International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 4 Issue 2 || February 2015 || PP.01-11

Gis Based Analysis of Supply and Forecasting Piped Water Demand in Nairobi

[1,]P. N. Wafula , [2,]T. G. Ngigi
Department of Geomatic Engineering and Geospatial Information System, Jomo Kenyatta University of Agriculture
And Technology, Nairobi, Kenya

ABSTRACT : *Predicting long term water demand is necessary to assess the future adequacy ofwater resources, to attain an efficient allocation of water supplies among competingwater users and to ensure long-term water sustainability. It assists in developing long term water supply infrastructure development. In order to predict future waterdemand and assess the effects of future population growth and other factors on water demand, suitable mathematical models are needed. The study uses GIS based regression model,that is Geographically weighted regression(GWR) and ordinary least square(OLS) to forecast monthly water demand in the western region of NCWSC Water Supply System, Nairobi. Vector dataset (spatial) of the study region by Itinerary levels and statistical data (non-spatial) on water consumption, household, Building density, Land value, connections and population data were used in this exploratory analysis. The result shows that GWR is a significant improvement on the Global model. Comparing both models with the AICc value and the R^2 value revealed that for the former, the value is reduced from 2801 (for OLS model) to 2694 (for GWR model). For the latter, OLS explained 83.46 percent while GWR explained 91.16 percent.The results of the studyshow that the GWR model is capable of predicting waterdemand more accurately than OLS regression model. This implies that local model's fitness is higher than global model. In addition, the empirical analysis revealed that water consumption and demand in the study region is significantly associated with population and Building density. This relationship, as detected by GWR, largely varies across the region. The GWR also achieved the water demand prediction for 2017 and 2020.*

KEY WORDS: *Water Demand, Hot spot analysis, GIS, regression analysis, Forecasting*

I. INTRODUCTION

Water is essential for human existence requiring extensive management, particularly in urban areas where supplies and infrastructure must meet the needs of a heterogeneous and growing population. As population grows in urban areas, ensuring a long-term, safe and reliable source of potable water becomes essential. Determining the amount of water needed by residents is an integral component to water management.Even though water is one of the most important natural good for maintenance of life in earth, the way man have been dealing with this resource is far from optimal. Some researchers point out that the instability between supply and demand of water (with supply lower than demand) will be one of the greatest problem faced by humanity in a not so far future (Glenn, Gordon, and Florescu, 2009; FAO, 2011). It is estimated that in 2025, around 3 billion people will not have access to fresh water. That will represent approximately 60% of the world population, according to Glenn, Gordon, and Florescu (2009).

Nairobi City Water and Sewerage Company (NCWSC), owned by the Kenyan Government, provide sewerage and water services in Nairobi. Nairobi's daily water supply is currently estimated at 580,000 cubic meters per day, against a daily demand of 750,000 cubic meters. Less than 50 per cent of Nairobi's residents have direct access to piped water, of which only 40 per cent have daily access to running water. Only 22 per cent of residents of the informal settlement, home to 60 per cent of Nairobi's residents, have access to piped water.Urban water demand modelling plays an important role in efficient planning, design and development of water supply systems. In order to ensure reliable water supply to the resident s of a city, an accurate estimate of future water demand is necessary. This estimate can help in planning a cost effective and reliable infrastructure expansion, developing alternative water supply sources and incorporating water demand management programs [House-Peters & Chang, 2011].This research has explored how spatial effects might influence water demand and how this can be utilized to predict future demand. In their paper Franczyk and Chang (2008) point that "the standard of water consumption cannot be explained by economic and population growth only, but also by biophysical and socioeconomic factors that usually have spatial dependence". Following the same line House-Peters, Pratt, and Chang (2010) suggest that "residential water consumption is not affected by socioeconomic, climate and physical variables only.

But it is also affected by geographical location and its interaction with nearby regions." Hence, the objective of this project is to rigorously explore GIS modeling techniques for the analysis of water demand and supply pattern in Nairobi at a detailed, sub-regional scale to examine the patterns for whole area of study.

II. STUDY AREA AND METHODOLOGY

Study Area : To enhance service delivery, Nairobi Water has divided Nairobi County into six regions, namely; Western, Eastern, Southern, Northern, North Eastern and Central (Figure 1). Specifically Western region was used for the analysis this study; this was due to its heterogeneous characteristics. For example; Kawangware is a slum area with high population and less piped water connection, Loresho is an area for the wealthy with very low population and well supplied with piped water, highridge represents a middle class area and Westlands is a commercial centre. In this study 155 itineraries of the western region are considered for study.

