|X_i\| = \sqrt{\sum_{j=1}^{m} x_{ij}^2}$$

The coefficients a_k, b_k, c_k, d_k, and μ_k can be determined by an optimization algorithm, minimizing, for example, the SSE

$$\mathrm{SSE}_k = \sum_{i=1}^{n} \left[F_k(X_i) - \hat{F}(X_i) \right]^2$$

where F_k and $\hat{F}$ = respectively, the predicted and the observed consumption value.

Diverse techniques were tested to optimize the weights and coefficients of DAN2, such as the quasi-newton method, genetic algorithm and also the Levenberg-Marquardt algorithm. All of

J. Water Resour. Plann. Manage. 2012.138:245-256.

them produced very similar results; however, the implementation and handling was easier with the Levenberg-Marquardt algorithm.

Table 4 shows the input and output vectors of DAN2 for the univariate model of Iguatemi HZ, which were on the basis of the results of ANN MLP-BP.

Hybrid Neural Network

Analyzing the literature, one can see the need to create a forecasting model that considers not only weather variables, but also the seasonality found in the time series. Some authors proposed to use models dividing the year in wet and dry seasons, or summer, autumn, winter, and spring. This work considered the time series as a dynamic sequence of values, with parameters continuously updated, as in Perry (1981) and Zahed (1990). The hybrid neural network forecasting model developed associated the Fourier series (FS) to the ANN. This association consisted of modeling by ANN the residuals of the FS forecast, (i.e., the difference between the FS forecast and the observed consumption). The joint ANN MLP-BP and FS model is called ANN-H, and the joint DAN2 and FS model is termed "DAN2-H". Table 5 exhibits the input and output vectors for the hybrid neural networks ANN-H and DAN2-H for the univariate case of Iguatemi HZ.

Forecasting Model on the Basis of FS

FS forecasting model fitted a harmonic equation to a time-series data and uses it to extrapolate for future forecasting values. The harmonic coefficients were updated on a daily basis with 7-day cycles. The dynamics of the model resides in updating the coefficients a_k and b_k, which modify the forecast equation, expressed as

$$\text{Fourier}(x_j) = a_0 + \sum_{k=1}^{NH}\left[a_k \cdot \cos\left(k \cdot \frac{\pi}{N_{\text{hour}}} \cdot j\right) + b_k \cdot \sin\left(k \cdot \frac{\pi}{N_{\text{hour}}} \cdot j\right)\right] \quad NH < N_{\text{hour}}$$

where a_k and b_k = coefficients to be determined by the least square method; a_0 = the average value of the function in the used range; NH = the number of harmonics of the series; $2N_{\text{hour}}$ = the period in hours, which coincides with the range $0 - 2\pi$; j = the order number given in the series and $x_j = \frac{\pi}{N_{\text{hour}}} \times j$, which represents the order number in the range $0 - 2\pi$.

The least square method was used to calculate the coefficients a_0, a_k, and b_k expressed as follows (Humes et al. 1984):

$$a_0 = \frac{1}{2N_{\text{hour}}}\sum_{j=1}^{2N_{\text{hour}}} f(x_j) \qquad a_k = \frac{1}{N_{\text{hour}}}\sum_{j=1}^{2N_{\text{hour}}} f(x_j)\cos\left(k\frac{\pi}{N_{\text{hour}}}j\right)$$

$$b_k = \frac{1}{N_{\text{hour}}}\sum_{j=1}^{2N_{\text{hour}}} f(x_j)\sin\left(k\frac{\pi}{N_{\text{hour}}}j\right)$$

where $f(x_j)$ = the jth value of consumption observed, which is an averaged calculated value on an hourly basis.

Thus, with a weekly cycle and hourly discretization of the consumption, $2N_{\text{hour}} = 24$ hours/day $\times$ 7 days $= 168$ h. For each harmonic degree k, two summations were needed to define the coefficients a_k and b_k. Unlike Zahed (1990), no correction factor was applied to the forecast consumption values to avoid introducing unnecessary fluctuations not inherent to the FS itself.

