The Binomial Logistic &Multiple Linear Regression-aided Mapping of Aquifer Sulphate Levels in the Landheer Area, Dadaab Sub-County

Meshack Owira Amimo[1], and Dr K.S.S. Rakesh[2,]

[1]Research Scholar, IIC University of Technology, Cambodia
[2]CEO, Gradxs, India

[1]bmoamimo@gmail.com
[2]kssrakesh@gmail.com

Abstract*-The Landheere centre is a newly unveiled settlement in the Dadaab subcounty, which is located some 6 to 9 kilometers away from the Laghdera flow course in Dagahaley, a refugee camp neighboring the Dadaab Axis zone, categorized as the central Merti Aquifer unit. The study was aimed at availing potable water to the community via Hydrogeological and Geophysical investigations. It was then decided that an aspect of water quality, sulphate levels, be also factored as waters with anomalous levels beyond 200mg/liter have been known to bear toxicity associated with health hazards for the nomadic pastoralist communities living in Garissa county. To predict the sulphate levels, secondary data of the Merti Aquifer was used to generate a predictive model based on Multiple Linear regression and Logistic regression. The original Merti hydrochemistry data was analysed using the correlation plot function in R and it gave a matrix table output showing the strongest predictors of sulphates, one of which was found to be the electrical conductivity, or EC. The EC of the areas closest to Landheer are well known. Dagahaley is one such area and has EC levels way below 900 mg/Liter, at the very worst. Since this EC may easily be measured in the field, alongside proposed site's coordinates and aquifer depths, the Models were used to predict sulphate levels expected in the borehole yet-to-be-drilled, and both algorithms determined that Landheer area is suitable for development of a well, to the extent that it is suitable in terms of water quality parameters. The MLR and Logistic Regression are thus useful Statistical Techniques ideal for the Water Resources Assessment for Development in the Merti Aquifer.

***Key words*- Merti Aquifer, MLR, Logistic Regression, Sulphate Levels, Electrical Conductivity, Correlation Plots**

I. INTRODUCTION

Sulphate levels in any groundwater systems poses a health risk and has to be factored before drilling and using any borehole water (Etim et al 2017). It was thus deemed imperative that a method of estimating sulphate levels be developed for use during exploration phase for landheere Area. The project area is located within Dadaab Township, off the Garissa-Dagahaley all-weather road. The Project targets a student population of at least 350 residents at the Education Centre station. Some 10.0-20.0 cubic meters per day yield is feasible from this well, given the hydraulics inferred from the geophysical curves generated. This is conditional on the depths recommended being attained.

The proposed Landheer Community borehole has the potential to supply more than up to 15000 liters per hour discharge, on average if done using mud drilling equipment.

II. PROJECT LOCATION

A. Location

The project area lies in the NEP Region of Kenya, within the Townshipsub-county. It is located on the southeastern sides of the main catchments course way. The area is defined by longitudes and latitudes shown in the geophysical curves analysed, and at an altitude of approximately 164m above sea level. Oblique dipping sediments litter the terrain alongside some zero degree dipping units of flood-prone Miocene Pliocene sediments.

B. Nature of the Project

This is a proposed well meant for sanitation and domestic use to run the newly unveiled rural settlement/ centre. The area has no water at the moment and survives mainly on water sourced via water bowsers from far away.

This will reduce the cost of water bills for the client, and shall greatly be of help in cushioning the business against operational halt, occasioned by lack of water.

The project shall have a daily requirement of approximately 18000 liters, computed from the projected population figures.

The borehole/shallow well will be developed upon drilling completion and should be preferably encased with steel-iron casing. Once the productivity of the borehole has been determined, a suitable submersible pump will be installed to pump water into the proposed storage tanks. The schematic design and the detailed itemization for the proposed borehole shall be the subject of phase two work for the planning and designing unit, but shall be predicated on the borehole performance in terms of aquifer yields and recharge.

C. Project Ownership

The proposed site is a public facility owned by the un-registered **Landheer Community Water Supply Project.** At an appropriate stage, the local community shall be trained on management related issues pertaining to sanitation, as well as borehole operations and maintenance. This shall be undertaken by the localWRMA office and the country government.

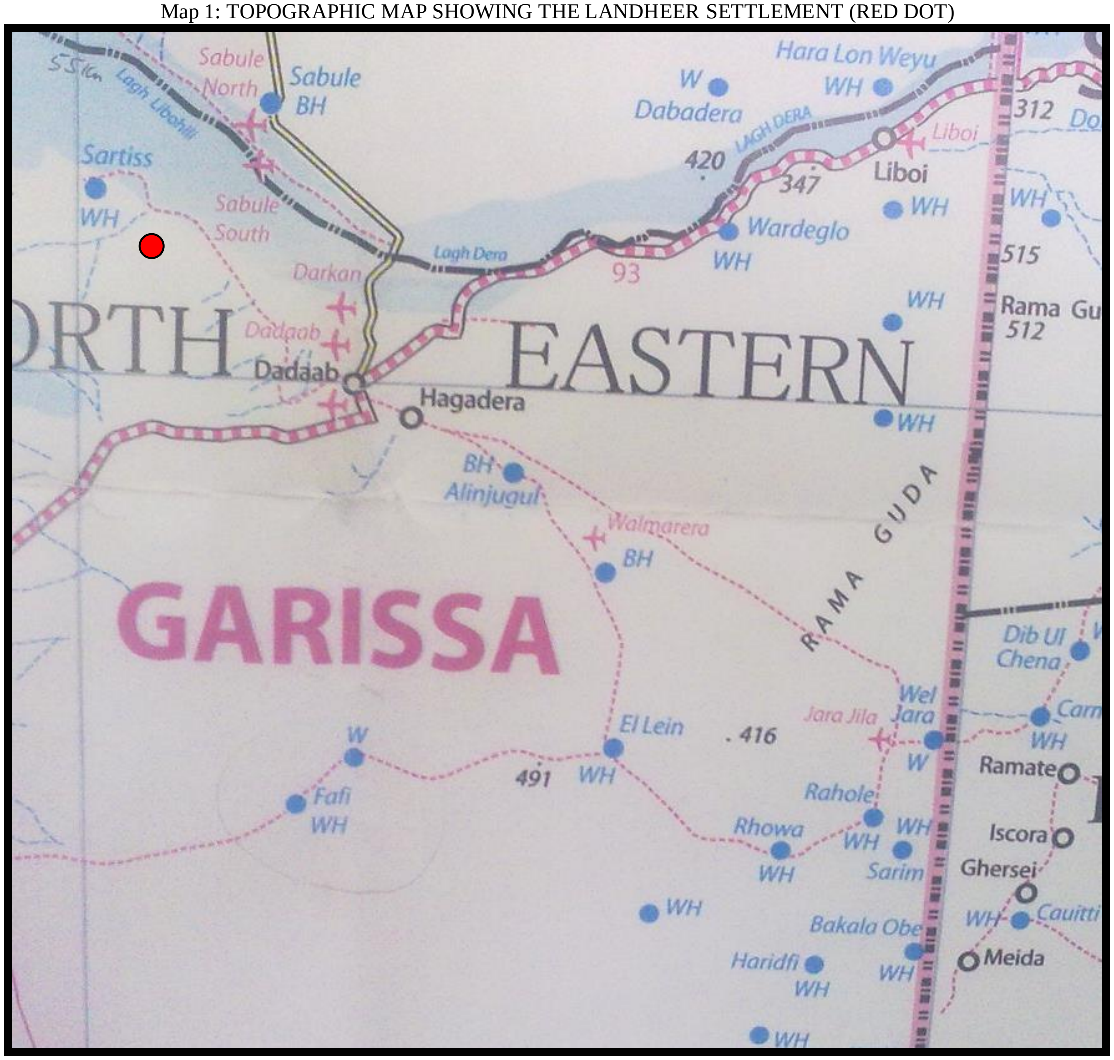
D. Site alternative and proposed action

Any site located within a radius of **1m** away from the pegged spot should have promising groundwater potential as the Stratigraphic units possess continuous permeability, both lateral and vertical.

