

Article

Use of Factor Analysis (FA), Artificial Neural Networks (ANNs), and Multiple Linear Regression (MLR) for Electrical Conductivity Prediction in Aquifers in the Gallikos River Basin, Northern Greece

Christos Mattas ¹, Lamprini Dimitraki ¹, Pantazis Georgiou ^{2,*} and Panagiota Venetsanou ^{1,3}

¹ Department of Geology, School of Geology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; cmattas@geo.auth.gr (C.M.); ldimitr@geo.auth.gr (L.D.); pvenetsn@geo.auth.gr (P.V.)

² Department of Hydraulics, Soil Science and Agr. Engineering, School of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

³ Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia 2121, Cyprus

* Correspondence: pantaz@agro.auth.gr; Tel.: +30-2310-998752

Abstract: Due to the fact of water resource deterioration from human activities and increased demand over the last few decades, optimization of management practices and policies is required, for which more reliable data are necessary. Cost and time are always of importance; therefore, methods that can provide low-cost data in a short period of time have been developed. In this study, the ability of an artificial neural network (ANN) and a multiple linear regression (MLR) model to predict the electrical conductivity of groundwater samples in the Gallikos River basin, northern Greece, was examined. A total of 233 samples were collected over the years 2004–2005 from 89 sampling points. Descriptive statistics, Pearson correlation matrix, and factor analysis were applied to select the inputs of the water quality parameters. Input data to the ANN and MLR were Ca, Mg, Na, and Cl. The best results regarding the ANN were provided by a model that included one hidden layer of three neurons. The mean absolute percentage error, modeling efficiency, and root mean square error were used to evaluate the performances of the methods and to compare the prediction capabilities of the ANN and MLR. We concluded that the ANN and MLR models were valid and had similar accuracy (using the same inputs) with a large number of samples, but in the case of a smaller data set, the MLR showed a better performance.



Citation: Mattas, C.; Dimitraki, L.; Georgiou, P.; Venetsanou, P. Use of Factor Analysis (FA), Artificial Neural Networks (ANNs), and Multiple Linear Regression (MLR) for Electrical Conductivity Prediction in Aquifers in the Gallikos River Basin, Northern Greece. *Hydrology* **2021**, *8*, 127. <https://doi.org/10.3390/hydrology8030127>

Academic Editor:
Ioannis Panagopoulos

Received: 30 June 2021
Accepted: 21 August 2021
Published: 24 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial neural network; multiple linear regression; groundwater; electrical conductivity; factor analysis

1. Introduction

The continuous increase in the global population over the last few decades and the improvement in human well-being in developed countries have led to increased demands for food production. According to Cay and Uyan [1], the total irrigated land area worldwide increased by more than five times from 1900 to 2000. The increase in agricultural and industrial production resulted in the increased introduction of chemical compounds into water resources [2,3]. Nowadays, nitrates and pesticides are among the most common pollutants of drinking and irrigation water resources [4]. In many areas of the world, water scarcity is mainly due to the fact of quality, rather than quantity. The supply of drinking water is a priority for modern societies [5]. Therefore, the optimal management of this resource is important to meeting increasing demands. The key element in effective management is the assessment of water quality, the identification of pollutants and their source, and the monitoring of the pollutants' fluctuations over time. Water quality is determined by assessing biological, chemical, and physical parameter values [6,7]. Assessment relies on standards developed by competent authorities of each country, which set

maximum permissible concentrations of certain chemicals allowed in water. Groundwater is one of the most important natural resources for drinking and irrigation purposes [8] and supports the socio-economic development of countries. The advantages of groundwater compared to surface water are its higher quality, lower rate of evapotranspiration, and lower vulnerability to contamination [9,10]. Globally, agriculture is the main consumer of groundwater [11]. Conventional investigations of groundwater quality are mainly based on data and measurements performed in the field and analysis of groundwater sample parameters carried out at the laboratory. The selection of parameters to be monitored depends on the objectives of the study and the available funding [12].

Despite the detailed planning and design of the sampling procedure, there are often restrictions regarding availability of time, sampling point accessibility, and lack of funds. Therefore, evaluation of water quality using conventional methods results in economic costs and reduce the decision-making capacity and effectiveness of management programs [13]. In order to overcome such data-scarcity problems, researchers have shown an interest in and increased use of descriptive statistical analysis, multivariate statistical analysis, and artificial neural networks for the evaluation of hydrochemical data in the field of hydrogeology. Multivariate statistics can be used to identify hydrochemical–hydrogeological procedures that determine groundwater quality characteristics [14,15] and distinguish anthropogenic from geological impacts on groundwater composition control [16].

In the last few decades, artificial neural networks (ANNs) have been widely applied in the area of water quality modeling. They are considered a prediction tool and have been widely used in various fields such as flood prediction [17,18], land use [19], and water quality [20], or to predict parameter values such as electrical conductivity and total dissolved solids based on other variables measurements [21–26]. They have also been used in hydrogeology to determine aquifer parameters [27–29], evaluate the qualitative characteristics of groundwater [30], and predict groundwater level [31–34]. ANNs are information processing systems consisting of nonlinear interconnected processing elements called neurons [35].

Regression models are best for establishing an association between dependent and independent variables, and they are considered the simplest and most straightforward form of model. They are based on the method of least squares and are usually considered for the first stage of an investigation of the relationship among variables.

The electrical conductivity value is an index of salinity, and it is often used as an indicator of water quality for agricultural, industrial, or domestic demands. It can highlight the changes over time and space in an aquifer and is usually measured in situ, but the measurement methods are usually timeconsuming [36].

The aim of this paper was the use of an ANN and multiple linear regression (MLR) to predict electrical conductivity (EC), a dependent variable, using independent variables. EC was selected as the most appropriate indicator of water quality in this paper since the GallikosRiver basin (northern Greece) is an area subject to intense anthropogenic agricultural activities, and it has a complex geological structure.

Descriptive statistics and multivariate statistics were enabled to identify the main hydrogeological and hydrochemical processes of the area. These methods were used to reveal the hidden relationships among variables and determine the parameters that were used as inputs in the ANN and MLR models. Finally, MLR and its prediction abilities were compared with the ANN models. To check the prediction accuracy, the coefficient of determination (R^2), mean absolute percentage error (MAPE, %), root mean squared error (RMSE), and modeling efficiency (EF) were used to select the best predictive model. To the best of our knowledge, there have been no prior studies in the GallikosRiver basin implementing prediction of EC using ANN and MLR models. This study introduces the coupling of different statistical methods, along with prediction tools, establishing a methodology that could be applied in the same area for the prediction of other parameters or in any other area of the world.

This paper is structured as follows. Section 2 describes the characteristics of the study area. Section 3 explains the methodology and the theoretical background of the models. Section 4 illustrates and explains the results. Section 5 concludes the paper.

