ISSN 0097-8078, Water Resources, 2022, Vol. 49, No. 3, pp. 503–514. © Pleiades Publishing, Ltd., 2022.

HYDROCHEMISTRY, HYDROBIOLOGY: ENVIRONMENTAL ASPECTS

Use of GIS, Statistics and Machine Learning for Groundwater Quality Management: Application to Nitrate Contamination

Tariq Judeh[a, *], Mohammad N. Almasri[b], Sameer M. Shadeed[c], Hanbing Bian[a], and Isam Shahrour[a]

[a] *Laboratoire de Génie Civil et Géo-Environnement (LGCgE), Lille University, Lille, 5900 France*

[b] *College of Engineering, Department of Civil Engineering, An-Najah National University, Nablus, Palestine*

[c] *Water and Environmental Studies Institute, An-Najah National University, Nablus, Palestine*

**e-mail: tariq.judeh.etu@univ-lille.fr*

Received August 10, 2021; revised October 21, 2021; accepted October 28, 2021

Abstract—Groundwater NO_3 contamination (GNC) threatens the drinkability of water in many countries worldwide. It could cause serious health problem and sometimes lead to death. This paper aims to introduce a comprehensive approach that combines GIS, statistics and Machine Learning (ML) for the groundwater quality management including both water quality assessment and prediction. The performances of this approaches are discussed through its application on assessing and predicting nitrate (NO_3) concentrations in the Eocene Aquifer, Palestine. Spatiotemporal records of NO_3 over the period 1982–2019 are integrated in a database and used in this research. The database includes the following factors: well depth, well use, anthropogenic on-ground activities, watersheds, soil type and land use. Geo-statistical assessment using GIS and statistical boxplot is employed to assess the variability of NO_3 concentrations and how they affected by the independent indicators. Assessment outcomes (NO_3 distribution and the influencing factors) were used to build the Random Forest (RF) prediction model. Such model is used to predict GNC level in groundwater based on multi-influencing factors. Assessment results indicate increasing and decreasing trends of GNC in the southern and middle parts of the study area, respectively. It also provides the RF model by the main influencing factors affecting GNC in the study area which are: well depth, well use, anthropogenic on-ground activities, watersheds and land use. Results indicate that RF has an average and maximum prediction accuracy of 88.5 and 91.7%, respectively. The well depth has the highest influence on GNC. This research could support water authority decision-makers toward the adoption of sustainable groundwater protection plans in Palestine.

Keywords: GIS, machine learning, groundwater, quality, assessment, nitrate, West Bank

DOI: 10.1134/S0097807822030162

INTRODUCTION

Groundwater nitrate contamination (**GNC**) is a common problem threatens the drinkability of groundwater worldwide [26]. The severity of the GNC depends on the importance of the groundwater as a main source of water for different uses among which, the domestic use is the most challenging [1, 2, 20, 28, 36]. GNC severity also depends on the nitrate (NO_3) concentration with reference to the maximum contaminant level (**MCL**). The MCL as set by the World Health Organization (**WHO**) is 50 mg/L as NO_3 [48]. The use of water beyond the MCL for drinking has been linked to methemoglobinemia and cancers [3, 9].

Groundwater quality management has become of great importance in all groundwater-dependent countries [47]. Given the widespread and riskiness of the GNC, it is considered as a fundamental aspect when adopting groundwater quality management strategies [47]. The efficient management of GNC requires both: GNC assessment and GNC prediction [4, 14, 21].

Assessment of GNC constitutes a substantial step to develop an efficient management system to control GNC [4, 14]. Different approaches were proposed to assess the GNC such as: descriptive statistical assessment [4], GIS-based groundwater NO_3 mapping [10], lumped-parameter models [14], Geodetector-Based Frequency Ratio [41], Optimized-DRASTIC Methods [41], parametric IPNOA [32], data-driven logistic regression models [32], water quality index [6], and groundwater potability index [17]. Groundwater quality strategies should rely on robust and accurate models for the prediction of groundwater contamination [21]. The artificial intelligence and machine learning (**ML**) methods have been successfully used as predictive tools for the GNC [7, 13]. Researchers adopted an algorithm to obtain an average of multiple predictions to decrease the bias and to improve the prediction accuracy [7, 13]. Logistic regression, random forest

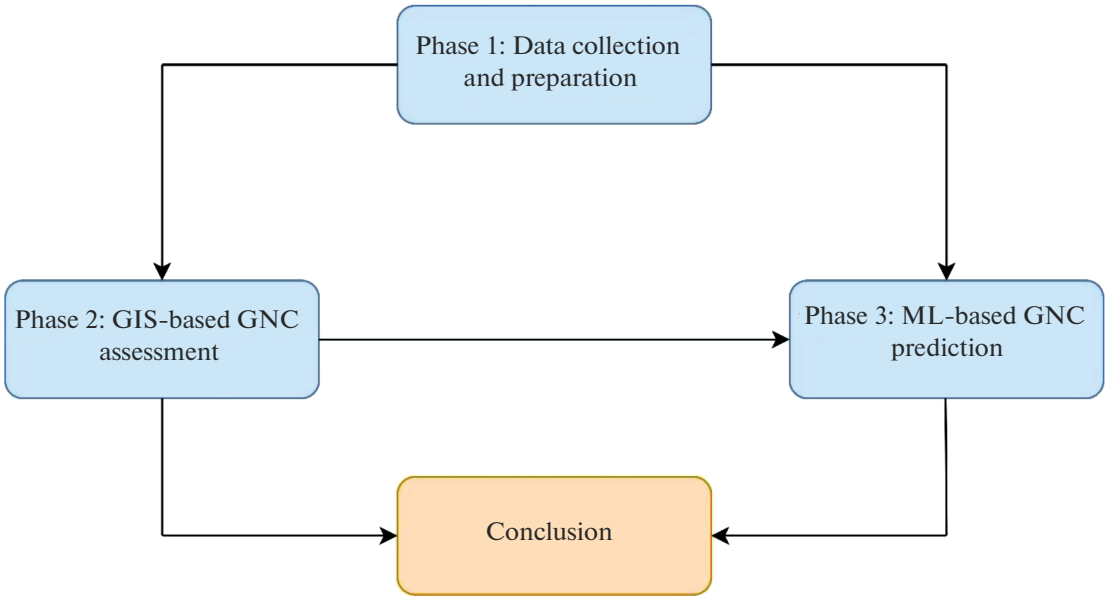


Fig. 1. Methodological framework.

(**RF**), decision tree, multiple linear regression, boosted regression trees, support vector machine, cubist, and Bayesian artificial neural network are commonly used predictive models for the GNC [5, 8, 21, 24, 31, 45]. The prediction capability, efficiency and accuracy of ML methods are dependent on the complexity of the case under consideration and quality of the available data. Therefore, it is important to conduct benchmarking analysis before the selection of ML method [5]. However, the comprehensive integration of spatiotemporal assessment and spatiotemporal prediction of GNC is still missing. This research will overcome the missing issue by integrating the GIS-based geospatial analysis and the descriptive statistical analysis for the development of spatiotemporal GNC assessment. Furthermore, assessment results will be used as input (GNC characteristics and its main influencing factors) for building the ML GNC prediction model.

