EXPLORATORY SPATIAL DATA ANALYSIS OF BUILDING ENERGY IN URBAN ENVIRONMENTS

Wei Tian[1,2], Lai Wei[1,2], Pieter de Wilde[3], Song Yang[1,2], QingXin Meng[1]
[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, China
[3] School of Architecture, Design and Environment, Plymouth University, Plymouth, UK

ABSTRACT

As energy data for buildings at urban scales becomes more widely available, it is possible to conduct spatial analysis for these energy data in order to understand and subsequently better model spatial patterns of energy use. A fundamental concept for spatial phenomena is spatial autocorrelation to explore how building energy data may be related to its location and neighbourhood. This study uses London as a case study to apply global and local spatial autocorrelation analysis for domestic electricity and gas at two different scales. The results, using Moran's I statistics for determining the degree of spatial autocorrelation, indicate that there are positive global spatial autocorrelations for both electricity and gas use in terms of per person or per household at two spatial scales. This means that spatial areas with similar energy use intensity would cluster together in London. Furthermore, the Moran's I statistics from bivariate analysis indicate that there is an apparent positive spatial autocorrelation between gas use in one area and electricity use in its neighbouring areas.

INTRODUCTION

As building energy data in urban areas is becoming increasingly available, there is an increasing amount of research on patterns of urban building energy use (Taylor et al., 2014; Caputo et al., 2013; Choudhary, 2012). Various methods have been used in detecting patterns of building energy in urban environment (Tian et al., 2014; Choudhary and Tian, 2014). The research methods can be categorized into exploratory data analysis and model-based inference analysis from a statistical perspective (Tian and Choudhary, 2012; Howard et al., 2012; Anselin, 2012). This exploration and the resulting understanding is a prerequisite for urban simulation efforts.

A model-based inference analysis is often involved in creating an engineering-based or statistical energy model to investigate the relationship between building energy and explanatory variables in urban settings (Tian et al., 2014; Bourdic and Salat, 2012). In contrast, explanatory data analysis is used to summarize the main characterises of a data set using both numerical and visual methods (Fischer and Getis, 2010). This research concentrates on the latter analysis. In the context of urban building energy assessment, GIS (geographical information system) can play an important role in exploring the trends of energy consumption for urban buildings, often called ESDA (exploratory spatial data analysis). It should be emphasized that it is only a preliminary step in ESDA to map the data in a GIS environment. Spatial autocorrelation is a central concept in the field of studying spatial phenomena (Fischer and Getis, 2010). For spatial data, a variable may be related to its location; such a relation is called spatial autocorrelation or spatial dependence.

For urban building analysis, these approaches can be used to relate energy use for buildings to the buildings' location or neighbourhood. The reasons for this spatial autocorrelation may include similarities between resident behaviours in one local neighbourhood, local urban microclimate conditions, similar income and expense in one small area, or the same building technology being used in nearby areas. It is necessary to understand these spatial patterns of building energy in order to make informed decisions on energy saving measures implemented in urban areas. However, there is only very limited research on this topic. Tian et al. conducted preliminary exploratory spatial analysis on energy use in London (Tian et al., 2014).

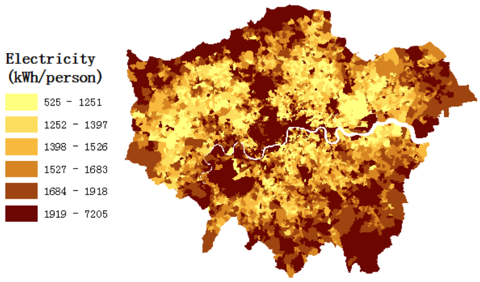
This paper implements spatial autocorrelation analysis to identify spatial patterns (clustering or dispersion) of energy use in urban areas. This study uses London as a case study to demonstrate the application of global and local spatial autocorrelation methods for domestic electricity and gas use. Both univariate and bivariate spatial analysis are conducted to explore the clusters of energy use for urban buildings. The results of spatial analysis may depend on the spatial scale used. Hence, two spatial scales are used in this research to provide robust results that will be explained in detail in section of "Data and Methods".

This paper is structured as follows. The energy used for exploratory spatial data analysis is introduced first. Then two types of statistical methods are briefly presented: spatial weight matrices and (global/local) spatial autocorrelation. The results are discussed for electricity and gas based on univariate spatial autocorrelation analysis. The next section describes the results in terms of bivariate spatial autocorrelation to explore bivariate spatial correlation structure for electricity and gas in London.

