

Urban land value, Accessibility and Night Light Data: A Case Study in China

Sumeeta Srinivasan, and Deng Yu

Abstract

In this paper, we use night light imagery data at two scales – the city and within the city. Since many of the urban characteristics of cities including traffic and paved area are correlated with each other we derive components for our sample of 286 cities in China. We then correlate these measures for average DN values from night light imagery at the center and periphery of the city (using buffers of 5km and 20km). We also test at four points of time in the period 2002-2012 to assess changes over time if any. We then focus on Beijing and assess the correlation between night light imagery derived DN values, land price and an accessibility indicator. In deriving spatial accessibility we use micro scale indicators such as time distance to the city center, and employment centers by transit and automobile. We use spatial regression and Geographically Weighted Regression (GWR) models.

S. Srinivasan (Corresponding author)
China Project, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA
Email: sumeeta@seas.harvard.edu

D. Yu
Institute of Geographical Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing, China,
Email: rain00788@163.com

1. Introduction

Urbanization is a worldwide process that has had profound effects on the environment and ecology. It is estimated that by 2050 the population will increase from 3.4 billion in 2009 to 6.3 billion (UN, 2010). The UN predicts that Asia is projected to see its urban population increase by 1.7 billion. In this context it is especially useful to measure the effects of growth on the livability of cities as measured by accessibility, housing price and exposure to traffic and other indicators or urban consumption. The IPAT identity suggests that population, affluence, and technology which are all consequences of urbanization are reflected in human impact as measured by energy and emissions intensity (Ehrlich and Holdren, 1971). Lenzen et al (2006) find that “urbanity”, which is defined as a dummy variable for households located in a large metro (above 1 million inhabitants), is positively correlated with per capita expenditure. Households located in large cities were found to have, on average, a 10% higher per capita energy requirement than those located in smaller urban areas. Lenzen et al suggest that instead of a uniform Kuznet’s curve the effect of increasing income varies considerably across countries, even when controlling for socioeconomic and demographic variations.

The National Geophysical Data Center within the U.S. National Oceanic and Atmospheric Administration (NOAA) makes available the Global DMSP-OLS Nighttime Lights dataset, which reports nighttime light intensity for each year 1992 to 2012. Many studies have used these data to investigate urban growth (Schneider et al, 2010) and economic activity (Ghosh et al, 2010). Studies have also looked at energy consumption by predicting electric power consumption in China and other countries using satellite data derived from the DMSP-OLS (Elvidge et al, 1997; Silvana et al, 2005; He et al, 2012). These studies have found that DMSP data can be an accurate proxy for population, electricity and GDP. Our study goes a step further in evaluating consumption as measured by traffic, gas, LPG, water, paved area and electricity used per capita at the cross-city scale and correlating it with nighttime light intensity. Within the city we use measures of accessibility and land value to look for more fine grained differences. Such estimates may be especially useful in situations where data on infrastructure are not routinely collected by governments in the global South.

2. Data and Methods

2.1 Data

Our study uses Version 4 of DMSP-OLS stable light data (2002-2012) which is part of the latest average DN data series and was obtained from the National Geophysical Data Centre (NGDC) website (National Geophysical Data Centre 2006). In this image, the data values which are annual visible band digital numbers range from 1-63 and it provides annual global composites of stable nighttime lights derived from the visible band. Background noise in the data is represented using zero while areas with no cloud-free observations are denoted by the value of 255. The data has a spatial resolution of 30 arc-seconds (approximately 1 km at the equator and 0.8 km at 40°N). The time series of DMSP/OLS nighttime lights for the period of 2002–2012 were based on three individual sensors F15 (2002, 2004, 2006), F16 (2008).and F18 (2010, 2012). There are yearly variations and differences among sensors therefore we do not calculate changes across different sensors and our models use data from the same sensor.