2. Study Area

The study area was the Gallikos River basin (868 Km²) in northern Greece (Figure 1). According to Mattas [37], 90% of the total area lies below an altitude of 600 m, while the mean altitude is 357.7 m. The length of the river within the boundaries of the study area is approximately 48 km. The mean annual precipitation over the basin is approximately 480 mm [37]. According to the Hellenic National Meteorological Service, the climate of the studied area is cold semi-arid (Bsk) [38]. The Gallikos basin belongs to the Serbo-Macedonian massif, Circum-Rhodope belt, and the zone of Peonia [39–42]. A vast area is filled with Quaternary fluvio-lacustrine sediments, due to the existence of the river, and Tertiary formations consisting of marls. The bedrock of the basin is formed from argillaceous schists, carbonate rocks (from limestones to dolomites), quartzites, amphibolites, and gneisses. In the study area, there are no significant surface storage constructions, and the majority of the irrigation demands are covered by groundwater. The main cultivations in the area are corn, tobacco, cotton, sunflower, cereals for forage, trees (mainly almonds and oil-producing olives), and vegetables [43]. Approximately 77% of pumped groundwater is used for irrigation according to the approved River Basin Management Plan-River Basin District GR10-Central Macedonia [44]. Two main aquifer systems have developed in the area. A granular system is formed in the sediments of the basin, and a fractured aquifer system exists in the crystalline rocks of the northeast part.

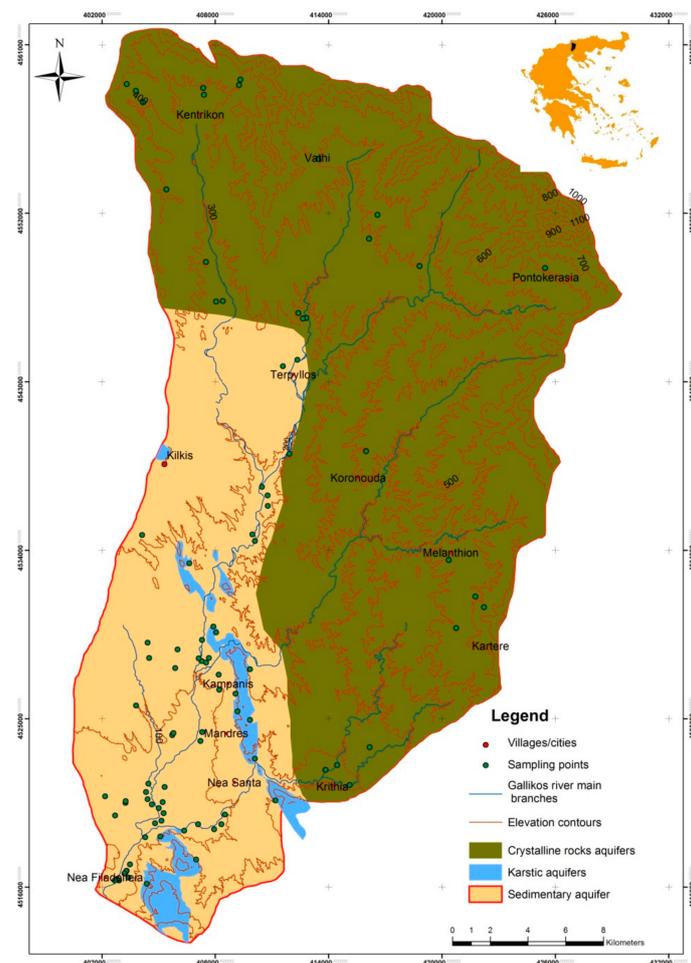


Figure 1. Map showing sampling points in relation to hydrogeological categories of aquifers within the boundaries of the Gallikos River basin [37].

There are two karstic aquifers that are smaller but are of great importance, since they provide good quality water for the drinking demands of the residents in the wider area [44]. Increased concentrations of nitrates (>50 mg/L), sodium (>200 mg/L), and chlorides (>250 mg/L) have been recorded in groundwater samples and are attributed to anthropogenic activities related to the agricultural and industrial sectors [45–48].

3. Materials and Methods

3.1. Water Sampling and Analysis

Hydrochemical data from 233 groundwater samples from 89 sampling points were examined and utilized for statistical treatment, multivariate statistics, multiple linear regression, and artificial neural network models. The IBM SPSS Statistics 25 software was used. Samples were collected over four different sampling periods (wet and dry periods in the years 2004 and 2005). In addition, 18 samples were collected from selected wells in the wet period of 2006. These samples were used as the verification dataset, to check the reliability of the MLR and ANN models. The sampling points had an adequate spatial distribution. In situ measurements of pH and EC were carried out, and the water samples were filtered through 0.45 μ m membrane filters. Each sample was refrigerated at 4 °C in the laboratory. Extra samples were collected and acidified at pH \2 using HCl. All analyses were conducted according to the Standard Methods for the Examination of Water and Wastewater [49] at the laboratory of Land Reclamation Department of the Soil and Water Resources Institute, which is accredited based on ELOT EN ISO/IEC 17025.

3.2. Multivariate Statistical Analysis

Multivariate statistical techniques can identify the factors that determine groundwater quality and are considered a reliable tool for finding pollutant sources and distinguishing anthropogenic or geogenic origins [50–52]. Multivariate statistical techniques, such as factor analysis, are widely employed in environmental studies [53,54]. One of the techniques commonly applied to identify the relationship between water quality parameters is the factor analysis method. In the present study, R-type factor analysis was performed. Selection of the input parameters for successful forecasting using an artificial neural network is crucial. The factor analysis outcomes were used to select the most suitable variables for the implementation of the ANN model.

3.3. Artificial Neural Networks

ANNs are used as a supplementary method to conventional statistics, contributing as an ultimate objective the elaboration and storage of the experimental knowledge and its modification into a useful form for the user to handle [55–57]. A typical ANN consists of artificial processing elements, called neurons or nodes, which interact with each other through synapses (see Figure 2).

The neurons are grouped in layers, and the encoding information is achieved during the process of training and learning. This structure is a widely used model in hydrogeology applications with the ability to recognize patterns among parameters. The most efficient transfer functions are the sigmoid logistic function and the hyperbolic tangent function, which are implemented in most ANN models [58]. Supervised training is based on an “external teacher” that provides the target value for each training phase. The model learns to adjust the synaptic weights, taking into consideration the targets. The objective is to minimize the error by searching for the optimal weights [59]. A standard statistical criterion that is used to evaluate an ANN’s performance is the mean squared error (MSE), which compares the predicted output with the desired output and the coefficient of determination (R^2).

