Environ Monit Assess (2018) 190: 633
https://doi.org/10.1007/s10661-018-7013-8

Assessment of groundwater nitrate contamination hazard in a semi-arid region by using integrated parametric IPNOA and data-driven logistic regression models

Hossein Mojaddadi Rizeei · Omer Saud Azeez · Biswajeet Pradhan · Hayder Hassan Khamees

Received: 7 June 2018 / Accepted: 26 September 2018 / Published online: 4 October 2018

Abstract Groundwater hazard assessments involve many activities dealing with the impacts of pollution on groundwater, such as human health studies and environment modelling. Nitrate contamination is considered a hazard to human health, environment and ecosystem. In groundwater management, the hazard should be assessed before any action can be taken, particularly for groundwater pollution and water quality. Thus, pollution due to the presence of nitrate poses considerable hazard to drinking water, and excessive nutrient loads deteriorate the ecosystem. The parametric IPNOA model is one of the well-known methods used for evaluating nitrate content. However, it cannot predict the effect of soil and land use/land cover (LULC) types on calculations relying on parametric well samples. Therefore, in this study, the parametric model was trained and integrated with the multivariate data-driven model with different levels of information to assess groundwater nitrate contamination in Saladin, Iraq. The IPNOA model was developed with 185 different well samples and contributing parameters. Then, the IPNOA model was integrated with the logistic regression (LR) model to predict the nitrate contamination levels. Geographic information system techniques were also used to assess the spatial prediction of nitrate contamination. High-resolution SPOT-5 satellite images with 5 m spatial resolution were processed by object-based image analysis and support vector machine algorithm to extract LULC. Mapping of potential areas of nitrate contamination was examined using receiver operating characteristic assessment. Results indicated that the optimised LR-IPNOA model was more accurate in determining and analysing the nitrate hazard concentration than the standalone IPNOA model. This method can be easily replicated in other areas that have similar climatic condition. Therefore, stakeholders in planning and environmental decision makers could benefit immensely from the proposed method of this research, which can be potentially used for a sustainable management of urban, industrialised and agricultural sectors.

Keywords Nitrate contamination · IPNOA · GIS · Logistic regression · Groundwater hazard assessment

H. M. Rizeei · B. Pradhan (✉)
Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia
e-mail: biswajeet24@gmail.com

e-mail: Biswajeet.Pradhan@uts.edu.au

O. S. Azeez · H. H. Khamees
Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, Seri Kembangan, Selangor, Malaysia

Introduction

Groundwater sources are the most crucial, dependable and valuable sources of water in all climatic regions in the world (Sacco et al. 2006). The demand for groundwater continues to increase because of population growth, agricultural requirements, urbanisation (Ettazarini 2007) and rapid industrialisation (Pradhan 2009). Groundwater has more benefits than surface water. Groundwater exhibits better quality, is less exposed to seasonal and perennial fluctuations and is more protected from pollutants and infections than surface water (Green et al. 2008). Hydro-technical facilities for surface water require larger investment than groundwater facilities that can be developed gradually (Nampak et al. 2014). Many models have been developed to assess groundwater quality and quantity, such as DRASTIC, neuro-fuzzy classifier and SINTAC, which are designed for groundwater risk and vulnerability assessment (Neshat and Pradhan 2015; Neshat et al. 2015). The Italian experience with SINTAC was a good example, particularly when it was associated with the hazard parameters for nitrate risk assessment (Sacco et al. 2007; Capri et al. 2009; Ghiglieri et al. 2009).

Padovani and Trevisan (2002) proposed IPNOA as the nitrate hazard index (HI) to detect the potential hazard to groundwater of nitrate contamination from agricultural activities at the provincial and regional scales. This approach involves the parametric analysis of all of the parameters under consideration in the potential groundwater contamination hazard (Raaz Maheshwari et al. 2012). Then, a progressive score is allocated to each parameter in accordance with its importance determined through the evaluation (Padovani and Trevisan 2002). This method is aimed at evaluating the nitrate HI from domestic sources (Capri et al. 2009).

Leakages from sewage, mishandling of wastewater effluents, improper denitrification during wastewater treatment and inadequate curing using fertilisers and wastes of animals are the major sources of nitrate contents in groundwater (Zhao et al. 2013). These sources lead to the contamination of aquifers, particularly where groundwater replenishment occurs directly from surface water (DeSimone and Howes 1998; Vikas et al. 2015). Nitrate contamination is often attributed to human activities on the ground that involve fertilisation of agricultural products (Gross 2008). Contaminated groundwater is not easily remediated after occurrence; thus, prevention is the most effective strategy in water quality management (Kowal and Polik 1987; Re et al. 2017). According to Into (2011), the concentrations of nitrate contamination in groundwater exceed the standard limits for freshwater (50 mg $N–NO_3/L$) in Europe in approximately 22% of planted areas (Sacco et al. 2007). Comparable levels, which can cause cancer for those who consume such water, were also observed in China and USA (Cantor 1997).

The application of geographic information systems (GIS) and remote sensing (RS) techniques allows the exploration of groundwater resources in a range of hydrogeological settings (Gupta and Srivastava 2010; Mishra et al. 2014; Nampak et al. 2014). Artificial intelligence and machine learning in GIS have been applied to a vast range of industries and applications with large amounts of data, and deep learning can help automate the extraction of information from visual datasets, which was previously impossible (Neshat and Pradhan 2015; Mezaal et al. 2017). The analysis of large amounts of hydrogeological data and the simulation of complex subsurface flow and transport processes can be conducted using the combination of GIS and RS systems (Mojaddadi et al. 2009; Althuwaynee et al. 2012, 2014; Bui et al. 2017a, b; Chen et al. 2017; Oh and Pradhan 2011; Pradhan 2013; Tehrany et al. 2013, 2014, 2015; Umar et al. 2014; Abdulkareem et al. 2018a, b; Rizeei et al. 2018a, b; Kordestani et al. 2018; Golkarian et al. 2018).

The integration of RS and GIS techniques to determine aquifer system suitability in Ghana was proposed by Gumma and Pavelic (2013). The main aim was to improve the development of groundwater in agricultural and urban purposes (Alwathaf and El Mansouri 2011). The significance of the integration of RS and GIS techniques in the assessment of groundwater projects has been emphasised in many studies (Shahid et al. 2002; Sener et al. 2005; Shaban et al. 2006; Nampak et al. 2014).

Most of the studies related to groundwater pollution concentrated on vulnerability assessment rather than hazard evaluation, that is, the high groundwater vulnerability does not necessarily represent the high level of hazard to groundwater (Spalding and Exner 1993; Lake et al. 2003; Raaz Maheshwari et al. 2012). Generally, most concentrations of nitrate contamination in groundwater are from agriculture areas, but residual and industrial areas contribute to nitrate contamination because of leakages from sewages and discharges from domestic activities not connected to sewage systems (Boy Roura 2013). The IPNOA model does not consider the soil types and roles of land use/land cover (LULC) in nitrate risk and hazard modelling. The IPNOA model runs based on parametric sample points from wells (Ghiglieri et al. 2009). The contributing hazard and control factors (HFs and CFs, respectively) are calculated with equal degree of importance, which is not an efficient method of conducting nitrate concentration analysis.

In the current research, we tried to overcome the previous drawbacks of groundwater nitrate modelling. Therefore, the IPNOA model was used to determine the nitrate concentration in agricultural lands using HFs and CFs. The soil type parameter was added to the IPNOA equation in addition to a detailed LULC extracted from SPOT-5 satellite imagery to improve the quality of contributing factors. The IPNOA model was optimised by the logistic regression (LR) model for the first time to assign the proper weight to HFs and CFs to deal with skewness and uncertainty from the standalone IPNOA model.

Material and methods

Study area

The investigation was conducted in Saladin Province of Iraq, which is located north of the capital Baghdad. Saladin Province covers an area of approximately 24,363 km^2 and has a population of 1,042,200 persons. Saladin Province has two main cities, that is, the capital city Tikrit and the city of Samarra. Saladin Province was a part of the capital Baghdad before 1976. The study area is geographically located at 43° 20′ 19.14″ E, 43° 60′ 59.74″ E and 34° 41′ 30.14″ N, 34° 28′ 30.74″ N. the area has a variety of LULC types (i.e. urban, bare and agricultural lands). The study area is composed of sandy to gravel soil types where average humidity is 33%. The amount of precipitation ranges from 600 to 800 mm/year, and the annual average temperature ranged between 15 and 23 °C. The elevation varies from 69 to 178 m. The formation of the study area shows almost similar lithological composition of limestone material (Abdula 2016).