The data for the cities were obtained from GIS point data for 287 major cities from the China Data Center at the University of Michigan. We estimate the average and standard deviation of DN values for 5km and 20km buffers. Since the value at the point representing the city would usually be high given it was centered in the downtown we used buffers to understand how much variation there was in DN values. We use panel data for 287 Chinese cities for the period 2002-2012. The data were obtained every two years and include the years 2002, 2004, 2006, 2008, 2010 and 2012. Measures for GDP, population, downtown population, traffic, paved road, construction, water and gas use were obtained from the China Statistical Yearbooks. Not all the energy measures were consistently available for all the years and many measures were highly correlated with each other and therefore we used a dimension reducing technique to derive components. Land price, accessibility and nightlight corrected data for Beijing were obtained from the China Academy of Sciences.

2.2 Methods

The aim of Principal Components Analysis (PCA) is to find a new set of multivariate variables that are uncorrelated and explain as much variance as possible (Shlens, 2014). If all the variables are in one data matrix, the aim of dimension reduction techniques like PCA is to find the best matrix created with fewer variables (lower rank) that explains the original data. If the data matrix X has each variable in a column (per capita water, LPG, paved roads

length) and each observation (city) in a row then the PCA is a "matrix decomposition" where the columns of U are orthogonal (left singular vectors), the columns of V are orthogonal (right singular vectors) and D is a diagonal matrix (singular values). U and V are the components that may include the several variables from the original data X .

$$X = UDV^T \quad (1)$$

The local Moran (Anselin, 1995) is as follows:

$$I_i = z_i \sum_j w_{ij} z_j \quad (2)$$

This is analogous to the global Moran's I , the observations z_i and z_j are deviations from the mean, and the summation over j is such that only neighboring values j are included. For ease of interpretation, the weights w_{ij} may be in row standardized form, though this is not necessary, and by convention, $w_{ii} = 0$. The results are mapped as "low-low" or "high-high" cases that contribute to positive autocorrelation and "low-high" or "high-low" cases that contribute to negative autocorrelation. For example, "high-high" cases are ones where the value of i is high and neighboring values are also high. While all cases are in one of the four quadrants the cases that are statistically significant are colored in the maps.

In a simple multivariate regression model, we model the relationship between one dependent variable and one or more independent variables. The mathematical model underlying multiple regression is

$$y = Xb + e \quad (3)$$

so that the value of the independent variable at each city is modeled as the sum of a constant, a sum of products of each independent variable value X , and a coefficient b , along with an error term e . The model is fitted to the observed data using a least squares regression procedure, which ensures that the sum of the squared errors at all locations in the data set is minimized. The least squares estimates of the regression coefficients can be calculated for this model as follows:

$$b = (X^T X)^{-1} X^T y \quad (4)$$

Geographically weighted regressions (GWR) allow the regression coefficients in the model to vary from place to place. Thus, in the same way that the variables themselves change from place to place, GWR assumes that the relationships among them may vary from place to place.

$$b = (X^T W X)^{-1} W X^T y \quad (5)$$

where W is a diagonal matrix of weights for each case in the data set. Off diagonal elements in W are zero, and the diagonal elements have values $\sqrt{w_{ii}}$

where w_{ii} is the weight we wish to associate with each observation. In GWR, the elements in W are based on the spatial association between the location at which the local regression is performed and the available points at which data are available, so that this is a locally weighted regression. Fotheringham et al. (2003) suggest using either a Gaussian or biweight kernel function at each location to assign weights to nearby observations in the data set. The bandwidth can be a fixed value chosen by the user or it can be an adaptive which is different at every point. A complex method for automatically choosing each local bandwidth has been implemented based on running multiple models at each location, omitting one point at a time and setting the bandwidth such that it produces the best estimates of the omitted location (O'Sullivan and Unwin, 2014).

3. Results

3.1 Dimension reduction of energy consumption across cities

As expected many of the variables that measured consumption in 2012 were highly correlated with each other. A principal components analysis of the variables revealed three distinct components which explained about 90% of the variance (Table 1). The first component explained about 62%, the second 23% and the third only 6% of the variance with drops thereafter. The first component loaded highly in each year on downtown population, total traffic, paved road length, construction expenditure, total traffic, per capita consumption of gas, and LPG. The second component loaded highly on per capita consumption of water and electricity. Total population and per capita GDP were also tested and the former correlated highly with downtown population and the latter was distinct in a separate third component. Since the components were likely to be correlated with each other, a Varimax rotation was used to generate components without the orthogonal restriction. This resulted in two distinct components in both 2002 and 2012. We also tested intervening years and the results were similar in all years. It should be noted that the variables were not reported exactly the same in every year. In 2002 most variables were only measured for the "downtown" portion of the city. However, the patterns of correlation remained similar with Component 1 loading highly on population, gas consumption, traffic, and paved road. We refer to this component in the paper as the "urbanity" component. The second component loaded highly on water and electricity consumption per capita. The third component and thereafter were yielded a different set of loadings for each year. In all years the first two components explained over 60% of the variance with a large drop thereafter and therefore we only use the