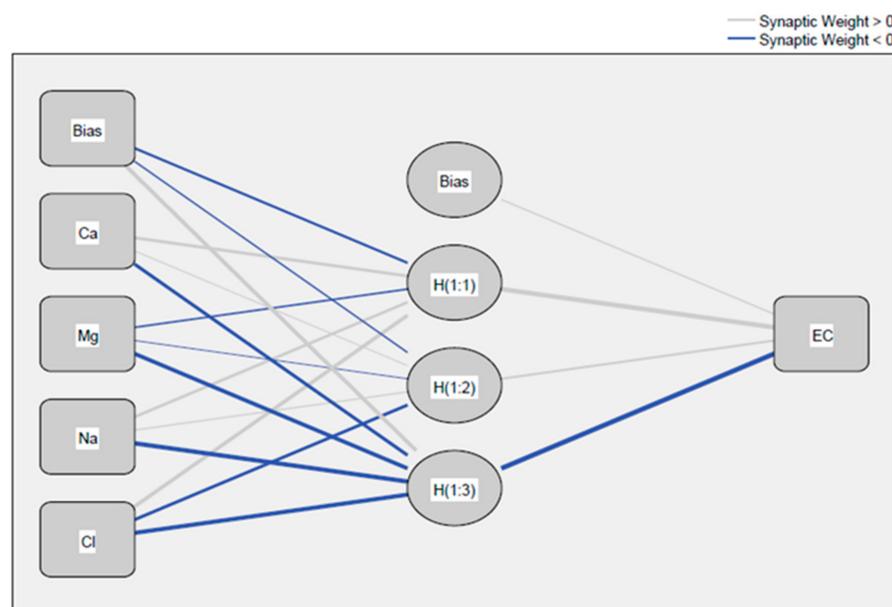


Figure 2. Structure of the artificial neural network based on the available data of the study area.

In the present study, a feed-forward, supervised, back-propagation learning algorithm ANN model was used for predicting the EC of groundwater, using data from the years 2004–2005, in the GallikosRiver basin. The ANN model consisted of one input layer with four elements (i.e., Ca, Mg, Na, and Cl), one hidden layer including three nodes, and the output layer where the EC value was calculated. Of the total sample, 80% was used for training and 20% for testing. The specific artificial neural network structure was selected because it showed the best performance after using the “trial and error” method by modifying the input parameters (number of hidden neurons, number of nodes, percentages of the training–testing sample sets, etc.).

3.4. Multiple Linear Regression

Multiple linear regression (MLR) is considered a very useful and accurate tool that provides equation linking between a dependent variable and a number of independent variables that act as predictors [60].

Different authors have successfully applied this method in hydrogeology and hydrochemistry to predict water quality [60,61] or to establish a statistical model [62]. In the present paper, MLR was employed to provide the equation to predict electrical conductivity. The predictors were selected after implementing the Pearson correlation coefficient, since selecting the appropriate predictor variables is necessary to improve the prediction level and minimize the required dataset [63]. The correlation coefficient (Pearson) is a statistical tool that is widely used to measure and establish the interrelationship and coherence pattern between two variables [63,64]. The advantage of MLR compared to ANNs is that it can provide an equation.

3.5. Performance Evaluation of the Models

The performance evaluation and, hence, the forecasting ability of the models was evaluated using the following statistical indexes:

The coefficient of determination (R^2) gives the percentage variation of variables on the y -axis, explained by variables on the x -axis. The range is from 0 to 1;

The mean absolute percentage error (MAPE, %) is a measure of prediction accuracy of a forecasting method, defined by Equation (1):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (1)$$

where A_t is the actual value, F_t is the forecast value, and n is the number of samples;

The root meansquare error (RMSE) is the square root of the mean of the square of the total error (Equation (2)):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}} \quad (2)$$

where O_i are the observations, S_i are the predicted values of a variable, and n is the number of observations. Thus, RMSE is a good measure of accuracy, but only to compare prediction errors of different models or model configurations for a particular variable and not between variables [65];

Modeling efficiency (EF) is used to compare predicted versus observed values (Equation (3)). A value equal to 1 indicates a perfect model performance. Generally, values that range between 0 and 1 indicate that the values predicted by the model's results are more appropriate for use than the mean value of a dataset, and negative values are worse [66]:

$$\text{EF} = 1 - \frac{\sum_{i=1}^n (O_i - S_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (3)$$

where O_i are the observations, S_i are the predicted values of a variable, and n is the number of observations.

A high R^2 and EF, a low MAE and MAPE, and a low RMSE indicate good model performance.

4. Results and Discussion

The results from the descriptive analysis of the samples for each period are presented in Table 1. The mean value of the nitrates was equal to 38.8 and is considered relatively high, and this cannot be attributed to natural causes. Given that this is an agricultural area, the high values are mainly related to the use of fertilizers and to the lack of a sewerage network for the settlements scattered in the study area during the sampling time periods. The GallikosRiver basin has been characterized as an area vulnerable to nitrate pollution from agriculture. Guidelines for fertilization practices that should be implemented for the protection of the water resources according to crop type, soil slope, and classification are described in the Official Government Gazette of the Hellenic Republic n.1496/v.2/3-05-2019 http://www.et.gr/idocs-nph/search/pdfViewerForm.html?args=5C7QrtC22wFqnM3eAbJzrXdtvSoClrL8JfWk9tSupxYfp1Rf9veiteJInJ48_97uHrMts-zFzeyCiBSQOpYnTy36MacmUFCx2ppFvBej56Mmc8Qdb8ZfRjQZnsIAdk8Lv_e6czmhEembNmZCMxLMtUhnwnTxyShEgwBm79OuvSkRyUUHxgpps8WhFndSwTj11 (Last accessed: 24 August 2021). The maximum values of many samples exceeded the maximum permissible value for potable water, set by the World Health Organization and National Legislation, for the following parameters: EC (7 samples), Na (16 samples), Cl (21 samples), and NO_3 (57 samples).

Table 1. Descriptive statistics of the groundwater samples.

Number of Samples	Period		EC ($\mu\text{S}/\text{cm}$)	pH	Ca (mg/L)	Mg (mg/L)	Na (mg/L)	K (mg/L)	HCO_3 (mg/L)	SO_4 (mg/L)	NO_3 (mg/L)	Cl (mg/L)
233	2004–2005	minimum	282	6.24	12.20	11.00	12.00	0.80	85.40	1.0	0.00	1.0
		maximum	6010	9.78	308.6	120.0	850.0	129.0	878.7	530.7	497.0	886.0
		mean	1139	7.38	112.4	44.16	77.25	12.8	410.8	61.8	38.8	103.5
		SD	686.8	0.4	56.04	21.79	84.66	16.60	117.2	62.9	62.5	135.0

This can be attributed to the operation of fabric dyeing units during the sampling period for Na and Cl, and fertilizers for NO_3 , as aforementioned. Except for the pH, the

values of other parameters varied in a wide range, as indicated by the high values of the standard deviations due to the different conditions that prevail in the different parts of the basin.

The Pearson correlation matrix identified the influence of Ca, Mg, Na, and Cl on EC, finding a significantly positive correlation (Table 2).