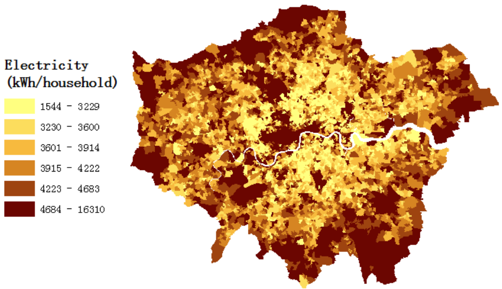
DATA AND METHODS

Data

The energy data used in this paper is domestic electricity and gas consumption for the year of 2013 in London, UK (DECC, 2015a; DECC, 2015b). Two spatial scales are MSOA (middle layer super output area) and LSOA (lower layer super output area) (ONS, 2015). MSOA and LSOA were designed for small area statistics used for the UK 2011 Census in terms of the numbers of both population and household. Note that the MSOA and LSOA statistics were released in 2004 for the first time, which have been renewed in 2011 due to the change of population in some areas. The 2011 version is used in this study to be in line with the data from the UK government (DECC, 2015a; DECC, 2015b). In London, the number of MSOA areas is 983 and the number of LSOA areas is 4835. The average populations in MSOA and LSOA are 8315 and 1691, according to the 2011 London census. The average household numbers in MSOA and LSOA are 3323 and 676 in London, respectively.



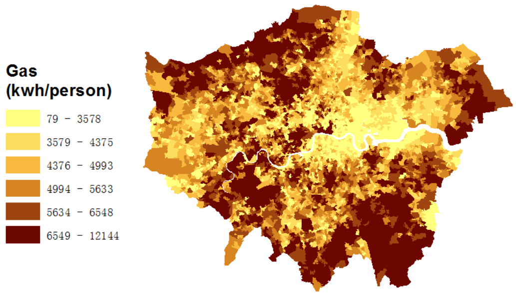
(a) Electricity per person



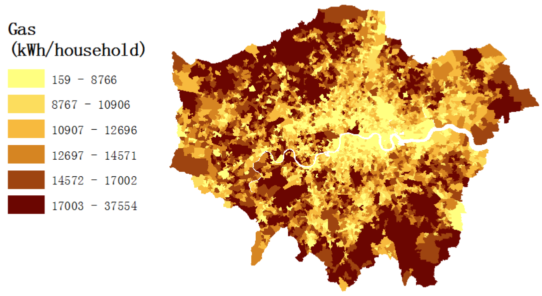
(b) Electricity per household

Figure 1 Electricity use per person and per household at LSOA (lower super output area) scale in London

Figure 1 shows the distributions of electricity per person and per household at LSOA scale in London. The spatial patterns are similar for electricity intensity in terms of per person and per household. In central London, there are some areas with high electricity intensity. When the city grows outwards, the electricity intensity per person or per household typically decreases. In outer London, electricity intensity would increase, again. This may be due to the combination of two factors: household size and urban heat island. More detailed analysis for this effect needs to be associated with both physical and social-economic explanatory variables, which is beyond the scope of this research.



(a) Gas use per person



(b) Gas use per household

Figure 2 Gas use per person and per household at LSOA (lower super output area) scale in London

Figure 2 illustrates the spatial distributions of gas use per person and per household at LSOA scale in London. This spatial pattern for gas use intensity is different from electricity from Figure 1. In the east part of central London, gas use intensity is very low and then gradually increases from the inner part to the outer part of London.

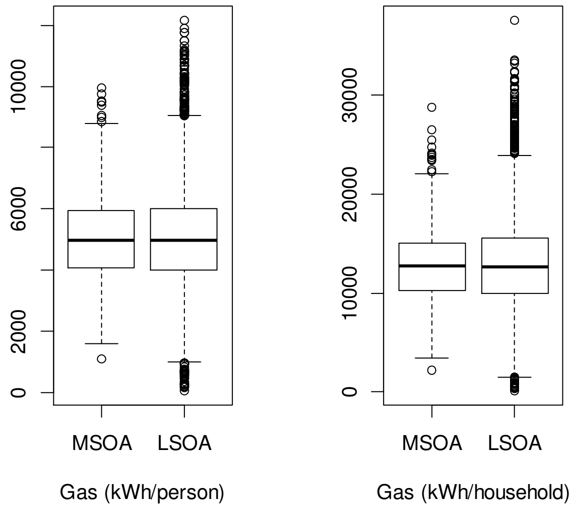


Figure 3 Boxplots for electricity use per person and household at MSOA and LSOA in London

Figure 3 shows the boxplots for electricity use per person and per household at two spatial scales in London: MSOA and LSOA. It may be expected that there are more outlier data in LSOA scale than those in MSOA scale. However, the variations of electricity intensity between MSOA and LSOA are small in the interquartile range (the difference between the upper and lower quartiles, the band inside the box in Figure 3).

