Groundwater Sodium Level Assessments in the Dadaab Subcounty in Northern Kenya: A Case Study of the Dertu Area

[1]Meshack OwiraAmimo, [2] Emmanuel Ombaka, and, [3]Jibril Shune

[1]Research Scholar, IIC University of Technology, Cambodia
[2]Lecturer, Department of Mathematics, Garissa University
[3]Hydrology Assistant, Northern Water Works Development Agency,Garissa Town

[1]bmoamimo@gmail.com
[2]eombaka2000@gmail.com
[3]jibreelshune@gmail.com

***Abstract*—Groundwater sodium content was studied in the Dertu area following the 2008 complaints that pregnant mothers were having problems relating to hypertension, still-births and stunted growth for the babies born there, alongside weak bones for these children. The local leadership opined that the water source being used could play a role in these calamities.**

Sodium content in the Merti aquifer waters close to and far from the study area were analyzed with a view to helping make informed predictions regarding potential groundwater development in the area.

To that effect, data table frames of historical data were assembled and analyzed using ML algorithms.

The study reveals that the anomalous levels of Na in the Waters is purely geo-genic, derived from the geological interactions between the flowing groundwater and the parent rocks in the study area. Samples of groundwater wells were analyzed and, were also examined using WHO and KEBS standards for water quality, which indicate that groundwater in a selected sampling sites, were unsuitable for domestic purpose and irrigation.

The study also reveals that the sodic mineralization varies from one place to another and as such, there would be some specific portion of the study area that one can pinpoint as being free of Na mineralization: there were as many places with anomalous sodic levels, as were those with acceptable Na+ levels. The prediction model so developed is deemed to be helpful in predicting the sodic content of an un-drilled proposed borehole site , given geospatial parameters and selected groundwater hydraulic variables. The model thus proves useful in making decisions under uncertainty, as far predicting sodium content is concerned.

Keywords— **Water quality, Groundwater, Chloride, Sodium, Potassium.**

I. INTRODUCTION:

Dertu Township is located 75 kilometers to the North of Garissa Township, and lies within the Dadaab subcounty. It is on a rough road, the Garissa-Dertu-Abakorrey-Habaswein Highway. The residents used to subsist on an Earthpans in the earlier years up to 1999, when a well was sunk for them.

The present sources of water are three wells that were sunk between 2001 and 2018. The wells serve both livestock and humans.

Since then, there has been a steady in the cases of miscarriages for pregnant mothers, and incidents of children taking too long to walk, alongside that of tooth problems amongst the residents. Weak or deformed bones have been noted in some of the children.For the male residents, blood pressure seems to have also risen, despite the fact that the residents walk long distances as they graze their livestock, thereby giving their bodies enough physical exercise, deemed healthy. They suspect that the cause may have to do with groundwater quality of the study area. The present study is not medical oinany way, but just seeks to establish a way of knowing bthe probable Na+ contents before drilling, then making decisions on whether to go ahead and drill, is site has acceptable levels of Na+ from the predictions made, thus.

Since human population has been increasing over the years, from less than 200 in 2001 and over 3000 at present, it has become necessary to undertake a study into the groundwater quality of the existing wells in the area, and study the possible means of predicting the Na+ levels beforehand, so as to intervene against the Non-Communicable Diseases deemed inherent, and also against undertaking irrigation using groundwater which is most likely already contaminating the soils with excess Na+ ions.

The objective of the present study is to apply **Machine Learning and the hydrogeological data** to predict localities within Dertu area, where potable groundwater with acceptable levels of Na+ ions may be abstracted, based on secondary data obtained from the Water Resources Authority (WRA) and the Northern Water Works Development Agency (NWWDA) on the longitudes (longtd) and latitudes (lattd), electrical conductivity (EC), Total Dissolved Solids (TDS), altitude (m) and well depths

(m) to determine the probability of a borehole being saline or fresh. The aforementioned factors have been proved to conclusively help determine the groundwater sodium levels in the study area.

The factors that determine groundwater quality are drainage hydrology, geology, climate, depth, and aquifer depths (wsl).

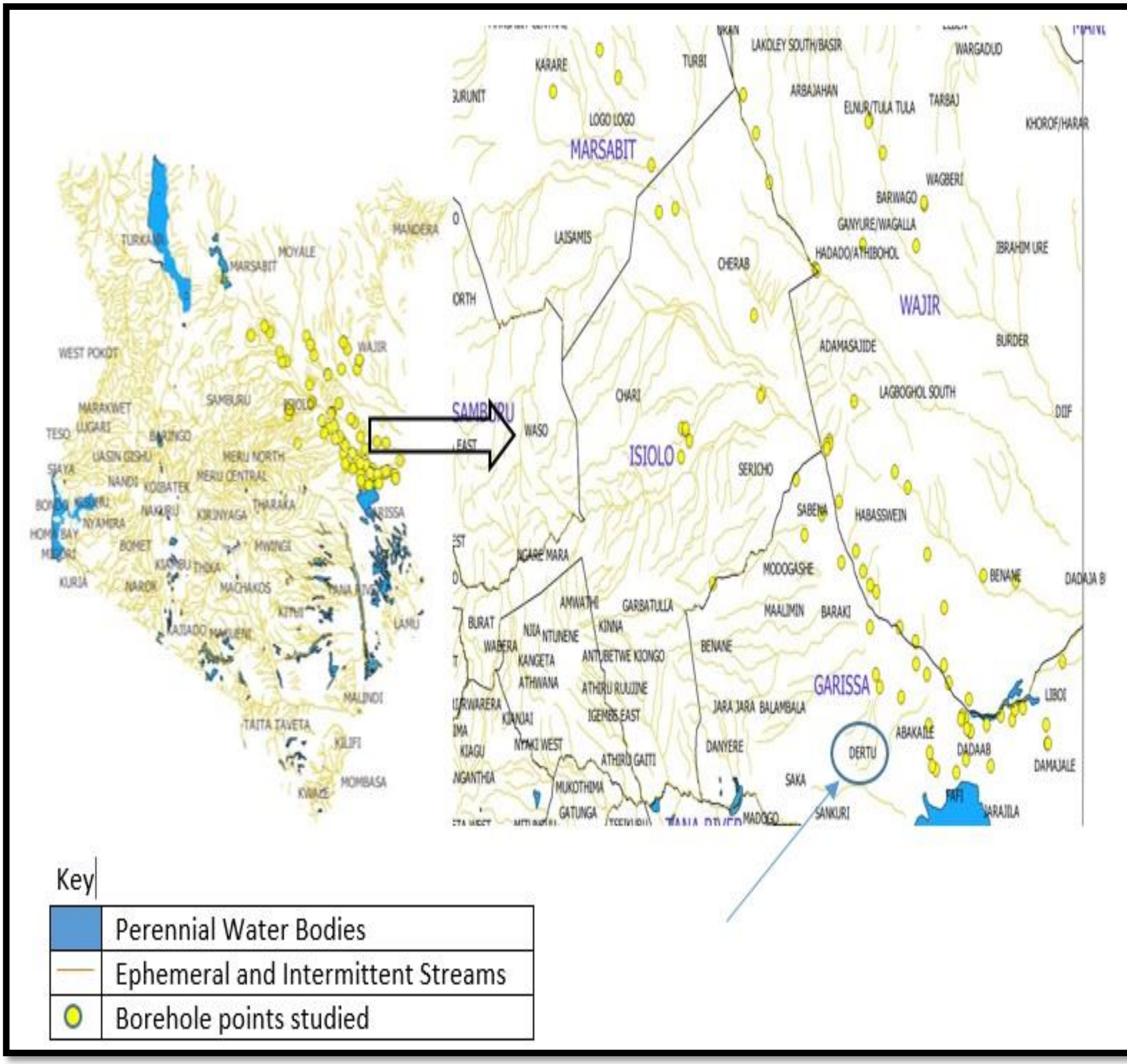


Fig 1: Map showing Borehole study points in the Merti aquifer area in the wells mapped for sodium levels

