ANN versus SARIMA models in forecasting residential water consumption in Tunisia

Maamar Sebri

ABSTRACT

Water scarcity and increasing water demand, especially for residential end-use, are major challenges facing Tunisia. The need to accurately forecast water consumption is useful for the planning and management of this natural resource. In the current study, quarterly time series of household water consumption in Tunisia was forecast using a comparative analysis between the traditional Box–Jenkins method and an artificial neural networks approach. In particular, an attempt was made to test the effectiveness of data preprocessing, such as detrending and deseasonalization, on the accuracy of neural networks forecasting. Results indicate that the traditional Box–Jenkins method outperforms neural networks estimated on raw, detrended, or deseasonalized data in terms of forecasting accuracy. However, forecasts provided by the neural network model estimated on combined detrended and deseasonalized data are significantly more accurate and much closer to the actual data. This model is therefore selected to forecast future household water consumption in Tunisia. Projection results suggest that by 2025, water demand for residential end-use will represent around 18% of the total water demand of the country.

Key words | artificial neural networks, residential water demand, SARIMA, time-series forecasting, Tunisia

Maamar Sebri
Faculty of Economic Sciences and Management of Sousse,
University of Sousse,
Tunisia
E-mail: *maamar.sebri@gmail.com*

INTRODUCTION

Tunisia is the northernmost country in Africa. It is bordered by Algeria to the west, Libya to the southeast and the Mediterranean Sea to the north and east. This geographic location means that it shares many features with most countries of the Maghreb region, such as climatic conditions and limited water resources. The climate varies from Mediterranean to semi-arid and arid; it is characterized by hot, dry summers and mild winters, receiving the major part of the annual precipitation. Water resources constitute the most precious environmental good. According to the Tunisian Ministry of Agriculture and Hydraulic Resources (MAHR 2010), the potential water resources in Tunisia for the year 2009 were estimated as 4,655 million m^3, of which 2,500 million m^3 come from surface water and 2,155 million m^3 are groundwater. The quantity of water available per capita was 460 m^3 in 2000 and it will decrease to only 345 m^3 by 2025 (Hamdane 2007). Such statistics place the country in an absolute water scarcity situation. Hydrologists typically assess scarcity by looking at the population-water equation. An area is experiencing water stress when annual water supplies drop below 1,700 m^3 per person. When annual water supplies drop below 1,000 m^3 per person, the population faces water scarcity and below 500 m^3 'absolute scarcity' (http://www.un.org/waterforlife-decade/scarcity.html).

Obviously, the adverse effect of climate change on water demand and supply is nowadays a common issue. Global warming is expected to cause, among other effects, a significant rise in mean temperature, a decline in rainfall and an increase in the frequency of droughts. In Tunisia, the sensitivity of water resources to climate changes concerns particularly: (i) the variability of surface water resources due to a decrease in precipitation levels and an increase in evaporation rate amplified by the rise in temperature and

doi: 10.2166/washdev.2013.031

(ii) the deterioration in groundwater quality particularly the shallow aquifers of the coastal region. These latter are affected by seawater intrusion due to the growing rise in sea level of the Mediterranean.

Despite the limited resources, water demand for the domestic, agricultural, industrial and tourist sectors has increased, and water scarcity problems are expected to intensify. Keeping in mind the shortage and the acute pressure on the available water resources, a reliable water consumption forecast (for both the long and short term) is mandatory for the implementation of water management schemes for different sectors, especially the residential, as it is straightforwardly related to human life. While short-term forecasting is useful in operation and management, long-term forecasting is required mainly for planning and design (Herrera *et al.* 2010).

Water consumption forecasting has generated a substantial body of literature in which various techniques and statistical methods have been proposed, leading to various results and conclusions. Early on, regression models and traditional time-series techniques dominated the water demand forecasting literature. In recent years, artificial neural networks (ANNs) have appeared as an efficient tool for forecasting. This technique has also been employed in a wide range of applications including financial, climatic and energy disciplines, and it mostly proves its superiority when compared with traditional techniques.

The current study focuses on the forecasting of residential water consumption in Tunisia using a comparative analysis between seasonal auto-regressive integrated moving average (SARIMA) and ANN models. The empirical analysis uses quarterly data of a residential water consumption time series and in addition to the SARIMA model, four different ANN models are built and compared based on data preprocessing type.

LITERATURE OVERVIEW

The ANN has been applied in a wide range of hydrology and water resource engineering areas. These include water quality (Zhang *et al.* 2002; Najah *et al.* 2009), groundwater level (Pallavi *et al.* 2009; Sreekanth *et al.* 2009), rainfall-runoff process (Jeong & Kim 2005; Aytek *et al.* 2008), sediment concentration estimation (Nagy *et al.* 2002; Williamson & Crawford 2011), streamflows (Birikundavyi *et al.* 2002; Ahmed & Sarma 2007) and waste water treatment (Chen *et al.* 2003a; Gamal El-Din *et al.* 2004).

