VOL. 09 NO. 4,
AUGUST, 2025
E-ISSN 3026-9210
P-ISSN 3026-815X

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Built Env. & Geological Research (JBEGR)

Assessing the Impact of Climate Change on Volume of Water Consumed by Poor Urban Households Using Geospatial Technology and Machine Learning Models

Taiwo, Tolu A.; Olusina, J. O.; Hamid-Mosaku, I. A.; Abiodun, O. E.; & Ayodele, E. G.

Department of Surveying and Geoinformatics, Faculty of Engineering, University of Lagos, Nigeria.

Corresponding Author: tolutaiwo75@gmail.com

DOI: https://doi.org/10.70382/ajbegr.v9i4.025

Abstract

The United Nations declares that climate change is primarily a water crisis. The impact is felt more by poor urban households who have to spend more time and scarce resources to access clean water. This study assesses the impact of climate change temperature variability on volume of water consumed by poor urban households. Spatial and seasonal variations of land surface temperature as a result of climate change was mapped in relation to the location of poor urban households using geospatial technology. The mapped pattern was fed into four machine learning models for modeling and prediction. The machine learning models predicted a difference of 0.1 liter between volume of water consumed by poor urban households with and without climate change data. Significance test shows that there is no significant difference in model performance when climate data was included and when climate data was not included in the modelling. In addition, weak correlation of 0.04 was found between LST and volume of water consumed, which conforms with the fact that there is no significant difference in volume of water consumed with or without temperature variability. That is, climate change temperature variability does not influence the volume of water consumed by poor urban households in the study area.

Keywords: Climate change, Water consumption, Poor households, Geospatial technology, Machine learning models, Land surface temperature, Nigeria, Household water demand.

Introduction

Background

The rise in greenhouse gases due to anthropogenic activities is increasing global temperatures. Variations in global temperatures influence the hydrological cycle and water consumption in households (Punia et al., 2022). Many studies have examined impacts of climate change on water supply (Ayanlade, 2023), water availability (Punia et al., 2022), renewable energy (Elum, 2017), household income and food security (Zakari et al., 2022). Influence of climate change on volume of water consumed by poor urban households is yet to be explored, though there have been evidences suggesting that water usage is a function of weather variability (Adamowski et al., 2012, Breyer et al., 2012, Guhathakurta and Gober, 2007). In addition, the influence of the climatic factors from that of the socio-economic factors have not been distinguished. Rather urban water consumption was distinguished into base water consumption (BWC); and seasonal water consumption (SWC) (Eslamian et al., 2016; Rasifaghihi et al., 2020). BWC was taken as indoor water use, while SWC was seen as outdoor water use. BWC was thought to be insensitive to climate effects, but sensitive to socio-economic factors. SWC was considered as dependent on changes in air temperature and precipitation. In most of the studies, the forecast results are site-specific; that is, they reflect the conditions of the study area.

Climate Change and Water Consumption

The United Nations declares that climate change is primarily a water crisis. Its impacts are seen in worsening floods, rising sea levels, and droughts. But water can fight climate change. Sustainable water management can help to build the resilience of societies and ecosystems and reduce carbon emissions (UN Water, 2025). Most impacts of climate change come down to water, in form of water-related hazards such as floods and droughts, as rising temperatures disrupt the entire water cycle (UN, 2025). Climate change is increasing water scarcity and making it even harder for poor people to get clean water. According to WaterAid (2021), people need reliable water supply that keeps pumping through flood, drought and natural disasters in order to meet their daily livelihood. It is estimated that half the world's population (~4 billion out of ~8 billion people) are currently subject to severe water scarcity due to climatic and non-climate factors. It is also estimated that limiting global warming to 1.5°C would minimize the increase in risks in the various water use sectors and keep adaptation effective (IPCC, 2022).

Understanding influence of climate change on water consumption will help managers and planners develop strategies to reduce water consumption (Dias and Ghisi, 2024). Qin et al., (2022) utilized a novel dataset where daily water usage from over 40,000 households in ten China provinces were collected over the years 2010–2019 using smart meters, to investigate how climate variability affected household water demand. They found increasing marginal impacts of water usage with respect to temperature intervals. Compared to a below 5°C day, water usage increased by around 16 *l*/household-day at the 5–10 °C interval to 45 *l*/household-day at the highest interval of above 30 °C. Gosling and Arnell (2013)

evaluated water consumption impact due to climate change. The study was based on climate change patterns from 21 global climate models in four scenarios, which were applied to a global hydrological model to estimate water resources in 1339 river basins. As a result, the authors found that between 1.6 and 2.4 billion people currently live in river basins exposed to water scarcity. The climate change effect by 2050 will place 0.5 to 3.1 billion people in situations of water scarcity. Zapata (2021) states that bottled water consumption is associated with an increased temperature. The author points out that the use of this resource is influenced by factors related to water availability (per capita water consumption and piped water availability) and the perception of the quality of water received by the public supply system. Regarding climatic conditions, the author states that a 1 °C increase in the average temperature will result in an increase of approximately one fifth in bottled water consumption. Chang, et al., (2014) discovered that the correlation between temperature and water consumption is consistently higher than that between precipitation and water consumption. Qin et al., (2022) found that households in China are using increasingly more water to cope with high-temperature days. Moreover, the studies utilized air temperature as one of the climatic factors instead of land surface temperature.

Land surface temperature (LST) is a measure of how hot the land is to the touch. Air temperature (AT), on the other hand, is the temperature of the environment above the ground surface. The difference in LST between urban and non-urban areas is called surface urban heat island. LST is used for effective climate modelling and climate-sensitive urban planning (Chayapong & Dasananda, 2012; Gago et al., 2013). Scientists monitor LST because the warmth radiating from Earth's landscapes influences, and is influenced by, global weather and climate patterns. LST has both spatial and temporal extent (Saleh et al., 2023) and urbanization usually increases its intensity (Peterson, 2003; Gaffin et al., 2008).

