Hybridisation of artificial neural network with particle swarm optimisation for water level prediction

Sarah J. Mohammed[1], and Salah L. Zubaidi [1]

Affiliations

[1] Department of Civil Engineering, Wasit University, Wasit 52001, Iraq.

Correspondence
Sarah J. Mohammed
Department of Civil Engineering, Wasit University, Wasit, Iraq.
Email:
saramustaf53@gmail.com

Abstract
Accurate water level (WL) prediction is essential for the efficient management of various water resource projects. The creation of a reliable model for WL forecasting is still a difficult task in water resource management. This study applies an artificial neural network (ANN) integrated with the particle swarm optimisation algorithm (PSO-ANN) for simulating monthly WL of the Tigris River in Alkut City, Iraq. Data pre-treatment methods are utilised for improving raw data quality and detect the optimal predictors. Monthly WL and climatic variables from 2011 to 2020, were used to construct and validate the proposed technique. The results showed that singular spectrum analysis (SSA) is a high-performance technique for denoising time series. The PSO-ANN model produces good results coefficient of determination (R^2) of 0.85.

الخلاصة: يعد التنبؤ الدقيق بمستوى المياه (WL) أمرًا بالغ الأهمية للإدارة الفعالة لمشاريع الموارد المائية المختلفة. لا يزال إنشاء نموذج موثوق للتنبؤ بـ WL مهمة صعبة في إدارة موارد المياه. طبقت هذه الدراسة الشبكة العصبية الاصطناعية (ANN) محسنه بخوارزمية تحسين سرب الجسيمات (PSO-ANN) لمحاكاة WL الشهرية لنهر دجلة في مدينة الكوت ، العراق. يتم استخدام طرق المعالجة المسبقة للبيانات لتحسين جودة البيانات وتحديد نموذج الإدخال المثالي. تم استخدام البيانات الشهريه لمنسوب المياه والمتغيرات المناخية من 2011 إلى 2020 لبناء التقنية المقترحة والتحقق من صحتها. أظهرت النتائج أن تحليل الطيف المفرد هو تقنية عالية الأداء لتقليل الضوضاء في السلاسل الزمنية. ينتج عن نموذج PSO-ANN نتائج جيدة معامل تحديد (R2) يبلغ 0.85 .

1. INTRODUCTION

A greater need for the development and use of water resources has been raised as a result of the intensification of global climate change, the regular occurrence of droughts, and the detrimental effects they have on social and economic growth [1, 2]. Water level (WL) prediction is a crucial duty of hydrologists, engineers, and relevant authorities in establishing a sustainable conceptual design of water infrastructures, drought management strategies, providing food, and also evaluating river/lake/reservoir behavior for operational purposes [3, 4].

Iraq is one developing country that is vulnerable to climate change, with significant negative consequences for water supplies, the economy, and the economy, particularly in the agricultural field. In general, the Middle East has been severely impacted by climate change, with drought and high record temperatures expected to increase and have a significant impact on populations [5]. The temperature in Iraq, which is located in the world's fastest-warming region, the temperature reached 54°C in Basrah City on July-22-2016, which is regarded one of the highest temperatures ever estimated in the Eastern Hemisphere [6]. The loss of water through evaporation can have serious consequences. These will cause a substantial decrease in the amount of precipitation in Iraq and lowering the discharge of the Tigris and Euphrates rivers, posing a serious problem for the country because Iraq's water security is based on the Euphrates and Tigris rivers, both of which are in deterioration [5].

WL change is a complicated hydrological phenomenon that is influenced by a variety of factors, including meteorological conditions. As a result, many tools for forecasting water levels while accounting for influencing factors have been developed [7]. Hydrological models are critical in decision-making systems for anticipating and estimating the effects of river floods, as well as for drought monitoring and water management [8]. In the past, conventional approaches used to predict water level, such as autoregressive integrated moving average (ARIMA), assumed that a given time series is the result of an underlying linear process. As a result, when modelling nonlinear hydrological time series, these techniques may not always perform well [9]. In recently, there has been a great deal of interest in the use of machine learning methods to improve WL prediction. Machine learning (ML) models have been discovered to be extremely effective in modelling nonlinear systems and prediction environmental phenomena [10]. There are many types of ML models, such as support vector machine (SVM) [11], adaptive neuron-fuzzy inference (ANFIS) [12], genetic programming (GP) [13], and artificial neural network (ANN) [14].

ANN have been extensively utilised in a variety of engineering and scientific problems, most notably in the modelling of nonlinear hydrological processes. Through their model structure, ANNs mimic the biological neural networks of the human brain by mapping the complicated nonlinear relationships and processes inherent among multiple influencing variables [3, 15]. The need for rising data-driven method reliability, capability, and accuracy has led to the development of hybrid models, which combine two or more techniques to outperform single models. One technique is usually designated as the primary one in these hybrid approaches, with the others serving as pre- or post-processing methods [16]. Several combined techniques have been utilised to forecast WL, for example Shafaei and Kisi [17], Xie, et al. [18], Ghorbani, et al. [19], Deng, et al. [3], Chen, et al. [20], Azad, et al. [21], Mohammadi, et al. [22], and Nguyen, et al. [23].

Various optimisation techniques can be employed to solve problems in applications. The purpose of optimisation algorithms is to find the best values for a system's parameters under various conditions [24]. The particle swarm optimisation (PSO) technique is used to find hyperparameters for the model. PSO algorithm has been employed in various hydrology fields, for example stream flow [25], water quality [26], water demand [24].

