International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 3, May-June 2022

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Application of Electrical Resistivity Techniques for Mapping Potential Groundwater Pollution Zones Using Random Forest Machine-Learning in Agbabu, Southwestern Nigeria

[1*]Alagbe Olufemi Adigun, [2]Iwueze Emeto Stephen

[1,2]Department of Applied Geophysics, Federal University of Technology, Akure Nigeria
[1]oaalagbe@futa.edu.ng
[2]iwuezees@gmail.com
*Corresponding author: oaalagbe@futa.edu.ng

ABSTRACT

The potential groundwater pollution zones within Bitumen impregnated Agbabu community an agglomerated town in Odigbo Local Government was mapped using electrical resistivity methods with Machine learning regression to produce a potential vulnerability map of the area. The Electrical resistivity techniques consisted of2D wenner imaging andvertical electrical sounding (VES)techniques were conducted along the established. Water samples from wells within the area were subsequently collected for physico-chemical studies. Twenty (20) Vertical Electrical Sounding data were acquired at different locations. Results of the depth sounding showed that the KQ and HA were the dominant type-curves constituting about 90% of the curve types obtained. Four to five geo-electric/geologic layers were delineated with Bitumen impregnated layer found within the third and fourth layer with resistivities ranging from 86 to 255Ωm. Results from the 2D Werner and geoelectric sequencing showed that the underlaying bitumen Impregnated layer was overlain by a protective clay layer which has some discontinuities and weak zones in Traverses 1 and 4 indicative of the most vulnerable zone(s) in terms of groundwater pollution while Traverses 2 and 3 are less vulnerable. Five (5) geophysically derived independent variables which represent intrinsic properties of the groundwater quality were computed, mapped and fed into the Random Forest Machine Learning Tool. These variables were permuted and the effect of the out of bag classification was measured. Longitudinal Conductance was ranked highest with a variable importance of 0.51. The model used 75% of the input dataset to train the model while 25% was used to test the trained model; the model trained from this subset had a mean square error (MSE) of 0.0982. The resultant output predicted values which ranged from 0.7251 to 2.83836. The predictive vulnerable map showed that the northern and the central part trending north east were regions predicted to be the most vulnerable regions while the southern part of the study area was predicted to be the least vulnerable area.

Keywords: Bitumen, Saturated zones, weak zones, Agbabu, Resistivity imaging, Clay formation, Wenner array

INTRODUCTION

Bitumen is one of the richly deposited minerals in Nigeria and just like crude oil, it is found in Ondo, Lagos, Ogun and Edo state (Odunaike *et al* 2009, Ojuri et al 2009, Ademilua 2014, Alagbe, 2020). Accidental discoveries of bitumen as a black viscous tar oozing out of river valleys and farmlands in areas of Ofosu, Agbabu, Mafowoku and Eregu has been dated back to several decades. Its scientific discovery in Nigeria is dated back to 1907 (Adewole, 2009). In its raw state, bitumen is a sticky and a viscous substance occurring mostly in sands and clays. Nigerian bitumen possesses relatively large amounts of naphthalene (10%),

aromatics (90%), asphaltenes (18–23%) and trace-metals which makes it unideal for consumption when it interacts with underground water (Adewole, 2009 , Olabemiwo et al., 2011).
It has been postulated that bitumen occurs in three forms, namely surface and near surface, bitumen-impregnated sand and bitumen seepages from wells (Adewole, 2009, Ojuri 2009, Ademilua 2014, Alagbe, 2020). However bitumen in Agbabu mainly occurs in the near surface mode of occurrence and consequently due to its shallow mode, contamination of the shallow wells becomes a possibility. Bitumen contamination is commonly associated with spills, leaks and seeps of the crude or oil products (Freeman and Cattell, 1990). Also of a great concern to man is the heavy metals and trace metals content of the heavy oil components of the bitumen which are capable being harmful to human health (Adewole, 2009), it is however noteworthy to state that in small amounts these trace metals are required in the body for maintaining good health but in large amounts they become toxic or dangerous (Olabemiwo *et al.*, 2011). The abundant presence of metals can be inimical to the health of humans, plants and marine life. Heavy metals toxicity can lower energy levels and damage the functioning of the brain, lungs, kidney, liver, blood composition and other important organs.

Groundwater is the water present beneath Earth's surface in soil pore spaces and in the fractures of rock formations (Alagbe, 2010, Alagbe *et al.* 2013). Researches over the years has shown that groundwater is the purest form of water, however several factors such as climate, characteristics of soil, human activities on the ground, circulation of groundwater through rock types, topography of the area, intrusion of saline water in coastal areas can have adverse effects on the quality of water (Ojeyemi *et al*, 2014, Abatyough *et al.*, 2016). It is therefore expedient to understudy the geology and formation of the subsurface to study the vulnerability of the groundwater.

Agbabu is located in the coastal region and its geology suggests a large deposit of bitumen and Cretaceous tar sand formations as well as the salt water intrusion occurring in the area (Omosuyi *et al*, 2008). Some heavy metals such as Cu, Fe, Mn, Ni and Zn are compulsory micronutrients for flora–fauna and microbes. Besides metals like Cd, Cr and Pb are harmful beyond a certain limit. Therefore, the heavy metal concentration in drinking water should be kept in lowparts per billion range.

Several authors have delineated the presence of bitumen in Agbabu community (Odunaike *et al.,* 2009, Amigun *et al.,* 2012, Ogungbemi *et al.,* 2019, Alagbe 2020,) and it has been instituted that this bitumen occurs at a shallow depth from the surface. Alagbe (2020) further stressed that the near-surface occurrence of natural bitumen coupled with the intrusion of saline water might hinder quality of water from the shallow hand-dug wells upon which majority of the community dwellers depend for water consumption and for domestic usage. It therefore becomes necessary to assess the water quality from various wells to ascertain the levels of pollution especially from the existing underlying bitumen in the study area.

Heavy oil production creates significant disturbances or distruptions of underground formations, groundwater hydrology, and land surface. Consequently, it is expected to affect the quality of ground and surface water resources at the production location and often in adjacent areas too.
Based on the fact that Agbabu coastal aquifers are vulnerable to contamination in view of their shallow nature, high porosity and permeability, the motivation for this research was basically to delineate the geological setting and the average depth of occurrence of the bitumen in the area; investigate if the presence of this bitumen currently influences the quality of the groundwater from hand dug wells within the location as wells as detect the most vulnerable areas when the mining activities commences in the area (i.e., when the subsurface is disturbed). This was achieved through the generation of a potential pollution predictive model

for the study area using electrical resistivity techniques, water chemistry analysis (physico-chemical analysis) and the Random Forest Machine Learning Algorithm Regression tool.

The electrical resistivity method is used in the study of horizontal and vertical discontinuities in the electrical properties of the ground (Ugbhor et al, 2020).The two most common arrays used for Electrical surveying are the Wenner array and the Schlumberger array (Cardimona., 1990).
Vertical electrical sounding (VES), using schlumberger is used mainly in the study of horizontal or near-horizontal interfaces. The technique is extensively used in geophysical surveys to determine overburden thickness and also in hydrogeology to define horizontal zones of porous strata. The use of Electrical Resistivity Imaging (ERI), using Wenner array to address a wide variety of hydrological, environmental and engineering issues is increasingly popular, as it allows for potential discrimination between various geological materials (Reynolds, 1997; Amidu and Olayinka., 2006; Aizebeokhai 2010).

Physico-chemical analysis involves the measurement of various physical properties of systems, most often phase transition temperatures and other thermal properties (thermal conductivity, heat capacity, thermal expansion), electrical properties (conductivity, dielectric permittivity), and optical properties (refractive index, rotation of the plane of polarization of light). Also measured are the density, viscosity, and hardness, as well as the dependence of the rate of the transformations occurring in a system on the system's composition. X-ray diffraction analysis and techniques of microscopic metallography are extensively used in physicochemical analysis.

In recent years, machine learning has experienced significant development and new methods have been proposed to solve some of the problems described for widely used methods (Khalil *et al*., 2005). An emerging type of machine learning techniques which utilizes ensembles of regressions is receiving highlighted interest in many fields of knowledge (Hansen and Salamon, 1990; Steele, 2000; Sesnie *et al*., 2008;). An ensemble learning tool called Random Forest (RF) is increasingly being applied in fields related to the environment and water resources (Herrera *et al*., 2010; Loos and Elsenbeer, 2011; McGinnis and Kerans, 2012; Rodriguez-Galiano *et al*., 2012). RF offers a new approach to the problem of vulnerability mapping, as it is relatively robust to outliers and it can overcome the "black-box" limitations of artificial neural networks, assessing the relative importance of the variables and being able to select the most important variables (features) and reducing dimensionality. At the same time the parameterization of RF is very simple and it is computationally lighter than other machine learning methods (neural networks or support vector machines) (Rodriguez-Galiano and Chica-Rivas, 2012). Although RF is being currently used as a remote sensing data classifier, its potential as a spatial modeling tool for vulnerability mapping is rated very high.

Random Forest is an ensemble method which combines multiple decision tree algorithms to produce repeated predictions of the same phenomenon. Decision trees can be divided into classification trees and regression trees. A regression tree (RT) represents a set of restrictions or conditions which are hierarchically organized, and which are successively applied from a root to a terminal node or leaf of the tree (Breiman, 1984; Quinlan, 1993). In order to induce the RT, recursive partitioning and multiple regressions are carried out from the dataset. From the root node, the data splitting process in each internal node of a rule of the tree is repeated until a stop condition previously specified is reached. Each of the terminal nodes, or leaves, has attached to it a simple regression model which applies in that node only. Once the tree's induction process is finished, pruning can be applied with the aim of improving the tree's generalization capacity by reducing its structural complexity. The number of cases in nodes can be taken as pruning criteria. RT involves first selecting optimal splitting measurement vectors. The process starts by splitting the dependent variable, or the

parent node (root), into binary pieces, where the child nodes are 'purer' than the parent node. Through this process, the RT searches through all candidate splits to find the optimal split that maximizes the 'purity' of the resulting tree as defined by the largest decrease in the impurity.

$$i(s,t) = i(t) - \rho L_{(t)} - \rho L_{(R)} \quad 1$$

where, s is the candidate split at node t, and the node t is divided by s into the left child node $L_{(t)}$ with a proportion of ρ, and right child node $L_{(R)}$ with a proportion of ρ.

i(t) is a measure of impurity before splitting.

The philosophy behind ensemble learning techniques is based upon the premise that its accuracy is higher than other machine learning algorithms because the combination of predictions performs more accurately than any single constituent model does.

To avoid the correlation of the different RTs, on the one hand RF increases the diversity of the trees by making them grow from different training data subsets created through a procedure called bagging. Bagging is a technique used for training data creation by resampling randomly the original dataset with replacement, i.e., with no deletion of the data selected from the input sample for generating the next subset {h(x,ϕk), k = 1, …, K}, where {ϕk} are independent random vectors with the same distribution. Hence, some data may be used more than once in the training of trees, while others might never be used. Thus, greater prediction stability is achieved, as it makes it more robust when facing slight variations in input data and, at the same time, it increases prediction accuracy (Breiman, 2001). Another characteristic of interest is that the trees of an RF grow with no pruning, which makes them light, from a computational perspective.

Location, Climate and Geology of the Study Area

The study area (Agbabu) is a settlement to the southern part of Odigbo Local Government Area of Ondo State of Nigeria on latitude 6° 35' 19" N and longitude 4° 50' 03" E (Fig. 1). It is bounded in the north by smaller farm settlements such as Mulekangbo, Ilu-binrin and Mile 2 and in the south by Igbaje community.

