Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
Bensafiddine & al. / Mor. J. Chem. 8 N°3 (2020) 665-672

Prediction of chemical water quality used for drinking purposes based on artificial neural networks

Ali EL Bilali[a*]***, Taleb Abdeslam***[a] ***, Nouhaila Mazigh***[a] ***,Mohammed Moukhliss***[a]

[a] *University of Hassan II, Casablanca, Faculty of Sciences and Techniques of Mohammedia, Morocco*

** Corresponding author:*
ali1gpee@gmail.com

Received 28 Feb 2020,
Revised 03 May 2020,
Accepted 14 Jun 2020

Abstract

The Groundwater resources generally have a good water quality and can be used for drinking purposes than water surfaces. However, the anthropogenic activities and climate change effects have been degrading the groundwater quality particularly in the arid and semi-arid areas. In addition, the monitoring of water quality in these regions is poor, as it is expensive and faces financial constraints, notably in rural areas. For this problem, we need to develop a new alternative that allows us to predict the water quality easily. Therefore, the solutions of this challenge would be to develop accurate and reliable models that would allow the prediction of chemical parameters commonly, used for evaluating the suitability of water for drinking uses. This study aims to develop Artificial Neural Networks (ANN) models for predicting the Total Dissolved Solid (TDS in mg/l), Total Hardness (TH), sulphate (SO_4^{2-}) mg/l and Chloride (Cl^-) mg/L parameters using Electrical Conductivity (EC), pH and Temperature as input variables. These models were developed based on the 42 samples collected and analyzed from Tanobart Groundwater in Morocco. Among the 42 samples, 30 samples were used for training of the models while the remaining data were used for the validation processes. The results showed that the ANN models are highly accurate for predicting the TDS, TH, Sulphate and of Chloride with coefficients of determination 0.962, 0.993, 0.986 and 0.957 for the TH, TDS, Sulphate and Chloride parameters respectively, for training processes. Also, the results during the calibration revealed a good accuracy for predicting theses parameters. Hence, these models can improve the water quality monitoring in rural areas to assess the chemical suitability of water for drinking purpose with low costs and in a short time.

Keywords*: Artificial Neural Network, Total Hardness, Total Dissolved Solid, Sulphate, Chloride*

1. Introduction

The groundwater resources play a viable role in supplying water for: drinking, industrial and agricultural purposes. In Morocco, as a semi-arid region, the groundwater is considered as a strategic water resource that can be used during drought conditions, notably in rural areas. In addition, the climate change effects combined with socio-economic development is increasing the water demand [1]. On the other hand, the anthropogenic activities and overexploitation of the groundwater across the country have degraded the water quality [2]–[5]. In that situation and to overcome the socio-economic issues, the country adopted a new strategy which includes the construction of new dams, the reuse of wastewater and desalination. Therefore, the monitoring of chemical water quality parameters is required to assess the water suitability for drinking purposes in such areas and to determine further actions to produce the water according to the World Health Organization (WHO) guideline for drinking water quality regulations and standards [6]. However, the basic chemical drinking water quality parameters are the Total Dissolved Solid (TDS), Chloride, Sulphate, Nitrate, Total Alkalinity (TAC), Total Hardness (TH)…. etc. Indeed the TDS is the sum of cations and anions in the water. For water that contains more than 1000mg/L of TDS is considered unpalatable[7], while the limits of the sulphate and TAC according to Moroccan regulations are 400mg/L and 200 meq/L respectively. The sampling protocol and analysis of all chemical parameters is expensive and laborious, which increases the cost of water quality monitoring particularly for the rural areas, as most population rely on wells for domestic water use. For this purpose, the prediction of the chemical parameters using physical parameters such as Electrical Conductivity through reliable and accurate model is considered a great challenge for this study. However, in recent decades, the Artificial Intelligence (AI) techniques have been used as efficient tools to model the complex systems. In water quality, the Artificial Neural Network has been used to predict the Nitrate Groundwater and have demonstrated good accuracy [8]. In some areas of Morocco, the population uses the Groundwater for both drinking and agricultural purposes. In addition, the monitoring of water quality is poor due to the large number of wells used and the cost of laboratory analysis. Therefore, the main objectives of this study are; (1) To calibrate the ANN model to predict the TDS, total Hardness, sulphate and the chloride; (2) in order to improve the water quality monitoring of Tanobart Groundwater in Khemesset, Morocco using the Artificial Neural Network approach through the EC, pH and the Temperature as inputs.