Evaluation Criteria

Precision of prediction models can be measured either, in terms of the errors in the predicted values in relation to those observed or through correlation existing between them. The following evaluation criteria were used:

MAE

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Pearson (r)

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

where N = the number of evaluated data; x_i and y_i = respectively, the observed and forecasted consumption; $\bar{x}$ and $\bar{y}$ = respectively, the mean of the observed and forecasted values.

At each time step, the next 24-h forecasting was compared with their respectively observed consumption values to evaluate the

Table 4. Input and Output Vectors of DAN2 for the Univariate Model of Iguatemi HZ

Model	Input	Output
DAN2 1	$C(t)$, $C(t-1)$, $C(t-2)$	$C(t+1)$
DAN2 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1)$
DAN2 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$	$C(t+1)$
DAN2 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1)$
DAN2 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $T(t)$	$C(t+1)$
DAN2 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and RH(t)	$C(t+1)$
DAN2 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $T(t)$ and RH(t)	$C(t+1)$

Table 5. ANN-H Models for the Univariate Model of Iguatemi HZ

Model	Input	Output
ANN-H 1	$C(t)$, $C(t-1)$, $C(t-2)$, Res(t), Res$(t-1)$, Res$(t-2)$	$C(t+1)$
ANN-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-168)$	$C(t+1)$
ANN-H 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-24)$	$C(t+1)$
ANN-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$	$C(t+1)$
ANN-H 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$ and $T(t)$	$C(t+1)$
ANN-H 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$ and $RH(t)$	$C(t+1)$
ANN-H 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$

J. Water Resour. Plann. Manage. 2012.138:245-256.

model performance according to the presented evaluations criteria. At the end of prediction process, an average value was obtained from all 24 h forecasts performed to obtain the average value of MAE and Pearson (r).

Results and Discussion

The performance of various architectural versions of the four different ANNs models for water consumption prediction, elaborated in the previous section, was presented in terms of the adopted evaluation criteria. These ANN models are as follows: ANN MLP-BP, the DAN2, and hybrid neural networks ANN-H and DAN2-H. Model architecture is expressed as: number of input variables, number of neurons in the hidden layers, and number of output variables.

Tables 6–10 present the results of the forecasts for the next hour obtained by the four models and Table 11 synthesizes their performance indices.

Table 6 presents results corresponding to the tests performed with ANN MLP-BP for which the number of neurons in the hidden layer varied from 1 to 8, for which only the best architecture is shown. It can be observed that the model ANN 8, which used $C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, and $\mathrm{RH}(t)$,

Table 6. DAN2-H Models for the Univariate Model of Iguatemi HZ

Model	Input	Output
DAN2-H 1	$C(t)$, $C(t-1)$, $C(t-2)$, Res(t), Res$(t-1)$, Res$(t-2)$	$C(t+1)$
DAN2-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-168)$	$C(t+1)$
DAN2-H 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-24)$	$C(t+1)$
DAN2-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$	$C(t+1)$
DAN2-H 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$ and $T(t)$	$C(t+1)$
DAN2-H 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, $Res(t-168)$ and $RH(t)$	$C(t+1)$
DAN2-H 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$

