

Article

Not peer-reviewed version

---

# Enhanced Forecasting of Groundwater Level Incorporating Exogenous Variable: Evaluating Conventional Multivariate Time Series and Artificial Neural Network Models

---

[Md Abrarul Hoque](#)\*, [Asib Ahmmed Apon](#), [Md Arafat Hassan](#), [Sajal Kumar Adhikary](#), [Md Ariful Islam](#)

Posted Date: 2 October 2024

doi: 10.20944/preprints202410.0144.v1

Keywords: groundwater level; exogenous variable; ANN; multivariate time series; ARIMAX



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Enhanced Forecasting of Groundwater Level Incorporating Exogenous Variable: Evaluating Conventional Multivariate Time Series and Artificial Neural Network Models

Abrarul Hoque <sup>1,\*</sup>, Asib Ahmmmed Apon <sup>2</sup>, Arafat Hassan <sup>3</sup>, Sajal Kumar Adhikary <sup>4</sup> and Ariful Islam <sup>5</sup>

<sup>1</sup> Department of Civil Engineering, European University of Bangladesh, 2/4 Gabtoli, Mirpur, Dhaka-1216, Bangladesh

<sup>2</sup> Graduate Student, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

<sup>3</sup> Department of Geography, Rutgers University, New Brunswick, USA

<sup>4</sup> Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh

<sup>5</sup> Department of Earth and Atmospheric Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska

\* Correspondence: abrarulmhoque@gmail.com

**Abstract:** Continuous and uncontrolled extraction of groundwater often creates tremendous pressure on groundwater levels. As a part of the sustainable planning and effective management of water resources, it is crucial to assess the existing as well as future groundwater level (GWL) condition. In the current study, an attempt was made to model and forecast GWL using artificial neural networks (ANN) and multivariate time series models. Autoregressive integrated moving average (ARIMA) and ARIMA incorporating exogenous variables (ARIMAX) were adopted as the time series models. Kushtia district in Bangladesh was selected as the case study area, and GWL data of five monitoring wells in the study are used to demonstrate the modeling exercise. Rainfall was taken as the exogeneous variable to explore whether its inclusion enhanced the performance of GWL forecasting using the developed models. The performance of each time series and ANN model was assessed based on various model evaluation criteria. It was evident from the results that the multivariate ARIMAX model (SSE of 15.361) performed better than the univariate ARIMA model with an SSE of 17.217 for GWL forecasting. This demonstrates the fact that the multivariate time series models generated enhanced forecasting of GWL compared to the univariate time series models. When comparing the time series and ANN models, it was found that the ANN-based model outperformed the time series models with the enhanced forecasting accuracy (SSE of 9.894). Results also exhibit a significant correlation coefficient value  $R$  of 0.995 (ANN 6-8-1) for the existing and predicted data. The current study conclusively proves the superiority of ANN over the time series models for the enhanced forecasting of GWL in the study area. Thus, the ANN approach was not only carried out for model building and simulation but also to provide a valuable tool for managing water resources amidst changing environmental conditions.

**Keywords:** groundwater level; exogenous variable; ANN; multivariate time series; ARIMAX

## 1. Introduction

Bangladesh is one of the most densely populated countries in the world. There is population growth and increasing urbanization in addition to climate change. For this reason, stress on groundwater (GW) is increasing rapidly. This earthly resource is important, particularly for developing countries that support agricultural necessity and crop production, drinking purposes, and ecosystems. For this reason, accurate and reliable forecasting of groundwater levels (GWL) is necessary. Furthermore, it is also necessary for sustainable management of water resources and to facilitate the consumptive use of. Additionally, it plays a crucial role in the earth's water cycle. Like many districts in Bangladesh, Kushtia relies heavily on agriculture as a large irrigation project named the Ganges-Kobadak (G-K Project, which covers an area of 197,500 ha) irrigation project serves the

region. In this project, pumps are used to supply water from the river Ganges to the irrigation areas. According to “Statistical Yearbook Bangladesh 2022” by the Bangladesh Bureau of Statistics (BBS), irrigation in Kushtia district is being carried out using different types of tube wells, including deep, shallow, and hand, as well as power pumps. The district’s approximately 68% area is covered by irrigation [1]. Forecasting of GW can help the farmers optimize water use and improve crop yields. Besides that, the area experiences a substantial variation of rainfall (a precipitation pattern that can impact GW recharge rates); GWL prediction will help in planning for dry periods to minimize the severe impacts of droughts and floods. At the time of drought, GW can play an important role when surface water is reduced by giving a buffer to help stabilize the water supply.

Groundwater levels in Bangladesh usually fluctuate on a seasonal basis due to the monsoon rains and dry periods, resulting in a significant impact on water availability all over the year. Urbanization (growth of population) is also a matter of concern because the demand for groundwater is increasing. GWL forecasting can help to manage this demand. Irrigation in this country heavily relies on groundwater due to the sufficient surface water in the dry season, and future water demand will greatly depend on groundwater resources. Groundwater is also affected by pollution; for instance, chemical fertilizers and pesticides used in fields contribute tremendously to its quality. Over-extraction of GW and the levels of the sea can lead to salinity intrusion problems in coastal areas in Bangladesh. Deforestation is a big problem in this country nowadays; it can change the recharge process of GW, which may decrease GW recharge. GW over-extraction can be caused by inefficient irrigation practices because of the use of outdated or inadequate technology, which is widely recognized in the country; as a result, it can affect GW accessibility. The quality and availability of groundwater are decreasing gradually due to factors (man-made or natural) such as over-extraction, contamination (due to various reasons such as poor management of industrial or domestic wastewater), and climate change. The attenuation of underground water levels is also a matter of concern. The goals of this study are to model and predict groundwater levels using the artificial neural network (ANN), the statistical time series autoregressive integrated moving average model (ARIMA), and to incorporate endogenous variables to analyze its effects. It is widely accepted that ANN and ARIMA models are powerful tools for predicting various data. In recent times, time series analysis has gained significant popularity as a statistical method for creating forecast models, and its application has expanded worldwide [2].

Groundwater is sometimes the only dependable supply of fresh water in many developing nations since it is widely accessible, relatively inexpensive to collect, and typically of higher quality than surface water [3,4]. This dependence on groundwater is especially noticeable in areas with little or highly contaminated surface water [5]. Over 70% of all irrigated land in Bangladesh is sustained by groundwater resources. The nation's dense population, expanding agricultural needs, and quick industrialization are the main causes of this widespread reliance [6]. But this dependence also puts a lot of strain on groundwater supplies, raising questions about how sustainable they will be in the long run. There could be serious repercussions from this groundwater depletion, such as decreased water availability, declining water quality, land subsidence, and ecological harm. Lowering groundwater levels in agricultural areas might result in reduced crop yields because of limited water supplies, which exacerbates problems with food security [7,8]. Apart from the physical loss of groundwater, another major worry is its quality, especially in areas where contamination is common [9]. The variety of elements influencing groundwater systems makes it difficult to forecast GWL changes. These variables include both anthropogenic (such as land use changes, irrigation techniques, and industrial water usage) and natural (such as rainfall, temperature, and evapotranspiration) influences [10,11].

Several researchers found that there are two main approaches used in GWL prediction, the data-driven models and the numerical models. They also found that data-driven models are useful in assessing different aspects such as uncertainty, variability, and complexities in water resources and environmental problems [12]. The ANN model is inspired by the human brain, more specifically, its function as structure; for example, the neurons. There are several neurons in NN to process the information given. Additionally, ANN has shown strong potential to capture complex patterns in

data and showed the ability to learn. Consequently, these characteristics make the ANN model a good choice for forecasting. “ In recent decades, artificial intelligence has been widely applied in studies related to water resources [12]. [13] used ANN for daily weather forecasting. To predict daily streamflow [14], researchers used both ANN and ARIMA models. However, this method (ANN) was not used widely until recently [15]. ARIMAX models are applied in predicting electricity usage [16], to forecast grain production [17], to forecast domestic water consumption [18], and to forecast drought [19]. Researchers used different kinds of ANN models, such as probabilistic neural networks (PNN), generalized RBF models (GRBF), globally recurrent neural networks (RNN), and input delay neural networks (IDNN), for GWL forecasting [20]. Because of the limited understanding of aquifer properties in Bangladesh, the conventional GWL prediction models’ applicability is limited; for this reason, soft computing tools are good alternatives that provide higher efficacy [21]. The ANN model has the capability to tell the connection between the historical data that cannot be seen, and this way, it helps to predict and forecast: authors [22] working on water quality forecasting. Many authors [23] used different types of ANN models, such as the threshold ANN (TANN), cluster-based ANN (CANN), and periodic ANN (PANN), to forecast streamflow. Depending on the existing data on hand, the main purpose of a time series study is to build models that can be used to predict the future, and it is often extremely hard to make good predictions [24]. “ARIMA models are well-known for their notable forecasting accuracy and flexibility in representing several different types of time series” [24]. The model has its key limitations; it assumes a linear system (a linear correlation structure, and this is not at all times acceptable in real cases), and for that reason, non-linear patterns cannot be implemented. The key benefit of ANN is the nonlinear capability of modeling [25]. The authors used the ARIMA technique to model various quality parameters of water and rainfall as hydrological variables [26]. Studies from authors [27] show that predicted river flow using automated ARIMA. To predict the discharge of rivers, some authors [28] tried to find a better predictor between ARIMA and ANN. The multiplicative ARIMA approach was attempted to forecast the standardized precipitation index (SPI) and standardized runoff index (SRI) [29].