LST used in this study has only spatial dimension since it is mapped in relation to the location of urban poor households. Moreover, it also has seasonal variation since it is measured in both dry season and wet season. Thus, this study utilizes LST instead of AT to study the impact of climate change on volume of water consumed in poor urban households. The aim is to assess the impact of climate change temperature variability on volume of water consumed by poor urban households. Spatial and seasonal variations of LST in the study area as a result of climate change is mapped in relation to the location of urban poor households using geospatial technologies. The mapped pattern is fed into machine learning algorithms for modeling and prediction.

Geospatial Technology and Machine Learning Modeling

Geospatial technology (GST) is a collective term for the various modern technologies that help in collection, analysis and interpretation of geospatial data (Muppudathi, 2014; Reed & Ritz, 2003). GST includes remote sensing (RS), Global Positioning System (GPS) and Geospatial Information System (GIS). GST, which represents the physical world in a virtual environment, is used in modelling, simulation and visualization of geospatial information, which can inform both present and future decisions. GPS and GIS systems are used in this

study to collect and process geospatial data respectively. The collected and processed data are then integrated with machine leaning algorithms to model water consumption by the urban poor.

A machine learning model is an algorithm for regression and/or classification. Four machine learning regression models (Multilinear Regression, Random Forest, Support Vector Regression, and Artificial Neural Network) were chosen for this work. The four models were chosen because they represent models in the supervised learning category for regression analysis, and it has been shown that Neural Network-based techniques outperform conventional techniques and provide effective solutions for many geospatial data analysis tasks (Kiwelekar et al., 2020).

Multilinear Regression (MLR)

Linear regression is a technique for investigating the relationship between an independent variable or feature and a dependent variable or outcome. Linear regression model describes the relationship between a dependent variable $\boldsymbol{y}$ and independent variables $\boldsymbol{x}$ with a straight line that is defined by equation 1:

$$\boldsymbol{y} = w_o + w_1\boldsymbol{x} \quad (1)$$

In this expression, $\boldsymbol{y}$ is the vector of the response values. The $\boldsymbol{x}$ symbol describes the matrix of features which the algorithm uses to predict the $\boldsymbol{y}$ vector. $\boldsymbol{x}$ is a matrix that contains only numeric values. w_o and w_1 are parameters that the linear regression uses to create the prediction. Features are real-world entities that are being modelled. The ordinary least squares (OLS) method is used to estimate w_o and w_1. The OLS method seeks to minimize the sum of the squared residuals. The distance from each data point to the regression line is calculated, squared, and all squared errors are summed together. As in most linear regression models, the objective is to minimize the sum of squared errors, that is,

$$\text{Min} \sum_{i=1}^{m} y_i^p - y_i^o,)^2 \quad (2)$$

In a multilinear regression (MLR), where there are multiple features, the space now spans multiple dimensions, with each dimension being a feature. For instance, for five features the space is five-dimensional and the regression equation in matrix form becomes equation 3.

$$y^p = w_1\boldsymbol{x_1} + w_2\boldsymbol{x_2} + w_3\boldsymbol{x_3} + w_4\boldsymbol{x_4} + w_5\boldsymbol{x_5} + w_o \quad (3)$$

MLR trains learning algorithms using labelled training data to understand the relationship between many features and an outcome. The result is a trained model. The trained model can then be leveraged to predict the outcome of new input data that were not seen before.

Random Forest (RF)

RF is an ensemble technique capable of performing both regression and classification tasks. Being an ensemble means that RF combine multiple decision trees through a technique called Bootstrap and Aggregation, commonly known as bagging**.** The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on

individual decision trees. RFs combine large number of decision tree models built on different sets of bootstrapped examples. Thus, the output does not depend on one decision tree but on multiple decision trees. RF has multiple decision trees as base learning models, and so we randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called bootstrap. In the case of a regression problem, the final output is the mean of all the outputs. This part is called aggregation.

According to Cutler et al., (2011), for a p-dimensional random vector $X = (X_1, \ldots, X_p)^T$ representing predictor variables and a random variable Y representing the real-valued target variable, an unknown joint distribution $P_{XY}(X, Y)$ is assumed. The goal is to find a prediction function $f(X)$ for predicting Y. The prediction function is determined by a loss function $L(Y, f(X))$ and defined to minimize the expected value of the loss

$$E_{XY}\,(L(Y, f(X))) \tag{4}$$

where the subscripts denote expectation with respect to the joint distribution of X and Y.

Intuitively, $L(Y, f(X))$ is a measure of how close $f(X)$ is to Y; it penalizes values of $f(X)$ that are a far away from Y. Typical choices of L are squared error loss $L(Y, f(X)) = (Y - f(X))^2$ for regression.

It turns out that minimizing $E_{XY}\,(L(Y, f(X)))$ for squared error loss gives the conditional expectation

$$f(x) = E(Y|X = x) \tag{5}$$

otherwise known as the regression function.