first two components. Table 1 shows the loadings for each variable for three of the years. The first two components explain at least 72% of the variance in 2002 increasing to 85% in 2012. The table only reports loadings higher than 0.4.

As seen in the mapped values by city in Figure 1 there is some variation over both time and space in China. GDP per capita values have increased over time and there are higher values in 2012 than in 2002 which suggested clustering of higher income as measured by GDP in certain regions of China near Beijing and along the coast. Urbanity (component 1) and electricity and water use (component 2) also appear to cluster in the larger cities such as Beijing and Shanghai with a few industrial cities also showing high values.

Table 1. Loadings of variables on components from principal components analysis with varimax rotation for 2002, 2010 and 2012 data

	2002		2010		2012	
	Comp 1	Comp 2	Comp 1	Comp 2	Comp 1	Comp 2
Downtown population	0.93		0.95		0.91	
Gas consumption per capita			0.92		0.93	
Gas consumption per household in downtown	0.90					
LPG consumption per capita			0.69	0.40	0.49	
LPG consumption per household in downtown	0.71	0.51				
Total Traffic			0.76		0.92	
Downtown Traffic	0.49					

Paved Road			0.90		0.85	
Paved road in downtown	0.92					
Construction Expenditure					0.79	
Electricity consumption per capita		0.70		0.94		0.96
Water consumption per capita		0.83		0.93		0.96
N	161		240		256	
Proportion of variance explained	48%	24%	53%	29%	61%	24%

3.2 Night light across cities

The city location data were available as points are located at the city center where DN values are always high. Therefore we used buffers of 5km and 20km to delineate polygons around each city. For this paper we report results at the 20km buffer since there was more variation in DN numbers at this scale in our sample. Figure 2 shows that there are distinct clusters of high average DN values and low average DN values at both scales. For the 20km buffer the red clusters of high surrounded by high mean DN values include Beijing, Shanghai and more recently Guangdong which was an outlier in before 2010. There is distinct clustering of high values along the coasts with high surrounded by low average values in Chengdu, Chongqing and other interior cities. It is also interesting that the standard deviations of the DN values in the 20km buffer are relatively high in northern cities in China as compared to southern cities. This could suggest either more concentrated urbanization within the northern cities but also more inequity in infrastructure within these cities between the downtown and the peripheral suburbs.

3.3 Night light data, income and consumption across cities

Regressions of DN value on the Urbanity, electricity and water consumption and GDP per capita suggest that for the 20km buffer around the cities urbanity is a significant predictor of average DN value even after accounting for GDP and spatial dependency (Table 2). This relationship is a bit weak in 2012 data and the effect of electricity and water consumption seems to be less significant in predicting DN values in the more recent data. The presence of a significant spatial lag suggests regional clustering. Table 3 shows that an increase in the urbanity component is accompanied by a fall in the standard deviation of DN values in the buffer on average suggesting that at this scale the high consumption cities have urbanized within the entire 20km buffer. Figure 3 shows that in contrast to 2002 when urbanity, electricity and water consumption as well as GDP per capita were positively correlated with the DN value the 2012 coefficient signs vary by region. In 2012 the cities where urbanity had high and positive coefficients were in the northeast near the Korean peninsula. By contrast, in the south near Guangzhou the urbanity coefficient was negative suggesting that increased urbanity was associated with lower average DN values in the 20 km buffer. This southern swath of cities also showed a similar trend with respect to electricity and water consumption. GDP per capita also showed regional differences in significance in predicting DN value though it was still positively correlated. Overall the R^2 values were higher in 2002 as compared to 2012. Also in 2012 R^2 values for the model was highest in the northern cities of China and near Shanghai. This suggests that when controlling for spatial effects DN values may not always be positively correlated with urban consumption as measured by the components.

