Simulating a Stochastic Signal of Urban Water Demand by a Novel Combination of Data Analytic and Machine Learning Techniques

Salah L. Zubaidi, Hussein Al-Bugharbee, Yousif Raad Muhsin, Sadik Kamel Gharghan, Khalid Hashim, Hussein Mohammed Ridha, Rafid Alkhaddar, Patryk Kot and Mawada Abdellatif

1 Department of Civil Engineering, College of Engineering, Wasit University, Iraq
2 Department of Mechanical Engineering, College of Engineering, Wasit University, Iraq
3 Department of Medical Instrumentation Techniques Engineering, Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq
4 Department of Civil Engineering, Liverpool John Moores University, UK
5 Department of Environmental Engineering/ University of Babylon, Iraq
6 Department of Electrical and Electronic Engineering, Universiti Putra, Serdang 43400, Selangor, Malaysia

* Corresponding author: salahlafta@uowasit.edu.iq

Abstract

In this research, a new methodology is presented to forecast the stochastic component of urban water demand for Baghdad City from 2003 to 2014. The methodology contains data pre-processing to analyse raw time series of water via Empirical Mode Decomposition (EMD) technique and select the best scenario of independent variables by a stepwise regression method. Artificial neural network (ANN) is integrated by Backtracking Search Algorithm (BSA) to find the best factors of the ANN model. The outcomes reveal that data pre-processing can detect the stochastic signal of water data and choice the best model input’s scenario. BSA successfully determines the parameters of the ANN model. The methodology accurately simulated the stochastic signal of water time series depend on different statistical criteria such as coefficient of determination and mean absolute relative error equal to 0.99 and 0.0208, respectively.

Keywords

Artificial neural network; backtracking search algorithm; Baghdad City; empirical mode decomposition.

1. Introduction

Water is a vital element that is responsible for continuous human life in the world. Freshwater resources correlated positively with smart cities and sustainable societies [1, 2]; it drives all types of industries. Hence, all cities are located beside the freshwater resource [3-5]. Iraq is one of the countries depending on Tigris and Euphrates rivers as the main potable water resources [6]. The variability of climate change led to depleting multi-resources of freshwater; surface water, groundwater, and stormwater in the centres of cities [7-9]. On the other hand, water demand has
increased due to the growth in population, increases in per capita consumption, and rapid economic development that complicated the problem [10, 11]. In addition, industries contributed to increasing pollution in the freshwater sources [12-16]. As a result, municipal water systems have become under huge stress in recent years. This issue has firmly emphasised on the need for better planning, design and operation water systems to guarantee that the requirements of the present and future inhabitants are met with the wanted level of satisfaction [17].

Accurate water demand prediction is a key aspect for proper planning, operation and development of municipal water system. The forecast could be short- and medium-term; medium-term forecasting is useful for planning, design and making the expansion of a municipal water system [18]. Municipal water demand depends upon climatic factors, socio-economic factors, public water policies and strategy related factors [19]. Iraq suffered from the impact of climate change, which includes an increase in temperature and decreases precipitation. This impact is likely to be worse when it goes to the future [20]. Accordingly, the main resources of freshwater are adversely affected. This increases the uncertainties for managers of water utilities. Baghdad City is a big city that their municipal water system has been affected by climatic factors.