2. Study area

The Groundwater of Tanobart in Khemesset region is used for supplying drinking water production of Maaziz City and agricultural purpose with a capacity of up to 0.2 billion m^3 per year. Compared to other Groundwater resources, the water in this area is characterized by acceptable quality even though it has been fluctuated in some of samples during the yearly monitoring by the River Agency of Bouregreg and Chaouia. However, this Groundwater is alluvial type that is fed by the rivers particularly Tanobart River as illustrated in Figure 1. Almost, all rural population in this area uses the groundwater for drinking, agriculture and livestock purposes. Since the vital role of this water resource in socio-economic development, the monitoring of water quality is highly required, that is a difficult task, due to the costs of analysis and the large number of wells used. This study used three wells for sampling protocol. The figure 1 presents the study area and the well localizations used in this study.

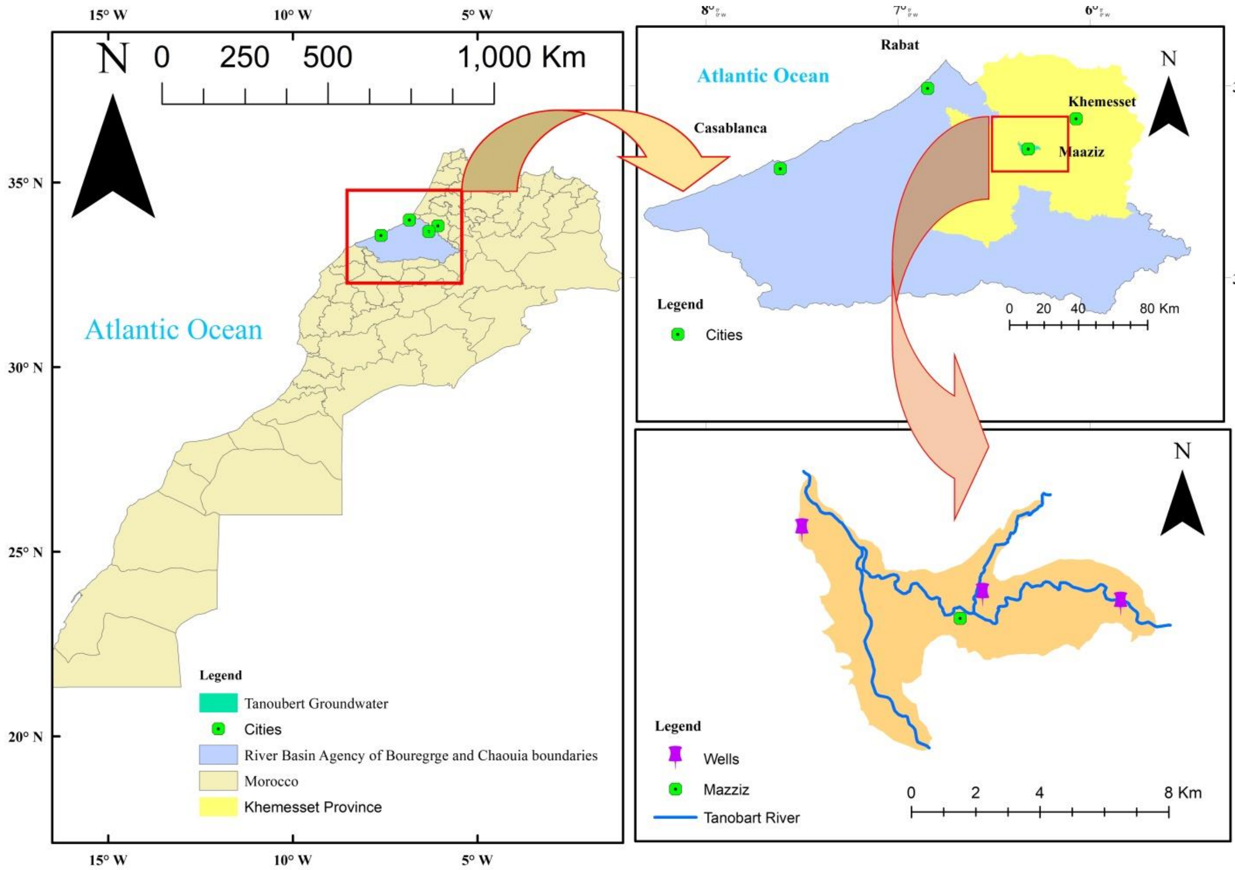


Figure 1: Study area mapping and well localisations

3. Method and Materials

3.1 Artificial Neural Networks

The Multilayer Perceptron Artificial Neural Networks (MLP ANN), as one of the most used ANN models, is a data-based model based on the learning algorithms such as Resilient Back propagation (RP), Levenberg-Marquardt (LM), and Scaled Conjugate Gradient (SCG). Indeed, the ANN model is characterized by three components including weight (w), bias (bi) and the activation function (f). To explain more, for an ANN model with single output, it is calculated by the following equation:

$$y = f\left(\sum_{i=1}^{n} wi * xi + b\right) \qquad (1)$$