Table 2. Pearson correlation matrix.

	pH	EC ($\mu\text{S}/\text{cm}$)	Ca (mg/L)	Mg (mg/L)	Na (mg/L)	K (mg/L)	HCO ₃ (mg/L)	SO ₄ (mg/L)	NO ₃ (mg/L)	Cl (mg/L)
pH	1	−0.075	−0.093	−0.063	0.063	0.135 *	−0.045	−0.142 *	0.075	−0.088
EC	−0.075	1	0.761 **	0.711 **	0.853 **	0.116	0.394 **	0.408 **	0.343 **	0.657 **
Ca	−0.093	0.761 **	1	0.721 **	0.450 **	0.150 *	0.436 **	0.451 **	0.488 **	0.525 **
Mg	−0.063	0.711 **	0.721 **	1	0.426 **	0.085	0.532 **	0.472 **	0.522 **	0.498 **
Na	0.063	0.853 **	0.450 **	0.426 **	1	−0.026	0.266 **	0.196 **	0.096	0.464 **
K	0.135 *	0.116	0.150 *	0.085	−0.026	1	0.139 *	0.000	0.456 **	0.144 *
HCO ₃	−0.045	0.394 **	0.436 **	0.532 **	0.266 **	0.139 *	1	0.102	0.362 **	0.188 **
SO ₄	−0.142 *	0.408 **	0.451 **	0.472 **	0.196 **	0.000	0.102	1	0.193 **	0.510 **
NO ₃	0.075	0.343 **	0.488 **	0.522 **	0.096	0.456 **	0.362 **	0.193 **	1	0.274 **
Cl	−0.088	0.657 **	0.525 **	0.498 **	0.464 **	0.144 *	0.188 **	0.510 **	0.274 **	1

* Correlation was significant at the 0.05 level (two-tailed). ** Correlation was significant at the 0.01 level (two-tailed).

The factor analysis was valid for the four periods, since the Keiser–Meyer–Olkin coefficient had a value of 0.681 (>0.5). At each period, three factors showed eigenvalues higher than 1, based on the selection criteria. These factors explain more than 68.2% of the total variance, which is statistically significant.

The results (Table 3) showed that Na, Cl, Ca, Mg, and EC participated in the first factor, revealing that the main processes defining groundwater quality are pollution from industrial activities in the area and carbonate rock dissolution. Nitrate pollution due to the agricultural activities did not have a strong impact on the EC value, since the nitrates and potassium from agricultural pollution participate in the second factor [67].

Table 3. Results of the R-type factor analysis of the groundwater samples from the years 2004–2005.

Component	Initial Eigenvalues			FACTORS		
	Total	Cumulative %		1	2	3
1	4.263	42.631	pH	0.047	0.138	−0.814
2	1.459	57.221	EC	0.951	0.152	0.053
3	1.097	68.195	Ca	0.719	0.401	0.259
4	0.961	77.802	Mg	0.701	0.425	0.268
5	0.749	85.289	Na	0.883	−0.141	−0.242
6	0.496	90.245	K	−0.069	0.760	−0.182
7	0.388	94.123	HCO ₃	0.391	0.466	0.047
8	0.314	97.258	SO ₄	0.451	0.100	0.572
9	0.240	99.660	NO ₃	0.221	0.843	0.064
10	0.034	100.000	Cl	0.692	0.144	0.259

The participation of SO₄ in the third factor can also be attributed to pollution from agriculture due to the fertilization [68,69].

Therefore, EC cannot be used for the detection of agricultural pollution.

A farmer's income depends on the crop yield which, in turn, relies on irrigation water availability and quality [70]. Application of saline irrigation water causes degradation of soil fertility, and crop problems can develop [71]. Factor analysis can be effectively employed to identify the main factors that affect irrigation water quality.

After the evaluation of the results using descriptive and multivariate statistics, the impact of Ca, Mg, Na, and Cl (independent variables) on electrical conductivity (dependent variable) were established.

The implementation of multiple linear regression using the outcomes of the correlation coefficient and factor analysis resulted in Equation (4) for the total number of samples:

$$EC (\mu\text{S}/\text{cm}) = 3.645 \text{ Ca} + 5.668 \text{ Mg} + 4.675 \text{ Na} + 0.728 \text{ Cl} + 42.839 \quad (4)$$

The MLR method results revealed that the prediction of the dependent parameter (EC) using the parameters that were indicated by the Pearson coefficient was valid, since the coefficient of determination (R^2) was statistically significant (0.94). As depicted in Figure 3, the measured versus predicted values using the MLR method were close to the 1:1 axis. In Figure 4, the error values are plotted very close to the horizontal axis.

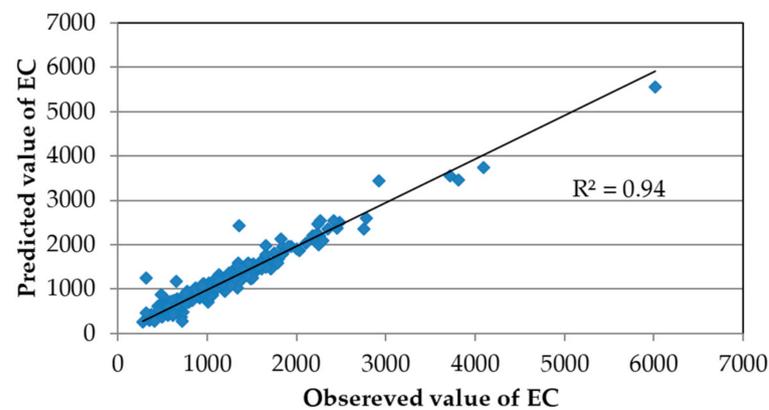


Figure 3. Coefficient of determination of observed versus predicted values of EC using the MLR model.

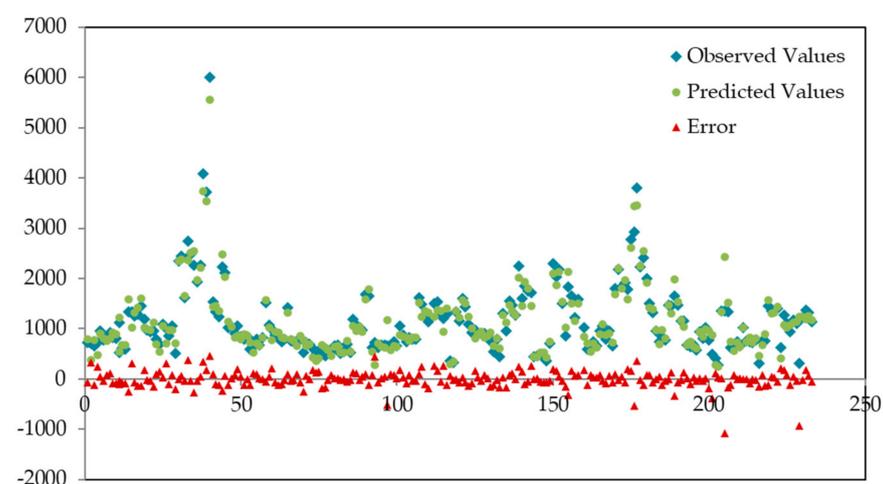


Figure 4. Observed, predicted, and error values of EC ($\mu\text{S}/\text{cm}$) using the MLR model.

The structure of the ANN model is depicted in Figure 2. The results of the ANN are depicted in Figures 5 and 6. According to these figures, the observed and predicted values for the majority of the samples were very close ($R^2 = 0.927$), and the absolute error was around the horizontal axis.

