The International Journal Of Engineering And Science (IJES)
|| Volume || 3 || Issue || 11 || Pages || PP -01-09|| 2014 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Comparison of Ordinary Least Square Regression and Geographically Weighted Regression for Estimating and Modeling the Electricity Distribution Using Geographical Information System (GIS)/Remote Sensing (RS)

[1]Kalai Selvi.J, Research Scholar, [2]Vidhya .R , Professor, [3]Manonmani.R,Research Scholar

[1]Institute of Remote Sensing,Anna University,Chennai,Tamil Nadu
[2]Institute of Remote Sensing,Anna University,Chennai,Tamil Nadu
[3]Institute of Remote Sesning, Anna University,Chennai,Tamil Nadu

ABSTRACT

Electricity is a key energy source in each country and an important condition for economic development especially in industrial area. The current energy situation in the region is characterized by a rapid increase in energy demand due to urbanization, rapid population growth and economic growth all add to rising energy demand . So it is very important to know the present spatial distribution of network. Understanding the relationship between the spatial distribution of electricity network and different land use types accounting for the spatial non stationarity can help electricity planers to better evaluate the assessment of network distribution. A relatively new technique, geographically weighted regression (GWR) has the ability to account for spatial non stationarity with space. While its application is growing in other scientific disciplines, the application of this new technique in electricity distribution has not been used elsewhere. The geographic information system (GIS) , along with the two different empirical techniques(GWR and Ordinary least square regression) was used to analyze the relationship between low tension (LT) distribution and various land use classes derived from recent high resolution satellite image quick bird for Manali (Industrial region) in Chennai. Low tension was spatially interpolated in ArcGIS using interpolation techniques with zonal statistics. The explanatory variables used are the Land use parameters like built-up area, scrub, agricultural land, industry etc and the socio economic factor population growth. The OLS model performed moderately well (AIC=31.665, R2=31.9% and Adjusted R2=31.8%), the Moran's I =0.66 for the residuals from the OLS model. The best results were obtained with the GWR model (AIC=19.08, R2=51.65% and Adjusted R2=50.42%) The results suggest that GWR provides an effective estimation for modeling the LT consumer's distribution network pattern.

KEYWORDS: *Remote Sensing, Geographical Information System, Geographically Weighted Regression, Ordinary least square component;*

Date of Submission: 27 October 2014 — Date of Accepted: 15 November 2014

I. INTRODUCTION

Energy has come to be known as a `strategic commodity' and any uncertainty about its supply can threaten the functioning of the economy, particularly in developing country like India. As demands for the electricity energy have increases in urban area due to tremdeous population growth as a result in change in land use patterns and industry establishment. Energy is critical, directly or indirectly, in the complete process of evolution, growth and survival of all living beings and it plays an important role in the socio-economic development and human welfare of a country . Hence, to offer high-quality service to the end user considering the urbanization growth an adequate planning for energy or demand estimation is very much essential.

In recent years Geographic information system has been an important development in the field of electricity network [1][2] and it is used for the spatial data management and manipulation. With the advent of remote sensing and GIS technologies, the mapping of electricity distribution network with considering socio economic and land use variable have been widely used[3]. There are many studies on electricity related to trend analysis[4] and many these studies applied regression method as non spatial[5] .

Recent new technology known as Geographically weighted regression is applied to study the spatial relationship between more than two variable. One of the nonparametric modeling method is the geographically weighted regression (GWR) technique[6]. GWR is among the new developments of local spatial analytical techniques. It is a local spatial statistical technique that relies on a form of kernel regression within a multiple linear regression framework to develop local relationships between the dependent and independent variables[7][8][9]. GWR was used to examine the spatially varying relationships between several urbanization indicators based on LULC changes. Geographically weighted regression is an exploratory technique mainly intended to indicate where non-stationarity is taking place on the map, that is where locally weighted regression coefficients move away from their global values. In another research[10] the relationship between precipitation verse irrigated and rain fed crop was carried out. Another research [11] is about analysis the relationship between agricultural landscape pattern and urban.