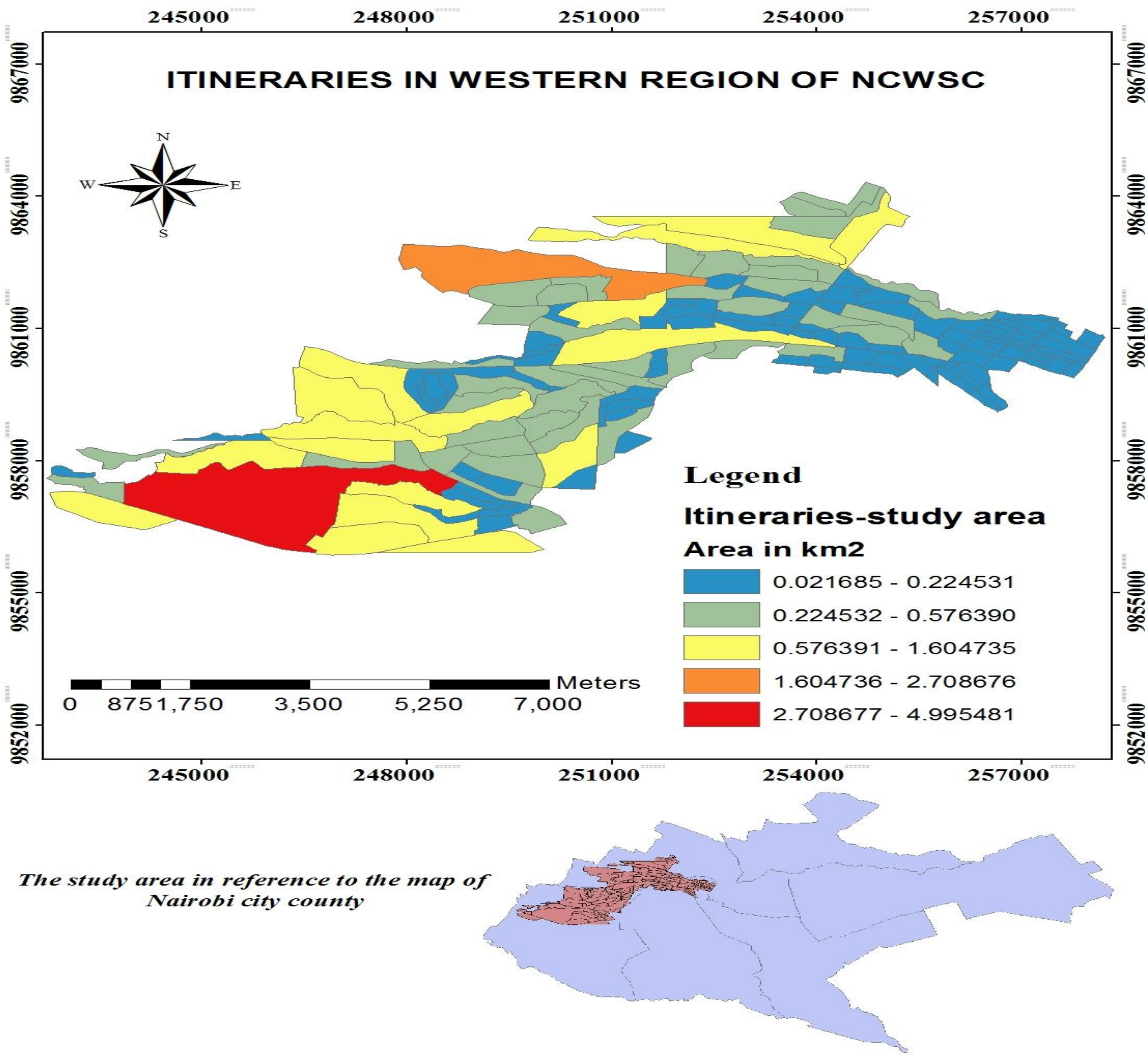

Figure 1: Map of the study area

III. METHODOLOGY

The data collected in order to achieve the objectives of this study included both Primary, and secondary data. Primary data were collected through the use of some selected informants, interviews and field observations. Face-to-face interviews were administered with the senior and junior staff at NCWSC. Observations were made through several visits to the sites, to measure actual water consumption through water meter reading. This enabled the researcher to have firsthand information on what was obtained with regard to consumption in Western Region. Demographic data was obtained from Kenya National Bureau of Statistics

(KNBS). GIS vector data in shapefile format of the water pipe network was obtained from the Engineering Department of NCWSC. Secondary data were obtained through perusing of relevant official records and reports kept at the NCWSC library. This was supplemented by critical perusal of books, publications such as newspapers and journals on the topic under discussion. A geodatabase of the water pipe network for the region was created using ArcGIS 10 and the information captured included; pipe size, monthly consumption, building density, land value, Households, and meter connections information. The geodatabase formed the platform on which various analysis were done. The dependent variable for this model is the documented water consumption for 2009, 2013 through 2014 by itinerary level. These statistical values were entered into the prepared GIS vector polygon map as non-spatial data. To visualize the spatial distribution of such data, a choropleth map was generated to show the water availability density, Demand density and supply density of the study area. It was normalized with the area coverage polygons (in km^2) by itinerary and a five-class natural breaks (Jenks) classification method was applied. In order to detect water consumption hotspots and show continuous distribution, inverse distance weighting (IDW) model was applied on the average water consumption in every itinerary to the map.

IV. ANALYSIS METHODS

In this paper, the OLS and GWR spatial statistical tools were employed for exploring the spatial relationships between water consumption or demand and the five predictors. The OLS was used as a diagnostic tool and for selecting the appropriate predictors (with respect to their strength of correlation with the criterion variable) for the GWR model. It can automatically check for multicollinearity (redundancy among predictors). The multicollinearity was assessed with the variance inflation factor (VIF) values of the OLS. If the VIF value(s) is greater than 7.5, it therefore indicates the existence of multicollinearity among the predictors. In addition, autocorrelation statistic was applied to detect whether there is spatial autocorrelation or clustering of the residuals which violate the assumption of OLS. Progressively, the spatial independency of the residuals was assessed with the global spatial autocorrelation coefficient Moran's I. A scatterplot matrix was used to determine the relationship between water consumption and the factors that were found to influence the formation of low or high consumption. Geographically Weighted Regression (GWR) is a local, spatial, regression method; it allowed the relationships being modeled to vary across the study area. GWR coefficients values, computed by the regression tool, reflected the relationship and strength of each explanatory variable to water demand. The result of OLS and GWR were compared. Basically, the first fundamental geographic question (the where question) regarding water demand distribution in the study region has been answered by Figure 3 (i.e. by displaying the location of water demand hotspots and the spatial pattern of distribution). The next logical geographic questions that follow are "why" such clustering pattern? And "what" are the likely factors that are associated with this observed pattern? The GWR is designed to answer such scientific questions and others like, does the relationship between the dependent variable and the predictors varies across space? Which explanatory variable shows stronger influence in a certain area?

Five factors that influenced water consumption in the study region were identified and selected for the analysis as explanatory variables. This included population density, land value, number of meter connections, household and building density.

V. RESULTS AND DISCUSSION

Water availability: Case study of the Western Region : This is the spatial distribution of water availability density and deficit in the study area. The availability of water or deficit of the same was arrived at by subtracting supply density from demand density. The water availability density is illustrated in figure 2 below. The blue colour showing areas with very low water supply whereas the red shows areas where water supply is adequate.

