Energy 111 (2016) 237–250

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Relative importance of factors influencing building energy in urban environment

Wei Tian [a,*], Yunliang Liu [a], Yeonsook Heo [b], Da Yan [c], Zhanyong Li [a], Jingjing An [c], Song Yang [a]

[a] Tianjin Key Laboratory of Integrated Design and On-line Monitoring for Light Industry & Food Machinery and Equipment, College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin 300222, China
[b] Department of Architecture, University of Cambridge, Cambridge, UK
[c] School of Architecture, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 22 February 2016
Received in revised form
25 May 2016
Accepted 26 May 2016

Keywords:
Variable importance
Correlated variables
Building energy
Urban environment

ABSTRACT

Energy assessment of urban buildings has become an active research field due to a large amount of energy consumed in cities as a result of fast urbanization. Hence, it is necessary to determine relative importance of variables for explaining variations of building energy use. However, two commonly used methods (correlation analysis and standardized coefficient) are only suitable for uncorrelated variables. This may not be the case for an extensive urban dataset containing social, economic, and physical variables. Therefore, this study proposes a two-stage approach to handle a large number of correlated variables in urban energy analysis. London has been chosen as a case study to determine influential factors affecting domestic energy use. The first stage applies two fast-computing methods (Genizi measure and correlation-adjusted score) to select important factors. The second stage implements two computationally intensive approaches (Lindeman Merenda Gold and proportional marginal variance decomposition) to further assess relative contributions of explanatory factors selected in the first step from conditional and marginal perspectives. The results indicate that this two-stage approach can deliver reliable results by explicitly accounting for correlations among variables in urban energy assessment.

1. Introduction

Over half of the global population lives in cities and urban areas are expected to absorb almost all the population growth by 2050 [1]. Accordingly, cities consume over 60% of global energy, which is projected to dramatically increase with an increase in the population in urban settlements. Therefore, it becomes a challenge to improve the quality of life and also reduce energy use and associated carbon emissions in urban environments [2,3].

There has been an increasing amount of studies on understanding key factors that impact energy consumption from building sector, particularly using variable importance techniques to understand complicated relationships between energy use and building characteristics in urban settings [4–7]. Chen et al. [8] investigated the impact of occupants on residential energy consumption in Hangzhou (China). They found that around 26.3% of variations of heating and cooling energy use can be explained by socioeconomic factors and the occupant age has negative correlation with energy consumption in Hangzhou. Tian and Choudhary [9] implemented two global sensitivity analysis methods (standardized regression coefficient and multivariate adaptive regression splines) to explore key variables influencing gas use for secondary schools in London. The SRC (standardized regression coefficient) is a method suitable for linear models, while the MARS (multivariate adaptive regression splines) can account for complicated non-linear relationships between inputs and outputs. He et al. [10] used local sensitivity analysis to analyse the influences of urban surrounding conditions (e.g., coverage, adjacent building height, surrounding with trees or no-trees) on both cooling load of a house and microclimate thermal environment in an urban block. Cheng and Steemers [11] applied a local sensitivity analysis method to investigate the relationship between inputs and dwelling carbon emissions for a district building energy model. The input factors considered are internal temperature, floor area, external air temperature, gas boiler efficiency, wall and window U-values.

* Corresponding author.
E-mail address: tjtianjin@gmail.com (W. Tian).

http://dx.doi.org/10.1016/j.energy.2016.05.106
0360-5442/

Abbreviations

CAR	Correlation-Adjusted coRrelation
CTA	the lowest council tax band (ranges from A to H)
CTH	the highest council tax band (ranges from A to H)
DmA	land area for domestic buildings
FV0	the most financial vulnerable households (band range from 0 to 11)
FV11	the least financial vulnerable households (band range from 0 to 11)
HH	household number
LMG	Lindeman, Merenda, and Gold method
LSOA	lower layer super output area
MSOA	middle layer super output area
NoE7	Number of Economy 7 domestic electricity meters
NoOd	number of ordinary domestic electricity meters
OwnM	households owned with a mortgage or loan
OwnO	households owned with outright ownership
PCC	partial correlation coefficient
PMVD	proportional marginal variance decomposition
Pop	population number
PriR	households in private rented homes
SocR	households in social rented homes
SRC	standardized regression coefficient
VIF	variance inflation factor
Wok	working-age population number

Santamouris et al. [12] found that an approximate linear relationship exists between annual expenses for electricity and family income in Athens. Wyatt [7] found that the income and number of resident adults are both positively linearly correlated with energy consumption in London using Pearson correlation coefficient. Tian et al. [13] reported that the number of households allocated to the council tax bands can explain a larger variation of electricity use compared to gas use in London using correlation coefficient method.