The study area is located in the north eastern region of Kenya covering Boroansis, Ahmed Tukalley, Baraki ,Gurufa, Dadaab, Dertu, and Shantabaq well Data, covering latitudes 0°05' south and 1°16' north and longitudes 38°40' and 41°52' east. The region has an altitude range of between 150m to 365m above sea level. However, the models generated shall be representative of the whole Merti aquifer.

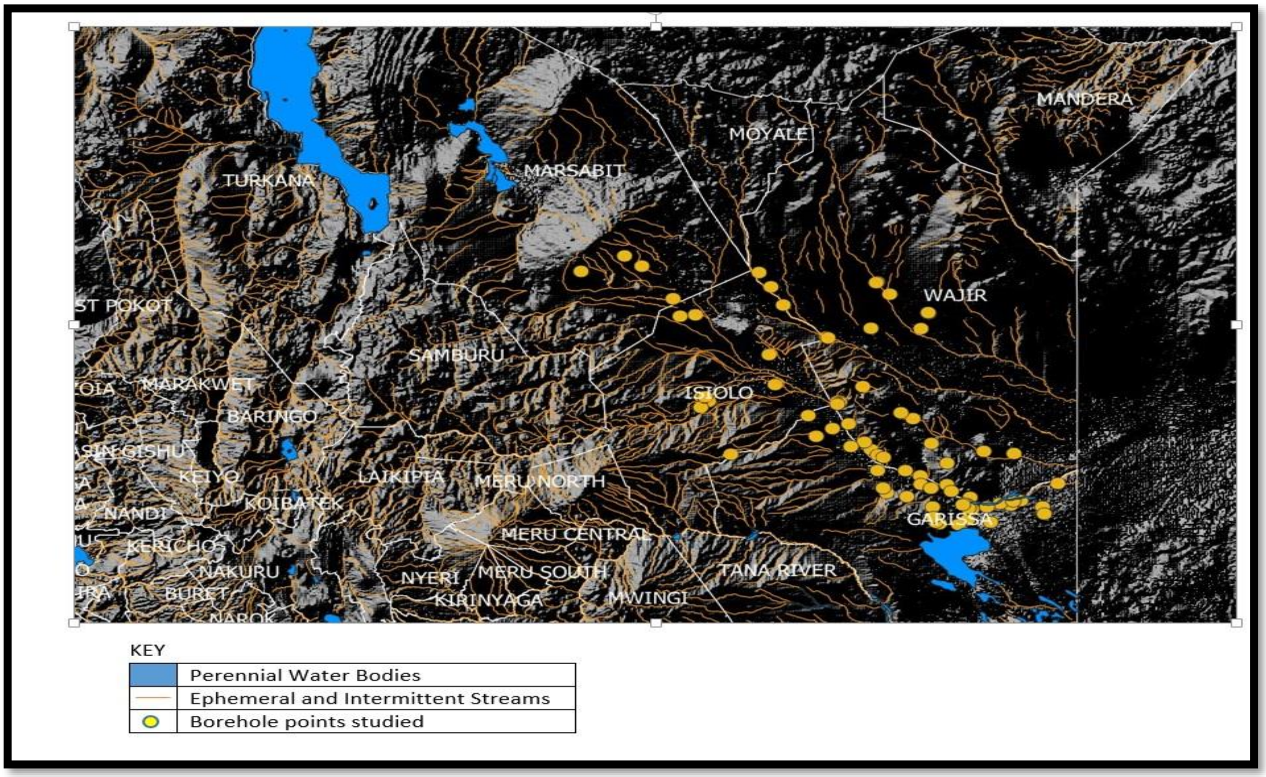


Fig 2: A hill-shade map showing the ephemeral and intermittent streams distribution in the study area

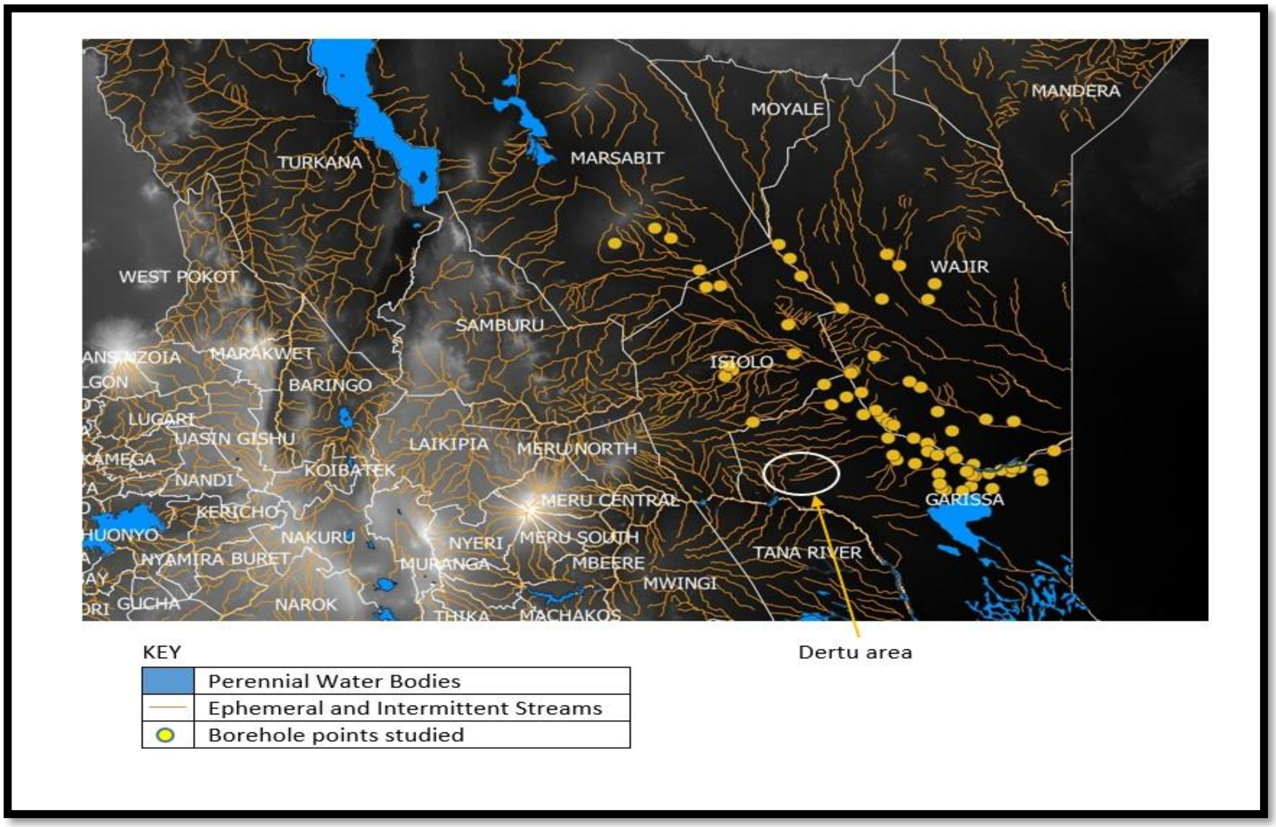


Fig 3: A single-band gray altitude map showing Dertu location in the Merti aquifer study area