Hajirahimi and Khashei [27] discuss various types of combined models for time series prediction and emphasise the significance of data pre-treatment techniques. The most significant benefit of pre-processing methods is that they provide more suitable inputs for prediction model, allowing them to best model the underlying data and, as a result, produce more reliable results. This is also frequently accomplished by decreasing and eliminating noise, normalising and cleaning raw data, and selecting the optimal independent variables, such as wavelet transform (WT) [28], tolerance method [29], singular spectrum analysis (SSA) [24], variance inflation factor (VIF) [30], and autocorrelation function (ACF) and partial autocorrelation function (PACF) [31].

Furthermore, various research in water level forecasting [12, 19, 32-34] recommend using climatic factors for forecasting WLs to enhance prediction precision, so climate factors will be involved as predictors in this study.

The purpose of this paper is to critically evaluate a novel hybrid methodology that combines data pre-processing techniques, ANN model, PSO algorithm to estimate the water level of the Tigris River in Al-Kut City (upstream of the Al-Kut barrage).

The primary objectives of this study were to:

1. Evaluate twelve weather factors over a ten-year period to estimate how climate change would affect WL.
2. Improve the quality of original time series and identify the optimal predictor scenario.
3. Combining the particle swam algorithm PSO with the ANN model to estimate monthly water level.

This paper has four sections. **Section 2** describes the study area and data set. **Section 3** explains the suggested approach to measure WL. **Section 4** presents the findings and data discussion. **Section 5** presents the conclusions.

2. Study Area and Data Collection

The city of AL-Kut is the center of Wasit Province in Iraq, is situated between two latitudes **(** 32 degree 29 minutes 51.2 seconds) north, and two longitudes (45 degree 48 minutes 56.9 seconds) east, and has an estimate height of approximately 20 metres. The province is 17,153 km^2 in area. Comparatively, the urbanised area of Kut is approximately 40 km^2. It is situated in a strategic site on the Tigris River, and from which two branches (Al-Gharaf and Al-Dujaili) emerge near to the north of the city [35]. These two branches deliver potable water to various cities in Wasit and Thi-Qar provinces for irrigational, commercial, residential, and industrial purposes [36] (Figure 1). The Kut Barrage regulates the flow of water from the Tigris River to the Dujaila and al-Gharraf branches by raising the WL in the barrage upstream. Consequently, forecasting upstream WL is important because it drives city growth and prosperity.

Figure 1. Location of the research area, AL-Kut City, Iraq.

Traditional monthly WL data (the KUT barrage's upstream) was produced by the water resources directorate, Wasit Province. In General, abnormal conditions caused data from Iraqi metrological stations to be lost (e.g., terrorism). Monthly climate variables were gathered from the National Oceanic and Atmospheric Administration (NASA) [37], in accordance with Ahmad, et al. [38] and Capt, et al. [39]. The time series was collected from 1st January 2011 to 31st December 2020. It consists of the water level (WL) in units of metre (m), minimum temperature (Tmin -°c), maximum temperature (Tmax -°c), mean temperature (Tmean -°c), dew forest (DF -°c), relative humidity (RH -percent), rainfall (rain -mm/day), specific humidity (SH -g/kg), surface pressure (P -kpa), maximum wind speed (WSmax - m/s), wind speed (WS- m/s), range wind speed (WSrange -m/s), and minimum wind speed (WSmin -m/s).

3. Methodology

The suggested technique for predicting monthly WL is comprised of four categories (Figure 2): (1) data pre-processing, (2) PSO algorithm, (3) ANN, and (4) model performance measures:
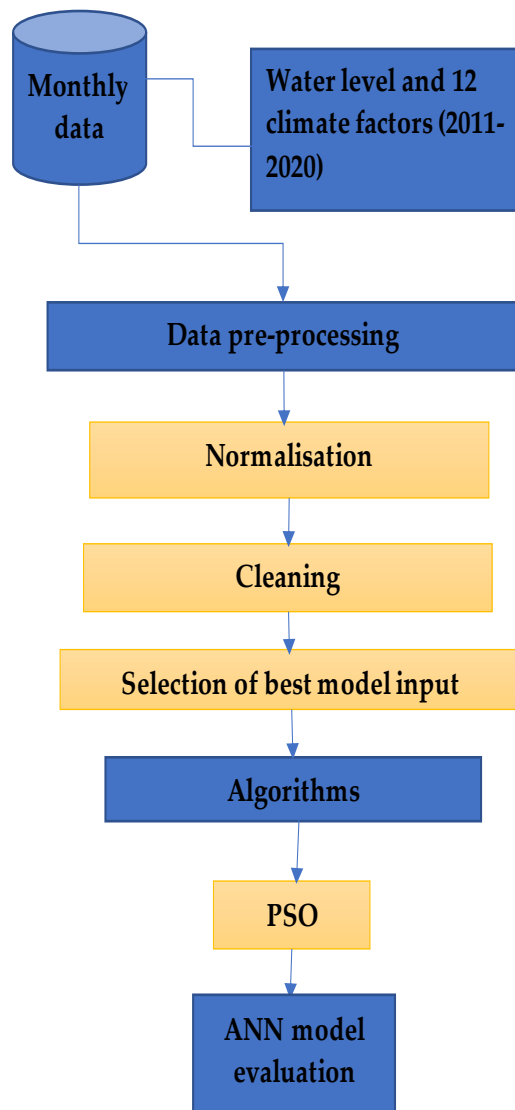
Figure 2. Flowchart showing the procedures needed to model potential WL.

3.1. Data Pre-processing

It is crucial to preprocess data into a suitable format prior to incorporating it into an ANN. These strategies ensure that each input receives equal consideration during the phase of learning [40]. Normalisation, cleaning, and determining the best model inputs are the three steps involved in this research. The natural logarithm was employed to normalise the raw WL data in order to remove the influence of outliers and multicollinearity between independent factors [41]. The cleaning method's goal is to find and remove noise components to boost the regression coefficient and reduce the scale of error [42]. Different types of noise are present in time series. One of the best procedures to denoise raw data by breaking it down into different components is singular spectrum analysis (SSA) [24].

