Cluster-Based Geoelectrical Zonation of Aquifers in Some Parts of Lower Orashi Area Southeastern Nigeria Using Ves Data.

[2]Onwuegbuchulam C.O [3]Ezebunanwa A.C, [4]Mbagwu E.C [1]Chinaka A.I

[1]*Mathematics Department, Alvan Ikoku Federal University of Education Owerri, Imo State, Nigeria.*

[2]*Teesside University Middlesbrough School of computing, engineering, and digital technologies.*

[3]*Department of Management, Business School Hertfordshire University Hatfield, United Kingdom*

[4]*Department of Earth, Energy and Environment, University of Calgary Alberta Canada*

Abstract - **Groundwater constitutes the prime source of potable water for Imo and Anambra States, Nigeria. However, the stratification of the Anambra Basin remains complex and heterogeneous. The above-mentioned elements lead to the difficulties associated with the exploitation of groundwater. The standard point-based interpretation of the Vertical Electrical Sounding (VES) data is known to be inaccurate for assessing large-scale regional hydro stratigraphic patterns and is associated with the infamous borehole yield discrepancies. An integrated, data-driven approach consisting of geophysical inversion, Dar–Zarrouk (DZ) parameterization, and unsupervised ML is further consolidated to delineate the aquifer systems for the entire basin. The data of 20 VES stations were collected using the Schlumberger array to a maximum electrode spacing (AB/2) of 300 m, and the data were inverted to produce the layered resistivity models. The models were used to compute transverse resistance (T) and longitudinal conductance (S) as proxies for aquifer facility and sealed protection. Multivariate features of the layers containing resistivities, thicknesses, depth to basement, and DZ parameters were standardized and subjected to K-means clustering analysis. Based on three importable electrical geo-domain, the algorithm sectioned the research area into three separate but uniformly defined. Cluster 0 was identified by its multi-layered formation sandwiched sequence consisting of moderate resistivity (50->500 [Ωm]), T>5000 [Ωm²], and S of 0.51000 [Ωm], number of layers <4, and low S of <0.3[\siemens], is directly related to the lateritic caps and shallow basement highs, and these have little to no prospects for groundwater. Cluster 2 low resistivity (<30 [Ωm]), S v and S have and S, thus clay confinement of the Imo shales with suspected highly saline geochemistry. Primary Cluster 0 correlates with the aquifer position within the central plain of Ideato, and secondary spatial analysis of the study area entails: Cluster 2, was found to depressions of the landscape, and lastly, the isolated upland regions was Cluster 1. Umuma Isiaku, VES18, Amanator, VES15, and Akokwa, VES2, were identified borehole sites auxiliary with the maximum value of the normalized T and S components for GW PI (Ground Water Potential Index). In the context of the study, the combination of VES, DZ parameters, and clustering mitigates the ambiguity associated with and local, 1-D equivalent, developing a viable decision-making model for the groundwater resource assessment within the complex sedimentary basin of the study area.**

Keywords: **Aquifer zonation, Dar–Zarrouk parameters, Geoelectrical stratigraphy, Groundwater potential, K-means clustering, Vertical Electrical Sounding, Anambra Basin, Hydrogeophysics.**

I. INTRODUCTION

During the dry season, surface water becomes insufficient and polluted in southeastern Nigeria, particularly Imo and Anambra States. Consequently, groundwater must be utilized for domestic, agricultural and industrial purposes, making it a vital resource in this region [1]. Hydrogeological architecture in this area is dominated by the Anambra Basin, a Cretaceous to Tertiary sedimentary basin and has a complex interdigitating sequence of arenaceous and argillaceous units. [2]. The Oligocene –Miocene Benin and Eocene Ameki formations, aquifers of this basin, are dominated by coarse-grained, unconsolidated to semi consolidated sands. Above these formations is the Palaeocene Imo Shale, a regional aquitard and roughly constitutes the vertical thickness of the aquifer system. The Shale is massively conductive and clayrich, consequently creating thick aquitards of significant depth [3]. The heterogeneous lithological description produces aquifers with different thicknesses and productivity, and significant variation in the protective cover to the aquifers, which produces dry or contaminated boreholes and socio-economic problems for the local community [4].