Spatiotemporal data is the core input of GNC assessment and prediction models [4]. Given its ability in processing, analyzing, interpolating and visualizing spatiotemporal data, GIS has a key role in the spatial assessment and prediction of groundwater quality [25]. It is widely adopted by GNC researches in the United States [35], Germany [33], Iran [40] and China [25].

Palestinians are facing two main water challenges: the low level of available groundwater and the deterioration of its quality [11, 18, 46]. The elevated GNC in Palestine is considered as an indicator of the quality deterioration. A recent field study of the GNC in Palestine showed that the NO_3 concentration in 91% of the groundwater samples exceeded the MCL [30]. However, there are many aquifers in Palestine that are not well characterized for recent occurrences of NO_3. Amongst them is the Eocene Aquifer.

This paper proposes an unprecedented approach that combines GIS, statistical analysis and ML prediction models for the comprehensive management of the GNC. Furthermore, the application of this approach in a groundwater dependent area (Eocene Aquifer, Palestine) forms an existential issue. Groundwater resources account for 95% of the gross water supply in the study area [18, 28]. This research will support the decision-makers of the Palestinian water sector toward the sustainable water resources management.

MATERIALS AND METHODS

Methodology

In order to outline the comprehensive management of GNC, this paper introduces an unparalleled methodology. It consists of three phases: data collection and preparation, GIS-based GNC assessment, and ML-based GNC prediction (Fig. 1).

Data Collection

This phase involves the collection of spatiotemporal NO_3 records for the groundwater wells in the study area. It also involves a gathering of spatial information concerning well uses, well depths, soil types, land uses, surrounding anthropogenic activities, and watersheds.

GIS-based GNC Assessment

GIS and statistical descriptive analysis treat the collected records to carry out a geo-statistical assessment of GNC. GIS has the powerful of considering the spatial variations concerning GNC and considering the factors causing such variations. Statistical descriptive analysis enables the frequency detection of any abnormal event (e.g. NO_3 above MCL). Boxplots of the NO_3 concentration (e.g. minimum, 1st quartile (Q1), median, mean, 3rd quartile (Q3), maximum) were used to present the assessment results. Factors affecting GNC are specified through the examination of various factors such as land use, soil type, watershed, groundwater flow direction, well use, well depth and etc.

Kriging Interpolation Method (**KIM**) employs assessment results (e.g. spatiotemporal mean GNC records) to develop GIS-based spatiotemporal GNC maps. KIM has the powerful of considering both spatial autocorrelation and statistical models [17, 23]. It is highly effective in cases where quality records are biased and spatially correlated [22]. Therefore, researchers confirm its suitability for groundwater contamination and hydrogeology researches [23].

Ml-based GNC Prediction

This research employs RF for predicting the probability to exceed the globally stated NO_3 MCL thresholds. RF is selected among other ML methods due to its ability and collectivity in integrating multiple decision tree algorithms [34]. It has the ability to generate repeated predictions of the same phenomenon [34]. Furthermore, it can quantify the relative importance of the input influencing factors [7].

RF model is built through the online platform Kaggle. GNC distribution and its influencing factors resulting from GNC assessment are used as inputs for RF prediction model. These factors include both natural and man-made ones. Accordingly, a CSV file includes both assessment results (influencing factors) and the NO_3 records from the different groundwater wells is executed using Kaggle platform. After that, data analysis, visualization and prediction were coded on the platform using Python Script. The coded prediction model is trained using randomly picked 80% of the database. The remaining 20% are used to test the prediction accuracy.

Study Area

The Eocene Aquifer is an unconfined aquifer located in the northern part of the West Bank, Palestine. It has a total surface area of 430 km^2 covering areas from Jenin, Nablus and Tubas governorates (Fig. 2) [19, 42]. It serves 36 communities with a total population of about 225000 capita [27]. The groundwater flows from the south to the north and northeast [44]. Rainfall plays a major role in recharging the aquifer. The annual rainfall varies between 400 and 600 mm and mainly occurs in winter [37, 39]. The West Bank climate is classified as hot and dry during the summer and cool and wet in winter [38].

The study area is characterized by intensive agricultural activities with an extensive use of agrochemicals [3]. It includes about 17,000 hectares of irrigable areas including 6,500 hectares of irrigated area [16]. The aquifer was exploited by 129 wells including 54 active ones (45 for agricultural and nine for domestic uses) [16]. 43 out of the 54 active wells are distributed among 12 communities. Table 1 summarizes the statistical descriptive parameters of the groundwater NO_3 in the 12 communities. Arrana, Jenin city and Ras Al-Far'a (54% of the population) have mean and median NO_3 concentrations exceeding the MCL. Furthermore, seven out of the 12 communities (82% of the population) use groundwater wells that have a maximum NO_3 concentration exceeding the MCL.

The ground surface elevation in the study area ranges from 100 m above mean sea level (mamsl) in the north to 925 mamsl in the south. The geologic cross section for the aquifer shows the following formations: limestone, dolomite and marl of Cenomanian to Turonian age, chalk and chert of Senonian age, chalk, limestone and chert of Eocene age and alluvium of Pleistocene to Recent age. However, the Eocene Aquifer overlies the Upper Cenomanian-Turnoian Aquifer, with a transition zone of chalk and chert that varies in thickness between 0 to 480 m [42] (Fig. 3).

Data Collection and Application to the Eocene Aquifer

Two datasets are used in this research: (i) Palestinian Water Authority (**PWA**) dataset for 27 wells over the period 1982–2012 [29] and (ii) field dataset for 24 wells which were sampled during the period 2017–2019. Eight mutual wells have recorded in the datasets which have both historical (1982–2012) and recent (2017–2019) NO_3 concentration values. However, the two datasets were combined into a one composite GIS-based data, which includes data for 43 wells. It includes the following indicators: well depths, coordinates of spatial location (X, Y), type (use), on-ground activities, NO_3 concentration and sampling date. The collected data was manipulated under the GIS environment where wells were spatially joined by communities, watersheds, soil types and land use shapefiles.

Accordingly, GNC assessment and mapping in the study area is conducted. Considering RF model, four levels of NO_3 contamination are identified (Table 2). Levels are selected considering both: WHO guidelines, and the natural breaks of the continuous NO_3 records in the database [48].