Recently, water consumption especially for residential use modeling and forecasting with ANN has also attracted many researchers who intended to test the success of this relatively new technique and compare its prediction accuracy with traditional techniques. Šterba & Hiľovská (2010) proposed a hybrid neural network model consisting of a combination of the auto-regressive integrated moving average (ARIMA) and ANN to predict the aggregate water consumption in Spain ranging from January 1984 to June 2007. The authors found that the proposed hybrid model outperforms ARIMA and ANN models estimated separately.

Bougadis *et al.* (2005) investigated the short-term weekly water demand forecasting in the Ottawa region during the period 1993–2002. The authors used only the data corresponding to months of summer season since they analyzed the peak demand which generally occurs in summer, and they also incorporated temperature and rainfall as explanatory variables. The authors employed various specifications of ARIMA, linear regression and ANN as forecasting methods, but their results indicate that ANN technique substantially outperforms time-series and linear regression methods in terms of forecasting accuracy.

The modeling of monthly municipal water consumption in the metropolitan area of Izmir (Turkey) was examined by Firat *et al.* (2009). Based on several socioeconomic and climatic variables, the authors used three different techniques of ANN such as generalized regression neural networks (GRNN), feed forward neural networks (FFNN) and radial basis neural networks (RBNN) in addition to the multiple linear regression (MLR). The more satisfactory results are given by the GRNN when judged according to the three performance criteria, namely normalized root mean square error (NRMSE), efficiency (*E*) and correlation coefficient (CORR).

In the same fashion of comparison of neural networks models, Firat *et al.* (2010) used the data from Firat *et al.* (2009) to compare the accuracy of GRNN, cascade correlation neural networks (CCNN) and FFNNs. They used different specifications based on the historic lags of water demand as inputs. The results show that CCNN with five antecedent values of water consumption data records is better than all the other investigated models.

Considering a large set of machine learning methods, Herrera *et al.* (2010) addressed the task of predicting hourly water demand in an urban area of a city in southeastern Spain. The authors used as predictive models: ANN, projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), support vector regression (SVR), random forests and a proposed heuristic model they called weighted pattern-based model for water demand forecasting. The experimental results indicate that the SVR is the most accurate model.

A more recent study of Azadeh *et al.* (2012) focused on forecasting the short-term water consumption in Tehran from April 2004 to March 2009. A hybrid approach consisting of ANNs, fuzzy linear regression (FLR) and analysis of variance (ANOVA) was used. The main results indicate that forecasts obtained using the hybrid approach are more accurate than those using ANN or FLR separately. The authors argued that this is an expected result since potential nonlinearity and uncertainty in the water consumption function are usually observed in a large metropolitan city such as Tehran.

According to the above literature review, one may suggest that the water consumption forecasting becomes a more and more attractive field of research. In most cases, various statistical and mathematical techniques are simultaneously used in order to select the best model. However, the ANN technique is often found to be superior to the traditional approaches in terms of prediction accuracy. In particular, for time-series analysis, the ANN technique appears as a new contender to the widely used Box–Jenkins method in forecasting sophisticated trend and seasonal data. In this context lies the current study by using the ANN and SARIMA models to forecast the quarterly time series of household water consumption in Tunisia.

SARIMA AND ANN MODELS FOR TIME-SERIES FORECASTING

In this section, the basic concepts and modeling approaches of the SARIMA and ANN models for time-series forecasting are briefly reviewed.

The SARIMA model

SARIMA is the version of the traditional ARIMA model when the series exhibit seasonal patterns. ARIMA, as introduced by Box & Jenkins (1976), includes autoregressive (p) and moving average (q) parameters as well as differencing in the formulation of the model (d). Thus the Box–Jenkins model is summarized as ARIMA (p,d,q). In addition to the non-seasonal parameters defined above, SARIMA model includes seasonal autoregressive (P), seasonal moving average (Q) and seasonal differencing passes (D), and it is referred to as SARIMA$(p,d,q)(P,D,Q)_s$, where s indicates the seasonal period length. According to Zhang & Qi (2005), the general form of the SARIMA$(p,d,q)(P,D,Q)_s$ model describing the current value y_t of a time series by its own past is:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{1}$$

where:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$
$$\Phi_P(B) = 1 - \Phi_s B - \Phi_{2s} B^{2s} - \cdots - \Phi_{Ps} B^{Ps},$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,$$
$$\Theta_Q(B) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \cdots - \Theta_{Qs} B^{Qs}$$

where B is the back shift operator defined by $B^k y_t = y_{t-k}$ and ε_t is white noise with zero mean and constant variance. $(1-B)^d$ and $(1-B^s)^D$ are the non-seasonal and seasonal differencing operators, respectively.