Thus, there is an urgent need for simple and affordable techniques that do not rely on single-point assessments but rather derive a conceptual understanding of an area's aquifers and their distribution, availability and vulnerability. The Schlumberger array Vertical Electrical Sounding (VES) technique is, due to its ease of implementation and responsiveness to litho-fluid difference among layers, a mainstay of groundwater exploration in sub-Saharan Africa [5]. However, the standard approach to VES interpret and the drawings of their corresponding curves, has substantial problems due to the principle of equivalence (where a combinations of layers that are present in the ground can produce the identical curves) and the suppression phenomena (thin layers of the ground are not picked) that leads to significant subjectivity in the interpretation of data [6].

In response to these constraints, investigators have begun employing the Dar–Zarrouk (DZ) parameters, longitudinal conductance (S) and transverse resistance (T), which, combined, collapse the resistivity and thickness of a layered sequence into a single, coherent quantity with relevant physical meaning. S (Σ h/ρ) is the total electrical conductance, which is correlated to the protective capacity of the aquifer against contaminants and is influenced greater by the conductive clay layer than by the other layers. T (Σ h∗ρ), on the other hand, is the total transverse resistance, which is dominated by the more resistive layer of the aquifer and is, in theory, proportional to the hydraulic transmissivity in granular porous media. While the DZ parameters do compress multi-dimensional data, their interpretation at different locations is still a multi-dimensional problem.

Fortunately, the explosion of computing power and clustering algorithms helps to refine this raw approach. K-means clustering is one of the unsupervised machine learning methods that can, without supervision, cluster a set of VES stations into a certain number of classes based on proximity in a multi-dimensional space defined by the integrated features (resistivity statistics, depth, and DZ parameters). This data-driven approach has optimized the identification of hydro stratigraphic domains by converting discrete soundings into continuous zonation. In the Anambra Basin, this approach has enormous potential and has not been fully researched.

Henceforth, this research sought to develop and assess a first-of-its-kind cluster-based geo-electrical zonation approach for Anambra Basin. The specific objectives were to (1) obtain and invert VES geo-electrical data from twenty strategically selected stations; (2) compute and analyze each site's Dar-Zarrouk parameters; (3) conduct K-means clustering to partition the geoelectrical features into distinct geoelectrical domains; (4) carry out a hydrogeological interpretation of these domains and spatial and statistical validation; and (5) equalize the findings into a borehole prioritization and a groundwater potential map. The adopted methodology offers reproducible, objective, and scalable models for ethically grounded groundwater resource management in challenging geological settings, anywhere in the world.

1.1 STUDY AREA

The study was carried out in some selected communities in Ideato North, Ideato South Local Government Areas of Imo State and the adjacent area of Anambra State, Nigeria. The geographical coordinates of the study area are latitude 5 45, 6 00 N and Longitude 7 00, 7 15 E. The area of study is located in the western part of Anambra Basin. The area is characterized by an undulating topography within the elevation range of 50-150m above sea level, with ridges capped with lateritic materials and valleys containing thicker deposits of sediment.

Geologically, the surface is the Benin Formation (Coastal Plain Sands), characterized by friable, coarse-to-medium-grained sands with some clay lenses and bands of ironstone [2]. The shales and sandstones of the Ameki Formation underlie and outcrop in the rivers valleys. The Paleocene Imo Shale constitutes a structural base. The climate, with a mean annual rainfall above 2000 mm, is tropical humid which enhances groundwater recharge. The land use is predominantly agricultural with the increasing urbanization of the area. This, however, increases the need for groundwater and the risks of groundwater contamination by pit latrines, agrochemicals, and other pollutants.

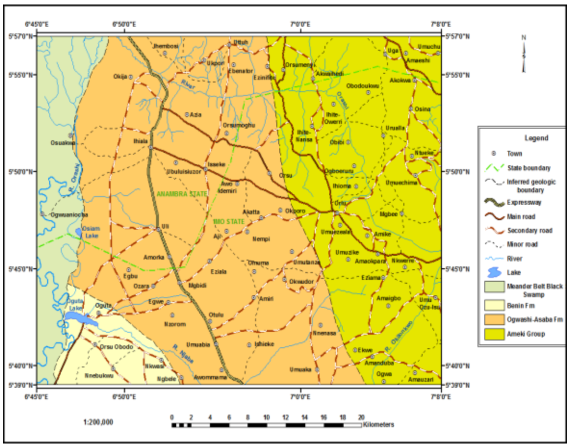

Fig 1 geology map of the area.