Because of its ease of use and capacity to represent linear connections in time series data, ARIMA models are widely used [30,31]. Researchers have created hybrid models that include ARIMA with other modeling techniques, like artificial neural networks (ANNs), to overcome the shortcomings of conventional time series models. Artificial neural networks (ANNs) are data-driven models that do not require explicit mathematical representations of the underlying processes to capture complicated, nonlinear correlations in data [32,33]. Because the interactions between variables in groundwater systems are frequently complex and poorly understood, ANNs are therefore especially well-suited for simulating these systems. ANNs are superior to conventional models in many ways [34]. They are very adaptable and able to capture dynamic interactions because, among other things, they can learn from data and adjust to changes in the system. Second, ANNs are useful for anticipating GWL fluctuations in complex systems because they can accurately describe nonlinear relationships to any desired degree. Third, ANNs can be useful in situations where the system is poorly understood or data on the physical processes are unavailable because they do not require the underlying physics of the process to be explicitly stated [35,36]. Groundwater levels can be significantly impacted by the fluctuation of rainfall, which is a key source of groundwater recharge [37]. It is generally acknowledged that one of the most important ways to increase the precision of GWL predictions is to incorporate rainfall data into both ANN and ARIMA models [38]. Because of Bangladesh's monsoon climate, the relationship between rainfall and GWL is especially significant there [39]. This study compares the ARIMAX and ANN models' performances to determine which modeling strategy is best for forecasting GWL variations in the area. In summary, groundwater is an essential resource that needs to be managed carefully to maintain its sustainability. A crucial aspect of groundwater management is the prediction of GWL changes, which offer important insights into groundwater availability in the future and assist in formulating policies for its sustainable use [40,41].

GW is an almost worldwide source of superior freshwater [42]. The study will focus basically on modeling groundwater levels using ANN and ARIMA models and finding the best models based on their various performance evaluation criteria. Furthermore, a key focus of the research will be to find



the future GWL conditions beyond the available well data to understand the upcoming situations in Kushtia, Bangladesh.

2. Study Area and Data Description

In the current study, Kushtia district, covering an area of 1608.8 km<sup>2</sup> in the Khulna division of Bangladesh, is selected. Figure 1 shows the location of the study area, which is positioned in southwest Bangladesh. Encompassing the coordinates of 23°42' to 24°12' north and 88°42' to 89°22' east. Located in this area, the Ganges-Kobadak irrigation project (or G-K project) is an extensive surface irrigation network that serves the southwestern districts of Kushtia, Chuadanga, Magura, and Jhenaidah. The district has six upazilas, of which Daulatpur is the largest, totaling a population of 2,149,692 (according to the Population and Housing Census 2022). For modeling and prediction, five groundwater monitoring stations with a rainfall (RF) station are selected from each upazila (Bheramara, Daulatpur, Kushtia Sadar, Kumarkhali, and Mirpur). The GWL data collected was on a weekly basis. Accordingly, data consisting of 414 weekly groundwater level readings from 1999 to 2006 were sourced from the Bangladesh Water Development Board (BWDB). The RF data obtained here is exogenous input and collected for the same periods.

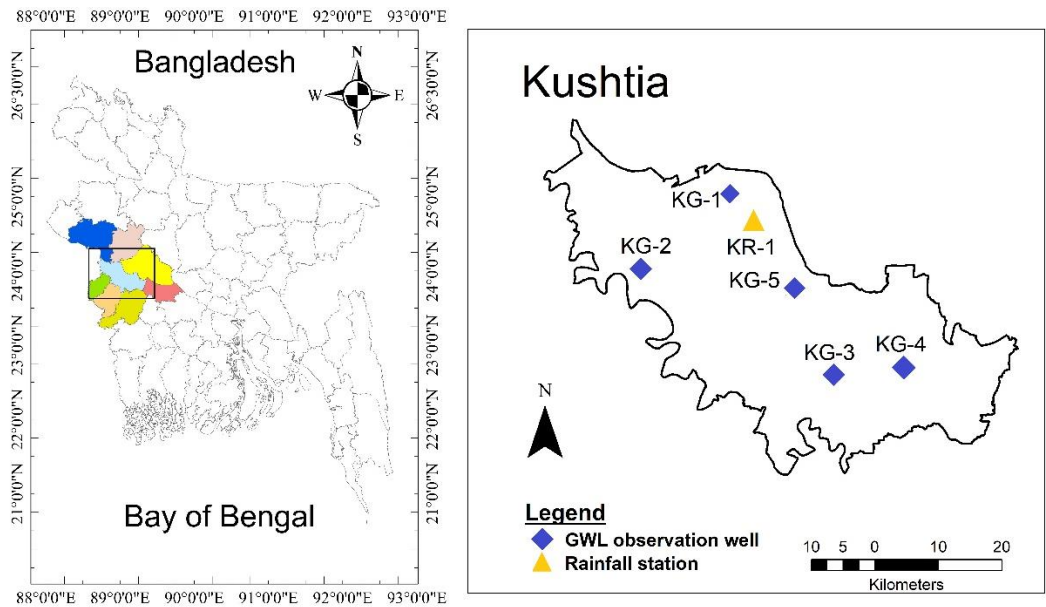


Figure 1. Kushtia district of Bangladesh showing the GWL and RF Stations.

The area is characterized by several rivers such as the Ganges, Mathabhanga, Kaligonga, and Kumar that flow across the district. Additionally, the district has a flat, alluvial landscape, which is typical for the delta region. The area is flat terrain and has fertile soil that makes it favorable to agriculture, which is highly dependent on irrigation. Moreover, Kushtia relies on groundwater as the main water source for irrigation, especially during the dry season when there is limited surface water availability. The area experiences an average maximum temperature of 37.8°C and an average minimum of 9.2°C. The yearly rainfall averages 1,467 millimeters.

Information on the selected groundwater level monitoring stations and rainfall stations with locations are provided is shown in Table 1.

Table 1. GWL and RF station details.

SL No	Station ID	Station Type	Sub-District (Location)	Latitude (°)	Longitude (°)	Data Point (Weeks)
1	KG-1	GWL	Bheramara	24.09	88.96	414
2	KG-2	GWL	Daulatpur	23.98	88.83	414
3	KG-3	GWL	Kushtia Sadar	23.83	89.10	414

4	KG-4	GWL	Kumarkhali	23.84	89.20	414
5	KG-5	GWL	Mirpur	23.93	89.02	414
6	KR-1	RF	Mirpur	24.05	88.99	414

3. Methodology

The primary objective of this study is to model and forecast groundwater level changes using an artificial neural network (ANN) and statistical autoregressive integrated moving average with exogenous variables (ARIMAX) time series models. To study the connection between climate variation and GWL and to forecast GWL, time series modeling is considered one of the most robust statistical methods. [42]. Both univariate and multivariate models are studied to find the best predicting model. Univariate models rely solely on groundwater level data, whereas multivariate models include rainfall data as an additional external factor alongside groundwater level data. The main advantage of using ANN models over conventional techniques is that it does not require the underlying processes, which are complex in nature, to be explicitly defined in mathematical form [43]. Previous research by authors [44] employed ANN models to forecast groundwater level changes for future periods.

Univariate and multivariate, both ANN models were constructed, and the top-performing model was identified. Results from other authors [14] showed that forecasting time series is more accurate in ANN models than in ARIMA-based models. Accepting all complex parameters as input, ANN models generate patterns during model training, and then they use the same patterns to generate the forecasts or predictions [13]. The strength of ANN models in prediction lies in their data-driven nature, their ability to detect previously unseen patterns, and their efficiency with large datasets. [45]. In both cases, model development and evaluation have been carried out. After that, 100 weeks of future prediction of the GWL data have been carried out based on the existing data and variables.

