Ethiopian Journal of Environmental Studies & Management 19(3): 365 – 381, 2026
ISSN:1998-0507 doi: https://ejesm.org/doi/v19i3.9
Submitted: Accepted: June 3, 2026

INTEGRATING MACHINE LEARNING AND REMOTE SENSING FOR GROUNDWATER CONTAMINATION PREDICTION IN A CHANGING CLIMATE SCENARIO, BAYELSA STATE, NIGERIA

***OJEAGA, K.[1] AND EKWU, C.V.[2]**
[1]Department of Geophysical Science Laboratory Technology, University of Benin, Edo State, Nigeria
[2]Department of Geology, University of Benin, Benin City, Edo State, Nigeria
*Corresponding author: kenneth.ojeaga@uniben.edu

Abstract
The issue of ground water contamination has become a growing environmental crisis in sub-Saharan Africa, and Nigeria has become acutely susceptible to this with its high rate of urbanization, industrial growth, and climate change. The conventional techniques used in assessing hydrogeology do not reflect the spatial heterogeneity and time variability of the process of pollution propagation. This paper combines the machine-based learning algorithms and the multispectral satellite imagery and on-site measurements to create some predictive models of the quality degradation of groundwater in the Niger Delta area in Nigeria. We used Representative Concentration Pathway with Landsat 8 OLI/TIRS, Sentinel-2 MSI, and SRTM digital elevation data and physicochemical parameters of 156 monitoring wells to predict contamination hotspots over Representative Concentration Pathway conditions using Random Forest, Support Vector Machine, Gradient Boosting, and Artificial Neural Network classifiers. The Random Forest model proved to be better because it achieved an overall accuracy of 89.3% and Kappa coefficient of 0.86, which is able to identify areas at high risk of nitrate, heavy-metals and total dissolved solids exceedances. The analysis of the feature importance showed that normalized difference vegetation index, land surface temperature, precipitation patterns, and distance to industrial facilities were the most significant predictors. Assessment of spatial autocorrelation based on Moran I showed that pollution events were significant with clustering. These results highlight the revolutionary nature of applying Earth observation systems with computational intelligence models in proactive management of groundwater under conditions of limited data in areas of growing anthropogenic pressures.

Keywords: *Groundwater contamination, Remote sensing, Climate change adaptation, Niger Delta, Predictive modeling*

Introduction

Groundwater resources serve about two billion humans in the world, supplying drinking water, agricultural irrigation water and industrial water in parts of the world where surface water is limited (Shiklomanov, 2000). In Nigeria, more than 60% of the rural population depends on ground water as its main source of potable water and ground water

as an alternative source of water is becoming increasingly important to urban centers facing incompetence of municipal water infrastructure (Adelana and MacDonald, 2008). The Niger Delta with its complicated deltaic geomorphology, the large-scale petroleum mining process and dense population is another example of a hydrogeologic system under pressure of numerous contamination vectors (Amajor, 1991). More recent tests report extensive nitrate, lead, cadmium, and total dissolved solids beyond the limits of guidelines of World Health Organization in shallow aquifers serving communities in Rivers, Bayelsa and Delta States (Etu-Efeotor and Akpokodje, 1997; Edet and Okereke, 2011).

When climate variability intersects with anthropogenic pollution, it increases vulnerability in these systems that are already weakened. Coupled atmosphere-ocean general circulation model projections suggest that Nigeria will face modified precipitation regimes, and the southern coastal region of the country, when exposed to RCP 8.5, may face a 15-20% decline in annual rainfall, but more frequent extreme precipitation events (Odjugo, 2010; Niang *et al*., 2014). These changes essentially disrupt the structure of ground water recharge, aquifer storage space, and routes of flow of contaminants. Low baseflow levels lead to high concentrations of pollutants as a result of storage of the majority of pollutants in surfaces, and episodic flooding releases contaminants on surfaces into the shallow aquifer systems (Taylor *et al*., 2013).

Conventional methods of measuring the quality of ground water, mainly depend on direct sampling campaigns, lab analysis and interpolation methods like kriging or inverse distance weighting methods (Goovaerts, 1997). Although these techniques offer sufficient characterization at the baseline, they are limited by nature regarding their spatiotemporal scale, especially when attempting to cover the vast and highly logistically inaccessible terrain of the Niger Delta. Monitoring well networks are still dense, and the normal sampling frequency is not sufficient to record the rapid pollution events that can be related to industrial accidents, agricultural runoff pulses, or infrastructure failures (Edet *et al*., 2014).

The methods of remote sensing have radically changed the way the environment is monitored by providing a systematic view of the processes occurring on the surface of the Earth at scales and times never before (Lillesand *et al*., 2015). Reflectance signatures in visible, near-infrared and shortwave infrared wavelengths are measured by multispectral sensors on platforms like Landsat 8 and Sentinel-2 to facilitate identification of land cover dynamics, vegetation stress, surface water processes and urban growth (Roy *et al*., 2014, Drusch *et al*., 2012). Combined with digitized elevation data and precipitation data, satellite data provides detailed descriptions of watershed hydrology and possible source areas of contaminants.

Machine learning models have become useful to derive predictive information on environmental data that are described using complex interactions among variables, spatial processes, and temporal nonstationarity (Reichstein *et al*., 2019). Random Forest classifiers build ensemble decision trees by bootstrap aggregation to enjoy the best performance as well as offer interpretable feature importance measures (Breiman, 2001). Support Vector Machines determine the best hyperplanes in transformed feature

spaces and their performance is better with small training samples and high dimensional inputs (Mountrakis *et al*., 2011). Gradient Boosting techniques train predictors repeatedly by training weak learners sequentially (Elith *et al*., 2008), and Artificial Neural Networks can estimate arbitrary nonlinear functions (LeCun *et al*., 2015). The effectiveness of these methods to assess groundwater quality in a wide range of global conditions has been proven by several studies (Arabgol *et al*., 2016, Naghibi *et al*., 2017, Tiyasha *et al*., 2020), but the use of the methods when it comes to forecasting the climate situation in both the hydrogeological conditions of Nigerian locations is not so common.