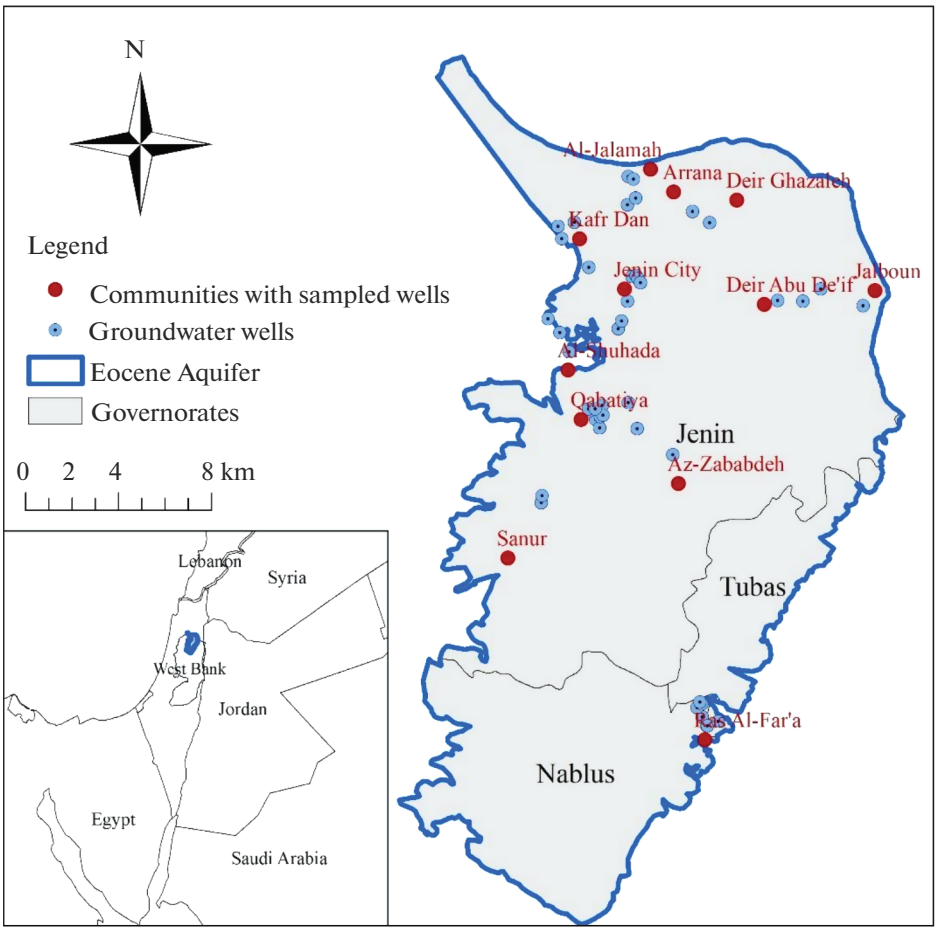


Fig. 2. The geographic setting of the Eocene Aquifer.

Table 1. Descriptive statistics of the NO_3 concentrations (mg/L) in different communities in the Eocene Aquifer (period 1982–2019)

Communities	Population in 2019	No. of samples	Mean	Median	Min	Max
Al-Jalamah	2343	41	48	50	11	64
Arrana	2498	8	57	59	32	71
Al-Shuhada	2375	4	4	4	2	5
Az-Zababdeh	4402	7	19	19	10	28
Deir Abu De'if	7278	8	11	11	1	22
Deir Ghazaleh	1166	39	45	46	7	80
Jalboun	2906	8	9	6	2	25
Jenin city	51 557	150	70	53	0	216
Kafr Dan	6809	63	47	45	6	85
Qabatiya	25 247	84	39	19	0	192
Ras Al-Far'a	1323	103	66	58	6	256
Sanur	5202	7	12	11	8	19

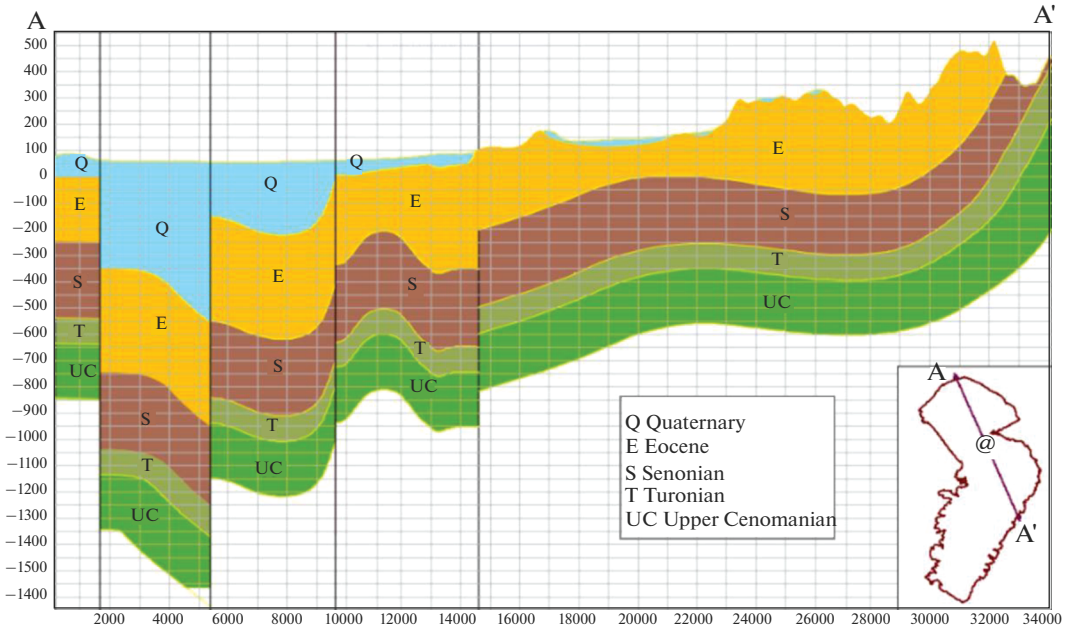


Fig. 3. Geologic cross section of the Eocene Aquifer.

RESULTS AND DISCUSSION

Assessment of GNC: Results and Discussion

Spatiotemporal Nitrate Distribution Across the Aquifer

Figure 4 shows the key statistics of the annual NO_3 concentrations for the period 1982–2019. It is noticed that the mean values are mostly above the MCL. Median values are always below the mean and close to the MCL. This observation indicates that approximately 50% of the NO_3 concentrations are above the MCL. The maximum values of the annual NO_3 concentrations are always higher than the MCL, approaching its up-limit value of 256 mg/L in 2012. Such concentrations clearly reflect the high groundwater contamination through the mean, median, 75th percentile and maximum NO_3 concentrations. However, Fig. 4 considers only the temporal GNC and neglects the spatial variation.

Figure 5 shows the spatiotemporal distribution of GNC in the Eocene Aquifer between 1982 and 2019. A significant increase in the GNC is noticed in the southern part of the aquifer. This is due to the intensive application of agrochemicals in this part, which provides citizens in the study area by the needed fruits and vegetables. On the other hand, the middle part shows a decreasing GNC trend. This is due to the construction and rehabilitation of sewer network in Jenin city and its suburbs.

Influence of the Soil type

The soil map of the Eocene Aquifer area contains five types of soil: Terra Rossa, Mediterranean Brown Forest, Alluvial, Colluvial-Alluvial and Brown Alluvial covering 48, 16, 14, 13 and 7% of the study area, respectively [16]. Figure 6 illustrates the minimum, 25th percentile, mean, median, 75th percentile and maximum NO_3 concentrations for the different soil types. It indicates that the mean NO_3 concentrations in the groundwater under all soil types are equal or exceed the MCL. Since the different soil types in the study area have a similar permeabilities, they have nearly the same mean, median and 75th percentile groundwater NO_3 concentrations. This result indicates a low influence of the soil type on the groundwater NO_3 concentration.