Nitrate contamination of groundwater in the study area is considered moderate with regard to its geological structure, that is, sandy dunes soil (AL-Dulaimi and Younes 2017). Several studies conducted in Tehama Basin have reported nitrate pollution in groundwater, which is one of the major sources of drinking water for inhabitants. The high concentration of nitrate is a result of decomposing organic matters and increasing use of fertiliser and sewage water because most of the cities and villages in the study area have inadequate sanitation system coupled with shallow water content. Figure 1 illustrates the location of study area.

Datasets

The data used in this work divides into three groups, that is, independent, dependent and analysis data. The dependent data comprised concentrated groundwater containing nitrate in mg/L sourced from 185 wells. The data were obtained from hydrogeological and hydrochemical investigations conducted by the government of Saladin. The samples were analysed in the laboratory. The field-data-collection campaign was commenced in 25 November 2014 and was completed in 31 December 2016. Parameters, such as rainfall, TEM, PH, EC and TDO, were recorded from the field as dependent data.

High-resolution SPOT-5 satellite image was used to extract LULC as independent explanatory data (Aalshamkhi et al. 2017). Other data, such as well depth (shallow and deep wells), were also recorded.

Nitrate groundwater quality in Saladin ranged from 1.3 to 203 mg/L, and the average of 46 mg/L was reported from 185 well samples, of which 65 showed values higher than 50 mg/L. The inventory points of nitrate samples were split into two groups, with the first group being the training group (70%) and the second group being the testing group (30%) (Fig. 2). The model was trained using 70% of the samples and tested using the remaining 30%.

Methods

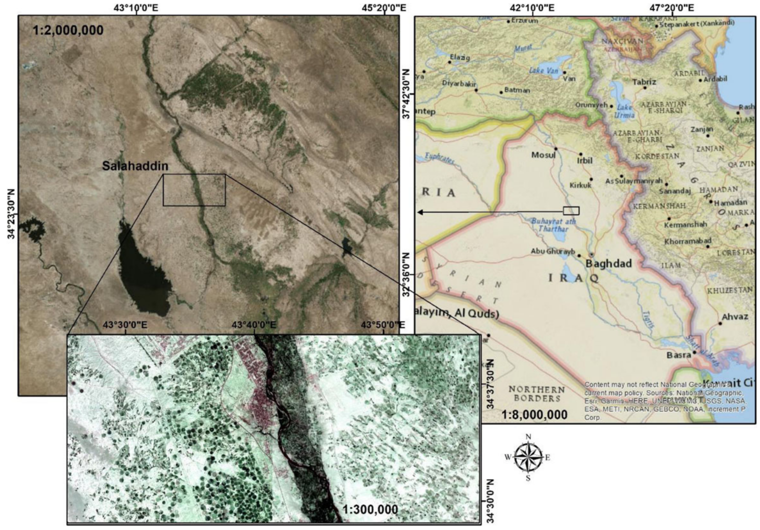
Three main technical parts comprise the framework of the current study, that is, satellite image classification, IPNOA parameter extraction and statistical model construction, as shown in Fig. 3.

Several programs and models were used in this analysis. ENVI 5.4 was used for satellite image processing, ArcMap 10.5 was used for spatial modelling and mapping and SPSS 9.2 was used for statistical-data-driven modelling.

LULC extraction

LULC is one of the primary factors that contributes immensely to the geohazard and risk in groundwater and surface water (Hossein Mojaddadi et al. 2017). The nitrates percolate through the soil depending on the type of soil and LULC. In a densely populated environment, nitrates from septic tanks seep down and fertilisers from LULC accumulate and contaminate shallow groundwater until it exceeds the safe drinking water standard level

Fig. 1 Location of study area

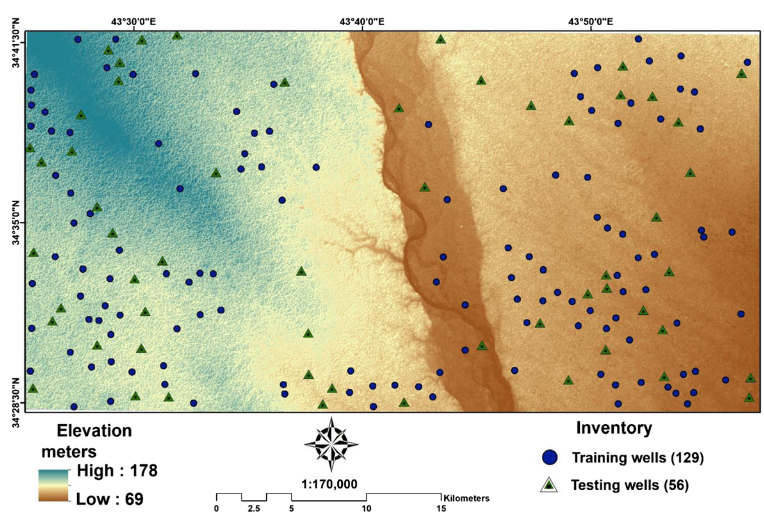


(Wick et al. 2009). For optimum outcome, the variables and subdivision should be selected precisely because of the sensitive correlation between nitrate concentration and LULC (Min et al. 2003).

High-resolution satellite image, such as IKONOS and SPOT satellite imageries, has been recommended to extract a detailed LULC map (Aal-shamkhi et al. 2017; Abdullahi et al. 2017). The SPOT-5 satellite image has high-resolution characteristics and provides a 5-m multispectral resolution. The image was captured in December 2015. All necessary pre-processing approaches were implemented on the image (i.e. radiometric, geometric and atmospheric corrections). In object-based image analysis (OBIA), multiresolution segmentation is the first step of processing that has three components, namely, scale, merge and compactness, which should be defined carefully before applying the support vector machine (SVM) classifier to the image (Rizeei et al. 2018a, b). The Taguchi method was employed to select the best combination of scale, merge and compactness factors to optimise the segments. Meanwhile, the radial base function kernel type was selected as the SVM classifier. The optimal parameters of segmentation and the applied classifier details are shown in Table 1.

The extracted LULC map in this study area is categorised into five classes, namely, urban land, irrigated agricultural land, bare land, water body and rain-fed agricultural land. However, the study area is mostly covered by agricultural lands, as shown in Fig. 4.