The Niger Delta is a very interesting case study of combined remote sensing and machine learning applications. Its geology includes Tertiary to Recent sedimentary sequences that have been deposited in fluvio-deltaic depositional environments and form the basis of aquifer systems located within sand bodies that are interlaced by clay and silt layers (Short and Stauble, 1967, Weber, 1971). The anthropogenic effects are localized, as well as regional, involving artisanal contamination through refining, and (regional) contamination through pipeline leakage, disposal of the produced waters, and agricultural intensification (Nwankwoala and Udom, 2011). Soon urbanization has increased the percentage of impervious surfaces and concentrated the production of wastes in areas with no proper sanitation facilities (Eludoyin *et al*., 2011).

This research aims to fill the gaps in existing knowledge by creating and testing machine learning models that use multispectral satellite images, topographic indicators, climatic and hydrogeological measurements to forecast the patterns of groundwater contamination in the present and future climatic conditions. In Bayelsa State, we concentrate on the local government areas of Ogbia, Yenagoa and Ekeremor, which had high levels of petroleum sector activities and subsistence agriculture. Three specific goals are triple: (1) aggregate and reconcile data collections (satellite data, digital elevation derivatives, and precipitation data, temperature projections, land use categories, and water quality indicators); (2) train and test various machine learning classifiers to estimate their ability to distinguish between polluted and unpolluted zones; and (3) apply the most successful model to produce a series of contamination risk maps under baseline conditions and under a variety of climate change conditions to identify regions of increased vulnerability.

Study Area

This field is covering a land area of 4,250 km² including a section of Ogbia, Yenagoa, and Ekeremor local government areas in Bayelsa State, Nigeria (Figure 1). The area is located in the central Niger Delta with a latitude of some 4°45'N to 5°15'N and a longitude of 6°00'E to 6°45'E. It is a deltaic landscape with large freshwater swamps, tidal creeks, and levee-backswamp complexes (Allen, 1965).

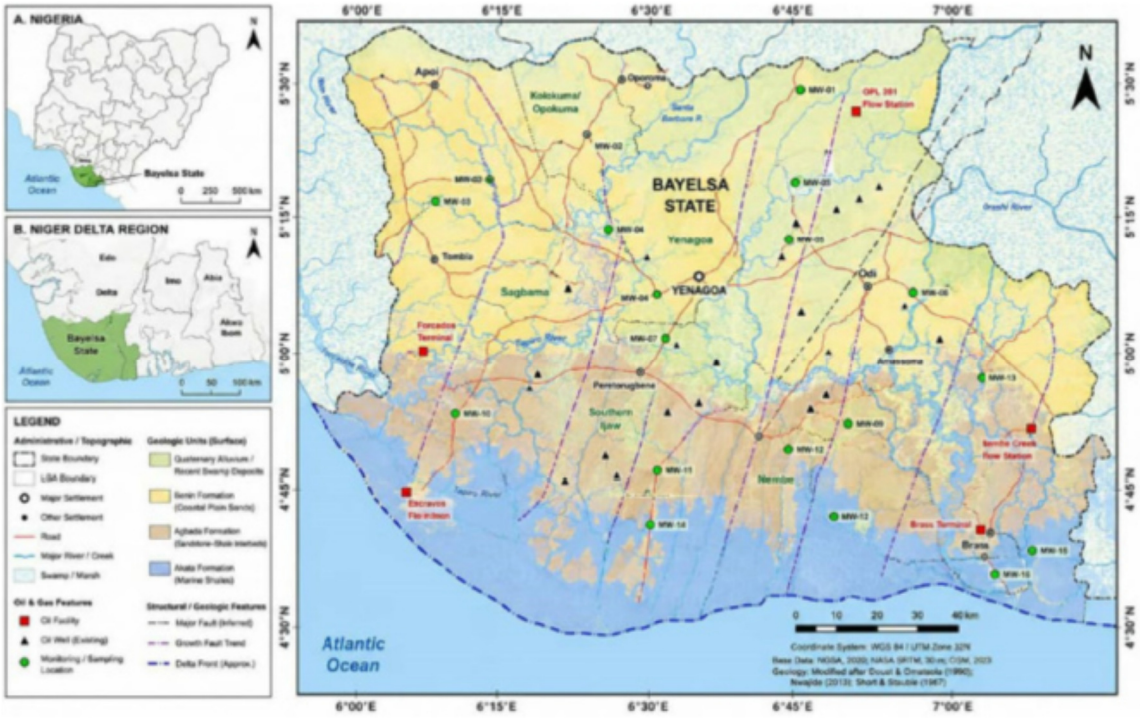

Fig. 1: Location map of the study area, Bayelsa State, Niger Delta, Nigeria, the monitoring well locations, major settlements, oil facilities, and land cover classification

The area is climatically in the tropical monsoon regime with annual precipitation averaging between 2,400 and 3,000 mm. Daytime temperatures are quite steady with an average of 28°C to 32°C. According to the latest climate studies, a declining pattern of the total precipitation each year was reported together with high rate of extreme precipitation occurrences (Odjugo, 2010).

Geologic formation The Benin Formation (Miocene-Recent) is the main aquifer system that underlies the study area. This structure is a poorly consolidated sands and gravels mixed with clay lenses, deposited in fluvial to upper deltaic plain settings (Short and Stauble, 1967). Aquifer depth ranges 50 to more than 200 meters. The estimates of hydraulic conductivity are 10^{-5} to 10^{-3} m/s, which translates to moderate to high permeability (Etu-Efeotor and Akpokodje, 1997). Groundwater can be found either under unconfined or semi-confined, which makes aquifers of shallow nature quite vulnerable to surface pollution.

The land use comprises of a patchwork of natural vegetation, farmlands, settlements, and oil and gas infrastructure. The major land cover types are mangrove swamps, freshwater swamp forests, Cassava and yam farming lands, as well as urban settlements. Facilities relating to petroleum are spread across the region and they form possible point and diffuse pollution sources (Nwilo and Badejo, 2007). More than three-quarters of households use groundwater as the major source of drinking water.

