SPATIAL VARIATION OF URBAN BUILDING ENERGY ANALYSIS

Wei Tian[1,2], Song Yang[1,2], QingXin Meng[1], Lai Wei[1,2]
[1] College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin, China
[2] Tianjin Key Laboratory of Integrated Design and On-line Monitoring for Light Industry & Food Machinery and Equipment, Tianjin 300222, PR China

ABSTRACT

Building energy use in various parts of a city can be visualized spatially using GIS (geographical information system) programs. When doing so, one would expect spatial variations due to socio-economic and/or physical factors. This paper investigates the local spatial variations of domestic energy use intensity (both gas and electricity) across London. Three local regression methods are considered: geographically weighted regression (GWR), mixed model, and Bayesian hierarchical computation. The results indicate that the Bayesian hierarchical model can produce reliable results compared to the other two frequentist methods. The 'Stan', a new full Bayesian statistical inference with Hamiltonian Monte Carlo (HMC), can reduce computational cost compared to the commonly used sampling methods (random walk Metropolis or Gibbs sampling). The Stan Bayesian method is expected to have more widely applications in energy assessment of urban buildings.

INTRODUCTION

It is apparent that building energy in cities can be visualized in various spatial scales. Then it is important to understand the variations of energy use for buildings in terms of these spatial distributions in order to make sustainable policies for reducing energy use and associated carbon emissions in urban environment. To this end, the Geographical Information System (GIS) is an important tool for spatial data visualization and spatial analysis for urban building energy use.

There is an increasing number of research papers on using GIS in assessing the patterns of energy use for buildings in urban context (Calderón et al., 2015; Tian et al., 2014a; Choudhary and Tian, 2014). Tian et al. (2014a) used spatial error model to investigate the spatial relationships between energy use and council tax bands in London. Howard et al. (2012) applied a robust multiple regression method to estimate energy use intensity by end use in terms of building function in New York City. Choudhary and Tian (2014) applied Bayesian spatial conditional autoregressive (CAR) models to explore the patterns of gas use at MOSA (middle layer super output area) scale in London. Tian and Choudhary (2012b) implemented spatial lag models to obtain energy use intensity in terms of building types in London. These previous studies have helped to understand the trends of energy use, more specially using spatial analysis methods, since there is only a few studies (Tian and Choudhary, 2012b; Tian et al., 2014a) on using these formal statistical spatial analysis approaches in the field of building energy. An important feature for these spatial analysis methods is to consider spatial dependence or spatial autocorrelation of energy use (Tian et al., 2014a; Fischer and Getis, 2010). To comprehend further spatial characteristics of building energy use, one question that needs to be asked, however, is whether there are spatial variations for energy use intensity per dwelling types or other criterion. To the best of the knowledge, there is still no research on assessing local variations of energy use per dwelling type (or other measures) using local spatial regression methods in cities.

Therefore, this research will use three local spatial regression methods to explore the patterns of energy use in urban environment and compare the advantages and disadvantages of these local methods. These three local methods are geographically weighted regression (GWR), mixed model, and Bayesian hierarchical computation, which will be described in detail in the section of "DATA and METHODS". Compared to global methods (such as spatial lag or spatial error models) that may mask spatial variations, the local regression methods can make full use of available data and account for local variations in urban buildings to obtain the locally varying coefficients for energy use intensity per household or other criteria. London has been chosen as a case study in this research since there are detailed data available for building energy use in several spatial scales collated by Greater London Authority (London Datastore, 2015).

This paper is structured as follows. Firstly, the energy data and the corresponding explanatory variables are presented in GIS map or numerical statistical measures. The next section is to introduce four statistical methods used in this analysis, including the ordinary least square (OLS), GWR, mixed model, and Bayesian hierarchical model. The result section is to compare the results (estimated energy use intensity: MWh/property) from these four methods for gas and electricity, respectively.