Fig. 2 Inventory points of nitrate samples in the study area



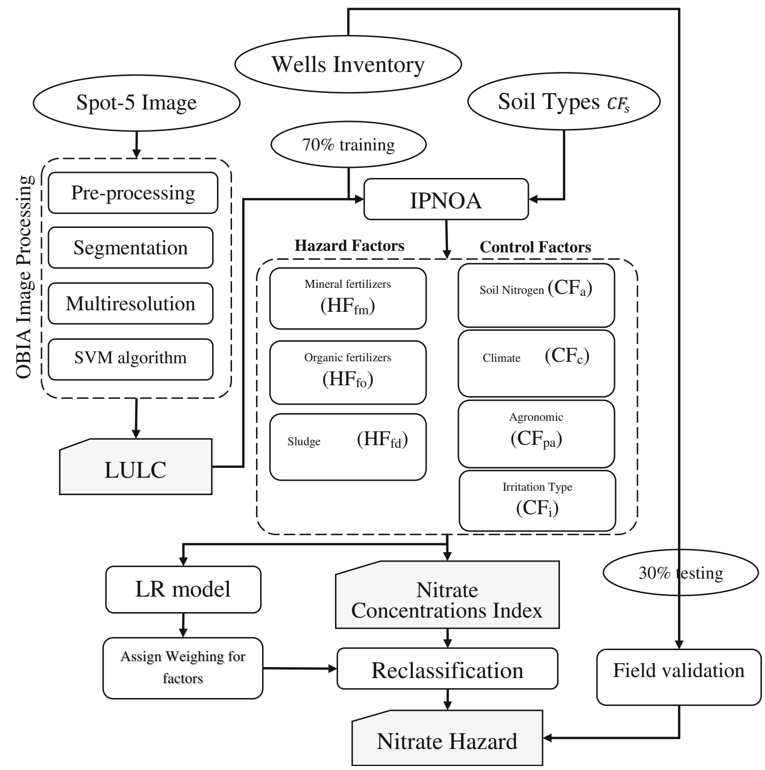


Fig. 3 Methodology flowchart applied in this study

Implementation of the IPNOA model

The distribution of nitrogen sources in agricultural lands exhibits a potential negative effect on groundwater (Bone et al. 2010). HFs comprise waste treatment sludge (HF_{fd}), organic fertilisers (HF_{fo}) and non-organic fertilisers (HF_{fm}). Conversely, CFs are employed to assess the influence of nitrate activities in terms of site and farm situation in the area under consideration (Padovani and Trevisan 2002). Leaching of nitrogen content in soil (CF_a), agronomic practices (CF_{pa}), climate (CF_c) and irrigation techniques (CF_i) are some of the factors that regulate water balance. The factors were combined by assigning a score that classifies the level of nitrogen or the effect (positive, negative or neutral) of the factors involved in the leaching of nitrates (CF). The scores assigned to the used factors can be calculated using the conversion tables shown in Tables 2 and 3. The HF score ranges from zero to five, whereas the CF score ranges from 0.94 to 1.10. A score of zero represents non-agricultural land usage (urban and natural environments). Therefore, nitrate contamination from agricultural hazard is finally estimated by multiplying the sum of the scores of the HFs by the product of the CFs.

HF of mineral fertilisers The average amount of nitrogen used for each crop grown was used to estimate the nitrogen input from mineral fertilisers (HF_{fm}), which is based on the agronomic practices plotted on the LULC map.

Table 1 OBIA segmentation and classification in details

Segmentation specifications				Classification algorithm				
Spectral layer weights	Scale parameter	Shape	Compactness	Algorithm	Threshold	kernel	Gamma in kernel	Penalty parameter
Green = 2 NIR = 2 Red = 1 Blue = 1	40	0.2	5	SVM	5	RBF	0.03	100

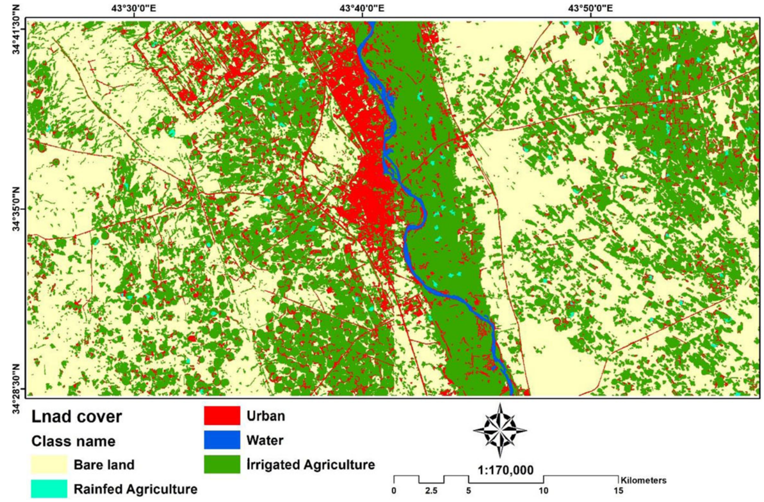


Fig. 4 Classified LULC map by OBIA-SVM classifier

HF of organic fertilisers The information related to organic fertilisers on agricultural lands was extracted from the site survey. Hence, the identification of cattle farm location, which is responsible for animal waste (HF_{fo}), is easy. However, only a few were observed to operate in such farms in the area under consideration and the disposal of animal waste was easily delineated geographically. The type and amount of waste and their corresponding scores were utilised to estimate nitrogen concentrations in farmlands. The organic fertiliser map was created based on the geostatistical interpolation method (i.e. kriging interpolation technique) using the spatial analyst tool in the GIS framework.

Table 2 Hazard factors score contributed in nitrate concentration

Hazard factors (HF)	Score
Mineral fertilisers (kg/ha/year)	HF_{fm}
0	1
1–25	2
26–100	3
100–180	4
> 180	5
Organic fertilisers (kg/ha/year)	HF_{fo}
0	1
1–150	2
151–300	3
300–500	4
> 500	5
Sludge (kg/ha/year)	HF_{fd}
0	1
1–150	2
151–500	3
500–1500	4
> 1500	5

HF of sludge Generally, sludge spreading (HF_{fd}) is not practiced in all farms in the study area. Therefore, a score of 1 was assigned to this factor. However, in several farms near streams, the amount of sludge was assigned a score of 2. Figure 5 shows the combined HFs of mineral fertilisers, organic fertilisers and sludge, which were implemented by the spatial analyst tool in ArcMap.

Nitrate contamination originating from agricultural activities can be estimated by the potential HI (Raaz Maheshwari et al. 2012).

CF of soil nitrogen content This factor indicated by nitrates of household origin and agricultural lands are designed to measure the N loading and leakage from wastewater pipes. Analytical data are collected during soil surveys to calculate the soil nitrogen content (CF_a). The total nitrogen content of the arable stratum was extracted from the data and spatialized during geostatistical interpolation using the kriging interpolation method (Vendrusculo et al. 2002). Figure 6a shows the map of soil nitrogen content.

CF of climate The spatial data covering a 20-year time series (from 1995 to 2015) were used for temperature and precipitation. In the original procedure, temperature and rainfall are classified in accordance with the

Table 3 Control factors score contributed in nitrate concentration

Control factor (CF)		Score
Soil nitrogen content (%)		CF_a
> 0.5		0.104
0.22–0.5		0.102
0.15–0.22		0.1
0.1–0.15		0.98
< 0.1		0.96
Climate		CF_c
Annual precipitation (mm)	Annual temperature (c)	
> 1200	6–15	1.10
1050–1150	13	1.08
950–1100	14–16	1.06
600–1000	15–16	1.02
500–900	14–16	1.01
500–600	< 16	098
< 600	> 16	0.94
Agronomic practices		CF_{pa}
Fertilisation	Tillage	1.04
Ferti-irrigation		1.00
Spread fertilisation	Traditional tillage	0.98
Leaf fertilisation	Minimum tillage	0.96
Local fertilisation	No tillage	0.94
Irrigation type		CF_i
Flood		1.06
Furrow		1.04
Aspersion		1.02
Drip		1.01
No irrigation		1.00
Soil types		CF_s
Sand and gravel		1.06
Peat		1.04
Quaternary alluvial		1.02
Sandy loam		1.00
Loam		0.98
Silty loam		0.96
Clay loam		0.94

baseline class and assigned a score of one. The amount of precipitation ranges from 600 to 800 mm/year, and the annual average temperature ranged between 15 and 23 °C. Therefore, areas with low annual mean temperature and/or high rainfall are regarded as more hazardous, meaning its score is higher than one. However, in the area under consideration, no spatial changes were observed in the small area of investigation. According to IPNOA climate scheme, 0.94 is achieved where rainfall is less than 600 mm/year and annual mean temperature is more than 16 °C. The entire area obtained the same value of 0.98 because no significant variation of rainfall was observed over the study area (Fig. 6b).

CF of agronomic practices The agronomic practices CF_{pa} were established based on LULC depending on field surveys conducted for each crop type. Traditional agronomic practices, such as fertile irrigation or tillage/local fertilisation, are usually implemented in several areas. In certain circumstances, the score of the most effective agricultural practice is used. Figure 6c shows the agronomic map of the study area.

CF of irrigation The integration of layer information and field observation was used to produce the irrigation control factor (CF_i) of LULC. This factor is appropriate for the irrigation technique because of its capability to transfer pollutants towards aquifers in direct connection with the efficiency of irrigation drainage. From the surveys conducted on different irrigation types on agricultural land, the scores assigned to LULC classes are shown in Fig. 6d. The detailed LULC extracted from satellite images can considerably enrich the accuracy of the CF_i value.