Historically, several traditional methods were used to predict water demand, e.g. time series analysis, regression analysis, or a combination of the two methods. The disadvantages of traditional methods are that the data is supposedly stationary and linear [21]. It is also tended to overestimate the water demand that leads to stress on the infrastructure of the water system [18]. To avoid these issues, the researchers applied artificial intelligence (AI) approach, e.g. artificial neural networks (ANN), which showed good ability to simulate water demand in comparison with conventional methods [22]. Also, it shows a successful application in different fields such as in [23-26]. Araghrinejad [27] reported that combined models are being developed to come upon the new needs of water simulation resulting from the climate factors’ variability. The hybrid models include several techniques. One of these techniques works as the main technique, whereas the others work to integrate and optimise the main technique. Rahim et al., [28] review different techniques of hybrid machine learning with data analytic that were used to predict water demand. They mentioned that additional studies are wanted to evolve new hybrid methods to enhance the results of prediction. Consequently, this study aims to estimate the stochastic component of urban water demand considering previous water consumption by using a novel methodology. This methodology comprises data pre-processing methods, empirical mode decomposition and stepwise regression, and a hybrid model of backtracking search algorithm with an artificial neural network (BSA-ANN). Depend on the literature review, this is the first time this procedure is applied to predict the monthly stochastic component of water consumption. The next sections of this study are as follows. The data set and general framework of the methodology are illustrated in section two. In section three, the results and their discussion are presented. Finally, section four draws the main conclusions of this study.

2. Methodology
The methodology of this study divides into three main sections counting data set, data pre-processing, and hybrid machine learning model with the optimisation algorithm.

2.1. Data Set
This study employs the historical monthly data of urban water consumption (in million cubic meters, M³) for Baghdad City over twelve years (2003-2014) to build and investigate a novel methodology. The data were gained by mayorality of Baghdad, which is responsible to deliver clean water to all types of customers; residential, industrial, institutional and commercial. Eight primary water treatment plants were used to supply water for indoor and outdoor uses inside the borders of Baghdad City depending on the Tigris River as a potable resource.

2.2. Data Pre-processing
It has a valuable impact on the quality of the prediction models. It can be divided into three stages: normalisation, cleaning and choice of best model inputs [29]. Normalisation data helps to reduce the effect of outliers and makes the time series to be normal or near-normal distribution [27]. In the present study, natural logarithm is applied to normalise the time series because of its capability to reduce the impact of multicollinearity between independent variables [30]. Cleaning data means to eliminate the trend (T), oscillatory (O), and noise (Ɛ) signals that represent the non-stationary components and detect the stochastic (S) signal, which refers to the stationary component. The components of raw time series can be represented as in the following formula [31].

\[ y_t = T_t + O_t + S_t + Ɛ_t \]  \hspace{1cm} (1)

In the present study, the empirical mode decomposition (EMD) technique is used for decomposing the raw time series data into a several of oscillatory components named Intrinsic Mode functions [32-35]. These IMFs are obtained using a technique called sifting process. They must satisfy two conditions:

i. At any interval of the component, the mean value of the envelope from minima and maxima should be zero.

ii. In the whole data set, the total extrema and the zero should be equal of different by one.

The sifting process, where IMFs are obtained can be summarized as:

- All the extrema points (e.g. maxima and minima) are identified.
- These extrema are connected to form the upper and lower envelope.
- The mean of this envelope is calculated and subtracted from the original signal.

\[ h_1 = x - m_1 \]  \hspace{1cm} (2)

- If \( h_1 \) meets the two conditions above, it is IMF1 if not, the above steps are repeated until the IMF, which satisfies the two conditions is found.
- The steps above are repeated to find other IMFs till the residual of subtracting be a monotonic function or it has only one extremum.

Finally, the stepwise regression method was considered to choose of the best independent variables based on the significant value that indicates if the variable is statistically significant (p-value < 0.05) or not (p-value ≥ 0.05) to avoid the redundant variables in the model input.

2.3. Hybrid Backtracking Search Algorithm-Artificial Neural Network (BSA-ANN)

This section briefly presents the methods applied in the present study, containing ANN model, BSA as an optimisation technique, and the combined BSA-ANN algorithm.