3.4 Night light data, income and consumption at the city scale

The first principal component for several variables in the study area including time to city center, time to central business district, time distance to employment centers, time to industry centers, time to subway station, time to highway explained 84.1% of variance and we use this component as a measure of accessibility for locations within the Beijing metropolitan area. Even after accounting for spatial dependency DN values were significantly positively correlated with land price and accessibility within the city of Beijing (Table 4). Figure 6 shows that these coefficients are not uniform

in their distribution across the city. Indeed the most interesting results appear to be for the periphery where there were negative coefficients in some locations suggesting that improved accessibility or higher land prices were negatively correlated with DN values. Along major highways leading outside the city the positive correlations with accessibility held but between them especially on the northern peripheries negative coefficients were evident. Land values also had negative coefficients in the northeastern peripheries but most locations tended to have positive correlations with DN values.

Figure 1 Mapped values for GDP per capita, Urbanity component and Electricity and Water consumption component

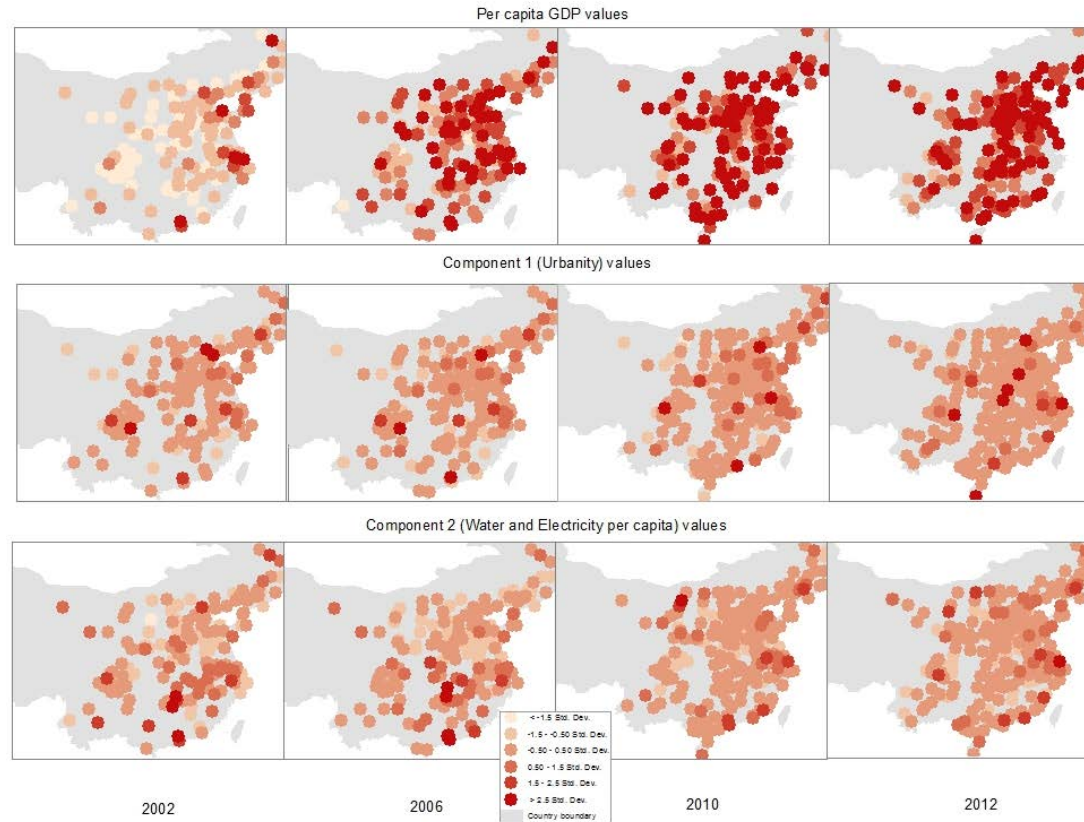


Figure 2 Local Moran's I for DN values in 5 and 20 km buffer

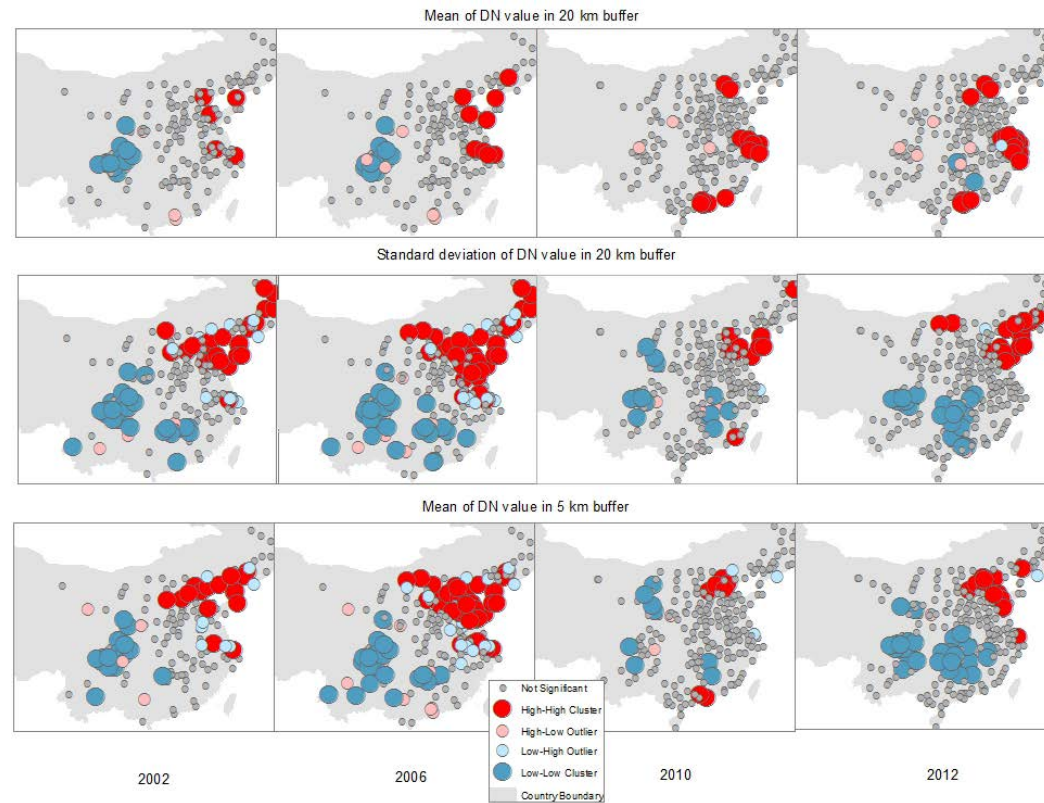


Table 2 Regression with dependent variable mean DN value in 20km buffer across cities in China

	2002 Coeff		2006 Coeff		2010 Coeff		2012 Coeff	
		(lag)		(lag)		(lag)		(lag)
Constant	6.4**	2.1**	7.6**	3.2	6.6**	3.4**	9.7**	1.9
Urbanity Component	5.8**	5.9**	7.7**	7.7**	7.2**	7.0**	1.3	1.4*
Electricity & Water component	2.4**	2.6**	2.3**	2.4**	2.5**	2.6**	-1.4	-0.9
GDP per capita	7e4**	6e4**	2e4**	2e4**	e4**	e4**	3e4**	2e4**
Spatial Lag	-	0.4x	-	0.4x	-	0.4**	-	0.48**
Adjusted R ²	0.80	0.81	0.75	0.76	0.75	0.78	0.24	0.34
Log likelihood	-487	-484	-603	-601	-743	-730	-983	-969
N	162		190		240		256	

x is significant at 10% * is significant at 5% and ** is significant at 1%

Table 3 Regression with dependent variable standard deviation of DN value in 20km buffer across cities in China