To build a SARIMA model, Box–Jenkins methodology requires three iterative steps: identification, estimation and diagnostic checking (Zhang & Qi 2005). The identification step consists of determining the number of seasonal and non-seasonal parameters. The major tools used in the identification phase are plots of the series, correlograms of autocorrelation function (ACF) and partial autocorrelation function (PACF). However, SARIMA models need the series to be stationary. Thus one should examine the stationarity property of the series before determining the parameters p, P, q and Q. Therefore, if necessary, the series needs to be non-seasonally and/or seasonally differenced to achieve stationarity, i.e. with a constant mean, variance and autocorrelation through time. The number of times of non-seasonal differencing reflects in the *d* parameter while the *D* parameter corresponds by the number of seasonal differencing. Once the series is stationary, the new plots of ACF and PACF are again used to decide on the number of *p*,*P*,*q* and *Q*. The decision is not straightforward and in less typical cases requires not only experience but also a good deal of experimentation with alternative models. Advice and some practical recommendations are given in many books including Pankratz (1983), Hoff (1983) and McCleary & Hay (1980). Alternatively to ACF and PACF plots, the identification of the model could be based on the minimization of some information criteria such as the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC).

Once the different SARIMA parameters are chosen, they are straightforwardly estimated making the overall measures of errors as small as possible. In general, nonlinear estimation is used to estimate the identified parameters by maximizing the likelihood (probability) of the observed series given the parameter values.

The last step of SARIMA building is the diagnostic checking of the adequacy of the estimated model. A first criterion is the dropping of the insignificant parameters based on the *t*-value statistics. A second common feature of the reliability of the model is the accuracy of the forecasts it generates based on partial data, so that the forecasts can be compared with the original observations. However, a good model should not only provide sufficiently accurate forecasts, it should also be parsimonious and produce statistically independent residuals that contain only noise. Therefore the third test of the model adequacy is the analysis of residuals. This is an important test of the model since the estimation procedure assumes that the residuals are not (auto-)correlated and that they are normally distributed. This test consists in plotting the residuals and inspecting them for any systematic trends, examining the autocorrelogram of residuals (there should be no serial dependency between them) and checking the normality assumption by examining the normal probability plots of residuals.

Once a set of models has been identified and estimated, it is possible that more than one of them is not rejected in the diagnostic checking step. The selection of one between these competing models is generally based on the minimization of some information criteria such as the AIC and the SIC.

The ANN model

The ANN model, inspired by studies of biological neural systems, is composed of processing elements called neurons or nodes (Firat *et al.* 2010). Neural networks are very sophisticated modeling and prediction techniques capable of modeling and approximating extremely complex functions and data relationships (Chen *et al.* 2003b; Zhang & Qi 2005). Thus, there has been an explosion of interest in ANNs over the past two decades. ANN models are comprised of user-defined inputs and desired output(s) that are connected by a set of highly interconnected nodes arranged in a series of layers (Bougadis *et al.* 2005).

There are many different ways of connecting artificial neurons together to create a neural network but two of the most popular neural network architectures, multilayer perceptrons (MLP) and radial basis function (RBF) are used in the current study. Both MLP and RBF are feed-forward networks. A feed-forward network has a layered structure. The units of each layer receive their input from the units of the previous layer and send their output to the units of the next layer using a mathematical function, for example, logistic, exponential, hyperbolic tangent, identity and Gaussian. Moreover, there are no connections between the units within a given layer. The feed-forward network uses the back-propagation algorithm for training. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) and scaled conjugate gradient are the most recommended back-propagation algorithms used for optimization of neural network architecture (Bishop 1995).

An MLP network is formed by one input layer of source nodes and one output layer of nodes. These two layers are connected via one or more layers of hidden neurons, which are so called because they are not directly accessible (Haykin 1999). A typical MLP with one hidden layer is depicted in Figure 1.

The RBF network, shown in Figure 2, is configured with a single hidden layer of neurons whose activation function (generally Gaussian) is selected from a class of functions called basis functions. The activation function of the hidden layer in an RBF network computes the Euclidean

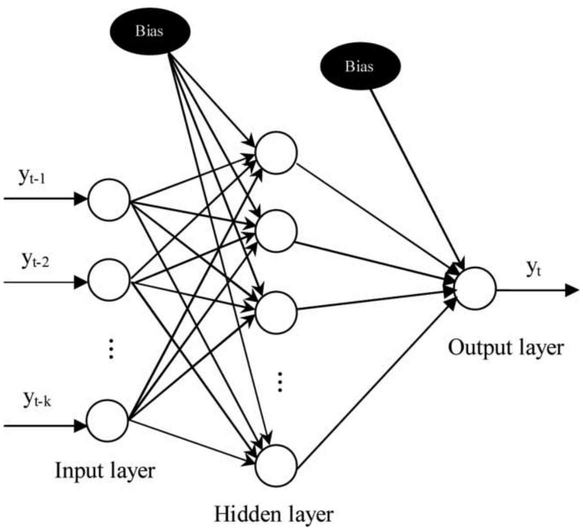

Figure 1 | Structure of MLP network for time series.