DATA AND METHODS

Data

The data used in this study includes the number of four dwelling types, domestic gas/electricity data, and digital spatial boundaries in London. Dwelling data are classified as four types based on UK census 2011, i.e. detached, semi-detached, terraced, and flats (London Datastore, 2015). Three spatial scales used in this paper are LA (Local authority), MSOA (middle layer super output area), and LSOA (lower layer super output area) (ONS, 2015). MSOA and LSOA were designed for small area statistics in the UK. In London, the numbers of MSOA and LSOA areas are 983 and 4835, respectively. Average household numbers in MSOA and LSOA are 3323 and 676 in London, respectively. There are 32 LAs and 1 City of London in Greater London. Since the domestic energy for either gas or electricity from the City of London is less than 0.2% of total London domestic energy, the discussion in this paper will ignore this area. In the City of London, there are a large number of commercial offices, but only a few of domestic properties.

Figure 1 shows the spatial distributions of the percentage of these four types of dwellings at LSOA level in London. The independent variables (also called explanatory variables) used in this paper are the numbers of four types of dwellings: detached, semi-detached, terraced, and flat properties. The detached and semi-detached houses are more clustered together in outer London, whereas flats are more concentrated in inner London. The spatial patters for detached dwellings are similar to those for semi-detached residential buildings in London. The characteristics for the terraced houses are quite different from the other three types of dwellings and the terraced properties are more distributed in northeast of London.

Table 1 Summary of domestic energy at three spatial scales in London (unit: kWh/property)

ENERGY	LA	MSOA	LSOA
Gas			
10% percentile	9803	8035	7710
Mean	12412	12775	13405
Median	12583	12615	12657
90% percentile	14967	17476	18610
Electricity			
10% percentile	3463	3204	3034
Mean	3956	3982	4280
Median	3993	3914	3885
90% percentile	4321	4774	4970

Notes: LA, local authority; MSOA, middle layer super output area; LSOA, lower layer super output area

Table 1 shows the descriptive statistics of energy use intensity for domestic buildings (kWh/property) at three spatial scales in London. The spread of domestic energy becomes larger when the spatial scales change from LA to LSOA. It is interesting to note that the differences between the mean and median energy data becomes larger from LA to LSOA. This indicates that the distributions of energy data become more asymmetric with a decrease of spatial scale. As a result, the statistical characteristic of energy data varies a lot at different spatial units. Moreover, it can be observed that the spread for electricity use intensity is smaller in comparison with gas use.

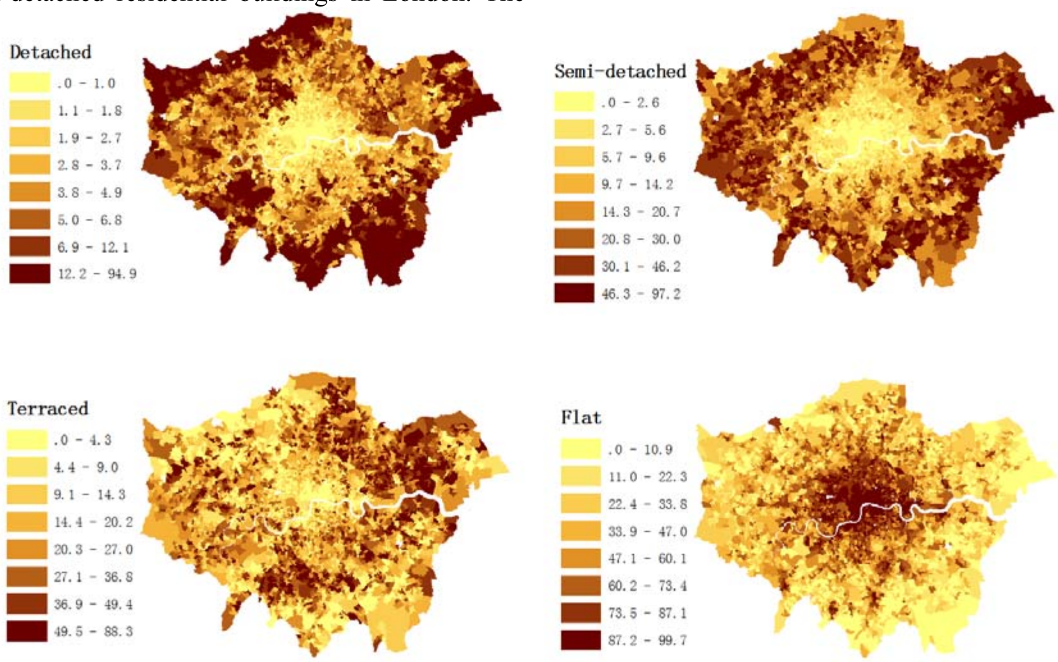
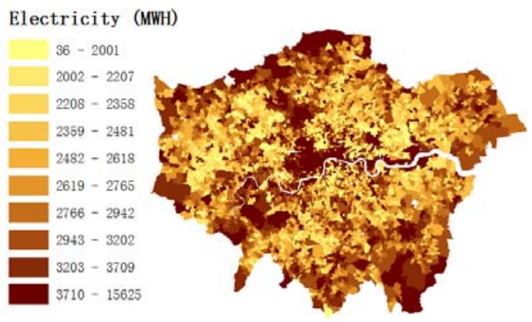
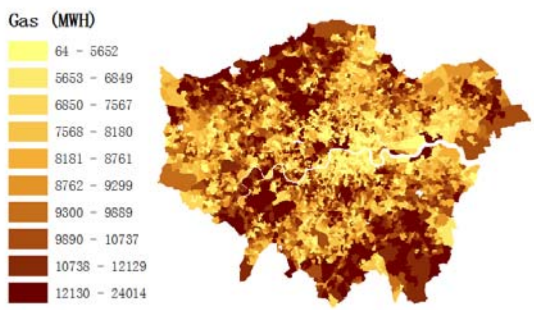


