Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2017, 4(2):98-113**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Modeling of Annual Municipal Water Demand System Using Soft Computing Techniques: A Case Study

Sahbi Boubaker[1,2]

[1]Department of Electrical Engineering, Community College, University of Hail, Saudi Arabia
[2]Research Unit on Study of Systems and Renewable Energy, National College of Engineering of Monastir, Tunisia

Abstract This paper addresses the problem of modeling the municipal water demand of Hail region, Saudi Arabia. Forecasting water demand is highly required for tactical decisions and efficient operation and management in both medium and long terms. In this study, the annual municipal water consumption has been modeled using three different approaches: artificial neural networks (ANN), stochastic time-series (STS) and particle swarm optimization (PSO). The three used modeling and simulation tools have been compared according to statistical performance indicators such as mean absolute percentage error (MAPE), coefficient of determination (R^2), root mean squared error (RMSE) and absolute relative error (AARE). For the current case study, the Feed-Forward Back Propagation ANN has been found to outperform the STS and PSO-based approaches because of the relatively small data-set (records) and its ability to model non-linear systems. Future perspectives of the present study can be built around developing more sophisticated models for monthly and daily municipal water consumption involving external parameters such as climatic and socio-economic variables.

Keywords Modeling, Forecasting, Water demand, Artificial Neural Network, Particle Swarm Optimization, Stochastic time-series, Performance indicators.

1. Introduction

Municipal water demand forecasting is considered nowadays as one of the key issues for water utilities to ensure both good planning and design in long-term and efficient operation and management in short and medium terms. In fact, many countries around the world are faced to severe water shortages and are consequently invited to manage adequately their water resources.

Arid climate countries, such as Saudi Arabia, are in particular experiencing low rain fall conditions and rapid urbanization associated with intensive use of water resources. Maintaining a steady and regular supply of water under climate change regimes and socio-economic factors is a major concern of water utilities.

In this paper, the annual municipal water demand of Hail region, Saudi Arabia, is considered. It is assumed to be affected by only its previous values. During the past few decades, the field of forecasting water demand has been under intensive consideration by the water researchers. Therefore, several papers considering both the models and the tools used to perform these models have been published. The models concern the description of the relationships that characterize the water demand and its different determinants; whereas, several approaches have been developed to calibrate these relationships. Box-Jenkins-based approaches [2-10, 30-32, 35, 37-38] in their different forms have been extensively (and continue to be) used. Artificial neural networks as universal modeling tool has been also used [11-19] although it suffers from many limitations such as the high training time. In recent years, a clear trend to combined approaches based particularly on swarm computation technique has been observed [20-29, 33-34, 36]. From a forecasting horizon point of view, short, medium and long term forecasters have been designed. In what follows, this paper will concentrate on the annual water demand prediction tools viewed as a dynamic system.

In their survey paper [1], the authors presented a summary of the papers on the water forecasting techniques between 2000 and 2010. They concluded that the wide variety of methods used depend on the forecast variables (climate and socio-economic) and on the forecast horizon (short, medium and long term). They also stressed the fact that the water forecasting problem is a difficult one because of the nature and the quality of the available data [1].

From both theory and practice findings, researchers have concluded that combining different tools is effective in forecasting [7]. Time series, as modeling tool, combined with artificial neural networks has been widely used. The problem of forecasting ARIMA models with ANN has been in particular addressed in [7]. The authors take benefit from the structure of the ARIMA to implement it with ANN. Results of three well-known time series have proved the efficiency of the proposed approach especially when higher accuracy is required.

Using IWR-MAIN and SPSS softwares, water demand has been forecasted in [10] for Umm Al-Quwain Emirate located in the north of the United Arab Emirates studying the effect of three independent variables, namely, population, average temperature and average rainfall. The obtained results revealed that population growth is the most significant variable that affects water demand. The available data have been divided into two sets. The first one is used to test correlation between water demand and the three explanatory variables and the second set is used to predict the consumption by comparing the real demands against the model demands.

Because of their ability to model any process without need of specific requirements, artificial neural networks (ANN) have been extensively used in forecasting [11-19]. Many case studies from countries around the world have been investigated. In general, depending on quality and quantity of available data, the results are case-sensitive. No universal model and no specific structure of the ANN have given similar results even for the same study area. In the study presented in [11] the future forecasted climatic variables have been used for forecasting future water demand. Both annual and monthly consumptions have been predicted using ANN. A sensitivity analysis conducted in this study has revealed that climatic variables have very little effect on annual demand. However, the climatic variables affect strongly monthly water demand. Root mean square error (RMSE), correlation coefficient (R^2) and average absolute relative error (AARE) have been used as performance indicators of the proposed approach. The novelty of this study is that, at the best of our knowledge, it is the first work that starts by forecasting climatic variables values and then uses them to forecast water consumption.

Swarm intelligence techniques become more and more used in forecasting future time series [20-25]. The problem of identifying the prediction model parameters is an identification one and can be posed as an optimization framework involving the error function as a fitness function and the model parameters as decision variables. The problem of forecasting short term electric load based on particle swarm optimization (PSO) and ARMAX (Auto Regressive Moving Average with exogenous variable) has been addressed in [20]. More than being difficult, the studied forecasting problem has many local minimum points. Therefore, classical optimization techniques can easily stall at their local minimum points and consequently can't find the global minimum. This study has taken benefit from the global search capabilities of PSO which has been found to outperform traditional stochastic time series approaches and evolutionary programming. Annual electric load forecasting in India, using PSO has been detailed in [23]. The population, per capital Gross Domestic Product (GPD) and import and export data have been used as variables affecting the load. The parameters of an exponential model have been identified by minimizing the squared error function between the observed loads and the predicted ones. The available data have been divided into two sub-sets: training and testing. The mean average percentage error (MAPE) was used as the performance evaluation criterion.

The study presented in [33] has investigated and compared the relative performances of a genetic algorithm and particle swarm optimization approaches in forecasting annual energy demand in residential and commercial sectors of Iran based on recorded annual economic indicators. Near-optima values for coefficients of a linear and an exponential model have been adjusted. The PSO algorithm combined to the simulated annealing (SA) has been proved to accurately forecast urban water demand in a region in China. The proposed PSO-SA approach [34] performed better than the PSO-based approach in term of mean squared error. The annual long-term forecasting problem for the desalinated dependent city of Riyadh in Saudi Arabia has been studied in [37]. A probabilistic model that incorporates explicitly the uncertainties associated with population growth, household size, household income as well as conservation measures has been developed and evaluated. As reported by the authors, the results show that future water demand is governed equally by socio-economic factors and weather conditions. Conservation measures, leakages' detection and efficient pricing policy are highly required. Water demand in the city of Mecca, Saudi Arabia, is explained by the number of pilgrims visiting the city for the annual Hajj pilgrimage. In fact, water consumption in this city is characterized by trends and cyclic variation [38]. Moreover, water demand has been demonstrated to be affected by the temperature especially in the lunar month of Hajj and consequently annual water demand is influenced. Since Saudi Arabia is facing a chronic water-shortage problem, the demand-supply gap needs to be minimized. The study presented in [39] suggests to water utilities to focus on the agriculture sector known to be the most responsible of water demand increase. The case study of modeling annual water demand and supply in the city of Istanbul, Turkey, has been investigated in [41]. Using an autoregressive integrated moving average (ARIMA); it has been found that the residential sector is responsible of 80% of total water use. A sustainable management plan based on reducing water demand by using efficient technologies in households and reducing water supply loss through investments on distribution infrastructure have been suggested by the authors for the coming years. In the technical paper [55], forecasting benchmark time-series has been conducted using a hybridization of the ARIMA and ANN to take advantage of the strength of each method (linear modeling for the ARIMA and nonlinear modeling for the ANN). Experimental results show an increasing accuracy of the combined method when compared to the accuracy achieved by either of the two techniques used separately.