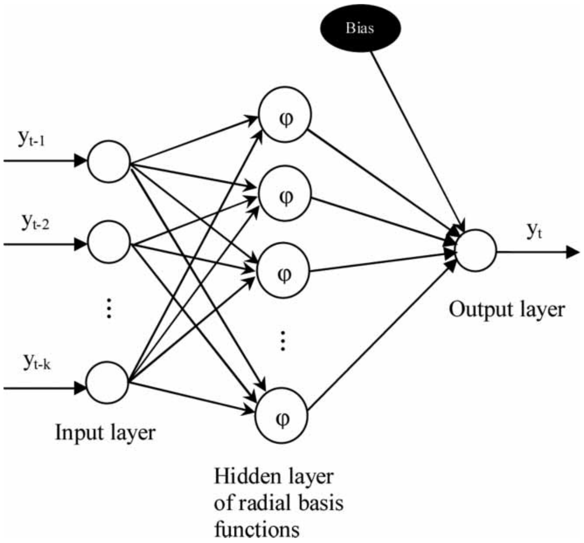

Figure 2 | Structure of RBF network for time series.

distance between the input signal vector and parameter vector of the network (Haykin 1999).

For time-series forecasting with ANNs, the task of identifying the correct number of units (lagged observations) to use in the input layer is a matter of great importance. By including irrelevant inputs, for example, one may inadvertently damage the performance of the neural network. In contrast, a data set with an insufficient number of inputs may never be accurately modeled. Different statistical tests and criteria are proposed in empirical studies to select the correct number of input units in neural networks modeling but most of them are model-dependent and none is privileged to others in all situations. While the simple selection method is still based on trial and error, in some studies, authors defined the number of input units according to the natural cycle of the time series. For instance, Sharda & Patil (1992), Tang *et al.* (1991) and Shabri (2001) used 12 inputs for monthly data and 4 inputs for quarterly data.

METHODS

Tunisia is characterized by low levels of water resources but the demand for this resource, especially for domestic purposes, has shown a remarkable increase through years. Figure 3 illustrates this feature well indicating that the water consumption time series presents a linear trend as well as a seasonal pattern. The time series consists of a total of 112 data records ranging from the first quarter of 1983 to the fourth quarter of 2010.

Despite the capability of neural networks to predict any type of relationship in the data with high accuracy, some empirical studies have highlighted the role of data preprocessing before using them as inputs to neural networks (Nelson *et al.* 1999; Zhang & Qi 2005). To investigate the effectiveness of modeling the water consumption time series with both seasonal and trend patterns using ANN, three type of data preprocessing were performed in the current study before training the data, namely: (i) detrending, (ii) deseasonalizing and (iii) both detrending and deseasonalizing. According to Figure 3, the time series of Tunisian

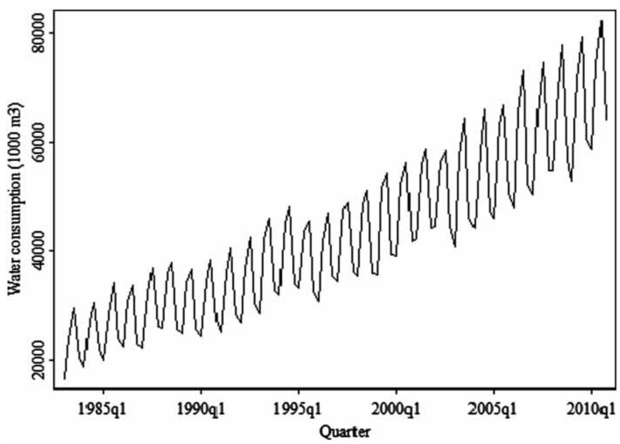

Figure 3 | Time-series plot of residential water consumption in Tunisia.

water consumption exhibits an upward trend and a seasonal pattern. Moreover the size of seasonal fluctuations increases with time, depending on the overall level of the series. Therefore, the simulated series could be generated according to the following simple multiplicative model:

$$y_t = T_t SI_t + \varepsilon_t \quad (2)$$

where T_t is the trend, SI_t is the seasonal component and the ε_t is the error term.

First, detrending of the data was performed by fitting a linear regression on t ($t = 1, \ldots, 112$) of the original data to obtain the estimated trend T_t defined by Equation (3), then dividing the original time series by the estimated trend to obtain detrended data.

$$T_t = 19805.97 + 411.937 * t \quad (3)$$

Second, deseasonalization was done by using the seasonal index based on the ratios-to-moving averages classical seasonal decomposition (Census Method I) following the multiplicative decomposition. The obtained seasonal index SI is given in Table 1. The deseasonalized series is obtained by dividing the original data by SI.

Finally, for detrended and deseasonalized data, the original time series was divided by both the trend and the seasonal index. This data preprocessing procedure was reserved to be used only for the ANN approach while for SARIMA model building, data preprocessing is a preliminary condition since these latter require the data to be stationary.

In the current study, Statistica version 8 was deployed to analyze the data and automatically select the best ANN topology while the statistical package StatGraphics Centurion version 16 was used to conduct an automatic SARIMA model building and evaluation.

After neural network modeling, the data were rescaled back following the reverse of the data transformation and, therefore, all the performance measurement criteria were calculated based on the original scale of the data.