3.1. ANN Model Development

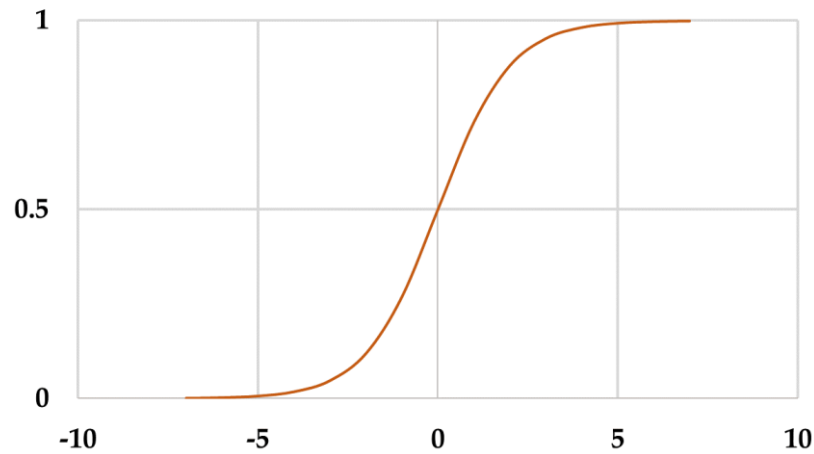
Many techniques exist for creating a neural network [20]. Different ANN models are developed, and several model architectures are analyzed by using a trial-and-error approach to find the most accurate prediction model. in the MATLAB platform using neural network toolbox. This toolbox offers proficiency in designing various neural network configurations with so many applications [20]. Creating an ANN model requires steps such as defining its type, structure, variable processing, training algorithm, and stopping criteria [23].

3.1.1. Dataset Processing, Model Architecture and Training

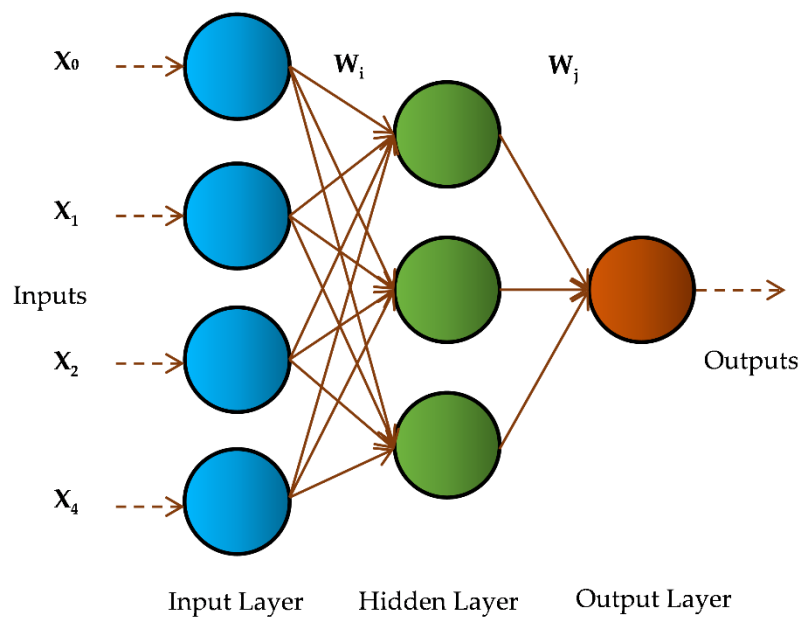
At first, the datasets are taken and prepared as input in the model. For univariate models, only GWL data were taken. The GWL data and RF data are on a weekly basis with 414 data points (weeks). Then, the model inputs are specified. For univariate models, only GWL data is used. However, for multivariate models, both GWL and RF data were taken as inputs. Typically, data is allocated as 70% for training, 15% for validation, and 15% for testing. The multilayer perceptron (MLP) feed-forward ANN model is applied, as it is widely used in hydrological modeling [23]. Some authors [12] selected the hidden layer node count using a trial-and-error. For the current study, hidden layer size was specified based on the best model performance by this approach. The usual design of an ANN model consists of an input layer, a hidden layer, and an output layer. In this study, the sigmoid activation function is used in the hidden layer of the ANN model with the linear activation function in the output layer. The input and output variables are standardized between 0 and 1, to make them fall within a specified range following the ANN modeling framework. The function introduces non-linearity into the network, enabling it to learn and model intricate patterns more effectively. The function works following Equation 1. A diagram is also provided in Figure 2. A typical ANN architecture is shown in Figure 3.

$$y_j = f(x) = \frac{1}{1 + e^{-x}}$$

(1)



**Figure 2.** Sigmoid function.



**Figure 3.** Typical ANN model architecture.

### 3.2. ARIMA Model Development

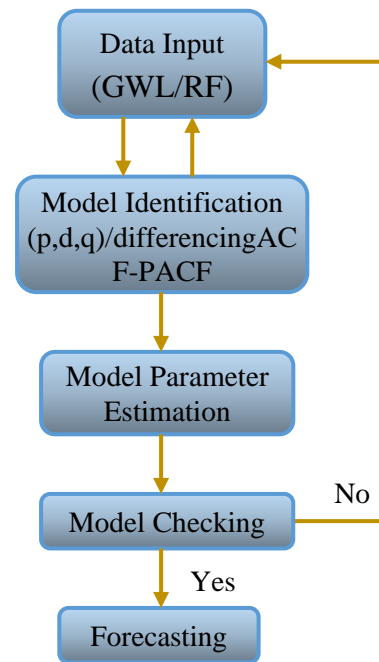
Besides ANN models, the study adopts time series models for predicting groundwater level changes: ARIMAX (autoregressive integrated moving average with exogenous input) and ARIMA (autoregressive integrated moving average). One of the most frequently used time series models for analyzing and forecasting hydrologic data is ARIMA, which combines autoregressive (AR) and moving average (MA) components. Since they are univariate, they cannot deal with exogenous variables.

An ARIMAX time series model is an extension of the ARIMA model that includes one or more external variables. The ARIMA model contains three parts ( $p$ ,  $d$ , and  $q$ ), where  $p$  = order of autoregression,  $d$  = order of integration (differencing), and  $q$  = order of moving average, and it can be expressed by. ARIMAX models proved effective in predicting various extreme weather events, including heavy rainfall and droughts [46]. ARIMAX models demonstrated superiority over multiple regression models in both calibration and validation periods [2]. The general mathematical form or multiplicative equation of an ARIMAX ( $p$ ,  $d$ ,  $q$ ) model with one exogenous variable is given by Equation 1.

$$\phi(L)(1-L)^d Y_t = c + \beta X_t + \theta(L)\epsilon_t \quad (1)$$

Where,  $C$  is the constant,  $\varepsilon_t$  is the error term.  $L$  = lag operator,  $\phi(L) = (1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p)$ , the autoregressive polynomial,  $\theta(L) = (1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q)$ , the moving average polynomial and  $\beta$  is the coefficients for the exogenous variables.

For ARIMA and ARIMAX models, the procedure involves specifying the model structure, estimating parameters, conducting residual diagnostics, and ultimately forecasting the data series. In this study, future predictions are also carried out to determine the upcoming GWL conditions. The study is carried out using the econometric modeler toolbox in the MATLAB platform. The ARIMA process is shown in Figure 4.



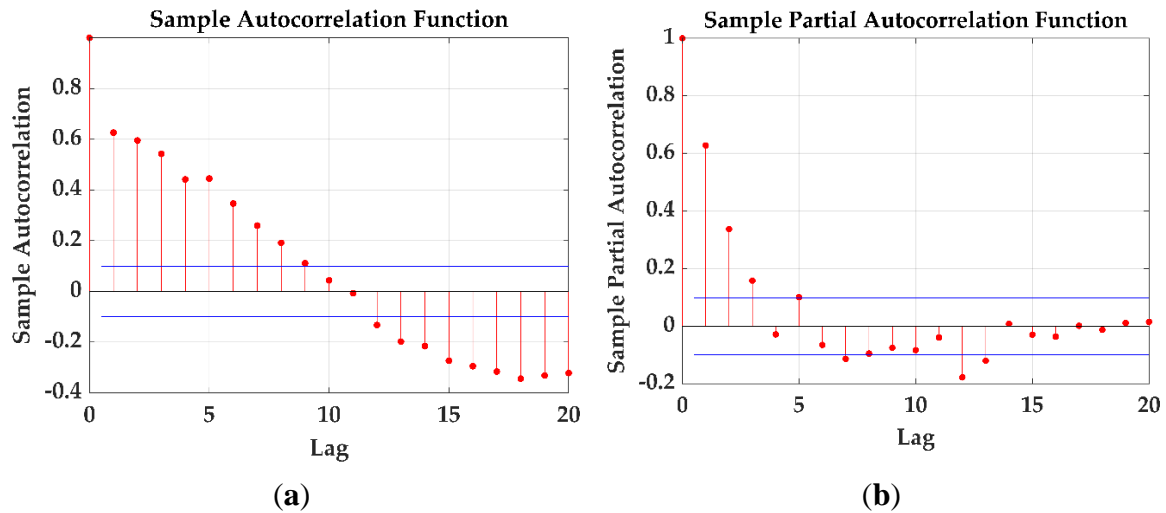
**Figure 4.** ARIMA Model Building Process.