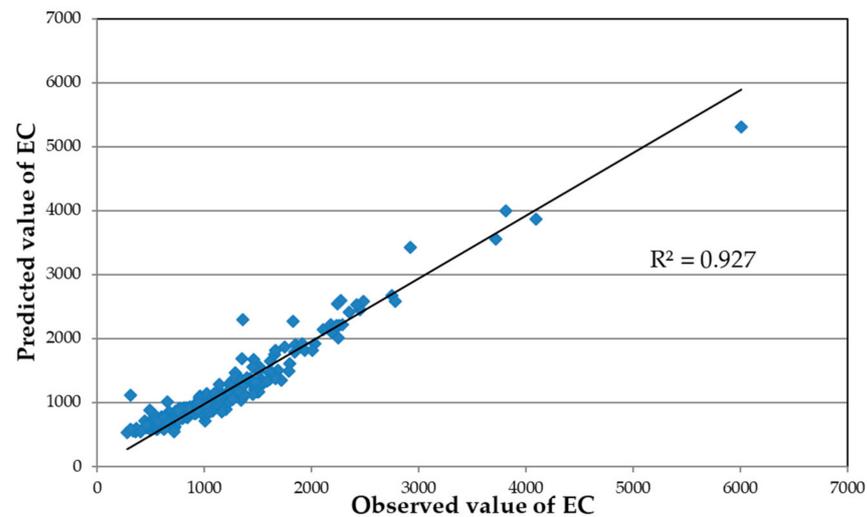


Figure 5. Coefficient of determination of observed versus predicted EC ($\mu\text{S}/\text{cm}$) values using the ANN model.

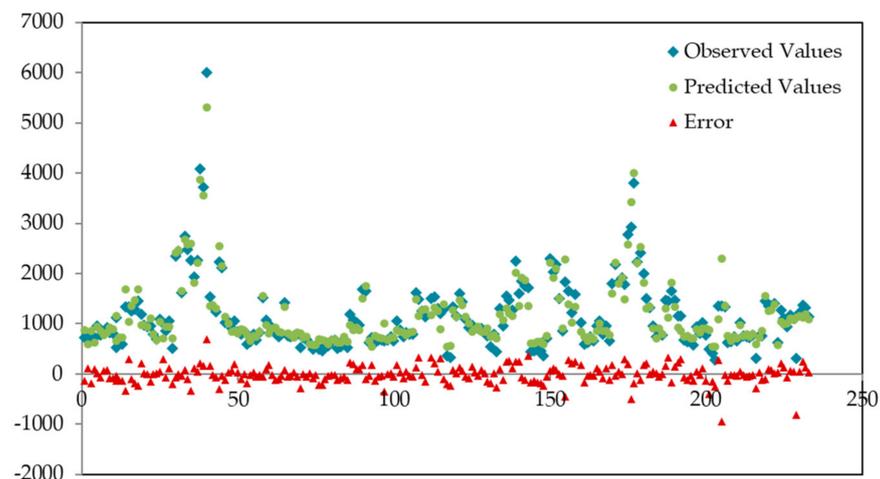


Figure 6. Observed, predicted, and error values of EC ($\mu\text{S}/\text{cm}$) using the ANN model.

In Table 4, the coefficient of determination, efficiency model, mean absolute percentage error, and root mean square error were calculated based on the results of the MLR and ANN models. The values of these indices were considered statistically significant, verifying that forecasting of electrical conductivity using Ca, Mg, Na, and Cl values was valid for the examined data set for both methods. The high R^2 values and the high EF values of both models indicate that they provided a reliable prediction of the EC, along with the small MAPE (%) and RMSE values. In addition, the comparison of the indices values highlights that the performances of ANN and MLR were similar on this large dataset, which included 233 samples taken from sampling points scattered around a large area with different geological conditions and land uses.

Table 4. Evaluation criteria of the EC prediction employing the ANN and MLR.

R^2	EF	MAPE (%)	RMSE
Artificial Neural Networks			
0.927	0.93	14.12	175.9
Multiple Linear Regression			
0.94	0.94	12.15	168

In order to verify the accuracy and reliability of the two methods, a small dataset of 18 samples collected during the wet period of 2006 was used.

The results of the MLR and ANN are depicted in Figures 7 and 8, respectively.

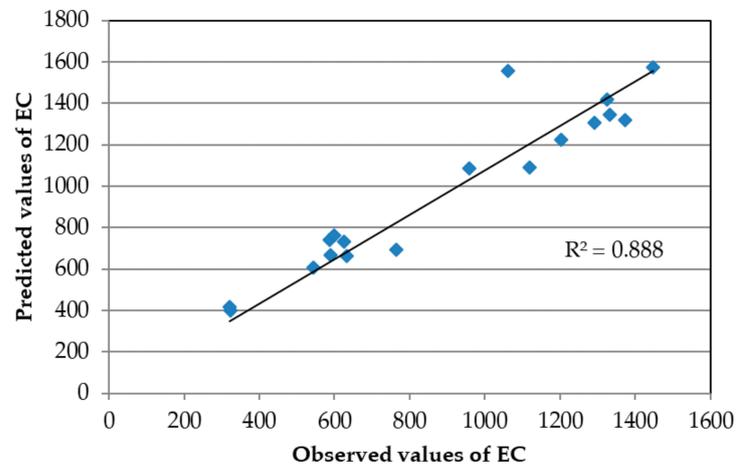


Figure 7. Coefficient of determination of observed versus predicted values of EC using the MLR model.

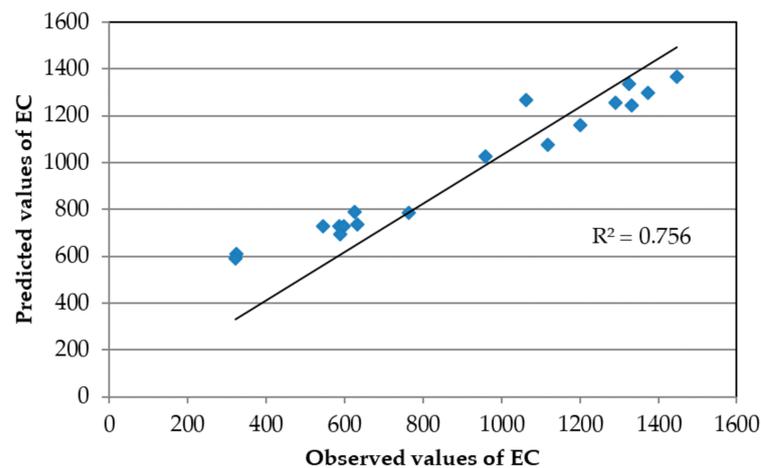


Figure 8. Coefficient of determination of observed versus predicted EC ($\mu\text{S}/\text{cm}$) values using the ANN model.

The evaluation criteria of the models' performance, depicted in Table 5, verify that the predicted values of EC are valid, but for a small set of data, the performance of the MLR than that of the ANN. The dependent variable (EC) was explained better in the MLR model by the independent variables (i.e., Ca, Mg, Na, Cl), since the coefficient of determination was much higher and the mean absolute percentage error was significantly smaller.

Table 5. Evaluation criteria of the EC prediction employing ANN and MLR.

R^2	EF	MAPE (%)	RMSE
Artificial Neural Network			
0.75	0.979	20	138.2
Multiple Linear Regression			
0.88	0.976	13.8	145.8

Forecasting models are very useful tools for water managers and can be used to predict the water quality with respect to changes in hydrological and hydrogeological regimes,

showing better performance than traditional statistical methods [72,73]. With the use of these models, complex data as a result of various natural or human processes are easily transformed into practical and understandable information for scientists, stakeholders, and policy makers involved in water management or even for the general population [74].