Table 7. ANN MLP-BP Results

				MAE (L/s)		r	
Model	Input	Output	Architecture	Train	Test	Train	Test
ANN 1	$C(t)$, $C(t-1)$, $C(t-2)$	$C(t+1)$	3-1-1	6.7	6.4	0.73	0.72
ANN 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$	$C(t+1)$	4-3-1	7.1	7.3	0.68	0.62
ANN 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1)$	4-7-1	4.9	4.8	0.92	0.88
ANN 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-4)$	$C(t+1)$	5-3-1	7.0	8.2	0.67	0.55
ANN 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$	$C(t+1)$	5-1-1	5.4	4.9	0.86	0.91
ANN 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1)$	5-8-1	4.9	4.7	0.92	0.89
ANN 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $T(t)$	$C(t+1)$	6-7-1	5.0	5.2	0.92	0.86
ANN 8	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $RH(t)$	$C(t+1)$	6-8-1	4.3	4.5	0.92	0.86
ANN 9	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$	7-8-1	4.4	4.6	0.92	0.87
ANN 10-1	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{ILZ}}$, $C(t-1)_{\mathrm{ILZ}}$, $C(t-2)_{\mathrm{ILZ}}$	$C(t+1)_{\mathrm{IHZ}}$	6-1-1	6.5	6.2	0.75	0.73
ANN 10-2	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{E}}$, $C(t-1)_{\mathrm{E}}$, $C(t-2)_{\mathrm{E}}$	$C(t+1)_{\mathrm{IHZ}}$	6-2-1	6.6	6.5	0.74	0.71
ANN 10-3	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{M}}$, $C(t-1)_{\mathrm{M}}$, $C(t-2)_{\mathrm{M}}$	$C(t+1)_{\mathrm{IHZ}}$	6-2-1	6.1	6.7	0.78	0.67
ANN 10-4	$C(t)_{\mathrm{IHZ}}$, $C(t-1)_{\mathrm{IHZ}}$, $C(t-2)_{\mathrm{IHZ}}$, $C(t)_{\mathrm{ILZ}}$, $C(t-1)_{\mathrm{ILZ}}$, $C(t-2)_{\mathrm{ILZ}}$, $C(t)_{\mathrm{E}}$, $C(t-1)_{\mathrm{E}}$, $C(t-2)_{\mathrm{E}}$, $C(t)_{\mathrm{M}}$, $C(t-1)_{\mathrm{M}}$, $C(t-2)_{\mathrm{M}}$	$C(t+1)_{\mathrm{IHZ}}$	12-2-1	6.1	6.8	0.78	0.72

Table 8. DAN2 Results

			MAE (L/s)		r	
Model	Input	Output	Train	Test	Train	Test
DAN2 1	$C(t)$, $C(t-1)$, $C(t-2)$	$C(t+1)$	5.5	4.6	0.76	0.85
DAN2 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1)$	4.1	3.5	0.87	0.90
DAN2 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$	$C(t+1)$	4.0	3.5	0.87	0.91
DAN2 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1)$	4.1	3.5	0.87	0.91
DAN2 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $T(t)$	$C(t+1)$	4.0	3.5	0.88	0.91
DAN2 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ and $RH(t)$	$C(t+1)$	4.1	3.5	0.87	0.91
DAN2 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$	4.0	3.5	0.88	0.91

J. Water Resour. Plann. Manage. 2012.138:245-256.

Table 9. ANN-H Results

Model	Input	Output	Architecture	MAE (L/s) Train	Test	r Train	Test
ANN-H 1	$C(t)$, $C(t-1)$, $C(t-2)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$	$C(t+1)$	6-1-1	6.7	6.5	0.75	0.71
ANN-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-168)$	$C(t+1)$	8-6-0	5.3	4.9	0.9	0.9
ANN-H 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-24)$	$C(t+1)$	10-4-1	5.3	6.6	0.75	0.70
ANN-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$	$C(t+1)$	10-4-1	5.3	5.0	0.89	0.87
ANN-H 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$ and $T(t)$	$C(t+1)$	11-8-1	5.4	5.7	0.80	0.78
ANN-H 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$ and $RH(t)$	$C(t+1)$	11-8-1	4.5	4.5	0.89	0.86
ANN-H 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$	12-3-1	5.0	5.3	0.89	0.86