Influence of the Land Use

Table 3 summarizes the key statistics of NO_3 concentrations for the different land use classes. It shows that the groundwater under the discontinuous urban areas has the highest mean NO_3 concentration with a value of 85 mg/L. This can be attributed to the extensive use of cesspits for wastewater (**WW**) disposal. Furthermore, fertilizers are used in the green spaces dis-

Table 2. Levels of GNC used in ML methods

Nitrate Level, mg/L	Description
<25	Not contaminated by NO_3
25–50	Reasonable NO_3 concentration
50–100	Contaminated by NO_3
>100	Extremely contaminated by NO_3

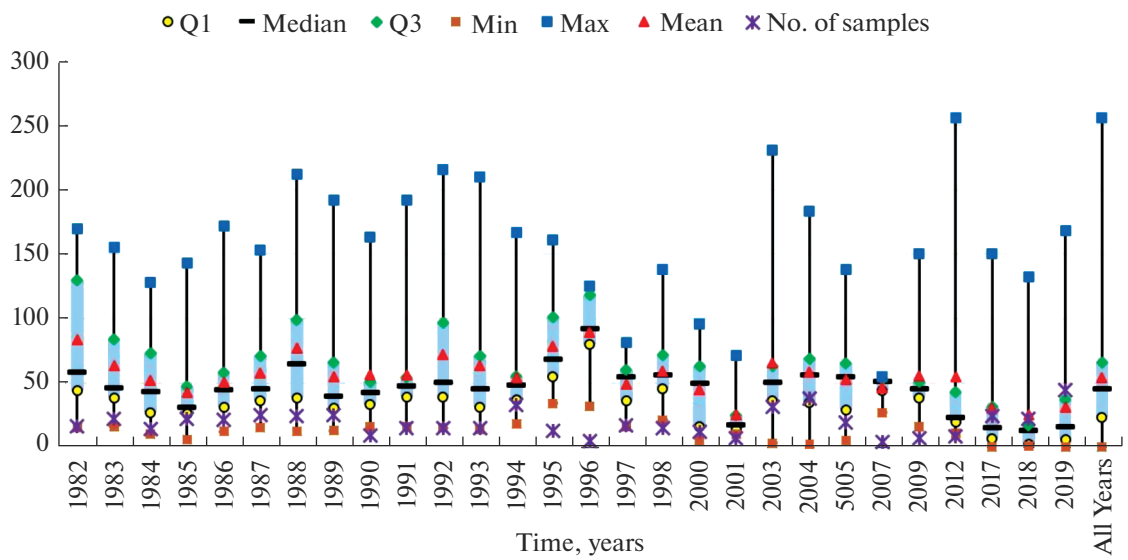


Fig. 4. Annual NO_3 concentration statistics in the Eocene Aquifer (Period 1982–2019).

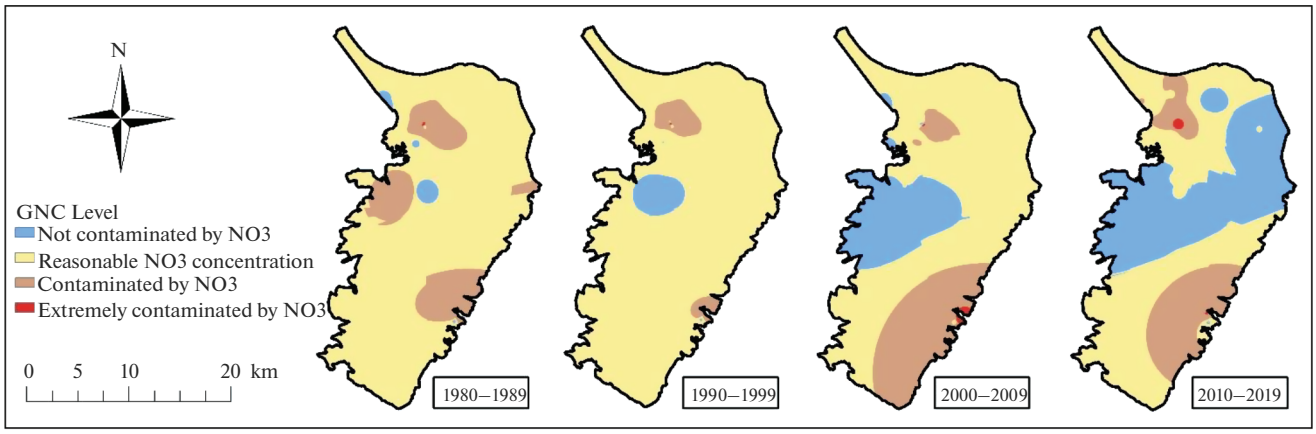


Fig. 5. Spatiotemporal distribution of GNC in the Eocene Aquifer between 1982 and 2019.

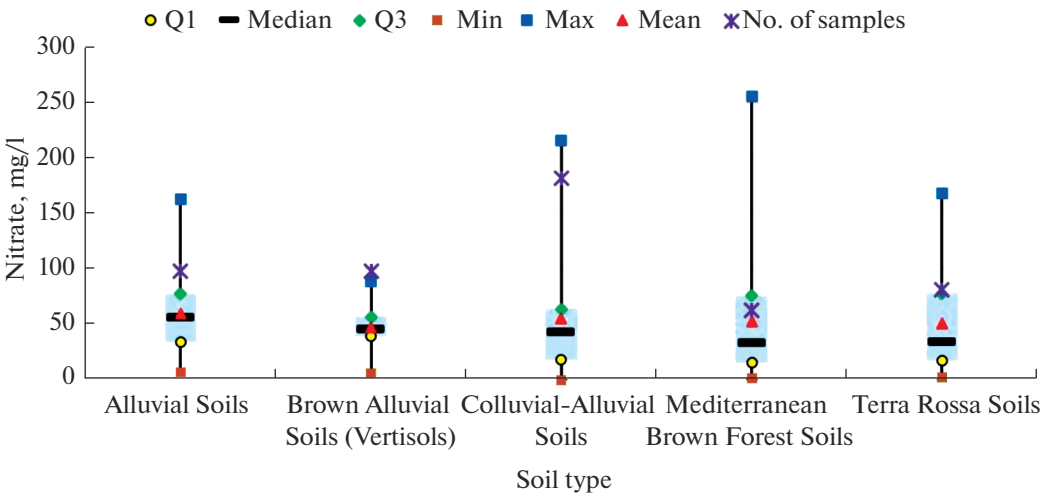


Fig. 6. Influence of soil type on NO_3 concentration (period 1982–2019).