Variable	2002 Coeff		2006 Coeff		2010 Coeff		2012 Coeff	
		(lag)		(lag)		(lag)		(lag)
Constant	10.2**	1.9	9.4**	1.6	8.8**	4.8**	12.9**	4.8**
Urbanity Component	-0.5	-0.37	-0.4	-0.4	-0.6**	-0.6**	-0.8**	-0.7**
Electricity & Water component	-0.4	-0.05	-0.9**	-0.5	0.7**	-0.5x	-0.5	-0.3
GDP per capita	3e4**	3e4**	2e4**	e4**	e4**	e4**	5e3**	4e4**
Spatial Lag	-	0.65**	-	0.6**	-	0.3**	-	0.6**
Adjusted R ²	0.2	0.3	0.3	0.3	0.3	0.3	0.1	0.3
Log likelihood	-462	-458	-534	-530	-679	-672	-715	-693
N	162		190		240		256	

* is significant at 5% and ** is significant at 1%

Table 4 Regression with dependent variable DN value for Beijing in 2012

Variable	Coeff	Coeff (Lag)	Coeff (Error)
Constant	243.2**	169.3**	198.2**
Land Price	25e-5**	8.7e-5**	8.7e-5**
Accessibility	0.02**	0.015**	0.016**
Spatial Lag	-	0.32**	
Lambda	-	-	0.93**
Adjusted R ²	0.17	0.47	0.61
Log likelihood	-41027	-38996	-38232
Likelihood ratio test		4063**	5589**
N	9166		

** is significant at 0.001%

Figure 3 GWR coefficients for mean DN values in 20 km buffer in 2002 and 2012

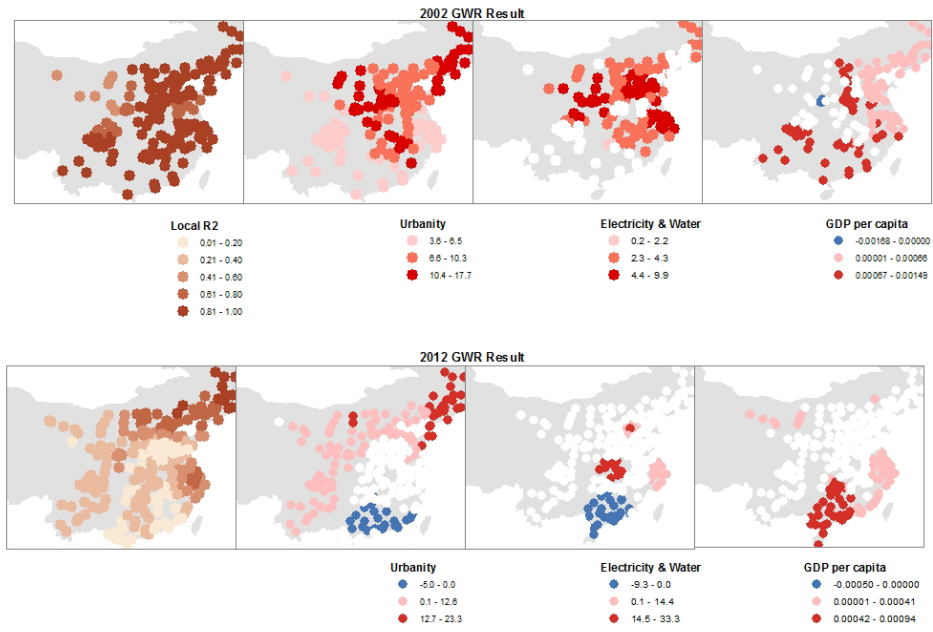
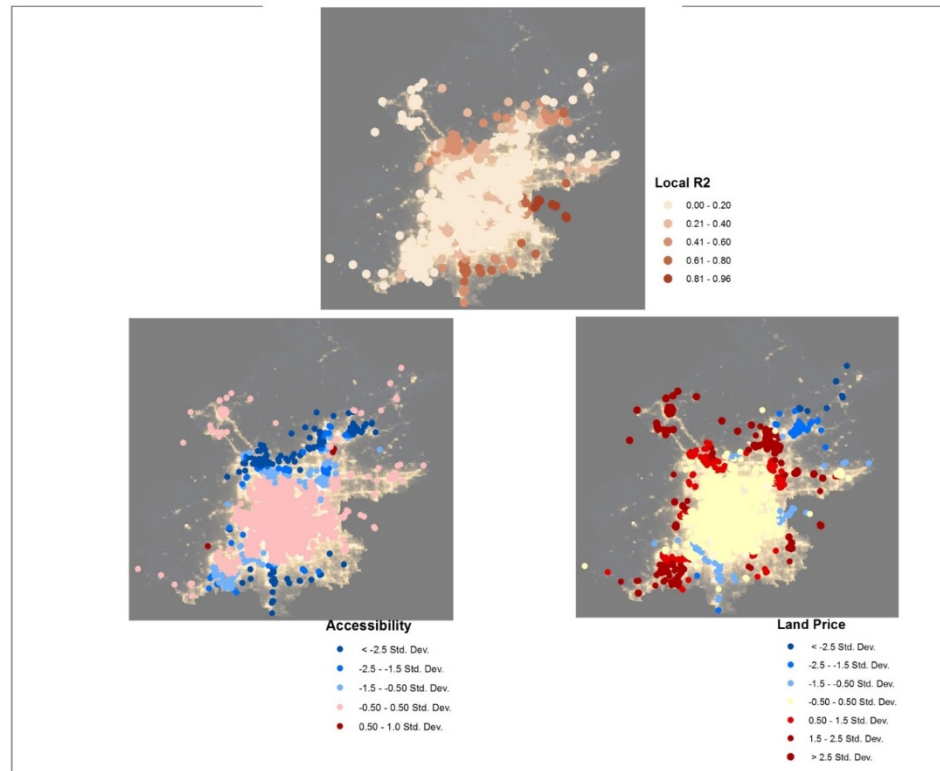


