Application of logistic regression analysis in prediction of groundwater vulnerability in gold mining environment: a case of Ilesa gold mining area, southwestern, Nigeria

K. A. N. Adiat · B. E. Akeredolu · A. A. Akinlalu ⬤ · G.
M. Olayanju

Abstract Reports of environmental problems occasioned from gold mining activities had prompted the groundwater vulnerability prediction/assessment of the study area. This was with a view to identifying factors responsible for the probability of groundwater contamination as well as developing empirical (LR) model and map that predict the probability of occurrence of contaminant(s) with respect to threshold level in the groundwater resources in the study area. In order to achieve the objectives of the study, logistic regression was applied to independent variables obtained from results of the analysis of remote sensing and geophysical data on one hand and dependent variables obtained from analysis of water samples on the other hand. The results of the analysis obtained from water chemistry established that all the physio-chemical parameters and major metallic ions are within the permissible limit. However, zinc concentration (Zn), being the only dependent variable that had two categorical outcomes, was the contaminant utilized for the study. Similarly, only five (5) independent (predictive) variables, which are percent clay in soil, drainage, slope, unsaturated zone thickness, and total longitudinal conductance, were established to have good correlation and statistically significant with the dependent variable, the contaminant, and thus utilized in logistic regression model development. The quantitative assessment of the developed model established that the overall model prediction accuracy was 85.7% suggesting that the model had a very good fit. The probability prediction model was also accurate and reliable with percentage reliability established to be 90%. In conclusion, it is evident from the results obtained from the study that since the model developed was assessed to be accurate and reliable, the model, and hence the technique, can be replicated in another area of similar geologic condition.

Introduction

Interest in predicting groundwater vulnerability has increased because of widespread detection of contaminants and the implications for human and aquatic health and resources. Report of environmental problems associated with mining communities had prompted the groundwater vulnerability study of basement aquifers in Ilesa gold mining area of southwestern Nigeria. The evaluation of the natural vulnerability of aquifers to contamination is a function of space and time (Civita 1987). In most cases, an accurate prediction of groundwater vulnerability is not feasible due to complexity of groundwater systems. In order to provide accurate and reliable vulnerability prediction in a given area, a

K. A. N. Adiat · B. E. Akeredolu · A. A. Akinlalu (✉) ·
G. M. Olayanju
Department of Applied Geophysics, Federal University of
Technology, Akure, Nigeria
e-mail: ayokunleakinlalu@gmail.com

e-mail: aaakinlalu@futa-edu.ng

suitable model that will account for the sub-surface geology, groundwater flow, and pollutant transport for the area needs to be developed.

A fundamental difficulty in groundwater vulnerability prediction model is the intertwined processes of groundwater flow and pollutant transport, which reflect in the influencing factors (Shih-Kai et al. 2013). Most of these factors are often evaluated by a number of experts using different approaches. It is important to note that the degree of contribution of one or more factors to groundwater vulnerability is not the same and this may also vary from one location to the other. Furthermore, the effects of all the important factors that can influence the groundwater contamination in the area must be integrated to develop a reliable model. Groundwater vulnerability study is a spatial problem that requires data input, processing, and solution from many experts.

A variety of methods have been developed and used for assessing aquifer vulnerability to contaminants (Twarakavi and Kaluarachchi 2005). Previous methods to estimate aquifer vulnerability to contamination may be classified into the following categories: hydrogeological complex and setting (HCS) methods, parametric system or overlay/index methods, numerical or process-based methods, and statistical methods. HCS methods which were developed based on criteria found to be representative of groundwater vulnerability under certain hydrogeological condition (Gogu and Dassargues 2000). The overlay/index models such as the multi criteria decision analysis (MCDA) in the context of analytic hierarchy process (AHP) (Adiat et al. 2012; Adiat et al. 2013; Akinlalu et al. 2017; Adiat et al. 2018) and DRASTIC model (Mohammad 2017; Malik and Shukla 2019; Hassan et al. 2019) are based on combining maps of various physiographic attributes and assigning weights to each attribute to obtain a final score (Connell and Van den Daele 2003; Thapinta and Hudak 2003; Twarakavi and Kaluarachchi 2005). The methods are largely dependent on data availability and expert judgment rather than the controlling physical processes (Twarakavi and Kaluarachchi 2005). Numerical or process-based methods are usually more elaborate than simple overlay or index methods. They require analytical and/or numerical solutions to the governing mathematical equations that represent coupled processes of contaminant transport. (Meeks and Dean 1990; Twarakavi and Kaluarachchi 2005). These methods are computationally costly and demand substantial data. Furthermore, the process-oriented numerical models

also suffer from flaws of being used for site-specific studies and not for evaluating vulnerability on a large scale. All the aforementioned methods suffer from flaws of inability to capture the probabilistic nature or the uncertainty of groundwater vulnerability consequent upon which validation may be inherently impossible for this category of methods that assess vulnerability outside of a probabilistic framework (Worrall 2002). On the other hand, statistical methods are flexible and better suited to accommodate uncertainty in the data than the former methods.

Uncertainty is inherent to predictions of groundwater vulnerability (Loague 1991; Loague et al. 1996), yet few groundwater vulnerability assessments have accounted for, or reported, associated uncertainty. Statistical methods are based on the concept of uncertainty, which is described in terms of probability distributions for the variable of interest (National Research Council NRC 1993). One possible goal in applying statistical methods to vulnerability assessment is to identify variables that can be used to define the probability of groundwater contamination (Burkart et al. 1999). Statistical methods use response variables such as the frequency of contaminant occurrence, contaminant concentration, or contamination probability.

Statistical methods range from simple summary or descriptive statistics of concentrations of targeted contaminants to more complex regression analyses that incorporate the effects of several predictor variables (Worrall 2002; Worrall and Kolpin 2003). A significant benefit of statistical method is that predictions of vulnerability are expressed in probabilistic terms. However, all uncertainty is not inherently represented within the resulting probabilistic predictions because unavoidable model and data errors propagate through its calculations make predictions of vulnerability best estimates. It is therefore reasonable to say that the prediction of groundwater vulnerability is best estimated using statistical approaches because they cater for series of uncertainties and complexities of the hydrogeological environment. Examples of statistical analysis methods utilized in groundwater resources research are cluster analysis, factor analysis, discriminant analysis, regression analysis, fuzzy recognition, and back propagation (BP) neural networks (Gui and Chen 2007; Chen et al. 2013; Adiat et al. 2020).

One of the common statistical methods to estimate aquifer vulnerability is the technique of binary logistic regression or commonly called logistic regression (LR).

LR models relate the probability of a contaminant concentration to exceed a threshold concentration to a set of possible influencing variables. LR analysis is a model structuring technique for modeling and analyzing several variables. LR analysis predicts the probability of a binary or categorical response based on independent or predictive (influencing) variables. LR analysis, with its advantage of being more simple than other analyses and its regression logic, has an important place in categorical data analysis. Therefore, LR is well suited for analysis of groundwater vulnerability assessment because the binary response or categorical response in the case of ordinal logistic regression can be established using a threshold that represents a drinking water standard, laboratory detection level, or relative background concentration (Twarakavi and Kaluarachchi 2005). Often, the objective of a groundwater vulnerability assessment is to predict the occurrence of a water quality constituent above a certain level or threshold. This method allows us to develop an acceptable model, which could define the correlation between dependent (predicted, i.e., contaminant) and independent (predictive) variables in best fit with the least variable. LR has been used by researchers to solve problems related to groundwater studies in different geologic environments in various parts of the world. Twarakavi and Kaluarachchi (2005) used ordinal LR to assess aquifer vulnerability to heavy metals in Washington, USA. Ozdemir (2016) adopted the methodology of LR to map sinkhole susceptibility in Konya, Turkey. Qian et al. (2018) used LR to predict water shortage risk in situations with insufficient data in Beijing, China. Chenini and Msaddek (2019) mapped groundwater recharge susceptibility using LR and bivariate statistical analysis in Tunisia. Kim et al. (2019) used the technique of LR to assess impacts of climate change on a complex river system in South Korea. However, within the context of the literature review done for this study, the application of LR to predict/assess groundwater vulnerability to contamination resulting from gold mining activities in a typical basement complex geologic environment has hitherto not been reported in the current study area. Consequently, attempt would be made to utilize the methodology of LR to predict/assess vulnerability of the aquifer to contaminant(s) in the gold mining area of Ilesa, a typical basement complex of southwestern Nigeria. The Ilesa Schist belt is one of the major schist belts in Nigeria that have been extensively mapped and studied in detail. The belt consists of several occurrences of primary and

alluvial gold workings. (Akinlalu et al. 2018). Gold mining operations started in the area in early 1950s (Makinde et al. 2014). This had resulted to various degree of land degradations (Adeoye 2016) and groundwater contamination (Makinde et al. 2016). The objectives of the study are the following:

i   generate factors/parameters (independent variables) that can be used to predict aquifer contamination if there is any
j   identify factors (dependent variable(s)) responsible for the probability of groundwater contamination
k   develop empirical (LR) model and map that predict the probability of occurrence of contaminant(s) (identified in ii above) with respect to threshold level in the groundwater resources in the study area and
l   quantify the prediction accuracy and reliability of the model developed.