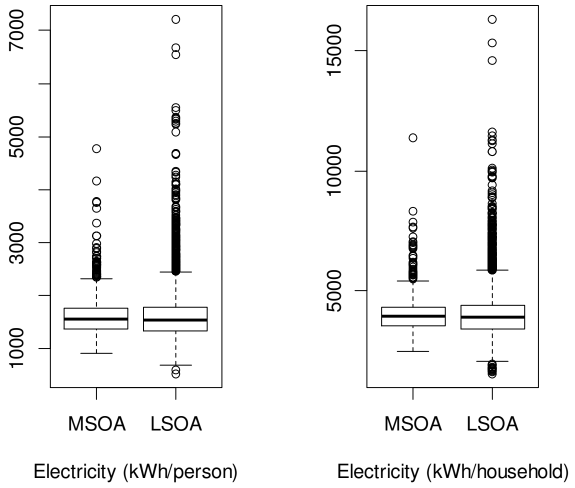


Figure 4 Boxplots for gas use per person and household at MOSA and LSOA in London

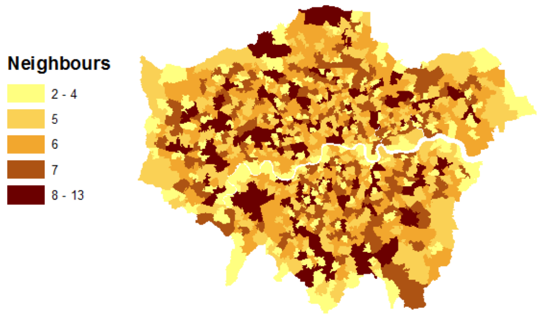
Figure 4 shows the boxplots for gas use per person and per household at both MSOA and LSOA in London. The characteristics for gas use intensity is similar to electricity use as shown in Figure 3. The difference is that there are more whisker data for higher gas use than electricity. This suggests that the distribution of gas use is more asymmetric compared to electricity intensity because there are more persons or households with high gas use relative to average users.

Methods

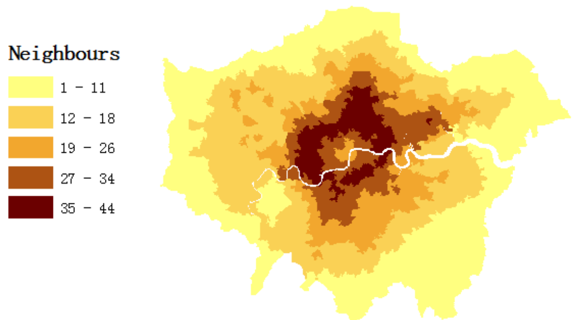
To define spatial autocorrelation in using spatial areal data, it must have proper spatial weight matrices (Bivand et al., 2013). A spatial weight matrix can assess the extent of similarity between locations. The first step is to determine whether two areas can be named as neighbour based on an adjacency criterion. The second step is to assign weights to the neighbour links identified from the first step.

The two spatial weight matrices are both implemented in this research in order to provide robust analysis: the queen and distance-based weight matrices as shown in Figure 5. Queen weight matrices belong to contiguity neighbours by defining a location's neighbours with either a shared vertex or border to other areas (Bivand et al., 2013). Thus, if an area has more nearby areas with shared borders, this area has more neighbours. In contrast, the distance-based matrix is used to calculate the centroids for polygons as the basis to determine which polygons can be identified as neighbours in terms of radius distances of these centroids in different spatial areas. Euclidean distance is used in this case and a default threshold distance is applied to ensure that every area has at least one neighbour. As a result, if the threshold radius range in one area includes more polygons, this area has more neighbours.