Reliable municipal water demand forecasting is one of the most important considerations for planning in long term. Satisfying water consumers' need is the main objective of managing water resources by the responsible utilities. Hail region, as a part of Saudi Arabia, is located in an extreme desert environment where the yearly rainfall is less than 100 mm. Water resources are limited while dispread development coupled with rapid increase in water demand. The present study will focus on:

- Analyzing annual water demand in Hail region, Saudi Arabia.
- Determining the effect of the historical records on the one-year-ahead water demand.
- Proposing, identifying and validating forecasting time-series models based on different techniques.
- Comparing the developed models according to different performance indicators.

The main contribution of this study is to apply Artificial Neural Network (ANN), Time Series (TS) and Particle Swarm Optimization (PSO) in order to model annual municipal water demand in Hail City, Saudi Arabia. Annual water consumption time series is first collected and analyzed. To measure the effect of the historical records on municipal water demand, correlation coefficients are determined and then the potential explanatory historic lags are selected. An annual MWD prediction model is defined as a function of its previous values. The problem of MWD forecasting is posed as an optimization problem where the error between the predicted values and the observed ones is to be minimized. The decision variables are the parameters of the proposed model structure. Known to be difficult, non-linear and stochastic, this problem solving is faced to multiple limitations when tackled using traditional techniques. In this study, we have resort to non-conventional techniques from the field of soft computing. Integrating different tools is found in the literature to be effective in predicting water demand. The efficiency of the developed models are measured by diverse performance indicators such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of determination (R^2), etc.

The remaining of the paper is organized as follows. In the next section, the basic concepts of stochastic time series (STS), artificial neural networks (ANN) and particle swarm optimization (PSO) are briefly reviewed as forecasting and modeling tools. In section 3, detailed application of the presented methodologies on modeling and simulation of annual municipal water demand in Hail region, Saudi Arabia, is presented. Results are then compared according to some performance indicators. Concluding remarks and recommendations are drawn in section 4.

2. **Proposed Methodologies**

2.1. Box-Jenkins stochastic time series

Municipal water demand measurements are usually recorded following certain frequency (daily, weekly, monthly or annual). The box-Jenkins based method has been used in water demand forecasting [1-10]. This approach is mainly used for modeling stationary (or stationarized time-series). The study of the autocorrelation function (ACF) of the successive records is used to determine the orders of the model.

In its general form, the Box-jenkins methodology is used to model the ARIMA (p,d,q) model [3-4].The ARIMA model for a time-series $y(t)$ is given by:

$$y(t) = C + a_1 y(t-1) + \cdots + a_p y(t-p) + e(t) + b_1 e(t-1) + \cdots + b_q e(t-q) \tag{1}$$

C is a constant value, $a_1, \ldots, a_p$ are the coefficients of the autoregressive part where p is its order and $b_1, \ldots, b_q$ are the coefficients of the moving average part and q is its order. $e(t)$ is a white noise sequence. In the present study, the autoregressive (AR) component of the ARIMA model is used. If the original time-series is not stationary, a differencing procedure is applied to stationarize it. The number of difference operator is usually determined by the study of the ACF plots. Once the values of $p, d\ and\ q$ are determined, the coefficients $C, a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q$ are estimated using any optimization tool.

2.2. Artificial Neural Networks

ANNs have been extensively used in predicting time-series related to water consumption [11-19]. Combined approaches including ANN remain the most used ones. Because of its universal feature, ANN has gained attention of researches in different fields. The finding of the ANN is based on the imitation of the human brain behavior. Usually, an ANN is composed of layers. The ANN neurons are fed by the inputs. The neurons of each layer are connected through weights. A set of transfer functions ensure the evolution of the signals between the layers. The ANN involves the choice of an appropriate architecture, the size of each layer and the suitable transfer functions between the layers. The ANN is trained using a special algorithm called training algorithm which much the inputs to the desired outputs.

2.3. Particle Swarm Optimization

PSO algorithm was first introduced by Kennedy and Eberhart in 1995. It is categorized as one of the population based swarm intelligence techniques such as genetic algorithms (GA), ant colony (AC), etc. Parameters tuning of the algorithm, its variants and its applications can be found elsewhere (e.g. [21], [56], and [27]). PSO is preferred in many applications for its simplicity, its universal aspect and the fact that it does not require any regularity of the studied problem (continuity, differentiability and convexity). Although it has been used in forecasting electric load ([20-23], [33]), its use in water demand forecasting remains limited.

In the present study, we intend to take benefit from the features of the PSO to forecast the MWD in Hail region. In this study, MWD forecasting is posed as (or transformed into) an optimization problem. This problem is defined as in equation (2).

$$\min_{\vec{X}} f(\vec{X}) \tag{2}$$

PSO is based on the concept of gradually evolving a swarm of possible solutions for an optimization problem. Figure 1 illustrates the move of the particle. The potential solution is coded as the position of the i^{th} particle in the d-dimensional search space:

$$\vec{X}^i = \left(X_1^i, X_2^i, \ldots, X_d^i\right) \tag{3}$$

Let $X_1, X_2, \ldots, X_d$ be the decision variables and k the iteration index of the optimization problem (2). During the search process, each particle in the swarm adjusts its position through a velocity operator defined as:

$$\vec{V}_{k+1}^i = w_k . \vec{V}_k^i + c_1 \times rand_1\left[\vec{P}_k^i - \vec{X}_k^i\right] + c_2 \times rand_2\left[\vec{G}_k^i - \vec{X}_k^i\right] \tag{4}$$

where:

- c_1, c_2 are two constant factors
- $rand_1, rand_2$ are random numbers in the [0,1] range
- $\vec{P}_k^i$ is the best previous position of the i^{th} particle

- $\vec{G}_k^i$ is the best previous position of the whole swarm
- $w_k = w_{max} - \frac{w_{max} - w_{min}}{k_{max}} \times k$ is a weighting function
- w_{max} *and* w_{min} are respectively the maximum and the minimum value of w_k

The position vector of the i^{th} particle is then given by (5).

$$\vec{X}_{k+1}^i = \vec{X}_k^i + \vec{V}_k^i \tag{5}$$

Figure 2 provides the flowchart of the PSO and Figure 3 provides its pseudo-code.

