Forecasting of monthly stochastic signal of urban water demand: Baghdad as a case study

Salah L. Zubaidi1*, Hussein Al-Bugharbee2, Yousif Raad Muhsin1, Khalid Hashim3,4, and Rafid Alkhaddar3

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 Civil Engineering, Liverpool John Moores University, UK
4 Department of Environmental Engineering, University of Babylon, Iraq
* Corresponding author: salahlafta@uowasit.edu.iq

Abstract. Forecasting of municipal water demand is essential for the decision-making process in the water industry in particular for countries that suffered from water scarcity. An accurate prediction of water demand improves the water distribution systems' performance. This study analyses the water consumption data of Baghdad city using a signal pre-treatment processing approach aiming at a stochastic signal extraction of such data. An autoregressive (AR) model is then applied to predict monthly water consumption. Our prediction model has been trained and tested using a water consumption data captured from Al-Wehda treatment plant between 2006 and 2015. The results reveal that applying signal pre-treatment method was an effective approach for detecting stochastics of our water consumption data, and the hybrid model was reliable for the prediction of water demand.

1. Introduction
Nowadays, many countries suffer from the harmful impact of climate change that led to reducing the freshwater resources, while on the other hand, the demand for freshwater is likely to increase due to the growing population. This increases the pressure that already being placed on water facilities [1, 2]. Further to this, water pollution has dramatically increased due to the fast-growing industrial activities [3-7]. Iraq locates in the eastern part of the Middle East, which is an arid to semi-arid region. This region, in general, suffers from water shortage that is expected to increase in the future because of the impact of several factors such as climate change, a high rate of population growth and pollution. The Tigris and Euphrates Rivers are the primary sources of freshwater in Iraq. These revisers experienced significant water shortages from 2009 to 2014 and this shortage of water is expected to increase because of the climate change and growing water demand upstream (Turkey, Syria and Iran) [8, 9]. In addition, freshwater resources are a fundamental element to the emergence of cities in Iraq, where all the settlements lived around Tigris and Euphrates rivers [10-12].
Accurate forecasting of urban water demand can play a significant role in the planning, designing, operating and managing of urban water supply systems. Also, it enhances the balance between water supply and demand in the future, which leads to supporting the sustainability [13, 14].

The need for reliable and accurate water demand prediction increases and motivates the researchers to continue the development of different prediction methods. In addition, studies show that artificial intelligence techniques outperform traditional methods (regression and time series) such as [15-19]. Araghinejad [20] stated that hybrid models can be more precise and reliable for predicting water demand than non-hybrid models, as proven in the previous researches such as [21-23].

This study aims to use a hybrid model of signal pre-treatment method and an autoregressive approach to detect the stochastic signal of water consumption time-series data, which can be used to predict the monthly water consumption for one year ahead.

2. Area of study and data set

The studied area in this paper was Baghdad City, which is a big city and capital of Iraq (Figure 1, Lat: 33.33333°, Lon: 44.43333°, Alt: 41 m). It locates in the central part of Iraq on the River Tigris, and it covers an area of 204.2 km². It has a population of about 8.5 million in 2016 that means approximately 21% of the total population of Iraq. Iraq faced a considerable climate change that impacts on the freshwater resources and affects the water demand. As a result, considerable stress applied to the urban water infrastructure. This issue encourages researchers to examine these impacts to reduce the uncertainty of decision-makers [8, 9, 24].

In Baghdad City, weather differs from summer that is dry, sweltering and clear to winter that is cold, mostly clear and dry. The hot days are from 27th May to 23rd September with an average daily high temperature reaches some times more than 45 °C in the hottest days that concentrated around 30th July. While the cold
days are from 23rd November to 5th March with an average daily high temperature reaches some times less than 5 °C in the coldest days that concentrated around 11th January. For rainfall, it occurs from 27th October to 21st April with an average total accumulation of at least 13 millimetres [25].

Mayoralty of Baghdad city is responsible for supplying municipal water inside Baghdad City border for all types of customers (residential, commercial, industrial and institutional). The mayoralty has eight main water treatment plants and uses the Tigris River as a potable resource. Al-Wehda water treatment plant was selected in this study. The plant delivers clean water to Al-Karradh and the regions around the University of Technology for indoor and outdoor uses [26]. Monthly data of water consumption for Al-Wehda water treatment plant over the period of 10 years starting from 2006 to 2015 was collected from the mayoralty of Baghdad.

3. Methodology

Data preprocessing has a substantial influence that improves the precision of forecasting outcomes [27, 28]. In this study, a multi-window of the pre-treatment signal processing method has been applied jointly with linear time-invariant coefficient autoregressive (AR) method approach to predict time series data of water demand. The following subsections brief the pre-treatment signal and the AR.