Where y is output, wi is weights, xi is input vector (i=1, 2, 3…n), b is bias, f is functioning transfers

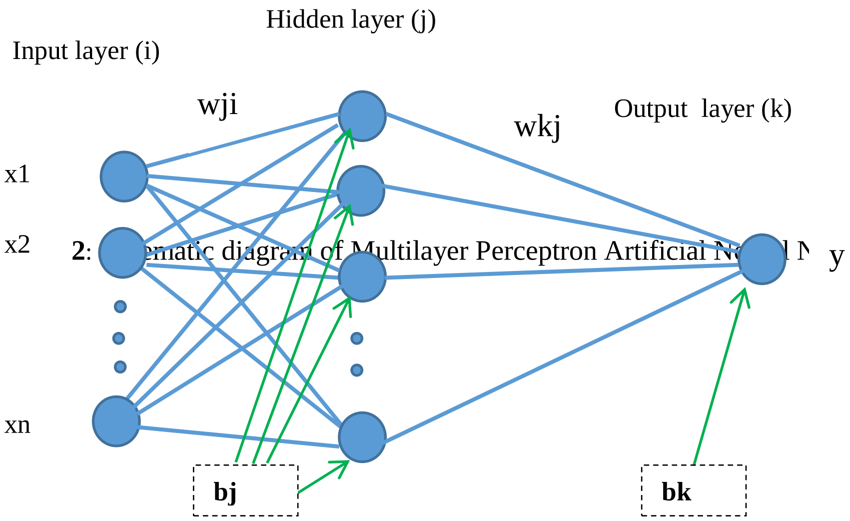
The inputs (xi) were presented to the model through the input layers and are transformed according to transfer function [9] . If the model contains more than one hidden layer, the jth hidden layer produces the values according to equation (1) to (j+1)th hidden layers until the outputs layer. It should be noted that the number of nodes in the hidden layer has an important role in ANN model performance. However, the number of nodes for each hidden layer was determined by trial and error procedures. In addition, the learning process is carried out by changing the weight and bais until the good agreement between the observed and predicted values is verified basing on LM algorithm using MATLAB tool. Then the models developed are used to predict the parameter concerned using the other inputs for validation. The Figure 2 presents the schematic diagram of MLP ANN. The performance of each ANN model was evaluated using root mean square error (RMSE), coefficient of determination R^2 and percent bais (RBAIS). These criteria were calculated for training and validation processes as follows:

$$RMSE = \sqrt{\frac{\sum (Si - Oi)^2}{n}} \qquad (2)$$

$$PBIAS = \frac{\sum_{i=1}^{n}(Oi - Si).100}{\sum_{i=1}^{n} Oi} \qquad (3)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(Oi - \tilde{O})(Si - \hat{S})}{[\sum_{i=1}^{n}(Oi - \tilde{O})^2 \sum_{i=1}^{n}(Si - \hat{S})^2]^{0.5}}\right)^2 \qquad (4)$$

Where Oi and Si are observed and simulated values respectively; Õ is represents the mean of observed values and n=number of values considered.