CF of soil characteristics Soil is the major channel for contaminations to penetrate groundwater, depending on the type of contaminants and the soil texture. Soil texture and nitrate contamination exhibit a positive correlation and are, thus, considered important factors affecting nitrate concentration in groundwater (DeSimone and Howes 1998). Conversely, soil texture is observed to be related to water input, nitrate application rate and evapotranspiration, which are regarded as the major factors determining the nitrate fluxes in groundwater (Liao et al. 2012). In this study, soil has six types, that is, from sandy to gravel, but mostly covers the gypsiferous gravel type (Fig. 6e).

The soil type factor is added to the IPNOA model to improve nitrate concentration analysis. The soil characteristic CF data were collected during the surveys and the classes were determined based on the soil media, as shown in Table 3 and Eq. (1).

The hazard index result is achieved by multiplying the HFs by the CFs, as shown in Eq. (1):

$$\mathrm{HI} = (\mathrm{HFf} + \mathrm{HFm} + \mathrm{HFs}) \times \mathrm{CF_a} \times \mathrm{CF_c} \times \mathrm{CF_{ap}} \times \mathrm{CF_i} \times \mathrm{CF_s}, \qquad (1)$$

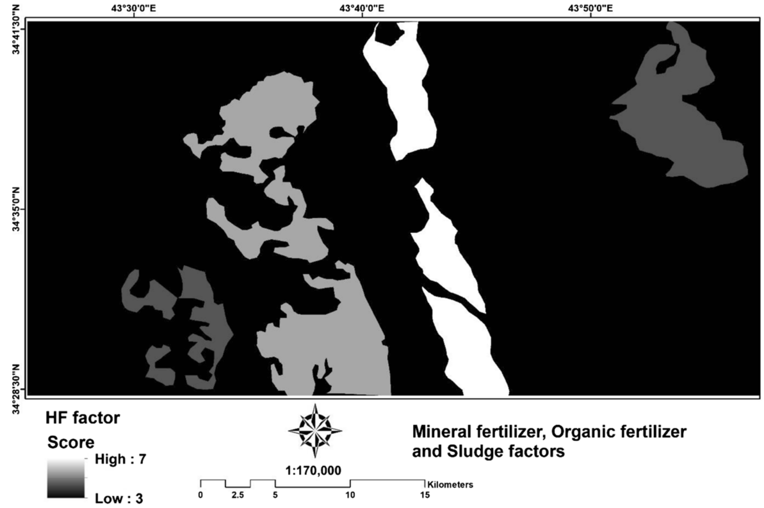


Fig. 5 Combined hazard factors (HF) index in study area

where the subscripts f, m, s, a, c, ap, i and s represent fertilisers, manure, sludge, nitrogen content, climate, agronomic practices, irrigation and soil, respectively. The total incidence of the CF_s is obtained by multiplying the single factors to restrict the weight of the parameters. The HFs are evaluated with respect to the real influence of the nitrogen load. The range between 0 and 5 is specified for HF and that between 0.94 and 1.10 is specified for CF. Finally, the IPNOA is estimated from the HI through the classification of the resultant values based on the percentile of the 135,125 possible combinations, which is scaled between 1 and 6 (Capri et al. 2009).

Therefore, the IPNOA raw data are ranked based on the 135,125 possible combinations into the classes previously mentioned (Capri et al. 2009) according to their hazard level shown in Table 4. A value of 1 represents areas with the least hazard, whereas a value of 6 represents the most hazardous zone.

Implementation of the LR model

The LR model is considered one of the most commonly used multivariate statistical models and is often cited as one of the most efficient data-driven techniques in several applications (Pradhan and Lee 2010; Hossein Mojaddadi et al. 2017). The LR model defines rigid assumptions, which are considered obstacles to the approaches used in this study (Benediktsson et al. 1990). The LR model is also difficult to use in real-life applications. However, statistical approaches based on LR could overcome these obstacles and create an easy approach for analysis that does not need a pre-assumption and can be used with other bivariate statistical analysis (BSA) methods, such as frequency ratio (Ayalew and Yamagishi 2005). Although the multivariate LR method is stronger than the other statistical methods, it still has several disadvantages that hinder the analysis of the classes of each nitrate conditioning factor. To overcome these weaknesses, many studies used LR as bivariate method to solve the problem; however, LR has some limitations in performing BSA as it uses the classes as an indicator and does not consider it in the analysis (Süzen and Doyuran 2004).

The dependent variable in this method is considered a binary variable that represents the absence or presence of nitrate (0 and 1, respectively). The binary model is a model that uses logical expressions to select spatial features from composite feature layers or multiple rasters. The output of the binary model is in binary format, that is, 1 (true) for spatial features that meet the selection criteria and 0 (false) for features that do not. This method also represents the probability in the range of [0, 1] on an S-shaped curve. The use of a few parameters with small pixels is recommended to obtain fast and reliable results (Bai et al. 2012). With the derived logistic coefficients, the probability (p) of nitrate concentration was calculated using Eq. (2):

$$p = 1/(1 + e^{-z}) \tag{2}$$

where p is the probability of flooding, with the value between 0 and 1 on an S-shaped curve. z denotes a linear combination, and it follows that LR involves fitting Eq. (3) to the data, as follows:

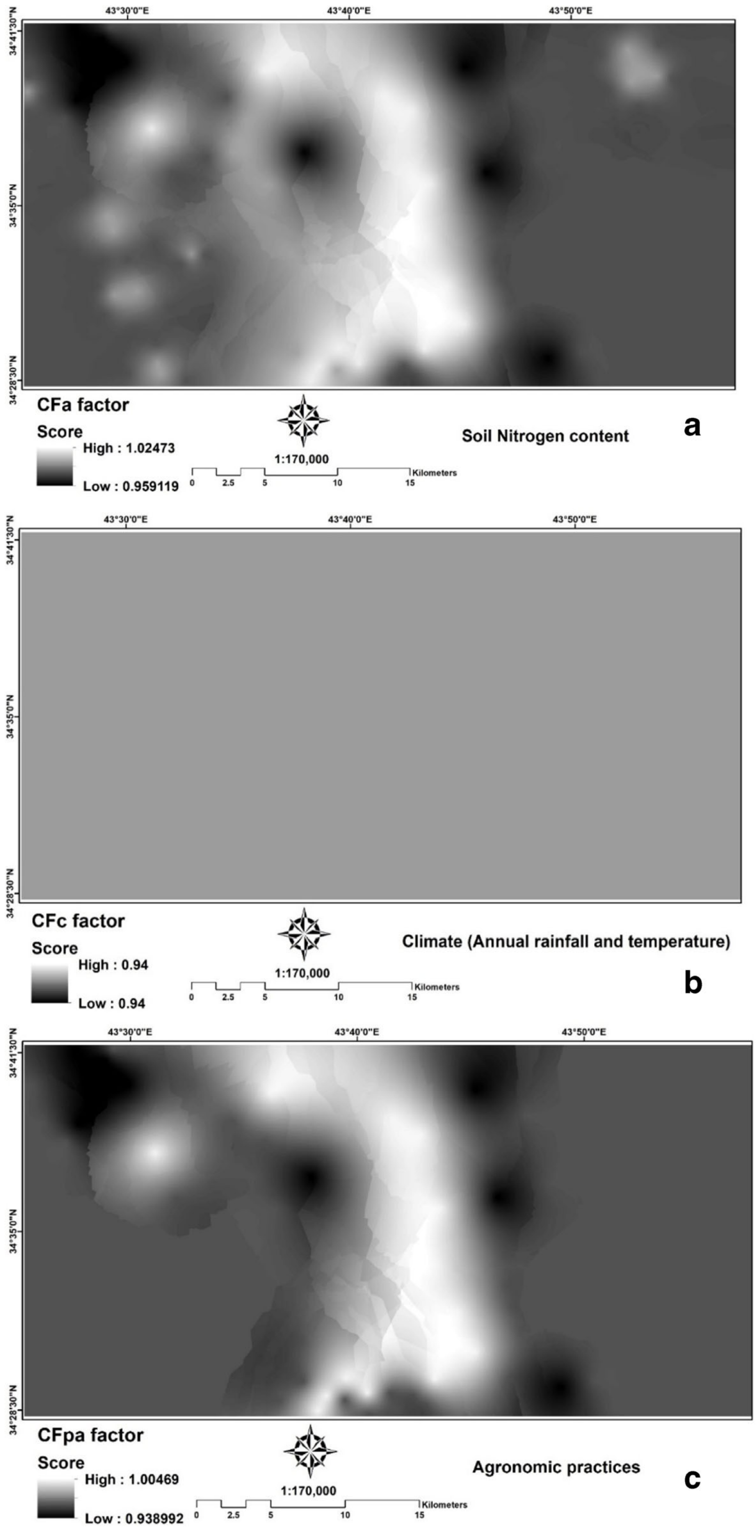


Fig. 6 CFs of **a** soil nitrogen content, **b** climate, **c** agronomic practices, **d** irrigation types, and **e** soil types