The area stands at an average altitude of **164m** above sea level within a gently dipping terrain punctuated with several ant hills and flood plains both on the south eastern and north western flanks. The river flows in the northwest-southeastern azimuth.

E. *Physiography*

Map 1: TOPOGRAPHIC MAP SHOWING THE LANDHEER SETTLEMENT (RED DOT)



III. HYDROGEOLOGY

A. *Geology and Stratigraphy*

The topography is undulating dotted with several anthills which are clayey rich, and support vegetations that comprise mainly thorny shrubs, under growths and acacia family trees.

The geology is defined by red to light toned sandy clayey sediments, the Jurassic Corallite formation, which overlies the carbonates – namely corallites, aragonitic sediments and calcite. The sandy clayey species are mainly the Mariakani Sandstones series-mainly to a depth of 80m.

From 80m onwards, we have fine to coarse grained sandstones dominating the geology, alongside silts and gravel.

The Miocene/Pliocene story terminates at the Basement contact zone beyond 300m bgl

The Jurassic limestone carbonates are fairly fractured and possess water at the shallow depths, though highly mineralized, via the fractures and karstification veins. Water also forms at the contact points between the carbonates and the Archaean metamorphic basement units.

Groundwater in the upper sediments shall enjoy annual precipitation recharge through direct infiltration, while the deep-seated zones shall be recharged via regional flow aided by the karstification channels and plate tectonics in the Jurassic – cretaceous period. Evapo-transpiration rates of up to 3,000mm per annum over shadow the annual rains of up to 400mm per annum.

IV. HYDROLOGY AND STRUCTURAL GEOLOGY

A. Recharge Mechanisms within the Aquifer Systems

Rahole area lies within the Merti aquifer.Evidences abound of jointing and fracturing of the carbonate sediments on the surface, alluding to intense forces of fracturing, carbonation and quaternary tectonic faulting. Much of the south westerly – north easterly directed stress fields helped sculpture the terrain into its present geological state.

Owing to the relatively high fractions of clays in then beds, there is no sufficient time available for maximum river bed infiltration into the sub surface zones lying on the adjacent sides of the river course. However, there are areas that were exceptionally karstified and fractured within the carbonate beds. These are the zones that store water upon seepage into their aquifer sediments systems alongside recharging the adjacent sub surface storage systems via the Darcyan flow mechanics.

B. Drainage

Owing to the relative flat nature of the terrain, there is flood rampancy. The semi-permanent civil structures on the ground to stand the risk of destruction added to the occasional loss of lives for both livestock and human persons. Most of the housing units are constructed through shrubs and dry acacia trees locally available, lightening the task of evacuation in the event of impending flood disasters.

C. Climate

The project area falls within zone 7 of the classification of climatic/ecological zones of Africa, that is to say arid to semi-arid with temperatures averaging 30 to 34 degrees per day and occasioning evapo-transpiration rates of up to 3000mm per annum. The rainfall average falls well below 500mm per year.

V. LITERATURE SULFATE

Yesilanacar et al (2012) studied the sulphate levels in aquifers in Turkey using ANN as the tool, to help assess the levels of Sulphates in the local groundwater systems. In many rural communities of the Harran Plains, groundwater is the major source of water for livestock, domestic as well as agricultural demands envisaged. It was found that the water quality had been severely compromised by activities of humans, at the levels of sanitation, disposal and irrigation. Against this backdrop, the study proposed a cost effective means of frequently monitoring the water quality variations and fluctuations, and for this to work out, neural networks was proposed for use to help masses the parameter such as Sodium Absorption ratio and sulphate levels of the Harran aquifer systems. The EC, pH, levels of groundwater (gwl), aquifer temperatures, Total Hardness and levels of chlorides were used as input variables to perform the predictive forecasts using the ANN-based back propagation algorithm, the Leveberg – Marquadt. Both experiments and assessments for the two variables (SAR and SO4 levels) yielded an accuracy level of 96 percent, hence proving that the secondary data of wells in the aquifers may be used to cheaply monitor the aquifer water quality dynamics, using the neural networks, thus.

A study into role of sulphates in aquifer water quality aspects was undertaken by Hossain et al (2016). The study highlighted the geologic sources, patterns of distributions, and major controlling variables for the process of mobilization of arsenic in Kumamoto basin catchment, in Japan. Sulphates are associated with high or anomalous levels of arsenic, a toxic heavy metal in some aquifer systems, as this study revealed. The area was studied in relationship to the determination of the prevalent redox processes active in the aquifer systems and sub-systems. The study involved analyzing both hydrogeological factors and geologic samples. The study revealed that arsenic levels in the local aquifers range between 0.1 and 60.6 milligrams per liter. The high arsenic levels dominate poorly aerated environments, with stagnant groundwater which flowed in zero to low velocities. These waters were endowed with high levels of dissolved iron, manganese, and aluminum, with moderate levels of dissolved bicarbonates , phosphates , and sulphates, as well as low levels of nitrates which suggests active reducing conditions prevalent in the of sub-surface hydrogeology systems of the study area. The study revealed that some three variables were considered crucial in understanding the major sources and cause of anomalous levels of Arsenic in the aquifer systems of Kumamoto area, namely, the high groundwater pH levels which imply highly carbonated waters, anoxic redox environments as well as the stagnant groundwater in water species considered of younger age, which have not been pumped out since the sediments were buried decades ago.

Redwan et al (2016) studied ground water chemistry in Egypt within the Tahta catchments. The areas groundwater was considered to suffer from various water quality degradation mechanisms such as extensive urbanization, agricultural and industrial activities, and common in many developing nations such as Egypt.

This study was undertaken out to map out the variables controlling the changes in the groundwater chemistry, especially in the area west of Tahta, Sohag, and Upper Egypt. The study involved the construction of piper diagrams which were subsequently used in the study and understanding of the aquifer. The diagrams indicated the predominance of Na–Cl (seventy five percent) with minor Calcium-Sodium-carbonate and the calcic-chloride water-types. The equilines were then studied to avail more info into the aquifer history and water chemistry. The equiline diagrams and ionic ratios indicated the dominance of $Ca^{2+} + Mg^{2+}$ over $Na^{+} + K^{+}$, as well as the $HCO_3^{-} + SO_4^{2-}$ over Cl^{-} (the carbonate-Sulphate-Chloride) , which indicated the limestone-sulphate-silica mineralization thermodynamics of the aquifer, and subsequent dissolution into the aquifer as a factor. Moreover, the reverse ion exchange reaction models and the Gibb's diagram did indicate that the bulk of the anions and cations in the aquifer are mainly derived from processes active in the geology:

i) Geological aspects of water–rock interaction and

ii) The geochemical evaporation-crystallization dominances.

The study further brought to the fore the fact that the Pliocene clays were the major sources both chlorides and sodic ions in the aquifers. Overall, the study revealed that geological–hydrogeological mapping of the aquifer information relating to anions and cations was useful in understanding the hydro-chemical patterns and mapping out the anthropogenic contributions, in impacting the groundwater quality.