2: ematic diagram of Multilayer Perceptron Artificial Ne l N

3.2 Methods

To conduct this study, 42 samples were recorded during the period from 2007 to 2018 at three wells as presented in the figure 1. The physicochemical parameters analyzed are: Electrical Conductivity (EC), pH, Temperature (T°), Calcium Ca^{2+}, Magnesium Mg^{2+}, Bicarbonate HCO_3^-, Carbonate CO_3^{2-}, Sulphate (SO_4^{2-}), Chloride (Cl^-), Sodium Na^+, Potassium K^+ and Nitrate NO_3^-. Basically, the physical parameters were analyzed in situ while the chemical parameters were analyzed in Laboratory using the Titration method for Ca^{2+}, Mg^{2+}, Cl^-, H_2CO_3, HCO_3^-, CO_3^{2-}, the flame Photometer for Na^+ and K^+ and UV Spectrophotometer for, NO_3^- and SO_4^{2-}. Moreover the evaluation of errors balance (EB) between positive and negative ions was carried out to control the data quality by equation (5). However, the descriptive statistic of these parameters is presented in the Table 1.

$$EB = \frac{\sum Cation - anions}{\sum Cations + anions} * 100 \qquad (5)$$

The Total Hardness in French degree (°F) was calculated using Ca^{2+} and Mg^{2+} by equation (6).

$$TH = \left(\frac{Ca2+}{4} + \frac{Mg2+}{2.4}\right) \qquad (6)$$

The Total Dissolved Solid was calculated by equation (7).

$$TDS = \sum Ions \qquad (7)$$

Where all ion concentrations in mg/l while TH is in French degree. The data is divided into two types namely: a training data set used for development of the models and a validation data set used to validate the models developed. Then, the criteria performances of each model were evaluated.

Table 1: Descriptive statistic of data used

	Cl^- (mg/l)	NO_3^- (mg/l)	HCO_3^- (mg/l)	SO_4^{2-} (mg/l)	NH_4^+ (mg/l)	Na^+ (mg/l)	K^+ (mg/l)	Ca^{2+} (mg/l)	Mg^{2+} (mg/l)	T°C	pH	EC *(µs/cm)*
max	463.90	75.90	495.87	697.30	0.30	268.00	6.90	272.50	126.36	25.20	7.90	2900.00
min	112.00	4.80	107.00	50.00	0.00	74.21	1.76	36.10	28.07	17.00	6.91	1020.00
mean	243.43	34.32	344.56	271.75	0.06	142.44	3.09	145.33	70.26	21.35	7.29	1678.92
SD	111.13	16.35	74.70	173.04	0.07	54.32	1.06	64.38	27.00	1.92	0.25	535.19
skew	0.67	0.46	-0.79	0.66	2.04	1.12	1.87	0.01	0.47	0.00	0.76	0.80
kurt	-1.16	-0.39	1.50	-0.32	4.58	0.19	4.16	-0.60	-0.92	-0.30	-0.23	-0.52

4. Results and Discussion

Firstly, a correlation analysis of inputs and outputs was carried out to have information about relationships between the parameters. As table 2 presents, the results show that the electrical conductivity (EC) is highly correlated with TH, TDS, sulphate and the chloride. Such result indicates that the EC is a key parameter to predict the chemical parameters studied. In addition, the pH and Temperature are poorly correlated with the outputs. Although these poor correlations, the Temperature and pH are selected as inputs, as they affect the conductivity and the solubility of some parameters.