Study area description

The study area is located in the south-western part of Ilesa, Osun state, Nigeria. It lies between longitude 4° 38′ 0″ E and 4° 43′ 0″ E and latitude 7° 31′ 30″ N and 7° 36′ 0″ N (Fig. 1). The area is sparsely inhabited, and most of the economic activities engage by the inhabitants are agriculture and mining. Numerous minerals such as gold (Au), lead (Pb), iron (Fe), nickel (Ni), cadmium (Cd), chromium (Cr), copper (Cu), zinc (Zn), and manganese (Mn) had been reported by the Nigeria Geological Survey Agency (NGSA) to be deposited in the area (Adekoya et al. 2003). The Ilesa Schist belt of southwestern Nigeria has complex geology and mineralization potential. The study area is located in one of the major schist belts in Nigeria and has been extensively mapped and studied in detail; others are Maru, Anka, Zuru, Kazaure, Kusheriki, Zungeru, Kushaka, Iseyin, Oyan, and Iwo schist belts. The belt consists of several occurrences of primary and alluvial gold workings. The primary gold commonly occur in quartz veins within several lithologies, and the host rocks to the veins include fine-grained mica schists, amphibolite schists, talc tremolite schists, and several varieties of gneisses (Akinlalu et al. 2018). Gold mining operations started in the study area in early 1950s (Makinde et al. 2014). More than fifty mining sites located in various parts of the study area were visited. Most of these mining pits

were open-pit, and the average depth of the mining pits was 3.4 m, while an estimate of 25.8 ha of land was degraded in the entire mining sites (Adeoye 2016).

In terms of structural features, lithology, and mineralization, the schist belts of Nigeria show considerable similarities to the Achaean green stone belts (Rahaman 1989; Olusegun et al. 1995). The area is known to have variable metamorphic mineral assemblages ranging from green schist—to amphibolite—facies (Ajibade et al. 1987). Four major rock types are present in the area, and these are the amphibolite and the amphibolite schist, the undifferentiated migmatite gneiss, the quartzite, and the quartz schist (Fig. 1). Geology is an important factor that controls groundwater accumulation in an environment especially in terms of quality and quantity. The schist belts, which form part of the Precambrian basement rock units, are notable for clay-rich weathered horizons. The degree of fracturing and weathering of rocks influence the rate of percolation and infiltration.

The topography of the area varies from heavily forested mountains, and gently rolling hills to a vast stream/river coastal plain. The topographic elevation of the area ranges from 278 and 490 m above mean sea level. The drainage pattern of the area is largely dendritic typical of highly fractured bedrock with flat and undulating terrain.

Methodology

The study was undertaken in two phases which include the data acquisition/processing phase and assessment of groundwater vulnerability through the application of logistic regression phase. The research utilizes the integration of ancillary data, water sample, remote sensing data, and subsurface geophysical data to derive dependent and independent variables. Logistic regression techniques were applied to the results obtained from the analysis of these data to develop groundwater vulnerability prediction models with a view to selecting a final model based on maximization of test statistics.

Data acquisition and processing techniques

The ancillary data utilized for the study were the geological map, soil distribution map, and the boreholes information of the available wells drilled across the area. These ancillary data were processed to extract the geological map, soil distribution map, and the boreholes

information of the well drilled across the study area. The geological map and soil distribution maps were georeferenced, clipped to required boundary and digitized.

The soil distribution map was categorized based on the two soil associations present in the area. The remote sensing data utilized for the study were the Landsat ETM image, Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER), and digital elevation model (DEM) image. The lineaments and drainage were extracted from the LANDSAT-TM images, while DEM was used for producing the slope map of the area. The remote sensing data were processed using ArcGis 10.1, Envi 4.5, and PCI Geomatica 2012. Computer-assisted methods for the detection of structural lineaments were exclusively based on edge enhancement or spatial filtering techniques (directional and/ or gradient filters). These methods produced edge maps requiring further processing for lineament segments to appear with one-pixel thickness. Optimal edge detectors, e.g., the Canny algorithm (Canny 1986), have already been successfully applied on natural scenes with satisfactory results. A composite band combination was used (Süzen and Toprak 1998). Directional filtering and edge sharpening enhancement algorithm of PCI Geomatica were utilized to extract the lineament for analyses (Abdullah et al. 2010). Slope was extracted from DEM using the slope algorithm of ArcGis 10.1. The density of the lineaments and the drainage were obtained by dividing the summation of the total lengths of the lineaments and drainage by the coverage area of the environment under consideration respectively (Adiat et al. 2012, 2013). Krigging technique was used to produce the lineament and drainage density maps.

Electrical resistivity data were acquired using the Ohmega Terrameter and its accessories. A total of seventy (70) Vertical Electrical Sounding (VES) stations were occupied (Fig. 1). The Schlumberger array was adopted with electrode spacing (AB/2) ranging from 1 to 100 m. The coordinates of measurement stations were taken using Garmin GPS 7.0. The data acquired were processed and plotted. Quantitative analysis, involving partial curve matching and computer iterations, using win RESIST software developed by Vander Velpen 1998, was adopted to determine the geo-electric characteristics of the study area. From this information, aquifer resistivity, aquifer thickness, unsaturated zone thickness, total longitudinal conductance of the unsaturated zone, and total transverse resistance of the unsaturated
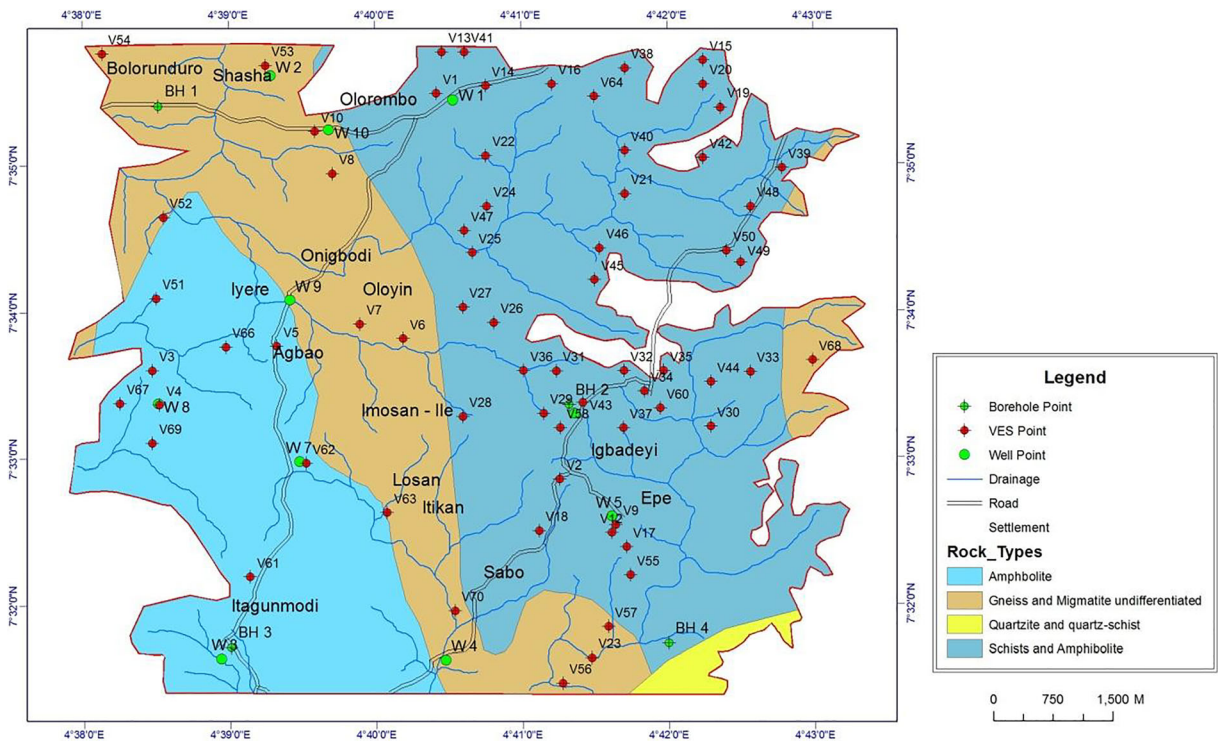
Fig. 1 Geological map of the study area showing borehole/well and VES locations (modified after geological map of Ilesa SW Sht. 243)

zone were estimated. The aquifers were identified by using resistivity range of the subsurface layer as the criteria. This was however guided by the well information obtained from the area. The unsaturated zone thickness was calculated using the summation of the thickness of the overlapping layers. The longitudinal conductance (S) and transverse resistance (TR) of the unsaturated zone were calculated from the results of resistivity data using Eqs. 1 and 2 below:

$$S = {}^h/_\rho \quad S = \sum_{i=1}^{n} \frac{h_i}{\rho_i} \tag{1}$$

Transverse unit resistance (TR) was determined from the layer parameters as (1):

$$TR = \rho * h \quad TR = \sum_{i=1}^{n} \rho_i * h_i \tag{2}$$

where ρ_i and h_i are resistivity and thicknesses of ith layer, respectively.