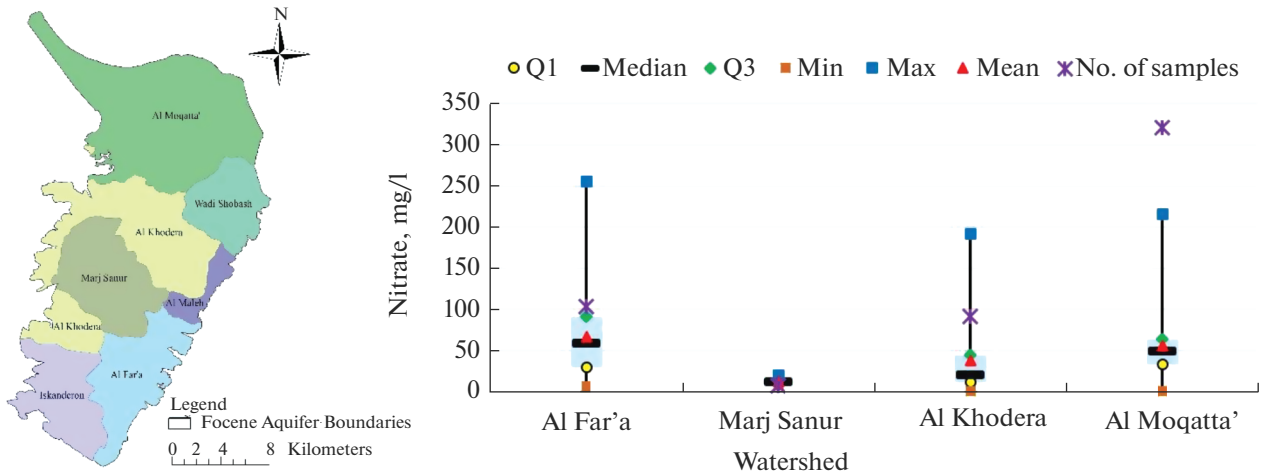
Table 3. Influence of land use on NO_3 concentrations (period 1982–2019)

Land use	Number of Samples	Concentration, mg/L			
		mean	median	min	max
Continuous Urban Areas	42	37	34	5	95
Discontinuous Urban Areas	148	85	78	0	256
Drip Irrigated Arable	59	49	45	19	85
Forest	11	24	24	0	53
Green Houses	5	41	50	16	54
Irrigated Complex Cultivation	73	55	43	6	169
Non Irrigated Arable Land	13	34	28	10	56
Non Irrigated Complex Cultivation	126	43	47	1	89
Olive Groves	45	15	16	1	42

tributed among this land use class. The second highest mean NO_3 concentration is observed in the groundwater under the irrigated cultivation with a value of 55 mg/L. This area is subjected to an extensive use of agrochemicals. The lowest mean value occurred under the olive groves as expected since olive trees are usually planted in the mountainous areas where either the depth to groundwater is high or contaminants are being washed away with surface runoff. Except for olive groves, all the maximum NO_3 concentrations exceed the MCL. The overall maximum NO_3 concentration is located in the groundwater under the discontinuous urban areas with a value of 256 mg/L.

Influence of Watersheds

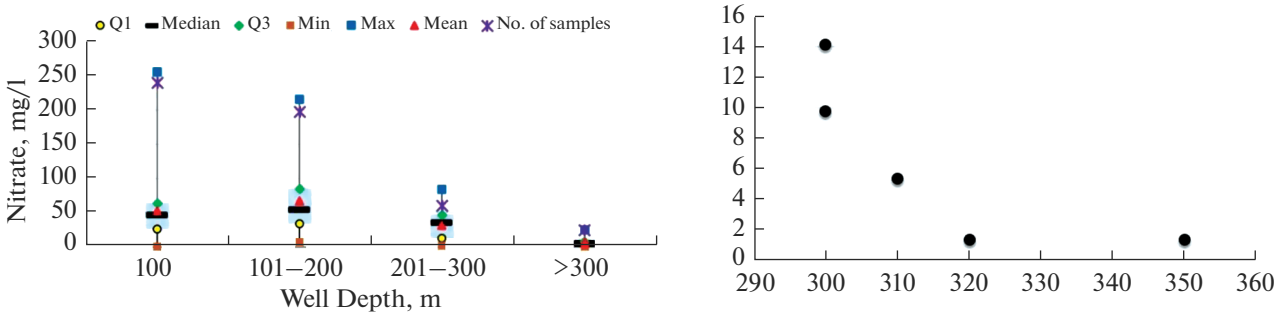
The Eocene Aquifer area overlaps seven different watersheds: Al-Moqatta', Al-Khodera, Marj Sanur, Iskanderon, Al-Far'a, Wadi Shobash and Al-Maleh (Fig. 7a). The first four watersheds drain into the west to the Mediterranean Sea, while the others drain into the east to the Jordan River. Actually, the available NO_3 concentration data were only distributed among four watersheds: Al-Far'a, Marj Sanur, Al-Khodera and Al-Moqatta'. Figure 7b illustrates the minimum, 25th percentile, mean, median, 75th percentile and maximum NO_3 concentrations for the different watersheds for the period 1982–2019. Marj Sanur watershed was excluded from the discussion as it has unrepresentative number of NO_3 concentration readings (7 readings). The groundwater wells located in Al-Far'a watershed has mean and median NO_3 concentrations that exceed the MCL. It also has the highest maximum and 75th percentile values compared to other watersheds. This is due to the intensive use of agrochemicals in the watershed as it is considered as 'the



(a) Watershed locations

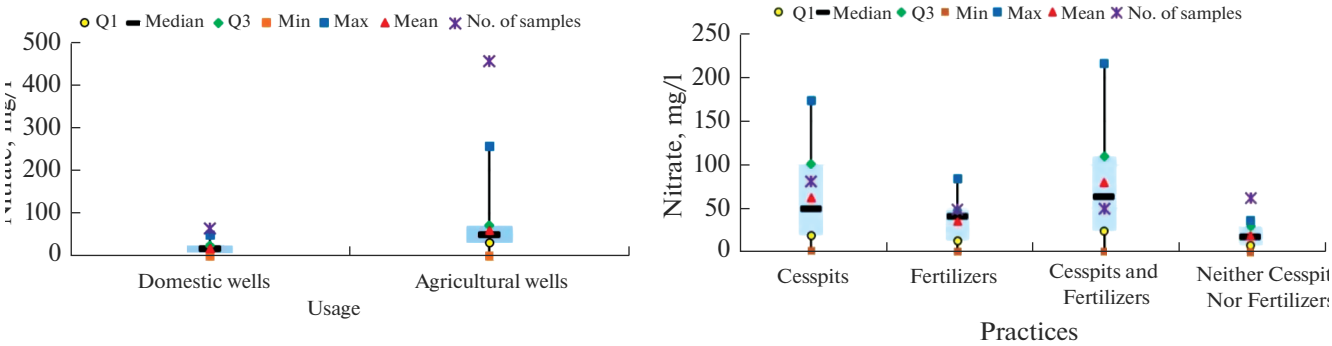
(b) GNC in the watersheds (2019)

Fig. 7. Influence of watersheds and their locations on the NO_3 concentration (period 1982–2019).



(a) for all groundwater wells (1982–2019) (b) for five adjacent groundwater wells (2017–2019)

Fig. 8. Influence of well depth on the NO_3 concentration.



(a) Well use (1982-2019) (b) Anthropogenic practices (2019)

Fig. 9. Influence of well use and anthropogenic practices on the NO_3 concentration.

food basket' for the northern part of the West Bank and the most dominated watershed by agriculture in the West Bank [3]. Mean and median NO_3 concentrations are approximately equal to MCL at Al-Moqatta' watershed while the lowest concentrations were recorded at Al-Khodera watershed.