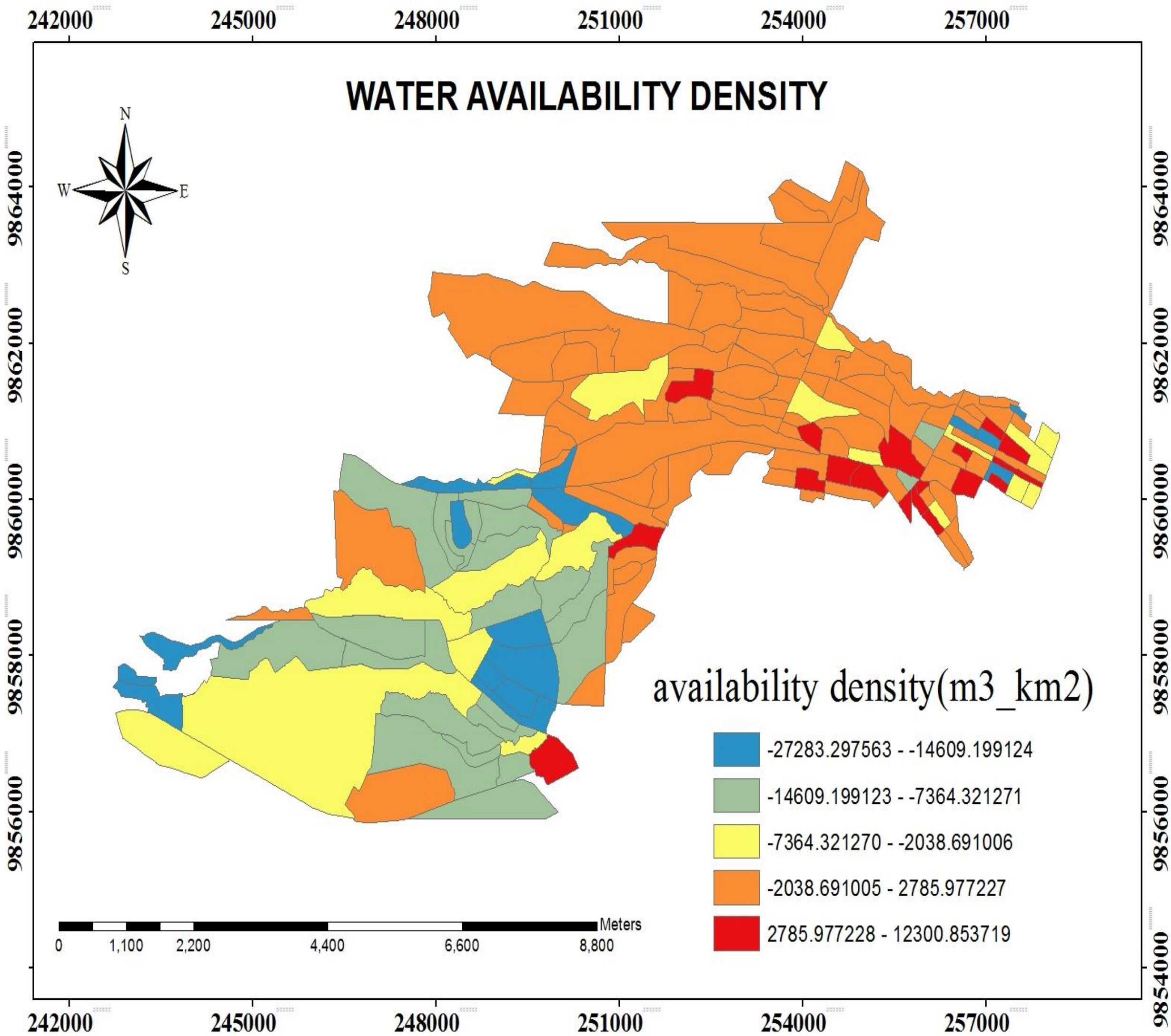

Figure 2: Map of the water availability density

Hot spot analysis and Inverse Distance Weighted Interpolation : Hot spots and cold spots were determined by using the water consumption data from NCWSC for each itinerary; these were the historical consumption information obtained from the NCWSC Western region. Hot spot data allows for the determination of high and low consumption. For conceptualization of spatial relationships; fixed distance band was used. Critical distance was used to decide what neighbours to include and thus the scale of analysis was consistent across the 155 itineraries of the Western region. The hot spot result was interpolated using Inverse Distance Weighted Interpolation (IDW), to form a continuous surface as shown in Figure 3. This was to improve the surface for visualization purposes; as an aid in decision making. It enables us to know the distribution of water consumption hence the question why such pattern of distribution. Interpolation was for visualization purposes only, otherwise, the true statistical analysis happened feature by feature. Showing both the surface and the true results of the hot spot analysis at the same time was a great way to present both the statistical results and the more approachable visualization in a better way. The areas that were had high z score indicated high water consumption and from the hot spot map, they were Muthaiga, Loresho, Spring Valley and Kileleshwa. Areas with low z score values were the cold spots; had low water consumption and these areas were Waithaka, Liruta, Kabiria, Kawangware and Kangemi. The variation in demand for water in the area mentioned above, are due to capacity and the explanatory variables to be considered in these study. From the study we noticed that the slum areas had low meter connections compared to high class areas like Muthaiga that had very many meter connection. Consumption of water varied depending on the region ranging from 60litres per person per day to 250litres per person per day.