For ARIMA and ARIMAX models, the procedure involves specifying the model structure, estimating parameters, conducting residual diagnostics, and ultimately forecasting the data series. In this study, future predictions are also carried out to determine the upcoming GWL conditions. The study is carried out using the econometric modeler toolbox in the MATLAB platform.

### 3.2.1. Model Identification

For the ARIMA model, at first, the time series (GWL) is to be transformed to stationary by differencing ( $d = 1, 2, 3, \dots$ ). The stationarity of a time series can be assessed using the Augmented Dickey-Fuller Test. The values for  $p$  (autoregressive order) and  $q$  (moving average order) are selected based on ACF and PACF plots. With no significant correlation beyond lag 3,  $p$  and  $q$  are both set to 3. Once the model was identified, its parameters were estimated through the maximum likelihood approach. In order to obtain the best model possible, these  $p$ ,  $d$ ,  $q$  terms (AR1, AR2, AR3,  $d = 1$ , MA1, MA2, and MA3) should be applied using the ACF and PACF plots. After that, each possible combination is analyzed (ARIMA (0,0,0) to ARIMA (3,2,3)) and evaluated for accuracy (using AIC and BIC) to pick the best model. The ACF and PACF plots are shown in Figure 5(a,b), respectively (Station ID: KG-1).





**Figure 5.** For Station ID: KG-1 (a) ACF; PACF plots.

### 3.3. Model Evaluation Criteria

In any model, during forecasting, accuracy is the main concern and not the processing time, and it is observed that there aren't any models that can forecast with full accuracy, but the errors can be reduced using various techniques [13]. There are different widely used model evaluation methods to assess the efficiency of the adopted techniques, including the mean squared error MSE (represented by Equation 2), root mean square error RMSE (represented by Equation 3), Nash-Sutcliffe efficiency NSE (represented by Equation 4), and sum of squared errors SSE (represented by Equation 5).

$$MSE = \frac{\sum_{q=1}^n [Y_{obs}(q) - Y_{est}(q)]^2}{N} \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

$$NSE = 1 - \frac{\sum_{q=1}^n [Y_{obs}(q) - Y_{est}(q)]^2}{\sum_{q=1}^n [Y_{obs}(q) - \bar{Y}_{est}(q)]^2} \quad (4)$$

$$SSE = \sum_i^n [Y_{obs}(q) - Y_{est}(q)]^2 \quad (5)$$

#### 3.3.1. ANN Model Performance Evaluation

In the current study, RMSE, NSE, and SSE were used for ANN models to validate their performance. Generally, the model with the lowest error will have the highest efficiency, and that will be the best model. The equations are shown below. The residual sum of squares (RSS), sometimes called the sum of squared residuals (SSR) or the sum of squared errors, is the aggregate of squared differences between predicted values and actual data. It serves as an indicator of model error in statistics.

#### 3.3.2. ARIMAX Model Performance Evaluation

In order to analyze the accuracy of ARIMAX models, in addition to MSE, SSE, and RMSE, the Akaike Information Criterion (AIC), presented as Equation 6, and Bayesian Information Criterion

(BIC), presented as Equation 7, were applied to evaluate the models. The model with the lowest error indicates the best model. The corresponding mathematical expressions are expressed as the following equations:

$$AIC = -2\log L + 2 \times \text{numParam}$$

(6)

$$BIC = -2\log L + \log(N) \times \text{numParam}$$

(7)

Where:  $-2\log L$  is the goodness of fit. The likelihood  $L$  reflects how well the model explains the data, and the logarithm of  $L$  is used to simplify the calculations and  $2 \times \text{numParam}$  penalizes the model for having too many parameters.  $\log(N) \times \text{numParam}$ : this term penalizes the model for having too many parameters  $Y_{obs}$  is the observed data,  $Y_{est}$  is the estimated data,  $\bar{Y}_{est}$  is the mean value of estimated data and  $N$  is the number of observations.

4. Results and Discussion

The current study focuses on analyzing ANN and multivariate time series ARIMAX models to predict groundwater levels in the Kushtia district in Bangladesh after developing, testing, and evaluating the models. In addition, an attempt has also been made to predict the future scenario of the groundwater levels. The results indicate that incorporating exogenous input (as RF data used in this study) provides better results. Also, the study showed that ANN models yield more accurate predictions compared to ARIMAX models. The modeling is performed considering both multivariate (GWL and RF data) and univariate (only GWL data) inputs.

4.1. Performance of ANNs

It was found from the results that the lowest sum of squared errors (SSE) for the station KG-1 was found to be 9.894 for the ANN-based multivariate models; the model architecture was ANN 6-8-1. On the other hand, for univariate models, the best-performing model was found for the same station (KG-1) with a value of 10.809 (SSE); the model architecture was ANN 3-7-1. It is clear that for GWL prediction in this study, the multivariate model performed better than the univariate model, which includes an exogenous input. Results of the best-performing models for each station are shown in Tables 2 and 3.

Table 2. Best ANN (MV) model results.

Station ID	Model Architecture	Model Performance		
		RMSE	NSE	SSE
KG-1	ANN 6-8-1	0.1546	0.988	9.8946
KG-2	ANN 7-8-1	0.1688	0.979	11.799
KG-3	ANN 10-4-1	0.2319	0.965	26.910
KG-4	ANN 6-7-1	0.2550	0.986	22.273
KG-5	ANN 8-9-1	0.1801	0.984	13.434

Table 3. Best ANN (UV) model results.

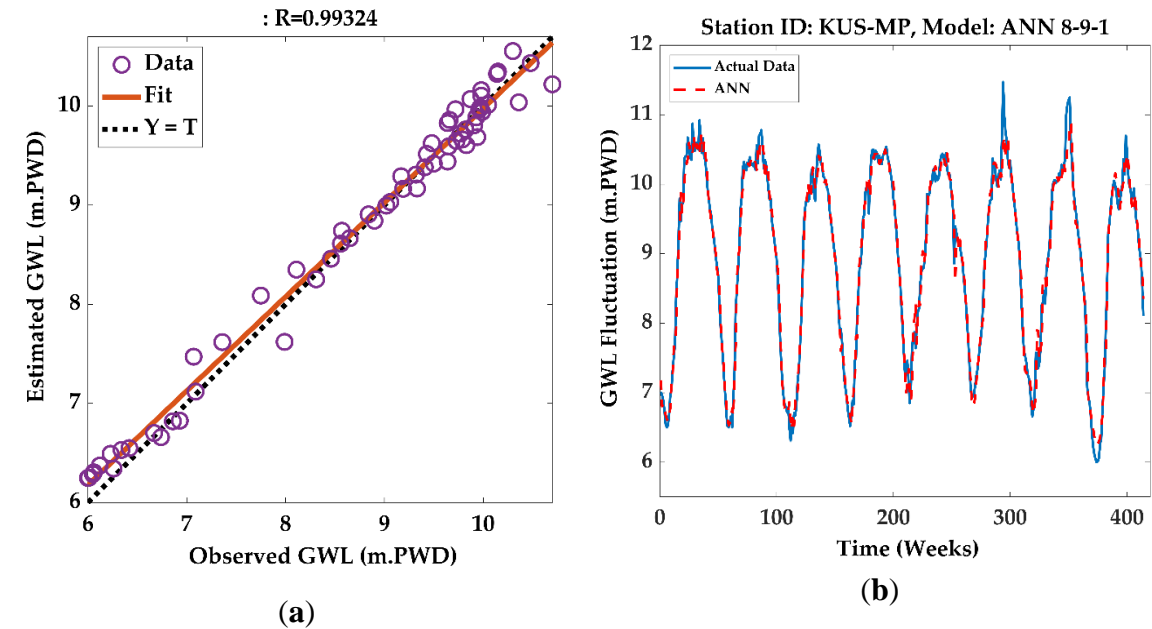
Station ID	Model Architecture	Model Performance		
		RMSE	NSE	SSE
KG-1	ANN 3-7-1	0.1616	0.987	10.809
KG-2	ANN 2-3-1	0.1715	0.979	12.171
KG-3	ANN 4-9-1	0.2588	0.957	26.802
KG-4	ANN 2-4-1	0.2544	0.986	27.725
KG-5	ANN 5-10-1	0.1812	0.9841	13.595

The model architecture was determined by a trial-and-error method for five GWL monitoring stations. Table 4 presents the thorough results for the KG-5 station for the models’ training, validation, and testing stages.