A total of ten (10) domestic drinking water wells were randomly collected from water sources available at the mining sites and their host communities (Fig. 1). The depths of the wells vary from 10 to 15 m. It is also important to add that all the wells tap water from

localized unconfined aquifer. The water samples were collected on April 21, 2016. A plastic bottle (2 l) was washed with dilute HCl acid of 0.5 mol/dm^3 and rinse with distilled water. These samples, stored in a distilled plastic bottles, were taken to the laboratory for analysis to determine the safety or otherwise of the groundwater resources of the area. In the laboratory, the samples were digested for water quality test. Physiochemical parameters test was performed on all the water samples. The following physiochemical parameters were tested: temperature, turbidity, conductivity, pH, chloride, total hardness, sulphate, nitrate, phosphate, total solids, total dissolved solids, total suspended solids, and total alkalinity. In addition to these parameters, some inorganic metals (Na, K, Ca, Mg, Zn, Fe, and Cu) were also tested. Also, atomic absorption spectrometer test was conducted on the samples to test for the presence of heavy metals such as Cd, Mn, and Pb. The tests were conducted at the Central Research Laboratory of Federal University of Technology, Akure, Nigeria. In order to determine whether the water in the study area was contaminated or safe for consumption, the water quality results obtained were compared with maximum permissible levels for safe drinking water by Nigerian Standard for Drinking Water Quality Threshold values guideline (NSDWQ) 2007. Kurtosis and Spearman's rank correlation analysis were employed to

determined non-normality of the physiochemical parameters and major metallic ions obtained from the water samples and relationship between the two input variables at the two-tailed significance (i.e., $\alpha = 0.05$) level. The results of the analysis will produce the dependent variables that will be utilized for groundwater vulnerability modeling.

Methodologies/steps of logistic regression as adopted in the study

The concept and procedures of logistic regression require several steps to be conducted and this has been explained in detail in Park (2013). Some of these steps found to be suitable to the nature and structure of the data set adopted for this study are presented as follows:

1   Examination of the basic assumptions of logistic regression which include:

  (a) Binary categorization of dependent variable and
  (b) Examination of the non-normality of the dependent variables and relationship between the dependent and independent variables

2   Development of logistic regression prediction model
3   Statistical assessment of the prediction model developed which involves (a) model significance; (b) results for the Hosmer–Lemeshow goodness-of-fit test statistic, R-square values, and model accuracy; and (c) assessment of the reliability of the prediction model.
4   Groundwater vulnerability prediction map and
5   Validation of the groundwater vulnerability prediction map

   The statistical package for social scientists (SPSS) was used for the statistical analysis.

Results and Discussions

Independent variables utilized in groundwater vulnerability modeling

The results of the ancillary data are discussed based on the independent variables utilized in groundwater vulnerability modeling. The result of the percentage of clay and particle size distribution present in each soil association was adopted to establish the top soil characteristics of the study area (Ogunsanwo 1989). Two types of soil series (Itagunmodi and Egbeda series) are obtainable in the study area.

   The borehole records show that there are two aquifer systems in the area and these are unconfined aquifer and confined aquifer. The depth of occurrence of the unconfined aquifer ranges from 10 to 15 m, and the depth of occurrence of the confined aquifer is at 30–40 m. It was however observed that most of the hand dug wells in the study area terminate in the unconfined aquifer layer, while the boreholes terminate in the confined aquifer layer. Therefore, hand dug wells are more susceptible to groundwater contamination than the borehole in the area.

   The independent variables obtained from the remote sensing data are lineament, drainage, and slope representing geomorphological parameters that influence groundwater vulnerability.

   The distribution of the lineaments in the study area concentrated in the southern and western parts of the study area, with few lineaments in the northern and eastern parts of the area. The study area is relatively dense in terms of lineament, and the lineament is denser in the eastern and central parts of the study location (Fig. 2). Groundwater of the area with high lineament and lineament density is relatively vulnerable to surface contaminants due to secondary porosity and permeability developed by the lineament features.

   The system of the drainage is largely dendritic typical of structurally controlled drainage along the sheared zone of metamorphic rock. The drainage system in an area is strictly dependent on the slope, the nature/attitude of bedrock, and the regional as well as local fractures pattern. The study area is well drained. Area of high drainage density is indicative of area with a relative poor groundwater infiltration (Fig. 3). This implies that groundwater in area with high drainage density is not vulnerable to surface contaminant. The dominant direction of the drainage pattern in the area is southeast–northwest direction. This suggests that the river/stream is structurally controlled. Four classes of slope obtained in the area are 0–2, 2–8, 8–15, and 15–30 representing flat, undulating, rolling, and moderately steep classifications, respectively (Adiat et al. 2012). The study area is largely characterized with flat to undulating slope, having small amount of runoff and high amounts of infiltration. Areas with low slope tend to retain water for long periods of time. This favors infiltration of water
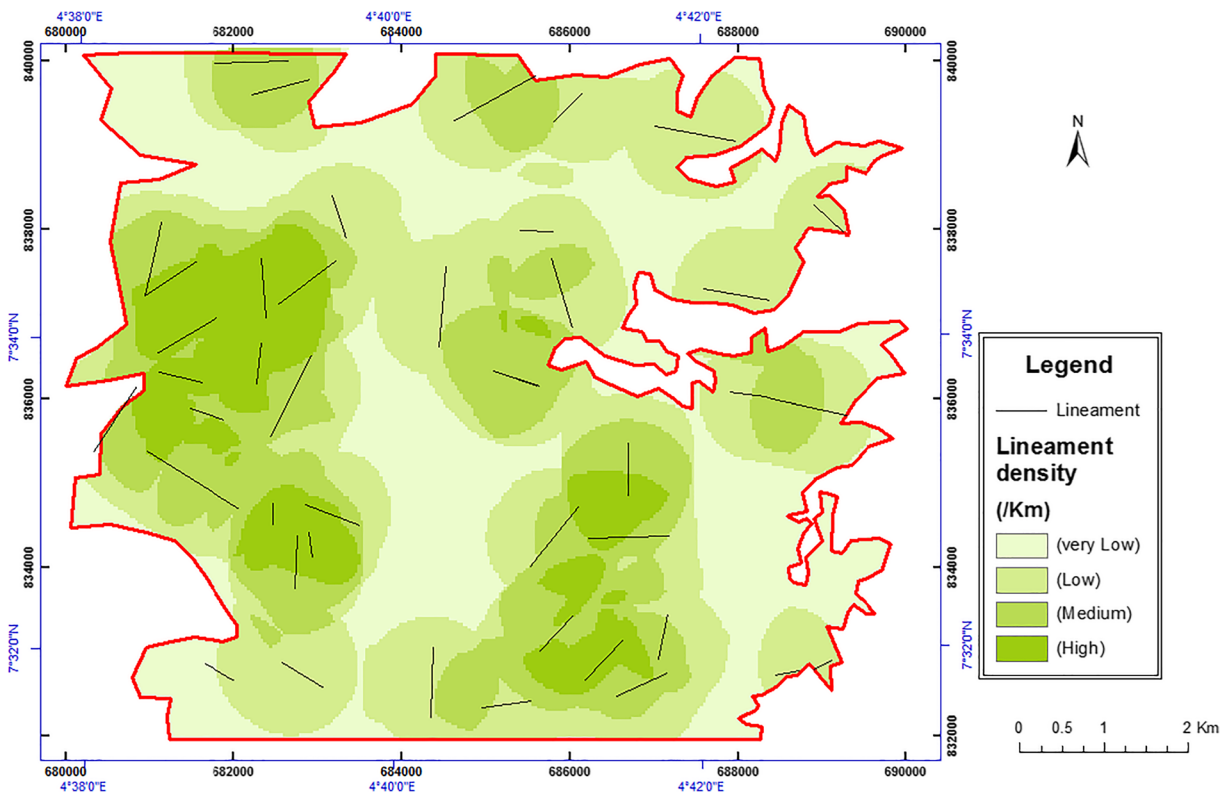
Fig. 2 Lineament density map of the study area

recharge and contaminant migration. Therefore, the flat to undulating slope characterizing the study area suggests that groundwater in most of the area is relatively vulnerable to groundwater contamination.

The geophysical parameters that influence groundwater vulnerability as obtained from the results of the interpretation of VES are the unsaturated zone thickness, aquifer resistivity, aquifer thickness, longitudinal conductance, and transverse resistance. Based on the depth of occurrence or thickness of the unsaturated zone, the aquifers in the area can be categorized into shallow and deep aquifers with thicknesses ranging between 1.2–10 m and 10.1–42.8 m, respectively. Deep seated aquifers are characterized by high thickness of unsaturated zone. Groundwater in the deep seated aquifers are more protected because the contaminants will take a longer time before they percolate into the aquifer, whereas the shallow seated aquifers are more vulnerable to groundwater contamination because the contaminants will percolate within a very short time.

In the study area, three aquifer types were identified. The aquifer media were delineated based on the

resistivity value of the geo-electric layers obtained from the study. The resistivity ranges of 67–150 Ωm, 150–600 Ωm, and 600–859 Ωm were classified as weathered basement, fractured basement, and partly weathered basement aquifers, respectively. The aquifers thicknesses vary between 1.2 and 42.8 m. In general, the larger the thickness of the aquifer, the higher the transmissivity of the aquifer media. Consequently, the greater the pollution potential. The unsaturated zone layer constitutes the main protective unit.