Numerous studies have described the application of the Geographical weighted regression method for land use change. In addition, the geographical weighted regression method has been applied to urban heat estimation [12], urban growth [13], crime mapping [14], Fisheries [15], population [16], Electricity consumption[17] and Groundwater subsidence [18]. Further GWR was used to investigated relationship between electricity consumption and household income. The results reveal that electricity consumption is useful for characterizing household income, a frequently used proxy for purchasing power. Stochastic model to estimate the efficiency analysis of electricity distribution[19] network using sample of about 500 electricity distribution utilities from seven European countries. Although Geographical regression method has been applied to crime mapping, fisheries, land use mapping, heat island, this approach have not yet been used in analysis relationship between distribution of low tension and land use type.

At present, energy modeling is a subject of widespread interest among engineers and scientists concerned with the problems of energy production and consumption. Modeling in some areas of application is now capable of making useful contributions to planning and policy formulation. GWR and OLS have the ability to predict the locations where the network expansion will occur is useful not only for proper utilization of the power , but also for policy makers who need to plan and manage the outcomes of spatial processes at regional or local levels. Spatially estimating the future demand is very important for the economical future expansion and safe operation of a distribution network

This study compares the accuracy in predicting energy consumption in the study area using OLS and GWR and also examines the relationship between energy consumption and diverse independent variables. The energy consumption estimation is calculated based on Land use parameters and also based on population. Results are compared and this study provides important reference materials for the utility companies in assessing energy consumption.

This will be helpful for a successful electricity planner.

II. AREA AND METHODS FOR RESEARCH

a. Objective

The main objective of this study includes

- Applying OLS and GWR to estimate the distribution pattern of LT Network for planning future expansion based on the independent variables.

- Evaluating the performance of OLS and GWR models.

- Validation of the results.

b. Area of research

The study area selected is situated in the city Chennai in the state of Tamil Nadu.The Chennai Metropolitan Area (CMA) comprises the city of Chennai, 16 Municipalities, 20 Town Panchayats and 214 Village Panchayats in 10 Panchayat Unions. Manali is an Industrial town and Municipality in Thiruvallur district in the state of Tamil Nadu. It is located in north of Chennai City . Manali is further divided into six wards and Manali had a population of 58,174 as per the census data. Manali is located in the Northern Suburb of Chennai city. Since this area consist both industrial as well as domestic loads this has been taken for case study and it's the one of the fast developing areas in Chennai whereas the load growth is in the rate of around 8% per year. Manali area is fed by the Manali 230/110KV Substation which in turn connected with the thermal generators located in north chennai. The study area Manali is shown in Figure[1].

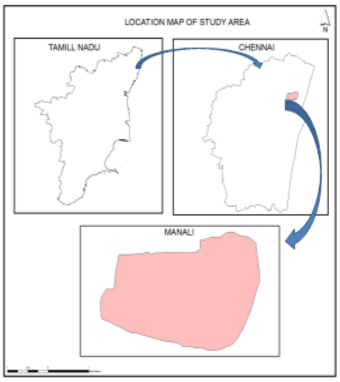

Fig.:1 Study area

III. OVERVIEW OF MODELS

A. Ordinary least squares (OLS) method

Various interpolation techniques are available to predict and interpolate information or variables within predetermined boundaries. The ordinary linear regression model the estimation of the coefficients by using Ordinary Least Squares. The residuals are assumed to be independently and identically distributed around a mean of zero. The errors are also assumed to be homoscedastic, i.e with constant variance. A regression model is expressed as inn equation(1).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{-----------(1)}$$

for $i=1....n$

Where Y is the dependent variable. The independent variables are known as predictor variables. The ε_i is the error term, and β_0 and β_1 are parameters which are to be estimated. The OLS estimator can be written in the form shown in equation(2)

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{---------------- (2)}$$

where $\hat{\beta}$ is the vector of estimated parameters, X is the design matrix which contains the values of the independent variables and a column of 1s, y is the vector of observed values, and $(X^T X)^{-1}$ is the inverse of the variance-covariance matrix. Weights can also be included in the OLS estimator and they are placed in the leading diagonal of a square matrix W , the estimator with weights are shown in equation(3).

$$\hat{\beta} = (X^T WX)^{-1} X^T Wy \quad \text{---------- (3)}$$

The ability of the model to replicate the observed y values is measured by the goodness of fit. This is expressed by the r^2 value which runs from 0 to 1 and measures the proportion of variation in the observed y.