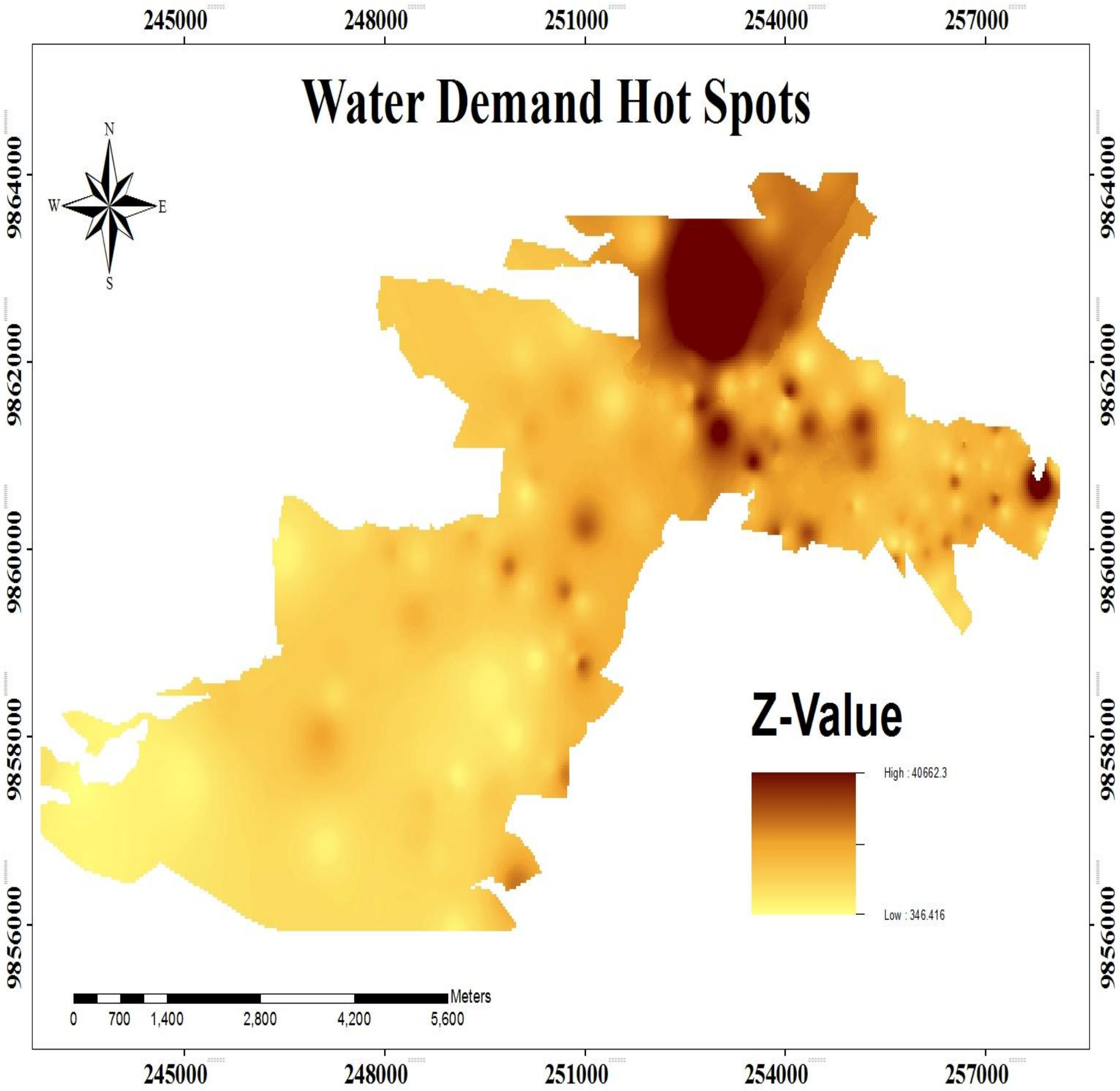

Figure 3: Hot spot analysis interpolated

Finding Key Variables : The next question this study sought to answer was, "Why are consumption so high in those hot spot areas?" and"What are the factors influenced the high consumption?" Explanatory variables influencing water consumption that were collected during the face to face interviews and field visits were assessed using scatterplot matrix graph in ArcGIS see (Figure 4). Variables such as population density, household, meter connections, building density and land value had a positive relationship with water consumption. This meant their increase led to increase water demand in the study region. Increase in population density was observed to increase water consumption in the study area and so was true to the other explanatory factor variables used in this study. For example high class areas like Muthaiga had high land values and the demand for water in these areas is also as high as 250 litres per person per day, compared to low class area with demand of about 60 litres per person per day.

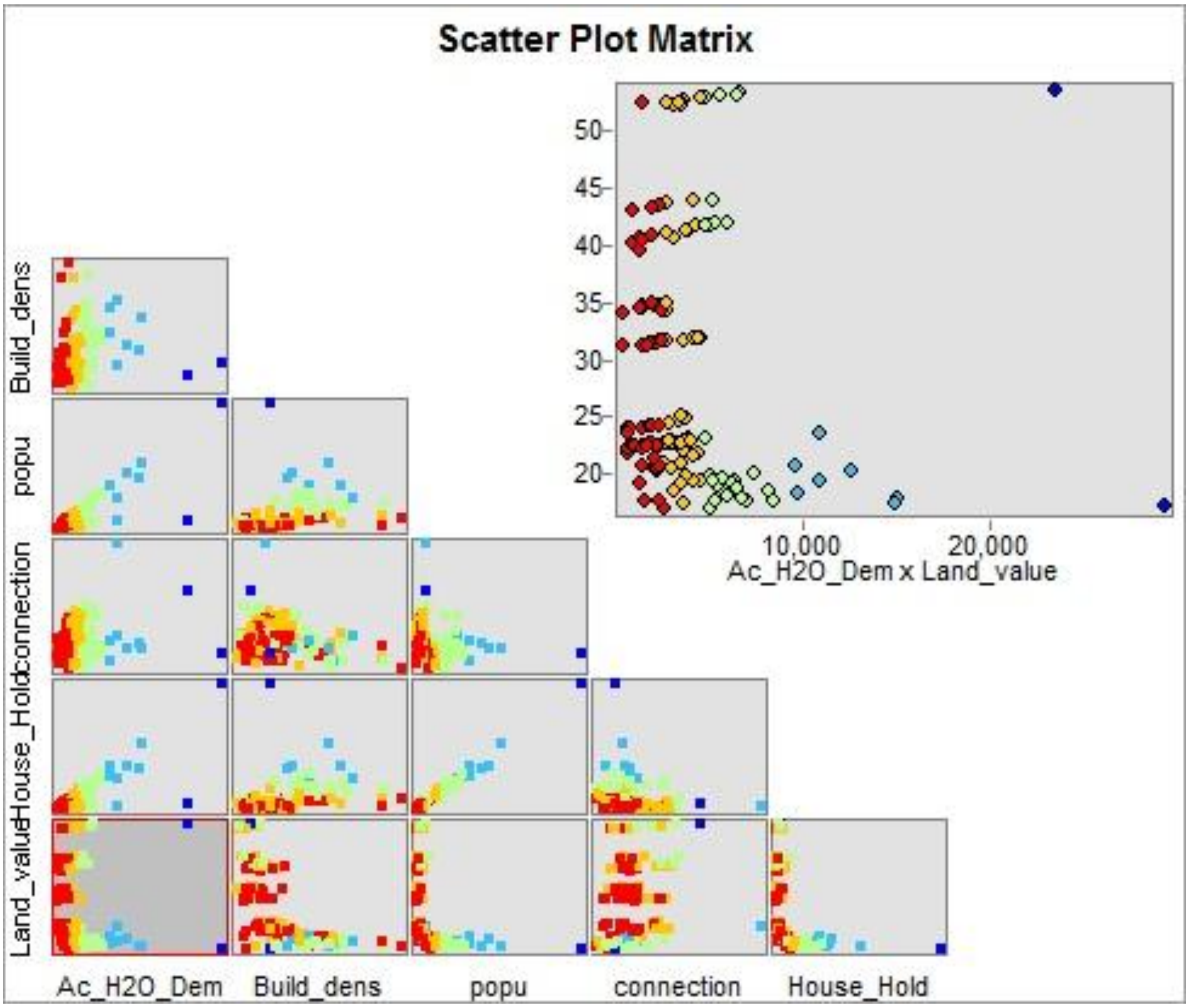

Figure 4: Scatterplot matrix investigating the relationship between water demand and factors influencing the consumption.

