



# Decomposing core energy factor structure of U.S. residential buildings through principal component analysis with variable clustering on high-dimensional mixed data



Endong Wang

Engineering Management and Technology Department, The University of Tennessee, EMCS 326B, Chattanooga, TN 37403, USA

## HIGHLIGHTS

- A novel integrated approach is developed for residential energy factor analysis.
- Severe homogeneity of U.S. residential energy factors is revealed (mean VIF > 15).
- A total of 32 key heterogeneous energy factors (Mean VIF = 1.21) are identified.
- Core factor subset sufficiently represents raw space (mean correlation = 0.86).
- Top three variables can interpret more than 33% residential energy variations.

## ARTICLE INFO

### Article history:

Received 25 December 2016  
Received in revised form 23 June 2017  
Accepted 28 June 2017

### Keywords:

Residential energy efficiency  
Energy factor structure  
Homogeneity decomposition  
Principal component analysis  
Variable clustering  
Mixed data

## ABSTRACT

Numerous energy computing frameworks were created with the aim of sustaining energy efficiency strategy to achieve residential sustainability in the U.S. While beneficial, without generic information on factor structure within building energy systems, most extant instruments are inclined to scope explanatory factor variables subjectively and diversely. Consequently, their intended utility often decreases with potential unstableness and limited generalizability among complicated energy system interactions. To overcome these issues, this paper develops a novel systematic homogeneity decomposition approach combining variable clustering and principal component analysis to identify key energy factor structure of residential buildings at the U.S. nation level. This study quantitatively results that, 32 key inter-heterogeneous energy variables (mean variance inflation factor = 1.21) appear sufficient to robustly profile the U.S. residential systems with an average Pearson correlation of 0.86 while reducing data burden by 68%. Top three significant variables relate to heating degree days, indoor environment and building vintage, respectively explaining 13%, 11% and 9% of energy variations. Thus, two major contributions are expected as follows. (1) These above obtained quantitative results can provide objective information for decision makers to sensibly select critical variables for robust energy computing with improved interpretability and generalizability by commonly using the above simplified 32-factor space (extracted from a 99-factor space) while saving data cost. (2) The developed novel approach can be useful in other countries for energy factor structure decomposing purpose since it has no geographical restrictions.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Achieving residential energy efficiency through effective planning and management embodies a significant opportunity for energy sustainability of the whole society in the U.S., given the large portion of energy consumed by dwellings. According to the update statistics from the U.S. Energy Information Administration (EIA) [1], in 2015, around 40 quintillion joules of energy was

consumed in buildings, accounting for almost 40% of the total U.S. energy use. Particularly, more than 54% of this amount was attributed to residential buildings. In viewing of these, such a large variety of strategic initiatives as efficiency rating programs [2], retrofitting incentives, energy efficiency policies, schemes and standards [3], have been proposed and implemented for residential energy performance upgrading.

Accordingly, in order to bolster the success of these above efficiency strategies, numerous energy computing models under diversified perspectives have been constructed to satisfy a wide

E-mail address: [Endong-Wang@utc.edu](mailto:Endong-Wang@utc.edu)

range of specific energy purposes, e.g. energy fault detection and diagnosis [4], dynamic projection of future trend [5], benchmarking of current efficiency pattern [6], at either component or whole building level. These models typically span from straightforward linear regression [7] to advanced machine learning techniques [8], e.g. case-based reasoning [5], artificial neural networks [9], support vector machine [9,10]. Regardless of algorithms and complexity of the models they use, appropriate selection and inclusion of valid energy indicators, variables or parameters are always of critical significance for a reliable modeling, accurate analysis and smooth results interpretation [11]. However, few or even no references regarding the dominating energy determinants at a generic level exist for variable selection (i.e. what variables should be primarily used) due to the lack of systematic energy factor structure analysis in residential domain. Thus, the majority of extant residential studies tends to empirically pick energy variables for model construction with significant subjectivity.

Nevertheless, building energy performance is of multidimensional essence and such a long list of individual factors as building shape, orientation, floor size, envelop materials, heating, ventilation and air-conditioning system (HVAC), lighting fixtures, indoor temperature setting, glazing, shading, ventilation, occupancy level, occupants' age and climate condition [12,13], potentially has influencing effects on energy performance of residential systems. This underlying complexity of energy factor structure may also be inferred by research findings in [14] that, more than 400 alternative retrofitting actions can be functional in enhancing building energy efficiency in either design or operation phase. This situation can be further complicated by the fact that these parameters are generally of varying characteristics. That is, while partial parameters are quantitative in essence and able to be expressed in numerical measurements, other parameters are naturally qualitative and generally difficult to be quantified at any quantitative scale.