Total longitudinal conductance and total transverse resistance of the unsaturated zone helped us to characterize the study area. Total longitudinal conductance map was grouped into four vulnerability classes based on the model of Antonio and Richard (2014). The four classes obtained are < 0.1, 0.1–0.3, 0.3–0.7, and 0.7–2.5 representing extreme, high, moderate, and low classifications, respectively. Areas with low longitudinal conductance value have high permeability and are more vulnerable. The study area is mainly characterized with extreme to high vulnerability class (Fig. 4).

The total transverse resistance of the unsaturated zone of study area was classified into high and low
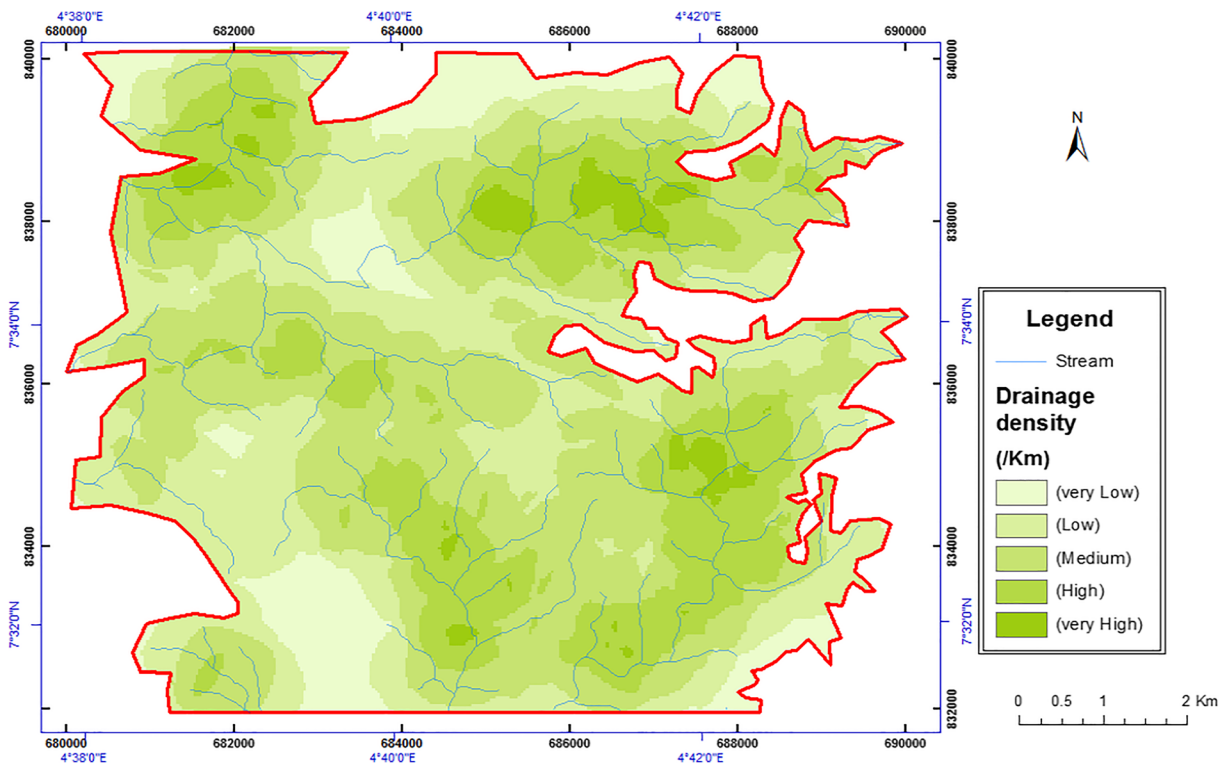
Fig. 3 Drainage density map of the study area

transverse resistance areas. Total transverse resistance value above 1000 Ωm^2 is classified as high transverse resistance, while values less than 1000 Ωm^2 are classified as low transverse resistance. High total transverse resistance dominated the entire study area with exception of some few pockets of low total transverse resistance at the central part of the study area. Areas with high total transverse resistance are classified as areas of low infiltration, due to their low permeability. Consequently, these areas are less vulnerable to surface contaminant.

The results of the analysis obtained from water chemistry laboratory are presented in Table 1. Physicochemical parameters evaluated were temperature, turbidity, conductivity, pH, chloride, total hardness, sulphate, nitrate, phosphate, total solids, total dissolved solids, total suspended solids, and total alkalinity. In addition to these, major metal concentrations, which include sodium, calcium, magnesium, iron, and heavy metals that include copper, zinc, cadmium, lead, and manganese, were also evaluated. The results of the major and heavy metals analysis obtained from the water chemistry are presented in Table 2. The results

were compared with the maximum permissible level for safe drinking water established by the Nigerian Standard for Drinking Water Quality Threshold values guideline (NSDWQ) 2007, to determine which of the physiochemical parameters and major and heavy metals present in the water samples had exceeded the maximum permissible level. It was observed that all the physiochemical parameters are within the permissible limit. The results of the comparison of major and heavy metals with maximum permissible level are presented in Table 3. The table shows that all water samples containing Mg, Cd, and Pb exceeded the permissible limit except where they were not detected.

All water samples containing Na, Ca, Fe, Cu, and Mn are within the permissible limit. On the other hand, some of the water samples containing Zn are within the permissible limit while some exceeded the permissible limit. This implies that zinc concentration (Zn) is the only dependent variable that had two categorical outcomes. It also established that there is relationship between the mining activities and high zinc concentration in the areas.
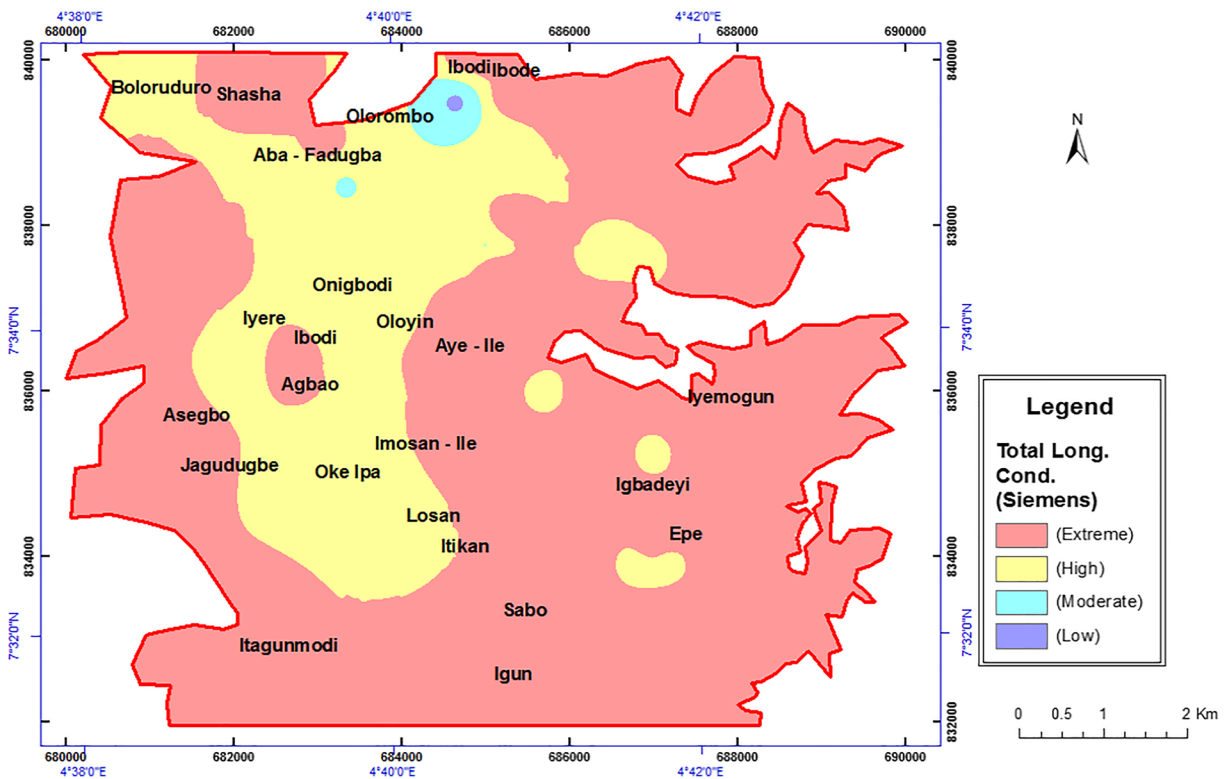
Fig. 4 Total longitudinal conductance map of the study area

Results of logistic regression as adopted in this study are as follows

Results of binary categorization of dependent variable

In logistic regression model development, two categorical outcomes of the dependent variable must be satisfied. From the results presented in Table 3, zinc concentration is the only dependent variable that satisfied the condition of an outcome variable with two possible categorical outcomes binarily categorized as 0 and 1 (Table 4). Thus, zinc concentration was selected to be the dependent (predicted) variable (i.e., the contaminant that would be utilized for the regression model development). The convention is to associate 1 with "success" (i.e., vulnerability test is passed; zinc concentration maximum permitted level is not exceeded), and 0 with "failure" (i.e., vulnerability test is failed; zinc ion concentration maximum permitted level is exceeded) as presented Table 4.