Support Vector Regression (SVR)

SVR is a regression algorithm that finds a hyperplane that best fits data points in a continuous space while minimizing the prediction error. This is achieved by mapping the input variables to a high-dimensional feature space and finding the hyperplane that maximizes the margin (distance) between the hyperplane and the closest data points, while also minimizing the prediction error. In contrast to OLS, the objective function of SVR is to minimize the coefficients, not the squared error. The error term is instead handled in the constraints, where we set the absolute error less than or equal to a specified margin, called the maximum error, ϵ (epsilon). We can tune epsilon to gain the desired accuracy of our model. Our new objective function and constraints are as shown in equations6 and 7.

$$\text{Min}\,\frac{1}{2}\|w\|^2 \tag{6}$$

$$s.t.\,|y_i - w_i x_i| \leq \varepsilon \tag{7}$$

The algorithm solves the objective function as best as possible but some of the points still fall outside the margins. As such, we need to account for the possibility of errors that are larger than ϵ. We can do this with slack variables. The concept of slack variables is simple: for any value that falls outside of ϵ, we can denote its deviation from the margin as ξ. We know that these deviations have the potential to exist, but we would still like to minimize them as much

as possible. Thus, we can add these deviations to the objective function, which becomes equations 8 and 9.

$$\text{Min}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}|\xi_i| \quad (8)$$

$$s.t.|y_i - w_i x_i| \leq \varepsilon + |\xi_i| \quad (9)$$

C and ξ (error penalty and slack variables respectively) are employed to prelude the influence of outliers and avoid overfitting.

SVR can handle non-linear relationships between the input variables and the target variable by using a kernel function to map the data to a higher-dimensional space, and it works with smaller amount of training samples and variables, and remain highly sensitive to variations in the variables (Karimi et al., 2016). This makes it a powerful tool for regression tasks where there may be complex relationships between the input variables and the target variable. A kernel is a set of mathematical functions that finds a hyperplane in the higher dimensional space that fits the data without increasing computational cost. The most widely used kernels include Linear**,** Non-Linear**,** Polynomial**,** Radial Basis Function (RBF) and Sigmoid. By default, RBF is used as the kernel.

Artificial Neural Network (ANN)

Recent advances in the field of deep-learning showed that Neural Network-based techniques outperform conventional techniques and provide effective solutions for many geospatial data analysis tasks (Kiwelekar, 2020). Multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP is a deep learning method that uses back-propagation algorithm for training the network. The input layer consists of the inputs to the network, followed by a hidden layer consisting of neurons or hidden units, and then an output layer representing classes or patterns. The input pattern is presented to the network via the input layer, and the input signals are passed to the nodes in the next layer in a feed-forward manner. The summation of the output is called the output layer. MLPs are controlled by setting and adjusting weights between nodes. Initial weights are usually set at some random numbers and then they are adjusted during training.

An activation function, $g(x)$, defines the output of a neuron in terms of the linear combination of inputs. Activation functions provide nonlinearity to the output because it enhances or damps the values passing through it in a non-proportional way. Activation function also controls how well the network model learns the training dataset, and also defines the type of predictions the model can make. Hidden units can be described as accepting a vector of inputs x, computing an affine transformation of equation 10,

$$z = W^T x + b, \quad (10)$$

and then applying an element-wise nonlinear function $g(z)$. Most hidden units are distinguished from each other only by the choice of the form of the activation function $g(z)$. There are different kinds of activation functions: rectified linear unit (ReLU), tanh function, and the logistic (sigmoid) function. The most commonly used in geospatial data analysis and water resources is the logistic activation function and, recently, ReLU.

Rectified linear unit (ReLU) is an excellent default choice of activation function (Goodfellow et al., 2016) because a model that uses it is easier to train and often achieves better performance. ReLU is a simple function that returns the value of the input if it is positive, or value 0 if the input is 0 or negative. The function is linear for values greater than zero, meaning it has a lot of the desirable properties of a linear activation function when training a neural networks using back propagation. Yet, it is a nonlinear function as negative values are always output as zero. ReLU uses the maximum shown in equation 11 as the activation function.

$$g(z) = \max\{0, z\}. \tag{11}$$

Cost function (or loss function) helps to fit parameters to data, as it gives the value of the parameters that fit the data well when it is optimized. It is the penalty paid by the learning algorithm for fitting the model. That is, cost function gives the cost of predicting outcomes after fitting the model. Minimizing the cost function, which is also the optimization objective, the learning algorithm's prediction becomes more accurate, and the algorithm better maps the function to the features. According to Ng (2013), cost function can be defined by equation 12.

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}(y^{p(i)} - y^{o(i)})^2 \tag{12}$$

that measures, for each value of the w's, how close the $y^{p(i)}$'s are to the corresponding $y^{o(i)}$'s. Feature $x^{(i)}$ and outcome $y^{(i)}$ represent the *i*th example in the training set, and m is the number of examples.

Implementing the Models

Four machine learning techniques, Multilinear Regression (MLR), Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Network (ANN), were employed for the modelling. The models were implemented with codes in JupyterLab® environment where the experiments were carried out.

Materials and Methods

Study Area

Nyanya-Mararaba is a border town between Federal Capital Territory (FCT) and Nasarawa State in Nigeria. It lies between Latitudes 8°24'54''N and 9°18'48''N, and Longitude 6°44'25''E and 7°35'15''E. The town's climate is largely sunny with diurnal temperatures ranging from 18.45 to 36.05 degrees Celsius. Rainfall varies from 0.0 to 400.0 mm/month. The climate is due to the town's location at the transition zone between the 'humid' south and the 'sub-humid' north. Without government presence, Nyanya-Mararaba Town is fast turning

into a slum where access to clean water is difficult because there is no water distribution network (WDN). Though there is no WDN in Nyanya-Mararaba Town but there is WDN in Karu, a town across the Abuja-Keffi Expressway, which demarcates the two towns as illustrated in Figure 1. Thus, the challenge is to accurately estimate volume of water consumed by poor households in Nyanya-Mararaba Town in order to extend piped water network to them.