B. Spatial Autocorrelation

Autocorrelation means that a variable is correlated with itself . The simplest definition of autocorrelation states that pairs of subjects that are close to each other are more likely to have more similar values , and pairs of subjects far apart from each other are more likely to have less similar values . Gradients or clusters are examples of spatial structures that are positively correlated, whereas negative correlation may be exhibited in a checkerboard pattern where subjects appear to repulse each other. When data are spatially auto correlated, it is possible to predict the value at one location based on the value sampled from a nearby location when data using interpolation methods. The absence of autocorrelation implies data are independent. Moran's *I* is a measure of spatial autocorrelation.

C. Geographically Weighted Regression (GWR)

GWR is the analysis of spatially varying relationship .GWR is to explore how the relationship between a dependent variable (Y) and one or more independent variables (the Xs) might vary geographically. Geographically Weighted Regression (GWR) is a recent contribution to modeling spatially heterogeneous processes. Using GWR the parameters may be estimated anywhere in the study area by the given dependent variable and a set of one or more independent variables which have been measured at places whose location coordinates are known. GWR model considers the differences of spatial location and the spatial correlation, which allows local rather than global parameter estimation, the estimated parameters are different with the spatial location varies. It can be regarded as a local model .

The GWR model equation would be:

$$y_i(\mathbf{u}) = \beta_{0i}(\mathbf{u}) + \beta_{1i}(\mathbf{u})x_{1i} + \beta_{2i}(\mathbf{u})x_{2i} + \dots + \beta_{mi}(\mathbf{u})x_{mi} \text{ ---------------- } (4)$$

A prediction/estimation may be made for the dependent variable if measurements for the independent variables are also available at the same location **u**. In general the GWR works by moving a search window from one point in a data set to the other, working through them all in a ordered list or sequence. The distance from one point to another can be defined by actual geographic distance or by its sequence of position i.e first nearest point or second and so on. The goodness of fit measured in GWR is the corrected Akaike Information Criterion.

$$AIC_c = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left(\frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})}\right) \text{ ------------------ } (5)$$

In AIC method, the user can choose a fixed bandwidth or a variable bandwidth that expands in areas of sparse observations and shrinks in areas of dense observations (Charlton et al., no date). Because the regression equation is calibrated independently for each observation, a separate parameter estimate, t-value, and goodness-of-fit is calculated for each observation. These values can thus be mapped, allowing the analyst to visually interpret the spatial distribution of the nature and strength of the relationships among explanatory and dependent variables.

IV. MODEL PARAMETER

A. Dependent variable

The feeder emanating from the substation has been mapped in GIS along with all roads and buildings. A very high resolution map taken from satellite is used us to map network elements in GIS and the spatial co-ordinates of poles ,transformers, individual service lines from the pole etc has been acquired. Further, the attribute data like make of distribution transformer, capacity of each DT, source feeder details, year of commissioning etc are also recorded along with each object which is given in Figure[3&4].

B. Independent variable

Further, land-uses of Manali area has been classified into nine categories. Viz. built up, canal, crop, Cooling pond, Industry, Plantation, Scrub, River and Tank. In this study supervised classification techniques were adopted. A supervised classification method was carried out using training sets and test data for accuracy assessment. Classified land use / land cover maps for the year 2006 and 2012 . After classified thematic maps were developed, accuracy was tested by different methods of accuracy assessment, and the post-classification process was the last process in classification. This was considered as input for the model.

C. Demographic profile

Population growth in the study area

Tamilnadu has emerged as the third largest economy in India.

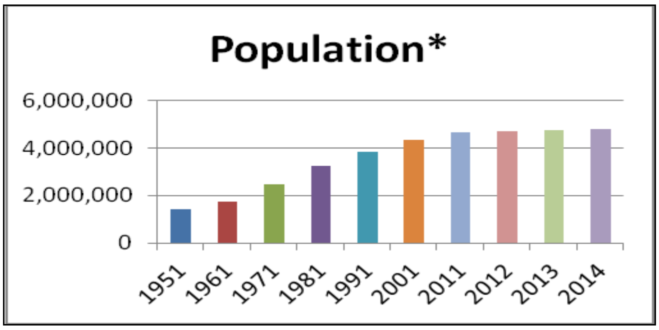

Fig.: 2 Chennai Population till 2014.