3.1 Autoregressive Model (AR)

The Autoregressive modelling is on the simplest but efficient time series models that represent a current reading in relation to some past readings. The number of previous points is also named model order. The weights used for relating the new and previous point are called model coefficients and they can be time-independent or dependent. It is used in different research fields such as condition monitoring [29, 30]. In the present study, the time-independent dependent model is utilised for simplicity. For a time series x of length n, it can be represented using the AR model of an order of (p).

\[ x(m) = a_o + \sum_{m=1}^{p} a_m \cdot x(m - l) + \epsilon(m) \]  

Where \( x(m) \) is the predicted value of a sub-signal at time \( m \); 
\( p \) is the order of the AR model; 
\( a_m \) \( (m = 0,1,2,\ldots,p) \) are the AR model coefficients which considered the weights of the past readings; 
\( \epsilon(m) \) is an error term which represents the difference between the real and predicted the \( m \) value of a sub-signal.

The term “time independent” means that the AR model coefficients are kept constant over time and did not change over time. This assumption is made for simplicity. For further details can be found in Al-Bugharbee and Trendafilova [30].

3.2 Signal Pre-treatment

As it was mentioned previously that in this research the stochastic portion of the original time series is extracted and subjected for modelling by AR. The extraction of the stochastic portion is conducted by decomposing the raw time series into a number of components using singular spectrum analysis technique [31]. The use of SSA enables transforming the original time series to a number of sub time series (e.g., components). When these components are added together, the original time series is obtained. Each component represents a portion of the original variance of the time series. The first component has the largest portion of variance information while the last component has a minimum. This technique does not need the imposition of any statistical assumptions such as normality or linearity, and can apply for short-, mid- or long-term time series [31]. Further details on the singular spectrum analysis and its application can
be found in [31-33]. The analysis of the time series by SSA and AR model is conducted using the MATLAB software.

4. Results and discussion
First, the box and whisker method is used to detect and treat then the outlier’s data points of water consumption time series over ten years period (N=120 data). After data cleaning, time series of water consumption is ready to be decomposed by signal pre-treatment to detect the stochastic signal. Ten different windows of signal pre-treatment were employed (w=3, 6, 9 …, 30) as we have options equal to ((N/2)-1) as advice in Golyandina et al., [34]. There are several samples divisions of training and forecasting in modelling with AR. For achieving the for generalisation in training and forecasting, fifty percent of the time series segments is utilised for training by the AR model with different windows of the pre-treatment signal, and four criteria of goodness of fit are used to select the best window. These criteria are as follows; correlation coefficient (R), mean absolute error (MAE), mean square error (MSE) and fitness (FIT). Figure 2 shows that w=21 of pre-treatment signal offers the best solution of the AR model to predict stochastic signal of water consumption one year ahead with R of 99.6% and both MAE and MSE approach to zero. It is important to mention that model prediction is quite same for windows when w goes beyond 21. However, the window 21 is selected is there is no need for selecting higher window size.

![Figure 2](image_url)

Figure 2. Illustrates the different fitness scale versus pre-treatment signal window.

As it was shown above, the hybrid SSA-AR model was built and precisely trained. Then the capability of the model is then examined within testing stage and for forecasting 12 months ahead. Data is shifted three months based on seasonal change the forecasting process, and this process was repeated twenty-two times to increase the reliability of our predictive model. Figure 3 shows the four methods of accuracy to check the
model 23 times. It can be seen that the technique reveals good outcomes e.g. R is changing between 0.988 to 0.999. In addition, table 1 shows the mean values of four fitness functions in the testing stage.

![Figure 3](image.png)

**Figure 3.** Shows the different fitness scale in the testing stage.

| R²  | MAE  | MSE     | RMSE  | %Fitness |
|-----|------|---------|-------|----------|
| 0.99| 3020.6| 1525723.3| 1019.5| 99.68    |

The results demonstrate that the pre-treatment signal method is a robust method to detect a stochastic signal of water consumption with w=21. The hybrid model of the pre-treatment signal and autoregressive Model is reliable and effective to forecast the stochastic signal of municipal water consumption based on four techniques of fitness functions.

5. Conclusions
Water demand prediction is a vital component in active freshwater resources planning and management because it can assistance to find suitable alternatives to confirm the balance between municipal water requested and delivered. In this research, the potential of hybrid pre-treatment signal and autoregressive techniques for monthly municipal water demand prediction was investigated. Historical urban water demand over ten years was employed. Based on the findings gained in this research, it can conclude that the pre-treatment signal processing method was able to detect the stochastic signal and the hybrid model provided a good forecast with a mean R of 0.996. Finally, the proposed methodology of this study can be used as an initial ground to base further studies.
6. References

[1] 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 IOP Conference Series: Materials Science and Engineering 584 1-9

[2] Zubaidi S L, Ortega-Martorell S, Kot P, Alkhaddar R M, Abdellatif M, Gharghan 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

[3] Alattabi A W, Harris C, Alkhaddar R, Alzeyadi A and Hashim K 2017 Treatment of Residential Complexes’ Wastewater using Environmentally Friendly Technology Procedia Engineering 196 792-9

[4] Hashim K S, Shaw A, Al Khaddar R, Pedrola M O and Phipps D 2017 Iron Removal, Energy Consumption and Operating Cost of Electrocoagulation of Drinking Water Using A New Flow Column Reactor J Environ Manage 189 98-108