**Table 4.** Selected ANN (MV) model results for Station ID: KG-5.

Model	Training		Validation		Test	
	MSE	NSE	MSE	NSE	MSE	NSE
ANN 8-2-1	0.030	0.984	0.066	0.964	0.046	0.978
ANN 8-3-1	0.032	0.982	0.061	0.967	0.045	0.978
ANN 8-4-1	0.046	0.975	0.069	0.963	0.076	0.963
ANN 8-5-1	0.036	0.980	0.074	0.960	0.037	0.982
ANN 8-6-1	0.031	0.983	0.078	0.958	0.042	0.980
ANN 8-7-1	0.030	0.983	0.201	0.892	0.040	0.981
ANN 8-8-1	0.035	0.981	0.079	0.958	0.042	0.980
ANN 8-9-1	0.032	0.983	0.083	0.955	0.032	0.984
ANN 8-10-1	0.027	0.985	0.071	0.962	0.039	0.981

The scatterplot shown in Figure 6(a) depicts the model's actual versus predicted data at the station. The close alignment of the circular points indicates a high degree of similarity between the observed and predicted values. The correlation coefficient R measures the linear relationship's strength and direction, with values between -1.0 and +1.0. A coefficient close to +1 or -1 reflects a strong correlation, whereas a value near 0 implies little to no correlation. Here, the R values are found to be greater than 0.90. So, they are considered very related. A graphical representation of the actual and predicted data plot for that station is shown in Figure 6(b) (with model architecture ANN 8-9-1).



**Figure 6.** For Station ID: KG-5 (ANN 8-9-1): (a) Scatterplots; (b) Actual and Predicted data plot.

4.2. Performance of ARIMA-Based Models

ARIMA-based multivariate models (ARIMAX) showed an SSE of 15.361 with the model architecture ARIMAX (3,0,3) for the station KG-1. However, the univariate models’ best performance was found in the same station (KG-1), model architecture ARIMA (2,0,1), and an SSE of 17.217. It is

noticeable from the SSE that multivariate models performed better than univariate models. Detailed results are provided in Tables 5 and 6.

Table 5. Best ARIMAX model results.

Station ID	Model Architecture	Model Performance (SSE)
KG-1	ARIMAX (3,0,3)	15.361
KG-2	ARIMAX (3,0,2)	18.721
KG-3	ARIMAX (1,0,3)	25.449
KG-4	ARIMAX (2,0,0)	63.680
KG-5	ARIMAX (3,0,2)	15.143

Table 6. Best ARIMA model results.

Station ID	Model Architecture	Model Performance (SSE)
KG-1	ARIMA (2,0,1)	17.217
KG-2	ARIMA (2,0,1)	26.880
KG-3	ARIMA (2,0,3)	28.207
KG-4	ARIMA (3,0,1)	64.582
KG-5	ARIMA (2,0,1)	16.585

Several ARIMA-based models are built after the identification of the model, and selected ARIMA (p, d, q) models detailed results for Station KG-5 are provided in Table 7 with AIC and BIC.

Table 7. Selected ARIMA (p, d, q) model results for Station ID: KG-5.

Model	SSE	MSE	RMSE	AIC	BIC
ARIMA (0,2,1)	20.688	0.050	0.224	-59.599	-47.521
ARIMA (1,2,2)	20.688	0.050	0.224	-55.600	-35.471
ARIMA (1,2,3)	20.680	0.050	0.223	-53.750	-29.594
ARIMA (2,0,0)	21.220	0.051	0.226	-47.077	-30.974
ARIMA (2,0,1)	16.585	0.040	0.200	-147.100	-126.971
ARIMA (2,0,2)	16.519	0.040	0.200	-146.764	-122.609
ARIMA (3,0,1)	16.537	0.040	0.200	-146.317	-122.162
ARIMA (3,2,1)	20.627	0.050	0.223	-54.820	-30.665
ARIMA (3,2,2)	20.563	0.050	0.223	-54.092	-25.911
ARIMA (3,2,3)	20.021	0.048	0.220	-63.151	-30.944

Additionally, the model parameters (constant, AR{1}, AR{2}, MA{1}, variance) and their statistics for ARIMA(2,0,1) of that station (KG-5) are provided in Table 8.

Table 8. Model Parameters for ARIMA (2,0,1) (Station ID: KG-5).

Parameters	Value	Standard Error	T Statistic	P Value
Constant	0.131533	0.009164	14.35275	1.02×10 <sup>-46</sup>
AR{1}	1.969639	0.008054	244.5408	0
AR{2}	-0.98431	0.007999	-123.057	0

MA{1}	-0.93165	0.020288	-45.9215	0
Variance	0.040065	0.002092	19.1474	1.02× <sup>-81</sup>

For the ARIMA model analysis, a main observation was the decay pattern in the ACF plot. Typically, the ACF shows a gradual decay. The PACF plot was used to determine the presence of significant partial autocorrelations. The initial inspection of the ACF and PACF plots provided information about the potential orders of the AR and MA components of the ARIMA model. For instance, if the ACF displays a significant spike at lag 1 and the PACF cuts off after lag 0, this suggests an ARIMA model with an AR(1) and MA(0) component might be suitable. The selection of AR and MA terms based on ACF and PACF helps in fitting the ARIMA model more accurately. Variations in the plots can impact the interpretation, and it requires careful consideration and validation of the ARIMA model chosen.

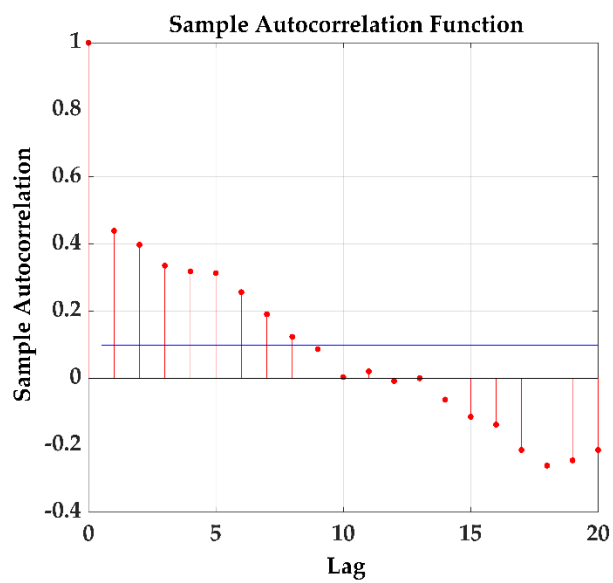
A histogram displays the frequency of data points within specified intervals consisting of bars. It is particularly valuable for assessing the residuals, which are the differences between the actual and predicted values by the model. A well-specified ARIMA model should yield residuals that are approximately normally distributed. The symmetry of the histogram about zero indicates that the ARIMA model does not systematically overestimate or underestimate the data. Residuals centered around zero suggest that the model’s predictions are unbiased on average. This visual inspection is crucial for identifying potential model inadequacies before proceeding to more formal statistical tests.

In a Q-Q plot, the quantiles of the residuals are plotted against the quantiles of the theoretical distribution. The points on the Q-Q plot will approximately lie on a straight line if the residuals are normally distributed. This plot of residuals provides valuable diagnostic information. ARIMA models typically assume that residuals are normally distributed, and the Q-Q plot helps verify this assumption by comparing the quantiles of the residuals to those of a normal distribution. A linear alignment of points along the reference line suggests that the residuals adhere to normality, and it indicates that the model adequately captures the underlying data patterns. This diagnostic tool provides a visual assessment of whether the residuals conform to the normality assumption and is crucial for the reliability of forecasts generated by the model.

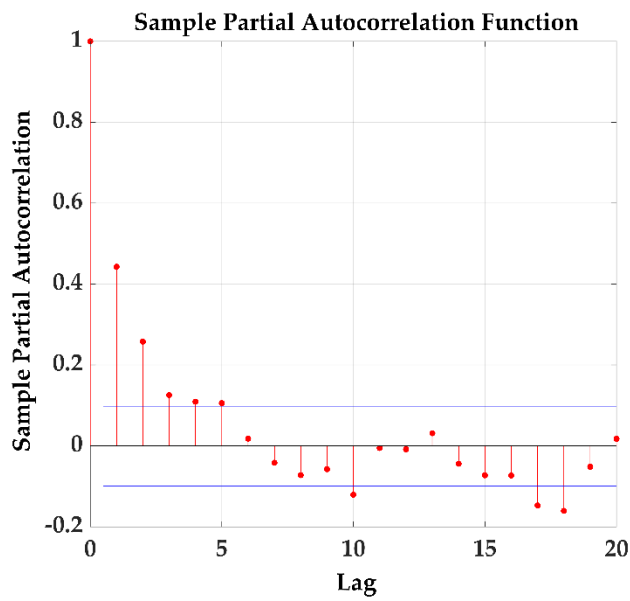
A residual plot displays the residuals (the differences between the observed values and the model's predictions) on the vertical axis and the predicted values or time on the horizontal axis. It helps evaluate the model's performance by showing various aspects of ARIMA models. Ideally, the residuals should be randomly scattered around zero. This randomness indicates that the ARIMA model has effectively captured the underlying patterns in the data. If the spread increases or decreases, it suggests that the model may need adjustments.

Detailed graphical representation (ACF, PACF, QQ Plot, Histogram, Actual and model prediction, and Residual Plots) of the models is shown in Figure 7 (a–f).