The choice of appropriate predictors is regarded as an essential step in the design of the prediction model's structure. This step helps improve the accuracy of the model by identifying the most pertinent explanatory factors [30]. Pallant [43] suggested employing a tolerance technique to select predictors with a tolerance coefficient value greater than 0.20, as values lower than 0.20 imply the presence of multicollinearity.

3.2. Particle Swarm Optimisation (PSO)

PSO is based on the behavior of swarms of birds and fish. It was created in 1995 by Eberhart and Kennedy. Each person in the population is referred to as a particle, and each particle reflects a potentially feasible solution, with the particle's position regarded as the global optimal solution. In d-dimensional space, the population seeks the global best solution. Meanwhile, each particle determines its own flight direction based on the value of the adaptive function and its velocity, gradually moving to a better region until it finds the global optimal solution [44]. PSO's main advantages are its simple structure, high flexibility, and quick convergence, making it ideal for solving complicated and nonlinear engineering problems [25]. In the iteration of the optimisation technique, the PSO formula can be stated as follows [45]:

$$P_{l+1} = P_l + V_{l+1} \tag{1}$$

P_l and V_l are the particle position and velocity, respectively.
According to the personal best value (P_b) and the best swarm location (P_g) in Eq. (2), the updated velocity can be calculated as follows:
$$V_{l+1} = a.V_l + c_1.r_1 (P_l - P_b) + c_2.r_2 (P_l - Pg) \tag{2}$$

Where the inertia weight is denoted by a, r_1 and r_2 are two randomly generated coefficients for each iteration, and c_1 and c_2 are the acceleration coefficients. The technique is repeated until the target criterion is reached.

3.3. Artificial neural network (ANN)

ANN is a tool for intelligence inspired by the biological nervous systems of animals and humans. ANN can learn a pattern and forecast the output of a model in dimensional space with relative ease. It is capable of managing complex datasets and associating inputs and outputs [26]. In ANN structures, layers are organized as input, hidden, and output layers. Within each layer are interconnected neuronal or nodal structures [46].

In this study, the water level was calculated using a multilayer perceptron (MLP), which is a feed-forward, backpropagation network. For hydrology and water resource forecasting applications, MLP has been used frequently and successfully. The ANN approach was trained using the Levenberg-Marquardt (LM) algorithm, which known for effectively simulating any predictor/response map and for minimising prediction error. The ANN's topology was divided into four layers of neurons. The first layer (i.e. the input layer), which contained predictor factors (i.e., climatic factors). The second and third layers are the hidden layers while, as well as the output layer, which contained response factors (i.e., water level) [42]. Multiple studies have effectively applied ANN with two hidden layers in a variety of settings, and the outcomes displayed that these models' appropriate components the nonlinear relation between dependent and independents [30, 47, 48]. According to Zubaidi, et al. [29] time series were divided into three datasets: training 70% (83 out of 119 data points), testing 15% (18 out of 119 data points), and validation 15% (18 out of 119 data points). The trial-and-error method does not always yield the optimal results. Accordingly, metaheuristic algorithms are used to determine the number of nodes in the first (N_1) and second (N_2) hidden layers, and the learning rate (Lr). The ANN model's performance was enhanced while also saving time by the hybrid techniques [48].

3.4. Model Evaluation

To demonstrate the prediction performance of the PSO-ANN model, four metrics are utilised to assess its prediction accuracy. The metrics for measuring the effectiveness of the model are mean bias error (MBE), mean absolute relative error (MARE), root mean squared error (RMSE), coefficient of determination (R^2).

$$MBE = \frac{1}{N} \sum_{i=1}^{N} (O_i - F_i) \tag{3}$$

$$MARE = \frac{1}{N} \sum_{i=1}^{N} \frac{|O_i - F_i|}{O_i} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - F_i)^2}{N}} \tag{5}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \bar{O})(F_i - \bar{F})}{\sqrt{\sum(O_i - \bar{O})^2 \sum(F_i - \bar{F})^2}} \right]^2 \tag{6}$$

Where Oi and Fi are actual and predicted data, respectively, $\bar{O}$ was the mean of the actual data, $\bar{F}$ was the simulated data's mean value, and N indicates the length of the data series. The ideal model [50 ,49] has values for the MBE, MARE, and RMSE criteria that are almost zero. Good model performance is indicated by a R^2 value greater than 0.85 [51].

In addition, in this research, the PSO-ANN algorithm's capacity to forecast the WL time series during the validation stage was verified using a graphical test. Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and Augmented Dickey–Fuller (ADF) tests were employed to evaluate the residual analysis.

4. Results and Discussion

4.1.Data Pre-processing Analysis

All-data were normalised to decrease the effect of outliers and achieve a normal or near-normal distribution [41]. Then, if any outliers remained after normalisation, they were rescaled. Figure 3 depicts the normalised WL data set and box plot.
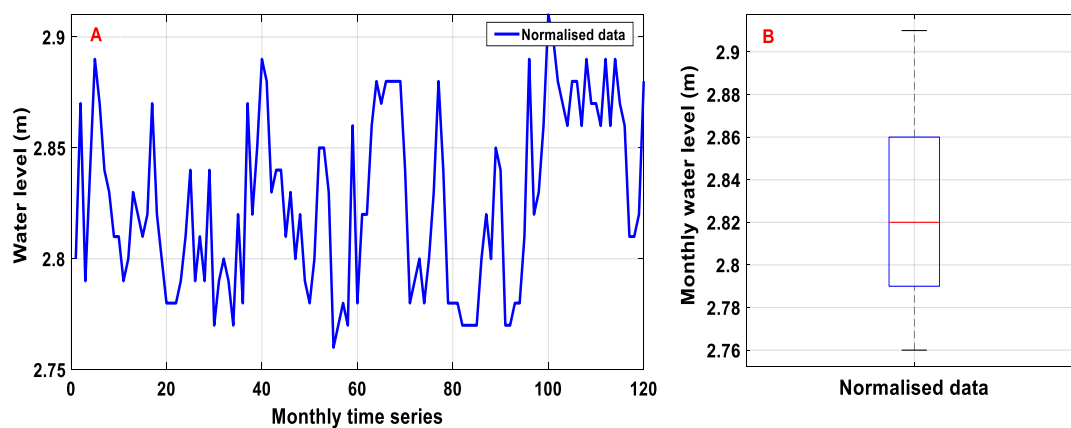

Figure 3: (a) Monthly WL data (b) WL normalised data are used in a box plot.