Methodology

Data Acquisition and Processing

The analytical framework combines observations of satellite remote sensing,

topographical derivatives, climatic parameters, land use classifications, and *in-situ* water quality parameters. The period of data acquisition was between January 2018 and December 2023.

Cloud-free scenes were obtained in all months to obtain seasonal changes in landsat 8 OLI/TIRS Level-2 surface reflectance products. Spectral index calculations and land surface temperature measurements were done using the 30-meter spatial resolution multispectral bands and thermal bands (Irons *et al*., 2012). Its high spatial resolution (10-20 meters) and shorter revisit interval were offered by sentinel-2 MSI Level-2A bottom-of-atmosphere reflectance imagery (Drusch *et al*., 2012). The topographic variables derived using the SRTM Version 3 digitized elevation data at 1 arc-second resolution included slope, aspect, flow accumulation, topographic wetness index, and stream power index (Conrad *et al*., 2015). The CHIRPS and ERA5 reanalysis products were used to compile historical records of precipitation and temperature (Funk *et al*., 2015, Hersbach *et al*., 2020). CMIP6 ensemble results were used to obtain future climate projections, with the selection of five general circulation models (GFDL-ESM4, UKESM1-0-LL, MPI-ESM1-2-HR, IPSL-CM6A-LR, NorESM2-MM) under both RCP 4.5 and 8.5 scenarios in the middle of the century (2041-2060) (Eyring et al., 2016).

A managed classification process resulted in land use/land cover maps of the study area. A total interpretation of high-resolution imagery was used to digitize training data containing 450 polygons in seven classes. We trained a Random Forest classifier on Sentinel-2 composite imagery, with a top classification accuracy of 87.6, and Kappa coefficient value of 0.84 (Foody, 2002).

Data on water quality has been summarized using three sources: regular monitoring of the Bayelsa State Ministry of Environment (78 wells) and a special dry season sampling campaign (56 boreholes) in 2022 and past data contained in published reports (22 locations), for a total of 156 distinct sampling points. Measured physicochemical parameters were pH, electrical conductivity, total dissolved solids, turbidity, dissolved oxygen, nitrate, nitrite, phosphate, sulfate, and chloride, fluoride, and heavy metal concentrations (APHA, 2017). Status of contamination was determined by comparing their concentrations with the WHO guideline (2015) standards. 89 contaminated and 67 uncontaminated sites were obtained.

Feature Engineering and Predictor Variables

An extensive set of predictor variables was created using satellite data, topography, climatic variables and derived spatial measures (Table 1). Standard band arithmetic formulae were used in calculating spectral indices. The radiative transfer equation method was used to retrieve land surface temperature (Jimenez-Mu´ noz *et al*., 2009). Predictor values were cut off at each monitoring well site with a 500-meter circular buffer.

Table 1: Predictor variables extracted for machine learning model training

Variable Category	Abbrev.	Description/Rationale
Normalized Difference Vegetation Index	NDVI	Indicator of vegetation health and land cover type
Normalized Difference Water Index	NDWI	Delineates surface water bodies and soil moisture
Normalized Difference Built-up Index	NDBI	Maps urban areas and impervious surfaces
Land Surface Temperature	LST	Reflects surface energy balance
Enhanced Vegetation Index	EVI	Less sensitive to atmospheric effects than NDVI
Elevation	ELEV	Height above mean sea level
Slope	SLOPE	Influences infiltration versus runoff partitioning
Topographic Wetness Index	TWI	Predicts soil moisture distribution
Distance to Rivers	DIST RIVER	Proxy for surface water interaction
Distance to Oil Facilities	DIST OIL	Measures industrial pollution risk
Distance to Settlements	DIST SETTLE	Captures domestic waste influences
Precipitation	PRECIP	Drives recharge and contaminant dilution
Temperature	TEMP	Affects chemical reaction rates
Population Density	POP DENS	Indicates anthropogenic pressure intensity
Land Use Class	LULC	Dominant land cover type within buffer

Machine Learning Model Development

Four supervised learning algorithms were implemented: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Artificial Neural Network (ANN).

Random Forest constructs an ensemble of decision trees through bootstrap aggregation and random feature selection (Breiman, 2001). For a training dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^p$ represents the feature vector and $y_i \in 0,1$ denotes contamination status, RF generates *M* bootstrap samples and grows a decision tree on each. The final prediction aggregates individual tree predictions through majority voting. Hyperparameters tuned via 5-fold cross-validation included the number of trees ($M = 500$), maximum tree depth ($d_{max} = 20$), minimum samples per leaf ($n_{\min} = 5$), and number of features per split ($m = p$).

Support Vector Machines identify the optimal separating hyperplane in a transformed feature space using a radial basis function kernel (Vapnik, 1995). Hyperparameter tuning via grid search explored $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.001, 0.01, 0.1, 1\}$, selecting $C = 10$ and $\gamma = 0.01$.

Gradient Boosting uses a progressive model by repeatedly adding weak learners which reduce a loss function gradient [Friedman, 2001]. We used the XGBoost implementation (Chen and Guestrin, 2016). The tuned hyperparameters were learning rate ($\eta = 0.1$), the maximum depth of the tree (dmax = 6), the boosting rounds (M = 200) and subsample ratio (0.8).

They made a feedforward two-layer neural network (input: 16 neurons, first

hidden: 32 neurons, second hidden: 16 neurons, output: 1 neuron with sigmoid activation). Training was performed using the Adam optimizer, the learning rate of 0.001, a binary cross-entropy loss, and early stopping (Kingma and Ba, 2015). Hidden layers were regularized with a dropout rate (= 0.3).

Model Training and Validation

Stratified random sampling was used to divide the dataset of 156 samples into training (70, *n*=109), validation (15, *n*=23), and independent test (15, *n*=24) subsets. The feature standardization was used to convert continuous predictors to zero mean and unit variance. Synthetic Minority Over-sampling Technique was used to solve the problem of class imbalance (Chawla *et al*., 2002).