Mangrove and freshwater swamps cover the area, which also contains brackish lagoons. Rainfall in this equatorial zone is high, frequently exceeding 3 m per year (Ola 1983). The altitude of the area ranges from 5-16m above mean sea level and is characterized by November to March(dry) and April to October (wet) seasons. (Akinmosin and Imo, 2016). The bituminous sediments in southwestern Nigeria occur in the cretaceous Abeokuta group, the basal sedimentary unit in the southwestern Nigeria miogeosyncline (Reyment 1965; Fayose 1978). The sediments occur on the eastern margin of a coastal sedimentary basin known as the Benin Basin which extends from the Ghana-Ivory Coast border through Togo and Benin Republics to western Nigeria. The basin was formed consequent to the opening of the south Atlantic probably during the Neocomian (Omatsola and Adegoke 1981). Agbabu bituminous belt sand deposits in south western Nigeria are naturally occurring in a sticky tar-like form (Akinmosin and Imo, 2016). Its geology suggests a large deposit of bitumen and Cretaceous tar sand formations. The Dahomey basin is an Atlantic margin basin containing Mesozoic-Cenozoic sedimentary succession reaching a thickness of over 3,000m. It extends from south-eastern Ghana to the western flank of the Niger Delta. Its stratigraphy is classified by various authors into Abeokuta Group, Imo Group, Oshosun Formation, Ilaro Formation and Coastal Plain sands and Alluvium (Jones and Hockey, 1964, Omatsola and Adegoke,1981). The Agbabu area is underlain by the sediments of the Imo Group (Fig. 2).

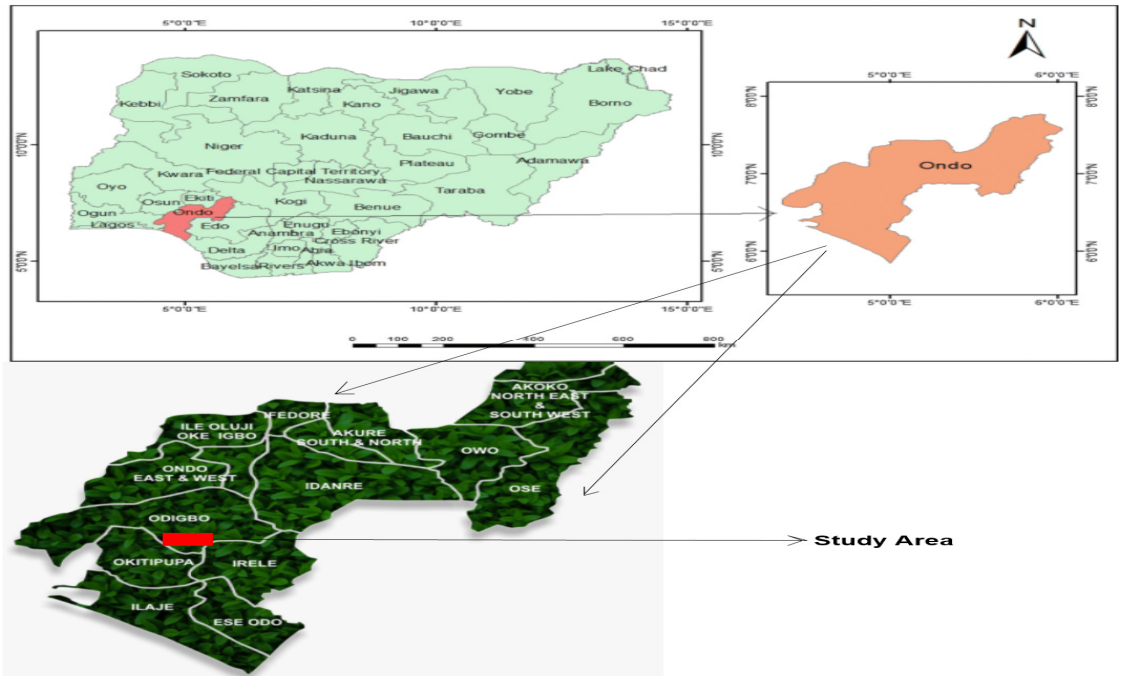

Figure 1: Location Map showing Ondo State and the study area

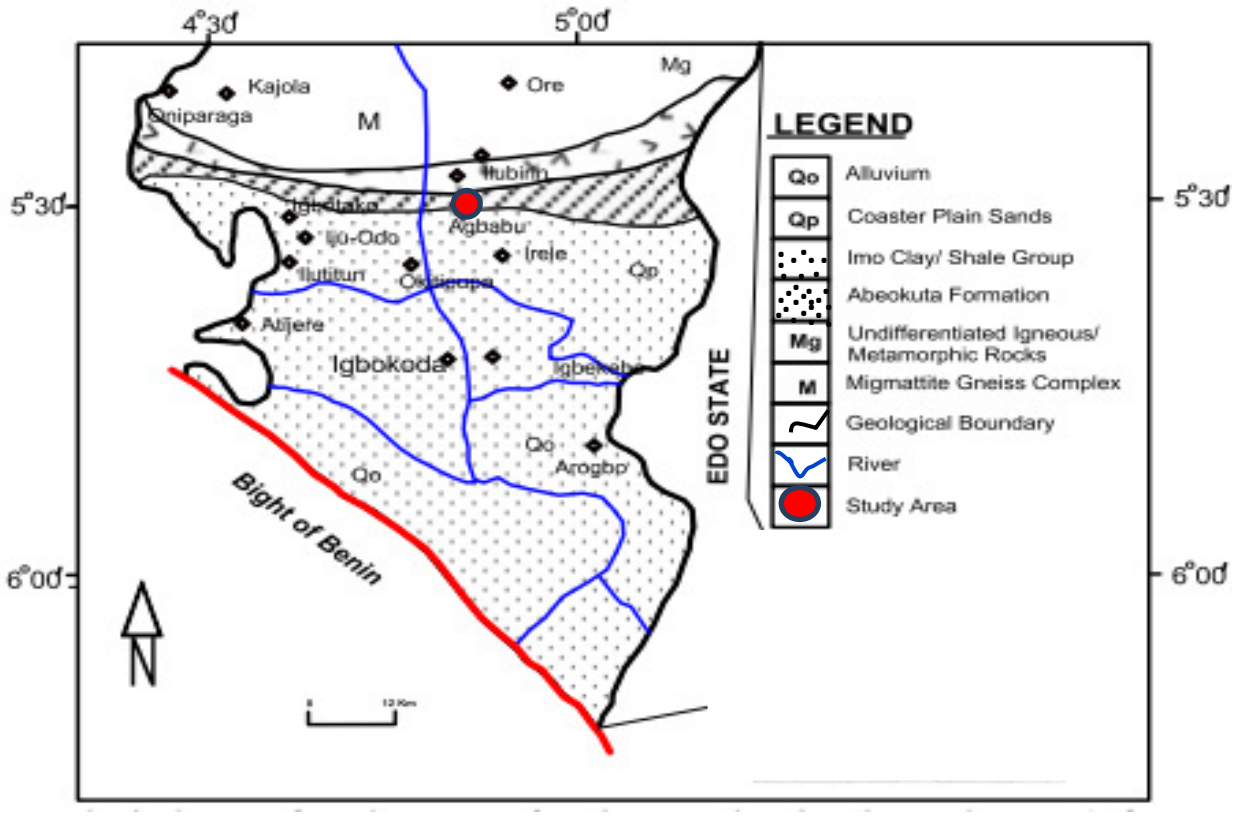

Figure 2: Geologic map of the Study area

MATERIALS AND METHODS

In order to achieve the objectives of this study, the research was executed in three phases. The methodologies adopted includes data acquisition, data compilation and extraction, data processing, development of several geophysical models, development of models for different vulnerability parameters as well as the Random Forest regression models for the potential groundwater vulnerability for the study area.

In the second phase of the study, these data were compiled, extracted and processed using computer software to generate several geophysical models and parameters. Also, dependent variables were derived from laboratory water chemistry analysis. In the final part of the study, the results obtained from phases one and two was used to developed various geophysical models of the subsurface stratification and the groundwater vulnerability prediction models using the random forest machine learning regression algorithm.

Data Acquisition

Electrical Resistivity Imaging

Four geophysical traverses were established across the study area in the NE – SW direction, covering a total distance between 150 and 165m with inter electrode spacing of 5 m, 10 m and 20 m, while inter dipole separation factor (n) was varied from 10 to 20 (Fig.3).

These traverses were targeted near existing wells to understand the subsurface lithology within and around these wells as well as tie the results of the geophysical studies with the hydro-chemical analysis. The traverses were set up to cover Agbabu community and neighboring Mile 12 where mining Activitieshad occurred in the past.

Transmitting dipole which is powered by low frequency dc source was stationed at station 0 and 30. The receiving dipole was initially stationed at station 10 and 20 (n-1) and subsequently moved to stations 30 and 40 (n-2), station 40 and 50 (n-2), 50 and 60 (n-2), 60 and 70 (n-2), stations 70 and 80 (n-2) ditto every other traverse.

The apparent resistivity values were calculated using $\pi an+1n+2n$ asthe geometric factor. The apparent resistivity values obtained were plotted at the intersection of lines drawn at 45 degrees from the mid points of the potential and current dipoles. 2D inversion modeling of the wenner data was carried out using DIPRO (a computer software package). It provides high resolution color or contoured images of both the field data Pseudosection and the 2D resistivity structure that resulted from the inversion.