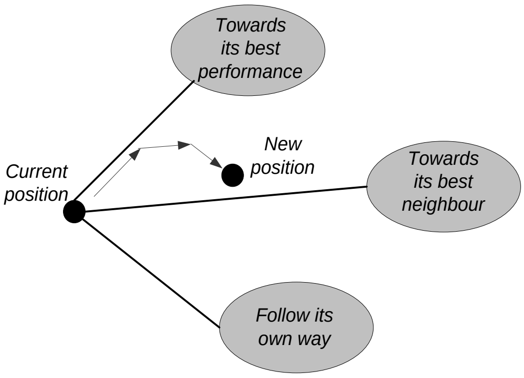

Figure 1: Particle's move

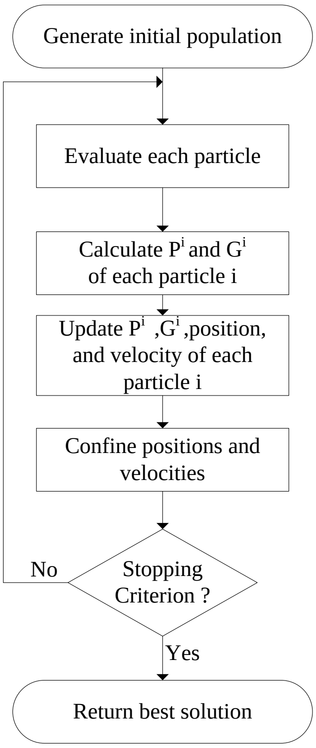

Figure 2: Flowchart of the PSO

(nb_part: swarm size; f :fitness function
Begin
Random initialization of positions
and velocities of each particle
***For** each particle i ,* $\vec{P}^i = \vec{X}^i$
While** stopping criterion not met **do
For** i=1 to nb_part **do
Update the particle position and velocity
according to equations (5) and (6)
Evaluate the new fitness functions
If $f(\vec{X}^i) < f(\vec{P}^i)$
$\vec{P}^i = \vec{X}^i$
If $f(\vec{P}^i) < f(\vec{G}^i)$
$\vec{G}^i = \vec{P}^i$
End for
End while

Figure 3: Pseudo-code of the PSO algorithm

3. Results and Discussions

This section discusses the progressive development of number of models for forecasting annual MWD in Hail region, Saudi Arabia (see Figures 4 and 5) using respectively ANN, STS and PSO. To compare agreement between the forecasted and observed consumptions, graphical plots and diverse performance indicators have been used. Moreover, comparison between different approaches has been investigated. Details about the study area can be found in [61].

Figure 4: Geographic location of Hail region, Saudi Arabia

Figure 5: Principal water distribution locations in Hail region (source: projects report from 1425 to 1433 H)

To model annual MWD in Hail region, we use a sample data composed of 15 records (from 2000 to 2014) collected from General Directorate of Water in Hail Region. The number of values is limited and does not meet the usefulness requirements of time series based approaches (at least 50 records are required for a reasonable analysis for stochastic TS). Nevertheless, the sample data are divided into model development part (from 2000 to 2009) and model validation part (from 2010 to 2014) depending on the case study. The plot of the autocorrelation function of the annual MWD (Figure 6 (a)) reveals that this last is non-stationary. Although, this function has no meaning in the case of small data-sets, stationnarizing this TS, decreases the sample size without guarantee to obtain stationary time series. In this research, the annual water demand TS should be used as it is in any task related to forecasting.

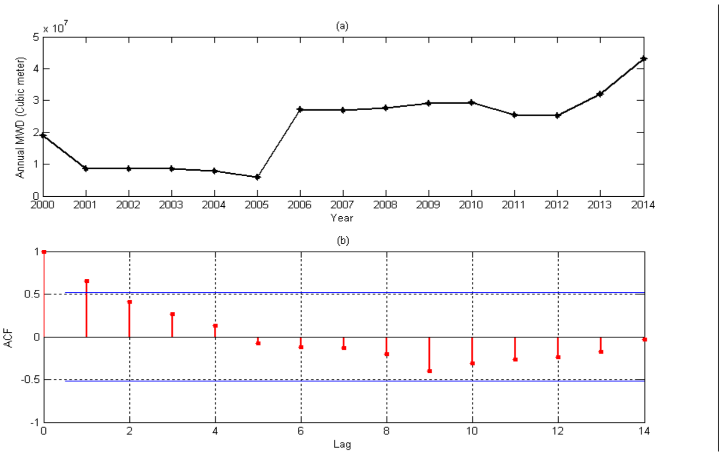

Figure 6: (a) Annual MWD in Hail region (b) Sample autocorrelation function of annual MWD time series

3.1. BP-FF-ANN

An ANN model is first developed for predicting annual MWD in Hail region. The proposed ANN is a feed-forward four layer perceptron (one input layer, two hidden layers and one output layer). The maximum number of inputs is chosen to be the five previous municipal water demands ($MWD(t-1), MWD(t-2), MWD(t-$ *3, MWDt−4, MWD(t−5)*). This choice is dictated by the limited number of available data. A schematic representation of the used ANN is depicted in Figure 7.

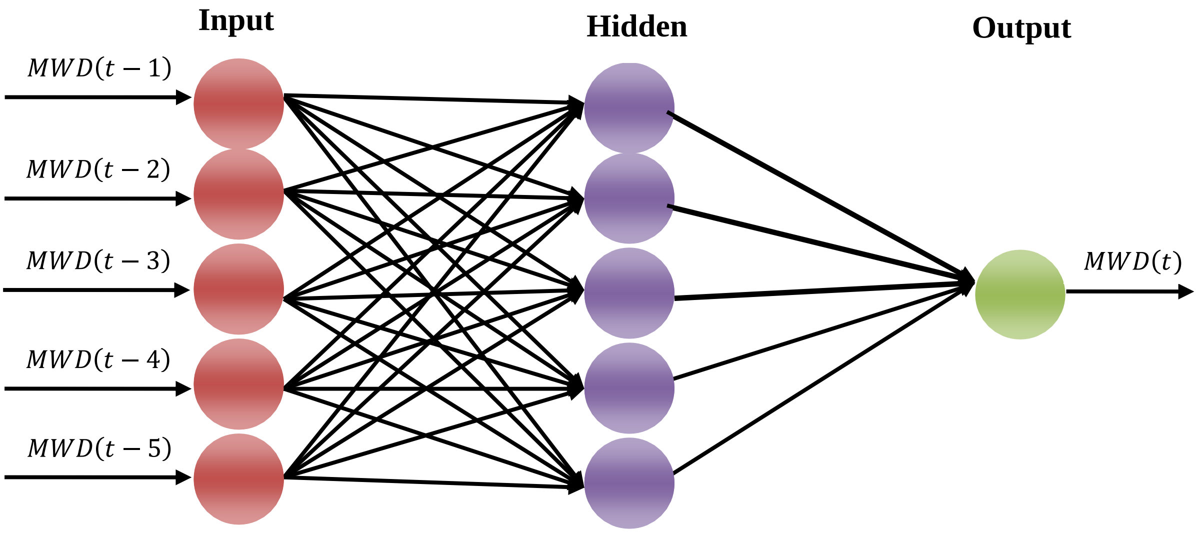

Figure 7: Block diagram of the proposed ANN with five inputs for annual MWD forecastin