Barzegal et al (2017) undertook a study of the aquifer hydrochemistry of some parts of Iran in 2017. The study had aimed at evaluating the geology and hydrology of the Tabriz plain aquifers in NW Iran, through both analytical chemistry and GIS methods to pick out spatial variations in the distributions of cations and anions in the aquifers. Aquifer levels of Sulphates were mapped out in the study and deemed to play a major role in water quality, meaning it significantly impacted both TDS and the EC. To undertake the goals therein, aquifer water sampling from thirty shallow wells and deep boreholes in the study area were targeted for mapping and analysis in 2012. The water samples were analyzed for various physicochemical parameters such as pH, Electrical conductivity, sodium levels, calcium, potassium, magnesium+, chlorides, carbonates, bicarbonates, sulphates and nitrates.

The study and analysis generally implied that in the aquifers, there is an active process of dissolving the following minerals into the water systems:

a) Gypsum,
b) Anhydrite,
c) Halite and
d) Silicate minerals.

These were observed to occur frequently across the study area, whilst the aquifers were deemed to be super-saturated, with respect to both calcites and dolomites, thanks to calcification and dolomitization processes, respectively, in the area's geology. The geochemical cross plots indicated that geo-weathering and dissolution of various rocks and minerals are the major co-determinants of water quality in the study area.

Miller et al (2021) focused on aquifer geochemistry and its influence on the groundwater quality in the study area. This study focused on the differences in sorption behavior of fluoride on activated alumina, between simple sodium fluoride solutions and multicomponent groundwater samples with high alkalinity. The study highlighted the influence of sulphates and bicarbonates on fluoride mineralization in an aquifer, albeit in real life situation as the experiment was in a laboratory where various levels of carbonates, fluorides and sulphates were set to react, and timing undertaken to measure rates of reactions and behavior, especially time it took to attain equilibrium.

Bordbar et al (2021) undertook a study involving the of ML techniques to assess water quality. The main objective of the study was to integrate adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural network (ANN) to design an integrated supervised committee machine artificial intelligent algorithm to assess the water quality of parts of northern Iran. The performance of each algorithm was noted. The values of correlation coefficients attained stood at percentage values of 88, 87, and 80, respectively for neural networks, support vector machines and the fuzzy inference models. The aspects of salinity intrusion threats being mapped had major contributions from bicarbonates and sulphates. Banadkooki (2020) also employed ML techniques to assess water quality aspects in ran and chief amongst the parameters deemed of significance was sulphates. The overall quality of the groundwater is important, since it helps establish the suitability of the aquifer sourced-water for drinking, irrigation, as well as for domestic purposes. In that study, three algorithms were employed, namely, the adaptive fuzzy interface system (ANFIS), support vector machines (SVM), and artificial neural network (ANN), primarily to help map the TDS of the aquifers. The hydrological data were sourced from Yazd plain study to help predict the Total Dissolved Solids (TDS). The ANFIS models were used to help predict TDS of the local aquifers. As opposed to the ANFIS models, the ANN multilayer perceptron (MLP), and SUPPORT VECTOR MACHINES or SVMs models, the hybrid ANFIS, ANN, and SVM demonstrated high accuracy in the training and testing stages of model development.

The study thus preferred the traditional ML algorithms over the new ANFIS models owing to time-tested and tried performance accuracies. It is against this background that the present study sets to predict sulphate levels in aquifers as well.

VI. THEORY AND METHODS USED

The present study employed the use of traditional statistical assessment, the Multiple Linear Regression, to map the sulphate levels at the Landheere area of Dadaab subcounty. The aim of the predictions was to assess the sulphate level expected in the well that is schedule to be drilled in November 2021, and whose technical study, Hydrogeological Surveys, has already been undertaken. The threshold limit as recommended by WHO is 200mg/L of sulphate in groundwater and this is deemed healthy. Any borehole whose water has a level above this is considered unsuitable, though some livestock and plants may tolerate it, suiting it for both animals and agriculture, respectively.

A. Multiple Linear Regressions

1) Multiple Linear Regression Analysis

The statistical procedure of multiple linear regressions, also known as MLR, is an algorithm which enables a researcher to establish the strength of association existing between two or more independent variables, and a single dependent variable, being predicted. Take the example of the use of aquifer depth, distance from a river and salinity.

It is known in hydrology that the total dissolved solids in an aquifer may be very strongly associated with depth to a well and its distance away from a river, as the river exerts dilution effect on the well during laminar flow recharge for a well located near a river.

This equation may be written as hereunder:

TDS=β_0 + $\beta_{1.}$ (Depth)+ $\beta_{2.}$ (***Distance to River)***

To explain MLR, let's assume that a research scientist assessing aquifer salinity levels has data had developed a scoring system that enabled her to predict an individual's body mass index with the TDS values as well as the depths and radii away from recharging stream.

The data is as hereunder:

A	B	C
depth-M	lateral dis	TDS
30	120	188
45	176	277
60	240	400
78	312	468
100	400	620
125	500	755
150	600	911
200	800	1190

Fig 1: Sample data to illustrate how logistic regression models work in predicting class.

Using the R software, the codes will yield the results appearing in the table hereunder:

```
> summary(model)

Call:
lm(formula = TDS ~ ., data = mydat)

Residuals:
         1          2          3          4          5          6
-1.385e+01 -5.387e-13  2.259e+01 -1.474e+01  8.519e+00 -2.778e+00  6.92
         8
-6.669e+00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     26.294     12.459   2.110   0.0886 .
depth.M         -6.777     17.093  -0.396   0.7081
lateral.dist..M  3.157      4.263   0.741   0.4923
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.78 on 5 degrees of freedom
Multiple R-squared:  0.9986,    Adjusted R-squared:  0.9981
F-statistic:  1830 on 2 and 5 DF,  p-value: 6.873e-08
```

Fig 2: Display of output predictions and correlation coefficients for the sulphate prediction model using Logit.

The figure above shows a strong correlation coefficient, R square value of 0.9986 which means we may use the MLR equation to predict any of the variables of given two of them , in future, using the calculator model generated by the dataframe we have above.

See the equation below.

TDS=β_0 + $\beta_{1.}$ (Depth)+ $\beta_{2.}$ (***Distance to River)***

We shall input the values generated in the model and the prediction equation therefore becomes as thus:

TDS=**26.294** +-6.78*(Depth)+ **3.157** *** (Distance to River).**

Suppose one now wishes to drill two wells. The first well is to be drilled to depth 135m. The other well is drilled to 180m bgl. Both are located 550m and 730m away, respectively from the river flow course. Use the equation above to estimate the TDS levels expected in the two wells.

TABLE I: TABLE SHOWING DEPTH OF TWO WELLS CLOSE TO A RIVER

SNo	Depth - m	Lateral dist away from River -m
1	135	550
2	180	730

Solving:

a) TDS=26.294 +-6.78*(Depth)+ 3.157 *(Distance to River).
Depth=135m, dist to river =550m
TDS=26.294+-6.78*135+3.157*550
TDS=847.34
TDS=847.3 mg/litre

One may look at the TDS of the well done 500m away from the river, in the table and find that the TDS was 755mg/L one concludes that salinity increases with distance away from the River, in this illustration case example.

b) TDS=26.294 +-6.78*(Depth)+ 3.157 *(Distance to River).
Depth=135m, dist to river =550m
TDS=26.294+-6.78*180+3.157*730
TDS=1110
TDS=1110 mg/litre

One may look at the TDS of the well done 800m away from the river, in the table and find that the TDS was 1190mg/L. our example is 730m away, and has 1110 mg/litre which is less than the 1190 mg/L as our radius away from river is also less at 730m.