As illustrated in Figure 5, there are less neighbours calculated from a queen weight matrix than from a distance-based matrix. The distribution of queen neighbour numbers is more dispersed in the city, while the distance-based neighbours show a series of interesting rings from central area to outer London. The neighbour numbers in central area are small since there are more commercial offices in this area where residence density is low. As discussed in the section of "Data", the division of LSOA is based on both population and household numbers. As the city grows outwards, there are more neighbours since more people live next to the city centre. In outer London, the population density becomes smaller again, and the corresponding neighbour numbers are becoming smaller.



(a) Queen weights matrix



(b) Distance-based weights matrix

Figure 5 The number of neighbours in terms of queen and distance-based spatial weights matrix in London

After obtaining spatial weight matrices, spatial autocorrelation can be computed globally or locally for explanatory variables considered in the study. The global analysis of spatial autocorrelation has been used to test for the clustering in the whole map, whereas the local analysis has been used to test for the clusters in local areas.

Moran's I is a common statistic for spatial autocorrelation by computing a weighted Pearson produce moment correlation of a variable against itself related to spatial weighting. This global Moran's I for a variable named y is defined as (Bivand et al., 2013),

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

Where y_i is the ith variable, $\bar{y}$ is the mean of the variable, w_{ij} is the spatial weights defined in the first part of this section, and n is the number of spatial units. For this global spatial autocorrelation measure, the analysis is interpreted based on a null hypothesis that the variable of interest is randomly distributed in the whole area being analysed. Hence, if the p-value obtained from a permutation test is statistically significant, we can reject this hypothesis, which means the variable considered in a project is distributed in a certain spatial pattern (not random). The range of this Moran index is from -1 to 1. If the Moran index is greater than zero, this is a positive spatial autocorrelation and indicates that variable with similar values clustered together. If the Moran index is less than zero, this is a negative spatial autocorrelation, which indicates the variables with similar values tends towards dispersion. If the Moran statistic is close to zero, it means there is no apparent spatial patterns (i.e. random distribution) for the observation of interested in a the analysis.

For local indicators of spatial association (LISA), the local Moran's I for the ith spatial unit is defined as (Bivand et al., 2013),

$$I_i = \frac{(y_i - \bar{y})\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})}{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} \quad (2)$$

All the terms are equal to those of Eq. (1). Note that the global Moran's measure is the average of the local Moran statistics. A positive local spatial autocorrelation means the clustering for high-high or low-low locations of a variable, named as "spatial clusters". In contrast, a negative local spatial autocorrelation indicates the high-low or low-high areas for a variable, referred to as "spatial outliers". The global measure from Moran's I is a single average value for the entire data set to describe the overall spatial pattern over the whole study area, while the local measures have a value for each spatial unit and thus represent the local variations in different parts of the study area.

The discussion so far is only limited to univariate analysis. The global and local Moran's I statistics can be extended to describe spatial autocorrelation for two variables (called bivariate) (Anselin, 2005). The univariate Moran's I is used to evaluate the association between a specific variable in one area and the same variable but in nearby areas (called lag-variable), which is focused on the same variable. In contrast, the bivariate Moran's I is used to assess the correlation between one variable at a location and the other different lag-variable at the nearby locations, which is used for two variables. Hence, it is necessary to specify the first and second variable in bivariate spatial autocorrelation analysis. If the order of these two variables is switched, the results from bivariate spatial analysis may be different.

Another relevant concept is a correlation coefficient that can analyse the linear relationship between two different variables in the same spatial unit. For the conventional correlation analysis, the change of order of first and second variables would not affect the results.

All the spatial analysis in this research are conducted using the GeoDa program, a free software package that provides relevant functions such as mapping, exploratory spatial data analysis, and spatial regression. The GeoDa was developed by the Spatial Analysis Laboratory of the University of Illinois at Urbana-Champaign under the direction of Luc Anselin (2005).

RESULTS AND DISCUSSION

Spatial autocorrelation for domestic electricity

Table 1 indicates that the spatial autocorrelation for domestic electricity both per person and per household is statistically significant for queen or distance-based spatial weights based on the p-values (less than 0.1%). This means that similar values for domestic electricity intensity would cluster together in London. It also shows that the electricity per person is more clustered in comparison with the electricity per household. These conclusions are consistent from both MSOA and LSOA spatial scales in London although the Moran's I indicators from MSOA are higher than those from LSOA.