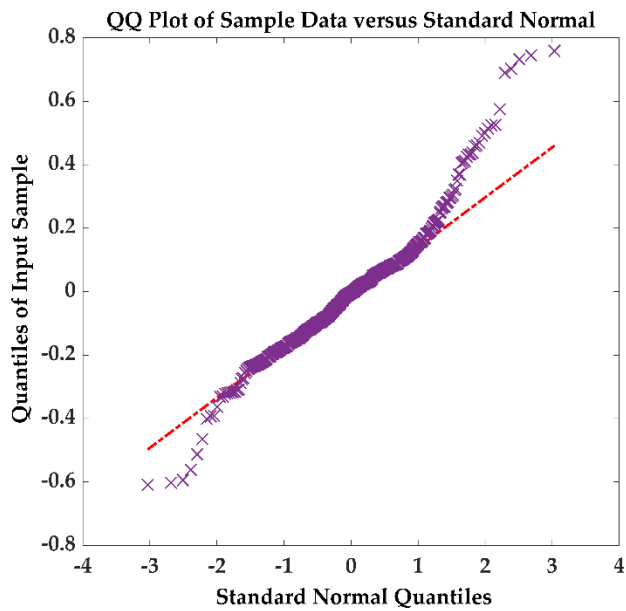




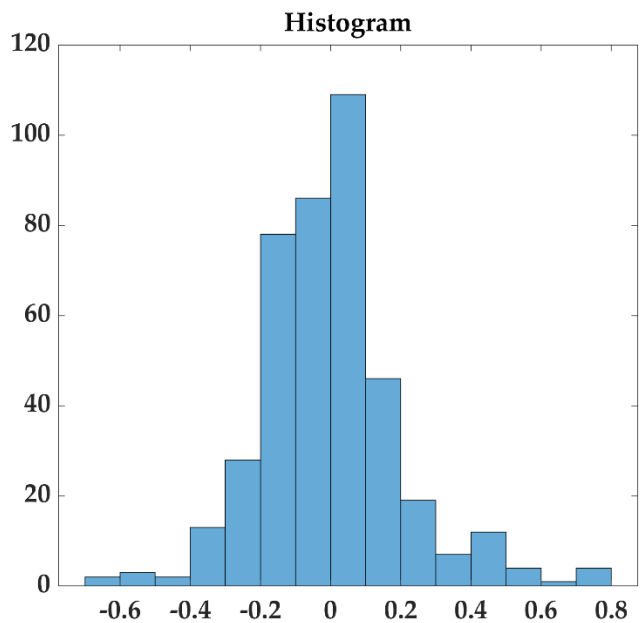
(a) ACF



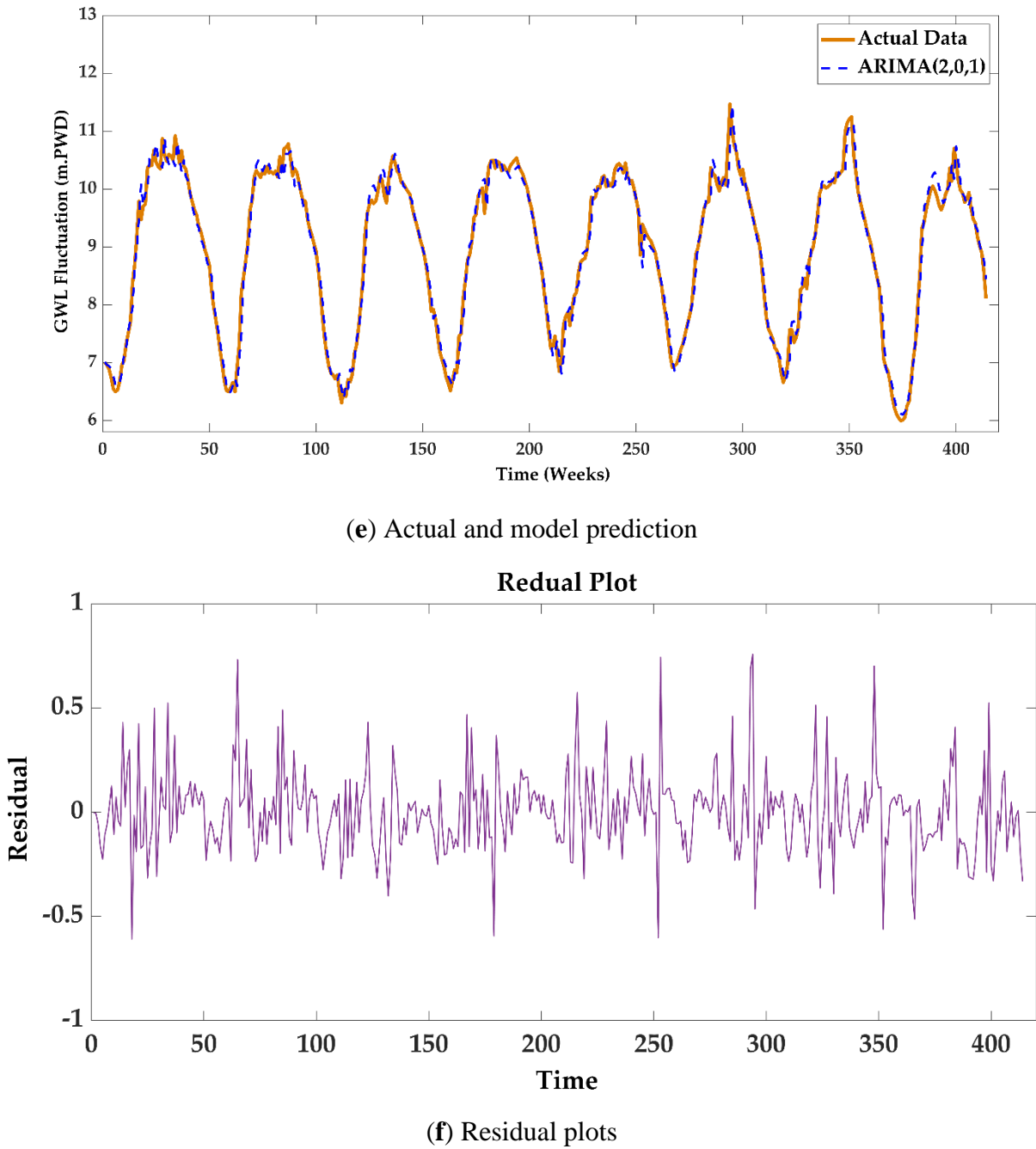
(b) PACF



(c) QQ plot



(d) Histogram



**Figure 7.** ARIMA (2,0,1) model plots for Station ID: KG-5 where (a) ACF; (b) PACF; (c) QQ plot; (d) Histogram; (e) Actual and model prediction; (f) Residual plots.

4.3. Forecasting of GWL

Finally, using the best-performing models, 100 weeks of future groundwater level data is predicted using the existing data. Future GWL values were forecasted using the ANN model. It was observed that the groundwater level (GWL) ranged from a maximum of 10.797 meters to a minimum of 5.875 meters. However, for the existing data, the highest and lowest values for the existing data were 12.61 m and 5.73 m, and besides that, the highest, lowest, and average future GWL values indicate differences in the predicted water levels compared to the observed raw data. The key findings are summarized in Table 9. The ANN models' performances are due to the model's architecture or training parameters. The ARIMAX results could be attributed to its incorporation of exogenous regressors, which might better account for factors influencing water levels. Future work should also consider incorporating a wider range of models and hybrid approaches to balance the

strengths of each predictive technique. Moreover, additional validation against independent datasets will be crucial to confirm the robustness of the models and ensure their generalizability.

The graphical representation of future (100 weeks beyond the current 414 weekly data points) GWL data is shown in Figure 8(a) for the ARIMAX (3,0,3) model and in Figure 8(b) for the ANN 6-8-1 model.

Table 9. Predicted highest, lowest, and average GWL values.