Table 1 | Seasonal index (SI) for water consumption time series

Quarter	q1	q2	q3	q4
SI	0.819	1.108	1.209	0.862

The data set was divided into two sub-samples, the training sample which makes up the first 80 data records (from the first quarter of 1983 to the last quarter of 2002) and the testing sample which uses the remaining 32 data records (from the first quarter of 2003 to the last quarter of 2010) for evaluating the performance of the developed SARIMA and ANN models. Following the approach (described above) adopted by Tang *et al.* (1991), Sharda & Patil (1992) and Shabri (2001), the number of time-series steps considered as inputs was set to four given that the natural cycle period of the water consumption time series is quarterly. It should be noted that other input lags ranging from 1 to 8 were also tried for the different neural network models but results provided at 4 input lags are still by far the best.

Performance measurement criteria

Three metrics were used in this analysis to compare the performance of the models: mean absolute errors (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). These criteria were defined respectively by the following expressions:

$$\text{MAE} = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n}; \quad \text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}};$$

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n} \times 100$$

where y_t and $\hat{y}_t$ indicate the observed and predicted values of water consumption respectively and n shows the sample size. These are the most frequently used criteria to compare models performance in terms of forecasting accuracy (Zhang *et al.* 1998). The model with the lowest success criteria has the better forecasting performance.

RESULTS AND DISCUSSION

SARIMA results

The first step in building SARIMA models is to check the stationarity of the series. Therefore, an ACF plot of the water consumption time series was made (Figure 4).

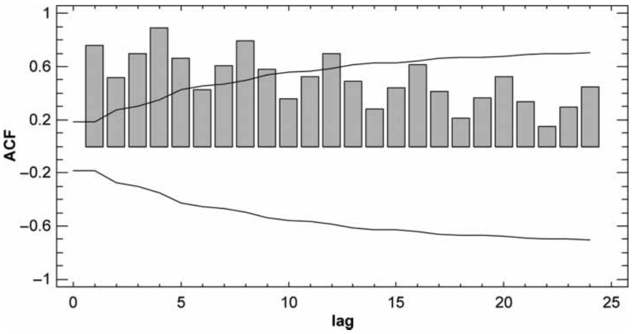

Figure 4 | Autocorrelation function (ACF) for original time series.

The plot indicates strong serial dependencies for lags of 1–12. In order to remove the serial dependency, Box–Jenkins methodology consists of differencing (non-seasonally and/or seasonally) the series until it becomes stationary. Based on the AIC criterion, the SARIMA$(1,0,0)(0,2,2)_4$ model was selected by the package to adjust the water consumption time series.

After deriving the model, the autocorrelation and PACFs of the residuals plotted in Figures 5(a) and 5(b) were used to check whether the errors were correlated or not.

According to Figures 5(a) and 5(b), it is clear that there was no serial dependency for the residual errors which indicates that the selected model could be used to adjust the data.

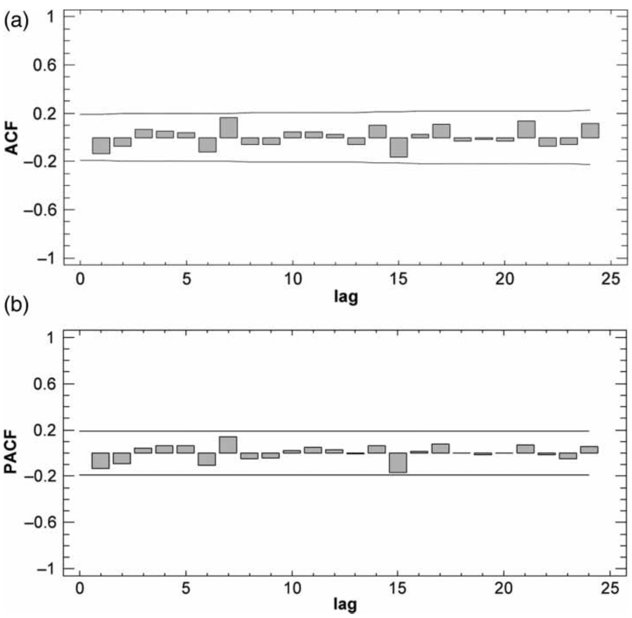

Figure 5 | (a) Autocorrelation function (ACF) for residual errors resulted from SARIMA (1,0,0)(0,2,2) model. (b) Partial autocorrelation function (PACF) for residual errors resulted from SARIMA(1,0,0)(0,2,2) model.

ANN results

The detailed results of ANN models according to the data preprocessing type are summarized in Table 2.

The best fit of original (ORG), detrended (DET) and deseasonalized (DES) data was obtained by MLP networks with a BFGS training algorithm. The number of hidden nodes was 22, 23 and 14, respectively. However, the RBF network (trained with radial basis function training algorithm, RBFT) produced the best fit when both detrending and deseasonalizing the data (DETDES) with 19 nodes in the hidden layer. As explained previously, for all the neural network models, the number of nodes in the input layer was set to four, while various activation functions were used in the hidden and output layers.

