Modelling water quantity parameters using Artificial Intelligence techniques, A case study Abu-Ziriq Marsh in south of Iraq.

Mustafa Al-Mukhtar¹, Fuaad Al-Yaseen¹,² and Jamal Sahib³

¹Civil Engineering Department, University of Technology-Baghdad- Iraq. ORCID: 0000-0002-8850-0899
²Directorate of Thi Qar Municipality – Iraq. ORCID: 0000-0001-6024-4558
³University of Thi Qar

Email: 42148@student.uotechnology.edu.iq

Abstract: The low water quantity has a significant impact on the ecosystem and the food chain of living organisms, thus causing a loss of biodiversity and a lack of natural food sources. Abu-Ziriq Marsh, located in the south of Iraq, is chosen as the case study for the application of the proposed methodology. The aim of this study was to assess the ability of using three different models of Artificial Intelligence (AI) techniques: Adaptive Neural-based Fuzzy Inference System (ANFIS), Artificial Neural Networks (ANN) and Multiple Regression Model (MLR) to predict and estimate the discharge of Abu-Ziriq Marsh by depending on flow release from upstream Al-Badaa regulator. Daily discharge of Al-Badaa regulator (Qₐ) and Abu-Ziriq Marsh (Q₇) were used in this study. The water quantity data, consisting of 720 records of daily data between the years 2017 and 2018, were used for training and testing the models. The training and testing data were randomly partitioned into 515 (70.5 %) and 215 (29.5 %) datasets, respectively. The performance of all models was assessed through the correlation coefficient (R), root mean square error (RMSE) and Nash–Sutcliffe efficiency coefficient (NSE). Results of RMSE, R and NSE for the calibration (validation) of ANFIS model were 4.11 (4.17), 0.87 (0.83) and 0.76 (0.70), respectively. The evaluation of the results indicates that ANFIS model is superior to other models. The identified ANFIS models can be used as tools for the computation of water quantity parameter (Q₇) in Iraqi Marshes.

Keywords: Artificial Intelligence Techniques; Data-Driven Models; Abu-Ziriq Marsh; Al-Badaa Regulator; Water Quantity Parameters.

1. Introduction
The water quantity in rivers, streams and wetlands is considered as essential source for drinking, irrigation and industrial use. For those living in wetlands, these serve as lifelines. The water discharge is one of the most important factors in any aquatic system as well as building strategies for water resources management [1]. Aquatic life is directly affected by the amount of stream flow that interferes with the physical and chemical properties of the river. The flow system is increasingly cited as a major variable that configures habitats and hydro life [2].

The data-driven techniques such as Artificial Intelligence (AI) have been identified as an efficient tool in predicting streamflow as they require much less development time, are useful for real-time applications, and prove capability of an accurate prediction [3]. Recently, AI techniques have been widely used in hydrological forecastings and water resource management e.g. [4], [5].
Kerh and Lee used ANN model to predict discharge at Kaoping River. The results exhibited that ANN model can be used to predict flood discharge. Meanwhile, the physiographical factors had a slight and positive influence on the accuracy of the prediction [6]. Besaw et al. developed two types of ANN models using the records of temperature and precipitation in order to predict discharge in an ungauged basin in the US. Results prove the capability of ANN in prediction river flow in the adjacent basin in a similar way as the basin in which the network trained. In other words, the methods used are widely applicable in other climate zones such as wet and temperate zones [4]. Wang et al. used four types of AI techniques ANN, ANFIS, support vector machine (SVM) and Genetic Programming (GP) to predict monthly discharge. The results showed that the best performance can be obtained through ANFIS, GP and SVM, respectively [7].

Artificial intelligence has been used in many water-related studies e.g. water quality modeling and water management applications [8]–[12]. Therefore, the main objective of the study presented in this paper is to investigate the applicability for some types of AI techniques for modeling daily discharge at Abu-Ziriq Marsh in south of Iraq, which include ANN, ANFIS and MLR models, and to compare their performance. A comprehensive detail about ANFIS, ANN and MLR functioning can be found in [13]–[15], respectively. Three quantitative standard statistical performance evaluation measures, i.e., coefficient of correlation (R), Nash–Sutcliffe efficiency coefficient (NSE), root mean squared error (RMSE), are used to validate all models.

2. Study area and data

2.1 Abu-Ziriq Marsh Description.
Abu-Ziriq Marsh, which covers 120 km², (about 3% of all Iraqi Marshes) lies at the tail end of Al Gharraf River southerly of Al Islah district at a location of latitude 31°09’ 54.9” N, longitude 46°36’33’’ E. The main source of water supply to the marsh is through Shatt Abo-Lihia and the channel of this river runs through the marsh until it dissipates at the tail end into the central marshes. The two main towns around the marsh are Al-Islah in the North and Al-Fuhod in the south of Thi Qar Governorate (Figure 1). Scattered villages of fishermen are located all along the embankments that surround the marsh. Highlighting the vitality role of Abu-Ziriq Marsh in sustaining the daily life of the local residents.