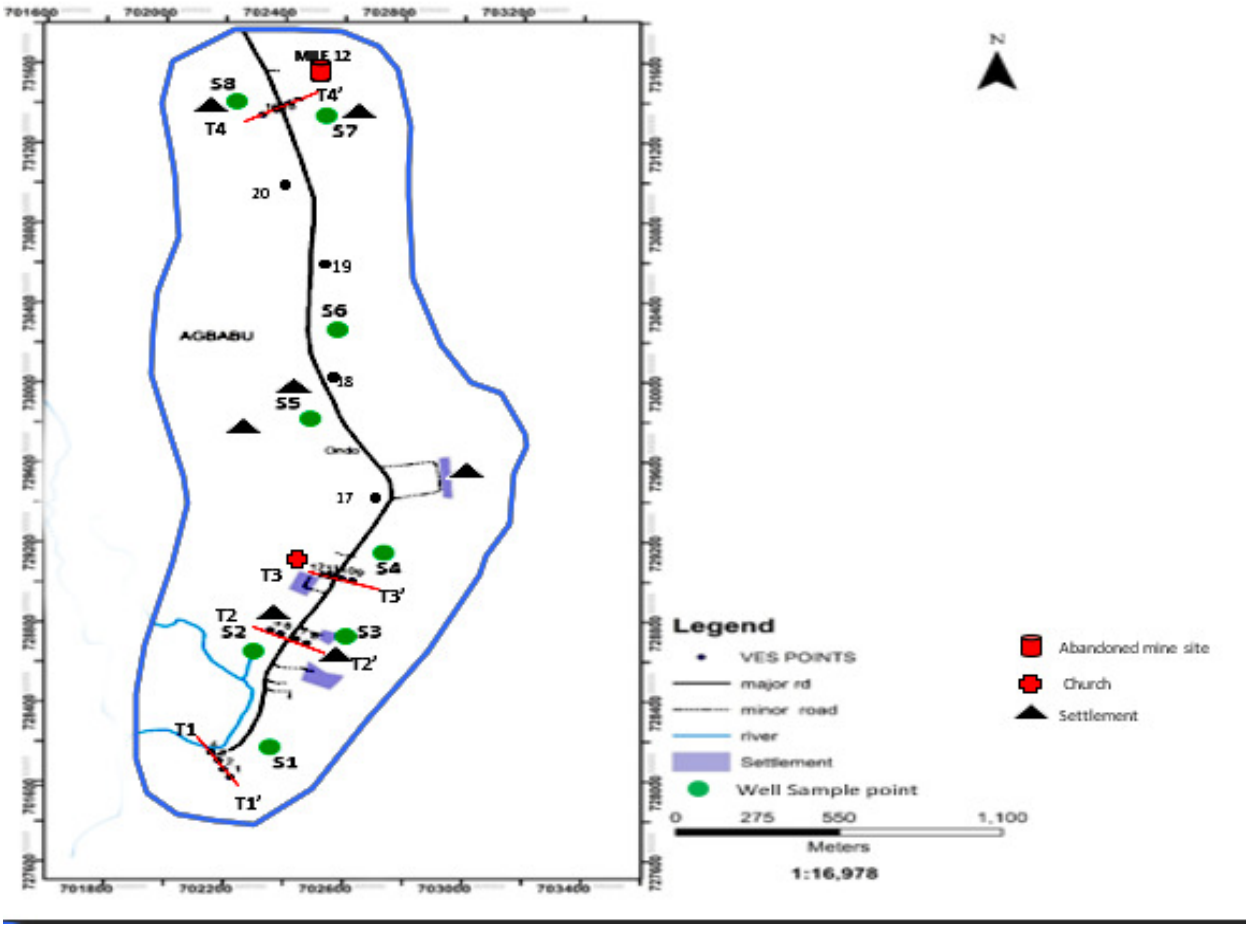

Figure 3: Base Map of the study area

Vertical Electrical Sounding (VES) data

The significant thickness relative to depth as well as the very high resistivity contrast with the host geology contributes to the success of geo-electric techniques in bituminous sands detection. As Agbabu's bituminous occurs at shallow depth of less than 100m; geo-electric methods have been used successfully in its detection (Bauman, 2005; Odunaike, et al, 2012; Alagbe, 2020). VES were carried out along some suspected points on ERI traveses.

Along Traverse 1, sounding stations were occupied at stations 20 m, 60 m, 105 m and 130m from the zero station (E-W) while along Traverse 2, stations 40 m, 90 m, 125 m and 165 m were all occupied along the E-W direction from the starting station. Furthermore, Stations 30 m, 85 m, 105 m and 135 m were all occupied along traverse 3 while stations 15 m, 55 m, 100 m and 165 m were sounded along traverse 4.

The current electrode spacing (AB/2) was varied from 1m to 130 m and to 165 m in traverse 1 and traverse 2 respectively while it (AB/2) varied from 135 m and to 165 m for traverse 3 and traverse 4 respectively. The survey was carried out at the beginning of dry season (November) when the ground was still sufficiently moist enough as to ensure good conduction of current.

The apparent resistivity measurements at each station were plotted against spacing on bi-logarithmic graph sheets. The resulting curves were then inspected visually to determine the nature of the subsurface layering. In this way, each curve was characterized depending upon the number and nature of the subsurface layers.

Partial curve matching was carried out for quantitative interpretation of the curves. The results of the curve matching (layer resistivities and thickness) were fed into the computer as a starting model parameter in an iterative forward modeling technique using RESIST version 1.0 (Vander velper., 1988). The result of the forward model was fed to the computer and various geoelectrical parameters were estimated. The geoelectric parameters estimated are aquifer resistivity, aquifer thickness, unsaturated zone thickness, total longitudinal conductance of the unsaturated zone, total transverse resistance of the unsaturated zone, hydraulic Resistance and hydraulic conductivity.

The unsaturated zone was identified by the overlapping layers to the aquiferous layer. Therefore,the results of overlapping layers resistivity and thickness to the aquiferous layer were used toestimate unsaturated zone thickness, total longitudinal conductance of the unsaturated zone andtotal transverse resistance of the unsaturated zone.

The unsaturated zone thickness was calculated using the summation of the thickness of the overlapping layers.

$Uz_t = \sum_{i=1}^{n} ti$ 2

Where t = the thickness of ith overlying layer

Longitudinal unit conductance (S), and transverse unit resistance (TR) are regarded

as Dar Zarrouk parameters.

Transverse Resistance (TR) is the resistance normal to the face and S as the conductance parallel to the face for a unit cross section area.

The longitudinal conductance (S) and Transverse resistance (TR) of the unsaturated

zone were calculated from the result of resistivity data (Table 2).

TR and S are associatedwith the thickness of the Layer (h in meters) and the resistivity of the layer(p in ohm-m). The longitudinal unit conductance (S) is given as

$$S = \frac{h}{\rho} \quad 3$$

$$s = \sum_{i=1}^{n} h/K \quad 4$$

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

$TR = \rho * h$ 5

$TR = \sum_{i=1}^{n}(\rho * h)$ 6

Where ρi and hi are resistivity and thicknesses of ith layer, respectively,
Layer resistivity (ρ) and thickness (h) are the two parameters that characterized a geologic unit and are fundamentally important both in the interpretation and in the understanding of the geoelectrical models. These parameters helped in deriving other parameters (hydraulic conductivity and hydraulic resistance) for the horizontal, homogenous and isotropic layers (Gemail *et al*. 2011)

The hydraulic conductivity (K) of the aquifer protective layers is a key parameter in assessing aquifer vulnerability.

$K = 386.40 R_{rw}^{-.93283}$ 7

where K is the hydraulic conductivity and R_{rw} is the aquifer resistivity
The estimation of Hydraulic resistance makes use of two parameters: the thickness (h) of the protective layers and the estimated hydraulic conductivity (K).
The hydraulic resistance (C) was computed using the two parameters and is expressed as

$C = \sum_{i=1}^{n} \frac{hi}{Ki}$ 8

where K_i is the hydraulic conductivity, while h_i is the thickness of the vadose zone materials.
The total longitudinal conductance of the unsaturated zone, total transverse resistance of the unsaturated zone, hydraulic conductivity and the hydraulic resistance maps were prepared by interpolation of the results. These values were interpolated using the computer software Surfer version 12 to provide values for all over the Agbabu. The values of each parameter are prepared on a value attribute table of their respective thematic map for statistical analysis.

WATER SAMPLE DATA ACQUISITION AND PROCESSING

A total of eight(8) privately owned drinking water wells within the community were randomly selected (Fig. 3) and water from these waters well sources were collected. The groundwater was collected over a day time period A plastic bottle (2 litres) was washed with HCL acid of 0.5mol/dm^3 and rinsed with distilled water and were stored in distilled plastic bottles. These samples were taken to the laboratory for analysis in less than 12 hours after collection. Well information was also taken. The samples were digested for water quality test. Physico-chemical parameterstest was performed on all the water samples. The following physio-chemical parameters weretested for: conductivity, pH, Chloride, Total Hardness, Sulphate, Nitrate,Phosphate, Total Solids, Total Dissolved Solids, Total Suspended Solids, Total Alkalinity. Inaddition to these parameters, some inorganic metals were also tested for (Na, K, Ca, Mg, Zn, Fe, andCu). Atomic Absorption Spectrometer test was conducted to test forthe presence of heavy metals such as Cd, Mn, and Pb.
In order to determine whether the water in the study area was contaminated or safe forconsumption, the water quality results were compared with a maximum permissible level for safe drinking water by the Environmental Protection Agency (EPA), Quality Threshold values guidelines arepresented in Table 1
Table 1: Table showing the maximum permissible limits for different parameters

Parameters	Maximum Permissible Limit (EPA, 2001)
Ph	6.5-8.5
Conductivity (µscm)	1200
Total Hardness (mg/l)	180
Ca Hardness (mg/l)	200
Mg Hardness (Mg/l)	50
Alkalinity (mg/l)	600
TDS (mg/l)	500
Chloride (mg/l)	250
Salinity (mg/l)	150
Sulphate (mg/l)	250
Na (mg/l)	180

DATA PRESENTATION AND INTERPRETATION

The data for this research work were presented in form of maps, curves , tables and the interpretation was done qualitatively and quantitatively.

RANDOM FOREST MACHINE LEARNING ANALYSIS

For Random Forest models induction, five explanatory variables (Longitudinal Conductance, Traverse resistance, Hydraulic resistance, depth to bituminous layer and Hydraulic conductivity) were used together. These derived information of the area from different sources were fed into the geodatabase in the Salford Model predictor v 8.2.2. RF facilitates several means of ranking the importance of input variables. In this instance, each variable was permuted and the effect on the out-of-bag classification accuracy was measured. Those variables which, when permuted, produced the greatest change to classification accuracy were ranked highest. In this instance, longitudinal conductance was the most ranked variable and this was used as the

basis for the target formation using vulnerability classes according to Antonio and Richard 2014, where **0** represents the least vulnerable and 3 represents the most vulnerable.

The parameters formed the input to the Random Forest algorithm and are known as input-feature vectors while the vulnerability class obtained from longitudinal conductance was used as target values for the training of the algorithm. A Random Forest (RF) comprising of 300 trees was used to train the data set with no limits on individual tree depth or subsequent pruning. The RF produced under these parameters required 12s to train. Breiman (1996) demonstrated that by increasing the number of trees the generalization error always converges; hence, overtraining is not a problem and the number of trees can be fixed once the error has converged. The resulting models were evaluated using the Out Of Bag (OOB) error estimation. Moreover, with the aim of reducing the dimensionality and improve the accuracy and interpretability of models, the most significant predictive features were selected by using the importance measures of RF.

The response values were fed into Surfer software to produce a potential vulnerability map for the area. Random Forest method within the package imbedded in the Salford Model Predictor statistical software was used. Figure 4, shows the architecture of the Random Forest classifying technique.

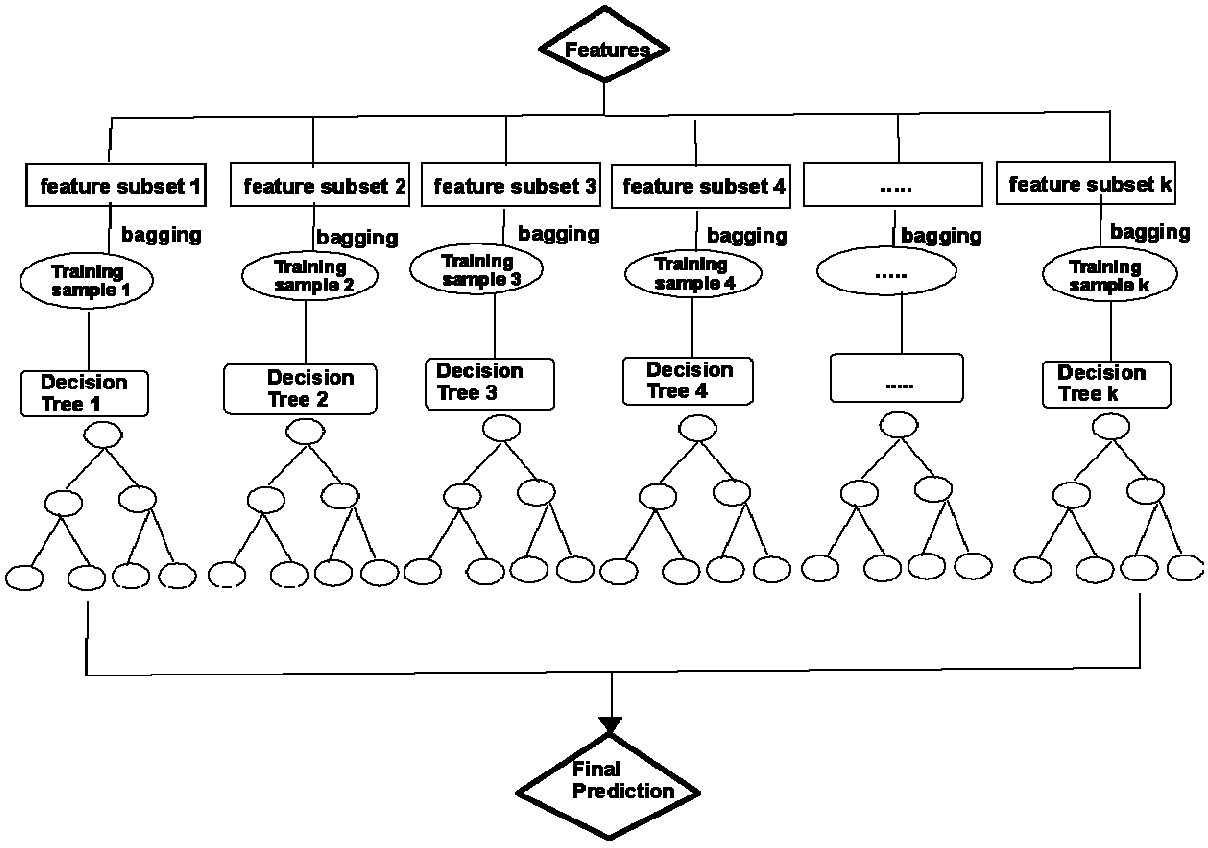

Figure 4:An Architecture of Random Forest Classifiers.