II. LITERATURE REVIEW

2.1 Geophysical Exploration for Groundwater in Sedimentary Basins

The Geoelectrical methods, particularly the resistivity sounding method, involve measuring the subsurface response to an injected electrical current which differs according to the subsurface materials lithology, porosity, pore-water salinity, and clay content [5]. In the sedimentary sequences of southern Nigeria, the resistivity contrast between freshwater-saturated sands (> 100 Ωm) and clay/shale (< 30 Ωm) is pronounced which greatly enhances the effectiveness of the VES [5]. The Schlumberger configuration is often preferred due to its deep penetration and less field work. The use of inversion software such as IPI2Win, RES1D, and ZondIP1D for 1D modelling is robust, but the issue of non-uniqueness of solutions continues to be a basic problem [6].

2.2 Dar–Zarrouk Parameters in Hydrogeophysics

The Dar–Zarrouk parameters provide a robust hydrogeophysical link, where S correlates with protective clay cover and T with aquifer transmissivity (Singhal & Niwas, 2014). Recent studies in the Niger Delta and Anambra Basin have successfully applied the DZ parameters for regional groundwater assessments [8][11]. Longitudinal conductance (S) has been used to map aquifer vulnerability, with higher values indicating thicker protective clay. Transverse resistance (T) has been correlated with borehole yield data, providing empirical relationships for transmissivity estimation in unconsolidated aquifers [4]. These parameters effectively "filter out" equivalence issues by considering the integrated electrical properties of the entire column.

2.3 Cluster Analysis in Geosciences

Cluster analysis is a subset of multivariate statistical technique, to find a natural groupings to the data. As a member of partitional clustering algorithms, K-means reduces the total within-cluster sum of squares, hence appropriately grouping observations with the same characteristic vectors [12]. Its use in geophysics is becoming more widespread. For example, [9] applied K-means to the classification of VES data in a basement complex, delineating areas of deeply weathered and fresh bedrock. In the sedimentary setting of the Dahomey Basin, [10] integrated electrical resistivity and seismic refraction data within a clustering framework to determine units of the confined aquifers. These works illustrate the value of clustering to minimize interpretive bias and uncover a broader pattern from geophysical surveys that are conducted in a dense point configuration.

2.4 Hydrogeology of the Anambra Basin

The Anambra Basin has complicated stratigraphy that determines the occurrence of and flow of groundwater. For instance, the stratigraphic sequence constitutes, from bottom to top, the marine Nkporo Group/Imo Shale (aquitard), arenaceous Ajali Sandstone (local aquifer), the alternating sandstone and shales of the Ameki Formation (major aquifer), and the predominantly sandy Benin Formation (extensively prolific aquifer). A recharge of groundwater that flows unconfined to Benin Sand Formation and semi-confined to Ameki Formation is predominantly from precipitation. The intercalation of these units leads to the formation of a heterogeneous groundwater system, which is characterized by high-yielding sands that are abruptly juxtaposed to clay-filled channels. This explains the uneven borehole success, as reflected in the overlying reports by the local water authorities [1].

III. METHODOLOGY

3.1 Vertical Electrical Sounding (VES) Data Acquisition

Out of a total of 20 VES stations, 20 were developed, and each covered a different geomorphological unit (upland lateritic, mid-slopes, and alluvial valleys). Schlumberger electrode configuration and current electrode distance spacing (AB/2) were incrementally increased from 1.5 to a maximum of 300 m to interrogate depths that would encounter major aquifer units and, to the extent possible, near the top of the resistive basement or conductivity (Imo Shale). A digital resistivity meter (ABEM Terrameter SAS 1000) was used to measure the apparent resistivity (ρa) of the ground. ρa was calculated from the formula [ρ_a = K(Δ V/I)] where K is the geometric factor for the Schlumberger array [K = π * \frac{\left((AB/2)^{2} - (MN/2)^{2} \right)} {MN}] Quality control procedures were implemented by repeating the measurements for some spacings, ensuring that the electrodes were properly grounded with saline water, and smoothed to check for minimal cultural noise.