The number of neurons in each layer and the activation functions have been decided by trial and error procedure. Table 1 presents the results of the annual MWD using feed-forward multilayer perceptron ANN with different architectures. For comparison purposes against other techniques, the number of inputs has been varied from 1 to 5 ($MWD(t-1), MWD(t-2), MWD(t-3), MWD(t-4), MWD(t-5)$). The good fit of the forecasted values is measured via the mean average percentage error (MAPE), the coefficient of determination (R^2), the root mean squared error (RMSE) and the average absolute relative error (AARE) defined as follows [61]:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| / \bar{y} \tag{6}$$

$$R^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{8}$$

$$AARE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{9}$$

where $\hat{y}_i$ is the forecasted water demand, y_i is the i^{th} actual water demand, $\bar{y}$ is the mean water demand and N is the number of observations. As no direct method exists to determine the best number of inputs for the ANN, we have resort to try different historical lags combinations. We started by using the two previous water demands ($MWD(t-1)$ and $MWD(t-2)$) as inputs and we have after that increased this number of neurons in the hidden layers as well as the activation functions have been decided through a trial and error procedure.

Table 1: Structure and Performance Indicators of the Best BP-FF-ANN(P)

Best model structure						Performance measures			
Model	**Architecture**	**Input variables**	**Neurons**	**Activation function**	**Training algorithm**	**MAPE**	**R^2**	**RMSE**	**AARE**
$BP-FF-ANN(2)$	$1-2-1$	$MWD(t-1)$ $MWD(t-2)$	$15-15-15-1$	tansig, tansig, tansig, purelin	CGB	13.6296	0.7162	584.2048	10.3031
$BP-FF-ANN(3)$	$1-2-1$	$MWD(t-1)$, …, $MWD(t-3)$	$15-15-15-1$	tansig, tansig, tansig, purelin	CGB	**6.4407**	**0.8693**	**382.699**	**5.7361**
$BP-FF-ANN(4)$	$1-2-1$	$MWD(t-1)$, …, $MWD(t-4)$	$15-15-15-1$	tansig, tansig, tansig, purelin	CGB	26.6907	-1.0216	1431.1	19.3547
$BP-FF-ANN(5)$	$1-2-1$	$MWD(t-1)$, …, $MWD(t-5)$	$15-15-15-1$	tansig, tansig, tansig, purelin	CGB	116.3321	-39.3343	5488.2	106.5136

Table 1 shows the structure and the performance indicators of the best $BP-FF-ANN(p)$, where p is the number of inputs, $p = 2, \dots, 5$ (inputs here are the p previous annual $MWDs$). Figure 8 shows the observed MWD in Hail region versus the forecasted values.

Due to the lack of variables that can affect the annual MWD such as average annual water bill, population, number of households, gross national product and number of subscribers, etc [14] we were limited only to the use of historical records of annual MWD.

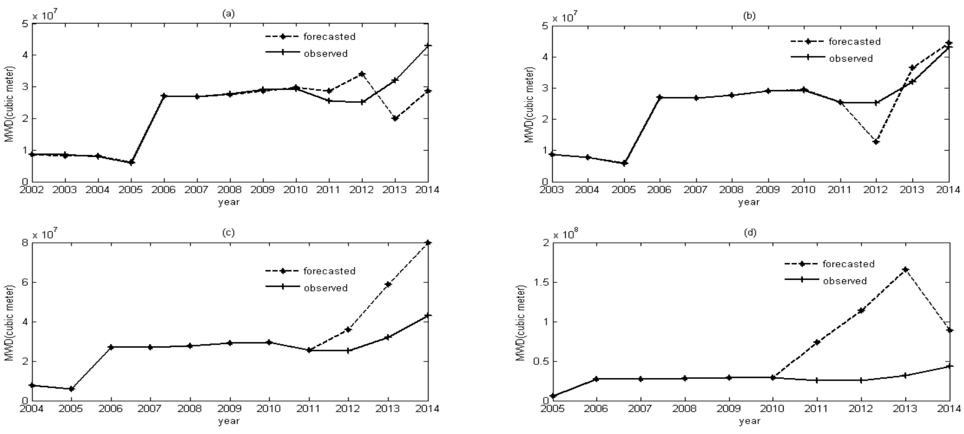

Figure 8: Forecasted and observed annual MWD for Hail region obtained with ANN based methods (a) BP-FF-ANN(2) (b) BP-FF-ANN(3) (c) BP-FF-ANN(4) (d) BP-FF-ANN(5)

$\mathrm{BP-FF-ANN(3)}$ is found to outperform $\mathrm{BP-FF-ANN(2)}$, whereas $\mathrm{BP-FF-ANN(4)}$ and $\mathrm{BP-FF-ANN(5)}$ deteriorate giving bad forecasts for both training and validation data. MAPE was found to be 6.4407% and R^2 equal to 0.8693. $\mathrm{BP-FF-ANN(3)}$ is found to be the best model because it has the highest R^2 and the minimum MAPE. We remember here, that, after using many training algorithms such as LM ($Levenberg-Marquardt$, RP ($Resilient\ Back\ Propagation$), CGB ($Conjugate\ Gradient\ Powell–Beale$), we found that the CGB training algorithm is the best one [17].

3.2. Auto-Regressive (AR) Time Series

In this section, the auto-regressive (AR) time series is used to predict annual water demand in Hail region. The AR time series has the advantage of forecasting values of a TS using only historical data. However, it is known to require stationary data having at least 50 values. Since the size of the records available for this study is limited to 15 values, and since the annual $MWDTS$ has been demonstrated to be non-stationary, we intend here to try the feasibility of the forecasting task even in these conditions. The adopted TS model is here given by:

$$A(z).\,y(t) = e(t) \tag{10}$$

- $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}$
- p is the AR model order
- z^{-1} is the delay operator
- $a_1, a_2, \dots, a_p$ are the regression coefficients
- $\mathrm{e(t)}$ is a white-noise signal
- $\mathrm{y(t)}$ is the actual annual MWD

Although its size is limited, the TS is divided into two subsets (one sub-set for developing the model and the other one for validating the model). Similarly to the study using the artificial neural networks, the maximum value for the order p is pre-set to five. Then, four AR models respectively denoted $AR(2), AR(3), AR(4)$ and $AR(5)$ are used to investigate the relationship between the current MWD and its p previous values. All computations are performed in Matlab environment.