Comparative analysis

The out-of-sample forecasting comparison between ANN and SARIMA models is presented in Table 3. This comparison is established in two dimensions: (i) raw versus preprocessed data for ANN models and (ii) ANN versus SARIMA models.

Obviously, it can be observed that the forecasting accuracy of the ANN models is input data-structure dependent, i.e. when the data is preprocessed, the ANN performs well. According to the three measurement criteria, the ANN model with DET and DES data gave better results

Table 2 | ANN results

Input data[a]	ANN topology	Training algorithm	Hidden activation function	Output activation function
ORG	MLP(4-22-1)	BFGS	Logistic	Identity
DET	MLP(4-23-1)	BFGS	Logistic	Hyperbolic tangent
DES	MLP(4-14-1)	BFGS	Exponential	Identity
DETDES	RBF(4-19-1)	RBFT	Gaussian	Identity

[a]ORG: original; DET: detrended; DES: deseasonalized; DETDES: detrended and deseasonalized; MLP: Multilayer Perceptrons; BFGS: Broyden–Fletcher–Goldfarb–Shanno; RBFT: radial basis function training.

Table 3 | The out-of-sample comparison between ANN and SARIMA models

Model	Input data	MAE	MAPE	RMSE
ANN	ORG	2,017.69	3.42	2,205.98
	DET	1,658.71	2.73	1,951.69
	DES	1,729.27	2.84	2,071.27
	DETDES	1,467.68	2.45	1,741.12
SARIMA	ORG	1,607.45	2.62	1,954.80

than the ANN with original data as they present lower values of different criteria. However, the ANN model estimated on DETDES data generated more accurate forecasts than all alternatives since it shows the smallest values of MAE, MAPE and RMSE criteria. These results lead to the conclusion that ANNs are not able to accurately forecast time series that exhibit trend and/or seasonality.

Moving on to the comparison between ANN and the benchmark SARIMA models, it is clear that SARIMA model shows good forecasting capability. Based on the three error functions, the SARIMA model outperformed the neural networks with ORG, DET and DES input data with the exception of the RMSE criterion, which gave a lower value for ANN with detrended data. However, the SARIMA model failed to give more accurate forecasts than the ANN model with DETDES data. These results suggest that with appropriate data preprocessing ANN significantly outperforms the traditional Box–Jenkins time-series technique.

Finally, one may argue that the neural network model with simultaneously detrended and deseasonalized data outperforms all the competing models in terms of forecasting accuracy. Accordingly, it has been selected to generate future forecasts of residential water consumption in Tunisia.

Forecasting future water consumption

Before simulating future water consumption forecasts, another diagnostic check was performed to confirm the robustness of the selected ANN model. In a study carried out in 2004, the Tunisian Water Company (SONEDE 2004) presented its quarterly forecasts of residential water consumption for the period 2003–2010. This period corresponds to the testing period in the current study that was not used in the model construction. Thus a comparison could be done between the two simulated data sets in order to compare the performance accuracy of the model selected in the present study (ANN with DETDES data) and the technique used in the SONEDE study, which is simply a linear trend-based forecasting model. The results in terms of observed and forecasted water use are depicted in graphical form in Figure 6.

It is clear from Figure 6 that the forecasts generated by the ANN model were dramatically better than those estimated by SONEDE especially for the last years of the period under consideration. The forecasts from ANN model were much closer to the actual data while the forecasts of SONEDE became unreliable for more distant predictions. In particular, SONEDE seemed to underestimate the water quantities demanded. Accordingly, the ANN model developed here proves again, its superiority in forecasting residential water consumption.

Using the best fit validated ANN model, which was determined above, the residential water consumption was forecasted in quarterly terms for the period 2011–2025. Results are presented in Table 4.

According to the forecast values, household water use will continue to rise with the same seasonal pattern. Particularly, by the end of 2025, the total water consumption will reach 393,326,330 m^3 which is very close to 0.4 km^3. According to the Blue Plan estimates (www.planbleu.org), (The Blue Plan is a mechanism developed by the 21 countries bordering on the Mediterranean and the European Union within a context of growing international action for the

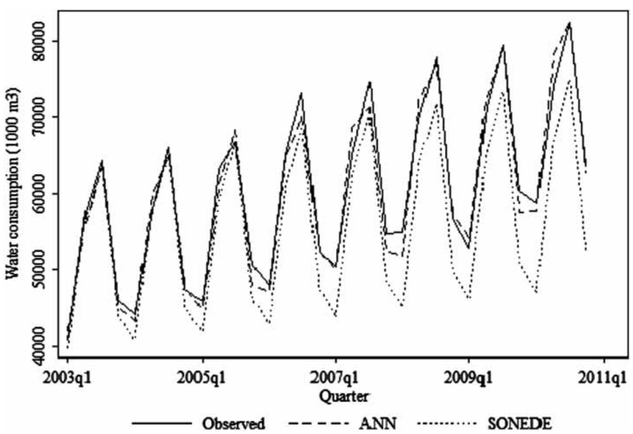

Figure 6 | Plots of observed water consumption, ANN forecasts and SONEDE forecasts for the testing period (2003q1–2010q4).