Ordinary Least square model : The OLS model was calibrated to diagnose multicollinearity among the explanatory variables and the result shows that the House hold variable and population variable returned VIF values of 47.891227 and 48.935721 respectively. Since these values are higher than the set redundancy threshold of 7.5 and the model indicated that house hold was not significant, the household variables was removed from the model and re-calibrated. Consequently, the R2 value increased from 0.816678 to 0.834698. From, result of recalibration it was shown that all the variables returned VIF values fairly greater than 1.0 indicating that none of the variables were redundant. The summary of the final result of the OLS model is presented in Table 1.

Variable	**Value**
AIC	2801.569696
Adjusted R-Squared	0.834698
Joint F-Statistic	168.210545
Joint Wald Statistic	1776.958849
Koenker (BP) Statistic	0.966035
Jarque-Bera Statistic	19156.609448

Table 1: Summary of global OLS results.

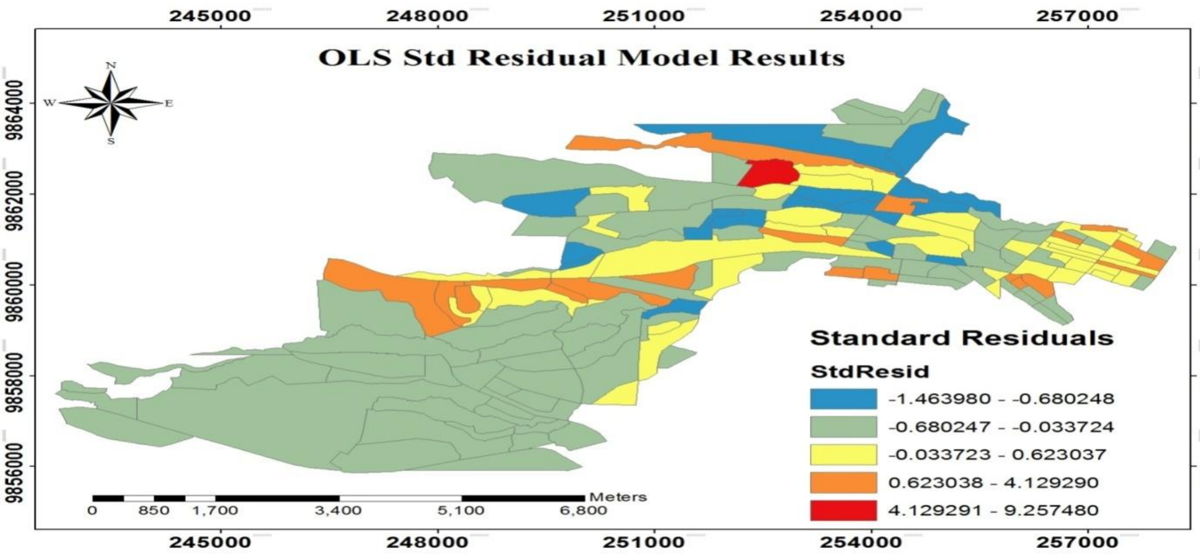

Figure 5: Output map of the Standardized Residuals after running the OLS

To investigate the distributive pattern of the residuals, the OLS generated residuals were mapped (Figure 5). A visual examination of the result shows that no pattern exist, instead the model's residuals exhibit a random noise meaning that there are no clustering of over predictions and under predictions in the model. The red color in Figure 5 depicts the under predicted residuals (positive) while the blue represents the over predicted (negative residuals). However, the result was further confirmed statistically by applying spatial autocorrelation statistic (global Moran's I). This will automatically detect significant clustering or random pattern in the residuals. The Moran's I reportrevealed that the pattern of the residuals is significantly different from random, with a Moran's index value = −0.05 and z-score value = −0.27. That is the residuals have no statistically significant spatial autocorrelation. In this case, all empirical evidence point to the fact that the OLS residuals fit properly.

Geographic Weighted Regression (GWR) Analysis : The identified explanatory variables were subjected to GWR model and this was observed to improve the result compared to that obtained by OLS model. The standardized residual of model are as mapped in figure 6 below.

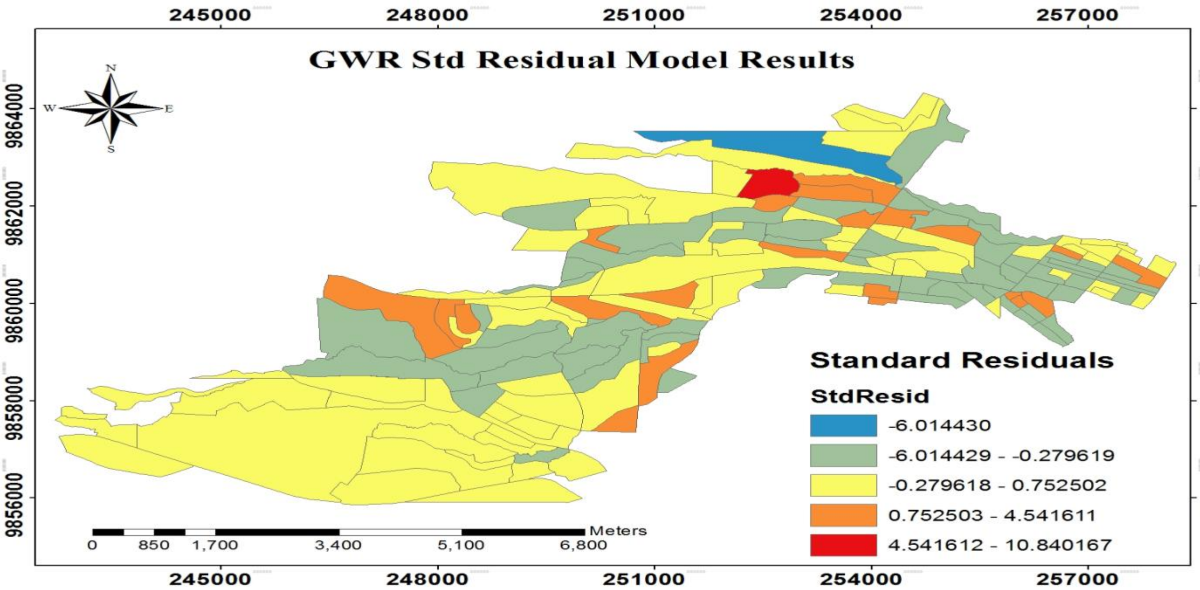

Figure 6: Output map of the Standardized Residuals after running the GWR

Mapping the residuals of GWR indicate that it is randomly distributed (Figure 6). This means the model is properly specified. Verifying with autocorrelation statistic (Moran's I) returned a randomly distributed residuals. The R^2 value as a spatial smoothing of GWR model showing the itineraries where the model's prediction and strength of relationship is improved. Overall, the R^2 value (0.95) shows a strong significant relationship between water consumption and the explanatory variables used in the investigation. Table 2 summarizes the result of the GWR model.