Table 2: Correlation analysis of input and output parameters

	Te(°C)	*pH*	*EC(µs/cm)*	*TH (F°)*	*TDS(mg/L)*	*SO_4^{2-}(mg/L)*	*Cl^-(mg/l)*
Te(°C)	1.00						
pH	-0.37	1.00					
EC*(µs/cm)*	-0.04	-0.50	1.00				
TH (F°)	-0.06	-0.59	0.90	1.00			
TDS (mg/L)	-0.07	-0.56	0.95	0.97	1.00		
SO_4^{2-}(mg/L)	-0.07	-0.56	0.91	0.92	0.96	1.00	
Cl^-(mg/l)	-0.18	-0.42	0.87	0.77	0.84	0.75	1.00

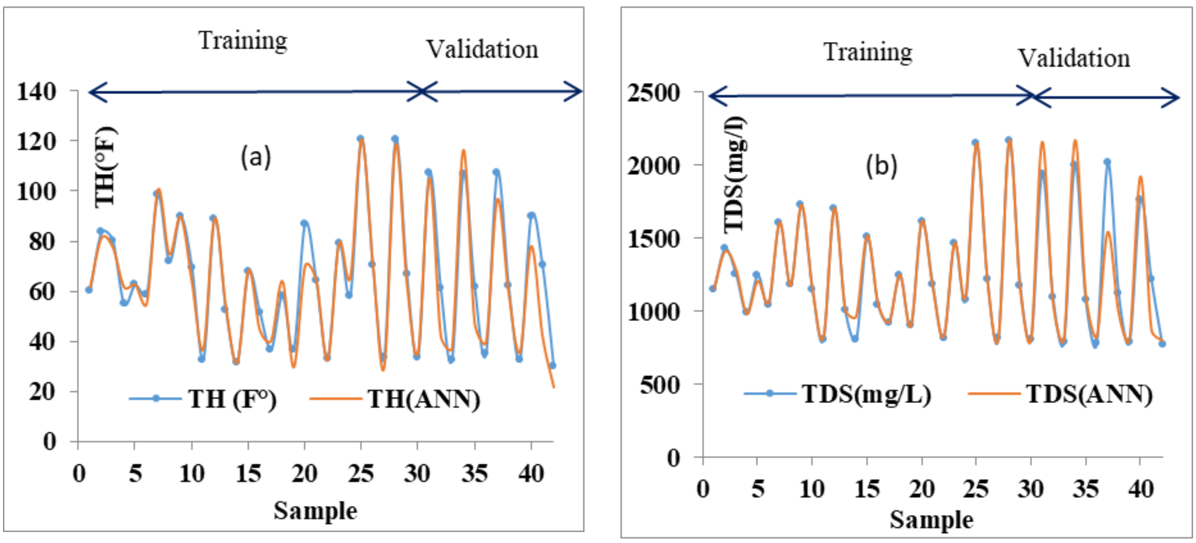
After data processing, the ANN models were developed to predict the TH, TDS, sulphate and Chloride parameters. The structure architectures of the best models trained and validated are (3-4-1 *with Logsig function*), (3-3-3-1 *with logsig function*), (3-5-1 *with logsig function*) and (3-43-1 *with tansig function*) for predicting the TH, TDS, sulphate and Chloride respectively. The Table 3 presents the performance results of training and validation process of the model developed. According to these results, it was observed that all ANN models have good prediction accuracy with coefficients of determination R^2, 0.962, 0.993, 0.986 and 0.957 for the TH, TDS, Sulphate and Chloride parameters respectively during the training process. On the other hand, these coefficients are 0.874, 0.858, 0.810 and 0.869 for TH, TDS, SO_4^{2-} and Cl^- respectively for validation process. Such results demonstrate that the ANN models predict these parameters well in terms of trends. Moreover, for the TH parameter, the ANN model (3-4-1) presented RMSE, 3.15 and 12.44 with data range of [21,120]°F for training and validation process respectively. Also, the results showed that this model overestimates the TH with RBAIS about 1% and 10% for training and validation respectively. These results revealed that the ANN model can predict the TH parameter with acceptable accuracy. Also, for the TDS parameter, the ANN model has good performance criteria in terms of errors where it presented an RMSE 9.67mg/l and 198.47 mg/L for a range of [797, 2170] mg/l and RBAIS 0.8% and 2% for training and validation respectively.