2.3.1. Backtracking Search Algorithm (BSA)

It is an evolutionary computation technique which depends on the nature system that is frequently used to remedy complex optimisation problems such as non-differentiable, highly nonlinear, constrained design issues and multimodal [17, 36]. BSA has been widely considered for solving various kinds of engineering optimisation problems during the last few years including, but not limited to, optimal control of nonlinear problems [37], wireless sensor [36], and hydrology [17]. Five population sizes, namely 10, 20, 30, 40 and 50 and 100 iterations were utilised to find the population, which could get the least fitness function value. Selecting the ANN parameters based on the trial-and-error is not secure and leads to over-or-under estimate the water demand. Hence, in this research, BSA algorithm is combined with ANN to locate the ANN factors.

2.3.2. Artificial Neural Network (ANN)

ANN technique, as a computational approach, is a parallel knowledge processing system that tries to emulate the workings of the brain’s neurons by using a set of layered neurons [37, 38]. ANN has the capability for adequately mapping the nonlinear relationships between dependent and independents variables such as municipal water consumption time series for short-, medium- and long-term [31].
Up to now, various ANNs kinds have been presented with different functions and conditions. Feed Forward Back Propagation (FF-BP) is widely applied for simulating and predicting complicated issues in science and engineering [37]. Accordingly, this study uses this type with four layers including input, two hidden layers, and the target layer. Additionally, the Levenberg–Marquardt algorithm is employed for training because it has a fast and high computational volume. Besides, the data is split into three sets including training (70%), testing (15%), and validation (15%) [39-41]. For additional information, readers are referred to Gharghan et al., [36].

3. Results and discussion

3.1. Model input scenario

Monthly time series data were prepared through normalising by natural logarithm and detection and treatment of outliers by box and whisker approach. Afterwards, the EMD technique was used for analysing the raw time series of water consumption into five components as shown in Figure 1. The figure reveals that the signals from the bottom represent the trend, seasonal, stochastic, noise for the rest of the signals and the raw time series of water. The stationary of the stochastic signal was examined by using Kwiatkowski–Phillips–Schmidt–Shin and Augmented Dickey-Fuller and tests.

![Figure 1. Raw data and the first five signals obtained by EMD.](image)

Five previous water consumption data (Lags) were determined to choose the best scenario of independent variables as reported in Firat et al.,[42] by using the stepwise regression method. The results reveal that $\text{Lag}_{t-1}$ and $\text{Lag}_{t-2}$ represent the best scenario to forecast the stochastic component of urban water demand with a correlation coefficient equal to 0.96 and 0.85, respectively.

3.2. Integrated ANN by BSA Algorithm

After splitting the data into three sets (training, testing and validation), BSA algorithm was used to determine the optimal number of neurons for both hidden layers (N1 and N2) and learning rate coefficient (Lr). To be with Gharghan et al., [36], this study employs five population sizes (i.e., 10, 20, 30, 40, and 50) to get the lower fitness function (RMSE). The outcomes present that Popsize 50 displays the minimum fitness function (RMSE=0.04101) after 60 iterations as shown in Figure 2. Consequently, the factors of ANN gained from 50 Popsize, $\text{Lr}=0.4195$, N1= 2 and N2=5.
3.3. Performance Assessment of the ANN Model

After integrating the ANN model by selecting optimum (Lr, N1 and N2), the ANN technique was run a couple of times to get the optimum architecture that can simulate accurately stochastic signal of water demand. Various statistical criteria were applied to assess the capability of a prediction model. The correlation between the observed and predicted stochastic signal of water data was tested via utilising $R^2$ as shown in Figure 3. It reveals that $R^2=0.99$ and the data are scattered match to slop line 1:1.

In addition, a graphical test was employed for assessing the performance of the forecast approach as presented in Figure 4. The figure reveals that the ANN model simulates accurately the observed pattern of stochastic signal of water that depends on the plot scale.
The performance of the suggested methodology was assessed via adopting different types of error evaluation including MAE, MARE, and CE as shown in Table 1. The table reveals that the procedure shows a good error scale and a good coefficient of efficiency according to Dawson et al.,[43].

|     | MAE   | MARE  | CE    |
|-----|-------|-------|-------|
|     | 0.041 | 0.0208| 0.999 |

Hence, this study offers a good scientific insight to support the mayoralty of Baghdad City for proper management for existing urban water system under the variability of socio-economic and weather factors.