Variable	Value
AICc	2694.6344790625367
R2	0.9517062523448627
R2Adjusted	0.911695329484053

Table 2: Summary of local GWR results.

Relationship and Strength of each Factor to the consumption : Both models (global OLS and local GWR) were able to capture and detect prominent factors (variables) that influence water consumption in the area of study. However, in this discussion session, only the useful predictors (those without bias that were entered into the model) will be analyzed. In the exploratory analysis using OLS, four predictors were entered into the models-population density, land value, building density, and Meter connections. All the predictors returned positive relationships. Coefficients values, computed by the regression tool, reflected the relationship and strength of each factor to the Water demand in the study region. These factors were symbolized to reveal their trend which showed that most were strong indicators of water demand in some locations and weak in other locations. The darkest coloured areas (Figure 7, 8 and 9) are locations in the study area where a Variable had the strongest relationship with water demand whereas the lightest coloured areas are locations where a variablehad the weakest relationship with the water demand. The dark coloured areas in (Figure 7) are locations in the study area where the population density was a very strong predictor of water demand. The light coloured areas were locations where the variable(population density) was less important. NCWSC can focus on this information especially in administering and planning for adequate water supply to its customers. This is particularly useful in controlling supply to specific areas, where region of high consumption are given first priority in terms of capacity of pipe network and water storage tanks.