Influence of Well Depth

Generally, there is an inverse relationship between well depths and NO_3 concentration [4, 12, 15, 43]. Figure 8a shows the variation of the statistical parameters with the well depth. It clearly indicates high NO_3 concentrations in the first 100 m of the aquifer. After that, NO_3 concentration significantly decreases with depth and becomes insignificant below 300 m. This NO_3 profile could be attributed to the following factors: denitrification in groundwater, groundwater movement and the associated NO_3 transport and mixing.

The NO_3 concentration is also influenced by other factors such as land use, groundwater recharge and fertilization practices. To reduce the influence of these factors, analysis was conducted on samples from five adjacent wells (distributed among an area of 2.37 km^2) in Qabatiya. Figure 8b illustrates the variation of concentrations with depth for these samples. It clearly shows that the NO_3 concentration decreases with the increase in depth.

Influence of the Wells Use and the Anthropogenic Practices

The groundwater wells are used for domestic and agricultural activities. Figure 9a shows that the agricultural wells have higher median (49 mg/L), mean (59 mg/L), 75th percentile (71 mg/L) and maximum (256 mg/L) values as compared to the domestic wells. The domestic wells have NO_3 concentrations in the range of 0 to 49 mg/L with a median, mean and 75th percentile that equal 16, 18 and 25 mg/L, respectively. The intensive use of agrochemicals in agriculture and cesspits for WW disposal are the main source of on-ground nitrogen loading in the study area [8]. Collected data was used to investigate the influence of these factors on GNC. Practices in the proximity of the sampled wells were classified into four categories: extensive presence of cesspits, extensive application of fertilizers, extensive use of cesspits and fertilizers, and neither cesspits nor fertilizers. Figure 9b illustrates the

Table 4. Prediction accuracy for RF model over 10 execution trials

Trial	1	2	3	4	5	6	7	8	9	10	Mean
Accuracy, %	89.6	91.7	91.7	83.3	85.4	87.5	89.6	91.6	83.3	91.7	88.5

Table 5. Confusion matrix for the RF prediction model (trial 10, Table 4)

Actual contamination level	Contamination Level	Predicted contamination level			
		not contaminated by NO_3, %	reasonable NO_3 concentration, %	contaminated by NO_3, %	extremely contaminated by NO_3, %
	Not contaminated by NO_3, %	27.08	2.08	0	0
	Reasonable NO_3 concentration, %	2.08	50	0	0
	Contaminated by NO_3, %	0	4.14	8.37	0
	Extremely contaminated by NO_3, %	0	0	0	6.25

statistics of NO_3 concentrations for the four practices. It shows that the highest concentrations are located in the category "extensive use of cesspits and fertilizers", followed by the category "extensive use of cesspits" and then the category "extensive use of fertilizers".

ML-Based GNC Prediction Model

GNC Assessment outcomes are integrated with the ML-based GNC prediction model. They provided the prediction model by the GNC distribution as well as by the main factors influencing this distribution. Such information is used as an input for the development of the prediction model. Generally, assessment results indicated that except soil type, all factors (well depth, well use, land use, watershed and fertilization and WW disposal practices) can be used as inputs for the ML prediction model. RF model was built considering these factors and executed 10 times. Table 4 shows that the obtained accuracy scores range between 83.3 and 91.7% with an average value of 88.5%.

Table 5 shows the RF confusion matrix for the best trial (accuracy = 91.7). It indicates two types of errors: safe error (error of predicting worse than contamination situation) and risk error (error of predicting less than contamination level). The safe error equals 2.08%, while the risk error equals 6.22%. Figure 10 shows the distribution of wells with correct, safe error and risk error prediction. It is noticed that the false predicted wells are not spatially correlated. They are distributed amongst the northern, middle and southern parts of the study area. Table 6 illustrates that both errors in the false predicted wells are occurred on the margins of the thresholds between two successive contamination levels. Actual and predicted GNC in wells 1 and 2 are close to the threshold of 25 mg/L. Moreover, actual and predicted values are around the threshold of 50 mg/L in wells 3 and 4. This reflects the low severity of the prediction model errors.

Table 7 shows the relative importance of the input factors used in the RF prediction model. It is obvious that the well depth (with 41% relative importance) is the most influencing factor, followed by anthropogenic-related factors. This reflects the high vulnerability of shallow groundwater wells to GNC. It also high-

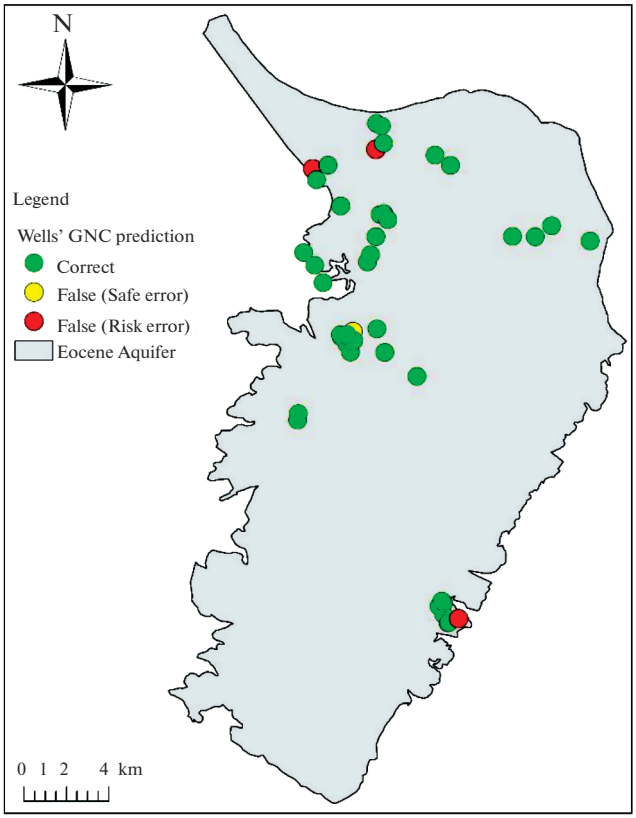


Fig. 10. Distribution of correct and false predicted wells concerning GNC.

Table 6. Types of prediction errors for the false predicted wells

Well	Actual NO_3 concentration, mg/L (level)	Threshold, mg/L	Predicted NO_3 concentration, mg/L (level)	Prediction error type
Well #1	23 (Not contaminated)	25	28 (Reasonable concentration)	Safe error
Well #2	27 (Reasonable concentration)	25	21 (Not contaminated)	Risk error
Well #3	54 (Contaminated)	50	47 (Reasonable concentration)	Risk error
Well #4	52 (Contaminated)	50	46 (Reasonable concentration)	Risk error

Table 7. Importance of input factors used in the RF prediction model

Factors	Well depth	Land use	Anthropogenic activities	Watershed	Well use
Relative importance, %	41	20	18	14	7

lights the severity of uncontrolled man-made practices (land use and fertilization and WW disposal method) upon the GNC. On the other hand, watershed and well use have the lowest influence with relative influence of 14 and 7% respectively.