5. Conclusions

The aquifers within the boundaries of the GallikosRiver basin have developed in an area with intensive agricultural activities and small-scale enterprises, receiving different types of pollutants. Agriculture determines the economy of the area and, hence, farmer's income, since it constitutes the most important employer. Crop yield and soil quality depend on irrigation water quantity and quality. Therefore, special management practices may be required.

Irrigation water salinity, which is a measure of quality, can be described through electrical conductivity. Artificial neural networks and multiple linear regression are commonly used with great success in the prediction of water parameters due to the fact of their good performance, simplicity, and low data requirements. This was the motivating factor for their application to the present study.

Samples collected during a three-year experimental period (2004–2006) were used for the calibration, validation, and evaluation of the models. The multiple linear regression and artificial neural networks models had similar performances in the case of a large dataset (233 samples). Both models provided reliable results, since all the evaluation indices that were used were statistically valid ($R^2 > 0.927$, $EF > 0.93$, $MAPE (\%) < 14.5$). In the case of implementing the two models on a smaller verification dataset (18 samples), the forecasting ability remained statistically significant ($R^2 > 0.75$, $EF > 0.976$, $MAPE (\%) < 20$) for both, but the MLR method achieved a better performance. Factor analysis is a suitable method for the selection of the input parameters for the MLR and ANN models, based on the evaluation of their accuracy and reliability.

The outcomes of this research in the specific case study area have practical importance, since the in situ measurement of EC is time consuming and costly. According to this study, these measurements could be avoided. The methodology followed in this study could be used as an effective tool for quality parameter forecasting in any other region that faces environmental problems. This study provides the necessary steps and techniques for parameter selection and model performance evaluation.

Author Contributions: Conceptualization, C.M. and P.G.; Methodology, C.M., L.D., P.G. and P.V.; Supervision, C.M.; Writing—original draft, C.M., L.D., P.G. and P.V.; Writing—review & editing, C.M., L.D., P.G. and P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data presented in this article are available through request to the first Author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cay, T.; Uyan, M. Spatial and temporal groundwater level variation geostatistical modeling in the city of Konya, Turkey. *Water Environ. Res.* **2009**, *12*, 2460. [[CrossRef](#)]
2. Ehteshami, M.; Biglarijoo, N.; Salari, M. Assessment and quality classification of water in Karun, Dez and Karkheh rivers. *J. River Eng.* **2014**, *2*, 23–30.
3. Napacho, Z.; Manyele, S. Quality assessment of drinking water in Temeke District (part II): Characterization of chemical parameters. *Afr. J. Environ. Sci. Technol.* **2010**, *4*, 775–789.
4. Kristensen, P.; Whalley, C.; Zal, F.N.N.; Christiansen, T. *European Waters Assessment of Status and Pressures 2018*; European Environment Agency Report; European Environment Agency: Copenhagen, Denmark, 2018.

5. Qaderi, F.E.; Babanezhad, E. Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network. *J. Clean. Prod.* **2017**, *161*, 840–849. [[CrossRef](#)]
6. Chatwell, G.R. *Environmental Water Pollution and Control*; Anmol Publication: New Delhi, India, 1989.
7. Kılıçaslan, Y.; Tuna, G.; Gezer, G.; Gulez, K.; Arkoc, O.; Potirakis, S.M. ANN-based estimation of groundwater quality using a wireless water quality network. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 458329. [[CrossRef](#)]
8. Sappa, G.; Ergul, S.; Ferranti, F. Water quality assessment of carbonate aquifers in southern Latium region, Central Italy: A case study for irrigation and drinking purposes. *Appl. Water Sci.* **2014**, *4*, 115–128. [[CrossRef](#)]
9. Souissi, D.; Msaddek, M.H.; Zouhri, L.; Chenini, I.; May, M.E.; Dlala, M. Mapping groundwater recharge potential zones in arid region using GIS and Landsat approaches, southeast Tunisia. *Hydrol. Sci. J.* **2018**, *63*, 251–268. [[CrossRef](#)]
10. Zanotti, C.; Rotiroti, M.; Fumagalli, L.; Stefania, G.A.; Canonaco, F.; Stefenelli, G.; Prevot, A.S.H.; Leoni, B.; Bonomi, T. Groundwater and surface water quality characterization through positive matrix factorization combined with GIS approach. *Water Res.* **2019**, *159*, 122–134. [[CrossRef](#)] [[PubMed](#)]
11. Siebert, S.; Burke, J.; Faures, J.-M.; Frenken, K.; Hoogeveen, J.; Döll, P.; Portmann, F.T. Groundwater use for irrigation—A global inventory. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1863–1880. [[CrossRef](#)]
12. Khalil, B.M.; Awadallah, A.G.; Karaman, H.; El-Sayed, A. Application of Artificial Neural Networks for the Prediction of Water Quality Variables in the Nile Delta. *J. Water Resour. Prot.* **2012**, *4*, 388–394. [[CrossRef](#)]
13. Ongley, E.D. Water quality management: Design, financing and sustainability considerations-II. Presented at the World Bank's Water Week Conference: Towards a Strategy for Managing Water Quality Management, Washington, DC, USA, 3–4 April 2000.
14. Alberto, W.D.; del Pilar, D.M.; Valeria, A.M.; Fabiana, P.S.; Cecilia, H.A.; de los Angeles, B.M. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, a case study: Suquia river basin (Cordoba–Argentina). *Water Resour.* **2001**, *35*, 2881–2894.
15. Lopez-Chicano, M.; Bouamama, M.; Vallejos, A.; Pulido, B.A. Factors which determine the hydrogeochemical behavior of karstic springs: A case study from the Betic Cordilleras, Spain. *Appl. Geochem.* **2001**, *16*, 1179–1192. [[CrossRef](#)]
16. Pereira, H.G.; Renca, S.; Sataiva, J. A case study on geochemical anomaly identification through principal component analysis supplementary projection. *Appl. Geochem.* **2003**, *18*, 37–44. [[CrossRef](#)]
17. Kourgialas, N.N.; Karatzas, G.P. A national scale flood hazard mapping methodology: The case of Greece—Protection and adaptation policy approaches. *Sci. Total Environ.* **2017**, *601*, 441–452. [[CrossRef](#)] [[PubMed](#)]
18. Falah, F.; Rahmati, O.; Rostami, M.; Ahmadisharaf, E.; Daliakopoulos, I.N.; Pourghasemi, H.R. Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 323–336.
19. Kesikoglu, M.H.; Atasever, U.H.; Dadaser-Celik, F.; Ozkan, C. Performance of ANN, SVM and MLH techniques for land use/cover change detection at Sultan Marshes wetland, Turkey. *Water Sci. Technol.* **2019**, *80*, 466–477. [[CrossRef](#)] [[PubMed](#)]
20. Singha, S.; Pasupuleti, S.; Singha, S.S.; Singh, R.; Kumar, S. Prediction of groundwater quality using efficient machine learning technique. *Chemosphere* **2021**, *276*, 130265. [[CrossRef](#)]
21. Alizadeh, M.J.; Kavianpour, M.R. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Mar. Pollut. Bull.* **2015**, *98*, 171–178. [[CrossRef](#)]
22. Lu, F.; Chen, Z.; Liu, W.; Shao, H. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecol. Eng.* **2016**, *95*, 422–429. [[CrossRef](#)]
23. Nathan, N.S.; Saravanane, R.; Sundararajan, T. Application of ANN and MLR Models on Groundwater Quality Using CWQI at Lawspet, Puducherry in India. *J. Geosci. Environ. Prot.* **2017**, *5*, 99–124. [[CrossRef](#)]
24. Rogers, L.L. Optimal Groundwater Remediation Using Artificial Neural Networks and the Genetic Algorithm. Ph.D. Dissertation, Stanford University, Stanford, CA, USA, 1992.
25. Maier, H.R.; Dandy, G.C. The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* **1996**, *32*, 1013–1022. [[CrossRef](#)]
26. Sandhu, N.; Finch, R. Emulation of DWRDSM using artificial neural networks and estimation of Sacramento river flow from salinity. In *North American Water and Environment Congress & Destructive Water*; ASCE: New York, NY, USA, 1996; pp. 4335–4340.
27. Aziz, A.R.A.; Wong, K.-F.V. A neural-network approach to the determination of aquifer parameters. *Ground Water* **1992**, *30*, 164–166. [[CrossRef](#)]
28. Abd-Elmaboud, M.E.; Abdel-Gawad, H.A.; El-Alfy, K.S.; Ezzeldin, M.M. Estimation of groundwater recharge using simulation-optimization model and cascade forward ANN at East Nile Delta aquifer, Egypt. *J. Hydrol. Reg. Stud.* **2021**, *34*, 100784. [[CrossRef](#)]
29. Tabari MM, R.; Azari, T.; Dehghan, V. A supervised committee neural network for the determination of aquifer parameters: A case study of Katasbes aquifer in Shiraz plain, Iran. *Soft Comput.* **2021**, *25*, 4785–4798. [[CrossRef](#)]
30. Isazadeh, M.; Biazar, S.; Ashrafzadeh, A.; Khanjani, R. Estimation of aquifer qualitative parameters in Guilans plain using gamma test and support vector machine and artificial neural network models. *J. Environ. Sci. Technol.* **2019**, *21*, 1–21.
31. Daliakopoulos, I.; Coulibaly, P.; Tsanis, I. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* **2005**, *309*, 229–240. [[CrossRef](#)]
32. Mohanty, S.; Jha, M.K.; Kumar, A.; Sudheer, K.P. Artificial Neural Network Modeling for Groundwater Level Forecasting in a River Island of Eastern India. *Water Resour. Manag.* **2010**, *24*, 1845–1865. [[CrossRef](#)]