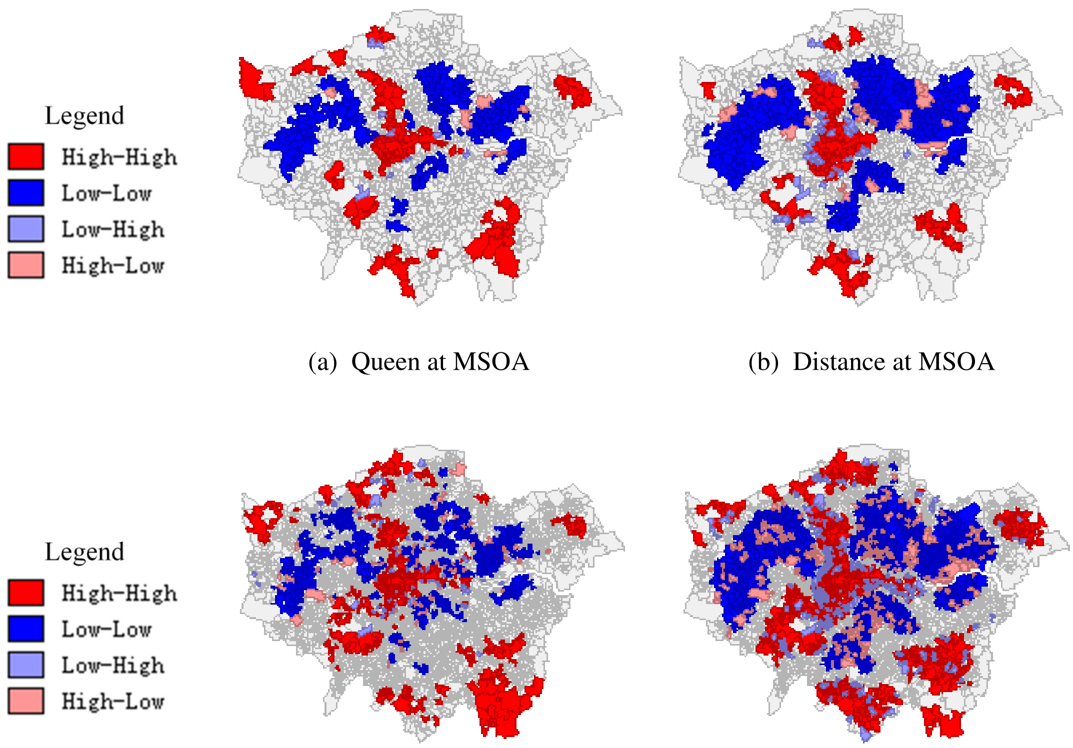
Figure 6 shows the cluster maps for domestic electricity intensity at two spatial scales: MSOA and LSOA, from local spatial autocorrelation analysis. The legend in Figure 6 has four codes for corresponding spatial association: dark red for high-high, dark blue for low-low, pink for high-low, and light blue for low-high. High-High means the areas with high electricity intensity are clustered with similar areas that also have large values for electricity intensity. The similar explanation is applied to low-low that the areas with low domestic intensity are clustered together. High-low indicates the areas with high-energy intensity that are surrounded by the areas with low energy intensity. The similar descriptions are used for low-high codes (low surrounded by high). The low-low areas are mainly in the northwest and northeast areas of London. Some small areas in the south also have clustering tendency with low electricity use per person. The high electricity area per person has more dispersed patterns across the study area, including small parts in central and outer London. The areas with negative local spatial autocorrelation for low-high and high-lower clustering are much less than those with positive local spatial autocorrelation.

Table 1 Global Moran statistics of spatial autocorrelation for domestic electricity and gas in London

ENERGY	BASE	SPATIAL SCALE	QUEEN WEIGHT		DISTANCE WEIGHT	
			MORAN	P-VALUE	MORAN	P-VALUE
Domestic Electricity	Population	MSOA	0.591	0.000	0.364	0.000
		LSOA	0.528	0.000	0.281	0.000
	Household	MSOA	0.462	0.000	0.266	0.000
		LSOA	0.445	0.000	0.199	0.000
Domestic Gas	Population	MSOA	0.640	0.000	0.479	0.000
		LSOA	0.573	0.000	0.386	0.000
	Household	MSOA	0.583	0.000	0.482	0.000
		LSOA	0.540	0.000	0.367	0.000

It should be no surprise that the distributions of local spatial autocorrelation are patchier from LSOA than those from MSOA scale. This is because LSOA has energy data for smaller areas and thus a higher resolution can be obtained. More areas with negative local spatial association can be found from LSOA scale compared to MSOA. This can explain why the global Moran's statistics are higher in MSOA than those in LSOA spatial scale in Table 1.