There is a difficulty in controlling the depths of existing and new wells due to political restrictions. Therefore, management options should be directed toward the anthropogenic-related factors. In which, the control of intensive use of fertilizers and cesspits are the most important ones. Decision makers are advised to embrace the land use planning as a strategic tool for mitigating GNC in the study area. Protection zones for groundwater wells have to be delineated and sources of GNC have to be prohibited in those zones.

CONCLUSIONS

This paper presented an unprecedented approach that combines GIS, statistics and ML modeling to assess and predict GNC in the Eocene Aquifer, Palestine. The integration of GIS and statistical analysis led to a comprehensive spatiotemporal assessment of GNC. This assessment works efficiently in providing the RF prediction model by the needed input influencing factors. Assessment results indicated serious NO_3 contaminations of the Aquifer. GIS maps indicated an increasing GNC trends in the southern part of the study area. Results showed the necessity to use five factors (well depth, well use, land use, watershed and fertilization and WW disposal practices) in the RF predictive model of the GNC. RF model successfully attained a maximum and average prediction accuracy of 91.70 and 88.54%, respectively.

Decision makers could use the approach presented in this paper to establish a knowledge-based method for the sustainable management of groundwater quality in Palestine. Collaboration is crucial between different stakeholders (e.g. PWA and MoA) to establish a groundwater monitoring programs. Such program should support the sustainable utilization of groundwater in Palestine by controlling the intensive use of agrochemicals and cesspits.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Almasri, M.N. and Ghabayen, S.M., Analysis of nitrate contamination of Gaza coastal aquifer, Palestine, *J. Hydrol. Eng.*, 2008, vol. 13, no. 3. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:3(132)
2. Almasri, M.N. and Kaluarachchi, J.J., Assessment and management of long-term nitrate pollution of ground water in agriculture-dominated watersheds, *J. Hydrol.*, 2004, vol. 295, nos. 1–4. https://doi.org/10.1016/j.jhydrol.2004.03.013
3. Almasri, M.N., Judeh, T.G., and Shadeed, S.M., Identification of the nitrogen sources in the Eocene Aquifer Area (Palestine), *Water*, 2020, vol. 12, no. 4. https://doi.org/10.3390/w12041121
4. Anayah, F.M. and Almasri, M. N., Trends and occurrences of nitrate in the groundwater of the West Bank, Palestine, *Appl. Geog.*, 2009, vol. 29, no. 4. https://doi.org/10.1016/j.apgeog.2009.01.004
5. Band, S.S., Janizadeh, S., Pal, S.C., Chowdhuri, I., Siabi, Z., Norouzi, A., Melesse, A.M., Shokri, M., and Mosavi, A., Comparative analysis of artificial intelligence models for accurate estimation of groundwater nitrate concentration, *Sensors (Switzerland)*, 2020, vol. 20, no. 20, pp. 1–23. https://doi.org/10.3390/s20205763
6. Barkat, A., Bouaicha, F., Bouteraa, O., Mester, T., Ata, B., Balla, D., Rahal, Z., and Szabó, G., Assessment of complex terminal groundwater aquifer for different use of Oued Souf Valley (Algeria) using multivariate statistical methods, geostatistical modeling, and water quality index, *Water*, 2021, vol. 13, no. 11. https://doi.org/10.3390/w13111609

7. Breiman, L., Random forests, *Machine Learning*, 2001, vol. 45, no. 1.
https://doi.org/10.1023/A:1010933404324
8. Canion, A., McCloud, L., and Dobberfuhl, D., Predictive modeling of elevated groundwater nitrate in a karstic spring-contributing area using random forests and regression-kriging, *Environ. Earth Sci.*, 2019, vol. 78, no. 9.
https://doi.org/10.1007/s12665-019-8277-1
9. Comly, H.H., Cyanosis in infants caused by nitrates in well water, *JAMA, J. Am. Med. Assoc.*, 1987, vol. 257, no. 20.
https://doi.org/10.1001/jama.1987.03390200128027
10. Ducci, D., An easy-to-use method for assessing nitrate contamination susceptibility in groundwater, *Geofluids*, 2018.
https://doi.org/10.1155/2018/1371825
11. El Baba, M., Kayastha, P., Huysmans, M., and de Smedt, F., Evaluation of the groundwater quality using the water quality index and geostatistical analysis in the Dier Al-Balah Governorate, Gaza Strip, Palestine, *Water*, 2020, vol. 12, no. 1.
https://doi.org/10.3390/w12010262
12. Freeze, R.A., Cherry, J.A., and Cherry, J.A., *Groundwater*, Prentice-Hall, 1979.
13. Friedl, M.A., Brodley, C.E., and Strahler, A.H., Maximizing land cover classification accuracies produced by decision trees at continental to global scales, *IEEE Trans. Geosci. Electron.*, 1999, vol. 37, no. 2.
https://doi.org/10.1109/36.752215
14. Hajhamad, L. and Almasri, M.N., Assessment of nitrate contamination of groundwater using lumped-parameter models, *Environ. Modelling Software*, 2009, vol. 24, no. 9.
https://doi.org/10.1016/j.envsoft.2009.02.014
15. Hallberg, G.R., Nitrate in ground water in the United States, *Developments in Agricultural and Managed-Forest Ecology*, Follett, R., Ed., Elsevier, 1989, vol. 21.
https://doi.org/10.1016/B978-0-444-87393-4.50009-5
16. Isam, N., *Geo-Database of the Ministry of Agriculture, Personal Communication*, Palestine, Ramallah, 2019.
17. Judeh, T., Bian, H., and Shahrour, I., GIS-based spatiotemporal mapping of groundwater potability and palatability indices in arid and semi-arid areas, *Water*, 2021, vol. 13, no. 9.
https://doi.org/10.3390/w13091323
18. Judeh, T., Haddad, M., and Özerol, G., Assessment of water governance in the West Bank, Palestine, *Int. J. Global Environ. Issues*, 2017, vol. 16, nos. 1/2/3.
https://doi.org/10.1504/IJGENVI.2017.083426
19. Khader, A.I., Rosenberg, D.E., and McKee, M., A decision tree model to estimate the value of information provided by a groundwater quality monitoring network, *Hydrol. Earth Syst. Sci.*, 2013, vol. 17, no. 5.
https://doi.org/10.5194/hess-17-1797-2013
20. Khayat, S., Geyer, S., Hötzl, H., Ghanem, M., and Ali, W., Identification of nitrate sources in groundwater by Δ15N nitrate and Δ18O nitrate isotopes: a study of the shallow Pleistocene aquifer in the Jericho Area, Palestine, *Acta Gydrochim. Hydrobiol.*, 2006, vol. 34, nos. 1–2.
https://doi.org/10.1002/aheh.200400615
21. Knoll, L., Breuer, L., and Bach, M., Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning, *Sci. Total Environ.*, 2019, vol. 668.
https://doi.org/10.1016/j.scitotenv.2019.03.045
22. Li, H., Smith, C.D., Wang, L., Li, Z., Xiong, C., and Zhang, R., Combining spatial analysis and a drinking water quality index to evaluate monitoring data, *Int. J. Environ. Res. Public Health*, 2019, vol. 16, no. 3.
https://doi.org/10.3390/ijerph16030357
23. Lodwick, W. A., Monson, W., and Svoboda, L., Attribute error and sensitivity analysis of map operations in geographical information systems: suitability analysis, *Int. J. Geogr. Inform. Syst.*, 1990, vol. 4, no. 4, pp. 413–428.
https://doi.org/10.1080/02693799008941556
24. Mair, A. and El-Kadi, A.I., Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA, *J. Contam. Hydrol.*, 2013, vol. 153.
https://doi.org/10.1016/j.jconhyd.2013.07.004
25. Maqsoom, A., Aslam, B., Khalil, U., Ghorbanzadeh, O., Ashraf, H., Faisal Tufail, R., Farooq, D., and Blaschke, T., A GIS-based DRASTIC Model and an adjusted DRASTIC Model (DRASTICA) for groundwater susceptibility assessment along the China–Pakistan Economic Corridor (CPEC) Route, *ISPRS Int. J. Geo-Information*, 2020, vol. 9, no. 5.
https://doi.org/10.3390/ijgi9050332
26. Mas-Pla, J. and Menció, A., Groundwater nitrate pollution and climate change: learnings from a water balance-based analysis of several aquifers in a Western Mediterranean Region (Catalonia), *Environ. Sci. Pollut. Res.*, 2019, vol. 26, no. 3.
https://doi.org/10.1007/s11356-018-1859-8
27. Palestinian Central Bureau of Statistics (PCBS). Population in Palestine. Available online: http://www.pcbs.gov.ps/site/lang__en/803/default.aspx?lang=en (accessed on 12 February, 2020)
28. Palestinian Water Authority (PWA), *Annual Status Report on Water Resources, Water Supply, and Wastewater in the Occupied State of Palestine*, 2011.
29. Palestinian Water Authority (PWA), National Water Information System. Available online: http://nwis.pwa.ps:82 (accessed on 18 January, 2020).
30. Qrenawi, L. and Shomar, R.T.A., Health risk assessment of groundwater contamination case study: Gaza Strip, *J. Eng. Res. Technol.*, 2020, vol. 7.
31. Ransom, K.M., Nolan, B.T., Traum, J., Faunt, C.C., Bell, A.M., Gronberg, J.A.M., Wheeler, D.C., Z. Rosecrans, C., Jurgens, B., Schwarz, G.E., Belitz, K., Eberts, S., Kourakos, G., and Harter, T., A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley Aquifer, California, USA, *Sci. Total Environ.*, 2017, vols. 601–602, pp. 1160–1172.
https://doi.org/https://doi.org/10.1016/j.scitotenv. 2017.05.192
32. Rizeei, H.M., Azeez, O.S., Pradhan, B., and Khamees, H.H., Assessment of groundwater nitrate contamination hazard in a semi-arid region by using integrated parametric IPNOA and Data-Driven Logistic Regression models, *Environ. Monit. Assess.*, 2018,