2.2. Data used.
The data for daily discharge time series of Al-Badaa regulator and Abu-Ziriq Marsh were used in this study. The daily discharge data were obtained from The Ministry of the Water Resources (MOWR), State Commission on Operation of Irrigation and Drainage Projects (SCOIPD) in the south of Iraq. Daily discharge data consists of 720 data sets, covering the period 2017 to 2018. Each dataset consists of two parameters, namely: inflow at Abu Ziriq (Q₂) and the outflow from Al-Badaa regulator (Q₀). These parameters are used to develop Data-driven models. The statistical parameters of the data are given in Table 1. In this study, the complete Abu-Ziriq water quantity data (730 samples) were divided into two parts by using cross-validation method; for training and testing data sets. The training and testing data sets comprised of 515 (70.5 %) and 215 (29.5 %) samples, respectively. Cross-validation was implied in this study by using the nearest neighbor method and k = 5.
Figure 1. General location of Abu-Ziriq Marsh.

Table 1. Summary of statistics parameters of water quantity parameters.

| parameters          | N  | Range | Minimum | Maximum | Mean | Std. Deviation | Variance |
|---------------------|----|-------|---------|---------|------|----------------|----------|
| Al-Badaa (Q_B)      | 730| 33.00 | 12.00   | 45.00   | 22.31| 7.06           | 49.87    |
| Abu-Ziriq (Q_Z)     | 730| 36.60 | 0.40    | 37.00   | 8.95 | 8.18           | 66.95    |

2.3. Evaluation criteria.
In this study, three performance criteria were used to assess the goodness of fit of the developed models: RMSE, R and NSE [16].

3. Results and discussion
3.1. Model Structure.
To select input parameters for modeling water quantity entering Abu-Ziriq Marsh, the discharge released from Al-Badaa regulator was used. The relationship between the two variables was initially tested using Pearson correlations coefficient. Based on the above test, it was found that there was a strong correlation between $Q_B$ and $Q_Z$ with $p < 0.01$ (Table 2).

Table 2. Correlation matrix among water quantity parameters (n=730).

| parameters          | Pearson Correlation | Sig. (2-tailed) | N | 730 | 730 |
|---------------------|---------------------|-----------------|---|-----|-----|
| Al-Badaa (Q_B)      | 0.78**              | 0.00            |   | 730 | 730 |
| Abu-Ziriq (Q_Z)     | 1                   | 0.78**          |   | 730 | 730 |

**Correlation is significant at the 0.01 level (2-tailed).
3.2. Models performance.

In this study, the discharge released from Al-Badaa regulator ($Q_B$) and water quantity entering Abu-Ziriq Marsh ($Q_Z$), south of Iraq, were used to develop artificial intelligence techniques. The $Q_Z$ models were created by utilizing ANFIS, ANNs and MLR.

A hybrid algorithm, Sugeno fuzzy Gaussian membership function type (gaussmf) and linear MFs were used in ANFIS models for predicted $Q_Z$. To find the optimum number of membership functions of the ANFIS model, the trial and error method was widely used to finding the final numbers of membership functions, since there is no special rule. The number of membership functions for each input of ANFIS was set to (6). The performance of the ANFIS model for the calibration and validation datasets are given in Table 3. As it can be seen in the figure, there was a satisfactory matching between both data sets. Moreover, values of RMSE, CC, and NSE were 4.11 (m$^3$/sec), 0.87 and 0.76 for the calibration and 4.17 (m$^3$/sec), 0.83 and 0.70, respectively for the validation (Table 3). Figure 2 shows the observed versus predicted $Q_Z$ from ANFIS model during the calibration and validation periods. In their study, Wang et al. reported superior performance of ANFIS in comparison to MLR, ANN and GP in modeling monthly river flow discharges in Lancangjiang River [7].

In ANN model, feed forward-backpropagation algorithm was used. The optimal number of neurons in the hidden layer was selected using the trial and error method, by experimenting with changing the number of neurons in the hidden layer from 1 to 20. The optimal number of neurons was determined in hidden layers that provide optimum structure as 5. Therefore, ANN structure (1, 5, 1) was selected as the optimum ANN model. The performance of the ANN model for the calibration and validation datasets were given in Table 3. Figure 3 shows the observed versus predicted $Q_Z$ from ANN model during the calibration and validation periods. As can be seen from the figure, there was less correspondence between the two datasets. The values of $RMSE$, $R$, and $NSE$ were 4.35 (m$^3$/sec), 0.85 and 0.73, respectively for the calibration and 4.26 (m$^3$/sec), 0.82 and 0.68, respectively for the validation. Ravansalar et al. used ANN, MLR and GP to predict streams flow. The results show the ANN model could not predict the stream flow [17].