4. Conclusion

The present study investigates the performance of a novel methodology to forecast the stochastic component of urban water demand. Monthly data of historical water consumption for Baghdad City over twelve years (2003-2014) were employed to build and examine the proposed methodology. The methodology includes data pre-processing (EMD and stepwise regression) techniques and hybrid BSA-ANN model. The results show that EMD is an effective technique for decomposing the raw water time series and locating the stochastic signal. The stepwise regression is a good method to select the best model input scenario. The BSA algorithm can locate the factors of the ANN model increasing the performance of the prediction model and decreasing the error scale. Generally, the methodology can precisely predict the stochastic signal of municipal water demand depend on a variety of statistical standards such as the coefficient of efficiency and mean absolute error equal to 0.999 and 0.041, respectively. Additional studies need to be implemented concerning municipal water demand forecast for other cities that have similar conditions.

References

[1] Farhan S L, Hashim I A J and Naji A A 2019 The Sustainable House: Comparative Analysis of Houses in Al Kut Neighborhoods-Iraq. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), ( Kazan, Russia: IEEE) pp 1031-6

[2] Farhan S, Akef V and Nasar Z 2020 The transformation of the inherited historical urban and architectural characteristics of Al-Najaf's Old City and possible preservation insights Frontiers of Architectural Research 1-17

[3] Farhan S L, Abdelmonem M G and Nasar Z A 2018 The Urban Transformation of Traditional City Centres: Holy Karbala as a Case Study International Journal of Architectural Research: ArchNet-IJAR 12 53-67
[4] Ethaib S and Zubaidi S L 2020 Removal of Methylene Blue Dye from Aqueous Solution Using Kaolin. In: IOP Conference Series: Materials Science and Engineering, (Nasiriyah, Iraq: IOP) pp 1-7

[5] Hashim K S, Ewadh H M, Muhsin A A, Zubaidi S L, Kot P, Muradov M, Aljefery M and Al-Khaddar R 2020 Phosphate removal from water using bottom ash: Adsorption performance, coexisting anions and modelling studies Water Science and Technology 3 1-17

[6] Zubaidi S L, Al-Bugharbee H, Muhsin Y R, Hashim K and Alkhaddar R 2020 Forecasting of monthly stochastic signal of urban water demand: Baghdad as a case study. In: IOP Conference Series: Materials Science and Engineering, (Najaf, Iraq: IOP Conference Series: Materials Science and Engineering) pp 1-7

[7] Ethaib S, Omar R, Mazlina M K S, Radiah A B D and Syafiie S 2016 Development of a hybrid PSO–ANN model for estimating glucose and xylose yields for microwave-assisted pretreatment and the enzymatic hydrolysis of lignocellulosic biomass Neural Computing and Applications

[8] Zubaidi S L, Dooley J, Alkhaddar R M, Abdellatif M, Al-Bugharbee H and Ortega-Martorell S 2018 A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks Journal of Hydrology 561 136-45

[9] Zubaidi S L, Kot P, Hashim K, Alkhaddar R, Abdellatif M and Muhsin Y R 2019 Using LARS–WG model for prediction of temperature in Columbia City, USA. In: IOP Conference Series: Materials Science and Engineering, (Najaf, Iraq: IOP) pp 1-9

[10] Farhan S L, Jasim I A, Al-Mamoori S K J o U R and Renewal 2019 The transformation of the city of Najaf, Iraq: Analysis, reality and future prospects 13 160-71

[11] Hashim K S, Kot P, Zubaidi S L, Alwash R, Al Khaddar R, Shaw A, Al-Jumeily D and Aljefery M H 2020 Energy Efficient Electrocoagulation Using Baffle-Plates Electrodes for Efficient Escherichia Coli Removal from Wastewater Journal of Water Process Engineering 33 1-7