33. Zhu, S.; Hrnjica, B.; Ptak, M.; Choiński, A.; Sivakumar, B. Forecasting of water level in multiple temperate lakes using machine learning models. *J. Hydrol.* **2020**, *585*, 124819. [CrossRef]
34. Hasda, R.; Rahaman, M.F.; Jahan, C.S.; Molla, K.I.; Mazumder, Q.H. Climatic data analysis for groundwater level simulation in drought prone Barind Tract, Bangladesh: Modelling approach using artificial neural network. *Groundw. Sustain. Dev.* **2020**, *10*, 100361. [CrossRef]
35. Diamantopoulou, M.J.; Georgiou, P.E.; Papamichail, D.M. Performance evaluation of artificial neural networks in estimating reference evapotranspiration with minimal meteorological data. *Glob. Nest J.* **2011**, *13*, 18–27.
36. Ghorbani, M.; Aalami, M.; Naghipour, L. Use of artificial neural networks for electrical conductivity modeling in Asi River. *Appl. Water Sci.* **2017**, *7*, 1761–1772. [CrossRef]
37. Mattas, C. Hydrogeological Research of Gallikos Basin. Ph.D. Thesis, Department of Geology, Aristotle University of Thessaloniki, Thessaloniki, Greece, 2009. (In Greek).
38. Mattas, C.; Anagnostopoulou, C.; Venetsanou, P.; Bilas, G.; Lazoglou, G. Evaluation of Extreme Dry and Wet Conditions Using Climate and Hydrological Indices in the Upper Part of the Gallikos River Basin. *Proceedings* **2019**, *7*, 3. [CrossRef]
39. Mercier, J. Sur l'existence de l'âge de deux phases régionales de métamorphisme alpin dans les zones internes des Hellenides en Macédoine Centrale (Grèce). *Ann. Geol. Des. Pays. Hell.* **1966**, *20*, 1–596.
40. Mountrakis, D.; Sapountzis, E.; Kiliass, A.; Elethriadis, G.; Christofides, G. Paleogeographic conditions in the western Pelagonian margin in Greece during the initial rifting of the continental area. *Can. J. Earth Sci.* **1983**, *20*, 1673–1681. [CrossRef]
41. Kockel, F.; Walther, H. Die Strymonlinie als Grenzzwischen Serbo-Mazedonischen und Rilla-Rhodope Massiv in Ost-Mazedonien. *Geol. Jb.* **1965**, *83*, 575–602.
42. Kockel, F.; Mollat, H.; Walther, H. Geologie des Serbo-Mazedonischen Massivs und seines mesozoischen Rahmens (Nordgriechenland). *Geol. Jb.* **1971**, *89*, 529–551.
43. Mattas, C.; Voudouris, K.S.; Panagopoulos, A. Integrated groundwater resources management using the DPSIR approach in a GIS environment context: A case study from the Gallikos river basin, North Greece. *Water* **2014**, *6*, 1043–1068. [CrossRef]
44. Official Government Gazette of the Hellenic Republic n.182/v.2/31-01-2014. Available online: http://www.et.gr/idocs-nph/search/pdfViewerForm.html?args=5C7QrtC22wEc63YDhn5AeXdvtvSoClrL8P4476sndBGZ5MXD0LzQTLf7MGgcO23N88knBzLCmTXKaO6fpVZ6Lx3UnKl3nP8NxdnJ5r9cmWyJWelDvWS_18kAEhATUkJb0x1LIdQ163nV9K--td6SluWb1VLgGCVmt3UIpyOluks9FzsCB7Qtn2t1Lb7YCWPL (accessed on 24 August 2021).
45. Official Government Gazette of the Hellenic Republic n.3282/v.2/19-09-2017. Available online: http://www.et.gr/idocs-nph/search/pdfViewerForm.html?args=5C7QrtC22wEsrjP0JAlxBXdtvSoClrL8J2n_NccUMzjnMRVjyfnPUeJInJ48_97uHrMts-zFzeyCiBSQQOpYnTy36MacmUFCx2ppFvBej56Mmc8Qdb8ZfRjQZnsIAdk8Lv_e6czmhEembNmZCMxLMtUT2iStv6LoLLokjzFTSjwL OlnEGr5bPE7bGxGF96wKi (accessed on 24 August 2021).
46. WHO, (World Health Organization). *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*, 4th ed.; WHO: Geneva, Switzerland, 2017. [CrossRef]
47. Mattas, C.; Papachristou, M.; Soulios, G. Assessment of groundwater quality characteristics from the upper part of Gallikos river basin (N. Greece) using factor and cluster analysis methods. In *Proceedings of the 10th International Hydrogeological Congress of Greece/Thessaloniki, Athens, Greece, 4–6 October 2014*; Voudouris, K., Stamatis, G., Mattas, C., Kaklis, T., Kazakis, N., Eds.; Volume 1, pp. 467–476.
48. Mattas, C.; Soulios, G.; Panagopoulos, A.; Voudouris, K.; Panoras, A. Hydrochemical characteristics of the Gallikos river water, Prefecture of kilkis, Greece. *Glob. Nest J.* **2007**, *9*, 251–259.
49. Clesceri, L.S.; Greenberg, A.E.; Trussell, R.R. (Eds.) *Standard Methods for the Examination of Water and Wastewater*, APHA, AWWA, WPCF, 17th ed.; American Public Health Association: Washington, DC, USA, 1989.
50. Esmaeili, S.; Moghaddam, A.A.; Barzegar, R.; Tziritis, E. Multivariate statistics and hydrogeochemical modeling for source identification of major elements and heavy metals in the groundwater of Qareh-Ziaeddin plain, NW Iran. *Arab. J. Geosci.* **2018**, *11*, 5. [CrossRef]
51. Machiwal, D.; Jha, M.K. Identifying sources of groundwater contamination in a hard-rock aquifer system using multivariate statistical analyses and GIS-based geostatistical modeling techniques. *J. Hydrol. Reg.* **2015**, *4*, 80–110. [CrossRef]
52. Selle, B.; Schwientek, M.; Lischeid, G. Understanding processes governing water quality in catchments using principal component scores. *J. Hydrol.* **2013**, *486*, 31–38. [CrossRef]
53. Han, Y.M.; Du, P.X.; Cao, J.J.; Posmentier, E.S. Multivariate analysis of heavy metal contamination in urban dusts of Xi'an, Central China. *Sci. Total Environ.* **2006**, *355*, 176–186.
54. Rahman, M.S.; Saha, N.; Molla, A.H. Potential ecological risk assessment of heavy metal contamination in sediment and water body around Dhaka export processing zone, Bangladesh. *Environ. Earth Sci.* **2014**, *71*, 2293–2308. [CrossRef]
55. Garrett, J.H. Where and Why Artificial Neural Networks Are Applicable in Civil Engineering. *J. Comput. Civ. Eng.* **1994**, *8*, 129–130. [CrossRef]
56. Haykin, S. *Neural Networks—A Comprehensive Foundations*; Prentice-Hall International: Hoboken, NJ, USA, 1999.
57. Bekas, G.K.; Alexakis, D.E.; Gamvroula, D.E. Forecasting discharge rate and chloride content of karstic spring water by applying the Levenberg–Marquardt algorithm. *Environ. Earth Sci.* **2021**, *80*, 1–12. [CrossRef]
58. Debes, K.; Alexander, K.; Gross, H.M. Transfer Functions in Artificial Neural Networks—A Simulation-Based Tutorial. *Brains Minds Media* **2005**, *2005*, 1–11.