II. RESEARCH METHODOLOGY

In our present study, the study team visited the Dertu area and its environs all the way from Bahuri to Fafi Kalala, Dertu centre and Boroansis, and also from Dertu to Shantabaq, baraki, Gurufa, Dadaab, Kamuthe, and Medina settlements.

This present study collected unpublished groundwater, geophysical and hydrological data **from over seventy** exploration undertaken within the period around **December 2019, all the way to December 2020** within the area of study. Additional data was obtained from the County Directorate of Water, Water Resources Authority (WRA) sub regional office in Garissa, as well as data from UNHCR Dadaab field office. The aim was to use groundwater chemical analysis data to analyze the sodium distribution geochemistry in the groundwater systems, and help make predictions on the expected sodium content. The aquifers are both carbonate and siliciclastic sediment-derived (Fisher et al, 2007).

This comes in the wake of at least three wells being abandoned as unsuitable for irrigation farming in the area, by the KRCS, way back in 2013, and also from the health challenges witnessed on delayed walking age, weak bones and miscarriages of babies, in the Dertu centre, between 2002 to date.

Due to the anomalous sodium levels occasioned thus, the KRCS had to incur extra expense drilling two extra wells near the EwasoNg'iro seasonal stream course, at extra expense, and the water quality was acceptable for the task of irrigation farming.

The present study is therefore meant to develop a prediction model that will in future guard against a repeat of the same mistakes occasion earlier**. Three wells in the area were sampled and analyzed for chemical anions and cations.**

III. GEOLOGY AND HYDROGEOLOGY OF THE STUDY AREA

A. *General Geology*

The geological history can be traced from the Pre-Cambrian era that is represented by metamorphic rocks of the Basement system. After a long period of uplift and erosion, a marine incursion took place in the early Jurassic which led to deposition of sandstones, shale, limestone and coarse to fine-grained sediments with calcite. This was followed by a further period of uplift and erosion, culminating in the maturation of Sub-Miocene peneplain, but a subsequent cycle of erosion removed traces of that level in some areas. During those periods of erosion, the Jurassic rocks were largely removed and a later cycle resulted in the deposition of marine and fluvial sediments on the Sub-Miocene surface. The Mid-Pleistocene to Holocene has been marked with rapid climatic fluctuations in the region. This is attributed to periods of intense weathering followed by erosion and deposition of the weathered material. Most of the colluvium and calcrete in the area is believed to have been deposited during that time.

Regional structural setting can be observed from the landsat interpretation map covering the area, the lineations are major features in a rather flat and featureless terrain. These features are more apparent and identified in colour composite (landsat MSS) images than on black and white. The most significant lineaments are LaghBoghol, LaghChoichuff, Habaswein-Wajir Bor and Hagardera-Liboi lineaments. There exists a lineament running across the North of Sericho township, and further trends in the direction azimuth defined as South West - North East direction. All these lineaments bear some hydrogeological significance, on the patterns of groundwater dynamics in the study area. The LaghaChoichuff is another major lineament in the study area, intercepting another major liniment cutting through Mt. Marsabit and trending in the North Eastern direction.

The LaghBoghol lineament is an important feature in the area since it marks the North East boundary of a confirmed gravity and magnetic (geophysical) sedimentary Basin. It also coincides in part with a subsurface fault, independently interpreted from the geophysical surveys. The Western border of this basin has been indicated by geophysical data as not being a major fault zone but a combination of faulting and tilting. In the middle part of landsat interpretation map by Miller (1975), there are large accurate features, which are interpreted as possible depositional patterns (bed forms) of clastic beds. These features play a significant role in hydrogeology.

B. *Precambrian Basement Rocks*

The Basement system forms the oldest rocks in the area and forms the floor upon which other rock formations rest. They comprise of sediments which were transformed by regional metamorphism into gneisses, schists, quartzites. No out crops are observed in the region since they are found at great depths.

i) Tertiary sediments: These consist of sandstones grits and conglomerates and are best exposed on the flanks of the Merti Plateau and Barchuma where they have been preserved from erosion by the capping of olivine Basalts. The other structural parameters lie on the South-East of the Yamicha Plateau, and also within the area to the north of the Eldere-Modogashe highway. This exposure is poor on the featureless plains that extend from Merti all the way to Habaswein and Dadaab sub-counties, as they easily erode on the account of their friable mineralogy, which leaves them covered by thick soil deposits. The grits are however exposed along rivers, notably the GalanaGof. The grey sandy soil is an indication of the grits presence as the soil is identical to that formed from them.

ii) Quaternary sediments: These are composed of lacustrine sediments with limestone, calcrete and superficial deposits belonging to both Pleistocene and Holocene periods. The texture of the lacustrine sediments is coarse to fine grained. The superficial deposits comprise alluvial sediments, containing sand and silt, along the ephemeral river flow courses, and colluvium comprised mainly of crudely stratified mixtures of clay, silt and rock fragments along the slopes of some large inselbergs or residual hills and hillocks.

iii) Geology of the Study Site: The geology of the Dertu area is predominantly Loams, fine-to- medium grained sands, holo-crystalline glass, red clays, alongside marls at shallow depths, but with some Jurassic clay alongside medium-grained sandstones at great depths. The study area and its immediate environs are underlain by sedimentary formations belonging to Plio-Pleistocene age. These formations increase in thickness towards east and south. Most of the sediments in the area belong to Tertiary age, being underlain at greater depths by Jurassic and Triassic formations. The oldest sedimentary succession exposed in the area is probably Upper Jurassic Limestone underlying the sandstones and limestone which outcrop locally.

The Triassic sediments include variable marls and sandy marls, friable sandstones, sometimes pebbly and ranging through grey, pale red or green, or mottled, in colour. Locally they include shale and thick brown clays. The area is entirely covered by young sediments, especially wind-blown sands and depositional alluvial sediments. The local geology has been delineated based on the few visible outcrops and geological logs of boreholes that have been drilled in the area.