Although it has been found to violate the stationarity condition, the annual water demand in Hail region TS, has been used to adjust the parameters of the AR models having the advantage of being simple and can be easily implemented [17]. The results obtained using the 'AR' command of Matlab are summarized in Table 2.

Table 2: Structure and Performance Indicators of the Auto-Regressive (AR) Time Series

Model	Variables					Training data	Validation data	Performance measures				
	a_1	a_2	a_3	a_4	a_5			$MAPE$	R^2	$RMSE$	$AARE$	AIC
$AR(1)$	-	-	-	-	-	from 2000 to 2010	from 2001 to 2014	23.5568	0.4999	7.9096×10^{10}	24.1668	32.0026
$AR(2)$	-0.8712	-0.006811	–	–	–	from 2000 to 2010	from 2002 to 2014	**20.0472**	**0.5202**	**7.5963×10^{10}**	**17.2452**	**32.1425**
$AR(3)$	-0.8687	-0.07489	0.00933	–	–	from 2000 to 2010	from 2003 to 2014	20.8710	0.4340	7.9654×10^{10}	18.7524	32.4828
$AR(4)$	-0.8596	-0.08748	-0.05088	0.09352	–	from 2000 to 2010	from 2004 to 2014	22.4161	0.2429	7.6473×10^{10}	21.0137	32.8541
$AR(5)$	-0.9756	-0.09482	-0.08161	-0.199	0.1724	from 2000 to 2010	from 2005 to 2014	21.2325	0.1603	7.9186×10^{10}	22.7122	33.0201

In addition to the performance criteria used for the ANN models, the Akaike Information Criterion (AIC) is used here to measure the relative quality of the developed AR models for the given set of data. It provides a mean of model selection regardless the quality of the models themselves.

Figure 9 shows the observed and forecasted annual MWD of Hail region obtained from auto-regressive TS models. From the obtained results, it appears that all the developed models have relatively low R^2 values ranging from 0.1603 (for $AR(5)$) to 0.5202 (for $AR(2)$). $AR(2)$ is found to be the best model with the best $MAPE, R^2, RMSE$ and $AARE$. In fact, it has the highest coefficient of determination. Moreover, $AR(2)$ has the smallest AIC value (32.1425) which means that regardless the relatively moderate quality of the AR models, it appears to be the best one. Intuitively, we have tried the $AR(1)$ and we discovered that the performance deteriorate in terms of $MAPE$ and R^2.

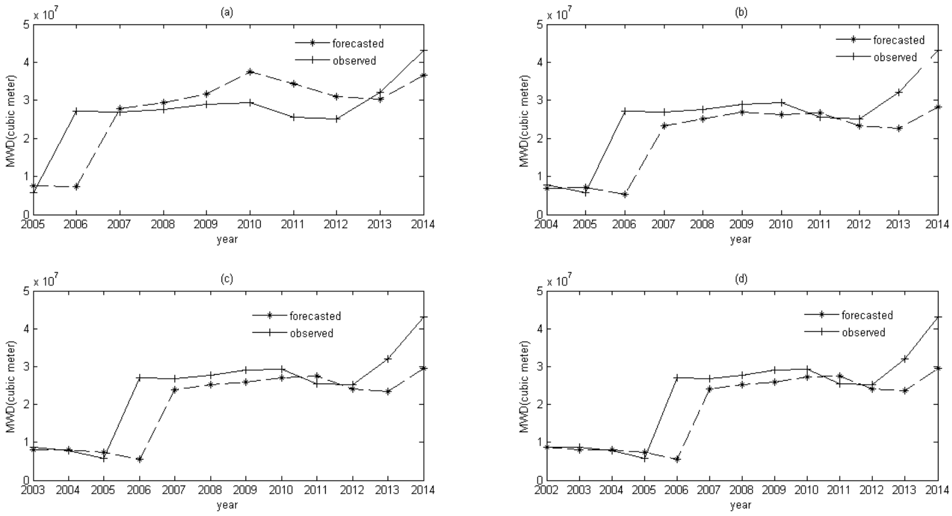

Figure 9: Forecasted and observed annual MWD for Hail region obtained with TS AR based methods (a) AR(5) (b) AR(4) (c) AR(3) (d) AR(2)

3.3. Particle Swarm Optimization for annual MWD prediction

Since it is based on stochastic TS, modeling annual MWD in Hail region is classified from an optimization point of view as a difficult problem. Moreover, minimizing the error between the forecasted and observed TS of annual MWD (function of the forecasting model parameters) is in general non-convex. Classical optimization techniques used to solve this problem can easily fall into local minima. For these reasons, we have resort to a general technique of optimization (the PSO) to tune the parameters of the forecasting models. In order to perform a comparative study between the proposed methods, the AR model structure is adopted here and the PSO is used to tune its parameters.

Let $\vec{X}=\left[a_1, a_2, \dots a_p\right]$ be the position of a particle in the swarm. $a_1, a_2, \dots$ and $\dots a_p$ are the coefficients of the autoregressive model presented in (11). Four models, respectively named $PSO-AR(2), \dots, PSO-AR(5)$ are developed and tested for fitting annual water demand. The fitness function to be minimized is defined as the squared sum error between the forecasted and the observed values. The problem to be solved is:

$$\begin{cases} min \sum_{i=1}^{N} \left(y_i(t) - \hat{y}_i(t)\right)^2 \\ \hat{y}_i(t) = \sum_{j=1}^{p} a_j \, y(t-j) + e(t) \end{cases} \tag{11}$$

The following synthesis parameters of the PSO have been adopted: $nb\_part = 20$, $nb\_var = p$, $c_1 = c_2$=1.49, $w_{max} = 0.9$ and $w_{min} = 0.4$. The stopping criterion is chosen to be a maximum number of iterations: $maxIter\_nb = 100$ [56].

For brevity, only results for the optimal parameters of $PSO-AR$ models and the performance indicators are presented without details of the PSO-based approach implementation. The results are summarized in Table 3. Plots of the estimated and observed values are depicted in Figure10. From the obtained results, it is concluded that the $PSO-AR(2)$ outperforms the three other approaches in terms of the four performance indicators. In fact, it has the best $MAPE$ (18.0845), $R^2(0.6210)$, $RMSE(6.7753 \times 10^4)$ and $AARE(18.9673)$. This result was predictable because this method operates using more data than the other methods. Moreover, minimizing the search-space dimensions provides to the PSO the opportunity to search in a less complicated environment.