3.2 Data Processing and 1D Inversion

The RES1D Software applies a least-square fitting procedure by employing a linear least squares optimization algorithm to reduce the curve misfits iterated, based on the models developed by [6]. Borehole logs and standard curve fit models (H, A, K, Q) were used to first manually build initial models for the various layers. Inversion, by iteratively changing the adjustments of the layers model parameters when the root mean square (RMS) of the misfit was determined between measured and the fit model of the curves to a target convergence of 5 percent, and to limit model parameters. Final 4 to 7 to a maximum of 7 layers, and for each of the layers true resistivity (ρi) in Ohm meters and thickness (hi) in meters, were determined.

3.3 Computation of Dar–Zarrouk Parameters

For each layered model, the Dar–Zarrouk parameters were computed for the sequence above the first continuous, highly conductive layer (interpreted as the top of the Imo Shale aquitard) or until the maximum probed depth was reached. The following formulas were applied:

- Longitudinal Conductance (S): S = Σ (h_i / ρ_i) (units: Siemens)
- Transverse Resistance (T): T = Σ (h_i * ρ_i) (units: Ωm^2)

These calculations were automated using a Python script to guarantee consistency. The S values were classified for protective capacity using the classification scheme validated by [8].

3.4 Feature Engineering and Standardization

A feature set was created for the number of layers, and minimal, maximal, and mean resistivity, max depth, S, and T for each site. These features underwent standardization before three clusters were created for individual geoelectrical domains using K-means (k=3).

3.5 K-Means Clustering Analysis

A standardized, 10-dimensional feature matrix (20 sites × 10 features) was used as the input for the K-means clustering algorithm from the Scikit-learn Python library [13]. To find the optimal number of clusters (k), the Elbow Method (where the within-cluster sum of squares is plotted against k) and Silhouette Score methods were used, both indicating the optimum number was three. The last clusters assigned to each VES station were 0, 1, and 2, signifying their respective clusters.

- Using the VES geographic coordinates, individual spatial cluster maps were produced. Preliminary T and S zonation maps were created using Inverse Distance Weighting (IDW) interpolation, and the cluster maps were used to validate them.

3.6 Groundwater Potential Assessment

An easy-to-use Groundwater Potential Index (GWPI) was developed to assess the sites within the most prosperous cluster (Cluster 0). As both high transmissivity (T) and high protection (S) are favourable, the GWPI was computed using the equation: GWPI = (Tnorm + Snorm) / 2, where Tnorm and Snorm are the min-max normalized values of T and S in Cluster 0, with both ranging between 0 and 1. Following this, the sites' GWPI scores ranked them in order.

IV. RESULTS AND INTERPRETATION

4.1 Inversion Results and DZ Parameters

The inverted VES models presented a characteristic 4 to 6-layer subsurface pattern. The first layer of topsoil (Layer 1) showed a high degree of variation (50–1500 Ωm), indicating a lateritic or clayey make-up. The subsequent layers consisted of a regular pattern with alternates of resistive units (40-250 Ωm), presumed to be sandy aquifers, and conductive units (<25 Ωm), presumed to be clay layers. The computed Dar-Zarrouk parameters were S ranging 0.08 to 2.45 Siemens and T from 850 to 18,500 Ωm^2, which showed a great degree of subsurface variability.

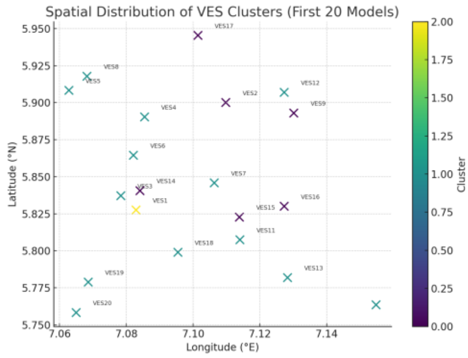

Figure 2. Spatial distribution of the VES clusters (20 models).

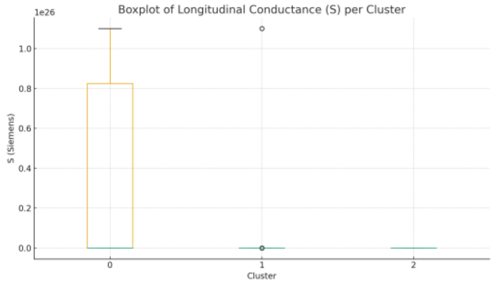

Figure 3. Boxplot of Longitudinal Conductance (S) per Cluster.