From field observations, the study area is underlain by a thick layer of sandy soils. These are directly underlain by a thin layer of sandy clays, limestone and assorted sands and sandstones. The thickness of each of these layers varies significantly from site to site. Further, each of these sediments are intercalate with each other up to depths in excess of 3000 metresbgl.

iv) Structural geology and allied parameters: Geo-structural parameters such as faults in the rocks often optimize storage, transmissivity and recharge, particularly when they occur adjacent to, or within, surface drainage systems. Faulting will have the highest impact on hard and massive rock types. Elastic formations such as tuffs and weakly consolidated deposits will bend (fold) rather than break (fault). As a result, their porosity will not increase in the area affected by the fault. Fractures and joints will break massive layers such as gneisses. This gives rise to increased (secondary) porosity, and consequently enhanced recharge, groundwater storage and transmissivity.

Only lineaments are well illustrated through the series of tree-lines along the river courses and allied neighborhoods and this is a sure proof that the areas bear some fractures that are transmitting juvenile and meteoric waters in the subsurface.

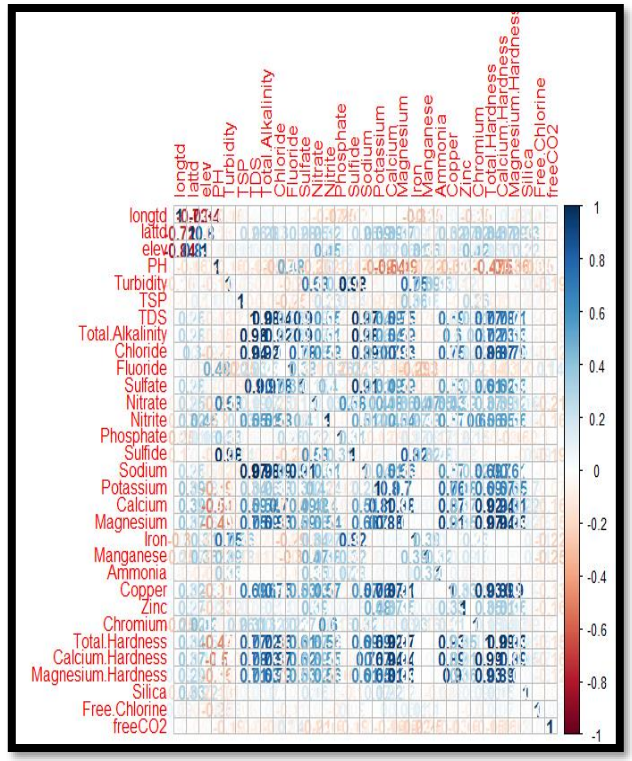
IV. SODIUM CONCENTRATIONS IN THE CHARACTERISTICS IN THE STUDY AREA

The results of the field assignments in the area coupled with secondary data that were obtained from partners give insight on the groundwater chemistry in the area.

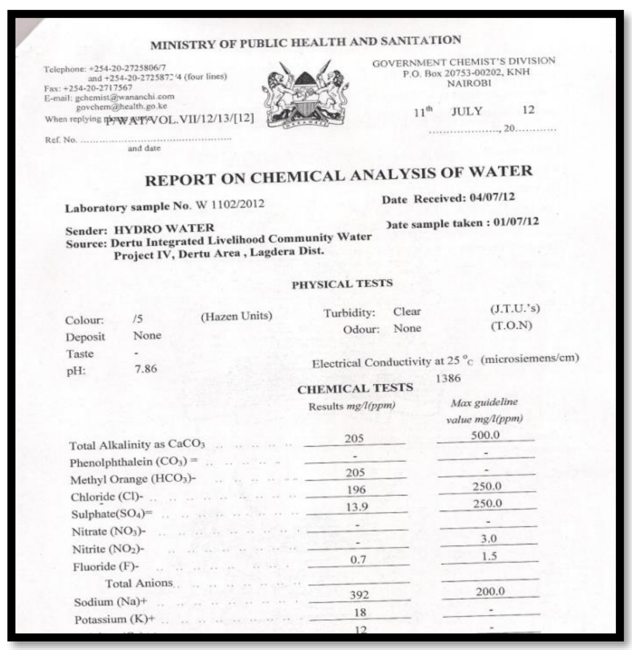
Various statistical analyses and plots have been carried out using chemical data to deduce a hydrochemical evaluation of the aquifer system based on the ionic constituents, geology, proximity to the central Merti aquifers and factors controlling sodium levels in the groundwater (Kumar,et al,2006).

The distribution of sodium ions in the study area was mapped/plotted using the statistical software, R, and the correlation plot revealed the main agents dictating the sodic concentrations in the area as mainly the sulphates and carbonates (Total Hardness, mg/L), for fieldwork, the water samples were collected after borehole completion report had been completed either from privately-owned or community wells. The sampled water from source was allowed to stabilize before measurements were taken. Sodium, Potassium, Fluoride, EC and TDS content in the groundwater samples were determined directly after dilution. All the experiments were carried out by qualified laboratory technicians. Presence of excessive Na ions in irrigation water promotes soil dispersion and structural break down when Na:Ca ratio exceeds 3. Infiltration problem will occur from such high ratio primarily due to lack of sufficient Ca2 ions to counter the dispersing effect of Na. Excessive Na ions also create problems in crop water uptake, poor seedling emergence, lack of aeration, plant and root **diseases (Mirza et al,2014).**

The results so achieved were then added to the existing historical data frames obtained from the WRMA and Ministry Headquarter offices. The model hereunder shows the analysis in R:



Model 1 (a) the model generated by the R software illustrating the correlation between sodium and other compounds in mg/L



MINISTRY OF PUBLIC HEALTH AND SANITATION

Telephone: +254-20-2725806/7 and +254-20-27258724 (four lines)
Fax: +254-20-2717567
E-mail: gchemist@wananchi.com govchem@health.go.ke
When replying please quote P/WAT/VOL.VII/12/13/[12]
Ref. No. .... and date

GOVERNMENT CHEMIST'S DIVISION
P.O. Box 20753-00202, KNH
NAIROBI

11th JULY 12, 20..........

REPORT ON CHEMICAL ANALYSIS OF WATER

Laboratory sample No. W 1102/2012 **Date Received: 04/07/12**

Sender: HYDRO WATER **Date sample taken : 01/07/12**

Source: Dertu Integrated Livelihood Community Water Project IV, Dertu Area , Lagdera Dist.