Table 4 | Forecast residential water consumption in Tunisia from 2011 to 2025 (1,000 m^3)

Year	Quarter	Water	Year	Quarter	Water	Year	Quarter	Water
2011	q1	59,419.878	2016	q1	67,018.719	2021	q1	73,432.586
	q2	79,438.992		q2	91,667.012		q2	100,634.078
	q3	88,906.573		q3	98,837.577		q3	111,617.627
	q4	64,852.626		q4	70,970.165		q4	78,996.392
2012	q1	60,679.128	2017	q1	68,944.005	2022	q1	74,940.256
	q2	81,589.232		q2	93,006.980		q2	103,355.506
	q3	91,419.276		q3	100,649.156		q3	113,404.156
	q4	65,617.934		q4	73,095.243		q4	80,039.627
2013	q1	61,622.385	2018	q1	70,406.893	2023	q1	76,854.802
	q2	84,683.643		q2	94,244.751		q2	105,791.349
	q3	94,391.206		q3	103,216.546		q3	114,813.292
	q4	66,618.349		q4	75,164.050		q4	81,501.189
2014	q1	62,988.914	2019	q1	71,459.323	2024	q1	78,876.778
	q2	87,296.854		q2	95,777.988		q2	107,566.711
	q3	96,089.088		q3	106,164.574		q3	116,208.876
	q4	67,635.334		q4	76,832.991		q4	83,444.065
2015	q1	64,995.563	2020	q1	72,378.442	2025	q1	80,760.276
	q2	89,862.190		q2	97,977.302		q2	108,902.945
	q3	97,468.905		q3	109,107.424		q3	118,090.191
	q4	69,022.053		q4	78,015.347		q4	85,572.921

environment. It is designed to provide information about the environmental risks and sustainable development issues in the Mediterranean and to shape future scenarios to guide decision-taking processes (www.planbleu.org)), the total water demand in Tunisia for different sectors will reach 2.2 km^3 in 2025. Consequently, residential water demand will represent around 18% of the total demand in 2025, while this share was only about 8.5% in 2000. It should be noted, in order to avoid possible confusion with other statistics, that the data used in the current study refer to residential water consumption for only the customers connected to the SONEDE network. Therefore, the significant increase in water consumption for residential end-use compared to other sectors, such as irrigation or industry, is apparent. The share of water consumption for irrigation was nearly 81% in 2000, and according to the Blue Plan projections, it will decrease to 72% by the end of 2025, while the share of water consumption for the industrial sector was 2.6% in 2000 and it will reach around 7.7% by 2025.

CONCLUSIONS

Tunisia, like most Middle East and North Africa countries, is characterized by very limited water resources. Thus, forecasts are quite important for implementation of effective water saving policies. Accurate forecasts of water consumption are vital when demand grows faster and the resources are vulnerable as is the case for Tunisia.

Throughout this study, an attempt has been made to forecast the residential water consumption time series of Tunisia using two techniques, SARIMA and ANN, then to use the model that generates more accurate forecasts to simulate the future quantities of residential water consumption. Four different ANN models are considered here based on the data preprocessing type. A comparative analysis between the competing models highlights that, as traditional statistical models, neural networks are not able to simultaneously handle many different components of the data well and hence data preprocessing is beneficial

(Zhang & Qi 2005). The empirical results suggest that ANN estimated on simultaneously detrended and deseasonalized data outperforms (in terms of forecasting accuracy) the traditional Box–Jenkins SARIMA model and the other ANN models estimated on raw, detrended or deseasonalized data. However, without appropriate data preprocessing, the SARIMA technique produces more accurate forecasts.

Based on the above findings, the ANN model using both detrended and deseasonalized data has been selected to simulate future water consumption forecasting. The results show that by the end of 2025, the total water consumption for residential use will continue to increase and it will represent nearly 18% of the total water demand of the country versus only 8% in 2000. These findings might be considered by the policy-makers in planning and reviewing the water resource allocation between different sectors. They also suggest that more effective policy measures should be implemented to regulate household water consumption. These policies range from pricing policies by reviewing the volumetric water pricing used by SONEDE to non-pricing policies by, for example, the promotion of water-saving devices.

REFERENCES

Ahmed, J. A. & Sarma, A. K. 2007 Artificial neural network model for synthetic streamflow generation. *Water Res. Manage.* **21**, 1015–1029.

Aytek, A., Asce, M. & Alp, M. 2008 An application of artificial intelligence for rainfall-runoff modeling. *J. Earth Syst. Sci.* **117**, 145–155.

Azadeh, A., Neshat, N. & Hamidipour, H. 2012 Hybrid fuzzy regression-artificial neural network for improvement of short-term water consumption estimation and forecasting in uncertain and complex environments: case of a large metropolitan city. *J. Water Res. Plan. Manage.* **138**, 71–75.

Birikundavyi, S., Libib, R., Trung, H. T. & Rousselle, J. 2002 Performance of neural networks in daily streamflow forecasting. *J. Hydrol. Eng.* **7**, 392–398.