Random forest generally involves 4 steps.

i. Randomly select N features from the full feature set.
ii. Perform the bagging algorithm on the training set and generate a training set with re-sampled instances.
iii. Employ a decision tree algorithm on the re-sampled training set and the randomly selected feature space, and build a decision tree, which serves as base classifier in Step 4.
iv. Repeat Steps 1, 2 and 3 for k times and generate k decision trees. Lastly, summarize the k decision trees and generate a final prediction.

RESULTS AND DISCUSSION

The results will be discussed in the order of the methodologies adopted for the study and the results presented as field curves, tables, charts, geoelectric sections, 2D resistivity inversion images and maps.

Electrical Resistivity Imaging

Traverse 1 (Figure 5) covers a total spread of 165m with an electrode spacing of 5m and runs in the East-West direction. The resistivity values range from 15.1 ohm-m to 1093 ohm-m along the traverse. Four major subsurface geologic layers are observed in the 2D resistivity structure along this traverse. The first layer (red color) with resistivity value ranging from 181 ohm-m to 525 ohm-m is interpreted to be a sandy layer. It has a uniform distribution between stations 0 to 40, covering a depth of about 5m. There is an observed downward extension of this layer between stations 40 to 60, 85 to 110 and 125 to 160 where it forms a downward facing cone shaped layering. Meanwhile two major discontinuities are observed along the traverse, the first occurs between stations 65 to 80 and the second discontinuity between stations 110 to 125. These discontinuity zones are possible weak zones for the upward migration of bitumen and therefore potential pollution zones. The second layer (yellow color) which is suspected to be a clay-cap over the underlying bitumen has resistivity values ranged between 106 ohm-m to 200 ohm-m occurs non-uniformly distributed along the traverse.

The third layer with green with a thickness of about 12m occurs as an isolated geologic body in 3 major zones along the traverse trending in the upward direction to the depth of about 5m from the surface especially towards the discontinuity zones on the first layer. This layer is characterized with resistivity values ranging from 33.9 ohm-m to 65.2 ohm-m. This third layer is suspected to be the bituminous zones and observed trending in the upward direction especially towards the discontinuity zones in the first layer. This suggest that the weak zones are easy pathways for bitumen migration once the surface is disturbed.

The fourth layer blue suspected to be saline water with resistivity value ranged from 15 ohm-m to 24 ohm-m occurring at the flanks of the traverse at stations 10-35 and 115-125 at a depth from 15m to the probed depth. Stations between 65 to 80 and 110 to 125 are potential active zones for pollution and should be avoided for groundwater development.

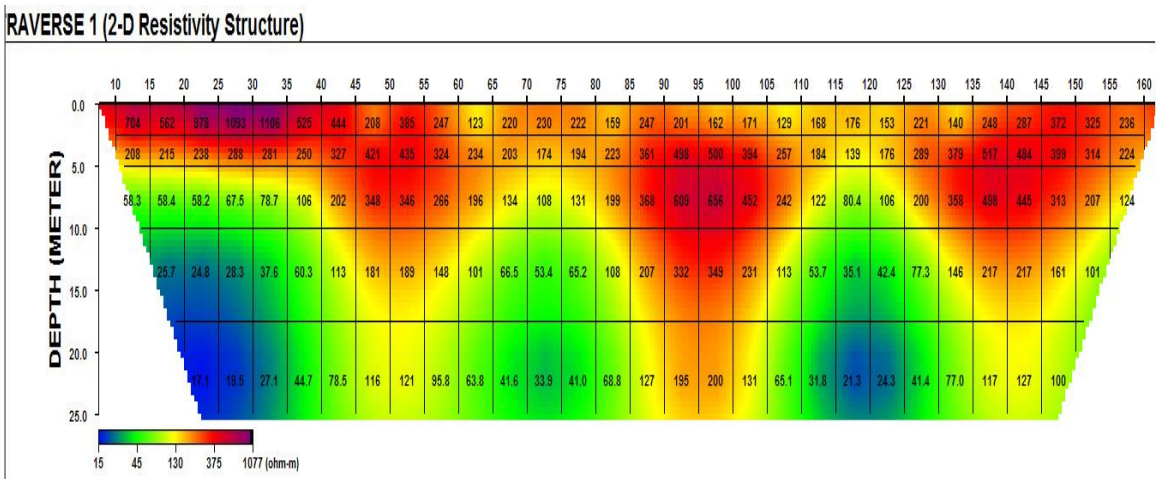

Figure 5: 2D Werner resistivity structure for traverse 1

Traverse 2 (Figure 6) covers a total spread of 175m with an electrode spacing of 5m and runs in the East-West direction. The resistivity value ranges from 0.96 ohm-m to 889 ohm-m along the traverse. Four major layers are observed in this 2D resistivity structure. The first layer (red color) with a resistivity value ranging from 59.9 ohm-m to 261 ohm-m is observed to be uniformly spread across the traverse with a thickness of about 6m except between stations 135 to 155 where it is exhibits an appreciable thickness as much as 30 m is suspected to be sandy layer. The second layer (yellow) reveals a clay cap protecting the underlying bitumen layer which has resistivity values ranging from 18.1 ohm-m to 51.6 0hm-m has a depth of 10 m between station 0 to 100 however from station 100 to 130, this layer has a depth of about 15 m. The third layer (green) is indicative of the bituminous layer which is not uniformly spread along the traverse has a resistivity value ranging from 6.69 ohm-m to 14.6 ohm-m. It can be inferred from the figure 6, that this layer only occurs from station 0 to 110, however stations 110 to 170 shows no indication of the presence of the bituminous layer therefore this region is the least active zone in terms of pollution within the traverse as such is best suited for groundwater studies. The fourth layer (blue) is suspected to be saline water with resistivity value ranged from 0.96 ohm-m to 2.13 ohm-m occurs at the eastern flank of the traverse between stations 0 to 55.

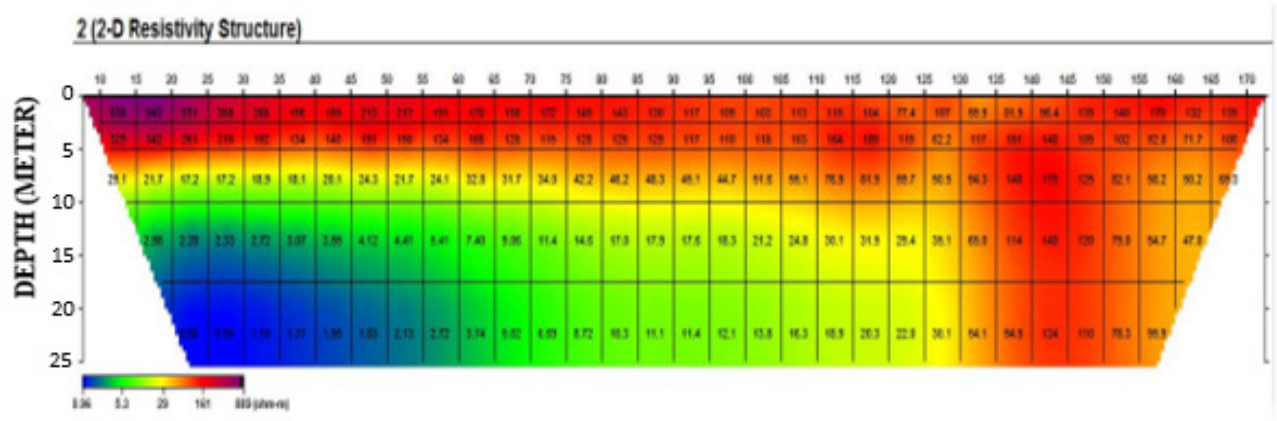

Figure 6: 2D Werner resistivity structure for traverse 2

Traverse 3 (Figure 7) covers a total spread of 140m with an electrode spacing of 5 m and runs in the East-West direction. The resistivity values range from 2.10 ohm-m to 606 ohm-m along the traverse. Firstly, a pocket of high resistivity obtained between stations 50 to 80 is suspected to be due to the tarred major road between the traverse. Four major layers are observed in this 2D resistivity structure. The first layer (red color) is observed uniformly across the traverse with a thickness of about 10m postulated with resistivity values ranging from 76.7 ohm-m to 229 ohm-m which suspected to be sandy layer. The second layer (yellow color) has resistivity value ranging from 27.4 ohm-m to 92.0 ohm-m which reveals a clay cap directing covering the underlying bituminous layer has a depth between 9 m to 15m at the eastern flank between stations 0 and 45, 9 m to 10 m at the centre of the traverse between stations 50 and 110 and at the western end of the traverse between stations 110 and 140, an increase in the depth was observed trending westward between 10m to 20m. The third layer (green color) is interpreted to be the bituminous layer appearing across the traverse but most pronounced at the flanks of the traverse with varying resistivity values between 6.79 ohm-m to 28.5 ohm-m. It occurs at a depth of about 15m to depth investigated between stations 0 to 50 while from stations 55 to 105, it occurs at a depth between 11m while between stations 105 to 135, this layer occurs at a depth of about 17m. The fourth layer suspected to be saline water impregnated layer with resistivity value ranged from 2.1 ohm-m to 3.35 ohm-m predominantly occurs at the center of the traverse between stations 50 to 105 forming a cone-like shape that does not extend to the flanks of traverse occurs at a depth of about 17m. Furthermore, stations 50-110 appears to be the most active zone in terms of pollution along the traverse.