Therefore, the TDS is easily predictable by ANN model with a good accuracy. For the Sulphate and Chloride parameters, the ANN models are also accurate with lower RBAIS indicating the good agreement between observed and predicted values. With regard to the Nitrate and the Total alkalinity (TAC) parameters, the ANN models presented poor performances with coefficients of correlation negatives. Therefore, for this parameter other models should be investigated. For this purpose, the Machine Learning Algorithms are suggested for future works[10].

Table 3: Criteria Performances of the models for training and validation processes

Criteria	TH(°F)		TDS(mg/L)		SO_4^{2-}(mg/l)		Cl^-(mg/l)	
	Training	Validation	Training	validation	Training	Validation	Training	validation
R^2	0.962	0.874	0.993	0.858	0.986	0.810	0.957	0.869
RMSE	3.15	12.44	9.67	198.47	0.77	94.63	8.24	43.94
RBAIS	1%	10%	0.8%	2%	0.7%	2%	1%	-1%

The Figure 3 presents the graphical comparisons of observed and predicted values during the training and validation processes. From the graph (a), it can be observed that, except three values, all other values of the observed TH are closer to the predicted values. The graph (b) shows the agreement between the observed and predicted values of TDS except one extreme value during the validation process. Whereas, the graph (c) and (d) illustrate the high accuracy of ANN approach to predict the sulphate and Chloride respectively except two extreme values. The main limitation of this study is that the models developed can predict only four chemical parameters, but the evaluation of suitability of water quality for drinking purposes requires more than four parameters notably biological parameters. However, even though this limitation, this work can improve the monitoring of water resources quality for drinking purposes particularly in rural areas. This work could be valued by decision makers by using these models in implementation of the Internet of Things –IOT to survey the water quality. These techniques of IOT require implementation of equipment enable to measure the Electrical Conductivity, pH and temperature parameters as main inputs of the models. Then the information about chemical water quality is exchanged and communicated through the internet. In addition, the monitoring frequency of water quality in this area is half-yearly. Consequently, measurement of the electrical conductivity, pH and temperature and applying the ANN models provide the information about chemical water quality with lower cost and in a short time and with acceptable uncertainty.