Results of examination of the non-normality of the dependent variables and relationship between the dependent and independent variables

The results of non-normality tests for physico-chemical parameters and major ion concentration show that all the kurtosis values deviated from zero; this indicates that the datasets are not normally distributed. This makes them applicable in logistic regression modeling. The non-normality implies that the relationship between the independent and dependent variables is non-linear. It is important to emphasis that non-linear relationship between independent and dependent variables is one of the assumptions of logistic regression (Park 2013).

It further implies that major ion concentration present in the water samples are not from the same aquifer system, and these established the disjointed relationship between the aquifer systems in the study area. It also depicts the non-parametric nature of the groundwater system in the study area. Since none of the data were normally distributed, the Spearman's rank correlation

Table 1 Physio-chemical parameters obtained from the samples collected from the study area on April 16, 2016

Well	Temperature (°C)	Turbidity (NTU)	Conductivity (μohms/cm)	pH	Chloride (mg/L)	Total hardness (mg/L)	Sulphate (mg/L)
W1	26	0.02	570	7.85	19.88	50	19.5
W2	26	0.03	720	8.01	8.88	66	19.65
W3	26	0.03	500	6.98	19.7	60	19.8
W4	26	0.04	510	7.5	31.95	96	12.5
W5	25	0.04	790	8.04	11.01	38	19.6
W6	26	0.03	520	7.9	11.54	170	19.8
W7	25	0.04	460	6.92	11.01	34	9.3
W8	25	0.03	730	7.12	19.53	100	14.5
W9	25	0.04	630	8.32	11.01	34	18.5
W10	26	0.03	540	7.57	7.99	46	12.5

Well	Nitrate (mg/L)	Phosphate (mg/L)	Total solids (mg/L)	Total dissolved solids (mg/L)	Total suspended solids (mg/L)	Total alkalinity (mg/L)
W1	6.5	6.2	155.2	140.3	14.9	112
W2	6.35	5.8	160.5	148.7	11.8	100
W3	6.2	6.6	112.6	98.4	14.2	84
W4	6.2	3.5	85.3	65.8	19.5	70
W5	6.3	5.8	113.2	96.5	16.7	60
W6	6.2	6.5	51.5	32.22	19.28	65.5
W7	3.3	3.5	150.5	132.3	18.2	111
W8	6.2	4.3	113.2	96.5	16.7	60
W9	6.2	5.4	87.6	69.8	17.8	68
W10	6.2	3.5	138.6	108.6	30	92

coefficient measure was used to determine the relationship between the dependent variable and each of the independent (predictive) variables. The result obtained from the Spearman's rank correlation shows that five (5) independent (predictive) variables (percent clay in soil, drainage, slope, unsaturated zone thickness, and total

Table 2 Major metals and heavy metal concentration obtained from the samples collected from the study area on April 16, 2016

	Major metal concentration (mg/L)				Heavy metal concentration (mg/L)				
	Na	Ca	Mg	Fe	Cu	Zn	Cd	Pb	Mn
W1	20	54.3	42.4	0.13	0.02	2.6	nd	nd	nd
W2	18.7	43.3	40.2	0.11	nd	2.3	0.43	0.13	nd
W3	21.3	38.5	50.3	0.13	nd	3.5	0.50	0.10	nd
W4	18.6	42.3	38.5	0.12	nd	3.5	nd	nd	nd
W5	20.3	53	35.5	0.13	nd	3.5	0.42	0.13	nd
W6	19.4	47.8	37.3	0.11	nd	2.6	0.40	0.10	nd
W7	18.5	37.5	45	0.1	nd	2.5	nd	nd	nd
W8	20.5	43.5	52.2	0.08	nd	3.5	nd	nd	nd
W9	20.3	50.3	45.5	0.12	0.01	3.5	0.4	0.1	nd
W10	19.3	45.5	37.6	0.11	nd	2.7	nd	nd	nd

nd not detected

Table 3 Comparison result of major and heavy metal concentration obtained from the water samples collected from the study area with maximum permitted level

		Major metal concentration (mg/L)				Heavy metal concentration (mg/L)				
Well samples	Max. Permitted level	Na	Ca	Mg	Fe	Cu	Cd	Zn	Pb	Mn
		200	150	0.02	0.3	1	0.003	3	0.01	0.2
	W1	20.0✓	54.3✓	42.4✗	0.13✓	0.02✓	nd	2.6✓	nd	nd
	W2	18.7✓	43.3✓	40.2✗	0.11✓	nd	0.43✗	2.3✓	0.13✗	nd
	W3	21.3✓	38.5✓	50.3✗	0.13✓	nd	0.50✗	3.5✗	0.10✗	nd
	W4	18.6✓	42.3✓	38.5✗	0.12✓	nd	nd	3.5✗	nd	nd
	W5	20.3✓	53✓	35.5✗	0.13✓	nd	0.42✗	3.5✗	0.13✗	nd
	W6	19.4✓	47.8✓	37.3✗	0.11✓	nd	0.40✗	2.6✓	0.10✗	nd
	W7	18.5✓	37.5✓	45.0✗	0.10✓	nd	nd	2.5✓	nd	nd
	W8	20.5✓	43.5✓	52.2✗	0.08✓	nd	nd	3.5✗	nd	nd
	W9	20.3✓	50.3✓	45.5✗	0.12✓	0.01✓	0.40✗	3.5✗	0.10✗	nd
	W10	19.3✓	45.5✓	37.6✗	0.11✓	nd	nd	2.7✓	nd	nd

nd not detected

✓ = indicates the heavy metals that passed in terms of relationship to maximum permissible level i.e falls below maximum limit

✗ = indicates the metals that failed i.e concentration exceeds maximum permissible limit

longitudinal conductance) have good correlation with the dependent variable (zinc concentration). Their respective correlation coefficients at two-tailed significance (i.e., $\alpha = 0.05$) are $-0.699, -0.047, -0.009, -0.535$, and 0.817. This makes them statistically significant, and consequently, they will be utilized in logistic regression model development.

Table 4 Binary categorization of zinc concentration with the maximum permitted level as the threshold for categorization

Zn concentration (mg/L)

Max. permitted level	3	Binary codes
W1	2.6✓	1
W2	2.3✓	1
W3	3.5✗	0
W4	3.5✗	0
W5	3.5✗	0
W6	2.6✓	1
W7	2.5✓	1
W8	3.5✗	0
W9	3.5✗	0
W10	2.7✓	1

✓ = implies that the Zn concentration was below the maximum permitted level

✗ = implies that the Zn concentration had exceeded the maximum permitted level

Results of logistic regression prediction model development

The final model has the following independent variables as members of its group: total longitudinal conductance, unsaturated zone thickness, slope, percent of clay in soil, and drainage as presented in the model in Eqs. 3 and 4.

$$\text{Logit}(p) = \ln\left(\frac{P}{1-P}\right) = \alpha + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 \quad (3)$$

Therefore,

$$P = \frac{e^{(\alpha + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5)}}{1 + e^{(\alpha + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5)}} \quad (4)$$

The constant (intercept) of the prediction model is 21.323, and the gradient coefficient of each predictive variable "b_i" is the log odds obtained for the independent variables of the final model. The log odds coefficients of total longitudinal conductance, unsaturated zone thickness, slope, percent of clay in soil, and drainage are 193.397, -2.481, -2.193, 25.156, and -11.933, respectively. From the log odd coefficient, each independent variable contribution to measure of variation of the dependent variable was estimated. Substituting these values shown in Eq. 4 give

Table 5 Probability prediction of zinc concentration not exceeding 3 mg/L in groundwater sample of the study area using the model in Eq. 5

Logistic Regression Constant (α) =21.232	Total Longitudinal conductance	unsaturated zone thickness	Slope	Percentage clay in soil	Drainage Density	Probability that zinc concentration ≤ 3mg/L utilizing all the variables (x_1, x_2, x_3, x_4, and x_5)
Gradient Coefficient	$(b_1) = 193.397$	$(b_2) = -2.481$	$(b_3) = -2.193$	$(b_4) = 25.196$	$(b_5) = -11.933$	
Independent variables	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	p
W1	0.49	20.94	3.84	0.74	0.32	1.00
W2	0.01	2.15	2.48	0.53	1.73	0.91
❖ W3	0.05	7.67	6.05	0.74	1.77	0.00
❖ W4	0.02	3.81	8.36	0.53	1.73	0.00
W5	0.06	4.84	8.10	0.53	1.47	0.69
W6	0.03	4.99	4.36	0.74	1.61	0.88
W7	0.29	22.04	2.37	0.53	1.47	0.99
❖ W8	0.01	15.57	1.73	0.53	1.55	0.00
❖ W9	0.11	11.79	3.40	0.53	1.57	0.02
W10	0.08	2.77	2.15	0.74	0.92	1.00

❖ - passed

$$P = \frac{e^{(21.323+193.397x_1-2.481\ x_2-2.193\ x_3+25.156x_4-11.933x_5)}}{1+e^{(21.323+193.397x_1-2.481\ x_2-2.193\ x_3+25.156x_4-11.933x_5)}}$$

$$(5)$$

The value of predictive variables for each well point was substituted in Eq. 5, and the result of the probability prediction (p) of dependent variable (zinc concentration) not exceeding 3 mg/L in groundwater sample of the study area is presented in Table 5. If the p results obtained (last column of Table 5) is approximately equal to 1 (i.e., $0.5 \leq p \leq 1.0$) as obtained in W1, W2, W5, W6, W7, and W10, it implies that the zinc concentration was below the maximum permitted level (i.e., zinc concentration $\leq$ 3 mg/L). On the other hand, if the p results obtained (as shown in Table 5) is less than 1 (i.e., $0 \leq p \leq 0.4$) as obtained in W3, W4, W8, and W9, it implies that the zinc concentration was above the maximum permitted level (i.e., zinc concentration $\geq$ 3 mg/L).