[12] Al-Marri S, AlQuzweeni S S, Hashim K S, AlKhaddar R, Kot P, AlKizwini R S, Zubaidi S L and Al-Khafaji Z S 2020 Ultrasonic-Electrocoagulation method for nitrate removal from water. In: IOP Conference Series: Materials Science and Engineering, (Najaf, Iraq: IOP Conference Series: Materials Science and Engineering) pp 1-9

[13] Alnaimi H, Idan I J, Al-Janabi A, Hashim K S, Gkantou M, Zubaidi S L, Kot P and Muradov M 2020 Ultrasonic-electrochemical treatment for effluents of concrete plants. In: IOP Conference Series: Materials Science and Engineering, (Najaf, Iraq: IOP Conference Series: Materials Science and Engineering) pp 1-10

[14] Alyafei A, AlKizwini R S, Hashim K S, Yeboah D, Gkantou M, Al Khaddar R, Al-Falufi D and Zubaidi S L 2020 Treatment of effluents of construction industry using a combined filtration-electrocoagulation method. In: IOP Conference Series: Materials Science and Engineering, (Najaf, Iraq: IOP Conference Series: Materials Science and Engineering) pp 1-8

[15] Ethaib S, Omar R, Kamal S M M, Awang Biak D R and Zubaidi S L 2020 Microwave-Assisted Pyrolysis of Biomass Waste: A Mini Review Processes 8

[16] Ethaib S, Omar R, Kamal S M M, Awang Biak D R and Zubaidi S L 2020 Toward Sustainable Processes of Pretreatment Technologies of Lignocellulosic Biomass for Enzymatic Production of Biofuels and Chemicals: A Review BioResources 15 10063-88

[17] Zubaidi S L, Gharghan S K, Dooley J, Alkhaddar R M and Abdellatif M 2018 Short-Term Urban Water Demand Prediction Considering Weather Factors Water Resources Management 32 4527-42

[18] Babel M S and Shinde V R 2011 Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok Water Resources Management 25 1653-76

[19] Firat M, Yurdusev M A and Turan M E 2009 Evaluation of artificial neural network techniques for municipal water consumption modeling Water Resources Management 23 617-32
[20] Osman Y Z, Abdellatif M, Al-Ansari N, Knutsson S and Jawad S 2017 Climate Change and Future Precipitation in An Arid Region of The Middle East: Case Study of Iraq *Journal of Environmental Hydrology* **25** 1-18

[21] Adamowski J, Fung Chan H, Prasher S O, Ozga-Zielinski B and Slusarieva A 2012 Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada *Water Resources Research* **48** 1-14

[22] Zubaidi S L, Kot P, Alkhaddar R M, Abdellatif M and Al-Bugharbee H 2018 Short-Term Water Demand Prediction in Residential Complexes: Case Study in Columbia City, USA. In: *11th International Conference on Developments in eSystems Engineering (DeSE)*, (Cambridge, United Kingdom: 11th International Conference on Developments in eSystems Engineering (DeSE)). IEEE pp 31-5

[23] Aljaaf A J, Hussain A J, Fergus P, Przybyla A and Barton G J 2016 Evaluation of machine learning methods to predict knee loading from the movement of body segments. In: *2016 International Joint Conference on Neural Networks (IJCNN)*: IEEE pp 5168-73

[24] Keight R, Aljaaf A J, Al-Jumeily D, Hussain A J, Ozge A and Mallucci C 2017 An intelligent systems approach to primary headache diagnosis. In: *International Conference on Intelligent Computing* (Springer) pp 61-72

[25] Aljaaf A J, Mallucci C, Al-Jumeily D, Hussain A, Alloghani M and Mustafina J 2018 *Applications of Big Data Analytics*, pp 121-34