Table 10. DAN2-H Results

Model	Input	Output	MAE (L/s) Train	Test	r Train	Test
DAN2-H 1	$C(t)$, $C(t-1)$, $C(t-2)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$	$C(t+1)$	5.0	3.9	0.80	0.89
DAN2-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-168)$	$C(t+1)$	3.2	2.9	0.92	0.93
DAN2-H 3	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-24)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-24)$	$C(t+1)$	3.3	3.0	0.91	0.93
DAN2-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$	$C(t+1)$	3.3	2.8	0.91	0.93
DAN2-H 5	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$ and $T(t)$	$C(t+1)$	3.2	2.9	0.91	0.93
DAN2-H 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$ and $RH(t)$	$C(t+1)$	3.2	2.9	0.92	0.93
DAN2-H 7	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$, $T(t)$ and $RH(t)$	$C(t+1)$	3.1	3.0	0.91	0.92

Table 11. Best Results of Each Neural Network

Model	Input	Output	MAE (L/s) Train	Test	r Train	Test
ANN 8	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$ *e* $RH(t)$	$C(t+1)$	4.3	4.5	0.92	0.86
DAN2 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1)$	4.1	3.5	0.87	0.90
ANN-H 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-3)$, $\mathrm{Res}(t-168)$ *e* RH(t)	$C(t+1)$	4.5	4.5	0.89	0.86
DAN2-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, $\mathrm{Res}(t)$, $\mathrm{Res}(t-1)$, $\mathrm{Res}(t-2)$, $\mathrm{Res}(t-168)$	$C(t+1)$	3.2	2.9	0.92	0.93

produced the best results according to the evaluation criteria used. It was expected that incorporating both weather factors in modeling consumption would render better results. However, the introduction of temperature did not improve the performance indices as compared with the model with only the consumption variables (ANN 6). Table 6 also presents the results for the spatially multivariate model, ANN 10-1 to ANN 10-4, which had a similar performance to the equivalent univariate model ANN 1.

Table 7 shows the results obtained for DAN2, in which all of its versions have 15 hidden layers. The values in this table suggest that DAN2 2 produced better results than DAN2 1, and almost the same performance as that of the other models presented in the same table.

J. Water Resour. Plann. Manage. 2012.138:245-256.

Table 12. Results of the ANN model with forecast horizon of 24 hours

Model	Input	Output	MAE (L/s) Training	MAE (L/s) Testing	r Training	r Test
ANN 6	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1,\ldots,t+24)$	7.0	6.9	0.79	0.82
DAN2 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$	$C(t+1,\ldots,t+24)$	6.2	6.0	0.74	0.77
DAN2 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$	$C(t+1,\ldots,t+24)$	3.8	3.5	0.87	0.90
ANN-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$	$C(t+1,\ldots,t+24)$	5.5	5.5	0.86	0.88
DAN2-H 2	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-168)$	$C(t+1,\ldots,t+24)$	3.4	3.3	0.89	0.90
DAN2-H 4	$C(t)$, $C(t-1)$, $C(t-2)$, $C(t-3)$, $C(t-168)$, Res(t), Res$(t-1)$, Res$(t-2)$, Res$(t-3)$, Res$(t-168)$	$C(t+1,\ldots,t+24)$	3.1	3.0	0.91	0.92

Therefore, model DAN2 2 was found to be superior considering parsimony criterion (minimization of number of variables, and hence parameters).

Table 8 displays the results for various versions of the first hybrid model ANN-H. Accordingly, ANN-H 6 produced the best results in terms of the evaluation criteria used.

The results for the second hybrid model DAN2-H are presented in Table 9, where DAN2-H 2 can be pointed out as the best model considering parsimony as an additional criterion.

MAE and r values obtained for each neural network model tested are presented in Table 10, where the best performance model is shown to be the neural network On the basis of DAN2, which surpassed the tested ANN MLP-BP models.

The modeling of FS residues did not improve the results for the next hour prediction of the ANN MLP-BP; in fact, some of these models presented a slightly deterioration of performance when compared with the nonhybrid version. It was believed that this was caused by the increase of the network complexity resulting from the introduction of more variables in the input of the ANN MLP-BP, which was not able to improve the network mapping given the tested architectures. Although the best model (DAN2-H 2) did not rely on weather variables, the models on the basis of MLP-BP used relative humidity as input.