Bishop, C. M. 1995 *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Blue Plan 2004 L'eau des Méditerranéens : situation et perspectives. No. 158 de la Série des rapports techniques du PAM, PNUE/PAM, Athènes.

Bougadis, J., Adamowski, K. & Diduch, R. 2005 Short-term municipal water demand forecasting. *Hydrol. Process.* **19**, 137–148.

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

Chen, J. C., Chang, N. B. & Shieh, W. K. 2003a Assessing wastewater reclamation potential by neural network model. *Eng. Appl. Artif. Intel.* **16**, 149–157.

Chen, A., Leung, M. T. & Hazem, D. 2003b Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Comput. Operat. Res.* **30**, 901–923.

Firat, M., Turan, M. E. & Yurdusev, M. A. 2010 Comparitive analysis of neural network techniques for predicting water consumption time series. *J. Hydrol.* **384**, 46–51.

Firat, M., Yurdusev, M. A. & Turan, M. E. 2009 Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Res. Manage.* **23**, 617–632.

Gamal El-Din, A., Smith, D. W. & Gamal El-Din, M. 2004 Application of artificial neural networks in wastewater treatment. *J. Environ. Eng. Sci.* **3**, 81–95.

Hamdane, A. 2007 La gestion de la demande en eau d'irrigation, vision intégrale de l'eau agricole et futurs possibles : cas de la Tunisie. Paper presented at the conference: Water demand management in the Mediterranean: progress and policies Zaragoza, Spain, 19–21 March 2007.

Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice-Hall, Englewood Cliffs, NJ.

Herrera, M., Torgo, L., Izquierdo, J. & Pérez-Garcia, R. 2010 Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **387**, 141–150.

Hoff, J. C. 1983 *A Practical Guide to Box–Jenkins Forecasting*. Lifetime Learning Publications, London.

Jeong, D. & Kim, Y. O. 2005 Rainfall–runoff models using artificial neural networks for ensemble stream flow prediction. *Hydrol. Process.* **19**, 3819–3835.

MAHR (Ministry of Agriculture and Hydraulic Resources) 2010 Potential water resources in 2009. A Summary sheet.

McCleary, R. & Hay, R. 1980 *Applied Time Series Analysis for the Social Sciences*. Sage, Beverly Hills, CA.

Nagy, H. M., Watanabe, K. & Hirano, M. 2002 Prediction of sediment load concentration in rivers using artificial neural network model. *J. Hydraul. Eng.* **128**, 588–595.

Najah, A., Elshafie, A., Karim, A. O. & Jaffar, O. 2009 Prediction of Johor River water quality parameters using artificial neural networks. *Eur. J. Sci. Res.* **28**, 422–435.

Nelson, M., Hill, T., Remus, W. & O'Connor, M. 1999 Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting* **18**, 359–367.

Pallavi, B., Prasad, R. K. & Singh, V. S. 2009 Forecasting of groundwater level in hard rock region using artificial neural network. *Environ. Geol.* **58**, 1239–1246.

Pankratz, A. 1983 *Forecasting with Univariate Box–Jenkins Models: Concepts and Cases*. John Wiley & Sons, New York.

Shabri, A. 2001 Comparision of time series forecasting methods using neural networks and Box–Jenkins Model. *Matematika* **17**, 24–32.

Sharda, R. & Patil, R. B. 1992 Connectionist approach to time series prediction: an empirical test. *J. Intell. Manufact.* **3**, 317–323.

SONEDE (Société Nationale d'Exploitation et de Distribution des Eaux) 2004 Etude de tarification de l'eau potable.

Sreekanth, P. D., Geethanjali, N., Sreedevi, P. D., Ahmad, S., Ravi Kumar, N. & Kamala Jayanthi, P. D. 2009 Forecasting groundwater level using artificial neural networks. *Curr. Sci.* **96**, 933–939.

Šterba, J. & Hiľovská, K. 2010 The implementation of hybrid ARIMA-Neural network prediction model for aggregate water consumption prediction. *J. Appl. Math.* **3**, 123–130.

Tang, Z., Almeida, C. & Fishwick, P. A. 1991 Time series forecasting using neural networks vs Box–Jenkins methodology. *Simulation* **57**, 303–310.

Williamson, T. N. & Crawford, C. G. 2011 Estimation of suspended-sediment concentration from total suspended solids and turbidity data for Kentucky, 1978–1995. *J. Am. Water Resour. Assoc.* **47**, 739–749.

Zhang, G., Patuwo, B. E. & Hu, M. Y. 1998 Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting* **14**, 35–62.

Zhang, Y., Pulliainen, J., Koponen, S. & Hallikainen, M. 2002 Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sens. Environ.* **81**, 327–336.

Zhang, G. P. & Qi, G. M. 2005 Neural network forecasting for seasonal and trend time series. *Eur. J. Operat. Res.* **160**, 501–514.

First received 29 February 2012; accepted in revised form 22 November 2012. Available online 15 February 2013