[5] Hashim K S, Al-Saati N H, Alquzweeni S S, Zubaidi S L, Kot P, Kraidy L, Hussein A H, Alkhaddar R, Shaw A and Alwash R 2019 Decolourization of Dye Solutions by Electrocoagulation: An Investigation of The Effect of Operational Parameters IOP Conference Series: Materials Science and Engineering 584 1-8

[6] Hashim K S, Hussein A H, Zubaidi S L, Kot P, Kraidy L, Alkhaddar R, Shaw A and Alwash R 2019 Effect of Initial PH Value on The Removal of Reactive Black Dye from Water by Electrocoagulation (EC) Method Journal of Physics: Conference Series 1294 1-6

[7] Hashim K S, Kot P, Zubaidi S L, Alkhaddar 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

[8] Aljanabi A A, Mays L W and Fox P 2018 A Reclaimed Wastewater Allocation Optimization Model for Agricultural Irrigation Environment and Natural Resources Research 8 55-68

[9] Aljanabi A A, Mays L W and Fox P 2018 Application of an Optimization Model for Assessing the Performance of Water Appropriation in Iraq Environment and Natural Resources Research 8 105-18

[10] 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

[11] 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): IEEE) pp 1031-6

[12] Farhan S L, Jasim I A and Al-Mamoori S K 2019 The Transformation of The City of Najaf, Iraq: Analysis, Reality and Future Prospects Journal of Urban Regeneration and Renewal 13 1-12

[13] 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

[14] 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

[15] Bougadis J, Adamowski K and Diduch R 2005 Short-term municipal water demand forecasting Hydrological Processes 19 137-48

[16] 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

[17] Aljaaf A J, Al-Jumeily D, Hussain A J, Dawson T, Fergus P and Al-Jumaily M 2015 Predicting the Likelihood of Heart Failure with a Multi Level Risk Assessment Using Decision Tree. In: 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECEE): IEEE) pp 101-6
[18] Aljaaf A J, Al-Jumeily D, Hussain A J, Fergus P, Al-Jumaily M and Abdel-Aziz K 2015 Toward an Optimal Use of Artificial Intelligence Techniques within a Clinical Decision Support System. In: Science and Information Conference, (London, UK pp 548-54
[19] Aljaaf A J, Al-Jumeily D, Hussain A J, Fergus P, Al-Jumaily M and Radi N 2015 Advanced Intelligent Computing Theories and Applications, pp 101-8
[20] Araghinejad S 2014 Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering (USA: Springer)
[21] 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: 2018 11th International Conference on Developments in eSystems Engineering (DeSE), (Cambridge, United Kingdom: IEEE) pp 31-5
[22] 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
[23] 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 1-18
[24] Osman Y, Al-Ansari N, Abdellatif M, Aljawad S B and Knutsson S 2014 Expected Future Precipitation in Central Iraq Using LARS-WG Stochastic Weather Generator Engineering 06 948-59
[25] Diebel J, Norda N and Kretchmer O 2013 Average Weather in Baghdad, Iraq. In: Weather Spark, (United States Cedar Lake Ventures, Inc.)
[26] Zubaidi S L, Al-Bugharbee H, Muhsen Y R, Hashim K, Alkhaddar R M and Hmeesh W H 2019 The Prediction of Municipal Water Demand in Iraq: A Case Study of Baghdad Governorate. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp 274-7
[27] Aljaaf A J, Mallucci C, Al-Jumeily D, Hussain A, Alloghani M and Mustafina J 2018 Applications of Big Data Analytics, pp 121-34
[28] Alloghani M, Aljaaf A J, Al-Jumeily D, Hussain A, Mallucci C and Mustafina J 2018 Data Science to Improve Patient Management System. In: 2018 11th International Conference on Developments in eSystems Engineering (DeSE), pp 27-30
[29] Al-Bugharbee H and Trendafilova I 2015 Autoregressive Modelling for Rolling Element Bearing Fault Diagnosis Journal of Physics: Conference Series 628 1-8
[30] Al-Bugharbee H and Trendafilova I 2016 A Fault Diagnosis Methodology for Rolling Element Bearings Based on Advanced Signal Pretreatment And Autoregressive Modelling Journal of Sound and Vibration 369 246-65
[31] Bugharbee H A and Trendafilova I 2018 A New Methodology for Fault Detection in Rolling Element Bearings Using Singular Spectrum Analysis MATEC Web of Conferences 148 1-5
[32] Al-Bugharbee H R 2016 Data-driven methodologies for bearing vibration analysis and vibration based fault diagnosis. University of Strathclyde)
[33] Al-Bugharbee H, Abolfathi A and Trendafilova I 2018 Vibration-Based Damage Detection of Structural Joints in Presence of Uncertainty MATEC Web of Conferences 148 1-6
[34] Golyandina N and Zhigljavsky A 2013 Singular Spectrum Analysis for Time Series (USA: Springer)