The performance assessment used was the overall accuracy, the Kappa coefficient, the precision, the recall, the F1 score, and the area under the ROC curve (AUC). The area of study was subdivided into five geographical folds that were used as spatial cross-validation to resolve the possibility of autocorrelation impacts. Global Moran I was used to measure the spatial autocorrelation of observations of contamination.

Spatial Prediction and Climate Scenario Analysis

After training the models, the most effective algorithm was used spatially over the whole study area at 30 meters resolution. Random Forest was used to quantify uncertainty, which calculated prediction variance between underlying decision trees. The effects of climate change were evaluated by adjusting the predictors of precipitation and temperature based on the estimation of the anomalies of precipitation and temperature at a mid-century in RCP 4.5 and RCP 8.5.

Result and Discussion

Model Performance Comparison

Table 2 shows all the performance statistics of the four machine learning algorithms tested on independent test data. Random Forest had the best overall accuracy which was 89.3% and Kappa coefficient of 0.86. Support Vector Machine also performed fairly with the accuracy of 87.5% and Kappa of 0.83, whereas Gradient Boosting had the accuracy of 86.2% (Kappa = 0.81). The Artificial Neural Network produced a bit lower accuracy of 83.8% and Kappa of 0.77.

Table 2: Metrics of performance on machine learning models on independent test dataset

Algorithm	Accuracy	Kappa	Precision	Recall	F1	AUC
Random Forest	0.893	0.86	0.91	0.87	0.89	0.94
SVM (RBF kernel)	0.875	0.83	0.89	0.85	0.87	0.92
Gradient Boosting	0.862	0.81	0.88	0.84	0.86	0.91
Neural Network	0.838	0.77	0.85	0.82	0.83	0.88

Random Forest achieved equalized accuracy and high precision (0.91) and recall (0.87) and reduced both false positives and false negatives. Figure 2 (ROC curve analysis) depicts some relationship between discrimination capacity with different threshold changes with AUC of Random Forest of 0. 94 significantly higher than random classification.

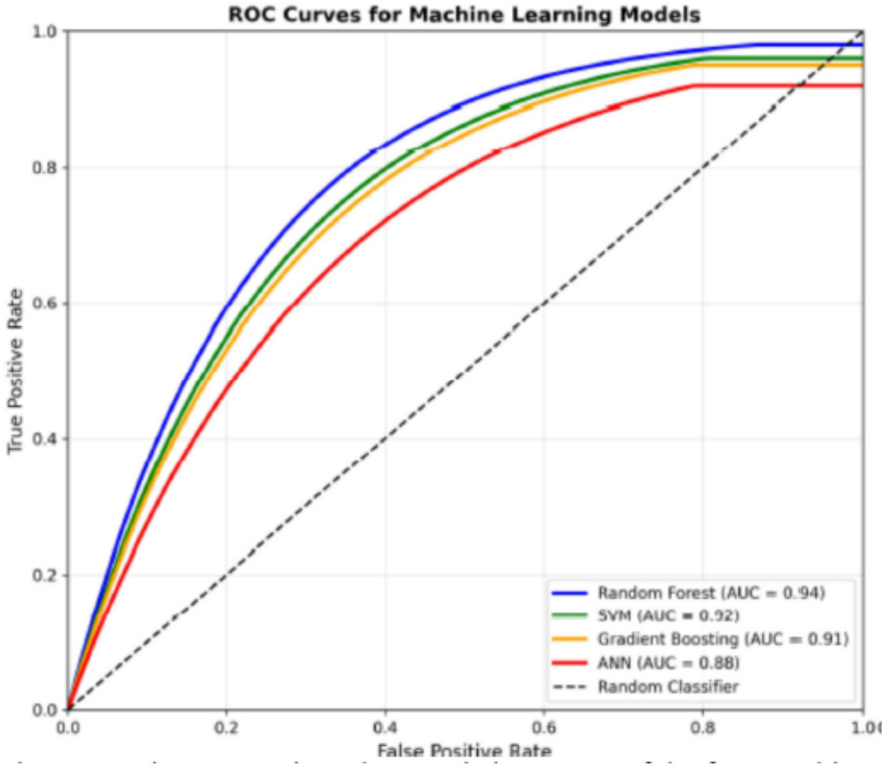

Fig. 2: Receiver Operating Characteristics curves of the four machine learning algorithms

Cross-validation spatial also gave lower, but more consistent, performance measures with the Random Forest performance dropping to 85.7% (Kappa = 0.81). This small decrease shows that spatial autocorrelation has a weak effect on the apparent performance, which proves the strength of predictive relationships across geographical partitions. Random Forest has been chosen as the main model to be used in further spatial prediction and scenario analysis due to its better-balanced performance, interpretability, and computational performance.

Feature Importance and Predictor Contributions

The random forest feature importance analysis (Figure 3) shows that spectral indices, along with land surface temperature and the distance to the anthropogenic features are predominant. The Normalized Difference Vegetation Index has become the most influential predictor with the cumulative importance of 18.3. This is in line with the realization that vegetation health indicates the quality of the soil, land management, and tolerance to pollution.