Figure 1 Percentage of four dwelling properties (detached, semi-detached, terraced, and flats) at LSOA spatial scale in London.



(a) Electricity use



(b) Gas use

Figure 2 Electricity and gas use at LSOA spatial scale in London.

Figure 2 shows the spatial distribution of electricity and gas energy at LSOA scale in London. There are some similarities in terms of spatial patterns for electricity and gas. The outer London has high energy use for both electricity and use. In contrast, the electricity use is much higher than the gas use in inner London.

Methods

Three local regression methods (GWR, mixed model, and Bayesian computation) are used in this study to obtain locally varying energy use per dwelling type. GWR stands for geographically weighted regression, a non-parametric technique of locally weighted regression for curve-fitting and smoothing application (Fischer and Getis, 2010). This method has close relationship with the global ordinary least square (OLS) method, which can be written in matrix notation,

$$Y = X\beta + \varepsilon \qquad (1)$$

where Y is a n × 1 vector of dependent variable, X is a n × k matrix of independent variables, β is a k × 1 vector of regression coefficients, and ε is a n × 1 vector of random errors. The n is the number of energy data and k is the number of independent variables, i.e. the number of spatial units dependent on the choice of spatial scales in this case. Y is the domestic energy data at different spatial units in this case and X is the numbers of four-type dwelling properties as described in the section of "Data". β is the energy use estimated by minimizing the sum of squared residuals using the OLS method,

$$\beta = [X^T X]^{-1} X^T Y \qquad (2)$$

where T denotes the matrix transpose and the remaining items have been explained in Eq. (1).

If using the GWR method (Fischer and Getis, 2010), the regression coefficient can be computed using the following equation,

$$\beta_i = [X^T W_i X]^{-1} X^T W_i Y \qquad (3)$$

where W_i is a n × n matrix of local spatial weight matrix for the i_{th} location and the other items are the same as above. This local spatial matrix is different from the one from the global spatial method (such as spatial error or spatial lag models) in which the spatial weight matrix W is obtained using the distance-based or contiguity methods, but does not vary with the location. In contrast, the local weight matrix is computed from a kernel function, which puts more weight on the locations that are closer to a given place. As a result, the non-stationary patterns for regression coefficients are expected using these changing matrixes. The kernel function has a bandwidth parameter to adjust how many spatial units will be included for calculating the estimated coefficients in GWR. Two methods are usually implemented, fixed and adaptive (Fischer and Getis, 2010). The fixed approach uses only one constant value irrelevant to the number of spatial units for a specific distance, whereas the adaptive kernel function can change the distance based on the density of data points. In this case, the adaptive kernel method is applied using the cross-validation approach for minimizing the prediction error.