In the recent past, liberalization, rapidly growing IT sector, an educated, hardworking and disciplined work force etc, accelerating economic development also contributed to the growth of urban areas in Tamilnadu. The extent of the State is130,058 sq.km. of which the urban area accounts for 12,525 sq.km. Tamilnadu is the most urbanized state in India. Chennai was established in 1639 and it has grown to the fourth largest Metro City in India.Fig.2 shows the population growth is Chennai. According to India Census , the study area Manali had a population of 58,174. Manali has an average literacy rate of 72 %, higher than the national average of 59.5 % .The population growth will also play an important role in energy consumption. Hence this also been considered in the analysis.

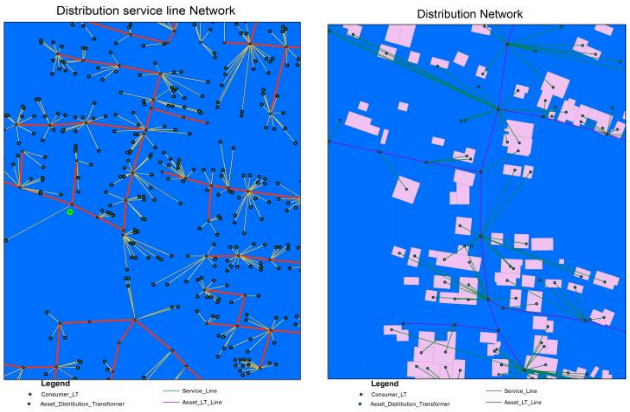

Fig.:3&4 Mapping of Distribution Network

V. RESULT AND DISCUSSION

The relationship between the electricity consumption pattern with the independent factors viz. Land use parameters and population growth the socio- economic factor are estimated using OLS regression model, auto correlation and GWR model. The land use map and the corresponding estimation results observed using OLS Regression model is shown in Fig(5) and (6). It is observed from the result that the spatial estimate result of LT Consumers using OLS model falls maximum under the built-up area and it is very minimum in scrub land.

Table 1.

R^2 Values from OLS MODEL

Object Id	Diag_Name	Diag_Value
1	AIC	31665.68
2	R^2	0.319
3	AdjR^2	0.318
4	F-Stat	674.759
5	Wald	305.241
6	K(BP)	1587.229
7	JB	259669.648
8	Sigma²	14.076

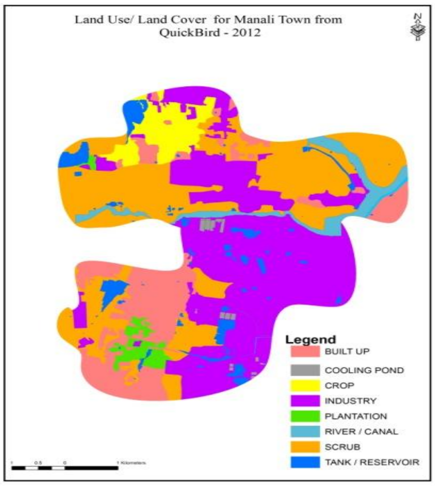

Fig.: 5. Land use and Land cover

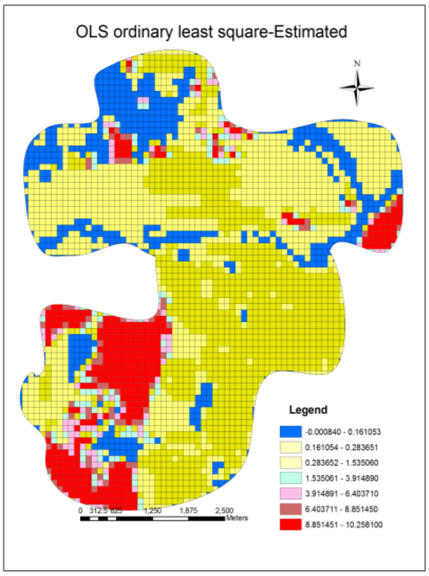

Fig.:6.OLS Model Regression

Under the built-up area and it is very minimum in scrub land. The R^2 values obtained from OLS is displayed in Table [1]. The t-statistics values observed from the coefficient table are shown in Table[2].