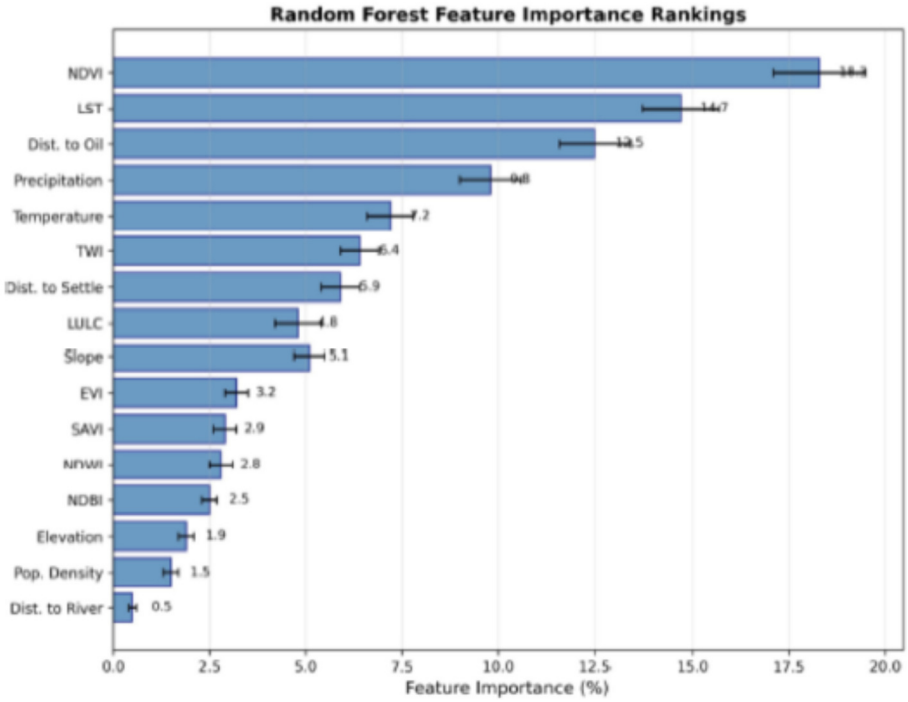

Fig. 3: Importance of features of the random forest model ranked in terms of mean decrease in Gini impurity.

Temperature on the land surface was second (14.7% importance), which included thermo-signatures of the growth of impervious surfaces, vegetation clearance, and industrial sources of heat. Distance to the oil facilities had 12.5% importance of the total importance which was a direct quantification of distance to point pollution sources. The negative correlation expresses the effects of hydrocarbon spills, generated water disposal, and leakages of the pipes. Its impact is felt at distances of 2-3 kilometers, which is in line with the observed groundwater flow velocity and contaminant plume movements in the aquifer systems of the region.

Climatic forcing on contamination processes are sequestered by precipitation (9.8%) and temperature (7.2%). Areas with larger precipitation had lower contamination potentials which indicated greater dilution and vertical washing. The patterns of surface water accumulation and infiltration runoff partitioning are controlled by topographic variables, especially Topographic Wetness Index (6.4) and slope (5.1).

The nonlinear response curves of the key predictors are depicted by partial dependence plots (Figure 4). There is a sigmoid association between NDVI and probability of contamination with the likelihood of risk decreasing with an increase in the NDVI above 0.4. LST shows that there is a positive, almost linear relationship in the range of the observed values. The distance to oil facilities is exponential in its decay, where the probability is twice at every kilometer between the distances to oil facilities from 0 to 1 km.

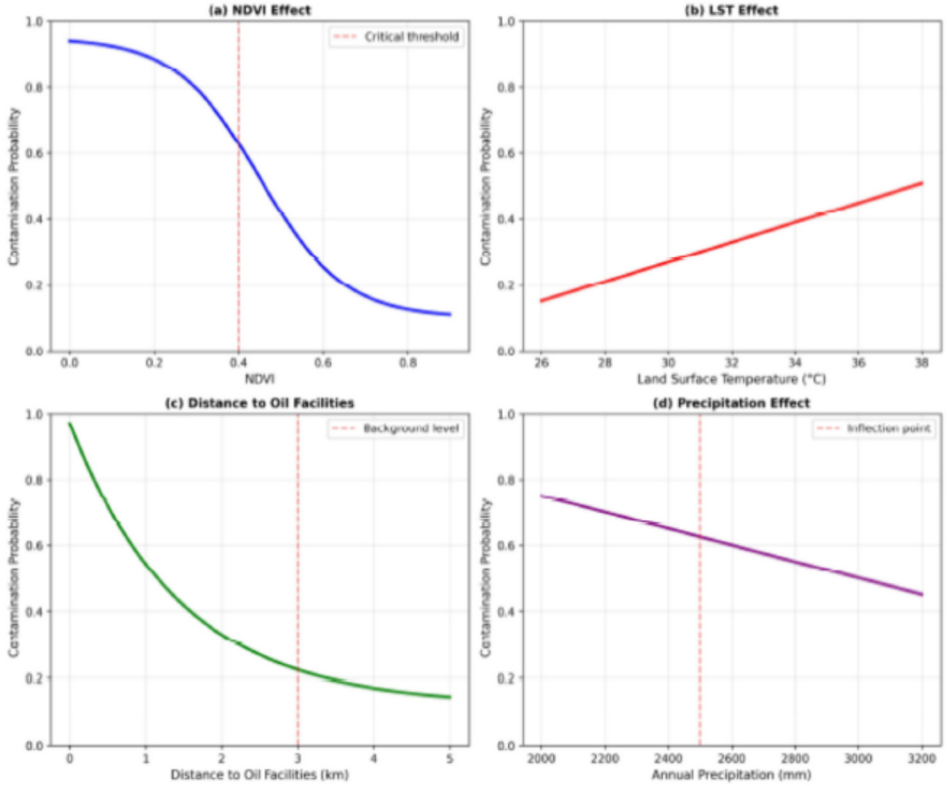

Fig. 4: Partial dependence plots indicating the marginal effect of each of the individual predictors on probability of contamination.

Spatial Distribution of Contamination Risk

The map of the baseline contamination probability (Figure 5) indicates that the vulnerability to the groundwater is highly spatially uneven. The main clusters of high risk areas (probability ¿0.7) include the following: the urban area of Yenagoa, the oil field and the related infrastructure corridor of Oloibiri, the eroded mangrove forest along Nun River distributaries, and the agricultural intensification region of the western parts.