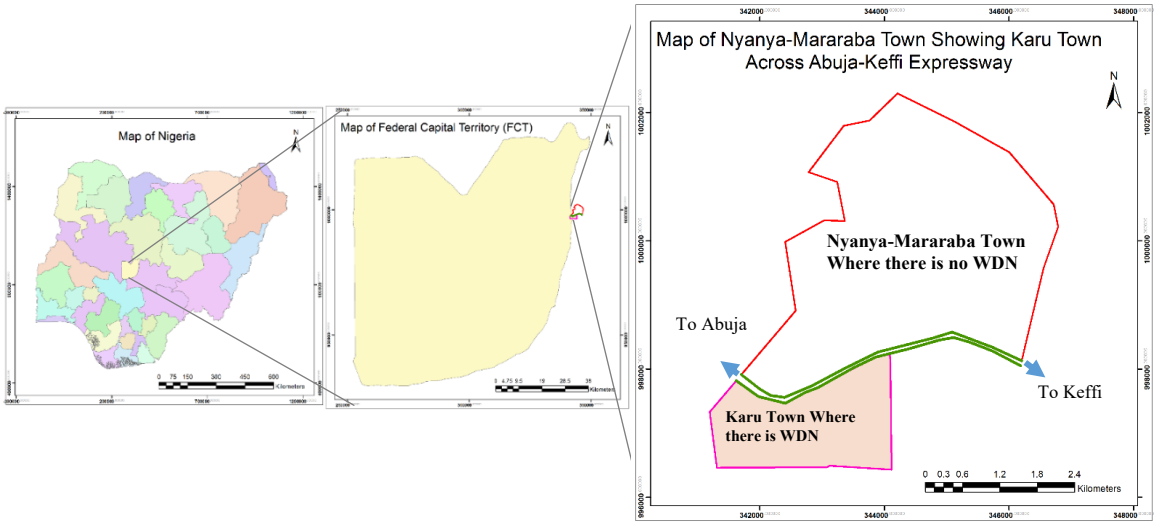


Figure 1: Map of Nigeria and Abuja, showing Nyanya-Mararaba Town

The growing population of the Federal Capital City is concentrated at Nyanya-Mararaba town where an estimated 70% of the people working in the city live. Covering an area of 14 square kilometers, Nyanya-Mararaba is a peri-urban town consisting of informal dwellings where many poor people find shelter.

Datasets

Nyanya-Mararaba Dataset

This study use Nyanya-Mararaba datasets described in Taiwo et. al (2023). The dataset is described in Table 1.

Table 1: Data and materials used in the study

S/N	Dataset	Format	Sources	Application
1.	Volume of water consumed from water meter readings in Karu Town (called Karu Dataset)	Numerical and categorical data in Excel csv file	FCT Water Board	Applied as input for modeling water consumption where there is Water Distribution Network (WDN)
2.	Socio-economic, property and historic data collected in Nyanya-Mararaba Town	Numerical and categorical data in Excel csv file	Questionnaires	Input for modeling water consumption where there is no Water Distribution Network (WDN)

S/N	Dataset	Format	Sources	Application
	(called Nyanya-Mararaba Dataset)			
3.	Coordinates of household locations in Nyanya-Mararaba Town	Numerical data with Easting and Northing values	GPS Survey	Applied in GIS to create point features of household locations
4.	Coordinates of water points in Nyanya-Mararaba Town	Numerical data with Easting and Northing values	GPS Survey	Applied in GIS to create point features of water points locations
5.	Rainfall, air temperature and wind speed values in the year 2023	Numerical data in Excel csv file	NiMET	Input for modeling influence of climate change on water consumption
6.	Land Surface Temperature (LST)	Imagery in TIFF format	RSLab at https://rslab.gr/Landsat_LST.html	Input for modeling influence of climate change on water consumption
7.	Digital elevation model (DEM) imagery	Imagery in TIFF format SRTM 1-arc second 30meter horizontal and 16meter vertical resolution	United States Geological Survey downloaded from https://earthexplorer.usgs.gov/	Applied in GIS to create DEM of study area and calculate height difference between every household and the nearest water point.
8.	Land Use Land Cover (LULC) imagery	Imagery in TIFF format. 10-meter resolution land cover	ESRI downloaded from https://maps.ArcMap.com	Applied in GIS to create land use land cover types of the study area

Source: Taiwo et al., (2023)

Control Datasets

Volume of water consumed per day for dry and wet seasons, represented by September to February and March to August respectively, were extracted from the questionnaire for each respondent. A dataset consisting of 1200 records was derived for each season of the year, resulting in two datasets, one for dry season and another for wet season. The two are the control datasets for this study.

Climate Change Datasets

The datasets used for climate change experiment is same as the complete dataset described above, the only difference being the addition of land surface temperature. This was included to account for climate variability within the poor urban environment. Figure 2 shows distribution of household water consumption and average LST in the climate datasets.

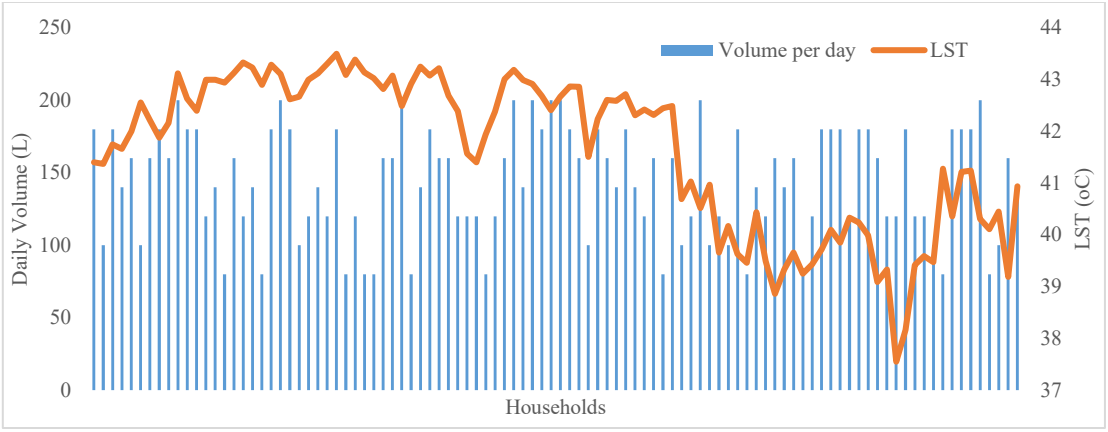


Figure 2: Distribution of household water consumption and average LST in the climate datasets