Table-2.

Coefficient Table

Id	Coef	StdError	t_Stat	Prob
1	0.021042	0.068954	0.305164	0.760262
2	0.102445	0.002004	51.11694	0
3	0.003086	0.001363	2.264861	0.023542
4	0.014374	0.011825	1.215536	0.224214
5	-0.00157	0.038385	-0.04089	0.967373

The t-statistics test the hypothesis that the value of an individual coefficient estimate is not significantly different from zero. Here for the variable Built-up area, t-statistics values are more statistically significant.

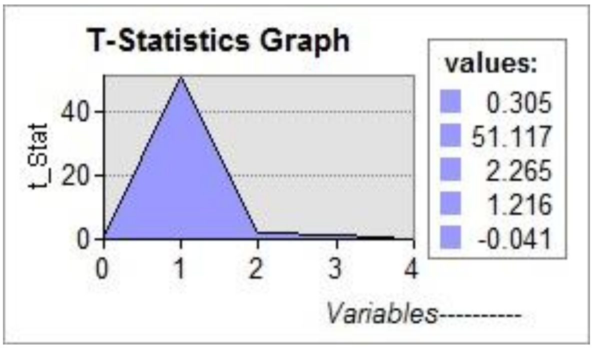

Fig.:7 t-statistics Graph

The report from the OLS advised that we should carry out a test to determine whether there is spatial autocorrelation in the residuals. If the residuals are sufficiently auto correlated, then the results of the OLS regression analysis are unreliable. An appropriate test statistic is Moran's I, this is a measure of the level of spatial autocorrelation in the residuals. Auto correlation measures spatial autocorrelation based on both feature locations and feature values simultaneously. It should be between -1.0 to 1.0 . The results of auto correlation estimates are shown in Table.[3]

Table : 3 Auto correlation result

Moran's Index : 0.66	
Z score : 53.61 standard Deviation	
Significance Level	Critical Value
0.01	-2.58
0.05	-1.96
0.10	-1.65
Random	-1.65 to 1.65
0.10	1.65
0.05	1.96
0.01	2.58

With the given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random. The calculated the Moran's I Index value and both a z-score and p-value to evaluate the significance of that Index. P-values are numerical approximations of the area under the curve for a known distribution, limited by the test statistic. To estimate the location of LT consumers after auto correlation the GWR model was used to identify the hotspots and the local results R^2 is shown in Fig.[8].

The hotspots in red colored squares Fig.[9] indicate the hotspot areas where the electricity consumers are more and their total consumption is also high compared with the other regions. Even the area around this hotspot is also built up area the reason for this estimation is, this area contains multi-storey buildings with more than one connection for an individual building. The Resultant table obtained from GWR is given as Fig.[4].

Table 4 : GWR Resultant Table**.**

Name	**Value**	**Description**
Bandwidth	895.430243	Dependent variable- LT Consumer
Residual squares	55865.531082	
Effective Number	80.742341	
Sigma	4.129999	
AIC	19084.301630	
R^2	0.516006	
R^2 Adjusted	0.504222	

The GWR table is for measuring the goodness of fit. It contains Residual squares , r^2 , adjusted r^2 and the sigma values. The r^2 measures the proportion of the variation in the dependent variable which is accounted for by the variation in the model, and the possible values range from 0 to 1. Values closer to 1 indicate that the model has a better predictive performance. However, its values can be influenced by the number of the variables which are in the model – increasing the number of variables will never decrease the r^2. The adjusted r^2 is a preferable measure since it contains some adjustment for the number of variables in the model. Goodness of fit measurements : the r^2 is 0.516 and the adjusted r^2 is 0.504. The comparative table is given as Table[5].