The correlation coefficients and the standardized regression coefficients (SRC) are widely used in building energy assessment in order to explore relationships between energy use and explanatory variables in urban areas. These two importance analysis methods assume that input variables at urban scale are uncorrelated. However, it is very likely that the factors affecting energy use may be correlated. For instance, the higher household income means the lower building age in Athens [12] and these two variables are usually regarded as explanatory variables to explain variations of energy use. Another example is that the housing tenure status is correlated with household income and floor area in the UK [7]. Although these two methods (SRC and correlation coefficient) have the advantages of being easy to use and fast to compute for initial studies, it has been found that these two approaches are unsuitable for variable importance analysis in the case of correlated factors [14,15]. More advanced statistical measures that could consider correlations among input variables are required to adequately explore the characteristics of urban building energy use.

This paper presents a two-stage variable importance method suitable for a large number of correlated factors in analysing urban building energy performance. The first stage is to implement two fast computing methods, Genizi measure and CAR (Correlation-Adjusted marginal correlation) score, in order to select the important factors influencing energy use. The second stage is to conduct two computationally intensive approaches, LMG (Lindeman, Merenda, and Gold) and PMVD (proportional marginal variance decomposition), to further assess relative contributions of explanatory factors using important variables selected from the first step. There are two reasons of using this two-stage method. One is that LMG and PMVD approaches cannot be directly implemented using a large number of variables available since they are too computationally expensive, and consequently the number of variables must be reduced from the first step. The other is that these four methods would provide more robust analysis from conditional and marginal perspectives that are necessary in the case of correlated variables as will be discussed in section 3. These four approaches (Genizi, CAR, LMG, and PMVD) will be described in detail in section 3. Furthermore, the paper demonstrates the relevance of using the proposed method to compare the ranking results from the standard methods that do not account for correlations among input factors. The standardized regression coefficients and correlation coefficients were selected as the standard methods, as they have been popularly used for determining important variables in building energy assessment [16]. London is chosen as a case study to demonstrate the method proposed because there are detailed data available on building energy and other relevant factors (such as dwelling types, population, households). In this study, the output variables are domestic gas and electricity use. The explanatory variables include population, household tenure, financial status, council tax bands, etc. For more detailed information on these variables, please refer to section 2.

This paper is structured as follows. Section 2 firstly describes both explanatory and response variables used in this research. Then the statistical approaches for assessing correlation among variables are presented in section 3. The methods of variable importance (Genizi, CAR, LMG, and PMVD) are also discussed in section 3. Section 4 contains four parts: results from correlation analysis; ranking results for gas use; ranking results for electricity use; discussion of computational time.

2. Data

The study is based on the data from London data store [17] and the Office for National Statistics [18] that provide demographic and energy use data at LSOA (lower super output area) spatial scale. The LSOA is statistical boundaries for small-area statistical analysis in the UK. Note that there are two versions of LSOA in terms of year released: 2001 and 2011. The 2001 spatial version was chosen for the study since there are significantly more data from 2001 LSOA than 2011 version. Note that the data used in this research is for the year 2011 although the spatial scale is based on 2001 LSOA. There are 4765 LSOA spatial units in London. However, only 4746 LSOA areas are used here due to the allocation issue [19]. The unallocated energy data are the consumption data that could not be allocated to the LSOA level because there is no information on a partial postcode or no postcode from the data suppliers. Since the energy data available has covered most of areas (also over 99% of total energy use) in London, this research can provide reasonable analysis.

2.1. Response variables

Fig. 1 shows the spatial distributions of energy data for domestic gas and electricity at LSOA level in London. The spatial patterns of gas and electricity use are very different. The electricity use is more concentrated at city centre and outside of Inner London, whereas a