The time series was then denoise using the SSA technique. The top row represents the normalised and clean data, the second signal is the trend signal, the third signal is the seasonal signal, the fourth signal is the stochastic signal, and the fourth signal is the noise signal. Figure 4 demonstrates the normalised water level time series along with the first four signals.
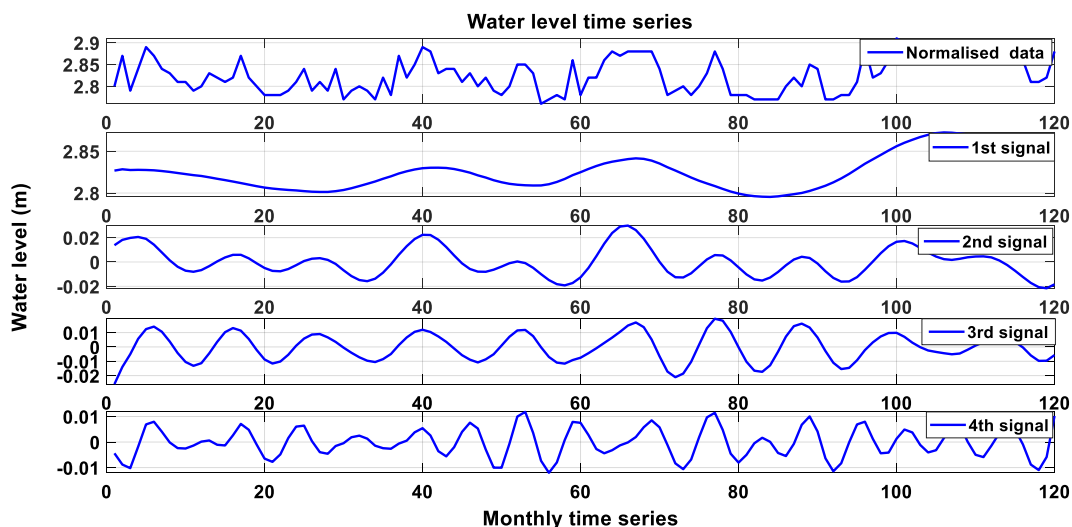
Figure 4. Normalised WL signal and the 1st four signals discovered by SSA.

The final phase of data pre-treatment methods, the best set of predictors (i.e., climate factors) to precisely predict WL data and prevent multi-collinearity were found using a tolerance method. Table 1 shows that WLt-1, DF, and WS are the best scenarios after the various scenarios were run, with tolerance coefficients greater than 0.2.

Table 1. The nominated predictors' collinearity statistics.

Predictor	Tolerance coefficient
WL_{t_1}	0.721
WS	0.420
DF	0.498

Table 2 demonstrates how data pre-treatment methods significantly improved the quality of the data for dependent and independent variables, e.g., the correlation coefficient (R) between WL data and lag of WL time series raises (from 0.643 to 0.993) and R between WL and WS time series increases (from -0.026 to -0.555). The values of R supported the association between WL and climatic factors.

Table 2. Correlation coefficient between water level and climatic variables for denoise data.

Data	WL_{t-1}	WS	DF
Original	0.643	-0.026	0.378
denoised	0.993	-0.555	0.457

4.2. Analysis of the PSO Technique

Three sets of data were created: one for training (70%), one for testing (15%), and one for validation (15%) [42, 52]. To determine the ideal ANN model hyperparameters, the hybrid algorithm PSO was applied using the MATLAB toolbox. The ideal swarm size and the most suitable learning rate and number of neurons in both hidden layers of the ANN model were determined using various swarm sizes (20, 30, 40, and 50). It is important to note that these population sizes do not refer to the sample size previously stated, but rather the size of the swarm. Five times were repeated for each swarm size to reduce uncertainty and broaden the range of predictions (see Figure 5 for PSO-ANN algorithm). The best implementation for each swarm size was selected and combined