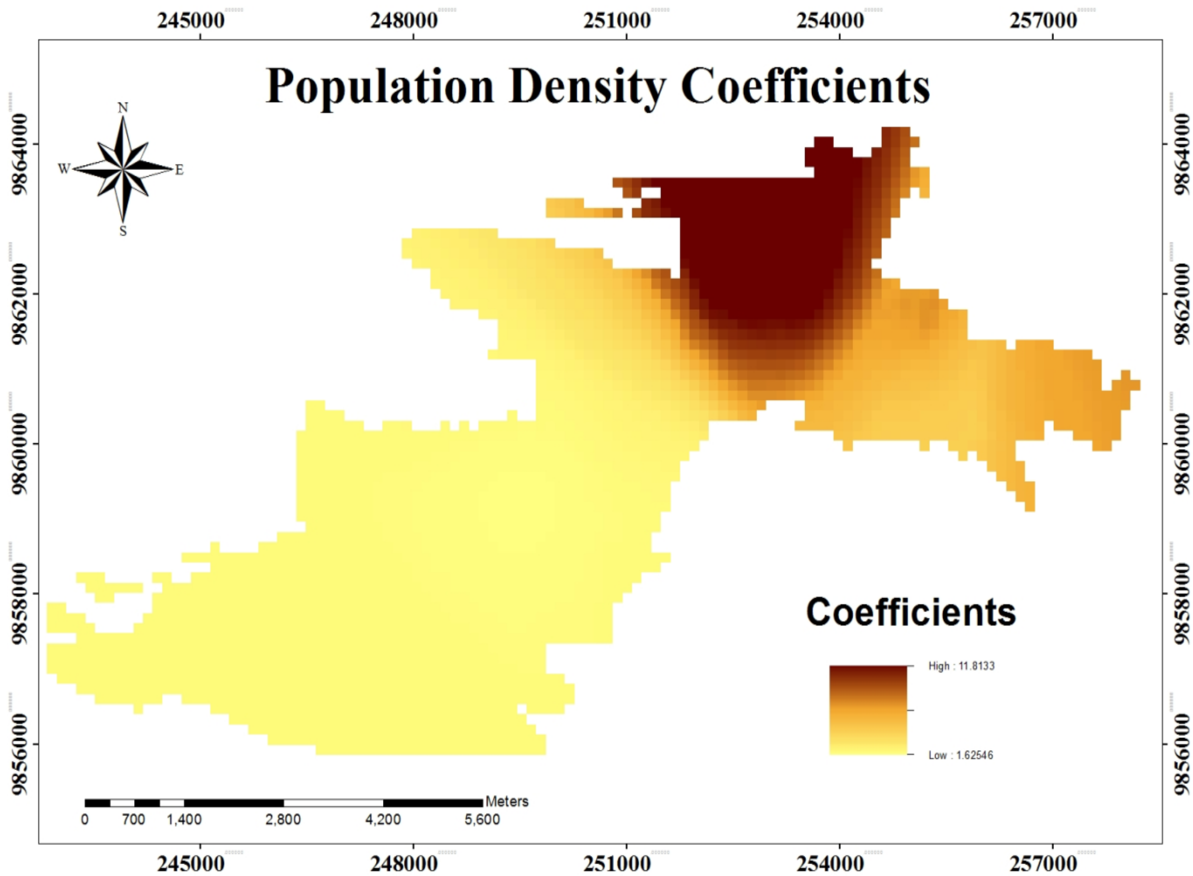

Figure 7: Spatial relationship of population Density to water demand

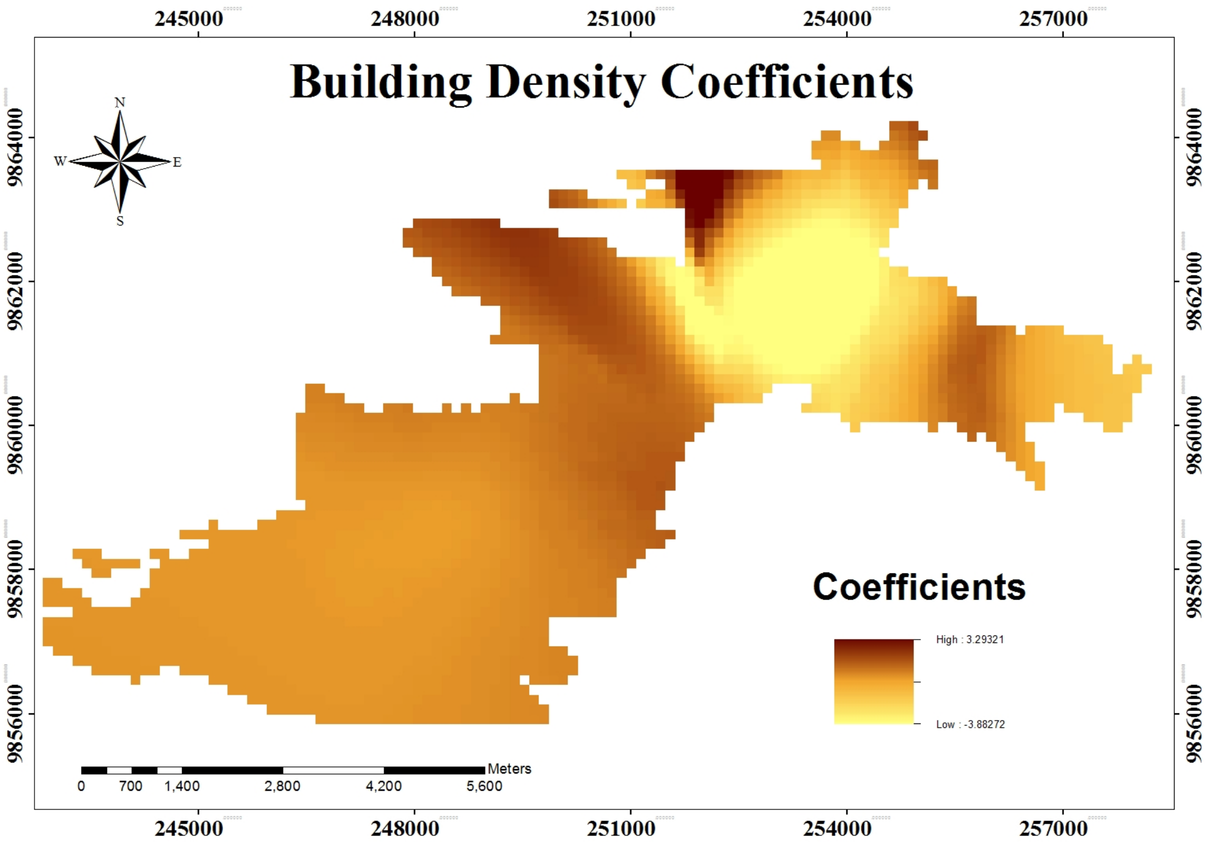

Figure 8: Spatial relationship of building Density to water demand

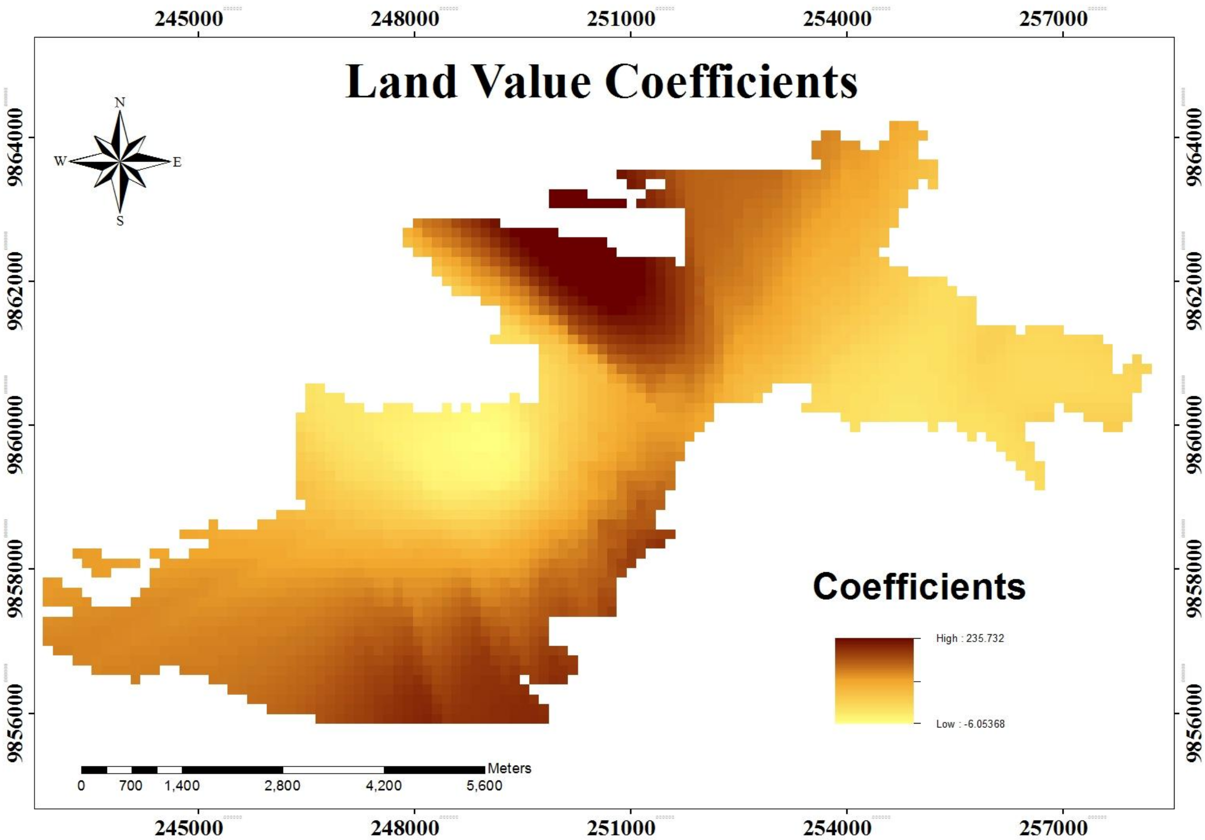

Figure 9: Spatial relationship of land Value to water demand

Prediction of future water demand : GWR was then used to predict values for a future time or for locations within the study area. Where in this case we had X (explanatory variables) values, but don't know what the Y (water demand) values are. We then applied GWR to predict the future water demand for the area of study using the explanatory variables for the model which were the X values for the model. The model was calibrated using the explanatory variables we were using all along, but that the explanatory variables for the predictions are new. The new variables represented projected population density, Building density, and land value variables for some time in the future that is for 2017 and 2020. To verify the result of the prediction we applied the GWR model on the consumption data for 2009 to predict what would be the consumption for 2013. The result from the model compared with the actual consumption recorded by NCWSC showed very minimal deviation from each other therefore making the result from the model reliable. Using 2009 as the base year we were able to predict the water consumption or demand for the study region to 2013, 2017 and 2020. The result fora few selected itinerariesis as shown in graph(figure 10) below.