Table 5: Validation of Models		
Name	Values from GWR	Values from OLS
AIC	19084.301630	31665.680
R^2	0.516006	0.319
R^2 Adjusted	0.504222	0.318
Sigma	4.12999	_
Sigma ^{4}	14.076	_

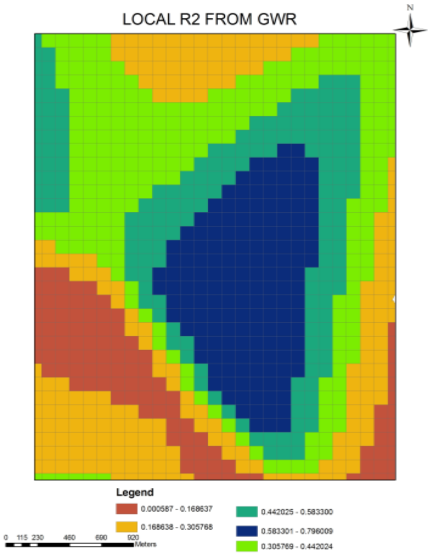

Fig. 8: GWR Local R^2

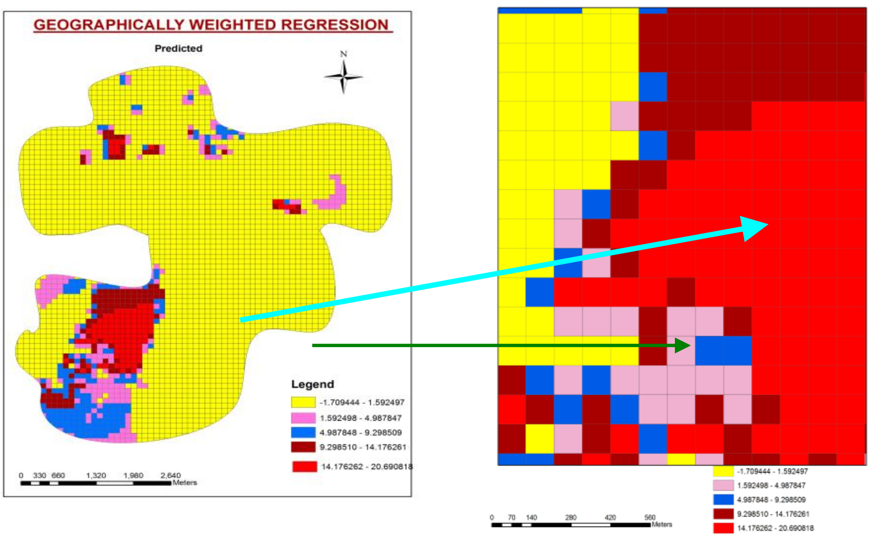

Fig. [9] Geographically weighted Regression predicted results

It is inferred from the above the results that the R^2 values close to the value 1 indicates better estimation results and the lowest Akaike Information Criterion (AIC) value which is the relative measure of goodness-of-fit indicates better results. From the result the OLS model performed moderately well (AIC=31.665, R2=31.9% and

Adjusted R2=31.8%), the Moran's I =0.66 for the residuals from the OLS model. The best results were obtained with the GWR model (AIC=19.08, R2=51.65% and Adjusted R2=50.42%) Wherein in GWR model gives better results of estimation than OLS model. The estimated values are compared with the actual physical distribution of network.

VI. CONCLUSION

This study explored the use of ordinary least squares (OLS) regression, spatial autocorrelation and geographically weighted regression (GWR) for modeling and analyzing the spatial varying relationships between LT Consumers and land use pattern in the study area. Results lead to the conclusion that GWR was more powerful and effective in interpreting relationships between LT consumer and land use pattern, particularly in relation to urbanization. Characters and strength of the relationships identified by GWR showed great spatial estimates. Given that impacts of different urbanization indictors of landscape patterns operated at different spatial scales, the OLS and GWR estimated the dependent variables distribution pattern from other independent variables or drivers. This study can be preceded further by using both the regression models for two different type of study areas and results can be analyzed based on its urbanization growth rate.

ACKNOWLEDGMENT

The author would like to thank TamilNadu Transmission Corporation Ltd., Chennai for providing data for this research work.