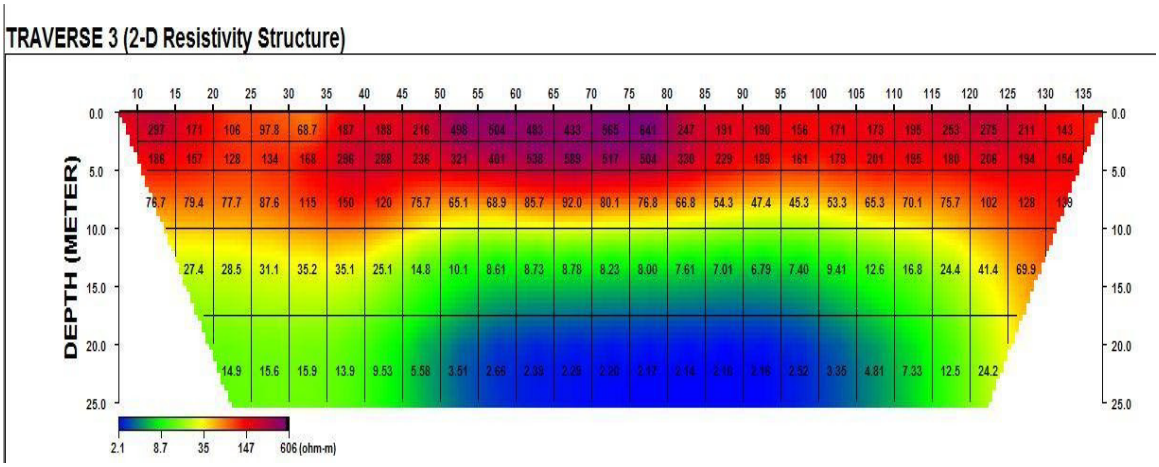

Figure 7; 2D Werner resistivity structure for traverse 3

Traverse 4,(Figure 8) fell along a part within the study area where mining activities had occurred in the past which led to the disturbance of the subsurface. This region shows the highest evidence of bitumen in the study area occurring at a very shallow depth. It shows a thin top sand layer at the extreme ends of the traverse with resistivities varying between 2109 and 346 ohm-m with a thickness of just few metres and is only well pronounced along stations 140 to 170 m while yellow color indicatives of the clay layer occurs sporadically within the bitumen impregnated layer (green color) showing that these subsurface strata had been mixed together due to the overtime disturbance of the subsurface. This totally agrees with the presence of natural bitumen which can be easily seen by visual inspection within this area. Another point of note from the 2D resistivity imaging is the upward movement of the bitumen layer at stations 135m-140m which is due to the activities of an injection pump used to pump the flowing bitumen over the years from the subsurface.

In terms of pollution, Figures 5 and 8 are the most active and vulnerable due to the fact that these traverses showed several discontinuities which could serve as weak zones and conduits for the possible migration of the underlying bitumen deposit. Traverse 4 shows that the bitumen has already spread across the traverse potentially contaminating any water source within the region. Two water wells (fig….) were seen within this area of the study area, one was completely abandoned while the other is still used for domestic consumption by the inhabitants of the area. Traverses 2 and 3 are shown to be less vulnerable to pollution because they showed less discontinuities and weak zones, and the clay cap is almost stable across these two traverses.

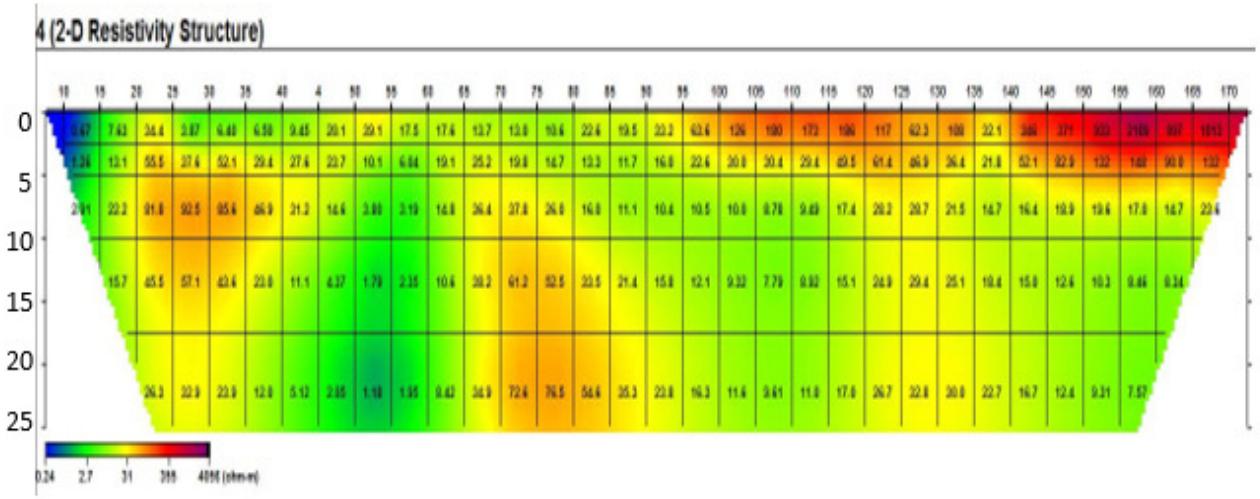

Figure 8; 2D Werner resistivity structure for traverse 4

4.2 Vertical Electrical Sounding Field Curve

Vertical Electrical Soundings were conducted within the study area at Twenty (20) different locations and four curve types namely KQ, HA, H and QH were identified (Table 2). The KQ and HA curve types are the most preponderant with both constituting about 90% of the curve types while H and QH accounts for about 10%. The VES curves typessummarized in Table 2 represents a subsurface condition within the study area and it was generally noticed that the resistivity values were decreasing as we probed deeper. This is suspected to be as a result of the underlying saline water present within the study area. Three to four sub-surface layers were generally delineated, where a resistive top sandy layer underlies a conductive clay cap which serves as a protective layer for a preceding bitumen impregnated layer in turn overlies the saline water intruded layer**.** Table 2 gives a summary of the results of the VES curves within the study area while figure 9 shows a Chart Showing the depth variation of the bitumen within the study area. Expectedly the bituminous layers are supposed to be characterized with high resistivity values but probably because of the intrusion of saline water into bituminous zones, there were observed low resistivity values and these results were in agreement with Omosuyi *et al,* 2008 andAlagbe, 2020. From the chart, it can be deduced that the bitumen within the study area has its deepest occurrence at VES locations 1,3,7,9,10,11 and 12 where it occurs at an average depth of 6m while it occurs at a moderate depth at stations 2,4,5,6 and 8 at an average depth of 3.4 m while at VES locations 13,14,17,19 and 20, it occurs at a really shallow depth of about 1.5m. It can be observed from the above deductions that the bitumen occurrence within the study area can be generally considered to be of shallow depth of occurrence and hence a possibility of easy contact with the shallow groundwater source(s)

Table 2: Summary of the VES Results

VES NO.	Curve Type	Resistivity (ρ1, ρ2…n)	Thickness (h1/h…hn)	Depth (m)
1.	KQ	140/539/179/27	1.9/3.1/1.4	1.9/5/6.3
2.	KQ	17/474/54/27	0.5/2.1/3.7	0.5/2.6/6.2
3.	KQ	51/407/100/4.2	1.5/6.9/1.4	1.5/8.3/9.7
4.	AK	49/159/441/17	1.2/1.1/5.6	1.2/2.3/7.9
5.	HA	89/46/86/1361	0.9/1.4/5	0.9/2.3/7.3
6.	KQ	58/206/162/58	1.3/0.5/5.6	1.3/1.8/7.4
7.	KQ	101/316/112/28	2.3/3.3/1.4	2.3/5.6/7
8.	AK	58/167/255/11	0.9/0.8/4.4	0.9/1.7/6.1
9.	KQ	94/127/79/36	1.1/5/16.6	1.1/6.1/23
10.	KQ	151/528/141/5.5	2.2/2.8/2.5	2.2/5/7.5
11.	KQ	63/261/90/9.3	1.7/3.2/1.4	1.7/4.9/6.2
12.	K	78/331/7.9	1.6/3.5	1.6/5.1
13.	H	83/12/38	0.8/27.5`	0.8/28.3
14.	H	214/12/20	0.6/15	0.6/16
15.	QH	125/31/11/24	1.8/1.9/16.4	1.8/3.7/20.1
16.	QH	617/198/7.5/117	2.5/1.3/13.1	2.5/3.8/16.9
17.	H	48/14/21	1/20	0.8/21.2
18.	H	90/13/16	1.5/20	0.7/2.0
19.	QH	97/33/14/20	2/4/20	2.2/20.1
20.	QH	540/206/19/34	2/4/32.5	1.0/1.8/19.7

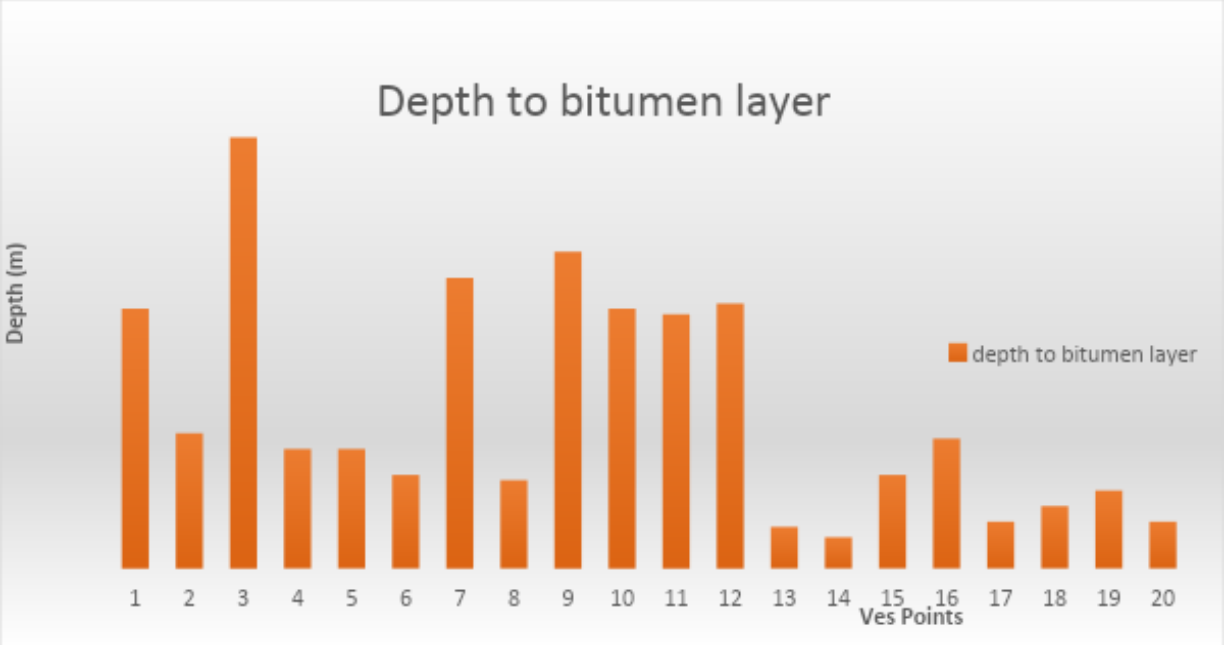

Figure 9: Chart Showing the depth to bitumen variation in the study area

Longitudinal Conductance

Total Longitudinal conductance and Total resistance of the unsaturated zone helps to characterize the study area.The data were grouped into four vulnerability classes based on longitudinal conductance model shown in tabl**e** 3.

Table 3**:** Relationship between longitudinal conductance and vulnerability classes of the method (After Antonio and Richard, 2014)

Longitudinal Conductance (Siemens) (Dar Zarrouk Parameters)	Vulnerability Classes
0.7 < 2.5	Low
0.3 -0.7	Moderate
0.1 – 0.3	High
< 0.1	Extreme

Areas with low longitudinal conductance values (-0.2 to 2 s) tend to have low protective capacity. The study area is majorly characterized with extreme to high vulnerability class. Hence, larger parts of southern and northern zones (pink/blue colouration) are vulnerable to pollution. Areas with high longitudinal conductance values (1.8 to 3.8 s) have low permeability and are less vulnerable. The total longitudinal conductance map of the study area is represented in figure 10.