Table 11 summarizes the results obtained by the best performance model simulation of each neural network type, which are illustrated in Fig. 5, along with the corresponding observed data. DAN2 produced better results when compared with ANN MLP-BP, especially the correlation between forecast and observed consumption in valleys (lowest values of the curve). It was assumed that one may obtain improvement in the ANN-H model with the introduction of more neurons in the hidden layer, or even more hidden layers.

Table 12 shows the performance results for the 24-h forecasting, which contains the best neural networks obtained for the next hour models on the basis of DAN2, DAN2 2, and DAN2-H 2. However, as the performance of DAN2 2 wasn't satisfactory, DAN2 4 was also evaluated, and DAN2-H 4. In the case of the neural network on the basis of MLP-BP (ANN 8 and ANN-H 6), they were not

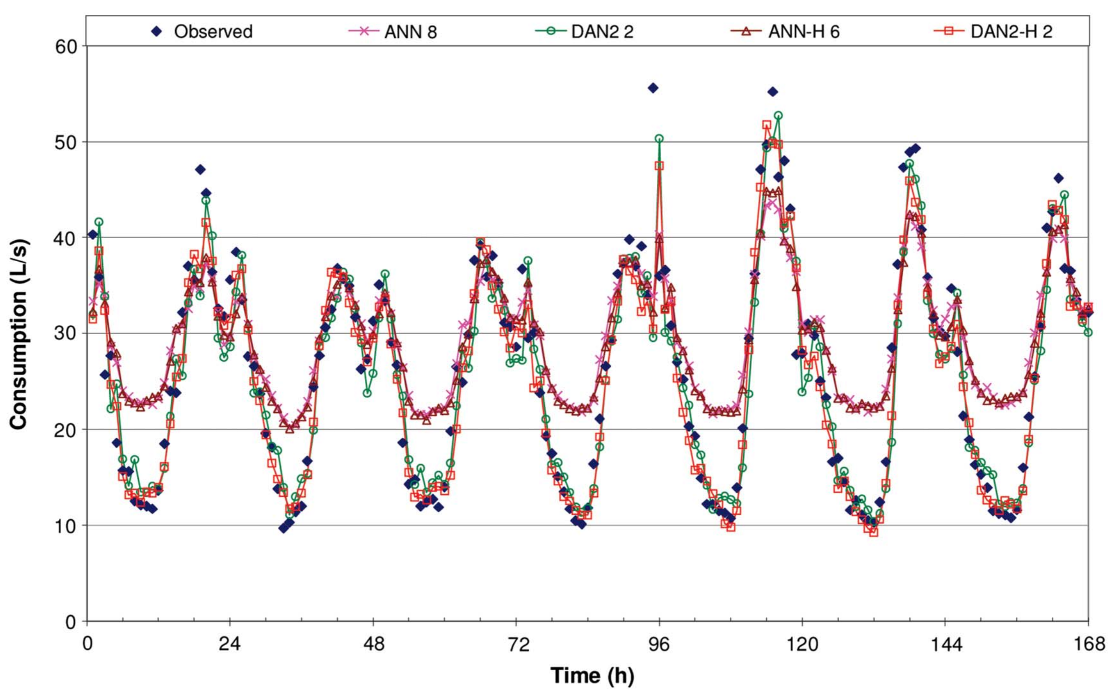

Fig. 5. Best results for each neural network

J. Water Resour. Plann. Manage. 2012.138:245-256.

presented in Table 11, as it would require weather forecasting, which was not straightforward and needed further investigation.

Regarding the prediction for the next hour, model DAN2-H 2 surpassed the other models evaluated so far. The best forecasting model for the next 24 h (DAN2-H 4) showed satisfactory results when compared with the forecast for the next hour, producing MAE (testing set) of 3.0 L/s, equivalent to 12% of mean consumption.