The performance of the MLR model and equation for the calibration and validation are illustrated in Table 3 and equation 1. Figure 4 shows the comparative plots of the results obtained from MLR model for $Q_Z$ during the calibration and validation periods. It can be noticed that both data sets were inaccurate. In other words, MLR model’s performance was not satisfactory in modeling $Q_Z$. $RMSE$, $R$ and $NSE$ values set were 5.28 (m$^3$/sec), 0.77 and 0.60 for the calibration dataset, respectively. While these values for the validation dataset were 4.9 (m$^3$/sec), 0.76 and 0.58, respectively. Ghorbani et al. applied MLR, ANN and SVM models to predict water discharge at Big Cypress River, USA. MLR model could not achieve a high efficiency to predict water discharge [18].

$$Q_Z = -11.189 + 0.896Q_B$$  \hspace{1cm} (1)

From aforementioned, it can be concluded that the ANFIS model outperformed the ANN and MLR models based on the three performance criteria: $RMSE$, $R$ and $NSE$ during the calibration and validation periods (Table3). This might be attributed to its sophisticated structure and the capability of eliminating the noisy data [19], ANFIS model (Sugeno) makes use of “IF– THEN” rules to produce an output for each rule [20]. This allows to learn from the data [21]. The neuro-fuzzy systems have an advantage of both ANFIS and ANNs, i.e. benefiting from the training ability of the ANN and the fuzzy IF– THEN rule generation and parameter optimization [22]. Our findings are in parallel with previous studies e.g. [8], [23]–[27], where they proved the superior performance of ANFIS in modeling hydrological and water resource management.
Table 3. Comparison of optimum ANN, ANFIS and MLR models for water quantity ($Q_z$)

| Estimated $Q_z$ (m$^3$/sec) | Model | Calibration | Validation |
|-------------------------------|-------|-------------|------------|
|                              |       | $RMSE$      | $R$        | $NSE$      | $RMSE$ | $R$ | $NSE$ |
| ANFIS                         | 4.11 (m$^3$.sec$^{-1}$) | 0.87 | 0.76 | 4.17 (m$^3$.sec$^{-1}$) | 0.83 | 0.70 |
| ANN                           | 4.35 (m$^3$.sec$^{-1}$) | 0.85 | 0.73 | 4.26 (m$^3$.sec$^{-1}$) | 0.82 | 0.68 |
| MLR                           | 5.28 (m$^3$.sec$^{-1}$) | 0.77 | 0.60 | 4.9 (m$^3$.sec$^{-1}$) | 0.76 | 0.58 |

Figure 2. Comparative plots of observed and predicted $Q_z$ values using ANFIS model for (a) calibration data set and (b) validation.
Figure 3. Comparative plots of observed and predicted $Q_z$ values using ANN model for (a) calibration data set and (b) validation.
Figure 4. Comparative plots of observed and predicted $Q_z$ values using MLR model for (a) calibration data set and (b) validation.

4. Conclusion
The performance of three methods used in this study was investigated in predicting inflow of Abu-Ziriq Marsh. The studied techniques were ANFIS, MLR and ANN models. This study used discharge of upstream (Al-Badaa regulator) for the prediction of discharge values downstream in Abu-Ziriq Marsh inlet, with the daily data spanning over the period of 2017–2018.

For the purpose of evaluation of the developed models, three criteria were used such as $R$, $RMSE$ and $NSE$. It was found that the ANFIS outperformed the other evaluated methods. In other words, ANFIS model led to the best fit with the observed data. Whereas the ANN model has come in second place in terms of performance measures. This could be attributed to the ANFIS structure. The ANFIS integrates the advantage of the simplifying function of fuzzy reasoning and the self-learning ability of neural networks and thus, gives a strong capability of eliminating noise [28]. ANFIS is recommended to be used as a predictive model for water quantity parameters ($Q_z$) in the Iraqi Marshes.