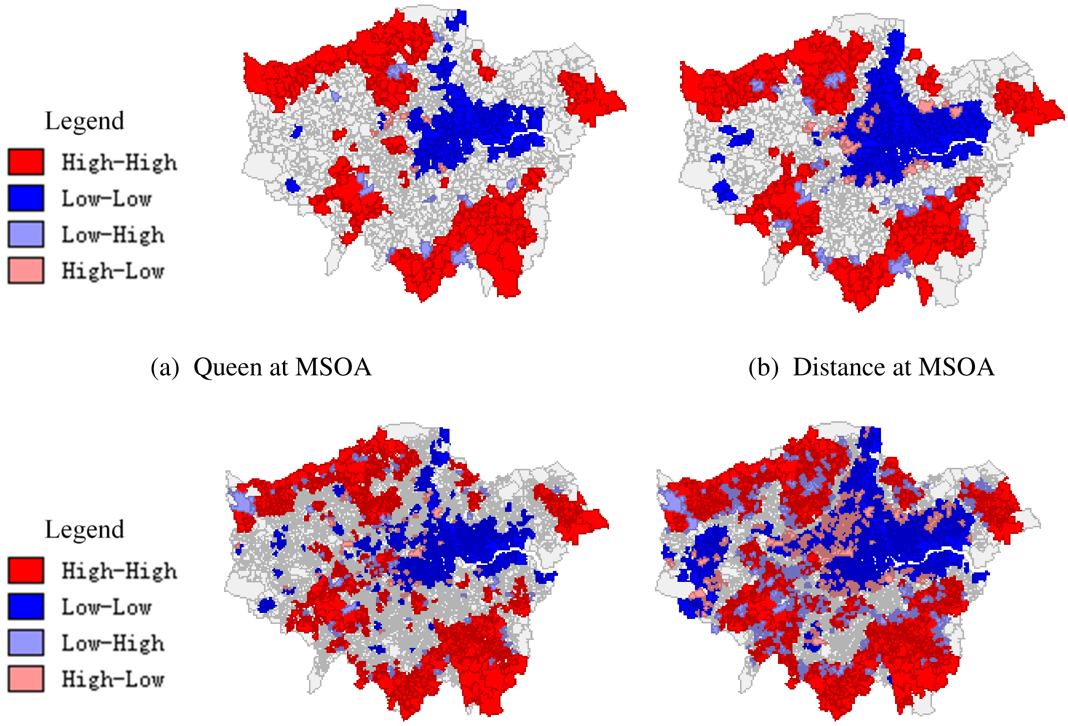
The spatial patterns for electricity per household (not shown here due to the limited space) are similar to those for electricity per person as described above. The main difference is that there are more high-high areas for the former, which are clustered in outer London. This is because the average household size is larger in outer London than that in inner London. As a result, the difference of electricity use per household is more prominent in comparison with electricity use per person.



(a) Queen at MSOA (b) Distance at MSOA

(c) Queen at LSOA (d) Distance at LSOA

Figure 6 Cluster map of local spatial autocorrelation for domestic electricity per person in terms of queen and distance-based spatial matrix at MSOA and LSOA scales in London.



(a) Queen at MSOA (b) Distance at MSOA

(c) Queen at LSOA (d) Distance at LSOA

Figure 7 Cluster map of local spatial autocorrelation for domestic gas per person in terms of queen and distance-based spatial matrix at MSOA and LSOA scales in London.

Spatial autocorrelation for domestic gas

Table 1 shows that a high degree clustering for domestic gas intensity is achieved above 0.5 from queen spatial matrices and more than 0.3 from distance-based spatial matrices in terms of global Moran's I. It is interesting to note that global spatial autocorrelation is more significant for gas use intensity than that for electricity use intensity. This means that the similar gas use between nearby areas is more clustered than the similar electricity use in its neighbours. Similar to domestic electricity, the global Moran's I values for gas use per person are higher than that for the gas use per household in London, except for the distance-based spatial matrix at MSOA scale.