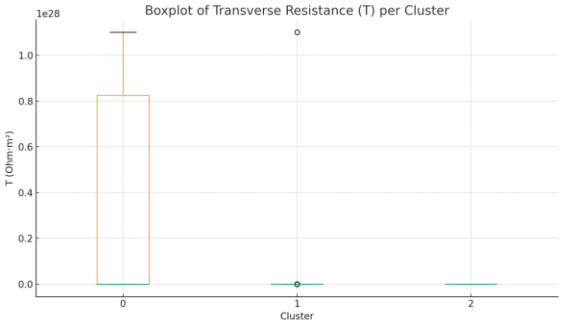

Figure 4. Boxplot of Transverse Resistance (T) per Cluster

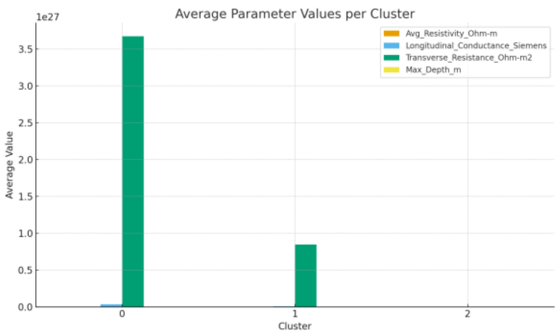

Figure 5. Average parameter values per cluster.

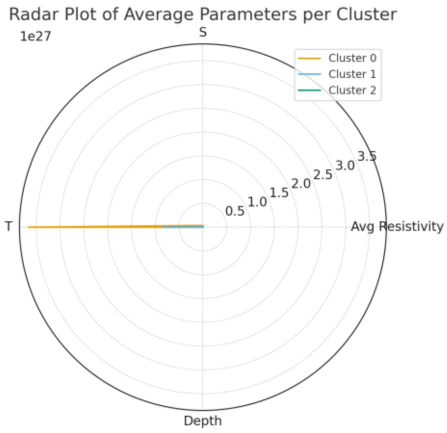

Figure 6. Radar plot of average parameters per cluster

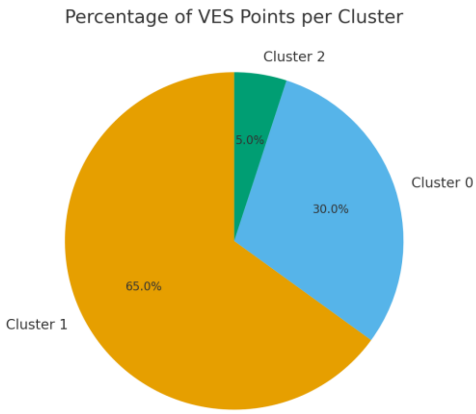

Figure 7. Percentage of VES points in each cluster.

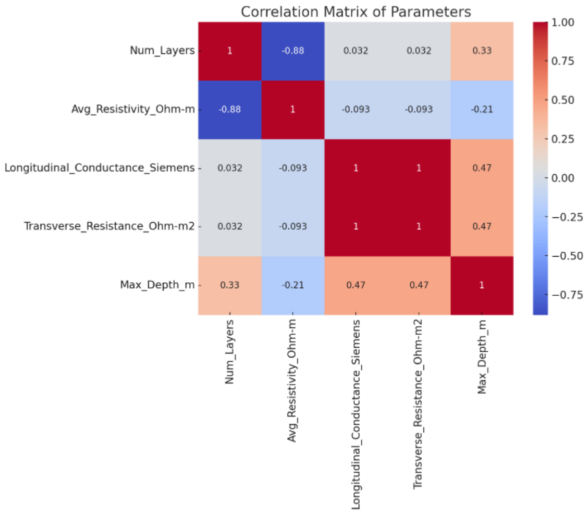

Figure 8. Correlation matrix of the parameters.

4.2 Cluster Characteristics and Hydrogeological Interpretation

K-means clustering partitioned the 20 sites into three distinct and coherent groups.

- Cluster 0 (Multi-layered Aquifer Systems – 9 sites): This cluster is defined from others by resistance being at moderate level average (120 (\pm) 45 (\Omega)m), high number of layers (5.6 (\pm) 0.7), and T \ were high at (9850 (\pm) 3200 (\Omega)m^{2}). While S values being moderate (1.1 (\pm) 0.4 simulation). The confinement statistics profile the succession of beds into a stratified sequence. a high T indicates the cumulative thickness of multiple water bearing sand beds with high resistivity and moderate S indicates there are clay beds over or interbedded. moderate S also indicates a protective mantlin well to moderate from overlying pollutants.