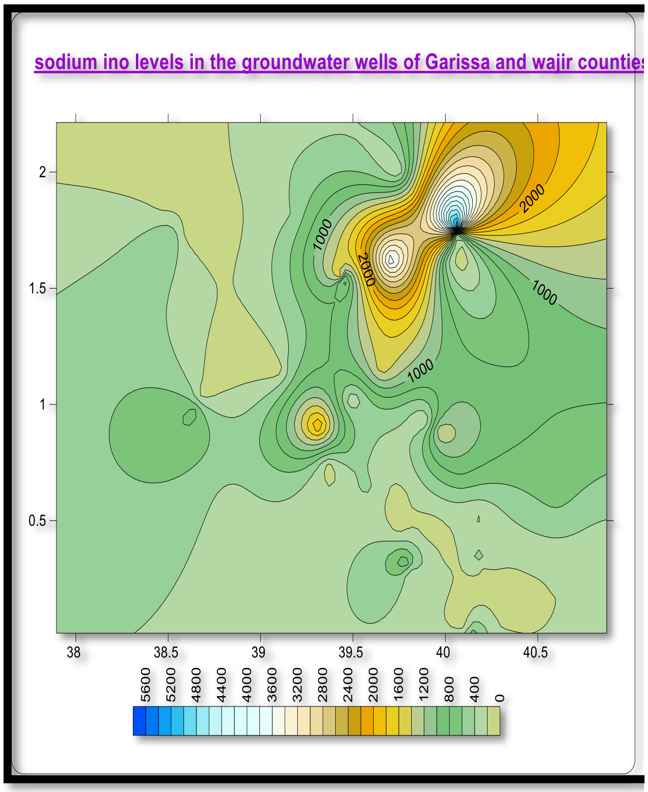
PHYSICAL TESTS

Colour: /5 (Hazen Units) Turbidity: Clear (J.T.U.'s)
Deposit None Odour: None (T.O.N)
Taste -
pH: 7.86 Electrical Conductivity at 25 °c (microsiemens/cm) 1386

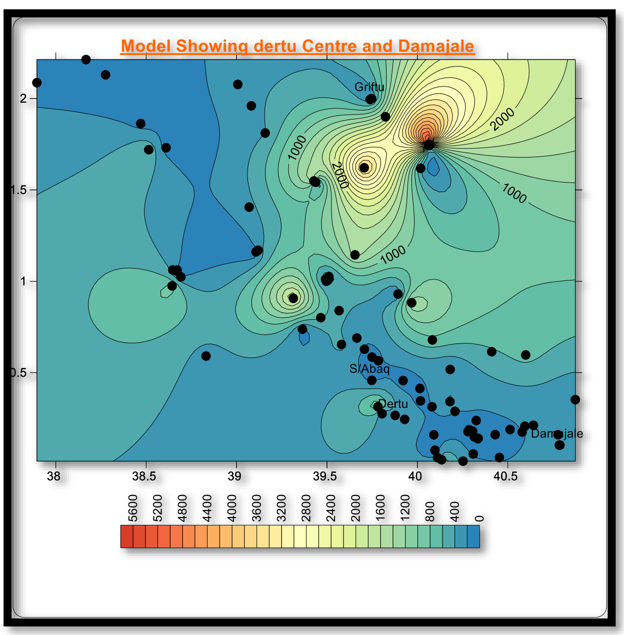
CHEMICAL TESTS

	Results *mg/l(ppm)*	*Max guideline value mg/l(ppm)*
Total Alkalinity as $CaCO_3$	205	500.0
Phenolphthalein (CO_3) =	-	-
Methyl Orange (HCO_3)-	205	-
Chloride (Cl)-	196	250.0
Sulphate(SO_4)=	13.9	250.0
Nitrate (NO_3)-	-	-
Nitrite (NO_2)-	-	3.0
Fluoride (F)-	0.7	1.5
Total Anions		
Sodium (Na)+	392	200.0
Potassium (K)+	18	-
[illegible]	12	-

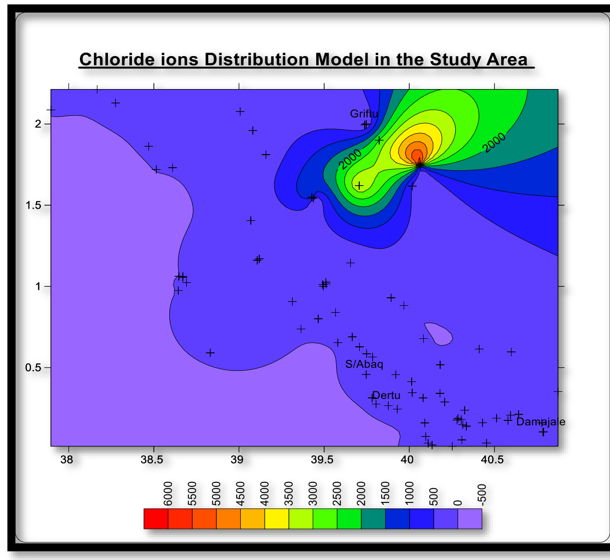
Table 1.Showing the Sodium Concentrations in the Study Area alongside other elements .the levels are anomalous, way above the WHO Guidelines



Model 1(b) : Model showing the sodium Concentrations in the Garissa and Wajir Counties harboring the Merti aquifer

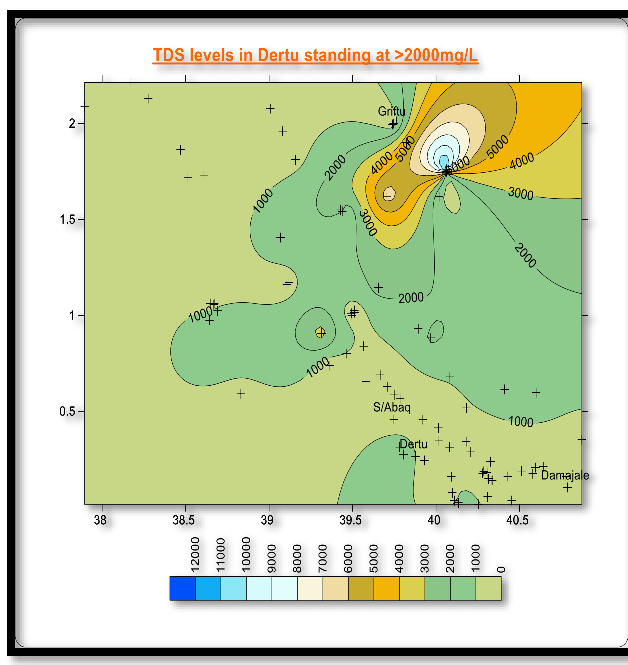


Model 2: Model indicating Sodium distriburion in the Dertu, Damajale and Griftu Settlements indicating Na+ mineralization of between 200 and 400 mg/L



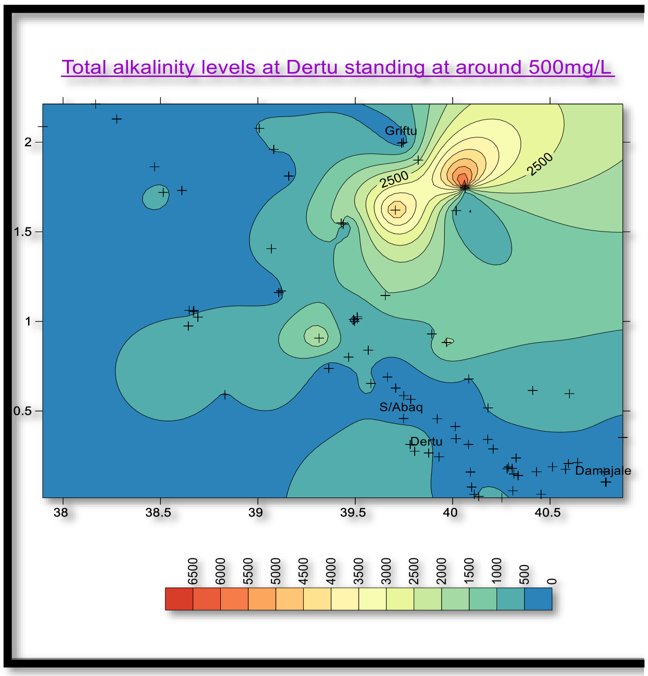
Model 3: Model Showing the Chloride Concentrations in the Study Area. Highest levels are in he central northern tip of the Merti aquifer