LST Data

LST for dry and wet seasons for each household location was extracted from LST images in ArcMap® using the Extract Values to Points tool among the Extraction tools under the Spatial Analysis toolbox in the ArcToolbox. This is shown in Figure 3.

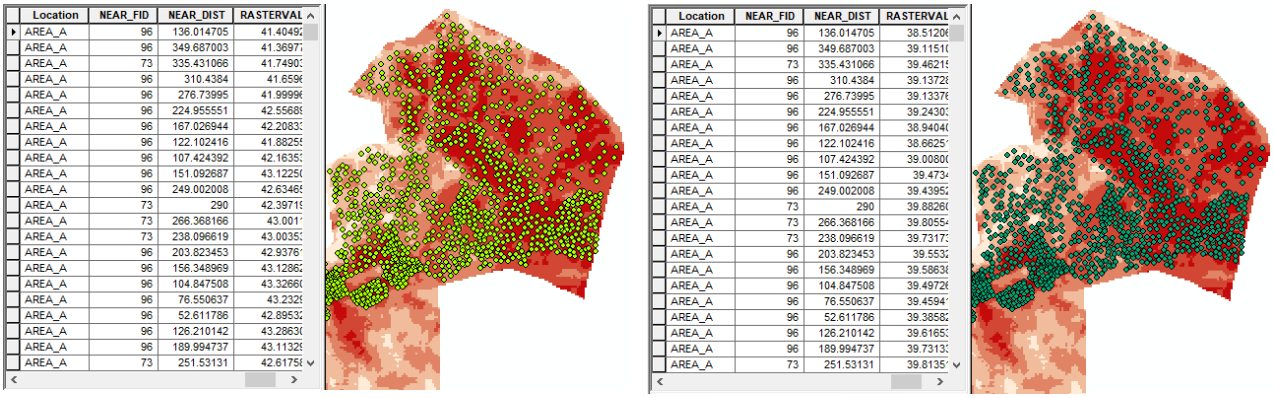


Figure 3: LST for dry and wet seasons in Nyanya-Mararaba Town in ArcMap®

Experimental Design

The MLR, RF, SVR and ANN models were coded with Python in JupyterLab®. They were each created as an instance of their respective classes in Scikit Learn® library: *SGDRegressor, RandomForestRegressor, LinearSVR and MLPRegressor*, which were each infused with a scaler to scale the dataset (Pedregosa et al., 2012). Thus, we had an empty model. The model was then trained or *fitted* with the training dataset. Training enables the model to learn the relationship between the explanatory variables and the target variable. After training, the model was ready for predictions. The results were evaluated. If the resulting errors were not within tolerance ($0.70 \leq R^2 \geq 0.99$), then the model needed refining through parameter tuning

and it was built again. If the errors were within tolerance, then the model should be built with the parameters obtained. Afterward, the model was validated and tested.

Training, Validation and Testing

Training a model is a process of fitting it with a dataset so that the learning algorithm can learn the data and the relationships among the features. Training prepares the model for and gives it ability to generalize what it has learnt to new dataset it has not "seen" before. This allows the model to make prediction based on what it has learnt. A major problem with training is called overfitting, a situation in which the model learns well the training dataset but unable to generalize and make good predictions with new dataset. To avoid overfitting, the dataset was split into three: training data, validation data and test data. After training, validation dataset was used to assess the model performance. If satisfied, then the model was tested with the test dataset, which acts as a new data that the model had never "seen" before. The performance of the model when fed with the test dataset shows how it will perform when it encounters new data in the world. Standard split ratio found in literature are 60-20-20 percent, 70-15-15 percent, and 80-10-10 percent. 80-10-10 was adopted for this study.

Model Performance Metrics

Three error metrics were used to measure the performance of the model: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R square (R^2). MAE is the average of the absolute difference between two continuous variables. For our purposes, we used the mean absolute error, expressed in equation 13, to measure the absolute difference between observed and predicted values of the target variable.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|Y_i^o - Y_i^p| \quad (13)$$

Mean Square Error (MSE) more greatly penalizes larger errors and is sensitive to outliers due to taking the squared difference between variables (Mueller, 2021). Also, MSE is good for values close to zero but not for large values. We want both large and small errors to be penalized equally so that we could obtain a robust model that would predict well. Therefore, we used root mean squared error, which is the square root of the value obtained from MSE.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(Y_i^o - Y_i^p)^2} \quad (14)$$

R^2 score is the amount of the variation in the target variable which is predictable from the explanatory variables as mathematically shown in equation 15. It is used to check how well observed results are reproduced by the model. Best possible score is 1.0, but the scores are stated as percentages in this study. A higher value for R^2 is desirable as it indicates better results.

$$R^2 = \left[\frac{\sum_{i=1}^{m}\left(Y_i^o - \bar{Y}_i^0\right)\left(Y_i^p - \bar{Y}_i^p\right)}{\sqrt{\sum_{i=1}^{m}\left(Y_i^o - \bar{Y}_i^0\right)^2 . \sum_{i=1}^{m}\left(Y_i^p - \bar{Y}_i^p\right)^2}}\right]^2 \quad (15)$$

In the three expressions above, Y_i^o, Y_i^p, $\bar{Y}_i^o$, and $\bar{Y}_i^p$ are observed and predicted values and their respective average values; m is the number of observations or examples. The three error metrics, MAE, RMSE and R^2 score, were hard-coded in Jupyter Notebook®, and the model performance was measured with the three error metrics each time the trained model made a prediction.