Station ID	Model/Data	Highest (m.PWD)	Lowest (m.PWD)	Average (m.PWD)
KG-1	Existing raw data	12.610	5.730	8.148
KG-1	ANN 6-8-1 (MV)	10.797	5.875	7.742
KG-1	ARIMAX (3,0,3)	11.526	6.270	8.375

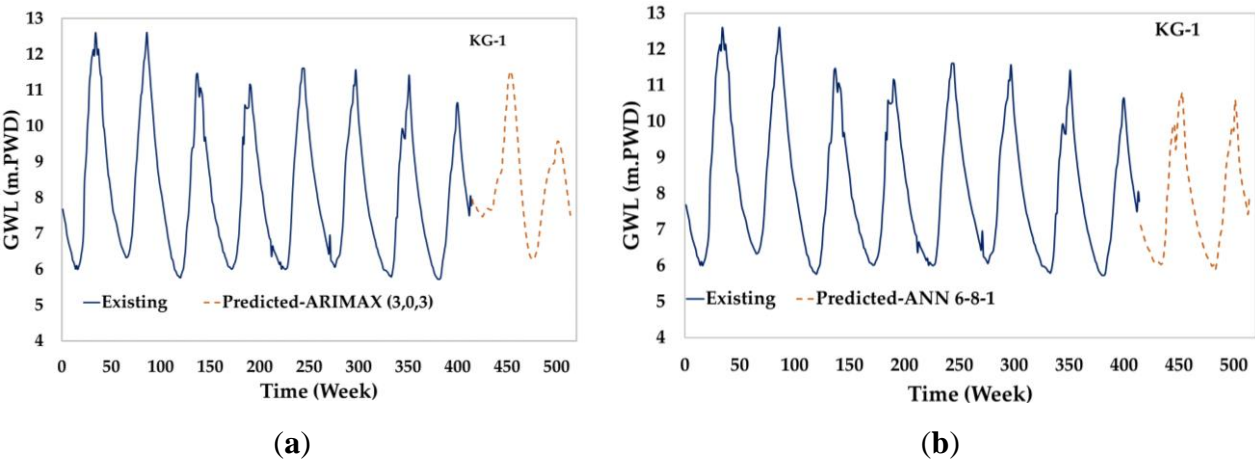


Figure 8. Existing GWL and predicted GWL (a) ARIMAX (3,0,3) (b) ANN 6-8-1.

4.4. Relative Enhancement of the Models' Performance

Comparative performance metrics of different forecasting models for stations are presented in Table 10. It is provided as evaluated based on three key performance metrics:  $\Delta_{PE}$  (performance enhancement, which indicates an improvement in predictive performance),  $\Delta_{PD}$  (performance degradation, which indicates a decline in predictive performance), and BRM (baseline reference metric).

Compared to the ANN model with architecture 6-8-1 (MV), the results of other models indicate a notable drop; specifically, the ANN (MV) outperforms the ANN UV, ARIMAX, and ARIMA models with reductions of -8.45%, -35.58%, and -42.53%, respectively. In contrast to the ANN model with architecture 3-7-1 (UV), ANN (MV) showed an enhancement of the performance with an increase of 9.24%, while ARIMAX and ARIMA experienced performance declines of -29.63% and -37.21%, respectively. It indicates that, although ANN (UV) delivers enhanced predictive accuracy, it also exhibits substantial prediction deviations.

For the ARIMAX (3,0,3) configuration, both ANN (MV) and ANN (UV) show significant performance improvements, with increases of 55.24% and 42.11%, respectively. However, the ARIMA model underperforms, demonstrating a performance decline of -10.78%. With the ARIMA (2,0,1) architecture, all models demonstrate a marked performance enhancement: ANN (MV) achieves a 74.00% improvement, ANN (UV) improves by 59.28%, and ARIMAX shows a 12.08% enhancement.

Table 10. Evaluation Metrics and Comparative Performance Improvement of the Predictive models.

Improvement in Performance
----------------------------

Station ID	Results (SSE)	Model Architecture	( $\Delta_{PE}$ , $\Delta_{PD}$ or BRM)				Comment
			ANN (MV) (%)	ANN (UV) (%)	ARIMAX (%)	ARIMA (%)	
KG-1	9.89	ANN 6-8-1 (MV)	0 (BRM)	-8.45 ( $\Delta_{PD}$ )	-35.58 ( $\Delta_{PD}$ )	-42.53 ( $\Delta_{PD}$ )	In contrast to ANN(MV) $\Delta_{PE}$ : -, $\Delta_{PD}$ : ANN (UV), ARIMAX, ARIMA;
KG-1	10.80	ANN 3-7-1 (UV)	9.24 ( $\Delta_{PE}$ )	0 (BRM)	-29.63 ( $\Delta_{PD}$ )	-37.21 ( $\Delta_{PD}$ )	In contrast to ANN (UV) $\Delta_{PE}$ : ANN (MV), $\Delta_{PD}$ : ARIMAX, ARIMA;
KG-1	15.36	ARIMAX (3,0,3)	55.24 ( $\Delta_{PE}$ )	42.11 ( $\Delta_{PE}$ )	0 (BRM)	-10.78 ( $\Delta_{PD}$ )	In contrast to ARIMAX $\Delta_{PE}$ : ANN (MV), ANN (UV), $\Delta_{PD}$ : ARIMA;
KG-1	17.21	ARIMA (2,0,1)	74.00 ( $\Delta_{PE}$ )	59.28 ( $\Delta_{PE}$ )	12.08 ( $\Delta_{PE}$ )	0 (BRM)	In contrast to ARIMA $\Delta_{PE}$ : ANN (MV), ANN (UV), ARIMAX $\Delta_{PD}$ : -;

5. Conclusions

The current study focuses on the modeling and prediction of GWL in Kushtia, a district of Bangladesh. For modeling and prediction, mainly two types of models (with two sub-categories each) were used, named ANN (UV, MV) and ARIMA and ARIMAX. Modeling was carried out in five separate sub-districts in Kushtia using GWL monitoring well data. The best model predicted that the future GWL will experience a change of approximately 0.5 m on average. Specially, it is worth noting that there is a large irrigation project named the Ganges-Kabodakh project in the district serving the water and food necessities in the area. ANNs achieved higher accuracy in forecasts. The correlation coefficient value (0.993) R is very significant and satisfactory for the actual and predicted GWL data for the selected model, ANN 8-9-1. In this study, climatic rainfall data was used as an endogenous variable with GWL data as it heavily influences groundwater. It was found that models incorporating exogenous variables generated the enhanced prediction in both types of techniques. It was also found that time series ARIMA-based models predicted with less accuracy than artificial intelligence-based ANN models. The addition of rainfall data leads to superior results, and it might help future works on other similar aquifers in the country, as GW is heavily affected by it. Apart from that, other endogenous variables can be studied other than rainfall, which is used in this study. The effects of other climate variables might produce better results. Only five GWL monitoring wells were considered in the current investigation. Other monitoring wells from the various parts of the country can be studied based on the framework developed in the current research. Further work can also be done in the coastal areas of the country where salinity intrusion is a crucial problem. Alongside that,

highly GW-contaminated areas should also be focused. Moreover, in these data-driven methods, long-term climatic data can also be collected for model development and prediction. Based on the findings of the current study, it is emphasized that other data-driven models, such as advanced statistical techniques and/or artificial intelligence-based models, including genetic programming (GP), gene expression programming (GEP), adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM), can also be tested to come up with a better modeling framework for the enhanced forecasting of GWL in the study area of Bangladesh or similar areas around the world.

**Author Contributions:** Conceptualization, M.A.H., A.A.A., M.A.H. and S.K.A., M.A.I; methodology, M.A.H., A.A.A. and S.K.A; software, M.A.H. and A.A.A.; validation, S.K.A.; formal analysis, M.A.H., A.A.A; investigation, S.K.A.; resources, S.K.A., M.A.I; data curation, S.K.A.; writing—original draft preparation, M.A.H., A.A.A. and M.A.H.; writing—review and editing, M.A.H., A.A.A., M.A.H. and S.K.A.; visualization, M.A.H.; supervision, M.A.H. and S.K.A.; project administration, M.A.H. and S.K.A.; funding acquisition, M.A.H., M.A.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Bangladesh Ministry of Science and Technology. Funding was provided for the years 2021-2022.