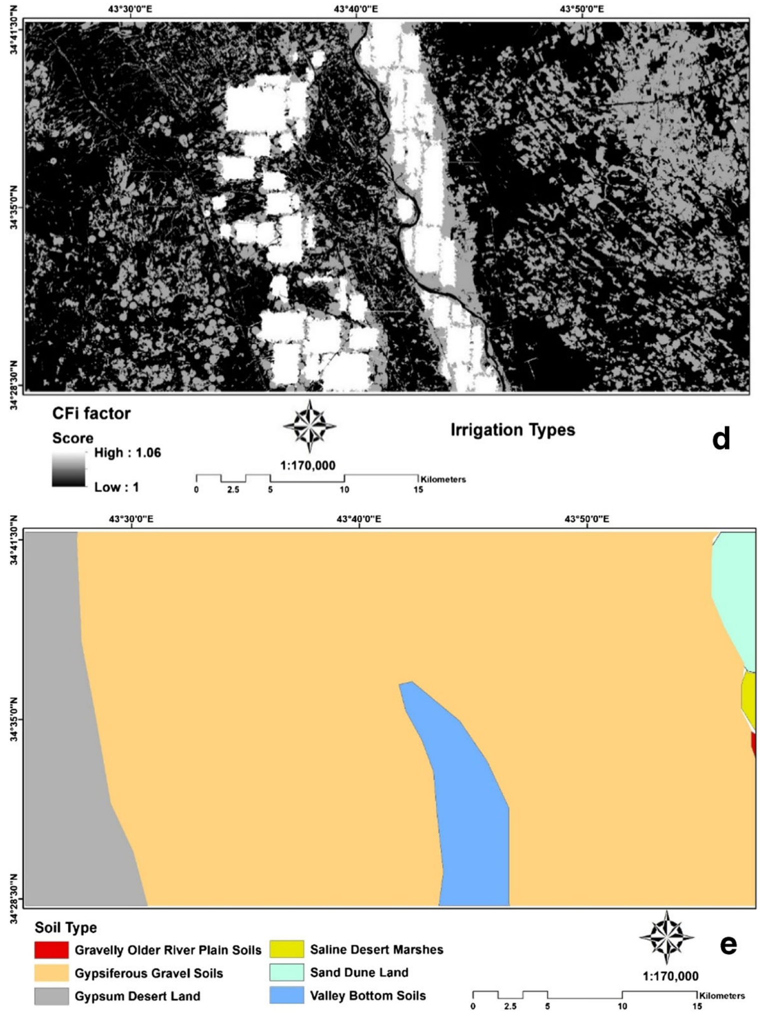


Fig. 6 (continued)

$$z = b_\circ + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \quad (3)$$

where $b_\circ$ is the intercept of the model, b_i ($i = 0, 1, 2, \ldots, n$) represents the coefficients of the LR model and x_i ($i = 0, 1, 2, \ldots, n$) denotes the conditioning factors (Lee and Sambath 2006).

Results and discussions

OBIA-SVM-extracted LULC result

As shown in Table 5, most of the study area is covered by bare land (681.43 km^2), followed by irrigated agricultural land (471.15 km^2). Rain-fed agricultural land is the smallest among all classes, showing that the agricultural industry in Saladin depends on watering and irrigation systems.

The LULC overall accuracy and kappa coefficient were calculated as 87.05% and 0.839, respectively, using the confusion matrix and ground truth points.

Table 4 Hazard indices, IPNOA, and relative classification by Capri et al. (2009)

Hazard index	IPNOA	Classification
2.54–3.18	1	Unlikely (u)
3.19–5.88	2	Very low (vl)
5.89–7.42	3	Low (l)
7.43–9.31	4	Moderate (m)
9.32–11.10	5	High (h)
11.11–17.66	6	Very high (vh)

Table 5 Extracted LULC areas using OBIA-SVM approach

Code	Class name	Area, m^2	Area, km^2
1	Bare land	681,433,748.34	681.43
2	Irrigated agriculture	471,153,764.54	471.15
3	Rain-fed agriculture	3,780,537.50	3.78
4	Urban	61,184,243.34	61.18
5	Water	7,019,962.50	7.02
Total		1,224,872,256.22	1224.87

IPNOA results

When the hazard and control parameters have been extracted, the nitrate concentration index map is calculated using the IPNOA model (Fig. 7a) and classified based on the hazard level scheme, as shown in Fig. 7b.

Correlation analysis was conducted between parameters in different forms and methods, and the p values were used to test the results of statistical analysis efficiency and usually calculated at the 95% significance level.

No parts of this study area were categorised as high or very high hazard, whereas most of the study area is in the safe zone to nitrate hazard (81.54%). However, the moderate zone to nitrate hazard with less than 1 percentage represents the smallest class, as shown in Table 6.

LR optimisation results

Weightage of the conditioning parameters was derived to determine the significance of the sequence of predictors. The rank of each parameter was determined by LR in the statistical software Weka and subsequently

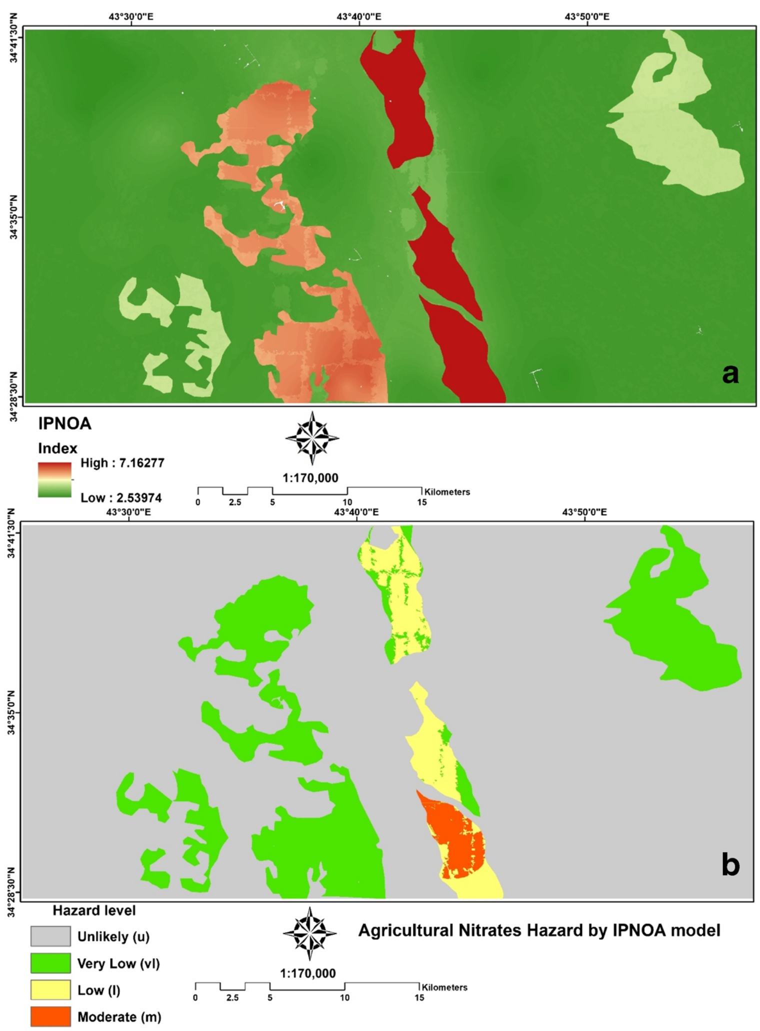


Fig. 7 **a** Nitrate concentration index map and **b** agricultural nitrate hazard map using IPNOA model

Table 6 Area of nitrate hazard classes derived from IPNOA model

Score	Class	Count	Area, m^2	Area, km^2	Percentage
1	Unlikely (u)	39,954,381	998,859,525	998.86	81.54
2	Very low (vl)	6,981,582	174,539,550	174.54	14.25
3	Low (l)	1,577,025	39,425,625	39.43	3.22
4	Moderate (m)	487,012	12,175,300	12.18	0.99
Total			1,225,000,000	1225.00	100.00

overlaid using spatial analyst tools. Figure 8 shows the final weightage derived by the LR model for all conditioning parameters.