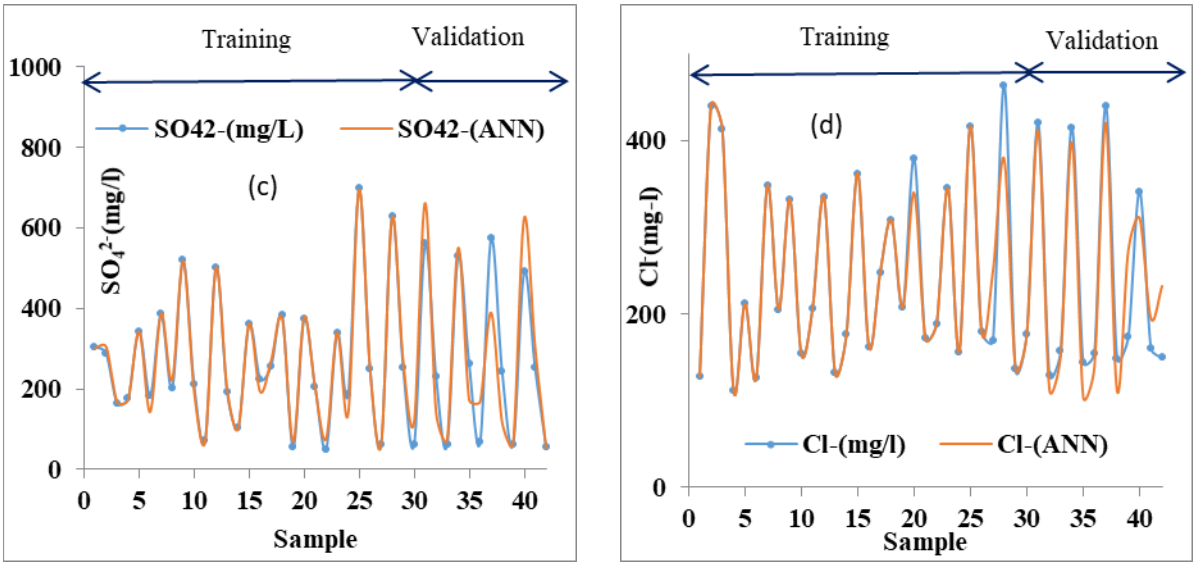


Figure 3: Model training and validation results; (a) for Total Hrdness, (b) for Total Dissolved Solid (TDS), (c) for Sulphate and (d) for Chloride

5. Conclusion

In this study a ANN approach is applied to predict the chemical parameters of Groundwater used for evaluating the suitability for drinking purposes in rural areas. The results showed high performance of the ANN models to predict the Total Hardness (TH), Total Dissolved Solid (TDS), Sulphate and the Chloride with R^2, 0.962, 0.993, 0.986 and 0.957 and, 0.874, 0.858, 0.810 and 0.869 for training and testing respectively. Also, the RBAIS was less than 2% except the Total Hardness where the RBAIS was 10%. Such results demonstrate that the implementation of ANN models by Groundwater quality managers can improve the monitoring using EC, the pH and Temperature that could be measured easily and in a short time.

Acknowledgements

This work has been supported by the River Basin Agency of Bouregreg and Chaouia by providing the data pertaining to the Tanobart Groundwater. The authors thank the team members of the Agency.

References

[1] P. Droogers *et al.*, “Water resources trends in Middle East and North Africa towards 2050,” *Hydrol. Earth Syst. Sci.*, vol. 16, no. 9, pp. 3101–3114, 2012.

[2] G. Busico *et al.*, “A novel hybrid method of specific vulnerability to anthropogenic pollution using multivariate statistical and regression analyses,” *Water Res.*, vol. 171, p. 115386, 2020, doi: 10.1016/j.watres.2019.115386.

[3] B. Alabjah, F. Amraoui, M. Chibout, and M. Slimani, *Assessment of saltwater contamination extent in the coastal aquifers of Chaouia (Morocco) using the electric recognition*, vol. 566. 2018.

[4] S. Najib, A. Fadili, K. Mehdi, J. Riss, A. Makan, and H. Guessir, “Salinization process and coastal groundwater quality in Chaouia, Morocco,” *J. African Earth Sci.*, vol. 115, pp. 17–31, 2016, doi: 10.1016/j.jafrearsci.2015.12.010.

[5] S. Najib, A. Fadili, K. Mehdi, J. Riss, and A. Makan, “Contribution of hydrochemical and geoelectrical approaches to investigate salinization process and seawater intrusion in the coastal aquifers of Chaouia, Morocco,” *J. Contam. Hydrol.*, vol. 198, pp. 24–36, 2017, doi: 10.1016/j.jconhyd.2017.01.003.

[6] M. G. Biyoune, B. Bouargane, H. Bari, and A. Marrouche, “Water quality depends on remineralization ’ s

method in the desalination plant," vol. 10, no. 2, pp. 162–170, 2020.
[7] W. H. O. (WHO) and W. H. O. Staff, *Guidelines for drinking-water quality*, vol. 1. World Health Organization, 2004.
[8] V. Wagh, D. Panaskar, A. Muley, S. Mukate, and S. Gaikwad, "Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India," *Groundw. Sustain. Dev.*, vol. 7, pp. 436–445, 2018, doi: 10.1016/j.gsd.2017.12.012.
[9] S. Haykin, "Neural Networks, a comprehensive foundation, Prentice-Hall Inc," *Up. Saddle River*, *New Jersey*, vol. 7458, pp. 161–175, 1999.
[10] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, and R. Irfan, "E ffi cient Water Quality Prediction Using Supervised," pp. 1–14, 2019.