Also, odds ratio of each independent variable was calculated by using the regression coefficient of the independent variables "b" as the exponent or exp (b).

$$\text{Odds ratio} = \exp^{(bi)} \qquad (6)$$

The odds ratios of total longitudinal conductance, unsaturated zone thickness, slope, percent of clay in soil, and drainage are 9.800e8, 0.084, 0.112, 8.416e10,

and 0.002, respectively. The significance of the odd ratio can be expressed in terms of the change in odds. When the independent variable increases by one unit, the odds that the case can be predicted increase by a factor of odds ratio times, when other variables are controlled. Therefore, increase in values of total longitudinal conductance and percent of clay in soil will significantly increase the odds of the groundwater sample not exceeding 3 mg/L by factors of 9.800e8 and 8.416e10, respectively; also, increase in unsaturated zone thickness and slope will slightly increase the odds of the groundwater sample not exceeding 3 mg/L by factors of 0.084 and 0.112, respectively. While increase in the values of drainage will significantly decreases the odds of the groundwater sample not exceeding 3 mg/L by factors of 0.002.

Results of statistical assessment of the developed prediction model

Results of model significance

Statistical assessments utilized to assess the predicted model are presented in Table 6. All values for the significant test for the model were statistically significant at $\alpha = 0.05$ level of significance. The Wald chi-

Table 6 Various statistical assessments utilized to assess the predicted model

Variables in model	Model significance		Hosmer–Lemeshow goodness-of-fit test statistic		R-squared values		Percentage of prediction accuracy
	Wald chi-square	P value	Chi-square	P value	Cox and Snell R square	Nagelkerke R square	Overall percentage
Total longitudinal conductance	1.07	.030	0.54	0.99	0.65	0.87	85.7
Unsaturated zone thickness	1.13	.029					
Slope	0.42	.052					
Percent of clay in soil	0.61	.035					
Drainage	0.41	.052					

square values of total longitudinal conductance, unsaturated zone thickness, slope, percent of clay in soil, and drainage were 1.07, 1.13, 0.42, 0.61, and 0.41, respectively, while the P values obtained were respectively 0.03, 0.029, 0.052, 0.035, and 0.052 indicating that all the independent variables are statistically significant ($P \le 0.05$); i.e., the independent variable has a significant effect.

Results of Hosmer–Lemeshow goodness-of-fit test statistic, R-square values, model accuracy

The model had 0.99 P value associated with the Hosmer–Lemeshow goodness-of-fit test. This value, being greater than 0.05, indicates that the estimates for the model fit the original data at an acceptable level. The R-square values (Cox and Snell R square and Nagelkerke R square) for the models were 0.65 and 0.87, respectively, indicating that the model had a moderately strong predictive power. The overall model prediction accuracy was 85.7%, meaning that the model had a good fit. Due to the model's satisfactory assessment and hence, strong prediction capability, the model was chosen as the final model for the study. Based on this level of reliability, the model can be used to predict the probability of zinc concentration above or below 3 mg/L in area that water samples were not taken, having knowledge of the independent variables in the area.

Results of the assessment of the reliability of the prediction model

Tables 4 and 5 are used to explain the results of the accuracy assessment of the model developed. Whenever the vulnerability test is passed (i.e., the Zn concentration maximum permitted level is not exceeded, as shown in the second column of Table 4), the value of the p, shown in the last column of Table 5, is expected to be approximately equal to one (i.e., $0.5 \le p \le 1.0$). If vulnerability test is failed (i.e., the Zn concentration maximum permitted level is exceeded, as shown in the second column of Table 4), the value of the p, shown in the last column of Table 5, is expected to be approximately less than one (i.e., $0 \le p \le 0.4$).

It was observed from the Table 5 that zinc concentration value obtained showed agreement with the model predicted probability that Zn concentration ≤ 3 mg/L in nine out of the ten locations. The disagreement observed at location W5 (Table 4) may be due to other hydrogeological factors, which though may not be significant in the final model, but might contribute to high zinc concentration being greater than 3 mg/L in the groundwater. On this basis, the probability prediction model is not only accurate but also reliable with percentage reliability of 90%.

Results of the groundwater vulnerability prediction map

The study area was gridded to grid size of 500 m with the center of each grid being used as the measuring point for the grid. The values of independent variables for each grid point were estimated and substituted into the model equation to obtain predicted probability used to produce the zinc concentration probability prediction (groundwater vulnerability prediction) map shown in Fig. 5. High concentration of zinc (i.e., above permissible level) typical of contamination dominated the
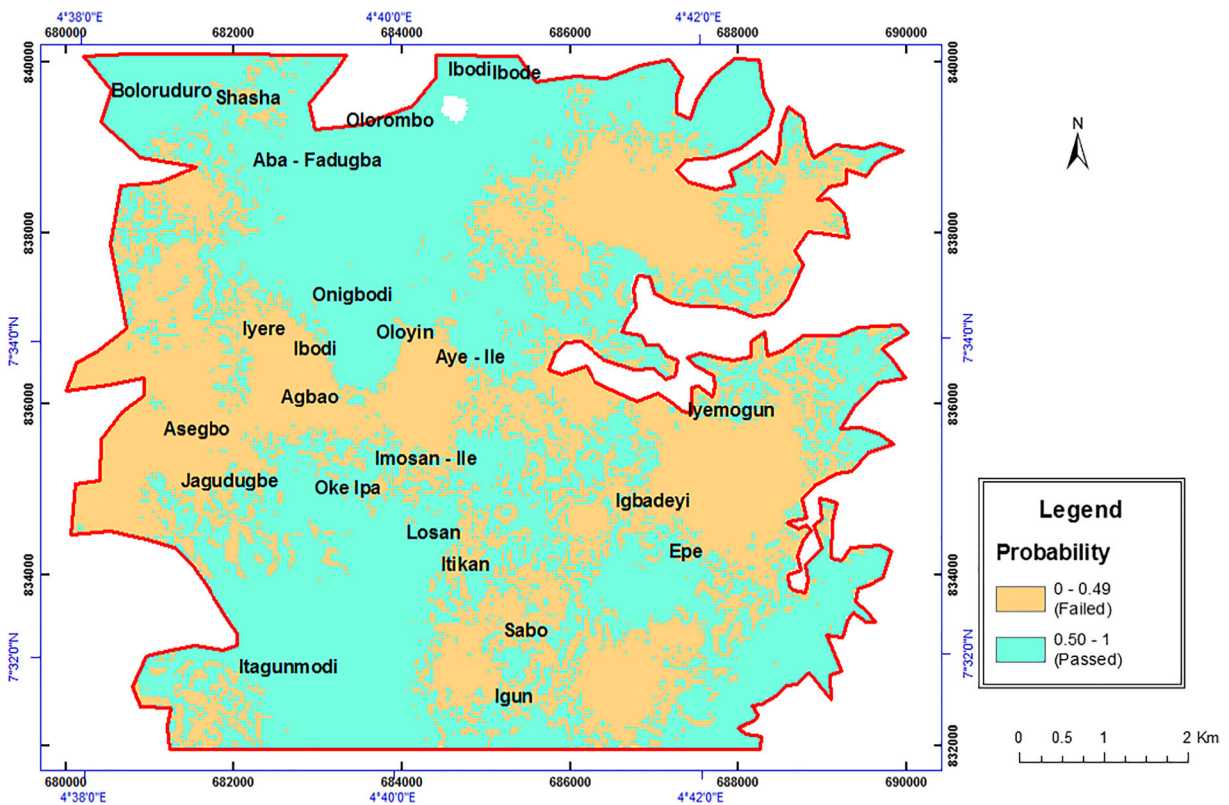
Fig. 5 Zinc concentration probability prediction (groundwater vulnerability prediction) map of the study area

eastern, western, central, south-western, and north-eastern part of the study area. It was observed that most of the parts dominated by high zinc concentration being predicted by the model are communities where gold mining activities are taken place.