The mixed models allow the parameters to be changed at more than one level. This method has several names, such as multilevel model, hierarchical linear model, nested models, and random-effects model. This model can be written after adding a new item from Eq. (1),

$$Y = X\beta + Zu + \varepsilon \qquad (4)$$

where u is a k × 1 vector of regression coefficients for random effects, Z is a n × k design matrix of variables including group level information. The random effect can represent the local influences in a specific spatial unit and then the regression coefficients would change with the locations in which the relationship may be similar among inputs and outputs in a nearby area. Therefore, the mixed model can take into account the variations of regression coefficients, i.e energy use per dwelling type in this case. Note that the results from the mixed model are discontinuous since the coefficients would change based on a specific spatial area containing a number of small spatial units. In contrast, the GWR can produce continuous coefficients as the kernel function can be moved in a continuous way.

Bayesian computation method has been gradually used in building energy analysis (Tian and

Choudhary, 2012a; Heo et al., 2014; Tian et al., 2014b). The Bayesian method used in this paper is an extended model based on the mixed model, called Bayesian hierarchical modelling. Compared to Bayesian linear model, this method needs two concepts for computing posterior distributions, estimated energy use intensity in this case. One is hyper-parameter that refers to the parameters of prior distribution and the other is hyper-prior to define a distribution of a parameter of prior distribution for considering random effects. Hence, the u in Eq. (4) is a random variable with a probability distribution if using Bayesian method, whereas it is assumed as a single constant value in the frequentist setting. The Stan (Homan and Gelman, 2014), a new probabilistic programming language implementing full Bayesian statistical inference with Hamiltonian Monte Carlo (HMC), is used for Bayesian hierarchical computation to obtain the locally varying energy use intensity. HMC is a Markov chain Monte Carlo (MCMC) algorithm that avoids the random walk behaviour and sensitivity to correlated parameters. These features allow it to converge to high-dimensional target distributions much more quickly than simpler methods such as random walk Metropolis or Gibbs sampling.

All the statistical analysis is conducted using R program (R Development Core Team, 2015), a software environment for statistical computing and graphics. Table 2 lists the R packages used in this paper for global and local spatial analysis.

Table 2 R packages used for statistical analysis

METHOD	PACKAGE	REFERENCE
Linear model	R base	(R Development Core Team, 2015)
GWR	R GWmodel	(Lu et al., 2014)
Mixed model	R lme4	(Bates D, 2015)
Bayesian hierarchical	Rstan (Bayesian) R parallel (multi core)	(Stan Development team, 2015)

Note that the GWR method is only used for MSOA spatial scale because the computational cost is too high for LSOA level due to a larger number of data. The Bayesian hierarchical models are used for both MSOA and LSOA to provide robust analysis. The prior distributions for Bayesian computation are derived from several sources, including previous research (Calderón et al., 2015; Hamilton et al., 2013), the results from the GWR and Mixed models in this research. The cases used in this paper are listed in Table 3.

The global ordinary least square (OLS) method is used for the purpose of comparison only. The input data is at LSOA scale and the regression coefficients obtained from this method are non-spatial. For the GWR method, the explanatory variables are at MSOA scale and this method is based on the adaptive kernel approach. As a result, the regression coefficients are continuous across the whole map. The computation of the mixed model is conducted at LSOA scale. The spatial group for the mixed model is LA (local authority) and the corresponding regression coefficients are the same spatial scale (discontinuous at LA). As for the Bayesian computation using the Stan language, two spatial scales for the input data are LSOA and MSOA, named as 'Baye1' and 'Baye2', respectively. The regression coefficients are also discontinuous at LA scale in London for both the Baye1 and Baye2.

Table 3 Summary of spatial regression methods used in this study

METHOD	DATA	SPATIAL GROUP	RESULT
Linear	LSOA	-	No spatial
GWR	MSOA	Adaptive	Continuous
Mixed	LSOA	LA	Discontinuous
Baye1	LSOA	LA	Discontinuous
Baye2	MSOA	LA	Discontinuous

Notes: LA, local authority; MSOA, middle layer super output area; LSOA, lower layer super output area

RESULTS AND DISCUSSION

Spatial variation for domestic gas

Table 4 shows the summary of domestic gas intensity (MWh/property) for various estimated dwelling types using global and local regression methods. The "REFS" column shows the gas use intensity based on the data in the National Energy Efficiency Data (NEED) framework (DECC, 2015). Note that these NEED values are used as reference only because the methods obtained from these values are quite different from the approach used in this paper. The NEED data are obtained from various sources using the data match method, which is more focused on individual dwelling properties (Wyatt, 2013; Hamilton et al., 2013). In contrast, the method used in this study is based on the areal data from UK DECC (Department Energy and Climate Change) and 2011 census aggregated data, which can better reflect the characteristic of dwelling building stock in one small area, but not at individual building level.