Figure 6 GWR coefficients for DN values in Beijing in 2012



4. Conclusions and Future Research

Urbanization is expected to continue in China for the near future. Eliot et al (2014) note in their province level study that urbanization may not deliver the reductions in energy intensity. Our results appear to show at both the across city and within city scale urban consumption is spatially variable and that it may not have a simple relationship with popular proxies such as night light data. Indeed night light data appear to be negatively correlated with some of the urban indicators in more recent data. Clearly more research is needed to confirm this for other cities since Beijing is likely to be a special case within China. While GDP per capita continues to be a positive predictor of the presence of night light it is less significant in recent data and is also prone to regional trends in more recent data. Thus it is likely that as countries grow wealthier night light data may not be sufficient as proxies for energy consumption, urbanity and income.

Acknowledgment

The authors would like to thank Xiao Yu for her research assistance in gathering data from the China Statistical Yearbooks.

References

- Anselin, L., 1995. Local indicators of spatial association - LISA. *Geographical Analysis*, **27**: 93-115.
- Ehrlich, P. R. and J. P. Holdren. (1971). "Impact of Population Growth". *Science*. **171**: 1212–1217
- Elliot, R. J. R., P. Sun and Tong Zhu, (2014), Urbanization and Energy Intensity: A Province-level Study for China, Department of Economics Discussion Paper 14-05, University of Birmingham, UK,
- Elvidge C, Baugh K, Kihn E et al., (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, **18**: 1373–1379.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2003). Geographically weighted regression. John Wiley & Sons, Limited.
- Ghosh, T., Powell, R. L., Elvidge, C. D., Baugh, K. E., Sutton, P. C., and S. Anderson. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, **3**: 148–161.
- He C., Ma, Q., Li T., Y. Yang, and Z. Liu, (2012) Spatiotemporal dynamics of electric power consumption in Chinese Mainland, *Journal of Geographical Science* **22**: 125-136
- Lenzen, M., Wier, M., Cohen, C. Hayami, H, S. Pachauri and R. Schaeffer. (2006). "A comparative multivariate analysis of household energy requirements in Australia, Brazil, Denmark, India and Japan" *Energy*, **31**: 181-207
- O'Sullivan, D., & Unwin, D. (2014). *Geographic information analysis*. John Wiley & Sons.
- Schneider, A., Friedl, M. A., and D. Potere. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment*, **114**: 1733-1746.
- Shlens, J. (2014), A Tutorial on Principal Component Analysis, arxiv.org/pdf/1404.1100v1.pdf Last accessed August 2014.
- Silvana A, Gilberto C, Antonio M et al., 2005. Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Computers, Environment and Urban Systems*, **29**: 179–195.

UN (2010). World urbanization prospects, the 2009 revision. New York: United Nations.