2) *Logistic Regression Model*

This is a predictive algorithm which utilizes independent variables to facilitate the prediction of the dependent variable class, unlike the traditional regression which predicts a continuous variable rather than class. The variables in question may be numerical or rather be factor/categorical variables, being used to predict class, and this means the dependent variable MUST always be categorical in Logistic regression. The logistic regression algorithm is therefore a statistical model which employs the logit or logistic function to model the conditional probability of an event using several variables.

Consider the case of for binary outcome logistic regression cases. In this instance, one computes the conditional probability of the dependent variable, which we may call Y, given independent variable X.

This may be expressed mathematically as thus: P(Y=1|X) or P(Y=0|X)

Then above expression is interpreted as thus: The conditional probability of Y=1, given X or conditional probability of Y=0, given X.
P(Y |X) is therefore approximated as a sigmoid function applied to a linear combination of input features. At a high level, logistic regression works a lot like good old linear regression. So let's start with the familiar linear regression equation:

Y = $\boldsymbol{\beta_0 + \beta_1 * X}$
In linear regression, the output Y is in the same units as the target variable (the thing you are trying to predict). However, in logistic regression the output Y is in log odds. Now unless you spend a lot of time sports betting or in casinos, you are probably not very familiar with odds.
Odds is just another way of expressing the probability of an event, P (Event).

Odds = P (Event) / [1-P (Event)]
Suppose a drilling company drilled 100 wells and got water in 70 of them. Using this as an example, the probability striking a productive aquifer in the area is 70%. My odds of striking water in the areas geological material may be calculated as:

Odds = 0.70 / (1–0.70) = 2.333
All mathematics of probabilities is bounded between 0 and 1, which becomes a problem in regression analysis. The Odds will always range from 0 to infinity.

3) Logistic Regression Function
Logistic regression uses logit function, also referred to as log-odds**;** it is the logarithm of odds. The odds ratio is the ratio of odds of an event A in the presence of the event B and the odds of event A in the absence of event B. The relationships are captured page hereunder:

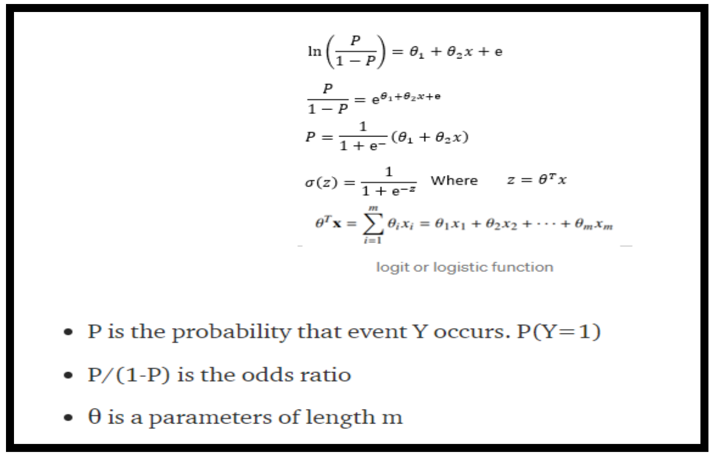


Fig 3: The Main components of a Logistic Regression Equation

The Logit function will always estimate probabilities to be between 0 and 1, and hence logistic regression is a non-linear transformation that looks like S- function shown below. It is thus named-Sigmoid curve.

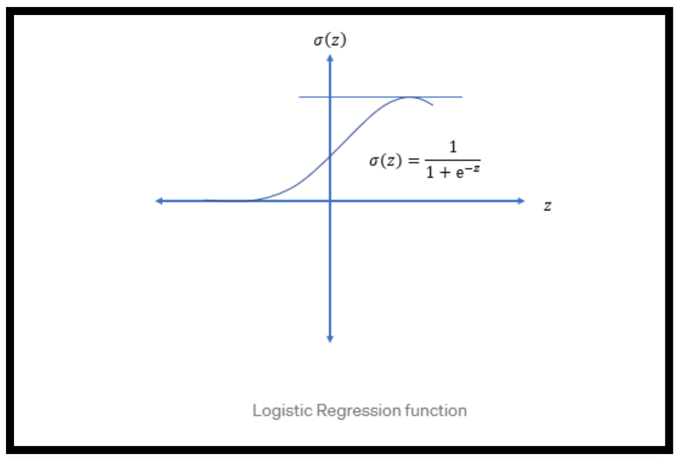


Fig 4. The logistic regression function which is sigmoid in geometry, tracing s-shaped figure in outlook.

The parameter denoted by Greek letter θ or theta of the logistic function algorithm may be estimated using the maximum likelihood estimation or MLE framework. The MLE searches for the parameters that best fit the joint probability of the independent variables denoted as X. Let's now use the 100 well drilled earlier as an example here. The MLE will give us values for parameter "θ" that would maximize the probability close to 1 for the borehole/well which will be having water, but a value close to zero for all wells that would have been declared dry after drilling.

4) Confusion Matrix to Evaluating the Performance of Binary Classification

Once one has performed the logistic regression, the model may be used to understand the complex interrelationships inherent in the dataframes. One way to simplify the analysis is via the use of a confusion matrix. A confusion matrix is a table that describes how many actual values and predicted values exist, for various classes captured /predicted by the model. In some technical literature, it is referred to as the Error Matrix.

	Actual Positive	Actual Negative
Predicted Positive	True Positive(TP)	False Positive(FP) (Type 1 Error)
Predicted Negative	False Negative(FN) (Type 2 Error)	True Negative(TN)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Population}} \qquad \text{Error Rate/Misclassification rate} = \frac{\text{False Positive} + \text{False Negative}}{\text{Total Population}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{Predicted Positive(TP+FP)}} \qquad \text{Sensitivity/Recall} = \frac{\text{True Positive}}{\text{Actual Positive(TP+FN)}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{Actual Negative(FP+TN)}} \qquad \text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}}$$

Fig 5: The Matrix Table Explaining how Classes in Logit Models Are interpreted in statistics.

Explaining the Above Table

- **True Positive**: When a borehole is productive, and we predicted that it is productive
- **True Negative**: When a borehole is not productive, and we predicted that it is not productive
- **False Negative**: When a borehole is productive (the well has water after drilling), but we predict that it is not productive. This is also known as the Type **2 error.**
- **False Positive**: A borehole cannot be productive (the well is dry after drilling), but we predict it is productive. This is also known as the Type **1 error.**

VII. GEOPHYSICS

A. Field Investigation Methods

1) Introduction

A great variety of geophysical methods are available to assist in the assessment of geological subsurface conditions. In groundwater exploration, the most widely applied techniques are electrical resistivity surveying, electro-magnetic (EM) profiling, seismic refraction and geophysical borehole logging. Other, less common investigation tools are induced polarisation (IP) surveys, magnetometer surveys, the gravity method and airborne geophysics.

In the present survey, the vertical electrical sounding technique was applied. An ATLAS-COPCO DC resistivity set was used, comprising an ABEM SAS 4000 Terrameter, connecting cables and clips, stainless steel non-polarising current electrodes and copper potential electrodes. This dedicated equipment measures both □*V* and *I* and presents a computed resistance. Using geo-metrical constants (K-factor), the apparent is calculated, and plotted as a function of the current electrode spacing AB.

The resistivity method is briefly discussed in the following sections.

2) Resistivity Method

Basic Principles of the Resistivity Method

The resistance R of a certain material is directly proportional to its length L and cross-sectional area A, expressed as:

R = □ * L/A (□) (1),

Where□ is known as the specific resistivity, characteristic of the material and independent of its shape or size. With Ohm's Law:

R = □V/I (□) (2),

Where□V is the potential difference across the resistor and I is the electric current through the resistor, the specific resistivity may be determined by:

□ = (A/L) * (□V/I) (□m) (3)

The electrical properties of rocks in the upper part of the earth's crust are determined by the lithology, porosity, and the degree of pore space saturation and the salinity of the pore water. These factors all contribute to the resistivity of a material (the reciprocal of the electrical conductivity).