Acknowledgments
The data used in this study were obtained from MOWR, SCOIDP in the south of Iraq. The author wishes to thank MOWR, SCOIDP in the south of Iraq, who made it easy to obtain the required data

References
[1] K P Singh, A Malik, N Basant and P Saxena 2007 Multi-way partial least squares modeling of water quality data Anal. Chim. Acta, vol. 584, no. 2, pp 385–396
[2] A Zahiri and H M Azamathulla 2014 Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels Neural Comput. Appl., vol. 24, no. 2, pp 413–420
[3] R S Govindaraju 2000 Artificial neural networks in hydrology. II: hydrologic applications J. Hydrol. Eng., vol. 5, no. 2, pp 124–137
[4] L E Besaw, D M Rizzo, P R Bierman and W R Hackett 2010 Advances in ungauged streamflow prediction using artificial neural networks J. Hydrol., vol. 386, no. 1–4, pp 27–37
[5] A P Piotrowski, M Osuch, M J Napierkowski, P M Rowinski and J J Napierkowski 2014 Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river Comput. Geosci. vol. 64 pp 136–151

[6] T Kerh and C S Lee 2006 Neural networks forecasting of flood discharge at an unmeasured station using river upstream information Adv. Eng. Softw., vol. 37, no. 8 pp 533–543

[7] W C Wang, K W Chau, C T Cheng and L Qiu 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series J. Hydrol., vol. 374, no. 3–4, pp 294–306

[8] M Al-Mukhtar and F Al-Yaseen 2019 Modeling water quality parameters using Data-Driven Models, a case study Abu-Ziriq Marsh in South of Iraq Hydrology, vol. 6, no. 1, p 24

[9] W B Chen and W C Liu 2014 Artificial neural network modeling of dissolved oxygen in reservoir Environ. Monit. Assess., vol. 186, no. 2, pp 1203–1217

[10] B H Khudair 2018 Water quality assessment and Total Dissolved Solids prediction using Artificial Neural Network in Al-Hawizeh Marsh south of Iraq J. Eng., vol. 24 no. 4, pp 147–156

[11] S Nemati, M H Fazelifard, Ö Terzi and M A Ghorbani 2015 Estimation of dissolved oxygen using data-driven techniques in the Tai Po River, Hong Kong Environ. Earth Sci., vol. 74, no. 5 pp 4065–4073

[12] M Salari et al. 2018 Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water Food Chem. Toxicol. vol. 118, pp 212–219

[13] S Haykin 1994 Neural networks: a comprehensive foundation. Prentice Hall PTR

[14] J S Jang 1993 ANFIS: adaptive-network-based fuzzy inference system IEEE Trans. Syst. Man. Cybern., vol. 23, no. 3, pp 665–685

[15] Ö Kisi and M Çobaner 2009 Modeling river stage-discharge relationships using different Neural Network Computing Techniques CLEAN–Soil, Air, Water, vol. 37, no. 2, pp 160–169

[16] P Krause, D P Boyle, and F Bäse 2005 Comparison of different efficiency criteria for hydrological model assessment Adv. Geosci., vol. 5, pp 89–97

[17] M Ravansalar, T Rajae and O Kisi 2017 Wavelet-linear genetic programming: A new approach for modeling monthly streamflow J. Hydrol. vol. 549, pp 461–475

[18] M A Ghorbani, R Khatibi, A Goel, M H FazeliFard and A Azani 2016 Modeling river discharge time series using support vector machine and artificial neural networks Environ. Earth Sci., vol. 75, no. 8

[19] M Al-Mukhtar June 2018 Integrated approach to forecast future suspended sediment load by means of SWAT and artificial intelligence models, a case study.

[20] M Khadr and M Elshemy 2017 Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt Ain Shams Eng. J., vol. 8, no. 4, pp 549–557

[21] A A Tofigh, M R Rahimipour, M O Shabani and P Davami 2015 Application of the combined neuro-computing, fuzzy logic and swarm intelligence for optimization of compocast nanocomposites J. Compos. Mater., vol. 49, no. 13, pp 1653–1663

[22] B Kosko 1997 Fuzzy engineering Prentice-Hall, Upper Saddle River, N J

[23] M Ay and Ö Kişi 2017 Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques KSCE J. Civ. Eng., vol. 21, no. 5, pp 1631–1639

[24] S Heddam 2014 Modeling hourly dissolved oxygen concentration (DO) using two different adaptive neuro-fuzzy inference systems (ANFIS): A comparative study Environ. Monit. Assess., vol. 186, no. 1, pp 597–619
[25] M Montaseri, S Zaman Zad Ghavidel and H Sanikhani 2018 Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques Stoch. Environ. Res. Risk Assess., vol. 32, no. 8, pp 2253–2273

[26] A Najah, A El-Shafie, O A Karim and A H El-Shafie 2014 Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring Environ. Sci. Pollut. Res., vol. 21, no. 3, pp 1658–1670

[27] S Zaman Zad Ghavidel and M Montaseri 2014 Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin Stoch. Environ. Res. Risk Assess., vol. 28, no. 8, pp 2101–2118

[28] T Rajaee, S A Mirbagheri, M Zounemat-Kermani and V Nourani 2009 Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models Sci. Total Environ., vol. 407, no. 17, pp 4916–4927