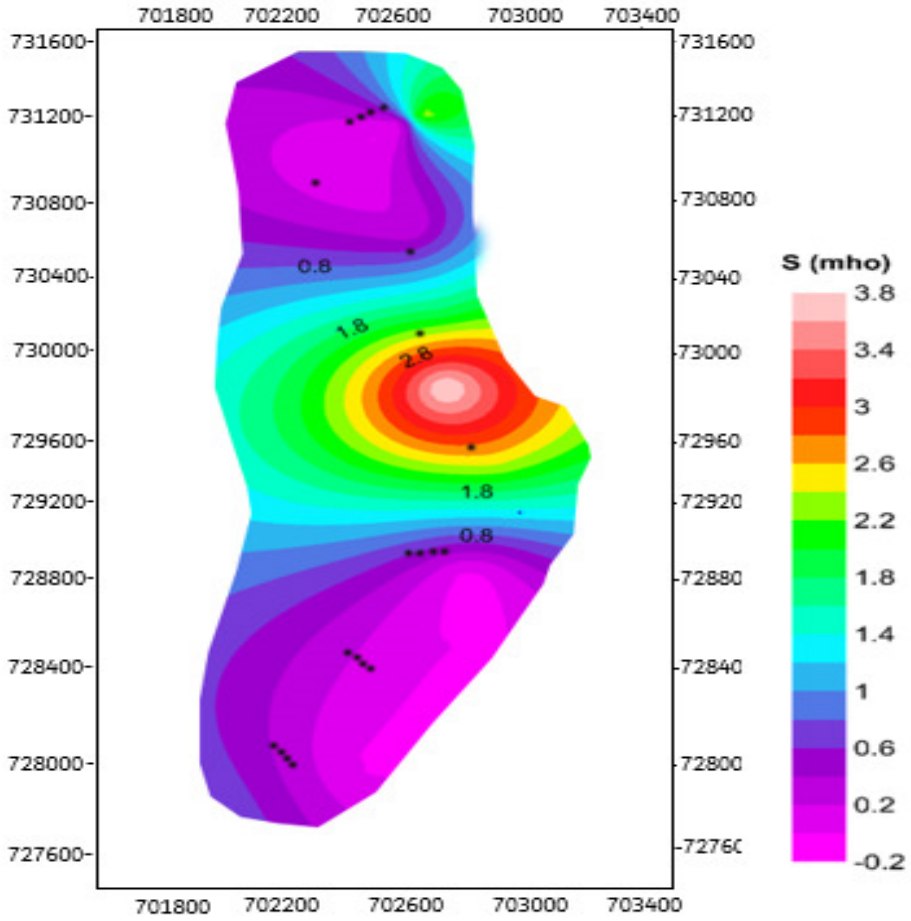

Figure 10: Longitudinal conductance map of the study area

Total Transverse Resistance

The total transverse resistance of the study area in map Figure 11, range between 64.12 to 1669.97 ohm-m. The study area was classified into high and low transverse resistance areas and total traverse value above 700 ohm-m are classified as high transverse resistance while values less than 700 ohm-m are classified as low transverse resistance. Low total transverse resistance dominated the entire study area with exception of some few pockets of high total transverse resistance at the southern and northern parts of the study area. (Fig.11). Areas with low total transverse resistance as observed in the northern, southeastern and the eastern parts and are classified as areas of high infiltration due to their high permeability. Consequently, these areas are highly vulnerable to surface contaminant. In general, it can be inferred that the study area is highly vulnerable to surface contaminants.

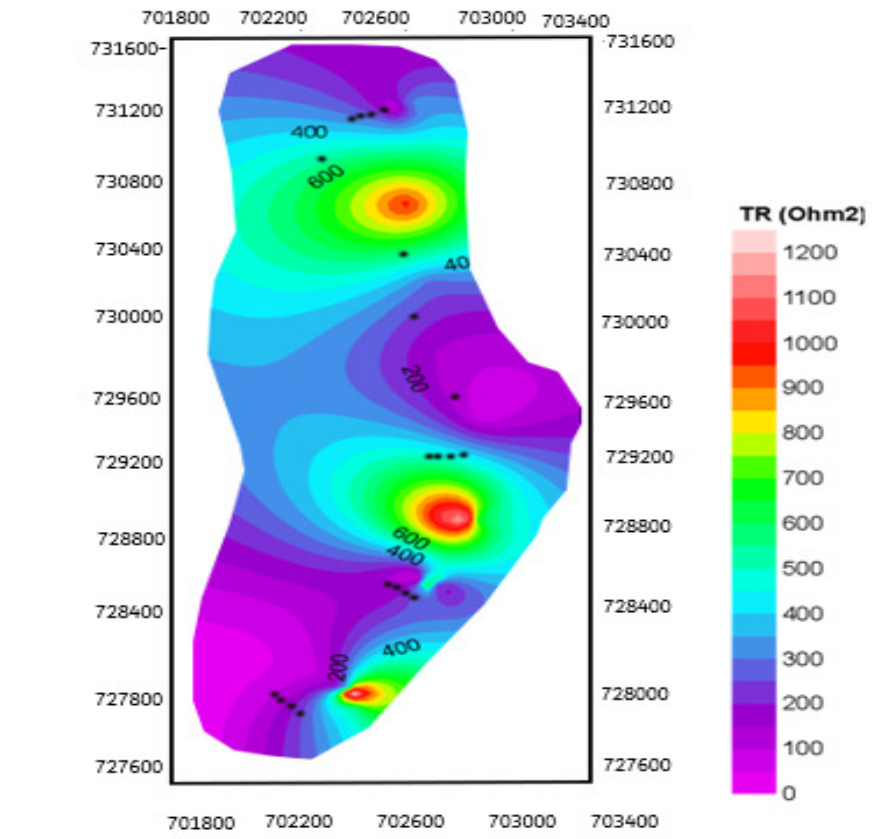

Figure 11: Traverse Resistance map of the study area

RANDOM FOREST MACHINE LEARNING ALGORITHM

Several independent (predictive) variables; total longitudinal conductance (LC), transverse resistance (TR), depth to the bitumen layer (DOB), hydraulic conductance (HC), resistivity and hydraulic resistance (HR) values for each of the VES locations were obtained. The values were extracted using respective computation formulae for each variable with a view of providing a framework for the establishment of relationships between the bitumenand the groundwater of the area.

Variable ranking and selection

Random Forest (RF) facilitates several means of ranking the importance of input variables. In this study, each variable was permuted and the effect on the out-of-bag classification accuracy was measured. Those variables which, when permuted, produced the greatest change to classification accuracy were ranked highest (Table 4). Longitudinal conductance was noted to be ranked highest and this was used as target variable in the Statistical Analysis.

Table 4: Table showing variables and corresponding variable importance

Variable	Variable Importance
LC	0.51
TR	0.22
HC	0.18
HR	0.06
DOB	0.03

The Random Forest Model

The RF model was trained with five (5) explanatory variables which represent intrinsic properties of the groundwater quality parameters in the study area (i.e. total longitudinal conductance, transverse resistance, depth to the bitumen layer, hydraulic conductance, resistivity and hydraulic resistance). The model trained from this subset had a mean square error (MSE) of 0.0982 which is good enough as it is close enough to zero. According to the model generated, Longitudinal conductance was still the most relevant variable associated with vulnerability of groundwater followed by hydraulic conductivity, traverse resistance, hydraulic resistance and depth to the bitumen layer. However, information regarding the depth to bitumen layer was least important as RF needed four (4) variables to explain the behavior of the potential vulnerability of the groundwater. The model used 75% of the input data set to train the model while 25% was used to test the trained model. The resultant output prediction is shown in the table 5. Figure 12, shows the vulnerability map obtained through the application of RF algorithm to the most relevant explanatory variables related to the potential vulnerability of groundwater in the study area when the underlaying subsurface stratification is disrupted.

Table 5: Input data used to train and the RF predicted values

CASEID	TARGET	PRED	SAMPLE
1	**1**	**0.33331**	**LEARN**
2	**0**	**0.20586**	**LEARN**
3	**1**	**0.9063**	**LEARN**
4	**1**	**0.8066**	**LEARN**

5	**1**	**0.53577**	**LEARN**
8	**0**	**0.25648**	**LEARN**
10	**0**	**0.76022**	**LEARN**
11	**1**	**0.95794**	**LEARN**
12	**1**	**0.99109**	**LEARN**
13	**3**	**2.2141**	**LEARN**
14	**2**	**2.36891**	**LEARN**
16	**1**	**0.84129**	**LEARN**
18	**3**	**2.83836**	**LEARN**
19	**2**	**1.68586**	**LEARN**
20	**2**	**1.76304**	**LEARN**
6	**0**	**0.49936**	**TEST**
7	**0**	**0.17251**	**TEST**
9	**1**	**1.02024**	**TEST**
15	**1**	**0.83945**	**TEST**
17	**3**	**2.73773**	**TEST**

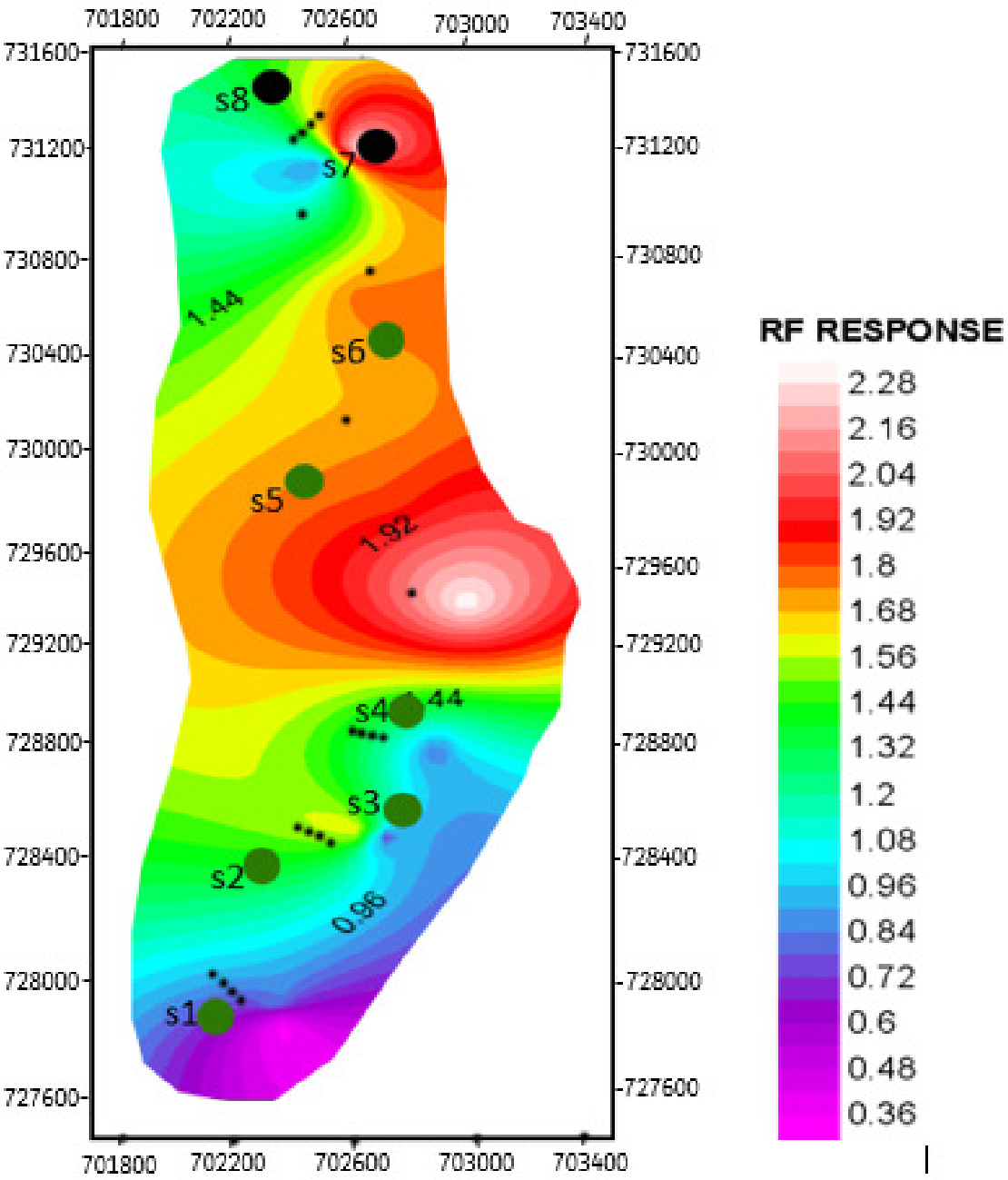

Figure 12: Potential Vulnerability map of the study area

Random Forest Potential Vulnerability Map

The model developed assisted in producing the potential Vulnerability map of the study area using the input features as criterion for classifying future highly vulnerable zones in the area. The generated map is presented in Figure 12**,** the southern part of the study area is observed to be the region with the least vulnerable part however moving northwards, it is observed that the central part trending north-eastwards of the study area represented with the red coloration ranges from 1.92 to 2.28 appears to be most vulnerable. This is of great significance because this part of the study area is the most inhibited by the residence of Agbabu; this points to the fact that when mining activities occurs in full strength, the water sources in the area will be grossly contaminated. Further northwards, it can be observed that high potential vulnerability ranging from 1.8 to 2.40 is predominant in the North-Eastern region, from the field observation, this North-Eastern region marked with high vulnerability corroborates with the area where mining activities had taken place in the past and this totally agrees with the presence of an abandoned well noted within this region. The effect of the past mining activities influenced the sub- surface stratification of this area hereby in turn influence the quality of groundwater in the area.