Table 3: Structure and Performance Indicators of PSO-Based Approaches

Model	Optimal parameters					Performance measures			
	a_1	a_2	a_3	a_4	a_5	$MAPE$	R^2	$RMSE$	$AARE$
$\boldsymbol{PSO-AR(2)}$	**1.0068**	**0.0662**	–	–	–	**18.0845**	**0.6210**	$\mathbf{6.7753\times 10^4}$	**18.9673**
$PSO-AR(3)$	1.0028	-0.0132	0.1105	–	–	20.7993	0.6066	7.4519×10^4	23.3042
$PSO-AR(4)$	0.9804	-0.1741	0.1847	0.1284	–	23.6603	0.5508	8.1203×10^4	27.6980
$PSO-AR(5)$	0.9805	-0.1699	0.1835	0.1273	0.0000	25.1913	0.4189	8.7947×10^4	28.2362

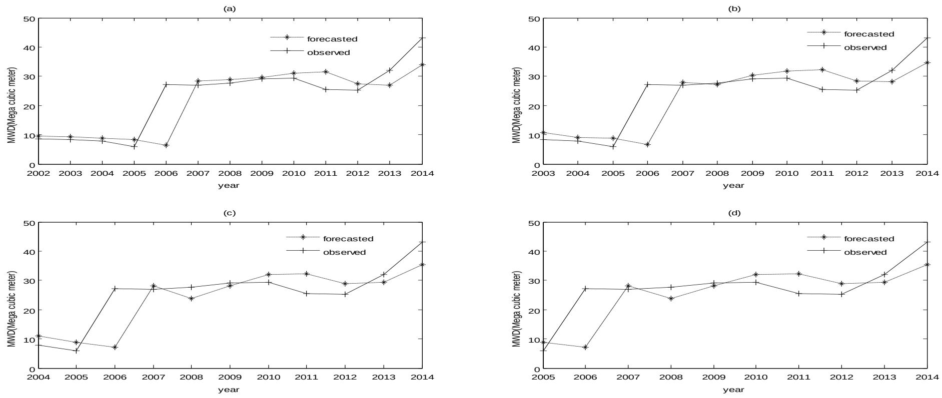

Figure 10: Forecasted and observed annual MWD for Hail region obtained with PSO-AR based methods (a) PSO-AR(2) (b) PSO-AR(3) (c) PSO-AR(4) (d) PSO-AR(5)

3.4. Comparative analysis of the annual MWD forecasts

In this section, the forecasting capabilities of the best $AR, ANN\ and\ PSO$ based models for predicting annual MWD in Hail region are compared. The performance indicators of the best $BP-FF-ANN(3), AR(2) and\ PSO-AR(2)$ are summarized in Table 4 below.

Table 4: Comparison of the Performance Indicators of the Three Best Forecasters

Model	$MAPE$	R^2	$RMSE$	$AARE$
$\boldsymbol{BP\text{-}FF\text{-}ANN(3)}$	**6.4407**	**0.8693**	**382.699**	**5.7361**
$AR(2)$	20.0472	0.5202	7.5963×10^{10}	17.2452
$PSO\text{-}AR(2)$	18.0845	0.6210	6.7753×10^4	18.9673

Results show that $\mathrm{BP-FF-ANN(3)}$ outperforms the two other approaches in terms of all performance indicators. In fact, it provides the best results with lowest $MAPE\ of\ 6.4407\%$ and lowest $AARE\ of\ 5.7361\%$. Moreover, it provides the highest $R^2\ of\ 0.8693$ (for both training and testing sets) and a lowest $RMSE\ of\ 382.699$. Although it takes more time in looking for the best synthesis parameters, the ANN based forecaster is found to be the best approach for predicting annual MWD in Hail region. The ANN technique is found to provide better results with limited amount of data. Since the data set has been partitioned into two sub-sets: training and testing, and both sub-sets have been used for model validation, the training data give a 100% degree of accuracy in fitting observed and forecasted annual MWD. The ANN- based approach is found here to be insensitive to the data non-stationarity. This approach produces a highest level of accuracy if we use a large number of samples, which is unfortunately unavailable. To finish this discussion, we can indicate (based in Table 4 values) that the $AR\ and\ the\ PSO-AR$ forecasters have moderate and comparable forecast accuracy with little superiority of $PSO-AR(2)$.

4. Conclusions and Recommendations

Water utilities in Hail region need to know what the water demand for today, tomorrow [1], next month and next ten years will be, to ensure the water needs in the best conditions.

In this study, three frameworks from the fields of traditional stochastic time-series (STS), artificial neural networks (ANN) and swarm intelligence (PSO) have been used to model one-year-ahead municipal water demand in Hail region, Saudi Arabia. A data set composed of 15 records (from 2000 to 2014) collected from Hail Directorate of Water (HDW) has been divided into two sub-sets: model development sub-set and model validation sub-set. Since many explanatory variables that affect annual MW consumption (such as water bill, population,…) are non-available, we were limited to the historical values of monthly MWD to model its future values. The ANN approaches require data re-scaling whereas the STS-based approaches require the original data to be stationary (or eventually stationarized). PSO-based approaches do not require any special condition since they are based on error function (between actual and forecasted water consumptions) computation. The best model is then selected on the basis of its performance as measured by the mean absolute of percentage error ($MAPE$), the coefficient of determination (R^2), the root mean square error ($RMSE$) and the average absolute relative error ($AARE$). Three approaches, namely BP-FF-ANN, AR, and PSO-AR have been developed and investigated. All these methods lead to relatively "good" fit and found to be adequate judged by their performance indices. Results show that by using the ANN-based approach, the overall forecasting errors have been significantly reduced. The ANN approach is found to be the most parsimonious model compared to other approaches. To the best knowledge of the author, this research is the first study that concerns with water resources in Hail region, Saudi Arabia. Future municipal water demand forecasting models are developed and tested. Consequently, water consumption is analyzed. From this study, the following recommendations can be mentioned:

- Since forecasting annual MWD is classified as long-range, it can participate to efficient planning and management of existing water supply system. Hail water utilities can take benefit from the results of this study to develop efficient plans for optimized system operation and eventual future expansions.
- Results obtained for annual MWD by similar studies ([10], [11]) demonstrated that using other explanatory variables (such as population statistics, water bill, subscribers' number, climatic variables, etc) can improve forecasting results accuracy.
- Hail water utilities can use this study in balancing water need and supply and reducing water loss.
- The developed annual MWD in Hail region are fundamental in taking decisions in water management issues [13]. Establishing efficient pricing policies, planning new system developments and optimizing sizes and operation procedures of the supply system (pumps, reservoirs) are among these issues.

As in [11], a future issue will be based on forecasting climatic and socio-economic variables first and after that use them to forecast water consumption.

Acknowledgment

This research has been granted by the Deanship of Scientific Research in Hail University, Saudi Arabia under the number E22 CC.

References

[1]. E. A. Donkor, T. A. Mazzuchi, R. Soyer, and J. A. Roberson, Urban water demand forecasting: Review of methods and models, Journal of water resources planning and management, vol. 140, pp. 146–159, 2014.