Figure 7 illustrates the cluster maps of domestic gas per person using local spatial autocorrelation analysis. The high gas use intensity area (high-high) is clustered in outer London, especially northwest and southeast. In contrast, the area with low gas consumption per person is more clustered in inner London. The resulting cluster map indicates that in this case there is a lack of low-high and high-low areas (i.e. negative local spatial autocorrelation) which is similar to electricity use as shown in Figure 6. However, more areas with high gas use clustering (high-high) can be found in comparison with high-high electricity use per person (Figure 6 and Figure 7).

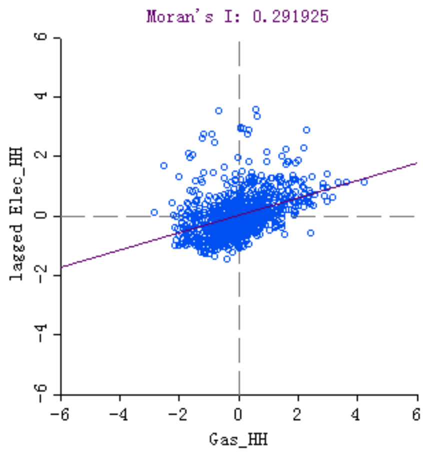
A similar spatial trend from electricity and gas use is that more low-high or high-low areas can be observed using LSOA data for London. This suggests that smaller area data can be very useful to explore the trends that are hard to identify based on the data from larger area.

The differences of gas use between per person and per household are very small. Hence, the local spatial cluster maps per household are not shown in this paper. Note that this pattern for gas consumption is different from electricity as discussed in the last section.

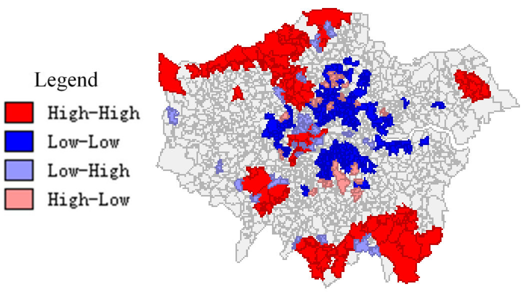
Spatial correlation between electricity and gas

As described in the methodology section, it is important to specify the first and second factor in bivariate spatial autocorrelation analysis. This section firstly analyses the bivariate spatial patterns of gas and electricity per household at MSOA in London. The gas is used as the first variable and the electricity is treated as the second lag variable. Subsequently, the order of these two variables is interchanged, and electricity is regarded as the first variable. The bivariate spatial analysis is used to explore the association between the electricity in one location and the lag variable of gas use in nearby areas.

Figure 8a shows the results of bivariate spatial autocorrelation for gas and electricity in London. The Moran's I of 0.292 indicates that there is an apparent association between gas use in one area and electricity use in its neighbouring areas. Bivariate LISA analysis as shown in Figure 8b can present the spatial distributions for these relationships. High-high means that the area with high gas use per household is surrounded by the households with high electricity use. These areas are mainly clustered in northwest and southeast of outer London. Between inner and outer London, some areas are identified as low-low code, low gas use households surrounded by low electricity households. There are also a few instances of low-high or high-low areas. Low-high here means that the households with low gas use are surrounded by the households with high electricity use. Then, high-low is the opposite of this explanation that the high gas use households have the neighbouring areas with low electricity use.



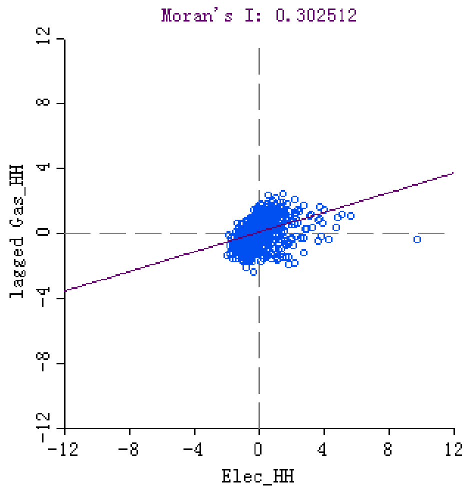
(a) Bivariate Moran's I: gas and electricity



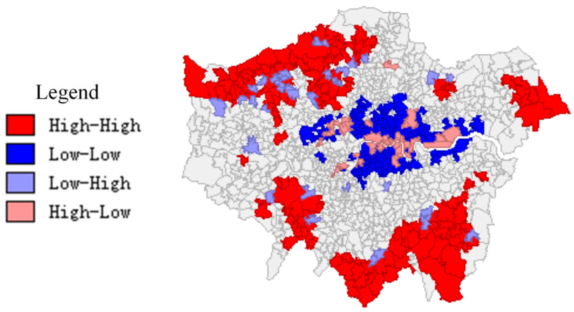
(b) Bivariate LISA cluster: gas and electricity

Figure 8 Bivariate global and local spatial autocorrelation analysis of gas and electricity for queen spatial matrix at MSOA in London.