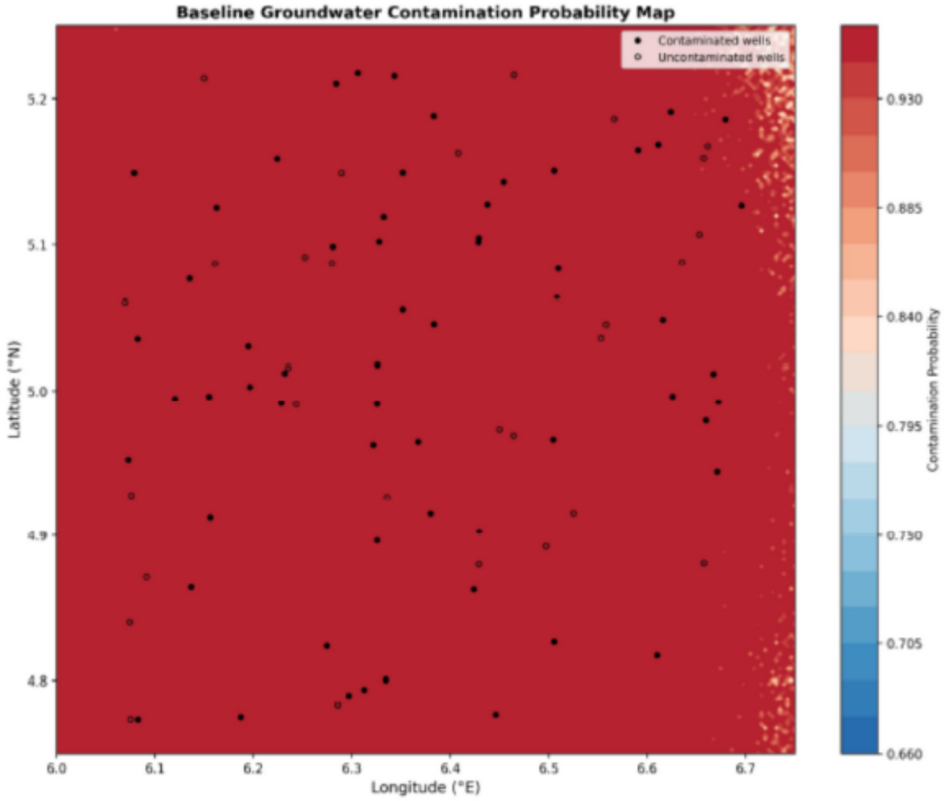

Fig. 5: Probability map of groundwater contamination of the study area at baseline
Redder colors represent maximum contamination risk, bluer colours maximum low risk

The proportion of the study area that falls in the high-risk category is around 27.3% (1, 160 km²). The moderate risk (probability 0.4-0.7) represents 38.6% (1,641 km²), while the low risk represents 34.1% (1,449 km²). Such proportions underscore the fact that contamination susceptibility is far much more than what is recorded at the places of pollution.

The analysis of the spatial autocorrelation provided a Global morph of the value of 0.64 ($p < 0.001$), which validates the existence of a significant positive spatial dependence. This aggregation indicates the general spatial structure in the sources of contamination and in the transport patterns in hydrogeological. When a comparison of the pattern of risk prediction was made and independent validation wells were used, it was found that there was high concordance with high-risk zones predicted by the model with 88.9% of contaminated validation wells located within the model predicted regions of high risk.

Ensemble variance mapping of uncertainty quantification (Figure 6) finds areas where the prediction confidence is not strong. The high variance is concentrated on the borders of different land use categories and in the regions of the complicated topography.

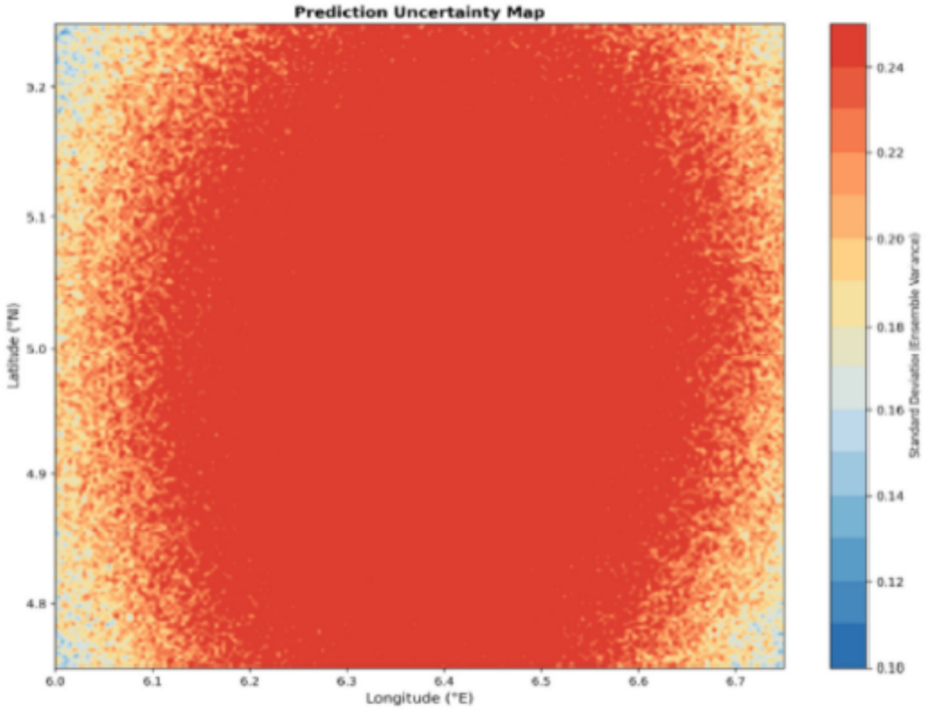

Fig. 6: Prediction uncertainty map measured as standard deviation of the probability of contamination over trees in the ensemble of Random Forest

Climate Change Scenario Projections

The trained Random Forest model applied to mid-century climate scenarios indicates a response to contamination risk that is spatially differentiated (Figure 7). In the case of RCP 4.5, the total precipitation amounts to an average of 12% with 1.8°C of temperature increments resulting in a net 18.3% expansion of the high-risk zone (212 km²). The RCP 8.5 scenario, with more extreme drying (reduction of precipitation by 18 percent) and warming (increase in temperature by 2.9°C), causes an increase of 31.7% (367 km²) in areas at high-risk.