From the results of LR modelling shown in Fig. 8, all HFs achieved less than 0.5 weightage, which is considered insignificant to nitrate concentration. Meanwhile, CF_{ap} achieves the highest value (3.28), which is the most significant factor to trigger nitrate concentration. CF_a and CF_s are ranked second and third, with values of 2.35 and 2.13, respectively. From the calculated weightages for each parameter using the statistical LR, the IPNOA model was optimised.

$$\begin{aligned} HI = {} & [(HF_f \times 0.09) + (HFm \times 0.07) + (HFs \times 0.04)] \\ & \times (CF_n \times 2.351) \times (CF_c \times 0.75) \\ & \times \left(CF_{ap} \times 3.287\right) \times (CF_i \times 1.567) \\ & \times (CF_s \times 2.127) \end{aligned} \tag{4}$$

Values were computed by the raster calculator of the ArcGIS software using Eq. (4).

Optimised LR-IPNOA results

For the nitrate concentration index, after calculating the LR coefficients from eight nitrate contributing factors, the optimised IPNOA was calculated (Fig. 9a) and classified based on the hazard level scheme, as shown in Fig. 9b.

The classes with moderate potential hazard (major scores) are in the central part of the plain where the most productive aquifers as well as irrigated agricultural lands are situated and are at the same time potential pollution concentration sources. Figure 9a shows the nitrate concentration index generated from the integrated LR and IPNOA model, and the nitrate index ranges from 2.8 to 8.04 for groundwater. This index indicates the estimated probability of groundwater at each pixel in the presence of a given set of conditioning factors.

The quantified information shown in Table 7 indicates that most of the study area is in an unpolluted area that is unlikely to be affected by the nitrate hazard (45.35%). Meanwhile, approximately 5 percentage of the entire area is in the moderate hazard zone for nitrate concentration.

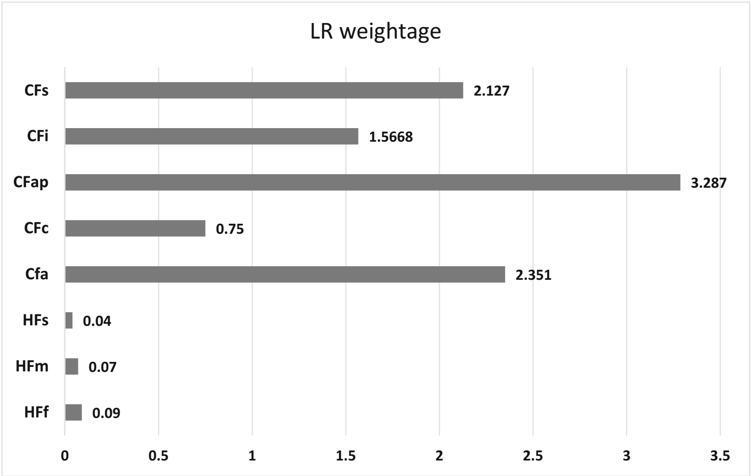


Fig. 8 Calculated weightage for CFs and HFs calculated by LR

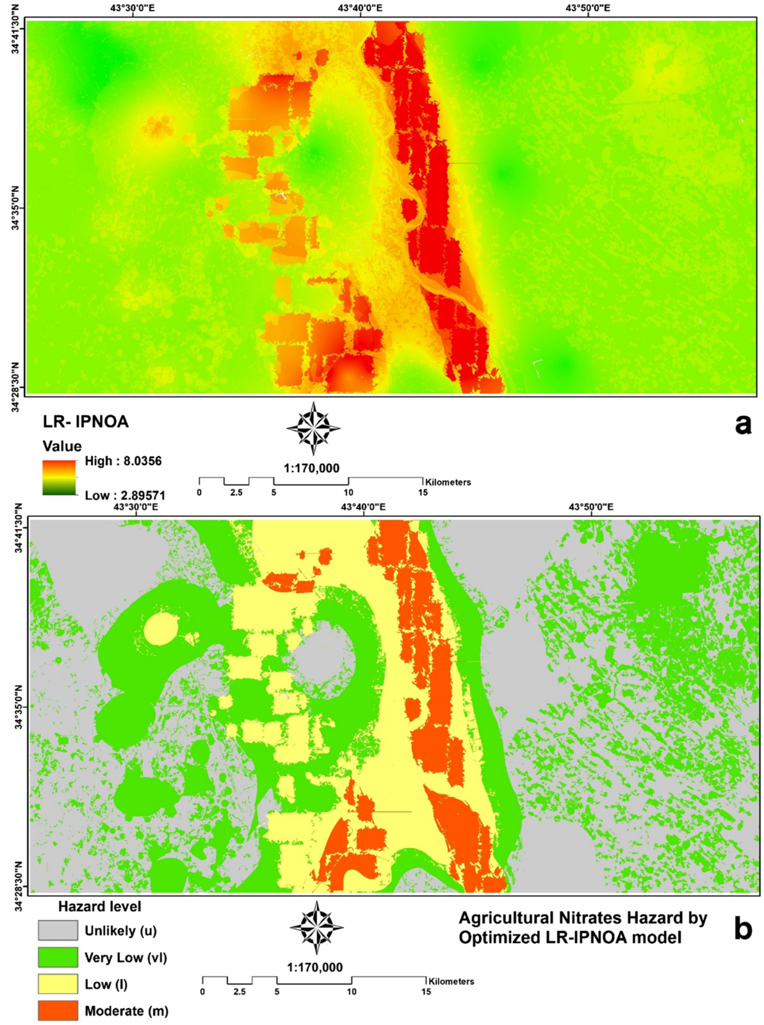


Fig. 9 Map showing **a** nitrate concentration index and **b** agricultural nitrate hazard index using LR-IPNOA model

No parts of this study area were categorised as high or very high hazard, whereas most of the study area is in the safe zone to nitrate hazard (81.54%). However, the moderate zone to nitrate hazard with less than 1 percentage represents the smallest class, as shown in Table 7.

Compared with the results obtained by the standalone IPNOA model, the area of the moderate zone derived using the LR-IPNOA model is 68 km^2 larger. Moreover, nearly 45% of the study area is classified as "unlikely to nitrate hazard" by the LR-IPNOA model, whereas 81%

Table 7 Area of nitrate hazard classes derived from LR-IPNOA model

Score	Class	Count	Area, m^2	Area, km^2	Percentage
1	Unlikely (u)	21,239,319	540,982,975	540.98	45.35
2	Very low (vl)	16,645,577	416,139,425	416.14	33.97
3	Low (l)	7,893,120	197,328,000	197.33	16.11
4	Moderate (m)	3,221,984	80,549,600	70.55	4.58
Total			1,225,000,000	1225.00	100

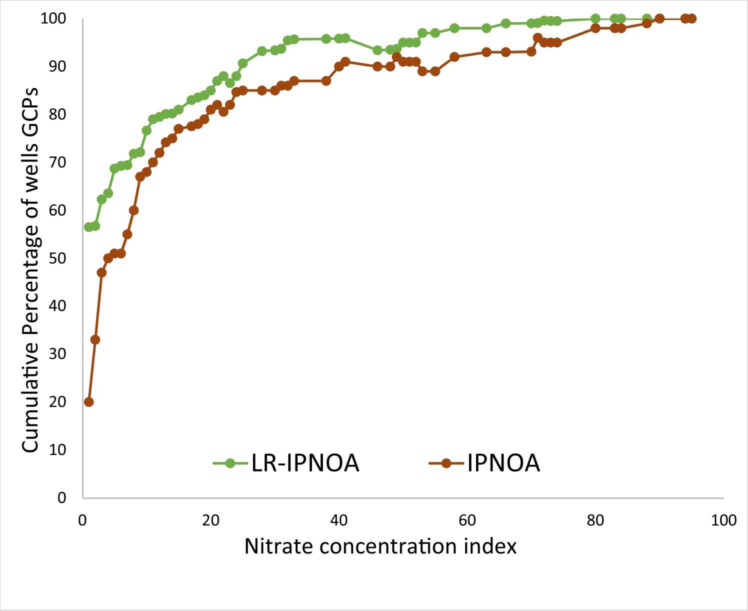


Fig. 10 ROC curve accuracy assessment (standalone IPNOA and integrated LR and IPNOA models)

is classified as the same class by the IPNOA model. The field survey shows that the LR-IPNOA accurately models the nitrate hazard where more than 5 percentage of the study area is affected by nitrate pollution.