Hydrogeological Interpretation: This cluster is assigned as the fully developed facies of the productive layers of the Ameki and Benin Formations. Given the combination of high yield and moderate S, this is the primary (optimal) area for sustainable groundwater development.

- Cluster 1 (Resistive Basement/Lateritic Highs – 5 sites): This cluster shows exceptionally high mean resistivity (950 ± 400 Ωm), with only a few number of layers (3.8 ± 0.4), moderate T (4200 ± 1500 Ωm^2), and very low S (0.15 ± 0.07 Siemens). The resistivity and simplicity of stacking is diagnostic.

Hydrogeological Interpretation: The sites sit on topographic high and are weathered basement or thick lateritic crusts exposed at the surface or are where the sedimentary cover is thin and is composed of iron-cemented sands. The low S suggests of no protective clay cover and leads to the groundwater being unconfined and very susceptible. The T is moderate and could be interpreted as the correct value or could be misleading and suggests the resistivity of a thick dry lateritic or compacted layer rather than a saturated aquifer. They are very poor prospects for groundwater boreholes, but may be adequate for engineering foundations.

- Cluster 2 (Conductive Clay/Saline Zones – 6 sites): Characterized by low average resistivity (22 (\pm 8) Ωm), certain layered structure, very low T (1250 '\pm 600' Ωm^2), and moderately high but fluctuating S (1.8 \pm 0.6 Siemens). The dominant characteristic was the low overall resistivity.

Hydrogeological Interpretation: This cluster correlates with thick, clay-rich deposits, likely the Imo Shale or alluvial clays in the lower parts of the landscape. The high S corroborates the substantial cumulative clay thickness, which could provide excellent confinement to any underlying aquifer. Nonetheless, a very low T suggests the intercalation of thick, resistive sandy layers within the depth of investigation is absent. The pore water within these clays is typically brackish or saline. These zones are aquitards and are of very low potential and high risk in terms of groundwater extraction due to likely poor water quality and low yield.

4.3 Spatial Distribution and Zonation Map

The spatial plot of the cluster assignments reveals a coherent and geologically sensible pattern:

- Cluster 0 sites are mainly clustered in the middle and northern portions of the study site, which line up with the regional distribution of the prolific Benin Formation sands.
- Cluster 1 consists of single, isolated "islands" on topographic highs, congruent with the presence of laterite capping.
- Cluster 2 is found in the southern and southeastern low-lying regions, indicating probable thicker accumulations of the Imo Shale or alluvial clays.

This spatial pattern grants a doubly predictive map for groundwater scouting and delineates the “green zones” (Cluster 0), “red zones” (Cluster 2), and “engineering zones” (Cluster 1).

The spatial coherence of the clusters is a reflection of the already documented sandy (resistive) and clayey (conductive) stratigraphy pattern in the Anambra Basin margin, which can also explain the historically irregular success rates of boreholes within short distances [2]. The map serves as a decision- grade tool for planned intervention, indicating the most sought-after aquifer zones and areas needing further field verification.

4.4 Groundwater Potential Ranking

The Groundwater Potential Index (GWPI) has been computed for the most promising Cluster 0. Thus ranks the sites most suitable for the drilling of boreholes:

1. Umuma Isiaku (VES18) – GWPI: 0.94
2. Amanator (VES15) – GWPI: 0.87
3. Akokwa (VES2) – GWPI: 0.81

   These sites combine the highest transmissivity potential (T) with a robust protective capacity (S) within the aquifer zone.

V. CONCLUSION

This research created and illustrated a full cluster-based framework for conducting geoelectrical zonation of aquifers in the complex Anambra Basin. The method used recognized and elucidated three different geoelectrical domains in an unbiased manner, which are: Productive Multi-layered Aquifers (Cluster 0); Resistive Basement/Lateritic Highs (Cluster 1); and Conductive Clay Aquitards (Cluster 2). These are geoelectrical domains which are congruent with the known geology and offer clear and unequivocal guidance for the evaluation of groundwater exploration and the determination of the degree of vulnerability of the groundwater resources in the region delimitable on a map. The efficiently constructed groundwater composite potential index ranked and spaced the communities of Umuma Isiaku, Amanator, and Akokwa for the groundwater development projects within the socioeconomically prioritized community and sustainable water supply projects.