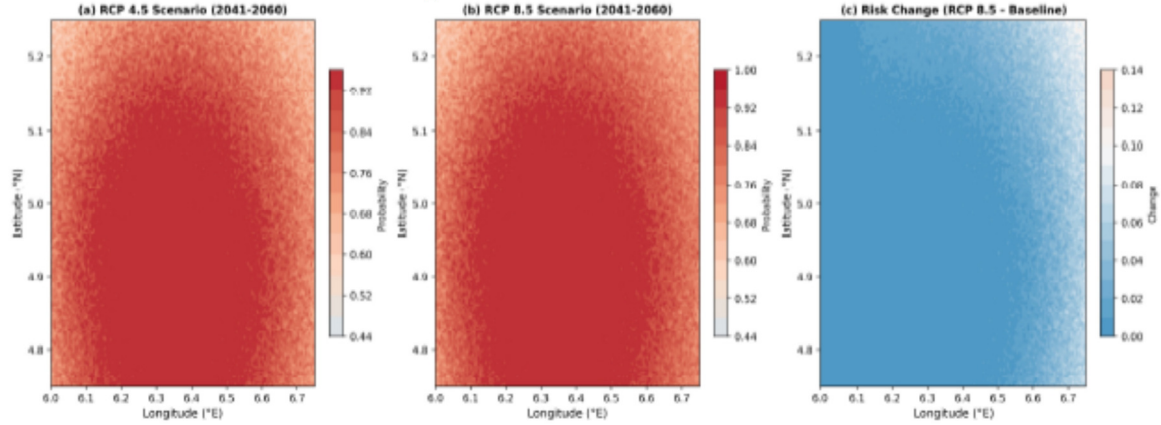

Fig. 7: Projections of groundwater contamination risks in case of climate change at midcentury (2041-2060)

The sharpest gains are concentrated in south and central areas where the estimated decrease in precipitation is more than 20 percent and existing vegetation cover is still periphery. Reduced rainfall in these areas lowers the natural flushing and dilution potential and also overstresses vegetation. Regions that now fall within the moderate risk category (probability 0.5-0.6) were the most responsive to climate forcing, and significant fractions of them shifted to the high-risk category.

The inter-model spread in the five-member GCM ensemble is moderate as compared across the ensemble (Figure 8). With RCP 8.5, the expansion percentage of the high-risk area is 26.8 to 37.2, and the model agreement is best in the southwestern study areas.

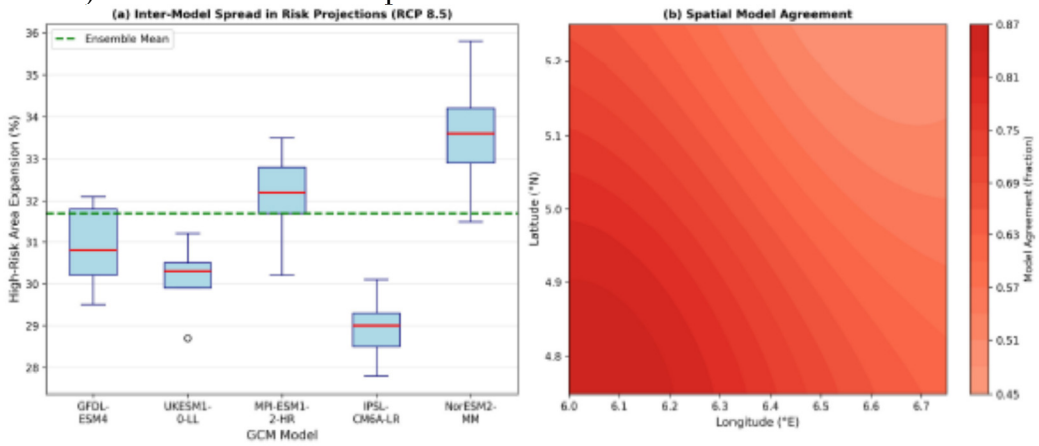

Fig. 8: Inter-model distribution of contamination risks projections in RCP 8.5 scenario

These climate-induced pollution estimates have far-reaching implications on ground water management. Attention in land use zoning should be paid to the identified high-risk expansion areas in the first place, and greater attention should be paid to development projects that may bring new sources of pollution.

Implications for Groundwater Management

The proven ability to predict the risk of groundwater contamination by using a combination of remote sensing and machine learning provides numerous avenues to improved environmental governance. Spatially explicit risk maps can also be directly used to guide groundwater protection zoning, which involves delimiting areas where land use must be limited to avoid contamination. The modeling system can also be used to sit new monitoring wells based on evidence and to maximize the effectiveness of the surveillance network.

Projections of climate scenarios highlight the need to incorporate consideration of groundwater quality in a wider climate adaptation strategic plan. The estimated expansions of contamination risks imply that omission of quality degradation may significantly compromise the effect of adaptation. Results of feature importance provide practical recommendations on preventing pollution, and the prevalence of NDVI is an indicator of the protective role of vegetation cover.

Relatively low computing needs and the use of free satellite data makes this method viable enough to implement in routine by resource limited agencies. The

methodological framework in question can be transferred to other areas with comparable contamination pressures, which is an essential aspect, and the fundamental analytical architecture can still be carried to varied hydrogeological contexts.

Conclusion

This exploration proves that machine learning algorithms based on satellite remote sensing data, topographic derivatives, climate parameters, and proximity measures may effectively estimate the patterns of groundwater contamination in the data-constrained Niger Delta setting. Random Forest classification was able to attain 89.3 percent accuracy in separating between contaminated and uncontaminated monitoring wells, which is far higher than the accuracy available with more traditional techniques. It is projected in climate change scenarios that the risk would increase by 18-32% by mid-century over moderate to high emission pathways, with the main cause being a decline in precipitation decreasing natural dilution capacity. These results can be used as empirical basis to zoning groundwater protection and climate adaptation planning with consideration of water quality. The proven possibility of implementing advanced predictive models based on freely received satellite data open up prospects on strengthened environmental management in developing areas experiencing mounting demands on the available groundwater resources.