with the optimal application for the remaining swarm sizes. For example, the ideal PSO-ANN performance is fourth for a 10-swarm size.
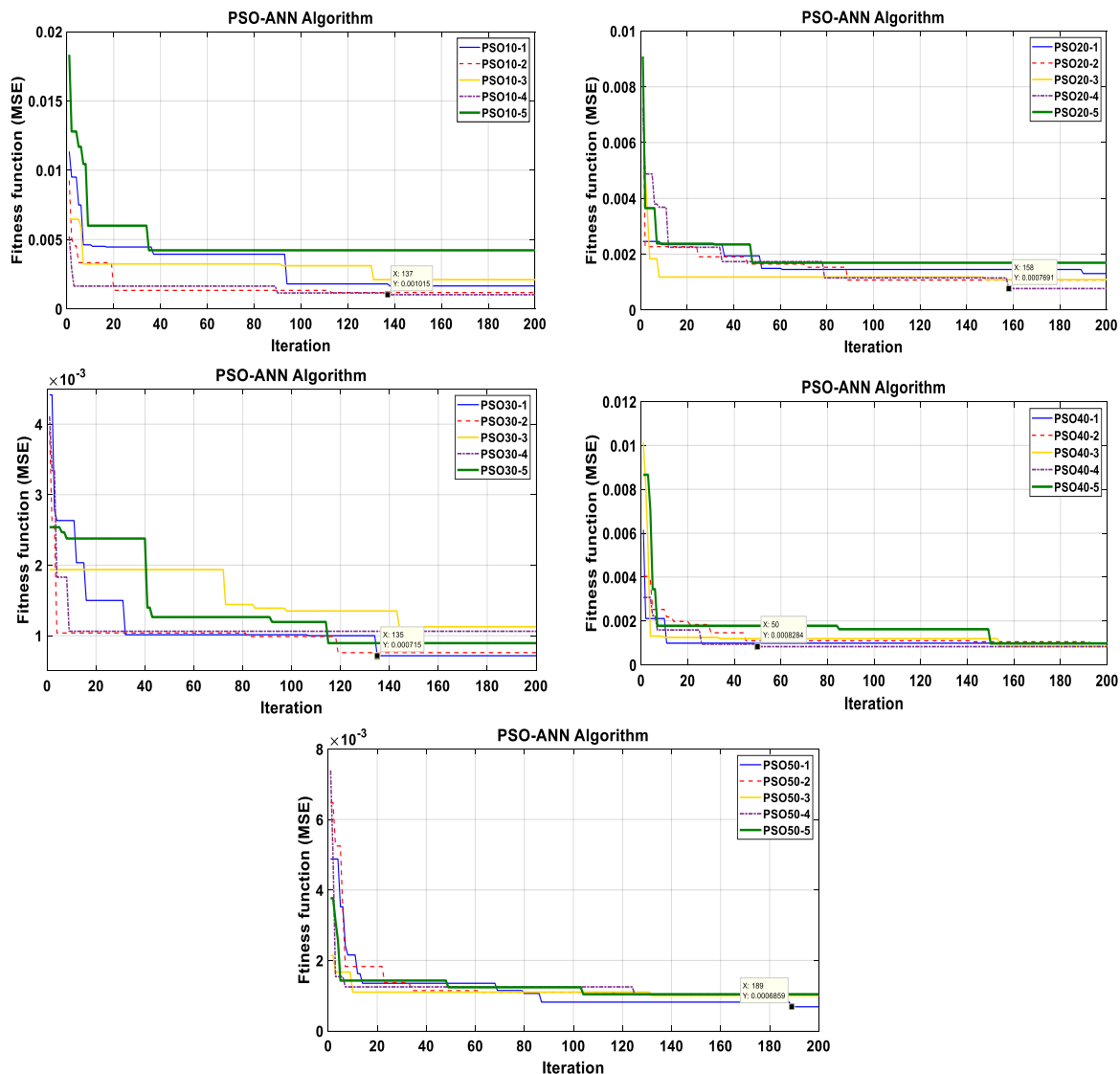

Figure 5. PSO algorithm performance.

Figure 6 demonstrates that, after 189 iterations, the (50_1) swarm size offers the optimal solution with the lowest fitness function (mean squared error (MSE) = 0.0006859). The PSO algorithm's results have been applied to ANN simulations of WL to enhance their capabilities. Accordingly, the learning rate (Lr) was 0.1488 and the number of nodes in the 1st and 2nd hidden layers were 13 and 8, respectively, for the ANN models.
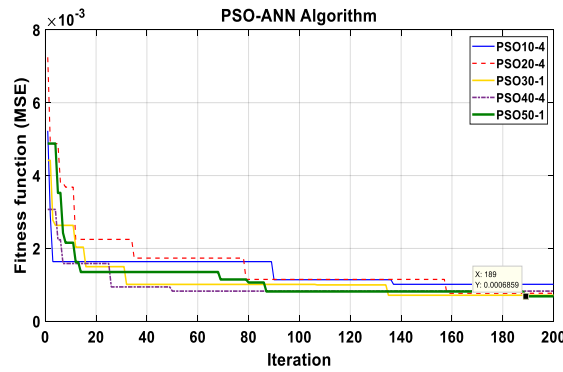
Figure 6. The best swarm for PSO algorithm.

4.3.Performance Evaluation

The ANN model was built to simulate WL after determining the best values for N_1, N_2, and Lr. To determine which network would accurately predict WL, the ANN technique was applied several times. To assess the effectiveness of the models, several different statistical criteria were utilised (validation stage). The R^2, MARE, MBE, and RMSE statistical indicators for WL are shown in Table 3. The PSO-ANN model displayed good prediction accuracy for WL, based on Dawson, et al. [51].

Table 3. PSO effectiveness evaluation criteria for WL in the validation phase.

Target	Model	R^2	MBE	RMSE	MARE
WL	PSO-ANN	0.85	0.0183	0.0184	0.006

Additionally, during the validation stage, a graphical test was utilised to show the hybrid model's ability to simulate the WL time series. Figure 7 depicts the measured WL data in blue, and PSO-ANN forecast WL data in red. It should be noted that the PSO-ANN predicted data follow the trend and periodicity of the actual data, and based on the scale of error, it is very close to the actual data.
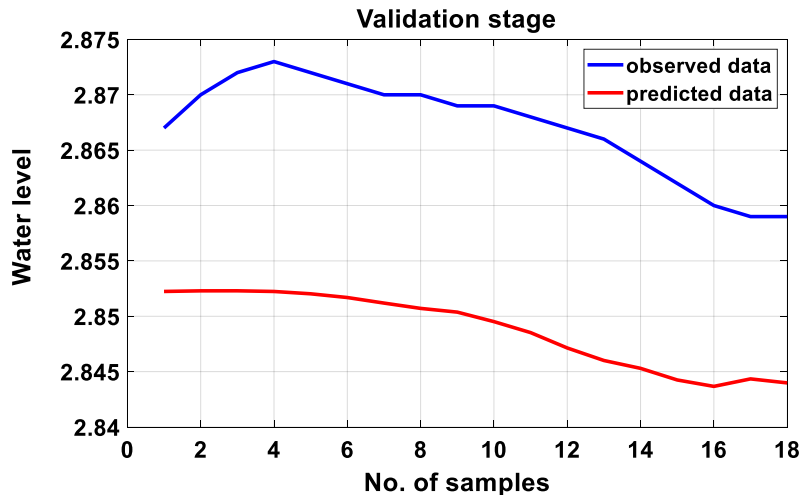

Figure 7. Comparison of actual and estimated WL data for PSO-ANN validation.