Figure 9a indicates that the global bivariate Moran's I for electricity and gas is a positive spatial autocorrelation of 0.302. From the local spatial autocorrelation analysis for these two variables, the areas with high-high code are similar to the results from gas and electricity bivariate analysis in Figure 8b. Low-low energy areas are similar as well. However, the high-low areas from two bivariate analysis results are very different, which means the households with high electricity use are surrounded by the households with low gas consumption. This type of areas is clustered in central London.



(a) Bivariate Moran's I: electricity and gas



(b) Bivariate LISA cluster: electricity and gas

Figure 9 Bivariate global and local spatial autocorrelation analysis of electricity and gas for queen spatial matrix at MSOA in London.

CONCLUSION

This paper explores the spatial patterns, especially spatial autocorrelation, of domestic energy use in London. The following conclusions have been drawn from this research.

1: Similar values for domestic electricity intensity in terms of per person or per household would cluster together in London. The electricity use per person is more clustered in comparison with the electricity per household. This conclusion is consistent from both MSOA and LSOA spatial scales although the Moran's I indicators from MSOA scale are higher than those from LSOA in London.

2: There is a positive global spatial autocorrelation for gas use per person or household at MSOA and LSOA scales in London. The global spatial autocorrelation is more significant for gas use intensity than that for electricity use intensity.

3: The Moran's I statistics from bivariate analysis indicate that there is an apparent positive spatial autocorrelation between the gas use in one area and the electricity use in its neighbouring areas in London. If the order of these two variables is changed, this is still the case.

The findings from this research will be useful for designing the combined heating and power (CHP) system to choose the suitable areas based on the spatial distributions of heating and electricity loads. The high-energy-use areas identified from the bivariate analysis may be selected as the sites for the CHP system since these areas need both higher heating and electricity use. Moreover, the results from this analysis can be combined with spatial regression analysis to determine the reasons for the high-energy use in some areas. This would help to determine the priority of energy saving measures.

The methods used in this paper can be also applied for urban modelling, and applied to other cities. As more data at urban scale are becoming available, the exploratory spatial data analysis (ESDA) method is expected to have more applications in various settings. For example, these ESDA techniques can deep understand complex spatial patterns of building energy use, thus helping to make informed decisions on energy saving measures at district or city scale. The ESDA methods can also be used as an initial step when conducting spatial regression methods.

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

Anselin L. (2005) Exploring Spatial Data with GeoDa: A Workbook, , Center for Spatially Integrated Social Science, Department of Geography, Universit of Illinois, Urbana-Champaign, USA, Accessed 04-05-2014, http://www.csiss.org/.

Anselin L. (2012) From spacestat to cybergis twenty years of spatial data analysis software. *International Regional Science Review* 35: 131-157.

Bivand RS, Pebesma EJ, Gómez-Rubio V, et al. (2013) *Applied spatial data analysis with R, second edition*: Springer New York.

Bourdic L and Salat S. (2012) Building energy models and assessment systems at the district and city scales: a review. *Building Research & Information* 40: 518-526.

Caputo P, Costa G and Ferrari S. (2013) A supporting method for defining energy strategies in the building sector at urban scale. *Energy Policy* 55: 261-270.

Choudhary R. (2012) Energy analysis of the non-domestic building stock of Greater London. *Building and environment* 51: 243-254.

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

DECC. (2015a) Lower and Middle Super Output Areas electricity consumption 2013, DECC (Department of Energy & Climate Change), UK.

DECC. (2015b) Lower and Middle Super Output Areas Gas consumption 2013, DECC (Department of Energy & Climate Change), UK.

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

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.

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.

Taylor S, Fan D and Rylatt M. (2014) Enabling urban-scale energy modelling: a new spatial approach. *Building Research & Information* 42: 4-16.

Tian W and Choudhary R. (2012) 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. (2014) Spatial regression analysis of domestic energy in urban areas. *Energy* 76: 629-640.