vol. 190, no. 11.
https://doi.org/10.1007/s10661-018-7013-8

33. Rodda, H.J.E., Demuth, S., and Shankar, U.D.E., The application of a GIS-based decision support system to predict nitrate leaching to groundwater in Southern Germany, *Hydrol. Sci. J.*, 1999, vol. 44, no. 2, pp. 221–236.
https://doi.org/10.1080/02626669909492219
34. Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., and Ribeiro, L., Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain), *Sci. Total Environ.*, 2014, vols. 476–477, pp. 189–206.
https://doi.org/https://doi.org/10.1016/j.scitotenv.2014.01.001
35. Schilling, K.E., and Wolter, C.F., A GIS-based groundwater travel time model to evaluate stream nitrate concentration reductions from land use change, *Environ. Geol.*, 2007, vol. 53, no. 2, pp. 433–443.
https://doi.org/10.1007/s00254-007-0659-0
36. Schröder, J.J., Scholefield, D., Cabral, F., and Hofman, G., The effects of nutrient losses from agriculture on ground and surface water quality: the position of science in developing indicators for regulation, *Environ. Sci. Policy*, 2004, vol. 7, no. 1.
https://doi.org/10.1016/j.envsci.2003.10.006
37. Shadeed, S.M., Judeh, T.G., and Almasri, M.N., Developing GIS-based water poverty and rainwater harvesting suitability for domestic use in the Dead Sea Region (West Bank, Palestine), *Hydrol. Earth Syst. Sci.*, 2019, vol. 23, no. 3, pp. 1581–1592.
https://doi.org/10.5194/hess-23-1581-2019
38. Shadeed, S., Judeh, T., and Riksen, M., Rainwater harvesting for sustainable agriculture in high water-poor areas in the West Bank, Palestine, *Water*, 2020, vol. 12, no. 2.
https://doi.org/10.3390/w12020380
39. Shadeed, S., Spatio-temporal drought analysis in arid and semi-arid regions: a case study from Palestine, *Arabian J. Sci. Eng.*, 2013, vol. 38, no. 9.
https://doi.org/10.1007/s13369-012-0504-y
40. Sheikhy Narany, T., Ramli, M.F., Aris, A.Z., Sulaiman, W.N.A., and Fakharian, K., Spatial assessment of groundwater quality monitoring wells using indicator kriging and risk mapping, Amol-Babol Plain, Iran, *Water*, 2014, vol. 6, no. 1, pp. 68–85.
https://doi.org/10.3390/w6010068
41. Shrestha, A. and Luo, W., Assessment of groundwater nitrate pollution potential in central valley aquifer using geodetector-based frequency ratio (GFR) and Optimized-DRASTIC Methods, *ISPRS Int. J. Geo-Inf.*, 2018, vol. 7, no. 6.
https://doi.org/10.3390/ijgi7060211
42. SUSMAQ, *Conceptual, Steady-State and Transient-State Models of the Eocene Aquifer in the North-Eastern Aquifer Basin*, Ramallah, Palestine, 2004.
43. Tesoriero, A.J. and Voss, F.D., Predicting the probability of elevated nitrate concentrations in the Puget Sound Basin: implications for aquifer susceptibility and vulnerability, *Ground Water*, 1997, vol. 35, no. 6.
https://doi.org/10.1111/j.1745-6584.1997.tb00175.x
44. Tubeileh, H., Shaheen, H., and Aliewi, A., Modeling the Eocene aquifer in northern West Bank, *An - Najah Univ. J. Res. (Science)*, 2006, vol. 20.
45. Uddameri, V., Silva, A., Singaraju, S., Mohammadi, G., and Hernandez, E., Tree-based modeling methods to predict nitrate exceedances in the Ogallala Aquifer in Texas, *Water*, 2020, vol. 12, no. 4.
https://doi.org/10.3390/w12041023
46. United Nations Development Programme (UNDP), *Water Governance in the Arab Region: Managing Scarcity and Securing the Future*, United Nations Development Programme, Sweden, 2013.
47. UN-Water, *Policy Brief: Water Quality*, 2011.
48. World Health Organization (WHO), *Nitrate and Nitrite in Drinking-Water: Background Document for Development of WHO Guidelines for Drinking-Water Quality*, World Health Organization, Geneva, 2003.S

SPELL: 1. OK