Results and Discussion

Performance of Machine Learning Models with and without Climate Change Data

The wet and dry seasons datasets which are the datasets without climate data serves as the control dataset. whereas the datasets with climate data serves as the complete datasets. Each of the models was trained, validated and tested with the complete dataset and the control data. The results are tabulated in Table 2 to Table 7.

Table 2: Model performances with and without climate data during wet season TRAINING

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	2.58	4.04	0.87	2.44	3.83	0.88
RF	0.15	0.97	1.00	0.15	0.89	1.00
SVR	1.89	5.24	0.82	1.83	4.83	0.84
ANN	0.58	1.35	0.99	0.44	1.12	0.99

Table 3: Model performances with and without climate data during wet season VALIDATION

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	2.88	4.18	0.85	3.17	5.11	0.75
RF	0.07	0.23	1.00	0.27	1.74	0.97
SVR	2.78	6.27	0.75	2.98	7.53	0.63
ANN	0.56	0.95	1.00	0.62	1.74	0.97

Table 4: Model performances with and without climate data during wet season TESTING

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	2.58	3.68	0.88	2.61	4.41	0.82
RF	0.30	1.91	0.97	0.04	0.09	1.00
SVR	2.35	5.11	0.84	2.66	6.48	0.77
ANN	0.64	1.20	0.99	0.42	0.74	0.97

Table 5: Model performances with and without climate data during dry season TRAINING

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	2.56	3.96	0.87	2.63	4.03	0.87
RF	0.16	0.89	1.00	0.16	1.00	1.00
SVR	1.97	5.24	0.82	1.96	5.33	0.81
ANN	0.62	1.50	0.99	1.18	2.11	0.97

Table 6: Model performances with and without climate data during dry season VALIDATION

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	2.75	4.28	0.84	2.66	3.53	0.88
RF	0.06	0.25	1.00	0.33	1.33	0.99
SVR	2.37	6.02	0.77	2.12	4.40	0.88
ANN	0.62	1.00	1.00	1.39	2.50	0.94

Table 7: Model performances with and without climate data during dry season TESTING

Model	With Climate Data			Without Climate Data		
	MAE (L)	RMSE (L)	R^2	MAE (L)	RMSE (L)	R^2
MLR	3.01	4.18	0.83	2.74	4.23	0.83
RF	0.28	1.80	0.97	0.14	0.68	1.00
SVR	2.05	5.25	0.80	2.42	6.19	0.74
ANN	0.53	0.96	1.00	1.06	1.76	0.97

Analyzing Table 1 to Table 6 shows that RF and ANN perform better than other models. Moreover, generally speaking, all the models perform better with climate data in both dry and wet seasons than without climate data. For example, with climate data, in wet season, during validation, MLR, RF, SVR, ANN have R^2 of 0.85, 1.00, 0.75, 1.00 respectively, whereas without climate data they have R^2 of 0.75, 0.95, 0.63, 0.95 respectively. In dry season, with

climate data, during validation MLR, RF, SVR, ANN have R^2 of 0.84, 1.00, 0.77, 1.00, whereas, without climate data, they have R^2 of 0.88, 0.99, 0.88, 0.94 respectively.

Significance Test on Difference in Model Performances with Climate Data and without Climate Data

The null hypothesis states that the difference between model performances when climate data was included and when climate data was not included is not significant.

Null hypotheses: $H_0 : \mu = \mu_0$, that is, there is no significant difference in model performance when climate data was included and when climate data was not included.

Alternative hypothesis: $H_a : \mu \neq \mu_0$, that is, there is significant difference in model performance when climate data was included and when climate data was not included.

Table 8: Two-tailed t-test at 95% confidence level

Model	*p* value	Degrees of Freedom	Test Result
MLR	0.25865	8	$H_0 : \mu = \mu_0$
RF	0.8632	8	$H_0 : \mu = \mu_0$
SVR	0.2172	8	$H_a : \mu = \mu_0$
ANN	0.8154	8	$H_0 : \mu = \mu_0$

The *p* value of each model is stated in Table 8. Since the *p* value is more than 0.05 for all the models, the test failed to reject the null hypothesis; that is, there is no significant difference in model performance when climate data was included and when climate data was not included. That is, climate temperature variability does not influence the volume of water consumed by the poor urban households in the study area.

Volume of Water Consumed in Dry and Wet Seasons with Climate Change Data

The experiment carried out here assesses the actual and predicted volume of water consumed in relation to climate variability in poor urban households. When the climate dataset was imputed in each of the model, the actual and predicted volume of water consumed are listed in Table 9. RF produced the least residuals in both dry and wet seasons. Average volume of water consumed in the households in the study area in a day in dry season is 147.35 liters, while the average consumed in wet season is 167.35 liters. Again, poor households in the study area consume more water during wet season than dry season.