The resistivity of earth materials can be studied by measuring the electrical potential distribution produced at the earth's surface by an electric current that is passed through the earth. Vertical electrical soundings are point measurements that provide information on the vertical resistivity layering at a certain location. Resistivity profiles, on the other hand, are carried out to obtain information on lateral changes in apparent resistivity.

3) Resistivity Sounding Technique

When carrying out a resistivity sounding, also called vertical electrical sounding (VES), an electrical current (I) is passed into the ground through two metal pins, the current electrodes. Subsurface variations in electrical conductivity determine the pattern of current flow in the ground and thus the distribution of electrical potential. A measure of this is obtained in terms of the voltage drop (□V) between a second pair of metal stakes, the potential electrodes placed near the centre of the array. The ratio (□V/I) provides a direct measurement of the ground resistance and from this, and the electrode spacing, the apparent resistivity (□$_a$) of the ground is calculated.

The measuring setup consists of a resistivity instrument (usually placed in the middle of the array), connected to two current electrodes (AB), and two potential electrodes (MN) towards the centre. Usually a so-called "Schlumberger" array is used for vertical electrical soundings, while profiles are generally carried out in "Wenner" configuration.

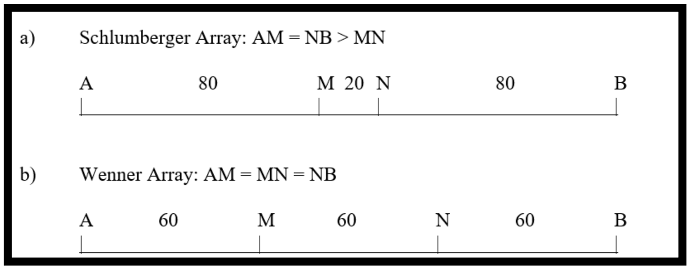


Fig 6: Examples of Schlumberger and Wenner Configurations
(For Resistivity Measurements, where: AB = current electrodes; MN = potential electrodes)

A series of measurements made with an expanding array of current electrodes (Schlumberger Array) allows the flow of current to penetrate progressively greater depths. The *apparent resistivity* as a function of the electrode separation AB provides information on the vertical variation in resistivity. The depth of penetration varies according to the electrode array, but is also affected by the nature of the material beneath the array. The point at which a change in earth layering is observed depends on the resistivity contrast, but is generally of the order of a quarter of the current electrode spacing AB (Milsom 1989). By contrast, in a homogeneous medium the depth penetration is of the order 0.12 AB (Roy & Apparao 1971).

The calculated apparent resistivity is plotted against current electrode half separation on a bi-logarithmic graph paper to constitute the so-called sounding curve. The curve depicts a layered earth model composed of individual layers of specific thickness and resistivity.

Interpretation of field data can be done with hand-fitted curves, but this method is time consuming, and practically limited to 3-layer solutions. Modern interpretation is computer-aided, using a curve fitting procedure based on a mathematical convolution method developed by Ghosh (1971).

While the resistivity method is a useful tool in groundwater investigations and borehole site surveys, its applicability and reliability should not be overestimated. The modelling of field data is often attended by problems of equivalence and suppression. Each curve has an infinite number of possible solutions with different layer resistivities and depths (this is known as *equivalence*). Mathematical convolution can easily lead to a well-fitting solution, which nonetheless does not correspond to reality. In general, the number of possible solutions is reduced by mutual correlation of several sounding curves, knowledge of the local geology and drilling data.

When deposits with similar resistivities border each other, it is usually not possible to make a differentiation. Intermediate layers, occurring between deposits of contrasting conductivity, may go undetected, as they tend to be obscured within the rising or falling limb of the sounding graph (*suppression*). Additional data, in the form of borehole records, air photography and geological field observations, are required to produce a realistic interpretation.

It should be noted that the layered earth model is very much a simplification of the many different layers, which may be present. The various equivalent solutions, which can be generated by a computer programme, should therefore be carefully analysed. In general, resistivity soundings should never be interpreted in isolation as this may lead to erroneous results.

4) Resistivity Profiles

Resistivity profiles are usually carried in Wenner configuration, i.e. an electrode set-up with a uniform distance between potential and current electrodes.

The entire array is moved across the area of interest. By doing so, lateral changes in apparent resistivity are measured, which reflect variations in the lithology, the depth of weathering or the water content. So-called "anomalies" may indicate the intersection of a fault (usually a negative anomaly), quartzite band (positive anomaly) or buried riverbed (anomaly depends on nature of surrounding deposits). Usually such lineaments, which may also be observed on aerial photographs, are linked to the occurrence of groundwater.

It must be noted that resistivity differences in a single profile array may largely reflect variations at the surface rather than underground. For this reason, it is usually not sufficient to carry out single-spaced profiles. The depth of penetration increases at greater electrode separations. A series of profiles at variable electrode separations will provide an indication of vertical resistivity trends. Moreover, by repeating the same profile at a different configuration, it will become clear if the observed resistivity patterns are caused by surface phenomena or underground features.

5) Geo-electrical Layer Response

Vertical electrical soundings (VES) provide quantitative information on electrical resistivity as a function of depth. The computer-interpretation of the sounding data produces a layered model of the underground. The derived resistivity layers are used to infer the presence of water-bearing strata, their texture and salinity.

Water-bearing and/or weathered rocks have lower resistivities than unsaturated (dry) and/or fresh rocks. The higher the porosity of the saturated rock, the lower its resistivity, and the higher the salinity (or electrical conductivity EC) of the saturating fluids, the lower the resistivity. In the presence of clays and conductive minerals the resistivity of the rock is also reduced. The relation between the formation resistivity (ρ) and the salinity is given by the “Formation Factor” (F):

$$\rho = F \times \rho_w = F \times 10{,}000 / EC\ (\mu S/cm),$$

Where: ρ_w = resistivity of water

In sediments or unconsolidated layers produced by weathering, the formation factor varies between 1 (for sandy clays) and 7 (for coarse sands).

Example: If a certain aquifer is considered with an average formation factor of 3, then an EC of 300 µS/cm will give a formation resistivity of 100 Ωm. The same material, when containing water with an EC of 1,500 µS/cm, will have a resistivity of only 20 Ωm. brackish water is marked by an EC of 2,000 to 10,000 µS/cm, and a ρ_w of 5 to 1. Deposits containing brackish water will therefore in most cases adopt a low formation resistivity (usually less than 10 Ωm). Saline water with an EC of about 30,000 µS/cm will reduce the resistivity of a formation to about 2 Ohm.M.

Clayey formations with fresh water will respond similarly to deposits with brackish or saline water: the fact that the same resistivity can be obtained for completely different hydrogeological units is known as the "equivalence-problem".

Fresh and dry Basement rocks are marked by very high resistivities, with a common range from 1,000 to 10,000 OhmM. Moderately to slightly weathered but dry layers are less resistive, and usually show values between 100 and 500 Ohm.M, depending on the portion of clays, the degree of weathering and the water content. The resistivity further decreases if the deposits are water-bearing, to 20 to 200 Ωm. The resistivity of impermeable clay layers (alluvial or produced by intensive weathering of clay-forming minerals) usually varies between 2 and 10 OhmM, while similar figures are recorded for aquifers with brackish to saline water.