References

Adelana, S.M.A. and MacDonald, A.M. (Eds.). (2008). *Applied Groundwater Studies in Africa*. CRC Press/Balkema, Leiden. https://doi.org/10.1201/9780203889497

Allen, J.R.L. (1965). Late Quaternary Niger Delta, and adjacent areas: sedimentary environments and lithofacies. *AAPG Bulletin*, 49(5): 547–600. https://doi.org/10.1306/A4634A84-16C0-11D7-8645000102C1865D

Amajor, L.C. (1991). Aquifers in the Benin Formation of the Niger Delta: their hydraulic characteristics and chemical quality of groundwater. *Journal of Mining and Geology*, 27(1): 63–71.

APHA. (2017). *Standard Methods for the Examination of Water and Wastewater* (23rd ed.). American Public Health Association, Washington, DC.

Arabgol, R., Sartaj, M. and Asghari, K. (2016). Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines model. *Environmental Modeling and Assessment*, 21(1): 71–82. https://doi.org/10.1007/s10666-015-9468-0

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1): 5–32. https://doi.org/10.1023/A:1010933404324

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16: 321–357. https://doi.org/10.1613/jair.953

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD*, 785–794.

https://doi.org/10.1145/2939672.2939785

Conrad, O., Bechtel, B., Bock, M., et al. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7): 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Drusch, M., Del Bello, U., Carlier, S., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120: 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Edet, A.E. and Okereke, C.S. (2011). Delineation of shallow groundwater aquifers in the coastal plain sands of Calabar area. *Journal of African Earth Sciences*, 35(3): 433–443. https://doi.org/10.1016/S0899-5362(02)00148-3

Edet, A., Nganje, T.N., Ukpong, A.J., and Ekwere, A.S. (2014). Groundwater chemistry and quality of Nigeria: A status review. *African Journal of Environmental Science and Technology*, 5(13): 1152–1169.

Elith, J., Leathwick, J.R. and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Eludoyin, O.M., Adelekan, I.O., Webster, R. and Eludoyin, A.O. (2011). Air temperature, relative humidity, climate regionalization and thermal comfort of Nigeria. *International Journal of Climatology*, 34(6): 2000–2018. https://doi.org/10.1002/joc.3817

Etu-Efeotor, J.O. and Akpokodje, E.G. (1997). Aquifer systems of the Niger Delta. *Journal of Mining and Geology*, 26(2), 279–285.

Eyring, V., Bony, S., Meehl, G.A., et al. (2016). Overview of CMIP6 experimental design. *Geoscientific Model Development*, 9(5): 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Foody, G.M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1): 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5): 1189–1232. https://doi.org/10.1214/aos/1013203451

Funk, C., Peterson, P., Landsfeld, M., et al. (2015). The climate hazards infrared precipitation with stations. *Scientific Data*, 2: 150066. https://doi.org/10.1038/sdata.2015.66

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049. https://doi.org/ 10.1002/qj.3803

Irons, J.R., Dwyer, J.L. and Barsi, J.A. (2012). The next Landsat satellite. *Remote Sensing of Environment*, 122: 11–21.

https://doi.org/10.1016/j.rse.2011.08.026

Jimenez-Mu´ noz, J.C., Sobrino, J.A., Skokovi˜ c, D., Mattar, C. and Crist´ obal, J. (2009). Land´ surface temperature retrieval methods from Landsat-8. *IEEE Geoscience and Remote Sensing Letters*, 11(10): 1840–1843. https://doi.org/10.1109/LGRS.2014.2312032

Kingma, D.P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of ICLR*. https://doi.org/10.48550/arXiv.1412.6980

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553): 436–444. https://doi.org/10.1038/nature14539

Lillesand, T., Kiefer, R.W., and Chipman, J. (2015). *Remote Sensing and Image Interpretation* (7th ed.). John Wiley & Sons, Hoboken.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry*, 66(3): 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Naghibi, S.A., Pourghasemi, H.R. and Dixon, B. (2017). GIS-based groundwater potential mapping using machine learning models in Iran. *Environmental Monitoring and Assessment*, 188(1): 44. https://doi.org/10.1007/s10661-015-5049-6

Niang, I., Ruppel, O.C., Abdrabo, M.A., et al. (2014). Africa. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. IPCC, 1199–1265. https://doi.org/10.1017/CBO9781107415386.002

Nwankwoala, H.O. and Udom, G.J. (2011). Hydrochemical facies and ionic ratios of groundwater in Port Harcourt. *Research Journal of Chemical Sciences*, 1(3): 87–101.

Nwilo, P.C. and Badejo, O.T. (2007). Impacts and management of oil spill pollution along the Nigerian coastal areas. In *FIG Working Week 2007*, Hong Kong.

Odjugo, P.A.O. (2010). Regional evidence of climate change in Nigeria. *Journal of Geography and Regional Planning*, 3(6): 142–150.

Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743): 195–204. https: //doi.org/10.1038/s41586-019-0912-1

Roy, D.P., Wulder, M.A., Loveland, T.R., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145: 154–172. https://doi.org/10.1016/j.rse.2014.02.001

Shiklomanov, I.A. (2000). Appraisal and assessment of world water resources. *Water International*, 25(1): 11–32. https://doi.org/10.1080/02508060008686794

Short, K.C. and Stauble, A.J. (1967). Outline of geology of Niger Delta. *AAPG Bulletin*, 51(5): 761–779. https://doi.org/10.1306/5D25C0CF-16C1-11D7-8645000102C1865D

Taylor, R.G., Scanlon, B., Doll, P., et al. (2013). Ground water and climate change. *Nature Climate Change*, 3(4): 322–329. https://doi.org/10.1038/nclimate1744

Tiyasha, Tung, T.M. and Yaseen, Z.M. (2020). A survey on river water quality modelling using artificial intelligence models. *Journal of Hydrology*, 585: 124670. https://doi.org/10.1016/j.jhydrol.2020.124670

Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-2440-0

Weber, K.J. (1971). Sedimentological aspects of oil fields in the Niger Delta. *Geologie en Mijnbouw*, 50(3): 559–576.

World Health organization (2015). Guidelines for drinking –water quality, p104-108