The Southern region of the study area is characterized with low RF response which infers that this region is the least vulnerable within the study area however moving northwards the potential vulnerability increases. This region of the study is currently lowly inhibited by the inhabitants of the community but when full mining commences within the area, this part of the study area might be the safest for citing of water wells for domestic purposes.

WATER ANALYSIS

A total of eight (8) water samples were collected in artesian wells. Five (5) samples were collected at Agbabu community, two samples were collected at Mile 12 and one (1) was collected at Ilubirin community which is to serve as control for the analysis. Sample 7 was an abandoned well in Mile 12 community.

Theresults of the analysis obtained from water chemistry laboratory analysis are presented in the table 6. Physico-chemical parameters evaluated were conductivity, total hardness, pH, salinity, alkanity, total dissolved solids, Calcium and Magnesium hardness, biocarbonate, Chloride, Sulphate and sodium. The conductivity and total hardness of the water samples were evaluated. The results of conductivity and total hardness ranging between 131µscm - 1400µscm and 29mg/l – 251mg/l respectively (Table 6). The conductivity indicates the ability of the water to conduct electric current and the total hardness indicates the ability of the water not to form lather easily with soap. The result of the conductivity analysis shows that every sample passed the test being within the permissible limits (Table 7) except for samples 3,6 and 7 where they exceeded the maximum permissible limit of 1200µscm and hardness shows that the samples ranged from 29mg/l – 251mg/l. All the total hardness results were all within the permissible limit except water sample result in well 2,5 and 7 with hardness value of 210 mg/L, 190 mg/L and 251 mg/L which all exceeds the maximum permissible of 180 mg/L. (Table 6). Total solids are the summation of total dissolved and total suspended solid in water. Total solid, total dissolved solid and total suspended solid of the water samples were evaluated, the results show the distribution of solid particles in the water sample and its solubility. the value of total suspended solid show the concentration of suspended material in samples 1,2,3,5 was all below the maximum permissible limit concentration of dissolved solids in for consumable water. However, the water samples result of total dissolved solid in samples 4,6,7,8 were above the permissible limit. Furthermore, alkalinity and PH values of the water samples were evaluated and the results obtained was between 24.5mg/l – 121.8mg/l and 6.40 -8.60 respectively (Table 6)**.** pH is a measure of how acidic/basic water is. The range goes from 0-14, with 7 being neutral. pHs of less than 7 indicate acidity whereas a pH of greater than 7 indicate a base. The pH of water is a very important measurement concerning water quality. The pH results were within the range of pH value for natural water (6.0 – 8.5) except for samples 1 and 7

which were below and above the maximum permissible Limits respectively(Table 1)**.** Complex ions such as biocarbonate, chloride and sulphate were evaluated, the results obtained was between 13mg/L- 252mg/L, 32.50mg/L –321mg/L and 7.14mg/L- 280.97mg/L respectively(Table 6).

The results of the samples passed the biocarbonate test except for the sample 8 which was above the maximum permissible limit. However, sample 6 was not within the permissible limit for Chloride and Sulphate while sample 7 exceeded the permissible limit for Sulphate.

The result of the major metals and heavy metals analysis obtained from the water chemistry are important in evaluating the concentration of element or ion presence in the water samples. Major metals such as Sodium, Calcium, Magnesium and heavy metals such as Calcium were studied. In the study area, all the major metals are present in varying concentrations in the groundwater especially Calcium, Magnesium were found well within the limits in all the studied samples however the results obtained for Sodium was within the permissible limit for samples 2,3,4,5,6 but samples 1,7 and 8 were either above or below the permissible limit. It ranged from 12.8mg/L to 207.8mg/L(Table 1)**.**

In summary, about 75% of the wells from which samples were collected averagely passed the physico-chemical tests with samples S1 to S6 parameters well within the permissible limits for a minimum of eight (8) parameters out of a total of eleven tested parameters. This shows that water from these well is currently consumable and for domestic purposes however 25% of the wells, samples S7 and S8 were distinct outliers well exceeding the maximum permissible limits for majority of the tested parameters depicting that water from these wells is not healthy for consumption. The latter wells, S7 and S8 are situated within the zones where mining had occurred in past years inferring that mining activity must have had a negative influence on the quality of the underground water in this area.

Table 4.5: Results of the analysis of physico-chemical parameters obtained from water chemistry

Samples	Northing	Easting	Conductivity (s)	Total Hardness (mg/l)	Ca Hardness (mg/l)	Mg Hardness (mg/l)	Ph	TDS (mg/l)	Bicarbonate (mg/l)	Chloride (mg/l)	Salinity (mg/l)	Sulphate (mg/l)	Na (mg/l)
S1	**702501**	**729172**	800 √	102 √ X	89.21 √	31.4 √	6.40 X	250.50 √	52.00 √	52.10 √	52.10 √	143.12 √	207.8 X
S2	**702612**	**729146**	679 √	210 √	32.08 √	15.2 √	6.59 √	300.50 √	32.00 √	32.50 √	24.53 √	63.44 √	43.7 √
S3	**702491**	**728901**	1251 X	36	102.09 √	51.3 √	6.61	128.50 √	14.00 √	50.45 √	33.21 √	86.87 √	10.5 √
S4	**702507**	**729079**	567 √	108 √ X	165.21 √ √	42.1 √	6.57 √	527.50 X	13.00 √	90.30 √	14.42 √	280.97 X	29.6 √
S5	**702449**	**729863**	1214 √	190 √	128.22	10.2 √	6.69 √	240.50 √ √	29.00 √	35.80 √	24.12 √	76.43 √	12.8 √
S6 (C)	**713450**	**751600**	700 √ X	29 √ X	56.21 √	9.3 √	8.55 √	150.10 X	19.00 √	300 √ X	25.21 √ X	255 X	25.9 √
S7	**702436**	**731540**	13(	√ -Pass	√	X	X	X- Fail	) √	321	208.00	76.57 √	251 X
S8	**702304**	**734145**	131 √	40 √	122.04 √	67.1 X	6.04 √	650.50 X	252.00 X	110.55 √	210.07 X	57.14 √	290 X

ACCURACY ASSESSMENT OF THE RF POTENTIAL PREDICTIVE MODEL

Model Validation using Confusion Matrix

Figure 12, shows the potential vulnerability map obtained through the application of RF algorithm to the most relevant geophysically derived explanatory variables related to the probability vulnerability of the groundwater within the area due to the occurrence of the underlying bitumen deposit.

Looking beyond bulk performance of the model, there is a wide range in performance with regard to predictive power of the RF as applied to individual classes. As shown in the confusion matrix in Table7, where the V_0, V_1 and V_2 are individual confusion matrix classes, the V_0, V_1 and V_2 classes produced accuracies with respect to the starting observed values, in the order of 90%, whereas the V_3 class exceeded 98%. It is likely that this excellent result noted with the V3 class is due to the spatially discrete and small area defined by the class, resulting in a very well-constrained class signature.

However, the overall accuracy of the RF model was computed to be 85% with an error value of 0.25. Omission error was computed to be between 0-20% while the commission error was computed to be between 0-33%. User accuracy (67-100%) and producer accuracy (75-100%) was also computed. Kappa value of 0.81 was obtained which according to table 8, means the model falls under the almost perfect region.

Model Validation using Result of Water Analysis

From Figure 13, the Northern part of the area which shows a region of high vulnerability coincides with sample 7 (S7) and Sample 8(S8). From the analysis of water from these wells, it was discovered that the water collected is not healthy for domestic consumption as it exceeded the permissible limits of several elements. It is however noteworthy to state that this is the region within the study area where mining operation used to occur which infers that the subsurface lithological stratification has been disturbed, this suggests reasons for the poor state of health of the water can be attributed to this subsurface.

Furthermore, samples 6 and 5 fall within a region of the potential map which suggests high vulnerability but the analysis of these samples showed water samples were within the permissible limits. This can be interpreted that these well sources would become unhealthy for consumption when the current subsurface stability of the lithology is disturbed. This area has a numerous number of people who reside.Samples 3,2 and 1 are within the region with low vulnerability but sample 4 infers medium vulnerability.

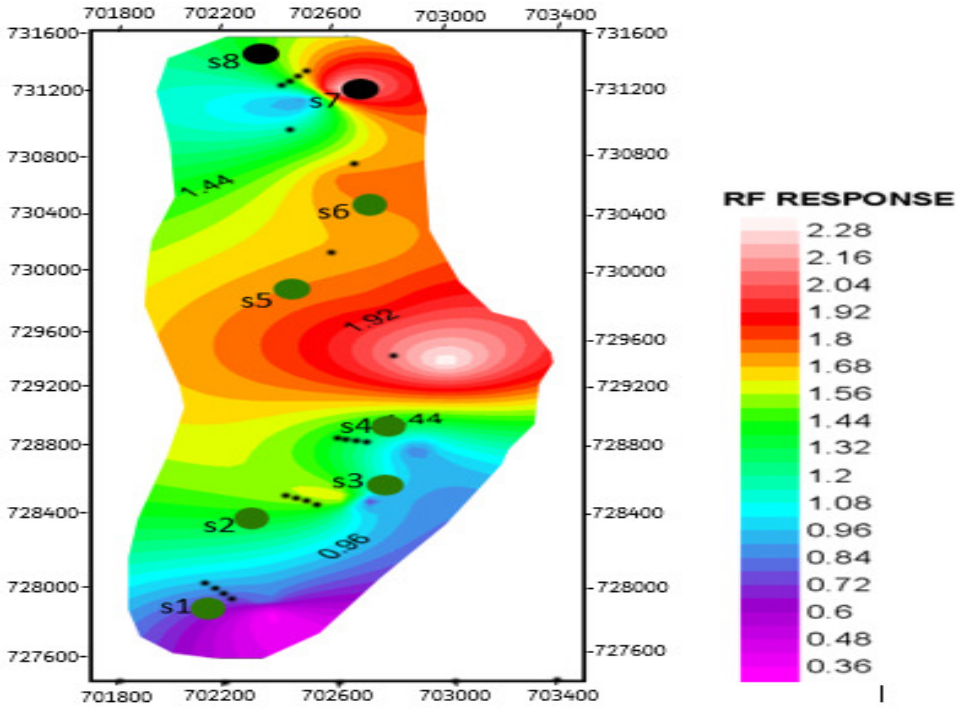

Figure 13: RF Potential Vulnerability model against Result of water chemistry

TABLE 7: Model Validation using Confusion Matrix

	V0	V1	V2	V3	Total	Commision Error	User's accuracy %
V0	4	1	0	0	5	20%	80%
V1	1	8	0	0	9	11%	89%
V2	0	0	3	0	3	0%	100%
V3	0	0	1	2	3	33%	67%
Total	5	9	4	2	20		
Omission Error	20%	11%	25%	0%			
Producer's Accuracy %	80.0%	88.9%	75.0%	100.0%			
Overall classification Error	85%						
Kappa Coefficient	0.81						

Table 8: Table showing Kappa values with their respective interpretation

Kappa Value	Interpretation
Below 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

CONCLUSION

In an attempt to predict the potential vulnerability of groundwater associated with present and future mining activities in Agbabu area, southwestern Nigeria, an integration of different geoelectric resistivity techniques was adopted to understand the hydrogeological variables that control the water quality of the study area.This informed the depth to which bitumen occurred within the area as well as the thickness of the lithologic layers across the area.