First, the central tendency of gas data is analysed using mean or median values. Table 4 indicates that the mean from all the regression methods is higher than the NEED values for detached and semi-detached houses. The mean values estimated from this analysis is lower than the NEED values for terraced houses and flats. These differences are because the definitions of mean values are different from the NEED and the values calculated from this analysis. The NEED values can be interpreted as average gas use for various dwelling types. In contrast, the mean values from this paper are the average gas data from all the 32 LA (local authorities) and City of London, i.e. the average data from 33 values. The mean values used here does not account the differences of dwelling numbers for the same type of residential buildings in different LAs, but focus on spatial influences. Apparently, the differences of property numbers are very significant at LA levels in London.

Table 4 Summary of coefficient values from global and local regression methods for gas use (MWh/property)

DWELLING	ITEM	REFS	GLOBAL	LOCAL			
		(DECC, 2015)	LINEAR	GWR MSOA	MIXED LSOA	BAYE1 LSOA	BAYE2 MSOA
Detached	Min	-	-	-45.5	12.7	10.0	10.0
	Q1 [a]	-	-	24.0	29.1	25.8	27.1
	Median	25.2	-	32.0	32.0	30.5	32.0
	Mean	26.0	30.9	31.2	32.4	33.1	32.5
	Q3 [b]	-	-	40.0	34.1	40.7	39.0
	Max	-	-	109.2	44.3	60.0	60.0
Semi-detached	Min	-	-	-45.0	8.4	8.0	5.0
	Q1	-	-	17.5	17.8	18.4	17.1
	Median	18.5		20.5	20.9	19.2	21.0
	Mean	19.5	18.3	21.6	22.9	22.1	24.2
	Q3.	-	-	25.7	24.0	22.5	25.0
	Max	-	-	80.9	44.5	50.0	60.0
Terraced	Min	-	-	-7.2	9.43	4.7	1.0
	Q1	-	-	11.5	11.26	9.8	11.0
	Median	13.8 [c] 15.5 [d]	-	13.6	13.70	11.7	13.3
	Mean	15.0 [c] 16.5 [d]	13.6	14.2	14.47	13.4	14.2
	Q3.	-	-	16.8	16.80	16.9	16.9
	Max	-	-	27.9	26.34	30.0	27.5
Flat	Min	-	-	-5.1	4.05	4.1	3.9
	Q1	-	-	5.7	6.09	6.2	5.9
	Median	8.2 [e] 9.6 [f] 15.8 [g]	-	6.9	6.61	7.7	6.5
	Mean	9.0 [e] 11.1 [f] 16.8 [g]	7.3	7.0	6.63	7.9	6.6
	Q3.	-	-	8.1	7.23	9.2	7.4
	Max	-	-	18.5	8.94	13.4	9.7

Notes: a, Q1, 25^{th} percentile, the first quartile; b, Q3, 75^{th} percentile, the third quartile; c, middle terrace; d, end terrace; e, purpose built flats; f, converted flats; g, bungalow