Results of the validation of groundwater vulnerability prediction map

The validation of the predictive model was achieved by using independent variables associated with a given location within the study area to predict for groundwater quality of the location. Imagine a hydrogeological system at well "W6" whose total longitudinal conductance was 0.0279, percent of clay in the soil was 0.74, the unsaturated zone thickness was 4.99 m, the slope of the area was 4.358, and the drainage density of the area was 1.6. In order to examine whether or not the groundwater quality would pass the test for zinc concentration permitted level (i.e., belong to category 1 or 0), the values of the independent variables for the location (i.e., "W6") are substituted into the model equation

thus obtain:

$$P = \frac{e^{(21.323+193.397x_1-2.481\ x_2-2.193\ x_3+25.156x_4-11.933x_5)}}{1 + e^{(21.323+193.397x_1-2.481\ x_2-2.193\ x_3+25.156x_4-11.933x_5)}} \quad (7)$$

$$P = \frac{e^{(21.323+193.397(0.0279)-2.481\ (4.99)-2.193\ (4.358)+25.156(0.74)-11.933(1.6)\ )}}{1 + e^{(21.323+193.397(0.0279)-2.481\ (4.99)-2.193\ (4.358)+25.156(0.74)-11.933(1.6))}} \quad (8)$$

$$P = \frac{e^{(21.323+5.395-12.380-9.557+18.615-19.093\ )}}{1 + e^{(21.323+5.395-12.380-9.557+18.615-19.093\ )}}$$

$$= \frac{e^{(4.303)}}{1 + e^{(4.303\ )}} = 0.98 \quad (9)$$

Therefore, the probability that groundwater quality of well "W6" passed the test for zinc concentration permitted level is 98%, or 98% of such independent variables will be expected to produced groundwater quality that passed the test for zinc concentration based on the threshold of the maximum permitted levels of inorganic concentration for safe drinking water by Nigerian
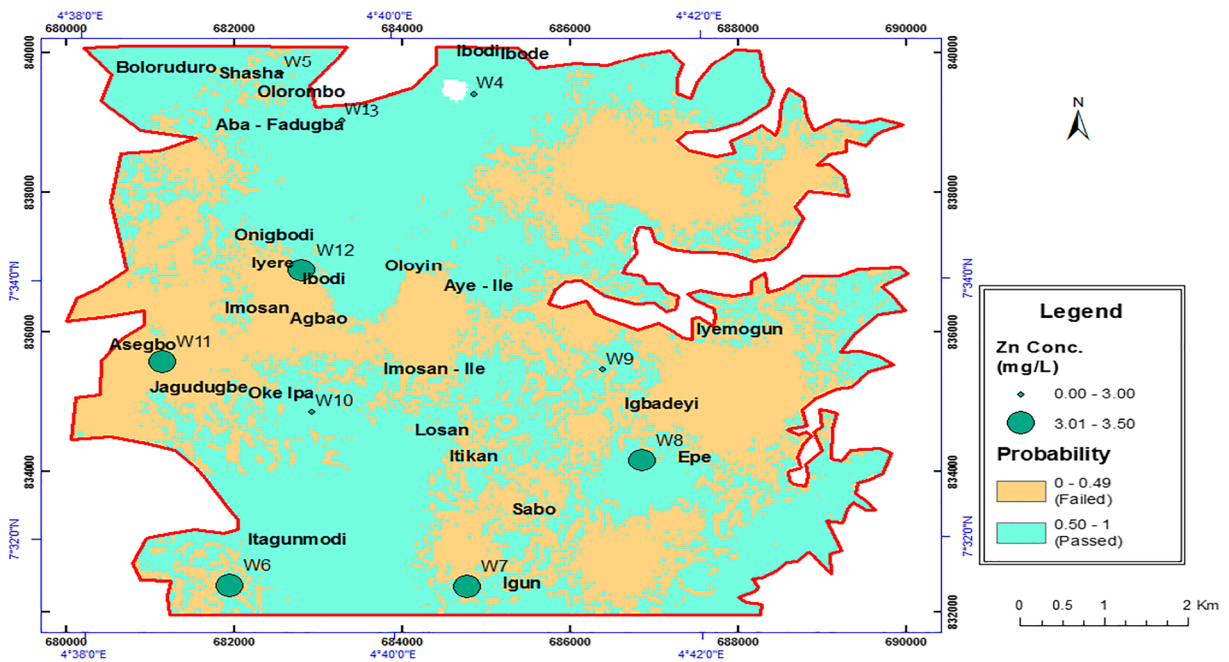
Fig. 6 Zinc concentration probability prediction map with binary categorizes of zinc concentration of the study area

Standard for Drinking Water Quality (NSDWQ) 2007 (Fig. 6).

Also, for well "W4" whose total longitudinal conductance was 0.023, percent of clay in the soil was 0.53, the unsaturated zone thickness was 3.806 m, the slope of the area was 8.356, and the drainage density of the area was 1.728. Substituting these values in model equation, we get

$$P = \frac{e^{(21.323+193.397(0.023)-2.481\ (3.806)-2.193\ (8.356)+25.156(0.53)-11.933(1.728)\ )}}{1 + e^{(21.323+193.397(0.023)-2.481\ (3.806)-2.193\ (8.356)+25.156(0.53)-11.933(1.728)\ )}} \quad (10)$$

$$P = \frac{e^{(21.323+4.448-12.380-9.443+13.333-20.620\ )}}{1 + e^{(21.323+4.448-12.380-9.443+13.333-20.620\ )}}$$

$$= \frac{e^{(-3.337\ )}}{1 + e^{(-3.337\ )}} = 0.03 \quad (11)$$

It therefore implies that the probability that groundwater quality of well "W4" passed the test for zinc concentration permitted level is 3%, or 3% of such explanatory variables will be expected to produced groundwater quality that passed the test for zinc concentration permitted level. Therefore, the groundwater quality of well "W4" failed the test for zinc concentration based on the threshold of the maximum permitted levels of inorganic concentration for safe drinking water

by Nigerian Standard for Drinking Water Quality (NSDQW) 2007 (Fig. 6).

Conclusions

Reports of environmental problems associated with mining communities had prompted the groundwater vulnerability study of the Ilesa gold mining area in Ilesa schist belt, southwestern Nigeria. The objectives of the study were to generate factors/parameters that can be used to predict aquifer contamination in the area, identify which of the factors generated is/are responsible for the probability of groundwater contamination, develop empirical (LR) model and map that predict the probability of occurrence of contaminant(s) with respect to threshold level in the groundwater resources in the study area, and quantify the prediction accuracy and reliability of the model developed. In order to achieve the objectives of the study, the integration of remote sensing, geophysical method, and chemical analysis of water samples was undertaken. Data management and result integration were carried out in GIS environment. The concept of logistic regression was applied to the results obtained to develop groundwater vulnerability model for the area. Analysis of remote sensing and geophysical data assisted in generating factors/parameters

(independent variables) that can be used to predict aquifer contamination in the area; these factors include lineament/lineament density, drainage/drainage density, slope, rock types (geology/lithology), and soil type association obtainable in the area, aquifer resistivity and thickness, longitudinal conductance, transverse resistance, and coefficient of anisotropy. On the other hand, analysis of water samples assisted in generating the dependent variables (contaminants) utilized in the study. Of all the dependent variables, zinc concentration (Zn) was the only variable that had two categorical outcomes, since two categorical outcomes of dependent variable(s) are a necessary condition for logistic regression model development; Zn was the contaminant utilized for the study. Similarly, only five (5) independent (predictive) variables, which are percent clay in soil, drainage, slope, unsaturated zone thickness, and total longitudinal conductance, were established to have good correlation and statistically significant with the dependent variable, the contaminant, and thus utilized in logistic regression model development. The quantitative assessment of the developed model established that the overall model prediction accuracy was 85.7% suggesting that the model had a very good fit. The probability prediction model was also accurate and reliable with percentage reliability established to be 90%. In conclusion, it is evident from the results obtained from the study that since the model developed was assessed to be accurate and reliable, the model, and hence the technique, can be replicated in another area of similar geologic condition.

References

Abdullah, A., Akhir, J. M., & Abdullah, I. (2010). The extraction of lineaments using slope image derived from digital elevation model: case study: Sungai Lembing-Maran Area, Malaysia. *Journal of Applied Sciences Research, 6*(11), 1745–1751.

Adekoya, J.A., Kehinde-Philips O.O., & Odukoya, A.M. (2003). Geological distribution of mineral resources in Southwestern Nigeria. Prospects for investment in Mineral Resources of Southwestern Nigeria, *A.A. Elueze (ed.)* pp 1–13.

Adeoye, N. O. (2016). Land degradation in gold mining communities of Ijesaland, Osun state, Nigeria. *GeoJournal. Vol. 81, No. 4 (2016), pp. 535–554.*

Adiat, K. A. N., Ajayi, O. F., Akinlalu, A. A., & Tijani, I. B. (2020, *2020*). Prediction of groundwater level in basement complex terrain using artificial neural network: a case of Ijebu-Jesa, southwestern Nigeria. *Applied Water Science, 10*(8). https://doi.org/10.1007/s13201-019-1094-6.

Adiat, K. A. N., Nawawi, M. N. M., & Abdullah, K. (2012). Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool-a case of predicting potential zones of sustainable groundwater resources. *Journal of Hydrology, 440-441*, 75–89. https://doi.org/10.1016/j.jhydrol.2012.03.028.