Consequently, the parameter variables adopted in existing energy computing frameworks vary dramatically from one to another in terms of types, coverage and contents even with the identical quantification purpose. For example, to reach the same objective of energy benchmarking among residential buildings, many have adopted simple energy use intensity (EUI) based single-criterion benchmarking approach where only the energy factor of floor size or occupancy level was considered [2,6]. Conversely, others may consider tens of parameters characterizing logical features of census division, income level, building structure and household social characteristics, as in [2]. The extreme inconsistency of accounted variables in previous residential energy studies can be reflected by the review information in [15,16] as well.

Apparently, subjective and diversified use of energy parameters renders extant frameworks uneasy to mutually communicate (e.g. contrasted with each other) which may substantially reduce their generalizability for further applications [17]. For instance, in very rare cases, two energy references are observed to use the same set of energy variables, which confronts the screening of a better computation model through accuracy comparison. In addition, while considering few parameters possibly omits important energy interactions [11], arbitrarily including too many variables may result in redundancy issue when the scoped variables are highly correlated and homogeneous (i.e. describing the same building energy aspect). Further, the use of redundant correlated variables in an individual quantitative modeling could produce unstable or even wrong results and findings [7] due to the potential multicollinearity risk [16]. Multicollinearity risk refers to a dilemma that the strong dependence between two or more input parameters may lead to the instability of common energy computation algorithms, e.g., multiple linear regression, artificial neural networks and thus the inaccuracy of associated outcomes [7,18,19]. Moreover, without a systematical quantitative analysis on the inherent

homogeneity and significance of residential energy factors, it is often challenging to design optimal questionnaires for an energy survey so as to acquire effective information for energy computation. For example, respondents often refuse to provide information to too many possibly redundant questions [20]. Eventually, energy surveys are often inefficient and expensive.

Despite its significance, no studies are observed particularly on the systematic factor analysis to identify central energy determinants which could benefit to guide variable use and data collection in residential energy computing, especially at a nation level. Several works, e.g. [11,21,23], significantly contributed to the area of variable selection by applying various methods during the course of model construction (Please refer to Section 2 for more details). While being valuable, prior research tended to identify energy determinants only with reference to their own cases which often analyzed a relatively small building pool representing a narrow geographical region. As a result, the energy influencers identified in these works are generally case-sensitive, i.e., the factors recognized as significant in one case characterizing a small-scale site may not be influential in others [11]. Eminent parameters could be omitted without an extensive holistic structure analysis at a large scale. More important, the selected important case variables may be correlated to be highly homogeneous with collinearity trap and then not cozy for use in energy computing.

Principal component analysis (PCA) has been classically adopted for dealing with factor homogeneity in building energy domain [11,15,23,24]. Over traditional PCA suffering from qualitative variables [23], principal component analysis for the mixture of numerical and categorical variables (PCAMIX) [25,26] appears more functional in energy domain to cope with homogeneous correlation structure. This is due to its superior capability of simultaneously handling a mixture of qualitative and quantitative parameter variables which is common to residential systems. Though useful, PCAMIX method alone has difficulty in identifying influential energy parameters that are uncorrelated to completely depict energy systems under various angles. Still, PCA-components are the complex composite integrations of original variables, which makes the results from the pure PCA-akin methods uneasy to be intuitively interpreted.

On the other hand, variable clustering (VC) [27] which centers on data space partitioning could provide a useful supplement to PCA by dividing a universal high-dimensional feature space into separated lower-dimensional subsets, per certain predefined clustering metrics, e.g. correlation distance [28]. Through VC technique, a whole set of energy parameter variables can be separated into intra-homogeneous clusters, and the variables allocated into the same group are tightly correlated, thereby embedding the identical parametric information on energy. Parameters from different clusters are kept as distant as possible to express differing energy aspects. Thus, with this space dividing capability of VC in residential domain, a large complicated group of energy variables can be partitioned into smaller simple subsets to characterize different energy angles. The new subspaces can be more easily tackled with an improved interpretability due to lower dimensionality. Among the typical methods available for variable clustering, factor analysis based approach, though relatively expensive, tends to generate more robust outcomes with eigenvalue decomposition and iterative optimization [29].

This paper develops an integrated homogeneity decomposition procedure to systematically explore the factor structure of residential energy systems at a nation level and objectively identify the core and representative subset of heterogeneous energy variables for efficient computing. It combines the technical merits of principal component analysis and variable clustering on the mixture of numerical and categorical variables. Particularly, the underlying factor structure of the U.S. domestic energy systems is explicitly

Download English Version:

<https://daneshyari.com/en/article/4915812>

Download Persian Version:

<https://daneshyari.com/article/4915812>

[Daneshyari.com](https://daneshyari.com)