[2]. S. L. Zhou, T. A. McMahon, A. Walton, and J.Lewis, Forecasting daily urban water demand: a case study of Melbourne, Journal of Hydrology, vol. 236, pp. 153–164, 2000.

[3]. M. Herrera, L. Torgo, J. Izquierdo, and R. Perez-Garcia, Predictive models for forecasting hourly urban water demand, Journal of Hydrology, vol. 387, pp. 141–150, 2010.

[4]. M. Herrera, J. C. Garcia-Diaz, J. Izquierdo, and R. Perez-Garcia, Municipal Water Demand Forecasting: Tools for Intervention Time Series, Journal of Stochastic analysis and Applications, Taylor & Francis, vol. 29, pp. 998–1007, 2011.

[5]. Y. Bai, P. Wang, C. Li, J. Xie, and Y. Wang, A multi-scale relevance vector regression approach for daily urban water demand forecasting, Journal of hydrology, vol. 517, pp. 236–245, 2014.

[6]. G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Journal of Neurocomputing, vol. 50, pp. 159–175, 2003.

[7]. M. Khashei, and M. Bijari, "An artificial neural network (p,d,q) model for time-series forecasting," Journal of Expert Systems with Applications, vol. 37, pp. 479–489, 2010.

[8]. A. Yasar, M. Bilgili, and E. Simsek, "Water demand Forecasting Based on Stepwise Multiple Nonlinear Regression Analysis," Arab Journal of Sciences of Engineering, vol. 37, pp. 2333–2341, 2012.

[9]. S. Gato, N. Jayasuriya, and P. Roberts, "Forecasting Residential Water Demand: Case Study," Journal of water resources planning and management, vol. 133, pp. 309–319, 2007.

[10]. M. Mohamed, and A. Al-Mualla, "Water Demand Forecasting in Umm Al-Quwain (UAE) Using the IWR-Main Specify Forecasting Model," Journal of Water Resources Management, vol. 24, pp. 4093–4120, 2010.

[11]. M. S. Babel, N. Mapron, and V. R. Shinde, "Incorporating Future Climatic and Socioeconomic Variables in Water Demand Forecasting: A Case Study in Bangkok," Journal of Water Resources Management, vol. 28, pp. 2049–2062, 2014.

[12]. J. Adamowski, H. F. Chan, S. O. Prasher, B. Ozga-Zielinski, and A. Sliusarieva, "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for water demand forecasting in Montreal, Canada," Journal of Water Resources Research, vol. 48, W01528, pp. 1–14, 2012.

[13]. J. Boughadis, K. Adamowski, and R. Diduch, "Short-term municipal water demand forecasting," Journal of Hydrological Processes, vol. 19, pp. 137–148, 2005.

[14]. M. Firat, M. A. Turdusev, and M. E. Turan, "Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling," Journal of Water Resources Management, vol. 23, pp. 617–632, 2009.

[15]. C. C. Dos Santos, and A. J. Pereira Filho, "Water Demand Forecasting Model for the Metropolitan Area of Sao Paulo, Brazil," Journal of Water Resources Management, vol. 28, pp. 4401–4414, 2014.

[16]. M. A. Al-Zahrani, and A. Abo-Monasar, "Urban Residential Water Demand Prediction Based on Artificial Neural Networks and Time Series Models," in Journal of Water Resources Management, DOI 10.1007/s11269-015-1021-z, 2015.

[17]. J. Adamowski, and C. Karapataki, "Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms," Journal of Hydrological Engineering, vol. 15, pp. 729–743, 2010.

[18]. M. H. Al Shamisi, A. H. Assi, and H. A. N. Hejase, “Using Matlab to develop Artificial Neural Networks Models for predicting Global Salar Radiation in Al-Ain City, UAE,” Journal of Engineering Education and Research using MATLAB, Dr. Ali Assi (Ed.); ISBN: 978-953-307-656-0, 2011.
[19]. A. Ajbar, and E. M. Ali, “Prediction of Municipal water production in touristic Mecca City in Saudi Arabia using neural networks” Journal of King Saud University-Engineering Sciences, 2013.
[20]. C-M. Huang, C-J. Huang and M-L. Wang, “A Particle Swarm Optimization to Identifying the ARMAX Model for Short-Term Load Forecasting,” IEEE Transactions on Power Systems, vol. 20, no. 2, pp. 1126–1133, 2005.
[21]. B. Wang, N-L. Tai, H-Q. Zhai, J. Ye, J-D. Zhu, and L-B. Qi, “A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting,” Electric Power Systems Research, vol. 78, pp. 1679–1685, 2008.
[22]. M. El-Telbany, and F. El-Karmi, “Short-term forecasting of Jordanian electricity demand using particle swarm optimization for,” Electric Power Systems Research, vol. 78, pp. 425–433, 2008.
[23]. S. Saravanan, R. Nithya, S. Kannan, and C. Thangara, “Forecasting India's Electricity Consumption Using Particle Swarm Optimization,” Power Electronics and Renewable Energy Systems, Lecture Notes in Electrical Engineering, Springer India, vol. 326, pp. 843–85l, 2015.
[24]. M. Nasseri, A. Moeini, and M. Tabesh, “Forecasting monthly urban water demand using Extended Kalman Filter and genetic Programming,” Journal of Expert Systems with Applications, vol. 38, pp. 7387–7395, 2011.
[25]. M. Fagiani, S. Squartini, L. Gabrielli, S. Spinsante, and F. Piazza, “A review of datasets and load forecasting techniques for smart natural gas and water grids: Analysis and experiments,” Journal of Neurocomputing, vol. 170, pp. 448–465, 2015.
[26]. Z. Beheshti, M. Firouzi, S. M. Shamsuddin, M. Zibarzani, and Z. Yusop, “A new rainfall forecasting model using the CAPSO algorithm and an artificial neural network,” Journal of Neural Computing & Applications, DOI 10.1007/s00521-015-2024-7, 2015.
[27]. M. Buyukyildiz, G. Tezel, and V. Yilmaz, “Estimation of the Change in Lake Water Level by Artificial Intelligence Methods,” Journal of Water Resources Management, vol. 28, pp. 4747–4763, 2014.
[28]. C. Sudheer, N. A. Shrivastava, B. K. Panigrahi, and S. Mathur, “Groundwater Level Forecasting Using SVM-QPSO,” B.K. Panigrahi et al. (Eds.): SEMCCO 2011, Part I, LNCS 7026, Springer-Verlag Berlin Heidelberg, pp. 731–74l, 2011.
[29]. K. Chau, “River Stage Forecasting with Particle Swarm Optimization,” R. Orchard et al. (Eds.): IEA/AIE 2004, LNAI 3029, Springer-Verlag Berlin Heidelberg, pp. 1166–1173, 2004.
[30]. T. H. D. Ngo, W. Bros, “The Box-Jenkins Methodology for Time Series Models,” Journal of Statistics and Data Analysis, SAS Global Forum, 2013.
[31]. The Box-Jenkins Method: Chapter 470. NCSS Statistical Software.
[32]. J. Meher, and R.Jha, “Time-Series analysis of monthly rainfall data for the Mahanadi River Basin, India,” Journal of Sciences in Cold and Arid Regions, vol. 5, pp. 0073–0084, 2013.
[33]. H. Nazari, A. Kazemi, M-H. Hashemi, M. M. Sadat, and M. Nazari, “Evaluating the performance of genetic and particle swarm optimization algorithms to select an appropriate scenario for forecasting energy demand using economic indicators: residential and commercial sectors of Iran,” International Journal of Energy and Environment Engineering, vol. 6, pp. 345–355, 2015.
[34]. B. Sun, J. Xie, and N. Wang, “Application of Urban Water Demand Prediction Model Using Particle Swarm Algorithm Based on Simulated Annealing,” Applied Mechanics and Materials Vols. 155-15 6, pp. 102–106, 2012.
[35]. T. Vantuch, and I. Zelinka, ,“Evolutionary Based ARIMA Models for stock Price Forecasting” © Springer International Publishing Switzerland 2015 A. Sanayei et al. (eds.), ISCS 2014: Interdisciplinary Symposium on Complex Systems, emergence, Complexity and Computation 14, DOI 10.1007/978-3-319-10759-2_25.