Chlorides are widely distributed in nature as salts of sodium (NaCl), potassium (KCl), and calcium ($CaCl_2$).The statistical model had shown strong correlation between Na+ ions ,and the Cl- ions with a coefficient value of almost 0.90. Dertu centre has well over 300 mg/L of the same.The chloride concentrations and distributrion model is as shown hereunder:



Model 4: Model Showing the TDS distributions in the Merti aquifer. Western portion of study area shows favorable levels of TDS

Since TDS also correlates well with sodic contents in the aquifers, the model for TDS was also developed thus.



Model 5: Model Showing the Total Alkalinity distributions in the Merti aqyifer

Total Alkalinity also correlates well with sodic contents in the aquifers, and the geospatial model for Total Alkalinity was thus developed to aid the study.

TABLE II
SALINITY STATUS CLASSIFICATIONS, BY TOTAL SALT CONCENTRATION

Salinity status	Salinity (mg/l)	Description and use
Fresh	< 500	Drinking and all irrigation
Marginal	500 –1 000	Most irrigation, adverse effects on ecosystems become apparent
Brackish	1 000 – 2 000	Irrigation certain crops only; useful for most stock
Saline	2 000 – 10 000	Useful for most livestock
Highly saline	10 000–35 000	Very saline groundwater, limited use for certain livestock
Brine	>35 000	Seawater; some mining and industrial uses exist

Classifications from Mayer, XM, Ruprecht, JK & Bari, MA 2005, Stream salinity status and trends in south-west Western Australia, Department of Environment, Salinity and land use impacts series, Report No. SLUI 38

V. THEORETICAL FRAMEWORK OF THE NEURAL NETWORKS MODELS: ERROR BACKPROPAGATION

A. *Background*

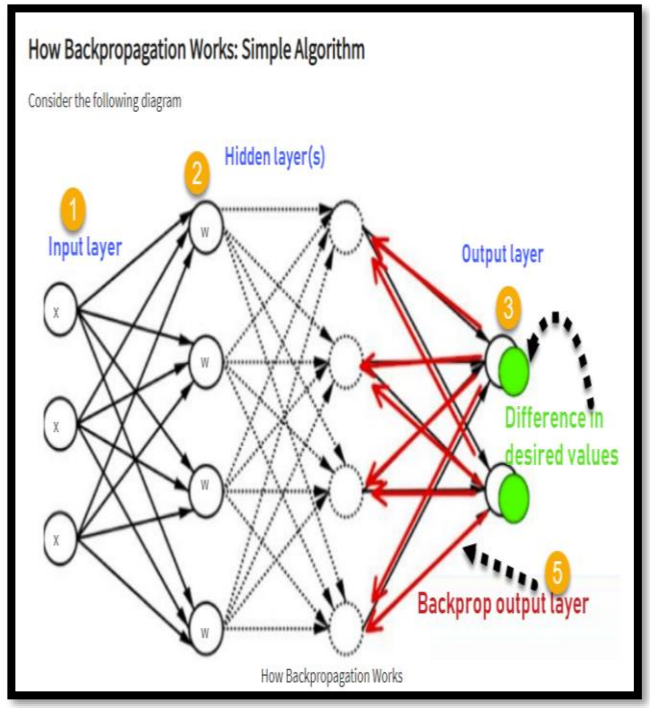
Neural Networks (NN) are important data mining tool used for classification and clustering. The technique may also be used for regression analysis and predictions of dependent variables from tables arranged in the dataframe formats. It is an attempt to build machine that will mimic brain activities and be able to learn (Cilimcovic et al, 2015). NN usually learns by examples. If NN is supplied with enough examples, it should be able to perform classification and even discover new trends or patterns in data. Basic NN is composed of three layers, input, output and hidden layer. Each layer can have number of nodes and nodes from input layer are connected to the nodes from hidden layer. Nodes from hidden layer are connected to the nodes from output layer. Those connections represent weights between nodes (Cilimcovic et al 2015).

It has been noted that the neural network is a group of connected input-output units, where each connection has a weight associated with its computer programs. It helps build predictive models from a statistical dataset. This model builds upon and mimics the human nervous system. It helps conduct image understanding, human learning, computer speech, regression analysis, et cetera.
The input in this case could be variables like the hand-held GPS generated longitudes, latitudes, and elevations, on one hand, and the field geoelectrical surveys generated depths to aquifers elevations.

B. *Back-propagation*

Back-propagation is the essence of neural net training. It is the method of fine-tuning the weights of a neural net based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and to make the model reliable by increasing its generalization. Back-propagation is a short form for "backward propagation of errors." It is a standard method of training artificial neural networks. This method helps in calculating the gradient of a loss function with respects to all the weights in the network.

The model hereunder shows how the method works.



Model 6: Model illustrating how neural network backpropagation works.

Inputs X, arrive through the reconnected path. In the study, Xis an array of variables.

X=longitude+latitudes+depth+Elevations+chloride+TDS+ alkanlinity
Y=the variable being predicted, in this case, Sodium level class of the aquifer.

The regression equation thus reads thus:

Y~X

Explanation: all the inputs are the list of variables used as predictors. These may be longitudes, Latitudes, elevations, alkalinity, chloride levels,depths to aquifer and sodium levels existing in all other wells in the area. Y may thus be the predicted variable. In the present study, the y variable is sodium content.

Input is modeled using real weights W. The weights are usually randomly selected.

Explanation: The weights are randomly selected using random number generator, or softwares like Matlab, python or R.
the output is calculated for every neuron from the input layer, to the hidden layers, to the output layer.

Explanation: all the numerical variables are usually normalized into variable ranging between 0 and 1.Again, this may be done using softwares like Matlab, python or R. The output generated is then subtracted from the expected output. The error is then noted.

C. *Calculate the error in the outputs*

Explanation: the next step is to calculate the error between the generated output and the expected output. The value generated is a normalized one.

Normalized output= (VALUE-MINIMUM VALUE)/MAXIMUM VALUE

All that has been done so far is iteration 1. At the fiftieth or hundredth (or so) iteration, the values may converge.
The process is repeated again, over and over, until the difference between the generated output vs the expected value is more or less equal to zero. The iterations are then stopped. The normalized values are then used to again determine the original, non-normalized value,X, of this fraction. This is the predicted variable.

VI. SOFTWARE IMPLEMENTATION IN PYTHON MULTIPLE LINEAR PERCEPTRON (MLP) CODING

The first step was to develop the data that was tom be modeled into an excel dataframe. The original dataframe in numerical values of sodium as continuous variables was converted into a dataframe of sodium levels (alongside the geospatial parameters) written out as factor variables defining the range of sodium concentration levels.