Table 9: Total volume of water consumed in dry and wet seasons with climate change data

	Dry Season			Wet Season		
Model	**Actual Volume (L)**	**Predicted Volume (L)**	**Residuals (L)**	**Actual Volume (L)**	**Predicted Volume (L)**	**Residuals (L)**
MLR	147.35	147.46	0.11	167.35	167.05	0.30
RF	147.35	147.36	0.01	167.35	167.35	0.00

	Dry Season			Wet Season		
SVR	147.35	143.64	3.71	167.35	163.64	3.71
ANN	147.35	147.50	0.15	167.35	167.34	0.01

Without climate change dataset, the RF model predicted 147.36 liters as the average volume of water consumed in each household in dry season, while the average volume of water consumed in the wet season is predicted as 167.35 liters. When climate change dataset was added to the features, the RF model predicted 147.36 liters as the total volume of water consumed in dry season, while it predicted 167.35 liters as the average volume of water consumed in the wet season. That is, at household level, difference in volume of water consumed in relation to climate change is negligible.

Table 10: Effect of climate change on volume of water consumed in dry and wet seasons

Model	Dry season with climate change (L)	Dry season without climate change (L)	Difference (L)	Wet season with climate change (L)	Wet season without climate change (L)	Difference (L)
MLR	147.46	147.30	0.16	167.05	167.36	-0.31
RF	147.36	147.35	0.02	167.35	167.35	0.00
SVR	143.64	143.65	-0.01	163.64	163.64	0.00
ANN	147.50	147.34	0.15	167.34	167.51	-0.17

Table 10 shows volume of water consumed as a result of climate variability in the study area. The machine learning models predicted a difference of 0.1 liter between volume of water consumed with and without climate change data. That is, the difference in water consumption as a result of climate change in the study area is negligible. Contrasting findings in literature suggest that volume of water consumed in households increases with increasing climate temperature change (Chang et al., 2014; Mardy et al., 2015; Qin et al., 2022). There are possible reasons for this contrast. First, those studies were not done among poor urban households, which suggests that poor households' challenges in accessing clean water are different. For example, Taiwo et al., (2023) found that poor households in Nigeria's Federal Capital Territory pay ten times more for drinking water because they are not linked with WDN. In addition, poor urban people spend more than 30-minute travel time to access clean water. These socio-economic challenges could explain why volume of water consumed by poor urban households does not change with climate variability.

This study found a very weak correlation (r = 0.04) between LST and volume of water consumed by the poor urban households in the study area. The very weak correlation conforms with the fact that there is no significant difference in the prediction of volume of water consumed in the study area with or without climate variability. Our results are in line with the findings of de Souza (2015) in which weak correlations between the variables were

discovered. A possible explanation is based on seasonality, with wet and dry seasons in the study area, and low temperature variability.

Conclusion

This study utilizes land surface temperature (LST) to study the impact of climate change on volume of water consumed in poor urban households. The aim is to assess the impact of climate change temperature variability on volume of water consumed by poor urban households. Spatial and seasonal variations of LST in the study area as a result of climate change was mapped in relation to the location of poor urban households using geospatial technologies. The mapped pattern was fed into four machine learning models for modeling and prediction. Significance test shows that there is no significant difference in model performance when climate data was included and when climate data was not included in the modelling.

The machine learning models predicted a difference of 0.1 liter between volume of water consumed with and without climate change data. That is, the difference in water consumption as a result of climate change in the study area is negligible. In addition, weak correlation of 0.04 was found between LST and volume of water consumed, which conforms with the fact that there is no significant difference in volume of water consumed in the study area with or without climate variability. That is, climate temperature variability does not influence the volume of water consumed by the poor urban households in the study area.

Impact of climate change variability on the volume of water consumed by urban poor households has been studied using the combination of geospatial technology and machine learning techniques with an integration of geospatial and non-geospatial data. The finding that climate change variability has negligible impact on the volume of water consumed by urban poor households has been highlighted, and should be a subject for consideration by policy makers. However, the study is limited to the impact of land surface temperature while other climate variables such as surface air temperature, precipitation, humidity and wind speed were not considered. Hence, further studies should be carried out among the urban poor in other cities investigating the impact of other climate variables or a combination of all climate variables on volume of water consumed.

References

Adamowski J. *et al. (2012).* Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res.

Ayanlade, A. (2023). Safe drinking water supply under extreme climate events: evidence from four urban sprawl communities. *Climate and Development*, *16*(7), 563–578. https://doi.org/10.1080/17565529.2023.2264270

Breyer Betsy, Chang Heejun, Parandvash G. Hossein, (2012). Land-use, temperature, and single family residential water use patterns in Portland, Oregon and Phoenix, Arizona, Applied Geography, Volume 35, Issues 1–2, 2012, Pages 142-151, ISSN 0143-6228, https://doi.org/10.1016/j.apgeog.2012.06.012.

Chang, Heejun; Praskievicz, Sarah; and Parandvash, Hossein (2014) "Sensitivity of Urban Water Consumption to Weather and Climate Variability at Multiple Temporal Scales: The Case of Portland, Oregon," International Journal of Geospatial and Environmental Research: Vol. 1: No. 1, Article 7. Available at: https://dc.uwm.edu/ijger/vol1/iss1/7

Chayapong, P., & Dasananda, S. (2012). Analysis of urban heat island phenomenon and its relationships with land use/land cover characteristics: Case study in Bangkok metropolitan administration area. 33rd asian conference on remote sensing 2012, Vol. 2, pp. 989–995

Cutler Adele, Cutler David Richard, Stephens John R. (2011). Random Forest. In the book "Esseble Machine Learning: Methods and Applications. 45(1):157-176. Eds: Cha Zhang and Yungian Ma. DOI:10.1007/978-1-4419-9326-7_5

Dias, T. F., & Ghisi, E. (2024). Urban Water Consumption: A Systematic Literature Review. *Water*, *16*(6), 838. https://doi.org/10.3390/w16060838

de Souza, A., Aristone, F., Sabbah, I., da Silva Santos, D. A., de Souza Lima, A. P., & Lima, G. (2015). Climatic variations and consumption of urban water. Atmospheric and Climate Sciences, 5(3), 292-301.