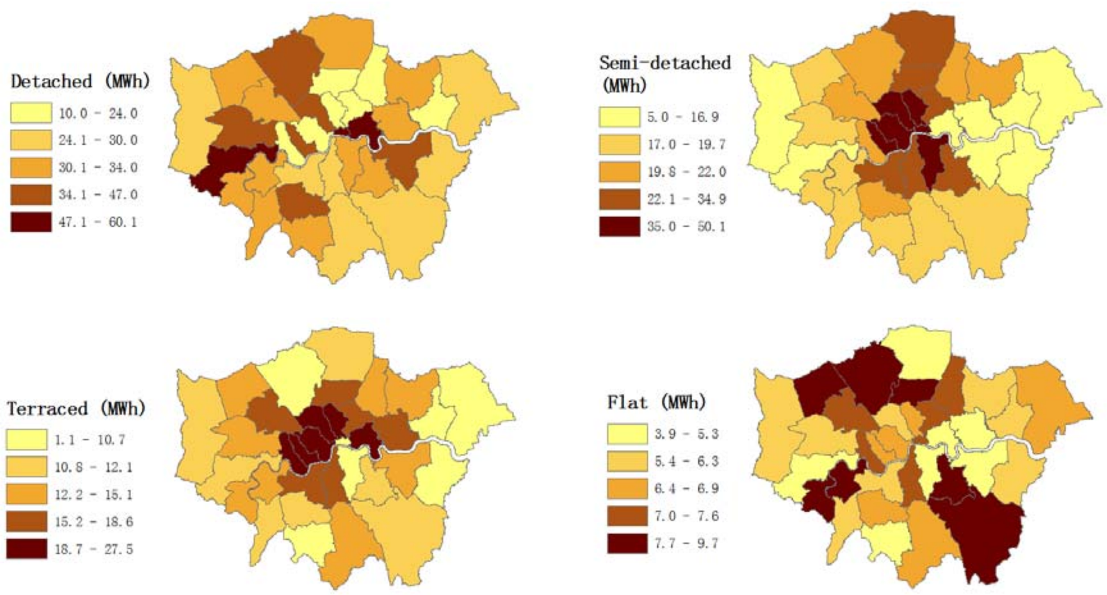


Figure 3 The gas use intensity per dwelling (MWh/property) from Bayesian hierarchical models at local authority scales in London.

Second, the statistical dispersion is analysed to show the spread of gas use intensity from local regression methods. The differences between minimum and maximum gas use intensity are significant, while the interquartile ranges (also called middle fifty between the upper and lower quartiles) are more similar from four local methods. Hence, the interquartile results are reliable for this case study. Among these approaches, the GWR is the most unstable method since there are some negative values for minimum gas use intensity. The results from the mixed model and Bayesian methods are similar. The specification of prior distributions can make the Bayesian computation more reliable. Then the following spatial analysis is based on the results from the Bayesian hierarchical model.

Figure 3 shows the spatial distributions of energy use gas per property at LA levels. The spatial patterns are very different for different dwelling types. For detached houses, the gas use intensity in outer London is higher than it is in inner London, while the trend is opposite for semi-detached houses similar to terraced properties. This can be explained using correlation analysis between the number of bedrooms and the gas use intensity. The correlation coefficients for semi-detached and terraced houses are 0.79 and 0.57, respectively. This means the number of bedrooms has significant influences on gas use intensity. For flat properties, the spatial distributions are very patchy. It is found that the correlation coefficient between gas use and household income is 0.44 although there is only very weak correlation between gas use intensity and the number of bedrooms for flat properties.

Caution should be taken in interpreting the results in some particular areas, which only have a few of specific of dwelling properties. For instance, central London has only a few of detached and semi-detached houses, so the energy use intensity of these two types of dwellings in these areas are unreliable. However, the impact of these unreliable results is expected to be small for other areas. This is because the method used in this analysis is intrinsically local, and the calculated results in other areas should be only slightly influenced.

Spatial variations for domestic electricity

Table 5 summarizes the electricity use in terms of dwelling types using various regression methods. The results are similar to the gas use. The electricity use intensity is greater than the reference values from the NEED project for detached and semi-detached houses, while the results for terraced properties and flats are similar to the NEED values.

The ranges between minimum and maximum electricity use are very different from four local regression methods. In contrast, the interquartile ranges are similar from four methods. This suggests that the spatial trend may be similar for all the method used in this paper. However, the calculation results from the GWR are unstable due to the negative values (Table 5), which is also the case for gas use intensity (Table 4). The negative values for electricity also occur for the mixed model, which indicates that the frequentist mixed models have limitations for different data sets. In this case, it is suitable for gas data, but not for electricity computation. The Bayesian hierarchical model among these three local methods is more adaptive for various settings, depending on the choice of prior distributions.