	A	B	C	D	E	F	G	H
1	longtd	lattd	elev	Chloride	TDS	alkalinity	Sodium	
2	38.647	1.062	293	8	477	196	341	
3	39.6547	1.14458	256	27	2370	1470	1496	
4	40.18014	0.34175	172	7	723	390	479	
5	38.69197	1.02384	291	465	1396	773	201.5	
6	39.31385	0.90669	214	18	3450	1965	2061	
7	40.20805	0.28844	135	7	480	321	171.8	
8	39.00675	2.07735	344	44	646	519	343	
9	40.32534	0.23798	159	132	572	286	140.4	
10	39.08145	1.96014	321	177	640	517	346	
11	39.7467	1.99968	312	3	470	293	214	
12	40.5794	0.175	124	20	859	374	185.6	
13	39.74757	0.458	164	33	506	379	137.8	
14	40.51227	0.18899	120	193	673	431	157.4	
15	40.018	1.618	260	444	1322	708	373	
16	39.15833	1.81137	300	157	700	381	393	
17	40.08295	0.67915	160	6	1094	480	754	

Fig 4: original data frame with sodic values depicted numerically

	A	B	C	D	E	F	G
1	longtd	lattd	elev	Chloride	TDS	alkalinity	sodium
2	38.647	1.062	293	8	477	196	201 to 400mg/L
3	39.6547	1.14458	256	27	2370	1470	over 600mg/L
4	40.18014	0.34175	172	7	723	390	between 401 and 600mg/L
5	38.69197	1.02384	291	465	1396	773	201 to 400mg/L
6	39.31385	0.90669	214	18	3450	1965	over 600mg/L
7	40.20805	0.28844	135	7	480	321	0 to 200mg/L
8	39.00675	2.07735	344	44	646	519	201 to 400mg/L
9	40.32534	0.23798	159	132	572	286	0 to 200mg/L
10	39.08145	1.96014	321	177	640	517	201 to 400mg/L
11	39.7467	1.99968	312	3	470	293	201 to 400mg/L
12	40.5794	0.175	124	20	859	374	0 to 200mg/L
13	39.74757	0.458	164	33	506	379	0 to 200mg/L
14	40.51227	0.18899	120	193	673	431	0 to 200mg/L
15	40.018	1.618	260	444	1322	708	201 to 400mg/L
16	39.15833	1.81137	300	157	700	381	201 to 400mg/L
17	40.08295	0.67915	160	6	1094	480	over 600mg/L
18	40.01372	0.41371	148	9	345	260	0 to 200mg/L

Fig 5: Modified Data frame with Sodic Values Depicted as Factors

The first row indicates Na^+ as 341mg/L in the borehole in the figure defined as Fig 1 above. The same row is now defined out as “between 201 and 400mg/L”. Predicting with fuzzy range as set out in table two is much more convenient than doing the same with numerical values. The dataset comprised eighty wells in Garissa and Wajir portions of the Merti aquifer that neighbor the Garissa County.

The data may also be in factor variable forms, namely, for the sodium contents such that class “1” (0-250) is acceptable class, "2" (251-700)mg/L ,is “high sodic content” and class 3 is “very high”(above 700mg/L) . See table hereunder

	longtd	lattd	elev	Chloride	TDS	alkalinity	sodium
2	38.647	1.062	293	8	477	196	2
3	39.6547	1.14458	256	27	2370	1470	3
4	40.18014	0.34175	172	7	723	390	2
5	38.69197	1.02384	291	465	1396	773	1
6	39.31385	0.90669	214	18	3450	1965	3
7	40.20805	0.28844	135	7	480	321	1
8	39.00675	2.07735	344	44	646	519	2
9	40.32534	0.23798	159	132	572	286	1
10	39.08145	1.96014	321	177	640	517	2

Fig 6:In the present study, the regression analysis ability of the neural networks is employed in making predictions as to whether a virgin site suits drilling of a well, or otherwise, based on the use of Back propagation algorithms.

The neural networks will be looking this:

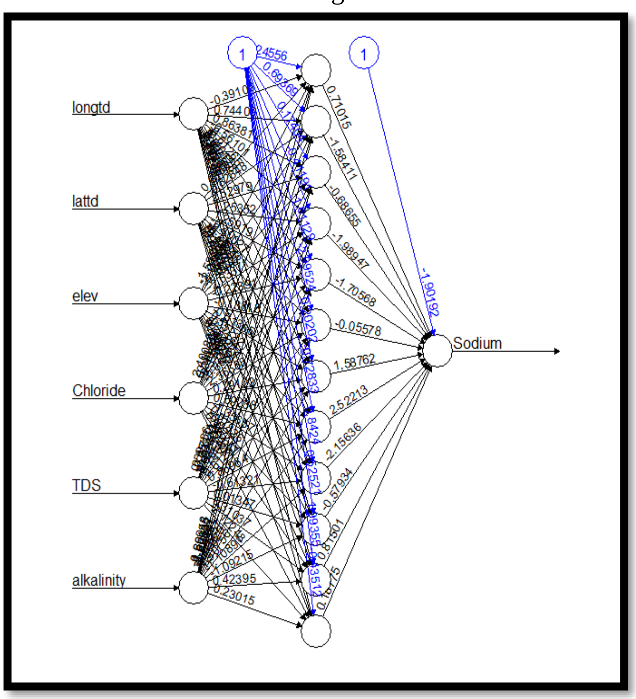


Fig 7: the neural network model indicating both inputs and the dependent variable being predicted, sodium

Example of Application: The research study Team was tasked with finding the estimated sodium levels /sodic level class, of the aquifer waters expected from new, proposed wells to be drilled in the Northern Kenya, using the tabulated parameters stated thus, respectively located in the vicinity of boreholes in the three final rows. This will help the donor determine the water suits his purpose or otherwise. The GPS meter values were used in the historical data bank (fig 7 hereunder). During the research, only GPS values of the new spots to be drilled were taken (fig 8).