Elum Z.A., Momodu A.S. (2017). Climate change mitigation and renewable energy for sustainable development in Nigeria: A discourse approach, Renewable and Sustainable Energy Reviews, Volume 76, 2017, Pages 72-80, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2017.03.040.

Eslamian S. Alireza, Li S. Samuel, Haghighat Fariborz, (2016), A new multiple regression model for predictions of urban water use, Sustainable Cities and Society, Volume 27, 2016, Pages 419-429, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2016.08.003.

Gaffin S. R., Rosenzweig C., Khanbilvardi R., Parshall L., Mahani S. (2008). Variations in New York city's urban heat island strength over time and space. Theor Appl Climatol 2008, 94:1–11. DOI: 10.1007/s00704-007- 0368-3.

Goodfellow I., Bengio Y., Courville A. (2016). *Deep Learning*. www.deeplearning.org

Gosling, S.N.; Arnell, N.W. (2013). A global assessment of the impact of climate change on water scarcity. *Clim. Change* **2013**, *134*, 371–385.

Guhathakurta S. and Gober, (2007). The impact of the Phoenix urban heat island on residential water use J. Am. Plan. Assoc.

IPCC (2022). Impact, Vulnerability and Adaptability. https://www.ipcc.ch/report/ar6/wg2/ Assessed on 29th August 2025

Karimi, S., Shiri, J., Kisi, O., & Shiri, A.A. (2016). Short-term and long-term streamflow prediction by using'wavelet–gene expression'programming approach. *ISH Journal of Hydraulic Engineering, 22 (2)*, 148-162.

Kiwelekar, A. W., Mahamunkar, G. S., Netak, L. D., & Nikam, V. B. (2020). Deep learning techniques for geospatial data analysis. *Machine Learning Paradigms: Advances in Deep Learning-based Technological Applications*, 63-81.

Mahdy Jarboo, Husam Al-Najar, (2015). Climate Change and its impact on domestic water consumption in Sub-urban regions in the Gaza Strip. International Journal of Climate Change Strategis and Management 16 March 2015; 7(1): 3 – 16. https://doi.org/10.1108/IJCCSM-072013-0092

Mueller, J. P., & Massaron, L. (2021). *Machine Learning for Dummies*® John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, Canada.

Muppudathi, G. (2014). GEO-SPATIAL TECHNOLOGIES IN TEACHING AND LEARNING. *Academia*, Page No: 145-150.

Ng, A. (2013). *Lecture Notes. Stanford University*. Retrieved from https://www.datasciencecentral.com/lecture-notes-by-ng-full-set/ Accessed on 24th June, 2025.

Peterson T. C. (2003). Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. J Clim 2003, 16:2941–2959. DOI:10.1175/1520-0442 (2003)0162.0.CO;2.

Punia Anita, Singh Saurabh Kumar, Bharti Rishikesh (2022). Effect of climate change on urban water availability and its remediation in different continents, Editor(s): Arun Lal Srivastav, Sughosh Madhav, Abhishek Kumar Bhardwaj, Eugenia Valsami-Jones, Current Directions in Water Scarcity Research, Elsevier, Volume 6, 2022, Pages 45-63, ISSN 2542-7946, ISBN 9780323918381, https://doi.org/10.1016/B978-0-323-91838-1.00002-6 .

Qin Ping, Chen Shuai, Tan-Soo, Jie-Sheng, Xiao-Bing Zhang (2022). Urban household water usage in adaptation to climate change: Evidence from China. Environmental Science and Policy 136 (2022) 486–496

Rasifaghihi N., Li S.S., Haghighat F., Forecast of urban water consumption under the impact of climate change (2020). Sustainable Cities and Society, Volume 52, 2020, 101848, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2019.101848

Reed, P., & Ritz, J. (2003). Geospatial Technologies. *The Technologies Teacher, 63*, 17-20.

Saleh Elisha Usman, Boyi J. Mairiga, Hyelduku Aliyu Mwada, Ojih Samuel, Ponsah E. Gwamzhi (2023). SPATIAL AND TEMPORAL CHANGES OF URBAN HEAT ISLAND IN JOS METROPOLIS, PLATEAU STATE, NIGERIA. GSJ: Volume 11, Issue 4, April 2023, Online: ISSN 2320-9186 www.globalscientificjournal.com

Taiwo Tolu A., Olusina J.O., Hamid-Mosaku A. I. and Abiodun O. E., (2023). An Investigation of the Performance of Geospatial Features in Machine Learning Feature Selection Techniques. *Nigerian Journal of Environmental Sciences and Technology*, 7(2), pp. 277-290. https://doi.org/10.36263/nijest.2023.02.0422

UN (2025) United Nations Climate Action. https://www.un.org/en/climatechange/science/climate-issues/water Assessed on 29th August 2025

UN Water (2025): Water and Climate Change. https://www.unwater.org/water-facts/water-and-climate-change Assessed on 29th August 2025

Water Aid (2021). What Has Climate Change Got to Do with Water? https://www.wateraid.org/ng/water-and-climate-change Assessed on 29th August 2025

Zakari, S.; Ibro, G.; Moussa, B.; Abdoulaye, T. (2022). Adaptation Strategies to Climate Change and Impacts on Household Income and Food Security: Evidence from Sahelian Region of Niger. Sustainability 2022, 14, 2847. https://doi.org/10.3390/su14052847

Zapata, O. (2021). The relationship between climate conditions and consumption of bottled water: A potential link between climate change and plastic pollution. *Ecol. Econ.* 2021, *187*, 107090.