Conclusion

In this work, several ANN models were compared to evaluate their potential to match or even surpass other forecasting models applied to the water consumption data. The ANN models tested were based on MLP-BP and DAN2, and the proposed hybrid neural networks which consisted of modeling the difference between the Fourier series forecasts and the observed consumption data together with weather variables. All data were statistically treated for the occurrence of missing and faulty data before use. The inputs to the neural networks were selected on the basis of the correlation analysis. The tested forecasting models were evaluated using the Mean Absolute Error (MAE) and Pearson Correlation Coefficient (r) criteria.

Through analysis and comparison, it was found that the best forecasting model—DAN2-H 2, for the first hour prediction and DAN2-H 4 for the 24 h prediction—did not require the use of weather variables, resulting in simpler and faster model to train, as it was self-constructive, and didn't need the use of trial and error procedures to determine the number of neurons in the hidden layer. The modeling of residuals of FS improved the performance of the dynamic neural network, but not for the ANN MLP-BP, for predicting the next hour. For 24-h forecasting, the FS residues improved the performance of the tested models. It is believed that the dynamic aspect of the forecast, (i.e., the hourly update of the FS coefficients on a weekly basis), was responsible for the improvement, because it may eliminate the seasonal and weekday effect.

The error criteria showed that the dynamic neural network was a good alternative to the neural network MLP-BP, because it produced a better fit for the observed data and showed a simpler training phase because of the self-constructive aspect of that neural network.

The 24-hour forecast models on the basis of DAN2 performed slightly better than the forecast of the first hour. Conversely, the models on the basis of MLP-BP network had the expected results, (i.e., the forecast of the first hour had better results than the 24-h forecast). Finally, there were enough elements to conclude that ANNs, developed and applied to several areas of knowledge, were also promising tools for the water-demand forecast purposes. In this work, the DAN2 stood out when compared with the tested ANN MLP-BP.

Acknowledgments

The authors acknowledge the support from the Brazilian Scientific and Technological Development Council (CNPq) and from the Research Support Foundation of São Paulo (FAPESP) for providing the scholarship and grant to the authors, and also from the Araraquara's Autonomous Department of Water and Sewage (DAAE-Araraquara, SP, Brazil) for providing the time-series data and assistance. The authors would also like to thank the anonymous reviewers for their helpful comments that permitted improvements to the manuscript.

References

Adamowski, J. F. (2008). "Peak daily water demand forecast modeling using artificial neural networks." *J. Water Resour. Plann. Manage.*, 134(2), 119–128.

Athanasiadis, I., Mentes, A., Mitkas, P., and Mylopoulos, Y. (2005). "A hybrid agent-based model for estimating residential water demand." *Simulation*, 81(3), 175–187.

Bougadis, J., Adamowski, K., and Diduch, R. (2005). *Short-term municipal water demand forecasting—Intersciences*, Wiley.

Crommelynck, V., Duquesne, C., Mercier, M., and Miniussi, C. (1992). "Daily and hourly water consumption forecasting tools using neural networks." *Proc. of the AWWA's Annual Computer Specialty Conference*, Nashville, TN, 665–676.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function." *Math. Control Signals Syst.*, 2(4), 303–314.

Funahashi, K. I. (1989). "On the approximate realization of continuous mappings by neural networks." *Neural Netw.*, 2(3), 183–192.

Ghiassi, M., and Saidane, H. (2005). "A dynamic architecture for artificial neural networks." *Neurocomputing*, 63, 397–413.

Ghiassi, M., Zimbra, D. K., and Saidane, H. (2008). "Urban water demand forecasting with a dynamic artificial neural network model." *J. Water Resour. Plann. Manage.*, 134(2), 138–146.

Haykin, S. (1999). *Neural network—A comprehensive foundation*, 2nd Ed., Prentice Hall, NY.

Hristev, R. H. (2000). "Matrix techniques in artificial neural networks." Master Science thesis. Univ. of Cantebury, UK.