The greatest difficulty in the interpretation of resistivity measurements in Basement rocks is formed by:

a) *Equivalence*: the similar geophysical properties of layers with contrasting hydrogeological characteristics (e.g. clay layers and layers with brackish water),
b) *Absence of distinct layer boundaries*: the decreasing degree of weathering with depth is usually not well-defined, but gradual. This will result in a gradual increase in resistivity, and not in a distinct set of geophysical layers.
c) *Suppression #1*: Potential aquifer layers of moderate thickness may fail to show a significant response in the recorded resistivity data (especially where these are deep). Thin aquifers embedded within a very thick deposit can easily remain undetected by surface geophysics. They will however show up in down-hole geophysical logs.
d) *Suppression #2*: The resistivity contrast between the (clayey) weathered zone and the fresh bedrock may be so high, that an intermediate sap rock aquifer cannot be distinguished in the graphic plot of the sounding.

Despite the problems of suppression attributed to the large resistivity contrast between fresh and weathered basement (point *d*), this is also a favorable attribute. Because of the large difference, the depth of weathering can be measured quite accurately. Considering that aquifers often occur towards the boundary of the weathered zone and the bedrock, the drilling depth can be determined, even if the actual aquifer does not show up as distinct geophysical layer.

To determine Project Areas Hydrostratigraphy and potential for a borehole, **1No. Vertical Electrical** Soundings were undertaken using an **ABEM SAS 4000B Terrameter**. The analyzed data is hereunder tabulated:
The geoanalysed data is presented in the table overleaf.

VES 001-this is the first site done at Landheere which lies near an anthill and shows excellent promise for groundwater. The site is located on a flat terrain with vegetation cover.

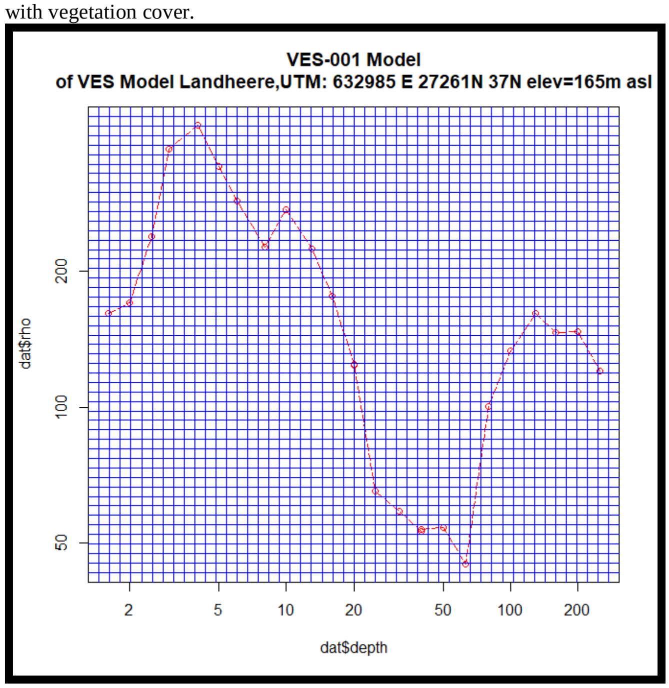


TABLE 2: LANDHEERE SURVEYS

Resistivity Curve No-R	Schlumberger Probe Depth Interval(m)	Resistivity In OhmM	Expected Geological sediment/Formation	Comments

001/2021 The first site near the main road shown by Mzee Ali Issa	0-1	161	Top Soils	Barren
	1-5	34	Subsoils	Barren
	5-20	125	Hard Corallites	Barren
	20-63	448	Fine sands /clays	Barren
	63-130	161	Fractured medium sst	Minor Aquifer
	130-160	145	Coarse sandstones	Minor Aquifer
	160-200	146.5	Medium sandstones	Major Aquifer
	200-250	147.1	Coarse sandstones	Major aquifer
	Over 250	Infinity	coarse medium sst	Major aquifer

6) MLR with R Software on the Sulphate Levels in Landheere Area

Model 1-Shows the correlation plot matrix table, displaying the correlation coefficient, relating EC and sulphates in Merti aquifer as 0.91. Since R^2 is this big, knowing the value of EC may enable to predict the value of sulphates, or even just the approximate values of sulphates, in an area during insitu tests for water quality in an area via the use Of Multiple Linear Regression.

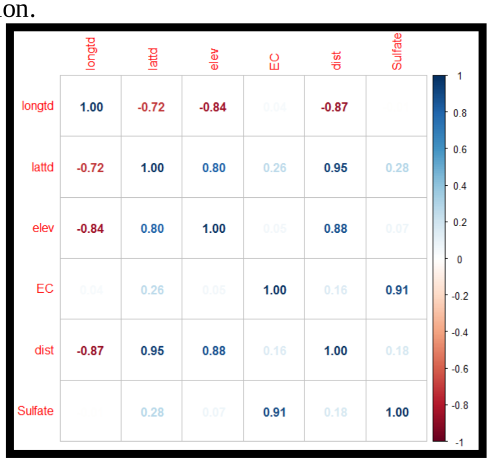


Fig 7: Correlation Plot Matrix Table

7) Discussion for the Modeling Using R for the Sulphate Predictions in Landheer Area of the Merti Aquifer

Sulphate data of the Merti aquifer was used in this study. Following the assessment that employed the correlation plot function, the major predictors of sulphate levels were derived and these stood out as longitudes, latitudes, elevations and Electrical conductivity. The distance between every well in the Merti away from the Laghdera flow course was measured as well and spelt bout in a separate column of its own.

The parent data class of the Merti aquifer which was used to generate predictive model for sulphates in MLR was as shown hereunder:

1	longtd	lattd	elev	EC	dist	CSulfate
2	38.647	1.062	293	973	204	verySafe
3	39.6547	1.14458	256	4540	133	Unsafe
4	40.18014	0.34175	172	1456	34	verySafe
5	38.69197	1.02384	291	2740	172	verySafe

Fig 8a: Data Extract Showing Parent Data that was analyzed

After training and testing of the data, Multiple Linear Regression model was developed in R. This model was subsequently employed to perform predictions on 6No sites, four of which had acceptable levels of sulphates and the two remaining did have unacceptable levels. The NWWDA has a potable field test kit which may easily measure the TDS and EC levels in the field. The areas under study have existing wells whose water was tested for EC, and the coordinates also taken using a GPS meter carried to the field to map each spot. Since it was possible to measure the EC and TDS, the new table looked like thus:

TABLE 3: COORDINATES OF THE STUDY AREA

SNo	Longitude-UTM /DD	Latitude - UTM/DD	Elevation-M	EC	Dist from Laghder a river
1	40.018	1.618	260	2600	180
2	39.15833	1.81137	300	1411	215

The column to be predicted, VALUE of sulfate LEVELS is missing. This is what the algorithm was used to predict.