Since this study has demonstrated the integration of classical geophysical parameters with unsupervised techniques, it has provided a robust paradigm that is unbiased and scaleable for the analysis of the potential of groundwater resources in a heterogeneous sedimentary basin. The proposed framework is readily transferable to other regions with similar hydrogeological problems and is a sphere that will empower ground water management systems to operate in an efficient and sustainable manner.

REFERENCES

[1] Ezeh, C.C. (2012). Hydrogeophysical Studies for the Delineation of Potential Groundwater Zones in Enugu State, Nigeria. International Research Journal of Geology and Mining, 2, 103-112

[2] Onu, F. The Southern Benue Trough and Anambra Basin, Southeastern Nigeria: A Stratigraphic Review. https://doi.org/10.9734/JGEESI/2017/30416

[3] Nwozor, R. N., Bassey, N. E., George, N. J., & Harry, T. A. (2025). Hydrogeophysical and Hydrogeological Characterization of Groundwater in Parts of the Benin Formation, Akwa Ibom State, Nigeria: Implications for Sustainable Water Resource Management. *Researchers Journal of Science and Technology*, *5*(1), 45–68

[4] Obasi, Ikenna & Bala, Jamilu & Aigbadon, Godwin & Anakwuba, E. & Akudo, Ernest & Onwa, Nneka. (2022). Assessment of aquifer vulnerability in the tectonically deformed sedimentary rocks in Abakaliki area, southeastern Nigeria, using geophysical and geological data. 10.21203/rs.3.rs-1725870/v1

[5] Aizebeokhai, A. P., & Oyeyemi, K. D. (2018). Geoelectrical characterization of aquifer types in the sedimentary terrain of southwestern Nigeria. *Environmental Earth Sciences, 77*(22), 763.

[6] Loke, Meng & Wilkinson, P. & Chambers, Jonathan & Meldrum, Phil. (2017). Rapid inversion of data from 2D resistivity surveys with electrode displacements. Geophysical Prospecting. 66. 10.1111/1365-2478.12522

[7] S. Niwas and D. C. Singhal (2012). Estimation of Aquifer Transmissivity from Dar-Zarrouk Parameters in Porous Media, Journal of Hydrology, Vol. 50, 1981, pp. 393-399. doi:10.1016/0022-1694(81)90082-2

[8] Okonkwo A, Ugwu GZ. (2015). Determination of Dar-Zarrouk parameters for prediction of aquifer protective capacity: A case of Agbani Sandstone aquifer, Enugu State, Southeastern Nigeria. Environmental Science, Geology. DOI:10.14303/IRJGM2015.107

[9] Nugraha, G.U., Bakti, H., Lubis, R.F. *et al.* Data driven resistivity zonation integrating inversion kriging and clustering for subsurface characterization in groundwater exploration. *Discov Water* 5, 62 (2025). https://doi.org/10.1007/s43832-025-00261-7

[10] Adiat, K.A.N. & Ajayi, O. & Akinlalu, Ayokunle & Tijani, Ismaila. (2020). Prediction of groundwater level in basement complex terrain using artificial neural network: a case of Ijebu-Jesa, southwestern Nigeria. Applied Water Science. 10. 10.1007/s13201-019-1094-6

[11] Laouini, G., Etuk, S.E. and Agbasi, O.E. (2017). Delineation of Aquifers Using Dar Zarrouk Parameters in Parts of Akwa Ibom, Niger Delta, Nigeria. Journal of Hydrogeology and Hydrologic Engineering, 6, Article No. 18

[12] Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. STHDA. *Available at:* https://www.datanovia.com/en/product/practical-guide-to-cluster-analysis-in-r/ (Book page with download option)

[13] Pedregosa, Fabian & Varoquaux, Gael & Gramfort, Alexandre & Michel, Vincent & Thirion, Bertrand & Grisel, Olivier & Blondel, Mathieu & Prettenhofer, Peter & Weiss, Ron & Dubourg, Vincent & Vanderplas, Jake & Passos, Alexandre & Cournapeau, David & Brucher, Matthieu & Perrot, Matthieu & Duchesnay, Edouard & Louppe, Gilles. (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12.