**Data Availability Statement:** The data and codes will be available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors acknowledge the necessary data support provided by Bangladesh Water Development Board (BWDB) for this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Hossain, M. B., Roy, D., Mahmud, M. N. H., Paul, P. L. C., Yesmin, M. S., Kundu, P. K. Early transplanting of rainfed rice minimizes irrigation demand by utilizing rainfall. *Environ. Syst. Res.* **2021**, *10*, 1-11.
- Islam, F., Imteaz, M. A. The effectiveness of ARIMAX model for prediction of summer rainfall in northwest Western Australia. *IOP Conf. Ser.: Mater. Sci. Eng.* **2021**, *1067*, 012037.
- Wada, Y., Van Beek, L.P., Van Kempen, C.M., Reckman, J.W., Vasak, S., Bierkens, M.F. Global depletion of groundwater resources. *Geophys. Res. Lett.* **2010**, *37*.
- Foster, S., Pulido-Bosch, A., Vallejos, Á., Molina, L., Llop, A., MacDonald, A.M. Impact of irrigated agriculture on groundwater-recharge salinity: a major sustainability concern in semi-arid regions. *Hydrogeol. J.* **2018** *26*, 2781-2791.
- Schmoll, O. (Ed.). Protecting groundwater for health: managing the quality of drinking-water sources, World Health Organization 2006.
- Shahid, S., Wang, X.J., Moshir Rahman, M., Hasan, R., Harun, S.B., Shamsudin, S. Spatial assessment of groundwater over-exploitation in northwestern districts of Bangladesh. *J. Geol. Soc. India.* **2015**, *85*, 463-470.
- Dangar, S., Asoka, A., Mishra, V. Causes and implications of groundwater depletion in India: A review. *J. Hydrol.* **2021**, *596*, p.126103.
- Jain, M., Fishman, R., Mondal, P., Galford, G.L., Bhattarai, N., Naeem, S., Lall, U., Balwinder-Singh., DeFries, R.S. Groundwater depletion will reduce cropping intensity in India. *Sci. Adv.* **2021**, *7*, Article Id. 2849.
- Jia, X., O'Connor, D., Hou, D., Jin, Y., Li, G., Zheng, C., Ok, Y.S., Tsang, D.C., Luo, J. Groundwater depletion and contamination: Spatial distribution of groundwater resources sustainability in China. *Sci. Total Environ.* **2019**, *672*, pp.551-562.
- Oh, Y.Y., Yun, S.T., Yu, S., Hamm, S.Y. The combined use of dynamic factor analysis and wavelet analysis to evaluate latent factors controlling complex groundwater level fluctuations in a riverside alluvial aquifer. *J. Hydrol.* **2017**, *555*, pp.938-955.
- Kisi, O., Alizamir, M., Zounemat-Kermani, M. Modeling groundwater fluctuations by three different evolutionary neural network techniques using hydroclimatic data. *Nat. Hazards.* **2017**, *87*, pp.367-381.
- Khedri, A., Kalantari, N., Vadiati, M. Comparison study of artificial intelligence method for short term groundwater level prediction in the northeast Gachsaran unconfined aquifer. *Water Supply.* **2020**, *20*, 909-921.
- Narvekar, M., Fargose, P. Daily weather forecasting using artificial neural network. *Int. J. Comput. Appl.* **2015**, *121*, 0975 – 8887



14. Kassem, A. A., Raheem, A. M., Khidir, K. M. Daily streamflow prediction for khazir river basin using ARIMA and ANN models. *ZJPAS*, **2020**, 32, 30-39.
15. Husna, N. E. A., Bari, S. H., Hussain, M. M., Ur-rahman, M. T., Rahman, M. Ground water level prediction using artificial neural network. *IJHST*. **2016**, 6, 371-381.
16. Rabbi, F., Tareq, S. U., Islam, M. M., Chowdhury, M. A., Kashem, M. A. A multivariate time series approach for forecasting of electricity demand in bangladesh using arimax model, In 2020 2nd International Conference on Sustainable Technologies for Industry 4.0, Dhaka, Bangladesh, 19-20 December 2020; IEEE, (pp. 1-5).
17. Lemos, J. D. J. S., Bezerra, F. N. R. ARIMAX Model to Forecast Grain Production under Rainfall Instabilities in Brazilian Semi-Arid Region. *GJHSS*. **2024**, 24, 1-9
18. Khairi, S. M., Aziz, I. A. Domestic water consumption forecasting with sociodemographic features using ARIMA and ARIMAX: A case study in Malaysia. *Platform: A Journal of Science and Technology*. **2022.**, 5, 16-30.
19. Jalalkamali, A., Moradi, M., Moradi, N. Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. *IJEST*. **2015**, 12, 1201-1210.
20. Chitsazan, M., Rahmani, G., Neyamadpour, A. Forecasting groundwater level by artificial neural networks as an alternative approach to groundwater modeling. *J. Geol. Soc. India*. **2015**, 85, 98-106.
21. Pham, Q. B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A. R. M. T., Talukdar, S., Nguyen, X.C., Ahmed, A.N., Anh, D. T. Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Computing and Applications*. **2022**, 34, 10751-10773.
22. Palani, S., Liong, S. Y., Tklich, P. An ANN application for water quality forecasting. *Mar. Pollut*. **2008**, 56, 1586-1597.
23. Wang, W., Van Gelder, P. H., Vrijling, J. K., Ma, J. Forecasting daily streamflow using hybrid ANN models. *J. Hydrol. Reg. Stud*. **2006**, 324, 383-399.
24. Khandelwal, I., Adhikari, R., Verma, G. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Comput. Sci*. **2015**, 48, 173-179.
25. Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. **2003**, 50, 159-175.
26. Katimon, A., Shahid, S., Mohsenipour, M. Modeling water quality and hydrological variables using ARIMA: a case study of Johor River, Malaysia. *Sustain. Water Resour. Manag*. **2018**, 4, 991-998.
27. Musarat, M. A., Alaloul, W. S., Rabbani, M. B. A., Ali, M., Altaf, M., Fediuk, R., Vatin, N., Klyuev, S., Bukhari, H., Sadiq, A., Rafiq, W., Farooq, W. Kabul River flow prediction using automated ARIMA forecasting: A machine learning approach. *Sustainability*. **2021**, 13, 10720.
28. Fashae, O. A., Olusola, A. O., Ndubuisi, I., Udomboso, C. G. Comparing ANN and ARIMA model in predicting the discharge of River Opeki from 2010 to 2020. *RIVER RES APP*. **2019**, 35, 169-177.
29. Achite, M., Bazrafshan, O., Azhdari, Z., Wałęga, A., Krakauer, N., Caloiero, T. Forecasting of SPI and SRI using multiplicative ARIMA under climate variability in a Mediterranean Region: Wadi Ouahrane Basin, Algeria. *Climate*. **2022**, 10, 36.
30. Rashid, A., Alamgir, M., Ahmed, M.T., Salam, R., Islam, A.R.M.T., Islam, A. Assessing and forecasting of groundwater level fluctuation in Joypurhat district, northwest Bangladesh, using wavelet analysis and ARIMA modeling. *Theor. Appl. Climatol*. **2022**, 150, pp.327-345.
31. Satish Kumar, K., Venkata Rathnam, E., Analysis and prediction of groundwater level trends using four variations of Mann Kendall tests and ARIMA modelling. *J. Geol. Soc. India*. **2019**, 94, pp.281-289.
32. Tealab, A., Time series forecasting using artificial neural networks methodologies: A systematic review. *FCIJ*. **2018**, 3, pp.334-340.
33. As'ad, F., Avery, P., Farhat, C., A mechanics-informed artificial neural network approach in data-driven constitutive modeling. *Int. J. Numer. Methods Eng*. **2022**, 123, pp.2738-2759.
34. Moraes, R., Valiati, J.F., Neto, W.P.G., Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl*. **2013**, 40, pp.621-633.
35. Khashei, M., Bijari, M., A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput*. **2011**, 11, pp.2664-2675.
36. Govindaraju, R.S., Rao, A.R. *Artificial neural networks in hydrology*, eds., 2013; Springer Dordrecht, (Vol. 36).
37. Mancini, S., Egidio, E., De Luca, D.A., Lasagna, M., Application and comparison of different statistical methods for the analysis of groundwater levels over time: Response to rainfall and resource evolution in the Piedmont Plain (NW Italy). *Sci. Total Environ*. **2022**, 846, p.157479.
38. Sekhar, P.H., Kesavulu Poola, K., Bhupathi, M., Modelling and prediction of coastal Andhra rainfall using ARIMA and ANN models. *Int J Stat Appl Math*. **2020**, 5, pp.104-110.
39. Elbeltagi, A., Salam, R., Pal, S.C., Zerouali, B., Shahid, S., Mallick, J., Islam, M.S., Islam, A.R.M.T., Groundwater level estimation in northern region of Bangladesh using hybrid locally weighted linear regression and Gaussian process regression modeling. *Theor. Appl. Climatol*. **2022.**, 149, pp.131-151.

40. Sun, J., Hu, L., Li, D., Sun, K., Yang, Z., Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management. *J. Hydrol. Reg. Stud.* **2022**, *608*, p.127630.
41. Iqbal, N., Khan, A.N., Rizwan, A., Ahmad, R., Kim, B.W., Kim, K. and Kim, D.H., Groundwater level prediction model using correlation and difference mechanisms based on boreholes data for sustainable hydraulic resource management. *IEEE Access* **2021**, *9*, pp.96092-96113.
42. Qadir, B. H., Mohammed, M. A. Comparison between SARIMA and SARIMAX time series Models with application on Groundwater in Sulaymaniyah. *SJCUS.* **2021**, *5*, 30-48
43. Porte, P., Isaac, R. K., Mahilang, K. K. S., Sonboier, K., Minj, P. Groundwater level prediction using artificial neural network model. *Int J Curr Microbiol Appl Sci.* **2018**, *72*, 2947-2954.
44. Ghazi, B., Jeihouni, E., Kalantari, Z. Predicting groundwater level fluctuations under climate change scenarios for Tasuj plain, Iran. *Arab. J. Geosci.* **2021**, *14*, 115.
45. Liu, Q., Zou, Y., Liu, X., Linge, N. A survey on rainfall forecasting using artificial neural network. *Int. J. Embed. Syst.* **2019**, *11*, 240-249.
46. Islam, F., Imteaz, M. A. Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model. *Hydrology* **2020**, *7*, 52.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.