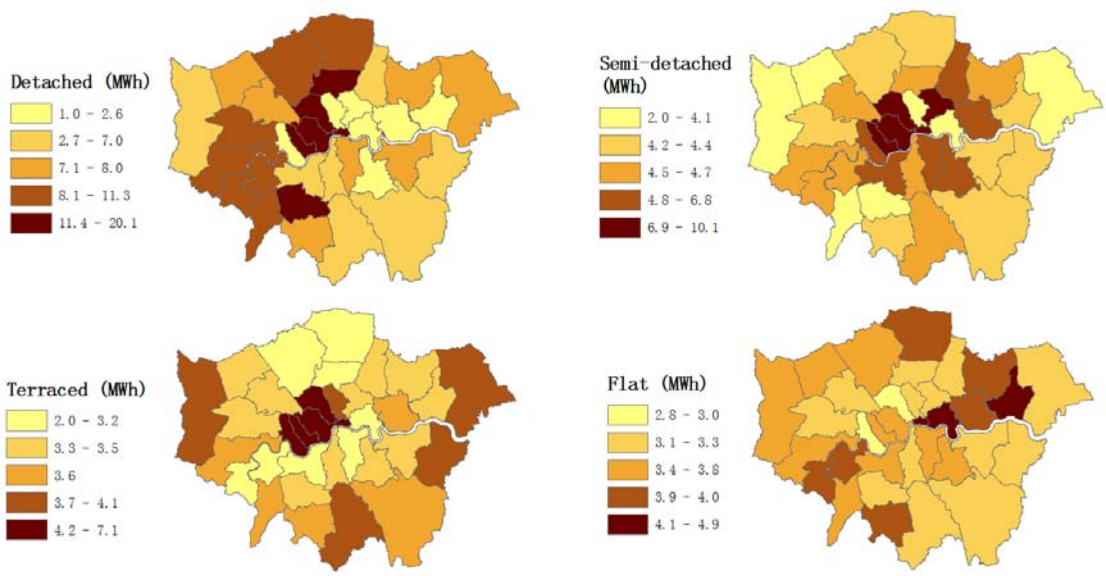


Figure 4 The electricity use intensity per dwelling (MWh/property) from Bayesian hierarchical models at local authority scales in London.

Table 5 Summary of coefficient values from global and local regression methods for electricity use (MWh/property)

PROPERTY	ITEM	REFS (DECC, 2015)	GLOBAL LINEAR	LOCAL GWR MSOA	LOCAL MIXED LSOA	LOCAL BAYE1 LSOA	LOCAL BAYE2 MSOA
Detached	Min	-	-	-100.5	-23.1	1.0	1.0
	Q1 [a]	-	-	3.3	4.2	4.7	2.8
	Median	5.4		6.7	7.5	7.2	7.1
	Mean	6.5	7.5	5.1	11.1	7.7	7.0
	Q3 [b]	-	-	9.5	9.2	8.7	8.1
	Max	-	-	98.0	217.1	20.0	20.0
Semi-detached	Min	-	-	-39.4	3.0	2.0	2.0
	Q1	-	-	4.2	4.3	4.2	4.0
	Median	4.3		4.8	4.5	4.5	4.5
	Mean	4.9	4.4	5.9	6.2	5.2	5.4
	Q3	-	-	6.7	5.6	5.6	5.7
	Max	-	-	44.5	29.9	10.0	10.0
Terraced	Min	-	-	-3.2	2.2	2.0	2.0
	Q1	-	-	3.0	3.4	3.4	3.2
	Median	3.6 [c] 3.8 [d]		3.4	3.6	3.5	3.5
	Mean	4.2 [c] 4.5 [d]	3.4	3.7	3.7	3.8	3.6
	Q3.	-	-	4.0	3.9	3.9	3.8
	Max	-	-	24.4	6.2	7.0	7.0
Flat	Min	-	-	-6.9	2.9	2.8	2.9
	Q1	-	-	3.0	3.3	3.2	3.3
	Median	2.6 [e] 2.5 [f] 3.3 [g]		3.3	3.5	3.5	3.6
	Mean	3.4 [e] 3.5 [f] 4.1 [g]	3.6	3.4	3.5	3.6	3.6
	Q3	-	-	3.8	3.7	3.8	3.8
	Max	-	-	7.9	4.3	4.9	5.0

Notes: a, Q1, 25th percentile, the first quartile; b, Q3, 75th percentile, the third quartile; c, middle terrace; d, end terrace; e, purpose built flats; f, converted flats; g, bungalow

Figure 4 illustrates the spatial patterns for the electricity use intensity at LA level in London. The spatial trends for gas and electricity use show similarity for detached, semi-detached, and terraced properties. The correlation coefficients between electricity use intensity and bedroom numbers are 0.62 and 0.61 for detached and semi-detached houses, respectively. Again, there are very weak correlation between electricity use and bedroom numbers for flats. There is no apparent correlation between electricity use and household income for flats. Further research is needed to clarify the reasons for the spatial patterns of both electricity and gas for flat properties in London.