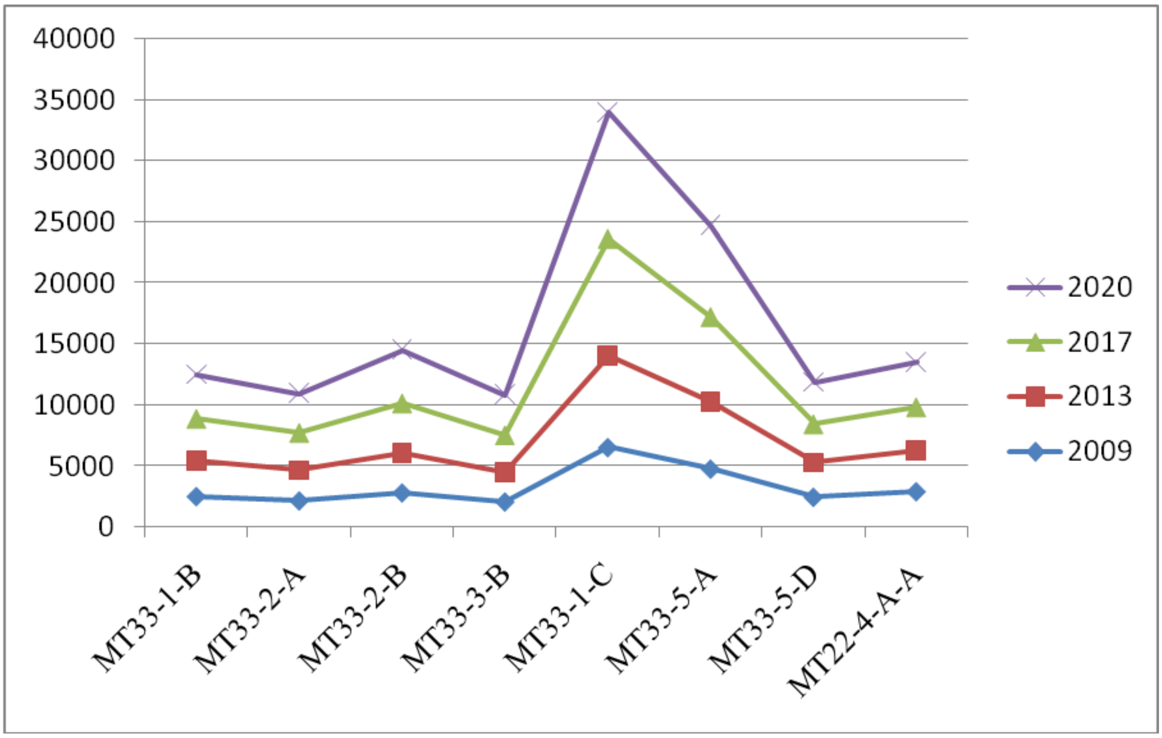

Figure 10: Graph of graph of actual consumption in 2009 and predictions

VI. CONCLUSION AND RECOMMENDATION

The exploratory analysis in this study explains the spatial variation in relationship among geographic dataset and across geographic regions. Using GIS-based local model and global statistic to explore the relationship between water demand and other variable affecting demand, it was able to detect and extract certain key information concerning stationarity and non-stationarity in spatial data. This study has explicitly shown that in spatial data, relationship is not static across geographic space by comparing the results of global OLS and local GWR model's fitness and parameter estimates, GWR model improved the result over the OLS model comparing R2 value. It has demonstrated that GWR can be used to predict water demand which is useful in capacity building for supply to ensure demand for water is met by the water supply. It was discovered that water demand in the study region is significantly associated with population and building density that is directly linked to meter connections and households. Others significant variable are land value, building density, and household. Household was found to be insignificant as a variable since it directly related to population.

Finally, this study is a contribution to the field of GIS, spatial statistics and urban water demand modeling. It presents essential evidence that water demand forecasting can be achieved using GWR as a predictor in modeling data.

VII. ACKNOWLEDGEMENTS

Authors would like to thank Nairobi City Water and Sewerage Company for providing the dataset, particularly Mr. Barasa Masinde, from the Western Region and Mr. Martin Irungu Karethi from the Headquarters.

REFERENCES

[1]. Glenn, J., T. Gordon, and E. Florescu (2009): 2009 State of the Future. Millennium Project.

[2]. House-Peters, L., B. Pratt, and H. Chang (2010): "Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in Hillsboro, Oregon," JAWRA Journal of the American Water Resources Association, 46(3).

[3]. Franczyk, J., and H. Chang (2008): "Spatial analysis of water use in oregon, USA, 1985-2005," Water Resources Management, 23(4), 755–774.

[4]. Kenya National Bureau of Statistics, 2001: Kenya Population and Housing Census 1999, Volume I: PopulationDistribution by Administrative Areas and Urban Centres. National Bureau of Statistics, Nairobi.

[5]. Kenya National Bureau of Statistics, 2001: Constituency Report on Well Being in Kenya, Based on Integrated Household Budget Survey 2005/06. National Bureau of Statistics, Nairobi.

[6]. Wentz, E. a., and P. Gober (2007): "Determinants of small-area water consumption for the city of Phoenix, Arizona," Water Resources Management, 21(11), 1849–1863.

[7]. Fotheringham, A.S., C. Brunsdon, M.E. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. Wiley: New York.

[8]. FAO (2011): "Climate change, water and food security," Discussion paper, Food and Agriculture Organization of the United Nations.

[9]. Chang, H., G. H. Parandvash, and V. Shandas (2010): "Spatial variations of single-family residential water consumption in Portland, Oregon," Urban Geography, 31(7), 953–972.

[10]. Brunsdon C, Fotheringham AS and Charlton M, 1996, Geographically weighted regression: a method for exploring spatial non-stationarity, *Geographical Analysis*, 28(4), 281-298

[11]. Anselin L, 1988. *Spatial econometrics: methods and models*, Dordrecht: Kluwer Academic Publishers

[12]. Chiuta, M. T., Johnson, P. and Hirji, R, (2002). "Water resources and the economy" In defining and mainstreaming environmental sustainability in water resources management in Southern Africa. SADC, IUCN, SARDC-CEP:Maseru/Harare.

[13]. Watson, D.F. and G.M. Philip., 1985. A refinement of inverse distance weighted interpolation. *Geo-Processing* 2: 315- 327.

[14]. C. Lin and T. Wen, "Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue," International Journal of Environmental Research and Public Health, Vol. 8, No. 7, 2011, pp. 2798-2814.