REFERENCES

[1]. A. Nagaraha Sekhar, K.S Rajan and Amit Jain, "Application of Geographical Information system and spatial informatics to Electric power systems" Fifteenth national power system conference ,NPSC,IIT Bombay, 2008
[2]. T.Jamal, , W. Ongsakul , M.S.H Lipu and M.S, Islam "An approach to integrate Geographic information system to the proposed smart grid for Dhaka, Bangladesh," International journal of scientific engineering and technology, vol No 3, PP124-129,2014
[3]. J. Kalai selvi. R.Vidhya R. Manonmani, "Geospatial technique in urban distribution network mapping and modeling" International journal of Asia Pacific research, vol 1, pp 63-70,2014
[4]. P. Chujai, N Kerdprasop and K. Kerdprasop , " Time series analysis of household electric consumption with ARIMA and ARIMA models" Proceedigs of the international Multi conference of engineers and computer scientists, vol IMECS 2013, 2013
[5]. V. Lepojevic and M. Andelkovic –Pesic " Forecasting electricity consumption by using holt-winters and seasonal regression models" Facta Universitatis Economics and Organization vol 8, no 4,pp421-431,2011
[6]. M.Cahill ,G. Mulligan ".Using geographically weighted regression to explore local crime patterns" journal Soc Sci Comput Rev , 25:174–193,2007
[7]. Fotheringham, A.S., Charlton, M. & Brunsdon, C., 1997a; Measuring spatial variations in relationships with geographically weighted regression, in Fischer, M.M. & Getis, A. (eds), pp 60-82. 1997
[8]. Brunsdon, C., Fotheringham, A.S. & Charlton, M.E., 1996; Geographically weighted regression - a method for exploring spatial nonstationarity, *Geographical Analysis*, 28 (4), 281-298. 1996
[9]. Robinson, D.P., Lloyd, C.D., McKinley, J.M. Increasing the accuracy of nitrogen dioxide (NO2) pollution mapping using geographically weighted regression (GWR) and geostatistics. Int. J. Appl. Earth Obs. 21, 374-383,2013
[10]. V. Sharma, S. Irmak, I. Kabenge "Application of GIS and GWR to evaluate the spatial non-stationarity relationship between precipitation Vs irrigated and rain fed maize and soybean yields" American society of Agricultural and biological engineers , vol 54(3) pp 953-972,2011
[11]. S. Su , R. Xiao and Y. Zhang 2012, Multi-scale analysis of spatially varying relationships between agricultural landscape patterns and urbanization using geographically weighted regression, Applied Geography 32 ,2012
[12]. M. Szymanowski and M. Kryza,2012 Application of geographically weighted regression for modeling the spatial structure of urban heat island in the city of Wroclaw (SW Poland) Procedia Environmental Sciences 3 pp 14–19,2011
[13]. N.M Shariff , S. Gairola and A. Talib " Modeling Urban Land Use Change Using Geographically Weighted Regression and the Implications for Sustainable Environmental Planning," International Environmental Modeling and Software Society (iEMSs) 2010 International Congress on Environmental Modeling and Software Modeling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada,2010
[14]. J. Lu , T .Guo-an The Spatial Distribution Cause Analysis of Theft, Crime Rate Based on GWR Model, Multimedia Technology (ICMT), 2011 International Conference 26-28 July 2011, Hangzhou ,IEEE,2011
[15]. Matthew J. S. Windle, George A. Rose, Rodolphe Devillers and Marie-Josée Fortin "Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic ,"ICES journal of Marine science vol 67, pp 145-154,2009
[16]. Eduardo Francisco (2008) , A Consumer Income Predicting Model Based on Survey Data: An Analysis Using Geographically Weighted Regression (GWR), Latin American Advances in Consumer Research, vol 2, pp 77-83
[17]. S.Rong-Kang ,S. Yi-Shiang , M.Kuo-Chen "Using geographically weighted regression to explore the spatially varying relationship between land subsidence and groundwater level variations: A case study in the Choshuichi alluvial fan, Taiwan" IEEE, 2011
[18]. Christian Growitsch, Tooraj Jamasb and Michael Pollitt "Quality of Service, Efficiency and Scale in Network Industries: An analysis of European electricity distribution," IWH-Diskussionspapiere ,2005
[19]. An Introduction to Spatial Autocorrelation Analysis with GeoDa Luc Anselin Spatial Analysis Laboratory Department of Agricultural and Consumer Economics University of Illinois, Urbana-Champaign.June 16, 2003