Geoelectric parameters (Total longitudinal conductance, Transverse Resistance, Hydraulic Conductivity and Hydraulic resistance) were determined to measure the vulnerability of the groundwater across the area. Results showed the region was generally dominated with low vulnerability however pockets of high vulnerability were deduced in the region where mining activities had once occurred in the past.

To produce a potential predictive vulnerability map, Random Forest (RF) Machine learning analysis was performed. Generated geoelectric parameters were used as input parameters for the model. These parameters were combined using the RF techniques to predict the region's most likely to be most vulnerable when the subsurface layering is disturbed by future mining activities. The central part trending North East zone of the study area which is currently densely inhabited by residence of the community is potentially most likely to be most vulnerable. This could also be probably being attributed to human induced activities within this region. Water samples were collected and sampled to obtain its chemistry and several physico-chemical parameters were measured. The result shows that the water within this area is averagely within the consumable limits however the region where mining had occurred in the past stood out as an outlier. This showed that mining activities had a direct effect of the quality of the groundwater while also stressing the result obtained from the Random Forest generated model. The results obtained in this study shows that when mining activities kicks off in full regard in the study area, groundwater in some parts of the region would become toxic and harmful for human consumption. This assertion is supported by Olabemiwo *et al*. 2011 who reported a statistical analysis of biochemical parameters of study rats fed with simulated bitumen leachates of Agbabu bitumen to have been significantly toxic to the health of the rats and in turn humans, it is therefore expedient that the results of this study be considered when developing the proposed Agbabu mine site.

REFERENCES

Abatyough T.M., Ibraheem A.O.,Christian, O. ,Adebiyi A., Yusuf J., Ahmed S.(2016) Multivariate Analysis of Under Ground Water Pollution Sources in Agbabu Bitumen Belt American *Chemical Science Journal 12(4): 1-10*

Ademilua O. L., Ojo O.F., Eluwole A.B., Ajayi C.A. (2014). Geophysical Investigation for Aquifer Potential Assessment and Groundwater Development at EKSU Staff Quartes GRA Ado Ekiti, Southwest Nigeria. *Journal of Environment and Earth ScienceISSN 2224-3216 (Paper) ISSN 2225-0948 (Online) Vol.7, No.5, 2017*

Adewole M.G. (2009). Environmental implications of bitumen seep induced pollution in parts of Ogun state, southwestern Nigeria*. Environ Earth Sci (2010) 59:1507–1514*
Aizebeokhai, A. P. (2010): 2D and 3D Geoelectrical Resistivity Imaging: Theory and Field Design. Scientific Research and Essays, 5 (23). pp*. 3592-3605. ISSN 1992-2248*

Akinmosin, A, Imo, D. (2016). Lithofacies types and influence on bitumen saturation in x-horizon of the nigerian tar sand deposits. *Ife Journal of Science vol. 18, no. 1 (2016)*

Alagbe O. A. (2020). Delineation of Bitumen Saturated Zones in Agbabu, Southwestern Nigeria, Using an Integrated Geophysical Methods. *Journal of Energy and Natural Resources. Vol. 9, No. 3, 2020, pp. 88-97. doi: 10.11648/j.jenr.20200903.11*

Alagbe O.A. and Sunmonu L.A., Olafisoye E.R., Adagunodo T.A.(2013). Geophysical and Hydro-physicochemical Evaluation of Hand-dug Wells near a Dumpsite in Oyo State, Nigeria*. Archives of Applied Science Research, 5(6), 29–40.*

Alagbe O. A (2010): Study on the Groundwater Accumulation of Oke-Ogba Area using Groundmagnetic Survey. Journal of Applied Sciences and Environmental Management *14(4). DOI:10.4314/jasem.v14i4.63251*
Amidu S.A. and Olayinka. A.I, (2006): Environmental Assessment of Sewage Disposal Systems Using 2D Electrical-Resistivity Imaging and Geochemical Analysis: A Case Study from Ibadan, Southwestern Nigeria*. Environmental and Engineering Geoscience (2006) 12 (3): 261–272. https://doi.org/10.2113/gseegeosci.12.3.261*

Amigun J. O., A. O. Adelusi and B. D. Ako (2012). The application of integrated geophysical methods in oil sand exploration in Agbabu area of Southwestern Nigeria. *International Research Journal of Geology and Mining (IRJGM) (2276-6618) Vol. 2(9) pp. 243-253, November 2012*

Antonio, C. de O.B., and Richard, F.F. (2014). Natural Vulnearbility Assessment to Contamiation of Unconfied Acquifers y Longitudinal Conductance- (s) Method. Journal of Geography an Geology, 6(4).

Bauman P (2005). 2-D Resistivity Surveying for Hydrocarbon-A Primer. CSEG Recorder, April , *pp.25-33*

Breiman L. (1984) Classification and regression trees. Chapman & Hall/CRC; 1984.

Breiman L. (2001) Random Forests. Mach Learn 2001;45(1):5–32.
Cardimona S. (1990): Electrical Resistivity Techniques for Subsurface Investigation Department of Geology and Geophysics, University of Missouri-Rolla, Rolla, MO.

Fayose, E.A. (1978). Depositional environments of carbonates of Calabar Flank, southeastern Nigeria-*Nigerian J. Min. Geol. 15: 1-13*

Freeman D.J. and Cattel F.C. (1990). Wood burning as a source of atmospheric polycyclic aromatic hydrocarbons. *Environ. Sci. Technol. 1990, 24, 10,1581 – 1585*

Gemail, K.S., El-Shishtawy, A.M., El-Alfy, M., Ghoneim, M.F., Abd El-Bary, M., (2011). Assessment of aquifer vulnerability to industrial waste water using resistivity measurements. A case study, along El-Gharbyia main drain, Nile Delta, Egypt. J. Appl. Geophys. 75, 140–150.
Hansen LK, Salamon P. Neural network ensembles. (1990). IEEE Trans Pattern Anal Mach Intell;12(10):993–1001.

Herrera M, Torgo L, Izquierdo J, Pérez-García R. Predictive models for forecasting hourlyurban water demand. *J Hydrol 2010;387(1–2):141–50.*

Khalil, M.A., Abd-Alla, M.A., (2005). An approach to estimate hydraulic parameters and water quality from surface resistivity measurements at wadi El-Assuity area, Egypt. NRIAG J. Geophysics., Special issue, 267–281pp.

Loos M, Elsenbeer H. Topographic controls on overland flow generation in a forest — an ensemble tree approach. *J Hydrol 2011;409(1–2):94–103.*

McGinnis S, Kerans BL. Land use and host community characteristics as predictors of disease risk. *Landsc Ecol 2012:1–16.*

Odunaike R. K., J. A. Laoye, O. O. Fasunwon, G. C. Ijeoma, and L. P. Akinyemi (2012). The application of integrated geophysical methods in oil sand exploration in Agbabu area of Southwestern Nigeria. *Internati6onal Research Journal of Geology and Mining (IRJGM) (2276-6618) Vol. 2(9) pp. 243-253, November 2012*

Odunaike R. K. , J. A. Laoye, O. O. Fasunwon, G. C. Ijeoma, and L. P. Akinyemi (2009). Geophysical mapping of the occurrence of shallow oil sands in Idiopopo at Okitipupa area, South-western Nigeria. *African Journal of Environmental Science and Technology Vol. 4 (1) pp. 034-044, January, 2010*
Ogungbemi, O.S., Ogunyemi, A. T.,Obaniwa, M. M. (2019). Geophysical Interpretation of Geological Features Constraining Bitumen Deposit In Agbabu, Southwestern Nigeria. *IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG) e-ISSN: 2321–0990, p-ISSN: 2321–0982.Volume 7, Issue 4 Ser. I (Jul. – Aug. 2019), PP 36-44*

Ojeyemi M. O., Adediran G.O., Adekola F.A., Adelowo O.O., Olajire A.A. (2014). Biodegradation of hydrocarbon compounds in Agbabu natural bitumen. *Africa Journal of Biotechnology vol. 13 (11), pp. 1257-1264, 12 March, 2014*

Ojuri O. O., Samuel A.O., David L. R.,Barker J.F.(2009): Contamination potential of tar sand exploitation in the western Niger-Delta of Nigeria: baseline studies. Bull Eng Geol Environ (2010) 69:119–128DOI 10.1007/s10064-009-0239-5

Olabemiwo O., Adediran G.O., FolahanA., Abass A.O. (2011). Impacts of Simulated Agbabu Bitumen Leachate on Haemotological and Biochemical parameters of Wistar Albino Rat. *Research Journal of Environment Toxicology 5 (3): 213-221, 2011*

Omatsola M.E., Adegoke O.S. (1981). Tectonic evolution and Cretaceous Stratigraphy of Dahomey Basin Nigeria*. J. Min. Geol. 1: 44 – 4*

Omosuyi G.O., Ojo J.S., and Olorunfemi M.O. (2008). Geoelectric Sounding to Delineate Shallow Aquifers in the Coastal Plain Sands of Okitipupa Area, Southwestern Nigeria. *The Pacific Journal of Science and Technology. Volume 9. Number 2. November 2008 (Fall)*
Quinlan J.R. (1993). Programs for Machine Learning. *Morgan Kaufmann Publishers, Inc.,*
Reyment, R.A. (1965) Aspects of the geology of Nigeria—The Stratigraphy of the Cretaceous and Cenozoic Deposits. *Ibadan University Press, 133 p.*
Reynolds, J. M. (1997). An Introduction to Applied and Environmental Geophysics. 796p.
Rodriguez-Galiano VF, Chica-Rivas M. (2012)- Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *International Journal Digital Earth 2012. http://dx.doi.org/10.1080/17538947.2012.748848.*

Rodriguez-Galiano VF, Chica-Olmo M, Abarca-Hernandez F, Atkinson PM, Jeganathan C. (2012)- Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens Environ 2012a;121:93–107*

Sesnie S, Gessler P, Finegan B, Thessler S (2008). Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sens Environ 2008;112(5): 2145-59.*

Steele B.M. (*2000)* Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote Sens Environ;74(3):545–56*

Ugbhor O.D., Ibuot C.J., Obiora N. D., Joy O.E. (2020). Geoelectrical investigation of groundwater potential and vulnerability of Oraifite, Anambra State, Nigeria. *Applied Water Science volume 10, Article number: 223 (2020)*

Vander-Velpen, B.P.A. 1988. "RESIST Version 1.0." M.Sc. Research Project. ITC: Delft Netherlands.