[36]. C. H. Aladag, U. Yolcu, and E. Egrioglu, "A New Multiplicative Seasonal Neural Network Model Based on Particle Swarm Optimization," Neural Process Lett (2013) vol. 37, pp. 251–262 , DOI 10.1007/s11063-012-9244-y, 2013.
[37]. I. Almutaz, E. Ali, Y. Khalid, and A. H. Ajbar, "A long-term forecast of water demand for a desalinated dependent city: case of Riyadh in Saudi Arabia" Journal of Desalination and Water Treatment, vol. 51, pp. 5934–5941, 2013.
[38]. R. McNown, O. S. Aburizaizah, C. Howe, and N. Adkins, "Forecasting annual water demands dominated by seasonal variations: the case of water demands in Mecca" Journal of Applied Economics , vol. 47, no. 6, pp. 544–552, 2015.
[39]. O. K. M. Ouda, "Water demand versus supply in Saudi Arabia: current and future challenges," Journal of Water Resources Development , vol. 30, no. 2, pp. 335–344, 2014.
[40]. M. K. Tiwari, and J. F. Adamowski, "Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet-Bootstrap Machine-Learning Approach," Journal of water resources planning and management, vol. 141, 2015.
[41]. M. Yalcintas, M. Bulu, M. Kucukvar, and H. Samadi, "A Framework for Sustainable Urban Water Management through Demand and Supply Forecasting: The case of Istanbul," Sustainability (2015) vol. 7, pp. 11050–11067 , DOI 10.3390/su70811050, 2015.
[42]. M. S. Voss, and X. Feng, "ARMA Model Selection Using Particle Swarm Optimization and AIC Criteria," 15th Triennial World Congress, Barcelona, Spain, Copyright © 2002 IFAC.
[43]. K. Z. M. Azmi, Z. Ibrahim, and D. Pebrianti, "Simultaneous Computation of Model Order and Parameter Estimation for ARX Model Based on Multi-Swarm Optimization" ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 22, pp. 17191–17196, 2015.
[44]. C. Bennett, R. A. Stewart, and C. D. Beal, "ANN-based residential water end-use demand forecasting model," Journal of Expert Systems with Applications, vol. 40, pp. 1014–1023, 2013.
[45]. X-P. Wang, and M-Xi. Shi, "Matlab for Forecasting of Electric Power Load Based on BP Neural Network," R. Chen. (Ed.): ICICIS 2011, Part I, CCIS 134, Springer-Verlag Berlin Heidelberg, pp. 636–642, 2011.
[46]. R. B. Billings, and D. E. Agthe, "State-Space versus Multiple regression for Forecasting Urban Water Demand," Journal of water resources planning and management, vol. 124, pp. 113–117, 1998.
[47]. O. K. M. Ouda, A. Shawesh, T. Al-Olabi, F. Younes, R. Al-Waked, "Review of domestic water conservation practices in Saudi Arabia," Journal of Applications Water Sciences, vol. 3, pp. 689–699, 2013.
[48]. C. Chowdhury, and M. Al-Zahrani, "Characterizing water resources and trends of sector wise water Consumptions in Saudi Arabia," Journal of King Saud University-Engineering Sciences, 2013.
[49]. M. Al-Zahrani, "Modeling and Simulation of Water Distribution System: A Case Study," in Arab Journal of Sciences of Engineering, vol. 39, pp. 1621–1636, 2014.
[50]. O. K. M. Ouda, R. Al-Waked, and A. A. Alshehri, "Privatization of water-supply services in Saudi Arabia: A unique experience," Journal of Utilities Policy, vol. 31, pp. 107–113, 2014.
[51]. A. Abdul Salam, I. Elsegaey, R. Khraif, and A. Al-Mutairi, "Population distribution and household conditions in Saudi Arabia: Reflections from 2000 Census," SpringerPlus 3:530, 2014.
[52]. R. G. Perea, E. C. Poyato, P. Montesinos, and J. A. Rodriguez Diaz, "Irrigation Demand Forecasting Using Artificial Neuro-Genetic Networks," Journal of Water Resources Management, vol. 29, pp. 5551–5567, 2015.
[53]. W. Froelich, "Dealing with Seasonality While Forecasting Urban Water Demand," © Springer International Publishing Switzerland 2015 R. Neves-Silva et al. (eds.), Intelligent Decision Technologies, Smart Innovation, Systems and Technologies 39, DOI 10.1007/978-3-319-19857-6_16, 2015.
[54]. A. A. Adebiyi, A. O. Adewumi, and C.K. Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction," Journal of Applied Mathematics Hindawi Publishing Corporation, Article ID 614342 , 7 pages, volume 2014.

[55]. G. Peter Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing

[56]. S. Boubaker, M. Djemai, N. Manamanni, and F. Msahli, "Active Modes and Switching Instants Identification for switched Linear Systems Based on Discrete Particle Swarm Optimization," Journal of Applied Soft Computing, Elsevier vol. 14, pp. 482–488, 2014.

[57]. General Directorate of Water in Hail Region, Report of the projects between 1425-1433 H.

[58]. https://en.wikipedia.org/wiki/Ha%27il

[59]. Mario Nosvelli & Antonio Musolesi, Water Consumption and Long-run Socio-economic Development: an Intervention and a Principal Component Analysis for the City of Milan, Environ Model Assess (2009) 14:303–314.

[60]. Iek In Ieong & Inchio Lou &Wai Kin Ung & KaiMengMok, Using Principle Component Regression, Artificial Neural Network, and Hybrid Models for Predicting Phytoplankton Abundance in Macau Storage Reservoir, Environ Model Assess (2015) 20:355–365.