Adiat, K. A. N., Nawawi, M. N. M., & Abdullah, K. (2013). Application of multi-criteria decision analysis to geoelectric and geologic parameters for spatial prediction of groundwater resources potential and aquifer evaluation. *Pure Applied Geophysics, 170*(2013), 453–471. https://doi.org/10.1007/s00024-012-0501-9.

Adiat, K. A. N., Osifila, A. J., Akinlalu, A. A., & Alagbe, O. A. (2018). Mining of geophysical data to predict groundwater prospect in a basement complex terrain of southwestern Nigeria. *International Journal of Scientific & Technology Research, 7*(5) ISSN 2277–8616.

Ajibade, A. C., Woakes, M., & Rahaman, M. A. (1987). Proterozoic crustal development in the Pan-African regime of Nigeria. In A. Kroner (Ed.), *Proterozoic lithoshperic evolution* (Vol. 17, pp. 259–271). *Washington DC*: American Geophysical Union.

Akinlalu, A. A., Adegbuyiro, A., Adiat, K. A. N., Akeredolu, B. E., & Lateef, W. Y. (2017). Application of multi-criteria decision analysis in prediction of groundwater resources potential: A case of Oke-Ana, Ilesa area southwestern, Nigeria. *NRIAG Journal of Astronomy and Geophysics, 6*(2017), 184–200.

Akinlalu, A. A., Adelusi, A. O., Olayanju, G. M., Adiat, K. A. N., Omosuyi, G. O., Anifowose, A. Y. B., & Akeredolu, B. E. (2018). Aeromagnetic mapping of basement structures and mineralisation characterisation of Ilesa Schist Belt, southwestern Nigeria. *Journal of African Earth Sciences, 138*(2018), 383–391.

Antonio, C. d. O. B., & Richard, F. F. (2014). Natural vulnerability assessment to contamination of unconfined aquifers by longitudinal conductance - (S) method. *Journal of Geography and Geology, 6*, 4.

Burkart, M. R., Kolpin, D. W., Jaquis, R. J., & Cole, K. J. (1999). Agrichemicals in ground water of the Midwestern USA: relations to soil characteristics. *Journal of Environmental Quality, 28*, 1908–1915.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE transactions on pattern analysis and machine Intelligence, 8*, 679–714.

Chen, S., Fu, X. F., Gui, H. R., & Sun, L. H. (2013). Multivariate statistical analysis of the hydro-geochemical characteristics for mining groundwater: A case study from Baishan mining, northern Anhui Province, China. *Water Practice and Technology, 8*, 131–141.

Chenini, I., & Msaddek, M. H. (2019). Groundwater recharge susceptibility mapping using logistic regression model and bivariate statistical analysis. *Quarterly Journal of*

Engineering Geology and Hydrogeology, 53(167–175), 13–175. https://doi.org/10.1144/qjegh2019-047.

Civita, M. V. (1987). La previsione e la prevenzione del rischio di inquinamento delle acque sotterranee a livello region-ale mediante le carte di vulnerabilità, [Forecasting and prevention of groundwater contamination risk at a regional level using vulnerability maps] *In Proceedings Conf. Inquinamento delle acque sotterranee: Previsione e prevenzione*, 9-17.

Connell, L. D., & Van den Daele, G. (2003). A quantitative approach to aquifer vulnerability mapping. *Journal of Hydrology, 276*, 71–88.

Gogu, R. C., & Dassargues, A. (2000). Sensitivity analysis for the EPIK method of vulnerability assessment in a small karstic aquifer, southern Belgium. *Hydrogeology Journal, 8*(3), 337–345.

Gui, H. R., & Chen, L. W. (2007). *Hydrogeochemistric evolution and discrimination of groundwater in mining district.* Beijing: Geological Publishing House.

Hassan, M., Islam, M. A., Hasan, M. A., Alam, M. J., & Peas, M. H. (2019). Groundwater vulnerability assessment in Savar upazila of Dhaka district, Bangladesh - A GIS-based DRASTIC modeling. *Groundwater for Sustainable Development, 9.* https://doi.org/10.1016/j.gsd.2019.100220.

Kim, D., Chun, J. A., & Choi, S. J. (2019). Incorporating the logistic regression into a decision-centric assessment of climate change impacts on a complex river system. *Hydrology and Earth System Sciences, 23*, 1145–1162. https://doi.org/10.5194/hess-23-1145-2019.

Loague, K. (1991). The impact of land use on estimates of pesticide leaching potential: Assessments and uncertainties. *Journal of Contaminant Hydrology, 8*(2), 157–175.

Loague, K., Bernknopf, R. L., Green, R. E., & Giambelluca, T. W. (1996). Uncertainty of ground-water vulnerability assessment for agricultural regions in Hawaii: review. *Journal of Environmental Quality, 25*(3), 475–490.

Makinde, O. W., Adesiyan, T. A., Olabanji, I. O., Tunbosun, I. A., Ogundele, K. T., Adelowotan, O., et al. (2016). Assessment of heavy metals bioavailability in stream sediments around a gold mining environment in south-western Nigeria. *Journal of International Environmental Application & Science, 11*(2), 139–147.

Makinde, W. O., Oluyemi, E. A. , & Olabanji, I.O. (2014). Assessing the impacts of gold mining assessing the impacts of gold mining operations on river sediments and water samples from Ilesa west local government area ILESA West Local Government Area of Osun State, Nigeria *E3S Web of Conferences 1, 41010.* published by EDP Sciences.

Malik, M. S., & Shukla, J. P. (2019). Assessment of groundwater vulnerability risk in shallow aquifers of Kandaihimmat watershed, Hoshangabad, Madhya Pradesh. *Journal Geological Society Of India, 93*, 199–206.

Meeks, Y. J., & Dean, J. D. (1990). Evaluating groundwater vulnerability to pesticides. *Journal of Water Research, 116*(5), 693–707.

Mohammad, A. H. (2017). Assessing the groundwater vulnerability in the upper aquifers of Zarqa River Basin, Jordan using DRASTIC, SINTACS and GOD methods. *International Journal of Water Resources and Environmental Engineering, 9*(2), 44–53.

National, Research Council NRC. (1993). *Groundwater vulnerability assessment: contamination potential under conditions of uncertainty.* Washington, DC: National Academy Press.

Nigerian, Standard for Drinking Water Quality (2007). Nigerian, Standard for Drinking Water Quality (NSDWQ). *NIS 554: ICS, 13.060.20*, pp 14–16.

Ogunsanwo, O. (1989). Some geotechnical propertiees of two laterite soils compacted at different energies. *Bulletin of Engineering Geology, 26*, 261–269.

Olusegun, O., Kehinde-Phillips, & Gerd, F. T. (1995). The mineralogy and geochemistry of the weathering profiles over amphibolite, anthophillite and talc-schists in Ilesa Schist Belt, southwestern Nigeria. *Journal of Mining and Geology, 31*(1), 53–62.

Ozdemir, A. (2016). Sinkhole susceptibility mapping using logistic regression in Karapınar (Konya, Turkey). *Bulletin of Engineering Geology and the Environment. Issue, 2.*

Park, H.-A. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing, 43*(2), 154–164.

Qian, L., Zhang, R., Bai, C., Wang, Y, & Wang, H. (2018). An improved logistic probability prediction model for water shortage risk in situations with insufficient data. *Manuscript under review for journal Nat. Hazards Earth Syst. Sci. Discuss.,* https://doi.org/10.5194/nhess-2018-56 *(assessed on the 7th June, 2020).*

Rahaman, M.A. (1989). Recent advances in the study of the basement complex of Nigeria. *In: Oluyide, P.O. (Ed.), Precambrian geology of Nigeria. Geological Survey of Nigeria*, pp. 11–43.

Shih-Kai, C., Cheng-Shin, J., & Yi-Huei, P. (2013). Developing a probability-based model of aquifer vulnerability in an agricultural region. *Journal of Hydrology, 486*(2013), 494–504.

Süzen, M. L., & Toprak, V. (1998). Filtering of satellite images in geological lineament analyses: an application to a fault zone in Central Turkey. *International Journal of Remote Sensing, 19*, 1101–1114.

Thapinta, A., & Hudak, P. (2003). Use of geographic information systems for assessing groundwater pollution potential by pesticides in Central Thailand. *Environmental International, 29*, 87–93.

Twarakavi, N. K. C., & Kaluarachchi, J. J. (2005). Aquifer vulnerability assessment to heavy metals using ordinal logistic regression. *Ground Water, 43*(2), 200–214.

Vander Velpen, B. P. A. (1998). *Win RESIST version 1.0 resistivity depth sounding interpretation software.* Delft: M. Sc Research Project, ITC.

Worrall, F. (2002). Direct assessment of groundwater vulnerability from borehole observations. In: Hiscock, K., Rivett, M.O., Davison, R.M. (Eds.), Groundwater sustainability. Geological Society of London Special Publication, 193, 245–254.

Worrall, F., & Kolpin, D. W. (2003). Direct assessment of groundwater vulnerability from single observations of multiple contaminants. *Water Resource Research, 39* (art. no. 1345).