[26] Alloghani M, Al-Jumeily D, Hussain A, Aljaaf A J, Mustafina J and Petrov E 2018 Healthcare services innovations based on the state of the art technology trend industry 4.0. In: *2018 11th International Conference on Developments in eSystems Engineering (DeSE)*: IEEE pp 64-70

[27] Araghinejad S 2014 *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering* vol 67 (USA: Springer)

[28] Rahim M S, Nguyen K A, Stewart R A, Giurco D and Blumenstein M 2020 Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review *Water* **12** 1-27

[29] Zubaidi S L, Hashim K, Ethaib S, Al-Bdairi N S S, Al-Bugharbee H and Gharhan S K 2020 A novel methodology to predict monthly municipal water demand based on weather variables scenario *Journal of King Saud University - Engineering Sciences* **1** 1-7

[30] Behboudian S, Tabesh M, Falahnezhad M and Ghavanini F A 2014 A long-term prediction of domestic water demand using preprocessing in artificial neural network *Journal of Water Supply: Research and Technology—AQUA* **63** 31-42

[31] Zubaidi S L, Ortega-Martorell S, Kot P, Alkhaddar R M, Abdellatif M, Gharhan S K, Ahmed M S and Hashim K 2020 A Method for Predicting Long-Term Municipal Water Demands Under Climate Change *Water Resources Management* **34** 1265-79

[32] Karatoprak E and Seker S 2019 An Improved Empirical Mode Decomposition Method Using Variable Window Median Filter for Early Fault Detection in Electric Motors *Mathematical Problems in Engineering* **2019**

[33] Zeiler A, Faltermeier R, Keck I R, Tomé A M, Punetone C G and Lang E W 2010 Empirical mode decomposition—an introduction. In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*: IEEE pp 1-8

[34] Nunes J-C and Deléchelle E 2009 Empirical mode decomposition: Applications on signal and image processing *Advances in Adaptive Data Analysis* **1** 125-75

[35] Tabrizi A A, Al-Bugharbee H, Trendafilova I and Garibaldi L A cointegration-based monitoring method for rolling bearings working in time-varying operational conditions *Meccanica* **1** 1-17

[36] Gharhan S K, Nordin R and Ismail M 2016 A Wireless Sensor Network with Soft Computing Localization Techniques for Track Cycling Applications *Sensors (Basel)* **16** 1-23

[37] Heidar A and Hashempour M 2018 Investigation of Mechanical Properties of Self Compacting Polymeric Concrete with Backpropagation Network *International Journal of Engineering* **31** 903-9
[38] Ghasemi A K, Mir M M, Nanvakenari S and Movagharnejad K 2019 Predicting the Coefficients of Antoine Equation Using the Artificial Neural Network (TECHNICAL NOTE) *International Journal of Engineering* **32** 1353-7

[39] Zubaidi S L, Abdulkareem I H, Hashim K, Al-Bugharbee H, Ridha H M, Gharghan S K, Al-Qaim F F, Muradov M, Kot P and Al-Khaddar R 2020 Hybridised Artificial Neural Network Model with Slime Mould Algorithm: A Novel Methodology for Prediction of Urban Stochastic Water Demand *Water* **12** 1-18

[40] Zubaidi S L, Al-Bugharbee H, Ortega-Martorell S, Gharghan S K, Olier I, Hashim K S, Al-Bdairi N S S and Kot P 2020 A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach *Water* **12** 1-17

[41] Zubaidi S L, Ortega-Martorell S, Al-Bugharbee H, Olier I, Hashim K S, Gharghan S K, Kot P and Alkhaddar R M 2020 Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study *Water* **12** 1-17

[42] Firat M, Turan M E and Yurdusev M A 2010 Comparative analysis of neural network techniques for predicting water consumption time series *Journal of Hydrology* **384** 46-51

[43] Dawson C W, Abrahart R J and See L M 2007 HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts *Environmental Modelling & Software* **22** 1034-52