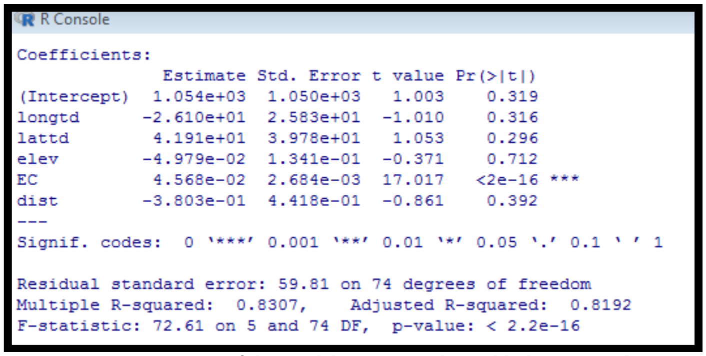
Model 2-coding for the model in R

```
# enter the data into R using the code below
sulphDat=read.csv("PRDMsulphate.csv",header=T,na.strings
="NA")

str(sulphDat)

# sample the data of sulphates in the Merti aquifer
head(sulphDat)
```

```
> head(mydat)
    longtd   lattd elev   EC dist Sulfate
1 38.64700 1.06200  293  973  204    31.0
2 39.65470 1.14458  256 4540  133   323.0
3 40.18014 0.34175  172 1456   34   180.0
4 38.69197 1.02384  291 2740  172     1.2
5 39.31385 0.90669  214 6520  143   406.0
6 40.20805 0.28844  135  978   30    21.0
```

Fig 8: R console showing the rows and columns of data as used in sulphate levels assessments.

```
# attach the dataset
attach(sulphDat)
#see the model generated by MLR
model=lm(Sulfate~.,data=sulphDat)
```



```
R Console
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.054e+03  1.050e+03   1.003    0.319
longtd      -2.610e+01  2.583e+01  -1.010    0.316
lattd        4.191e+01  3.978e+01   1.053    0.296
elev        -4.979e-02  1.341e-01  -0.371    0.712
EC           4.568e-02  2.684e-03  17.017   <2e-16 ***
dist        -3.803e-01  4.418e-01  -0.861    0.392
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.81 on 74 degrees of freedom
Multiple R-squared:  0.8307,    Adjusted R-squared:  0.8192
F-statistic: 72.61 on 5 and 74 DF,  p-value: < 2.2e-16
```

Fig 9:Summary Statistics of the Model in R Console Highlighting Correlation Coefficient of More Than 80 Percent and Which Is Suitable For Use to Predict

# View Model summary statistics
summary(model)

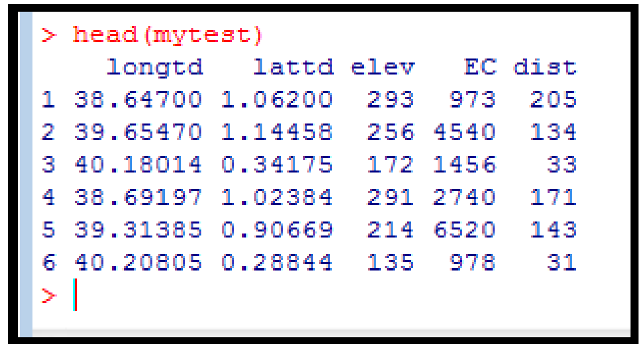
# call in the library dplyr
library(dplyr)

# rename the data as df1 now and df2
df1=sulphDat
df1
df2 <- mutate_all(df1, function(x) as.numeric(as.character(x)))
df2
# generate the corrplot library to compute relationships
# between various variables in a matrx table
library(corrplot)
see the outputs
M <- cor(df2)
corrplot(M, method = "number")

TABLE 4: SULPHATE VALUES IN WATER

SN	Sulphate value-in mg/L	Comment
1	31	Safe
2	323	Unsafe
3	180	Safe
4	1.2	Safe
5	406	Unsafe
6	21	Safe

Interpretation of table above: The original values of sulphate levels in the waters existing already in the six towns are captured above. The replacement wells in the same towns will naturally bear the same levels of sulphates as for the existing wells. The prediction model using MLR predicted the results shown hereunder.



```
> head(mytest)
    longtd    lattd elev   EC dist
1 38.64700 1.06200  293  973  205
2 39.65470 1.14458  256 4540  134
3 40.18014 0.34175  172 1456   33
4 38.69197 1.02384  291 2740  171
5 39.31385 0.90669  214 6520  143
6 40.20805 0.28844  135  978   31
> |
```

Fig 10:- The field data used to test algorithm.

TABLE 5: SULPHATE LEVELS IN WATER

SN	Predicted Sulphate value-in mg/L	Comment
1	41	Safe
2	210	Unsafe
3	65	Safe
4	132	Safe
5	298	Unsafe
6	42	Safe

Interpreting above table- the data shows correct predictions in that the actual values predicted vary, but still within the WHO thresholds or limits defined as either safe or unsafe. The results conclusively put to rest the fears that the MLR algorithm is unsuited to the data.

B. Binomial Logistic Regression

In order to compare the results obtained using MLR, a classification algorithm was used on the same dataset. A field dataset was tested on the algorithm and it generated more or less similar results. There are two main methods used in logistic regression analysis. When the data sets has only two outcomes of, say, YES or NO, it is essential that one uses the binomial Logistic regression. This binomial logistic regression will predict the probability to an observation falling into one of two classes in a dataset with a dichotomous response variable, usually based on one or more predictors, which may either be continuous or factor variables(also known as categorical variables in the event that dependent variable refers to a count, one may use Poisson regression algorithms. The present task employed the use of multinomial logistic regression to make the informed predictions of sulphate class. From the foregoing, the following is true:

a) One may use binomial regression algorithm to predict a data analysis task with just two classes.
b) One may use multinomial logistic regression algorithm to make predictions on a data with either TWO classes or MORE classes. The multinomial algorithm for logistic regression can handle many classes, from two onwards, without compromising accuracy levels.

The dataset expressed sulfate levels as a class rather than continuous variables, this time as capture hereunder:

1	longtd	lattd	elev	EC	dist	CSulfate
2	38.647	1.062	293	973	204	verySafe
3	39.6547	1.14458	256	4540	133	Unsafe
4	40.18014	0.34175	172	1456	34	verySafe
5	38.69197	1.02384	291	2740	172	verySafe
6	39.31385	0.90669	214	6520	143	Unsafe
7	40.20805	0.28844	135	978	30	verySafe
8	39.00675	2.07735	344	1307	256	verySafe
9	40.32534	0.23798	159	1073	28	verySafe
10	39.08145	1.96014	321	1294	222	verySafe

Fig 11: Parent Dataset Extract for Predicting the Sulphate Class

PART TWO-PREDICT CLASS NOW using logit models

 # predict sulphate levels in a new area using the model generated

mydatXX=read.csv("sulphateClass.csv",header=T,na.strings= "NA")

head(mydatXX)

Logistic Regression

library(nnet)

 # where CSulphate is a class of binomial order-Verysafe or Unsafe

longtd, lattd, elev, EC, an dist are continuous predictors

fit <- multinom(CSulfate~.,data=mydatXX)

 # predictions of class of sulphates

The **nnet R**library was sued to generate models for logit classification schemes for the Merti sulphate levels. A new field data was used here to test efficacy of the algorithm. It was decided that three datasets be from areas with unsafe sulphate levels and three others be from areas with safe levels of sulphate. The model correctly predicted each case. See the codes hereunder:

mypredd=read.csv("testSO4.csv",header=T,na.strings="NA")

head(mypredd)

predd=predict(fit,mypredd)

predd

predd=as.data.frame(predd)

predd

```
> predd
     predd
1 verySafe
2 verySafe
3 verySafe
4   Unsafe
5   Unsafe
6   Unsafe
>
>
```

Fig 12: Output table shows the results of the analysis using logistic models with amazing accuracy of prediction of sulphates for all the sites tested.

summary statistics

summary(fit) # display results

The output stats shows the equation which generated the prediction model in R as shown hereunder and which is what generated the probability values assigned , so that values of 1.00 or almost one (greater than 0.5) were assigned the class of Verysafe, whereas the other one which was below 0.5 or less assigned Unsafe levels of sulphates.

```
> summary(fit) # display results
Call:
multinom(formula = CSulfate ~ ., data = mydatXX)

Coefficients:
                   Values   Std. Err.
(Intercept) 12.381944424 0.003479926
longtd       0.281547337 0.147558041
lattd        1.547958522 0.003620907
elev        -0.010413636 0.012351981
EC          -0.006143099 0.001429339
dist        -0.032497965 0.014761516

Residual Deviance: 26.60325
AIC: 38.60325
>
```