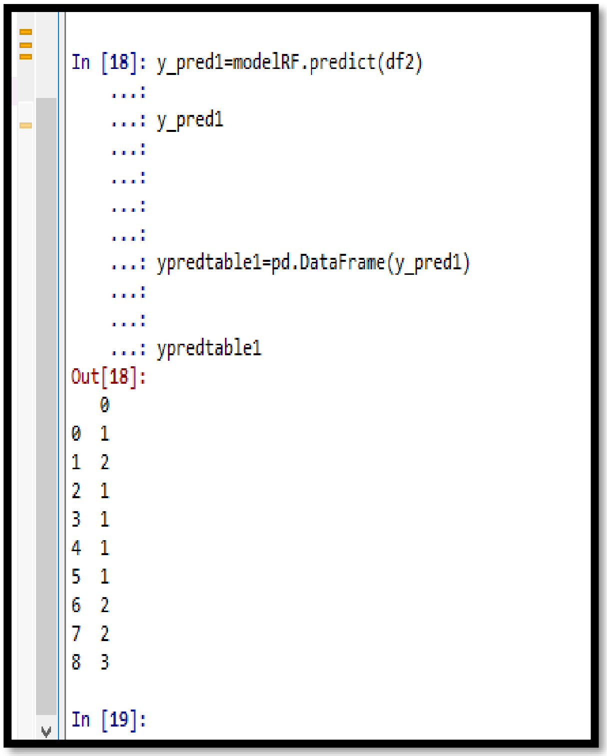
	longtd	lattd	elev	sodium
2	38.647	1.062	293	2
3	39.6547	1.14458	256	3
4	40.18014	0.34175	172	2
5	38.69197	1.02384	291	1
6	39.31385	0.90669	214	3
7	40.20805	0.28844	135	1
8	39.00675	2.07735	344	2
9	40.32534	0.23798	159	1
10	39.08145	1.96014	321	2
11	39.7467	1.99968	312	1
12	40.5794	0.175	124	1
13	39.74757	0.458	164	1
14	40.51227	0.18899	120	1
15	40.018	1.618	260	2
16	39.15833	1.81137	300	2
17	40.08295	0.67915	160	3
18	40.01372	0.41371	148	1
19	40.09103	0.15972	156	1
20	37.89534	2.0868	769	1
21	40.01724	0.34578	134	1
22	39.70587	1.6211	257	3
23	39.56672	0.83955	194	2
24	40.64177	0.21205	108	1

Figure 8:-screenshot the parent dataset.

	longtd	lattd	elev
2	40.32634	0.23898	161
3	39.07145	1.97014	322
4	39.7567	1.98968	311
5	40.5894	0.1775	127
6	39.7757	0.45598	166
7	40.54123	0.18999	122
8	40.0178	1.628	262
9	39.15783	1.82137	301
10	40.0813	0.67815	160

Figure 9: to be replaced possess classes array-1, 2, 1, 1, 1, 1, 2, 2, and 3. the new sites to be predicted for sodium levels using geospatial parameters, whose nearby boreholes

Indeed, the vicinity of the areas where the proposed wells are to be done bear the following sodium contents, in the excel data frame:


```
In [18]: y_pred1=modelRF.predict(df2)
    ...:
    ...: y_pred1
    ...:
    ...:
    ...:
    ...:
    ...: ypredtable1=pd.DataFrame(y_pred1)
    ...:
    ...:
    ...: ypredtable1
Out[18]:
   0
0  1
1  2
2  1
3  1
4  1
5  1
6  2
7  2
8  3

In [19]:
```

Fig 10: output in the Python Console. The predicted output is nearly copycat of the tabulated class dataframe showing sodic levels

The package, RandomForest Classifier in python predicts the same values as the MLP algorithm

VII. CONCLUSIONS

The sodium levels in the proposed study area have been predicted using a combination of both secondary and primary data, gleaned from the hydrogeological databanks, and this is of two advantages to a potential developer: there shall be an obvious reduction in the number of sodium contaminated wells in the project area, and water quality appropriate for activities like farming will be known well before any financial loss is occasioned. The two open-source softwares , R and python, have thus aided the analysis and predictions to aid sodium level analysis in existing groundwaters and the Na+ levels in virgin terrains in the study area.

VIII. RECOMMENDATIONS

i) Both The algorithms, neural networks and random forest classifiers, have been able to help predict all the three classes of sodium ion concentrations, correctly in the study area. The areas with class 1 of excellent /acceptable levels of sodium have been correctly identified. The others of moderate and extreme sodic levels have also been identified. These areas should be developed for groundwater abstractions.

ii) Donors have in the past sank boreholes with levels of sodium considered unacceptable for agriculture. A donor like KRCS will not in the future lose money sinking wells for irrigation, only to end up having anomalous levels of the Na+, thereby incurring unnecessary waste.

iii) The relationship between Na+ contents and the medical disorders should be studied further in the area, and the study should incorporate the fluoride levels as well. The local Public health office, together with civic leadership should launch a detailed study and testing to establish a correlation between the health challenges and the groundwater sodic contents. It could also be that one of the anion/cation that bear a strong correlation with Na+ ions is the offending hydrochemical variable responsible for the feared health hazard, or it may be that the source of health hazard is a heavy metal, way beyond the scope of our present study. Detailed study relating to Na+ should thus be undertaken.

iv) Boreholes likely to be of acceptable sodic levels are to be found even in the Dertu area, not necessarily in the EwasoNgiro river-bed area. Boroansis and its Environs nearing Dertu would be thus appropriate. During exploration mapping using geoelectrical arrays, all one would need to do would be to feed in the elevations, longitudes and latitudes,depths and aquifer resistivity of the proposed sites/drilling spots, in the Modeling software, and predict the Na+ contents, before any drilling is contemplated. The cost-free, open source softwares like R or Python should thus be employed for key decision making considerations such as the whether or not to sink wells in sensitive areas like Dertu.

REFERENCES

[1] Cilimkovic, M. (2015). Neural networks and back propagation algorithm. *Institute of Technology Blanchardstown, Blanchardstown Road North Dublin*, *15*.

[2] Fisher, R. S., &Mullican III, W. F. (1997). Hydrochemical evolution of sodium-sulfate and sodium-chloride groundwater beneath the northern Chihuahuan Desert, Trans-Pecos, Texas, USA. *Hydrogeology journal*, *5*(2), 4-16.

[3] Gholami, V., Khaleghi, M. R., &Salimi, E. T. (2020). Groundwater quality modeling using self-organizing map (SOM) and geographic

information system (GIS) on the Caspian southern coasts. *Journal of Mountain Science*, *17*(7), 1724-1734.

[4] Haselbeck, V., Kordilla, J., Krause, F., &Sauter, M. (2019). Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. *Journal of Hydrology*, *576*, 610-619.

[5] Kaya, Y. Z., Üneş, F., Demirci, M., Taşar, B., &Varçin, H. (2018). Groundwater level prediction using artificial neural network and M5 tree models. *AerulsiApa. Componente ale Mediului*, 195-201.

[6] Kumar, M., Ramanathan, A. L., Rao, M. S., & Kumar, B. (2006). Identification and evaluation of hydrogeochemical processes in the groundwater environment of Delhi, India. *Environmental Geology*, *50*(7), 1025-1039.

[7] Mirza, A., M. TanvirRahman, M. Saadat, MdSafiqul Islam, Md Abdullah Al-mansur, and Shamim Ahmed. "Groundwater characterization and selection of suitable water type for irrigation in the western region of Bangladesh." *Applied Water Science* 7, no. 1 (2017): 233.

[8] Naser, A. M., Higgins, E. M., Arman, S., Ercumen, A., Ashraf, S., Das, K. K., &Unicomb, L. (2018). Effect of groundwater iron on residual chlorine in water treated with sodium dichloroisocyanurate tablets in rural Bangladesh. *The American journal of tropical medicine and hygiene*, *98*(4), 977-983.

[9] Olago, D. O. (2019). Constraints and solutions for groundwater development, supply and governance in urban areas in Kenya. *Hydrogeology Journal*, *27*(3), 1031-1050. Pittalis, D. (2010)..