Humes, A. F. P. C, Melo, I. S. H., Yoshida, L. K., and Martins, W. T. (1984). *Noções de cálculo numérico São Paulo*, McGraw-Hill.

Jain, A., and Kumar, A. M. (2007). "Hybrid neural network models for hydrologic time series forecasting." *Appl. Soft Comput.*, 7(2), 585–592.

Jain, A., Ormsbee, L. E. (2002). "Short-term water demand forecast modeling techniques: Conventional methods versus AI." *J. AWWA*, 94(7), 64–72.

Jain, A., Varshney, A. K., and Joshi, U. C. (2001). "Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks." *Water Resour. Manage.*, 15(5), 299–321.

JBPCAfill. (2004). "Missing value estimator for Java. V.2." ⟨http://hawaii.aist-nara.ac.jp/~shige-o/tools/BPCAFill.html⟩ (Nov. 30, 2009).

Jowitt, P. W., and Xu, C. (1992). "Demand forecasting for water distribution systems." *Civ. Eng. Syst.*, 9(2), 105–121.

Lim, E. A., and Zainuddin, Z. (2008). "A comparative study of missing value estimation methods: Which method performs better?" *International Conference on Electronic Design 2008*, IEEE, New York, 1–5.

Maidment, D. R., Miaou, S. P., and Crawford, M. M. (1985). "Transfer function models of daily urban water use." *Water Resour. Res.*, 21(4), 425–432.

Miaou, S. P. (1990). "A class of time series Urban water demand models with non-linear climatic effects." *Water Resour. Res.*, 26(2), 169–178.

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., and Ishii, S. A. (2003). "Bayesian missing value estimation method." *Bioinformatics*, 19(16), 2088–2096.

Odan, F. K., Ferrero, C. A., Reis, L. F. R., Monard, M. C. (2009). "Análise comparativa dos modelos kNN-TSP e Série de Fourier para previsão de demanda horária para abastecimento de água." *XVIII Simpósio Brasileiro de Recursos Hídricos—ABRH*.

Perry, P. F. (1981). "Demand forecasting in water supply networks." *J. Hydraul. Div.*, 107(HY9), 1077–1087.

Pulido-Calvo, I., Gutierrez-Estrada, J. C. (2009). "Improved irrigation water demand forecasting using a soft-computing hybrid model." *Biosystems Eng.*, 102(2), 202–218.

Shvartser, L., Shamir, U., and Feldeman, M. (1993). "Forecasting hourly water demand by pattern recognition approach." *J. Water Resour. Plann. Manage.*, 119(6), 611–627.

Smith, J. A. (1988). "A model of daily municipal water use for short-term forecasting." *Water Resour. Res.*, 24(2), 201–206.

Tang, Z., and Fishwick, P. A. (1993). "Feedforward neural nets as models for time series forecasting." *ORSA J. Comput.*, 5(4), 374–385.

J. Water Resour. Plann. Manage. 2012.138:245-256.

Valdes, J. B., and Sastri, T. (1989). "Rainfall intervention analysis for on-line applications." *J. Water Resour. Plann. Manage.*, 115(4), 397–415.
Zahed, K. F. (1990). "Previsão de demanda de consumo em tempo real no desenvolvimento operacional de sistemas de distribuição de água." *Tese de doutorado*, Escola Politécnica da Universidade de São Paulo, São Paulo.
Zhang, G. (2001). "An investigation of neural networks for linear time-series forecasting." *Comput. Oper. Res.*, 28(12), 1183–1202.
Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing; Variable Star Bull.*, 50, 159–175.
Zhang, J., Song, R., Bhaskar, N. R., and French, M. N. (2006). "Short-term water demand forecasting: A case study." *Proc., 8th Annual Water Distribution Systems Analysis Symp.*, ASCE, Reston, VA, 1–14.

J. Water Resour. Plann. Manage. 2012.138:245-256.