Also, the residual data are stationary depended on the KPSS and ADF tests.

Overall, the results of the statistical analysis above demonstrate that:

1. These findings demonstrate the SSA and tolerance methods' potential utility. By removing the structureless noise, the first method improves the raw data quality by raising the correlation between the

target and predictor factors. With the latter approach, the multi-collinearity assumption is maintained while selecting the ideal model input scenario.

2. The proposed methodology effectively predicted the WL time series, according to various statistical analyses (such as MBE, RMSE, R^2, MARE, and graphical tests).

3. The findings show a connection between climate factors and water level.

4. Running each swarm five times improved the precision of choosing the ideal ANN model hyperparameters.

5. These results are limited to the city of Al-Kut only and are not generalised to the rest of the provinces in Iraq.

6. The study's conclusions offer useful scientific data that decision-makers can use to reduce the degree of uncertainty in WL forecast data.

5. Conclusion

Accurate measurement and early warning of river water level is a critical measure to ensure the safety of river basin residents' lives and property, and high precision forecasting of river water level is a necessary prerequisite to meeting this requirement. To estimate the monthly WL of the Tigris River in Al Kut City, this research utilises a hybrid model that combines data pre-processing with the PSO algorithm and the ANN model. The methodology was developed and evaluated using historical time series of WL and climatic variables from 2011 to 2020. The findings indicate that data pre-treatment methods effectively improved data quality, the SSA technique was effective in removing noise from the data, and the tolerance method was efficient in selecting the best model inputs. The suggested methodology performs well in predicting monthly WL with $R^2 = 0.85$, MBE = 0.0183, RMSE = 0.0184, and MARE = 0.006. In the future, socio-economic and additional climatic factors can be employed as predictors to simulate water level, it is also recommended to use various data pre-processing approaches, such as wavelet transformation (WT) to uncover the noise from data and employing updated metaheuristics to simulate water levels in rivers and lakes. In recent years, the utilisation of combined metaheuristic algorithms and ML approaches in WL forecasting has increased significantly. Nevertheless, there is still room for improvement regarding the WL forecast.

REFERENCES

[1] T. Doeffinger and J. W. Hall, "Water Stress and Productivity: An Empirical Analysis of Trends and Drivers," *Water Resources Research,* vol. 56, no. 3, 2020, doi: 10.1029/2019wr025925.

[2] F. F. Li, H. Cao, C. F. Hao, and J. Qiu, "Daily Streamflow Forecasting Based on Flow Pattern Recognition," *Water Resources Management,* vol. 35, pp. 4601–4620, 2021, doi: 10.1007/s11269-021-02971-8.

[3] B. Deng, S. H. Lai, C. Jaing, P. Kumar, A. El-Shafie, and R. J. Chin "Advanced water level prediction for a large-scale river–lake system using hybrid soft computing approach: a case study in Dongting Lake, China," *Earth Science Informatics,* vol. 14, pp. 1987–2001, 2021, doi: 10.1007/s12145-021-00665-8.

[4] L. He, S. Chen, Y. Liang, M. Hou, and J. Chen, "Infilling the missing values of groundwater level using time and space series: case of Nantong City, east coast of China," *Earth Science Informatics,* vol. 13, no. 4, pp. 1445-1459, 2020, doi: 10.1007/s12145-020-00489-y.

[5] Z. A. Al-Salihi, A. H. Kamel, and I. M. Abdulhameed, "Effects of the Climate Change on the Tigris River Basin in Iraq," *International Journal of Design & Nature and Ecodynamics,* vol. 17, no. 4, pp. 585-593, 2022, doi: 10.18280/ijdne.170413.

[6] S. L. Zubaidi *et al.*, "Assessing the Benefits of Nature-Inspired Algorithms for the Parameterization of ANN in the Prediction of Water Demand," *Journal of Water Resources Planning and Management,* vol. 149, no. 1, 2023, doi: 10.1061/(asce)wr.1943-5452.0001602.

[7] Q. Wang and S. Wang, "Machine Learning-Based Water Level Prediction in Lake Erie," *Water,* vol. 12, no. 10, 2020, doi: 10.3390/w12102654.

[8] M. L. Lineros, A. M. Luna, P. M. Ferreira, and A. E. Ruano, "Optimized Design of Neural Networks for a River Water Level Prediction System," *Sensors,* vol. 21, no. 19, p. 6504, 2021, doi: 10.3390/s21196504.

[9] M. Imani, R. J. You, and C. Y. Kuo, "Caspian Sea level prediction using satellite altimetry by artificial neural networks," *International Journal of Environmental Science and Technology,* vol. 11, no. 4, pp. 1035-1042, 2013, doi: 10.1007/s13762-013-0287-z.

[10]   H. Tao, N. K. Al-Bedyry, K. M. Khedher, S. Shahid, and Z. M. Yaseen, "River water level prediction in coastal catchment using hybridized relevance vector machine model with improved grasshopper optimization," *Journal of Hydrology,* vol. 598, p. 126477, 2021, doi: 10.1016/j.jhydrol.2021.126477.