CONCLUSION

This paper investigates the locally varying energy use intensity for electricity and gas in London using three local regression methods, including geographically weighted regression (GWR), mixed model, and Bayesian hierarchical model. The results indicate that the Bayesian hierarchical model can produce reliable results in comparison with the other two frequentist methods. Care should be put in interpreting the results from the GWR and mixed model since they may create unreasonable results. Another disadvantage for the GWR method is high computation cost for a large number of data set. The Stan, a new full Bayesian statistical inference with Hamiltonian Monte Carlo (HMC), can reduce computational cost compared to the commonly used sampling methods (random walk Metropolis or Gibbs sampling).

Note that this research is focused on the relationship between the energy use intensity and the number of various types of dwellings. The local regression method used in this paper can be also applied to explore the characteristics of building energy due to different types of explanatory variables, such as the floor area of dwelling properties, the population number, depending on the availability of input variables. It is also possible to combine different types of input variables together in local regression analysis, which would explain a large proportion of variations of energy use in urban buildings. The disadvantage of this combination analysis is that the interpretation of results would be more difficult due to a high correlation among input data.

ACKNOWLEDGEMENT

This research is supported by the Tianjin Research Program of Application Foundation and Advanced Technology (No. 14JCYBJC42600) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China.

REFERENCES

Bates D MM, Bolker B, Walker S. (2015) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Calderón C, James P, Urquizo J, et al. (2015) A GIS domestic building framework to estimate energy end-use demand in UK sub-city areas. *Energy and Buildings* 96: 236-250.

Choudhary R and Tian W. (2014) Influence of district features on energy consumption in non-domestic buildings. *Building Research & Information* 42: 32-46.

DECC. (2015) National Energy Efficiency Data-Framework (NEED) table creator updated 30 June 2014, DECC (Department of Energy & Climate Change), UK, Accessed 03-03-2015, https://www.gov.uk/government/statistical-data-sets/need-table-creator.

Fischer MM and Getis A. (2010) *Handbook of applied spatial analysis: software tools, methods and applications*: Springer Verlag.

Hamilton IG, Steadman PJ, Bruhns H, et al. (2013) Energy efficiency in the British housing stock: energy demand and the Homes Energy Efficiency Database. *Energy Policy* 60: 462-480.

Heo Y, Augenbroe G, Graziano D, et al. (2014) Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. *Building and environment*: doi:10.1016/j.buildenv.2014.1012.1016, in press.

Homan MD and Gelman A. (2014) The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research* 15: 1593-1623.

Howard B, Parshall L, Thompson J, et al. (2012) Spatial distribution of urban building energy consumption by end use. *Energy and Buildings* 45: 141-151.

London Datastore. (2015) LSOA atlas, Greater London Authority, UK, Accessed 03-02-2015, http://data.london.gov.uk/dataset/lsoa-atlas.

Lu B, Harris P, Charlton M, et al. (2014) The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science* 17: 85-101.

ONS. (2015) Digital boundaries for 2011 output areas (OAs) and super output areas (LSOAs and MSOAs) for England and Wales, ONS (Office for National Statistics), UK.

R Development Core Team. (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. , URL http://www.R-project.org/.

Stan Development team. (2015) RStan: the R interface to Stan, Version 2.5. http://mc-stan.org/rstan.html. .

Tian W and Choudhary R. (2012a) A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London. *Energy and Buildings* 54: 1-11.

Tian W and Choudhary R. (2012b) Spatial analysis of gas consumption by non-domestic buildings in Greater London. *The 5th International Building Physics Conference, Kyoto, Japan, 28-31 May.*

Tian W, Song J and Li Z. (2014a) Spatial regression analysis of domestic energy in urban areas. *Energy* 76: 629-640.

Tian W, Wang Q, Song J, et al. (2014b) Calibrating Dynamic Building Energy Models using Regression Model and Bayesian Analysis in Building Retrofit Projects. *eSim 2014.* May 7 to 10, Ottawa, Canada.

Wyatt P. (2013) A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England. *Energy Policy* 60: 540-549.