59. Graupe, D. *Principles of Artificial Neural Networks*, 2nd ed.; World Scientific Publishing Co. Inc.: River Edge, NJ, USA, 2007; ISBN 9812706240.
60. Chenini, I.; Khemiri, S. Evaluation of ground water quality using multiple linear regression and structural equation modeling. *Int. J. Environ. Sci. Technol.* **2009**, *6*, 509–519. [[CrossRef](#)]
61. Ali Khan, M.M.; Umar, R.; Baten, M.A.; Lateh, H.; Kamil, A.A. Evaluation of Groundwater Quality Using Linear Regression Model. *J. Appl. Sci. Res.* **2012**, *8*, 251–260.
62. Ghasemi, J.; Saaidpour, S. Quantitative structure–property relationship study of n-octanol–water partition coefficients of some of diverse drugs using multiple linear regression. *Anal. Chim. Acta* **2007**, *604*, 99–106. [[CrossRef](#)] [[PubMed](#)]
63. Batabyal, A.K. Correlation and multiple linear regression analysis of groundwater quality data of Bardhaman District, West Bengal, India. *Int. J. Res. Chem. Environ.* **2014**, *4*, 42–51.
64. Bodrud-Dozaa, B.M.; Islam, A.R.M.T.; Ahmed, F.; Das, S.; Saha, N.; Rahman, M.S. Characterization of groundwater quality using water evaluation indices, multivariate statistics and geostatistics in central Bangladesh. *Water Sci.* **2016**, *30*, 19–40. [[CrossRef](#)]
65. Neill, S.P.; Hashemi, M.R. *Fundamentals of Ocean Renewable Energy: Generating Electricity from the Sea*; Academic Press: Cambridge, MA, USA, 2018.
66. David, M.B.; Del Grosso, S.J.; Hu, X.; Marshall, E.P.; McIsaac, G.F.; Parton, W.J.; Tonitto, C.; Youssef, M.A. Modeling denitrification in a tile-drained, corn and soybean agroecosystem of Illinois, USA. *Biogeochemistry* **2009**, *93*, 7–30. [[CrossRef](#)]
67. Craswell, E. Fertilizers and nitrate pollution of surface and ground water: An increasingly pervasive global problem. *SN Appl. Sci.* **2021**, *3*, 1–24.
68. Cirkel, D.G.; Van Beek, C.G.E.M.; Witte, J.P.M.; Van der Zee, S.E.A.T.M. Sulphate reduction and calcite precipitation in relation to internal eutrophication of groundwater fed alkaline fens. *Biogeochemistry* **2014**, *117*, 375–393. [[CrossRef](#)]
69. Sharma, M.K.; Kumar, M. Sulphate contamination in groundwater and its remediation: An overview. *Environ. Monit. Assess.* **2020**, *192*, 1–10. [[CrossRef](#)] [[PubMed](#)]
70. Suresh, K.R.; Nagesh, M.A. Experimental studies on effect of water and soil quality on crop yield. *Aquat. Procedia* **2015**, *4*, 1235–1242. [[CrossRef](#)]
71. Lekakis, E.H.; Antonopoulos, V.Z. Modeling the effects of different irrigation water salinity on soil water movement, uptake and multicomponent solute transport. *J. Hydrol.* **2015**, *530*, 431–446. [[CrossRef](#)]
72. Oyeboode, O.; Stretch, D. Neural network modeling of hydrological systems: A review of implementation techniques. *Nat. Resour. Modeling* **2019**, *32*, e12189. [[CrossRef](#)]
73. Jain, A.; Kumar, A.M. Hybrid neural network models for hydrologic time series forecasting. *Appl. Soft Comput.* **2007**, *7*, 585–592. [[CrossRef](#)]
74. Gupta, R.; Singh, A.N.; Singhal, A. Application of ANN for water quality index. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 688–693. [[CrossRef](#)]