[11]   M. Çimen and O. Kisi, "Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey," *Journal of Hydrology,* vol. 378, no. 3-4, pp. 253-262, 2009, doi: 10.1016/j.jhydrol.2009.09.029.

[12]   P. Páliz Larrea, X. Zapata-Ríos, and L. Campozano Parra, "Application of Neural Network Models and ANFIS for Water Level Forecasting of the Salve Faccha Dam in the Andean Zone in Northern Ecuador," *Water,* vol. 13, no. 15, p. 2011, 2021, doi: 10.3390/w13152011.

[13]   M. Ali Ghorbani, R. Khatibi, A. Aytek, O. Makarynskyy, and J. Shiri, "Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks," *Computers & Geosciences,* vol. 36, no. 5, pp. 620-627, 2010, doi: 10.1016/j.cageo.2009.09.014.

[14]   K. Park, Y. Jung, Y. Seong, and S. Lee, "Development of Deep Learning Models to Improve the Accuracy of Water Levels Time Series Prediction through Multivariate Hydrological Data," *Water,* vol. 14, no. 3, p. 469, 2022, doi: 10.3390/w14030469.

[15]   E. S. K. Tiu, Y. F. Huang, J. L. Ng, N. AlDahoul, A. N. Ahmed, and A. Elshafie, "An evaluation of various data pre-processing techniques with machine learning models for water level prediction," *Natural Hazards,* vol. 110, pp. 121–153, 2021, doi: 10.1007/s11069-021-04939-8.

[16]   S. L. Zubaidi *et al.*, "A Method for Predicting Long-Term Municipal Water Demands Under Climate Change," *Water Resources Management,* vol. 34, no. 3, pp. 1265-1279, 2020, doi: 10.1007/s11269-020-02500-z.

[17]   M. Shafaei and O. Kisi, "Lake Level Forecasting Using Wavelet-SVR, Wavelet-ANFIS and Wavelet-ARMA Conjunction Models," *Water Resources Management,* vol. 30, no. 1, pp. 79-97, 2015, doi: 10.1007/s11269-015-1147-z.

[18]   Z. Xie, Q. Liu, and Y. Cao, "Hybrid Deep Learning Modeling for Water Level Prediction in Yangtze River," *Intelligent Automation & Soft Computing,* vol. 28, no. 1, pp. 153-166, 2021, doi: 10.32604/iasc.2021.016246.

[19]   M. A. Ghorbani, R. C. Deo, V. Karimi, Z. M. Yaseen, and O. Terzi, "Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey," *Stochastic Environmental Research and Risk Assessment,* vol. 32, no. 6, pp. 1683-1697, 2017, doi: 10.1007/s00477-017-1474-0.

[20]   N. Chen, C. Xiong, W. Du, C. Wang, X. Lin, and Z. Chen, "An Improved Genetic Algorithm Coupling a Back-Propagation Neural Network Model (IGA-BPNN) for Water-Level Predictions," *Water,* vol. 11, no. 9, 2019, doi: 10.3390/w11091795.

[21]   A. S. Azad *et al.*, "Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study," *Sustainability,* vol. 14, p. 1843, 2022, doi: 10.3390/su14031843.

[22]   B. Mohammadi, Y. Guan, P. Aghelpour, S. Emamgholizadeh, R. Pillco Zolá, and D. Zhang, "Simulation of Titicaca Lake Water Level Fluctuations Using Hybrid Machine Learning Technique Integrated with Grey Wolf Optimizer Algorithm," *Water,* vol. 12, no. 11, p. 3015, 2020, doi: 10.3390/w12113015.

[23]   D. H. Nguyen, X. Hien Le, J.-Y. Heo, and D.-H. Bae, "Development of an Extreme Gradient Boosting Model Integrated With Evolutionary Algorithms for Hourly Water Level Prediction," *IEEE Access,* vol. 9, pp. 125853-125867, 2021, doi: 10.1109/access.2021.3111287.

[24]   S. L. Zubaidi, J. Dooley, R. M. Alkhaddar, M. Abdellatif, H. Al-Bugharbee, and S. Ortega-Martorell, "A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks," *Journal of Hydrology,* vol. 561, pp. 136-145, 2018, doi: 10.1016/j.jhydrol.2018.03.047.

[25]   S. Samantaray and A. Sahoo, "A Comparative Study on Prediction of Monthly Streamflow Using Hybrid ANFIS-PSO Approaches," *KSCE Journal of Civil Engineering,* vol. 25, pp. 4032–4043, 2021, doi: 10.1007/s12205-021-2223-y.

[26]   M. I. Shah, M. F. Javed, A. Alqahtani, and A. Aldrees, "Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data," *Process Safety and Environmental Protection,* vol. 151, pp. 324-340, 2021, doi: 10.1016/j.psep.2021.05.026.

[27]   Z. Hajirahimi and M. Khashei, "Hybridization of hybrid structures for time series forecasting: a review," *Artificial Intelligence Review,* 2022, doi: 10.1007/s10462-022-10199-0.

[28]   Y. Seo, S. Kim, O. Kisi, and V. P. Singh, "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques," *Journal of Hydrology,* vol. 520, pp. 224-243, 2015, doi: 10.1016/j.jhydrol.2014.11.050.

[29]   S. L. Zubaidi *et al.*, "Hybridised Artificial Neural Network Model with Slime Mould Algorithm: A Novel Methodology for Prediction of Urban Stochastic Water Demand," *Water,* vol. 12, no. 10, 2020, doi: 10.3390/w12102692.

[30]   S. Zubaidi *et al.*, "A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach," *Water,* vol. 12, no. 6, p. 1628, 2020, doi: 10.3390/w12061628.

[31]   C. Zhong, Z. Jiang, X. Chu, T. Guo, and Q. Wen, "Water level forecasting using a hybrid algorithm of artificial neural networks and local Kalman filtering," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment,* vol. 233, no. 1, pp. 174-185, 2017, doi: 10.1177/1475090217727135.