Validation

According to Shahid et al. (2000), the existing groundwater borehole wells were used to assess the precision of the GIS model of the anticipated map of groundwater potential. (Pradhan 2009) mentioned that validation is one of the most important processes of modelling, and without validation, the models will be of no scientific significance (Neshat et al. 2014).

The validation of groundwater nitrate concentration was conducted using information extracted from 30% of inventory wells where 56 different wells were sampled over the study area. The nitrate concentrations indices, which were projected by the IPNOA and LR-IPNOA models, were compared with the sampling measurements for the sake of validation. As shown in Fig. 10, the calculated area under the curve shows 85.3% and 91.32% success rates of accuracy for nitrate concentration indices from the IPNOA and LR-IPNOA models, respectively. This result is reasonable for regional nitrate hazard analysis, which indicates that the optimised LR-IPNOA model more accurately determines the nitrate hazard contamination than the standalone IPNOA model.

Conclusion

An approach for the complicated issue of groundwater management was presented in this study. During periods of acute water shortages in the catchment area, groundwater remains as the most viable source of potable drinking water, domestic, irrigation, livestock watering and industrial usage. Therefore, the protection of groundwater is of paramount importance.

For the reduction of uncertainty and skewness, the parametric IPNOA model was empowered by soil and LULC types and optimised by using the data-driven LR model. We examined conditioning factors, such as waste treatment sludge (HF_{fd}), organic (HF_{fo}) and non-organic fertilisers (HF_{fm}), climate (CF_c), nitrogen content in soil (CF_a), irrigation techniques (CF_i), agronomic practices (CF_{pa}) and soil types (HF_s), to determine their degree of contribution to groundwater nitrate hazard. LULC was depicted from high-resolution SPOT-5 satellite image using the OBIA-SVM method with 87.5% overall accuracy.

No part of this study area was categorised as "high" or "very high" hazard area for nitrate concentration. However, the moderate zone to nitrate hazard represented the smallest class. The moderate potential hazard classes were in the irrigated agricultural plain, where the most productive aquifers were situated. The optimised LR-IPNOA predicted a larger area with moderate hazard than the IPNOA model. The field surveys showed that the LR-IPNOA accurately modelled the nitrate hazard concentration, and more than 5% of the study area was affected by nitrate pollution. With regard to receiver operating characteristic assessment, the optimised LR-IPNOA model showed better accuracy than the standalone IPNOA model for regional nitrate hazard analysis (i.e. 91.32% and 85.3%, respectively).

Our results showed that CFs and hazard indices do not contribute equally to nitrate concentration modelling. Thus, IPNOA should be integrated with either machine leaning, deep learning, data-driven or multivariate statistical models to achieve the ideal performance.

This research was necessary because of the rapid development of human activities in the Saladin area, which increase water demand during certain times of the year and/or during periods of drought. This demand often exceeds the supplies stored in reservoirs. However, groundwater pollution remains a challenging phenomenon and must be examined in large scale with several models at different climate conditions.

References

Aal-shamkhi, A. D. S., Mojaddadi, H., Pradhan, B., & Abdullahi, S. (2017). Extraction and modeling of urban sprawl development in Karbala City using VHR satellite imagery. In *Spatial Modeling and Assessment of Urban Form* (pp. 281–296): Springer.

Abdula, R. A. (2016). Stratigraphy and lithology of Naokelekan Formation in Iraqi Kurdistan-review. *The International Journal of Engineering and Science (IJES), 5*(8), 45–52.

Abdulkareem, J. H., Sulaiman, W. N. A., Pradhan, B., & Jamil, N. R. (2018a). Long-term hydrologic impact assessment of non-point source pollution measured through land use/land cover (LULC) changes in a tropical complex catchment. *Earth Systems and Environment, 2*(1), 67–84. https://doi.org/10.1007/s41748-018-0042-1.

Abdulkareem, J. H., Pradhan, B., Sulaiman, W. N. A., & Jamil, N. R. (2018b). Quantification of runoff as influenced by morphometric characteristics in a rural complex catchment. *Earth Systems and Environment, 2*(1), 145–162. https://doi.org/10.1007/s41748-018-0043-0.

Abdullahi, S., Pradhan, B., & Mojaddadi, H. (2017). City compactness: assessing the influence of the growth of residential land use. *Journal of Urban Technology.* 1–26.

AL-Dulaimi, G. A., & Younes, M. K. (2017). Assessment of potable water quality in Baghdad City, Iraq. *Air, Soil and Water Research*, https://doi.org/10.1177/1178622117733441.

Althuwaynee, O. F., Pradhan, B., & Lee, S. (2012). Application of an evidential belief function model in landslide susceptibility mapping. *Computers & Geosciences, 44*, 120–135.

Althuwaynee, O. F., Pradhan, B., Park, H. J., & Lee, J. H. (2014). A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena, 114*, 21–36.

Alwathaf, Y., & El Mansouri, B. (2011). Assessment of aquifer vulnerability based on GIS and ARCGIS methods: a case study of the Sana'a Basin (Yemen). *Journal of Water Resource and Protection, 3*(12), 845–855.

Ayalew, L., & Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology, 65*, 15–31.

Bai, S., Wang, J., Zhang, Z., & Cheng, C. (2012). Combined landslide susceptibility mapping after Wenchuan earthquake at the Zhouqu segment in the Bailongjiang Basin, China. *Catena, 99*, 18–25.

Benediktsson, J. O. N. A., Swain, P. H., & Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing, 28*, 540–552.

Bone, J., Head, M., Jones, D. T., Barraclough, D., Archer, M., Scheib, C., … Voulvoulis, N. (2010). From chemical risk assessment to environmental quality management: the challenge for soil protection. In: ACS Publications.

Boy Roura, M. (2013). Nitrate groundwater pollution and aquifer vulnerability: the case of the Osona region.

Bui, D. T., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., & Trinh, P. T. (2017a). A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology, 233*, 32–44.

Bui, D. T., Tuan, T. A., Hoang, N. D., Thanh, N. Q., Nguyen, D. B., Van Liem, N., & Pradhan, B. (2017b). Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides, 14*(2), 447–458.

Cantor, K. P. (1997). Drinking water and cancer. *Cancer Causes & Control, 8*(3), 292–308.

Capri, E., Civita, M., Corniello, A., Cusimano, G., De Maio, M., Ducci, D., et al. (2009). Assessment of nitrate contamination risk: the Italian experience. *Journal of Geochemical Exploration, 102*(2), 71–86.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena, 151*, 147–160.

DeSimone, L. A., & Howes, B. L. (1998). Nitrogen transport and transformations in a shallow aquifer receiving wastewater discharge: a mass balance approach. *Water Resources Research, 34*(2), 271–285.

Ettazarini, S. (2007). Groundwater potentiality index: a strategically conceived tool for water research in fractured aquifers. *Environmental Geology, 52*(3), 477–487.

Ghiglieri, G., Barbieri, G., Vernier, A., Carletti, A., Demurtas, N., Pinna, R., & Pittalis, D. (2009). Potential risks of nitrate pollution in aquifers from agricultural practices in the Nurra region, northwestern Sardinia, Italy. *Journal of Hydrology, 379*(3–4), 339–350.

Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. *Envrionmental Monitoring & Assessment, 190*, 149. https://doi.org/10.1007/s10661-018-6507-8.

Green, C. T., Fisher, L. H., & Bekins, B. A. (2008). Nitrogen fluxes through unsaturated zones in five agricultural settings across the United States. *Journal of Environmental Quality, 37*(3), 1073–1085.

Gross, E. L. (2008). Ground water susceptibility to elevated nitrate concentrations in South Middleton Township, Cumberland County, Pennsylvania. Shippensburg University of Pennsylvania.

Gumma, M. K., & Pavelic, P. (2013). Mapping of groundwater potential zones across Ghana using remote sensing, geographic information systems, and spatial modeling. *Environmental Monitoring and Assessment, 185*(4), 3561–3579.

Gupta, M., & Srivastava, P. K. (2010). Integrating GIS and remote sensing for identification of groundwater potential zones in the hilly terrain of Pavagarh, Gujarat, India. *Water International, 35*(2), 233–245.

Into, H. D. I. G. (2011). Nitrate in Drinking Water.

Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeology Journal*. https://doi.org/10.1007/s10040-018-1848-5.

Kowal, A., & Polik, A. (1987). Nitrates in groundwater. In *Environmental Technology* (pp. 604–609): Springer.

Lake, I. R., Lovett, A. A., Hiscock, K. M., Betson, M., Foley, A., Sünnenberg, G., et al. (2003). Evaluating factors influencing groundwater vulnerability to nitrate pollution: developing the potential of GIS. *Journal of Environmental Management, 68*(3), 315–328.

Lee, S., & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology, 50*(6), 847–855.

Liao, L., Green, C. T., Bekins, B. A., & Böhlke, J. (2012). Factors controlling nitrate fluxes in groundwater in agricultural areas. *Water Resources Research,* 48(6).

Mezaal, M. R., Pradhan, B., Shafri, H., Mojaddadi, H., & Yusoff, Z. (2017). Optimized hierarchical rule-based classification for differentiating shallow and deep-seated landslide using high-resolution LiDAR data. Paper presented at the Global Civil Engineering Conference.

Min, J. H., Yun, S. T., Kim, K., Kim, H. S., & Kim, D. J. (2003). Geologic controls on the chemical behaviour of nitrate in riverside alluvial aquifers, Korea. *Hydrological Processes, 17*(6), 1197–1211.

Mishra, N., Khare, D., Gupta, K., & Shukla, R. (2014). Impact of land use change on groundwater—a review. *Adv Water Resour Protect, 2*, 28–41.

Mojaddadi, H., Habibnejad, M., Solaimani, K., Ahmadi, M., & Hadian-Amri, M. (2009). An investigation of efficiency of outlet runoff assessment. *Journal of Applied Sciences, 9*(1), 105–112.

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., & Ghazali, A. H. B. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomatics, Natural Hazards and Risk, 1–23.

Nampak, H., Pradhan, B., & Manap, M. A. (2014). Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology, 513*, 283–300.

Neshat, A., & Pradhan, B. (2015). An integrated DRASTIC model using frequency ratio and two new hybrid methods for groundwater vulnerability assessment. *Natural Hazards, 76*(1), 543–563.

Neshat, A., Pradhan, B., Pirasteh, S., & Shafri, H. Z. M. (2014). Estimating groundwater vulnerability to pollution using a modified DRASTIC model in the Kerman agricultural area, Iran. *Environmental Earth Sciences, 71*(7), 3119–3131.

Neshat, A., Pradhan, B., & Javadi, S. (2015). Risk assessment of groundwater pollution using Monte Carlo approach in an agricultural region: an example from Kerman Plain, Iran. *Computers, Environment and Urban Systems, 50*, 66–73.

Oh, H. J., & Pradhan, B. (2011). Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Computers & Geosciences, 37*(9), 1264–1276.

Padovani, L., & Trevisan, M. (2002). *I nitrati di origine agricola nelle acque sotterranee: un indice parametrico per l'individuazione di aree vulnerabili*: Pitagora.

Pradhan, B. (2009). Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. *Open Geosciences, 1*(1), 120–129.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences, 51*, 350–365.

Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software, 25*(6), 747–759.

Raaz Maheshwari, B. R., Yadav, A. S. R. K., & Sharma, S. (2012). Nitrate ion contaminated groundwater: its health hazards, preventive & denitrification measures. *Bull Env Pharmacol Life Scien* Volume, 1, 26–33.

Re, V., Sacchi, E., Kammoun, S., Tringali, C., Trabelsi, R., Zouari, K., & Daniele, S. (2017). Integrated socio-hydrogeological approach to tackle nitrate contamination in groundwater resources. The case of Grombalia Basin (Tunisia). *Science of the Total Environment, 593*, 664–676.

Rizeei, H. M., Pradhan, B., & Saharkhiz, M. A. (2018a). Surface runoff prediction regarding LULC and climate dynamics using coupled LTM, optimized ARIMA, and GIS-based SCS-CN models in tropical region. *Arabian Journal of Geosciences, 11*(3), 53.

Rizeei, H. M., Shafri, H. Z., Mohamoud, M. A., Pradhan, B., & Kalantar, B. (2018b). Oil palm counting and age estimation from WorldView-3 imagery and LiDAR data using an integrated OBIA height model and regression analysis. *Journal of Sensors, 2018*.

Sacco, D., Zavattaro, L., & Grignani, C. (2006). Regional-scale predictions of agricultural n losses in an area with a high livestock density. *Italian Journal of Agronomy, 1*(4), 689–704.

Sacco, D., Offi, M., De Maio, M., & Grignani, C. (2007). Groundwater nitrate contamination risk assessment: a comparison of parametric systems and simulation modelling. *American Journal of Environmental Sciences, 3*, 117–125.

Sener, E., Davraz, A., & Ozcelik, M. (2005). An integration of GIS and remote sensing in groundwater investigations: a case study in Burdur, Turkey. *Hydrogeology Journal, 13*(5–6), 826–834.

Shaban, A., Khawlie, M., & Abdallah, C. (2006). Use of remote sensing and GIS to determine recharge potential zones: the case of Occidental Lebanon. *Hydrogeology Journal, 14*(4), 433–443.

Shahid, S., Nath, S., & Roy, J. (2000). Groundwater potential modelling in a soft rock area using a GIS. *International Journal of Remote Sensing, 21*(9), 1919–1924.

Shahid, S., Nath, S. K., & Maksud Kamal, A. (2002). GIS integration of remote sensing and topographic data using fuzzy logic for ground water assessment in Midnapur District, India. *Geocarto International, 17*(3), 69–74.

Spalding, R. F., & Exner, M. E. (1993). Occurrence of nitrate in groundwater—a review. *Journal of Environmental Quality, 22*(3), 392–402.

Süzen, M. L., & Doyuran, V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental Geology, 45*, 665–679.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology, 504*, 69–79.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology, 512*, 332–343.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena, 125*, 91–101.

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N., & Tehrany, M. S. (2014). Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena, 118*, 124–135.

Vendrusculo, L., Magalhaes, P., & Vieira, S. (2002). *Strategies for soil sampling and multi-layers maps generation through geostatic tools: development of a computational system.* Paper presented at the World Congress of Computers in Agriculture and Natural Resources, Proceedings of the 2002 Conference.

Vikas, C., Kushwaha, R., Ahmad, W., Prasannakumar, V., Dhanya, P., & Reghunath, R. (2015). Hydrochemical appraisal and geochemical evolution of groundwater with special reference to nitrate contamination in aquifers of a semi-arid terrain of NW India. *Water Quality, Exposure and Health, 7*(3), 347–361.

Wick, K., Heumesser, C., & Schmid, E. (2009). Agriculture and nitrate contamination in Austrian groundwater: an empirical analysis. Rollen der Landwirtschaft in benachteiligten Regionen, 19.

Zhao, Y., De Maio, M., & Suozzi, E. (2013). Assessment of groundwater potential risk by agricultural activities, in North Italy. *International Journal of Environmental Science and Development, 4*(3), 286.