Fig 13: Model showing the standard errors of logistic regression as well coefficients of the equations used.

explaining the confidence intervals of the coefficients generated by nnet

confint(fit) # 95% CI for the coefficients

```
AIC: 38.60325
> confint(fit) # 95% CI for the coeffic
                   2.5 %       97.5 %
(Intercept) 12.375123895 12.388764953
longtd      -0.007661110  0.570755783
lattd        1.540861676  1.555055369
elev        -0.034623073  0.013795802
EC          -0.008944552 -0.003341646
dist        -0.061430004 -0.003565925
>
```

Fig 14: Output table showing the ranges of values highlighting minima and maxima for the coefficients of all the variables employed to generate the model

exponentiated the coefficients

This was done to make the values more easily to understand and interpret in terms of how they relate to the model. The lowest range is for EC which seems to be arguably the best predictor of sulphates in the model. The values are -0.008944552 all the way up to -0.003341646 for EC, respectively for 2.5 and 97.5 CI intervals

```
# where F is a binary factor and
# longtd, lattd, elev, EC, an dist are continuous predictors
fit <- multinom(CSulfate~.,data=mydatXX)


## predict all pr

pr=predict(fit,mydatXX)
pr


##use the table command to evaluate accuracy
tab=table(Predicted=pr,Actual=mydatXX$CSulfate)

tab

TOTAL=100+2+1+147
```

Fig 15:Codes for Computing Model Accuracy.

When the codes above are run in R, the output is 98.8 percent and this is indicative of reliability of using Logistic regression to map the sulphate levels in the study area.

sees class based on probability calculated

Here one may see the classes predicted by using the model for all the rows entered in the dataframe. The word 'class' as used here indicates the fact that we are predicting the levels of sulphate in terms of whether deemed acceptable or otherwise for each point.

classPreds1=predict (fit, type="class") # predicted values
classPreds1

```
> tab=table(Predicted=pr,Actual=mydatXX$CSul
>
> tab
          Actual
Predicted  Unsafe verySafe
  Unsafe      100        2
  verySafe      1      147
>
> TOTAL=100+2+1+147
>
> CORRECT=100+147
>
>
> accur=CORRECT/TOTAL
>
> accur
[1] 0.988
> |
```

Fig 16: Show Reliability of The Model For Prediction Based On High Accuracy Levels Posted.

From the model, it is noted that the category of wells deemed unsafe has two misclassifications, whereas the category of Verysafe possess just one misclassification. Overall the accuracy is overwhelmingly reliable at **nearly 99%.**

Real Life Applications for Sulphate Level Predictions in Landheer Area

The Landheer area was mapped using geophysical and hydrological methods to determine suitability for drilling water which is proposed for use by humans and animals alike. The Two algorithms were used to compute its drilling suitability. The data predicted was as shown hereunder.

TABLE 6:

SNo	Longitude-UTM /DD	Latitude-UTM/DD	Elevation-M	EC	Dist from Laghdera river
1	40.1950317	0.24659	165	900	9

The area is around nine kilometers from Dagahaley area refugee camp, whose water has around 900 mg/L EC levels. Due to meandering of the roads, the distance from Landheer to the Laghdera course is estimated at 9 kilometers. The first attempt was done using MLR and results presented in the outputs hereunder:

```
> head(mytest)
   longtd   lattd elev  EC dist
1 40.19503 0.24659  165 900    9
> sulphatePred=predict(model,mytest)
>
> sulphatePred
       1
44.24952
>
> sulphatePred=as.data.frame(sulphatePred)
>
> sulphatePred
  sulphatePred
1     44.24952
> |
```

Fig 17: Output Table for MLR Prediction of Sulphates for Landheer Area

a) The area has acceptable levels of sulphates estimates o be at approximately 44.2 mg/liter. This shows that the water shall be acceptable for both livestock and humans alike. The water is thus not of any threats to human life. This info is now known and may be used for planning purposes by the entity that wishes to sink the wells.
b) Part B now involves predicting of class of sulphate levels using logistic regression. Once again, the model returns an output of 'very safe' on the new data for landheer, coded as 'landherTest'. The results that were generated as shown hereunder:

```
> head(mypredd)
     longtd   lattd elev  EC dist
1 40.19503 0.24659  165 900    9
>
> predd=predict(fit,mypredd)
>
> predd
[1] verySafe
Levels: Unsafe verySafe
>
> predd=as.data.frame(predd)
>
> predd
     predd
1 verySafe
>
```

Fig 18**:** The Data is named Landheer test And Is Predicted as 'Verysafe'.

VIII. CONCLUSION AND RECOMMENDATIONS

The MLR and Logistic regression Modeling techniques have proved to be an efficient modeling tool and may be used to map out the sulphate models in the Merti aquifer areas , both the ones already drilled and the areas yet undrilled , taking advantage of the strong correlation existing between Electrical Conductivity and the levels of Sulfate in the Merti. Both methods of regression are efficient for use in Water Resources Management and Development Planning. The R software was used and also proved to be a powerful mapping tool for aquifer hydrochemistry. The study found out that evaluating sulphate levels is also an indirect way of estimating aquifer EC levels.

REFERENCES

[1] Amimo Meshack Owira, Hydrogeological & Geophysical Investigations of Landheer Area, 2021, Northern Water Works Development Agency (Technical Surveys Report Repository,NWDDA)

[2] Banadkooki, F. B., Ehteram, M., Panahi, F., Sammen, S. S., Othman, F. B., & Ahmed, E. S. (2020). Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *Journal of Hydrology*, *587*, 124989.

[3] Barzegar, R., Moghaddam, A. A., & Tziritis, E. (2017). Hydrogeochemical features of groundwater resources in Tabriz plain, northwest of Iran. *Applied Water Science*, *7*(7), 3997-4011.

[4] Bordbar, M., Neshat, A., Javadi, S., Pradhan, B., Dixon, B., & Paryani, S. (2021). Improving the coastal aquifers' vulnerability assessment using SCMAI ensemble of three machine learning approaches. *Natural Hazards*, 1-22.

[5] Etim, E. U. (2017). Occurrence and distribution of arsenic, antimony and selenium in shallow groundwater systems of Ibadan Metropolis, Southwestern Nigerian. *Journal of Health and Pollution*, *7*(13), 32-41.

[6] Hossain, S., Hosono, T., Ide, K., Matsunaga, M., & Shimada, J. (2016). Redox processes and occurrence of arsenic in a volcanic aquifer system of Kumamoto Area, Japan. *Environmental Earth Sciences*, *75*(9), 740.

[7] Millar, G. J., Couperthwaite, S. J., Dawes, L. A., Thompson, S., & Spencer, J. (2017). Activated alumina for the removal of fluoride ions from high alkalinity groundwater: new insights from equilibrium and column studies with multicomponent solutions. *Separation and Purification Technology*, *187*, 14-24.

[8] Redwan, M., & Moneim, A. A. A. (2016). Factors controlling groundwater hydrogeochemistry in the area west of Tahta, Sohag, Upper Egypt. *Journal of African Earth Sciences*, *118*, 328-338.

[9] Yesilnacar, M. I., & Sahinkaya, E. (2012). Artificial neural network prediction of sulfate and SAR in an unconfined aquifer in southeastern Turkey. *Environmental Earth Sciences*, *67*(4), 1111-1119