[32]   R. Adnan , F. A. Ruslan , A. m. Samad, and Z. M. Zain "Artificial Neural Network Modelling and Flood Water Level Prediction Using Extended Kalman Filter," presented at the International Conference on Control System, Computing and Engineering, 23 - 25 Nov. 2012, Penang, Malaysia, MALAYSIA, 23-25 November 2012, 2012.

[33]   P. Coulibaly, "Reservoir Computing approach to Great Lakes water level forecasting," *Journal of Hydrology,* vol. 381, no. 1-2, pp. 76-88, 2010, doi: 10.1016/j.jhydrol.2009.11.027.

[34]   M. El-Diasty and S. Al-Harbi, "Development of wavelet network model for accurate water levels prediction with meteorological effects," *Applied Ocean Research,* vol. 53, pp. 228-235, 2015, doi: 10.1016/j.apor.2015.09.008.

[35]   S. F. Balket and N. M. Asmael, "Study the Characteristics of Public Bus Routes in Al Kut City," *Journal of Engineering and Sustainable Development,* vol. 25, no. Special, pp. 3-186-3-194, 2021, doi: 10.31272/jeasd.conf.2.3.18.

[36]   S. Ethaib, S. L. Zubaidi, N. Al-Ansari, and S. L. Fegade, "Evaluation water scarcity based on GIS estimation and climate-change effects: A case study of Thi-Qar Governorate, Iraq," *Cogent Engineering,* vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2075301.

[37]   NOAA. "National Oceanic and Atmospheric Administration. Data tools: Find a station." https://www.ncdc.noaa.gov/cdo-web/datatools/findstation (accessed 1 December 2021, 2021).

[38]   H. Q. Ahmad, S. A. Kamaruddin, S. B. Harun, N. Al-Ansari, S. Shahid, and R. M. Jasim, "Assessment of Spatiotemporal Variability of Meteorological Droughts in Northern Iraq Using Satellite Rainfall Data," *KSCE Journal of Civil Engineering,* vol. 25, no. 11, pp. 4481-4493, 2021, doi: 10.1007/s12205-021-2046-x.

[39]   T. Capt, A. Mirchi, S. Kumar, and W. S. Walker, "Urban Water Demand: Statistical Optimization Approach to Modeling Daily Demand," *Journal of Water Resources Planning and Management,* vol. 147, no. 2, 2021, doi: 10.1061/(asce)wr.1943-5452.0001315.

[40]   H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling & Software,* vol. 15, pp. 101–124, 2000, doi: PII: S1364-8152(99)00007- 9.

[41]   B. Tabachnick, G and L. Fidell, S., *Using Multivariate Statistics*. Boston, MA,USA, 2013.

[42]   S. L. Zubaidi *et al.*, "Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study," *Water,* vol. 12, no. 7, 2020, doi: 10.3390/w12071885.

[43]   J. Pallant, *SPSS survival manual: a step by step guide to data analysis using spss for windows* Open University Press/McGraw-Hill: London, UK, 2016, 2005.

[44]   J. Yan, Z. Xu, Y. Yu, H. Xu, and K. Gao, "Application of a Hybrid Optimized BP Network Model to Estimate Water Quality Parameters of Beihai Lake in Beijing," *Applied Sciences,* vol. 9, no. 9, 2019, doi: 10.3390/app9091863.

[45]   R. M. Adnan, R. R. Mostafa, O. Kisi, Z. M. Yaseen, S. Shahid, and M. Zounemat-Kermani, "Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization," *Knowledge-Based Systems,* vol. 230, p. 107379, 2021, doi: 10.1016/j.knosys.2021.107379.

[46]   G. M. Mulualem and Y.-A. Liou, "Application of Artificial Neural Networks in Forecasting a Standardized Precipitation Evapotranspiration Index for the Upper Blue Nile Basin," *Water,* vol. 12, no. 3, 2020, doi: 10.3390/w12030643.

[47]   F. Farzad and A. H. El-Shafie, "Performance Enhancement of Rainfall Pattern – Water Level Prediction Model Utilizing Self-Organizing-Map Clustering Method," *Water Resources Management,* vol. 31, no. 3, pp. 945-959, 2016, doi: 10.1007/s11269-016-1556-7.

[48]   S. L. Zubaidi, P. Kot, R. M. Alkhaddar, M. Abdellatif, and H. Al-Bugharbee, "Short-Term Water Demand Prediction in Residential Complexes: Case Study in Columbia City, USA," presented at the 2018 11th International Conference on Developments in eSystems Engineering (DeSE), Cambridge, UK, 02-05 September 2018, 2018.

[49]   K. Ahmed, S. Shahid, X. Wang, N. Nawaz, and K. Najeebullah, "Evaluation of Gridded Precipitation Datasets over Arid Regions of Pakistan," *Water,* vol. 11, no. 2, p. 210, 2019, doi: 10.3390/w11020210.

[50]   E. Eze, S. Halse, and T. Ajmal, "Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm," *Water,* vol. 13, no. 13, p. 1782, 2021, doi: 10.3390/w13131782.

[51]   C. W. Dawson, R. J. Abrahart, and L. M. See, "HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts," *Environmental Modelling & Software,* vol. 22, no. 7, pp. 1034-1052, 2007, doi: 10.1016/j.envsoft.2006.06.008.

[52]   S. L. Zubaidi, K. Hashim, S. Ethaib, N. S. S. Al-Bdairi, H. Al-Bugharbee, and S. K. Gharghan, "A novel methodology to predict monthly municipal water demand based on weather variables scenario," *Journal of King Saud University - Engineering Sciences,* vol. 34, no. 3, pp. 163-169, 2022, doi: 10.1016/j.jksues.2020.09.011.