Energy and Buildings 117 (2016) 301–312

Contents lists available at ScienceDirect

Energy and Buildings

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

Reconstructing building stock to replicate energy consumption data

Fei Zhao [a], Sang Hoon Lee [b,*], Godfried Augenbroe [c]

[a] Retroficiency Inc., United States
[b] Lawrence Berkeley National Laboratory, United States
[c] Georgia Institute of Technology, United States

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form 1 October 2015
Accepted 1 October 2015
Available online 9 October 2015

Keywords:
Building stock model
Normative energy model
Energy simulation
Inverse model

ABSTRACT

The paper introduces an approach to replicate building stock energy data using energy survey data. For demonstration of the approach, the research uses energy consumption data for office buildings in Chicago from Commercial Building Energy Consumption Survey (CBECS) 2003. The replication starts from derivation of the energy use distribution for a building stock in a specific location from the survey data. Then probabilistic methods are used to map building stock model space to real-world data space reflecting a weather adjustment of the energy survey data. The approach leverages a linear surrogate model of the physics-based reduced order normative energy model. The normative building energy model can rapidly estimate the building energy performance with respect to its design and operational characteristics. The research investigates a statistical procedure to inversely estimate building parameters using regression and Bayesian inference model based on the Markov Chain Monte Carlo (MCMC) sampling techniques. The research serves a new paradigm of the building stock aggregation that can lead to an efficient energy model, which contributes the body of knowledge of energy modeling beyond the single building scale.

1. Introduction

For a physical system like the building stock of a city, it needs a scientific research procedure when dealing with observable parameters such as building design and operation and to relate them to the building energy. Due to limitations of onsite measurement, energy consumption data is inadequate for reproducing all descriptive input parameters and resulting parameters of the building stock. There are two levels of data inadequacy preventing us from fully reproducing the building stock energy profile based on energy survey data. The first level of the data inadequacy is caused by a scale. Lack of data points in a city scale makes the composition of buildings with different characteristics such as geometry, material, or principle use not easy. There were efforts for clustering and aggregation of buildings at a city scale [1–3]. In the United States (U.S.), weighting factors for different building types have been developed for buildings built after 2003, based on construction data in the McGraw-Hill's construction database [4,5]. Tuominen et al. used prototype building types that represent the whole building stock of Finland to evaluate energy efficiency measures to analyze energy savings from retrofits [6]. These efforts provide the possibility of using a small portion of prototypical buildings to reflect a large population of buildings. The second level is the building stock energy data inadequacy across individual cities. Building energy survey data are only applicable in a few cities. The Commercial Building Energy Consumption Survey (CBECS) database in U.S. provides basic statistical information about energy consumption and information about energy-related characteristics of commercial buildings [7]. However, building energy consumption survey data have limited amount of buildings at limited locations. The CBECS database does not provide city information where sampled buildings are located, but instead provides location information by U.S. census zones, heating degree days (HDD) and cooling degree days (CDD). To overcome the data inadequacy of the building energy at a city scale, the paper introduces an approach that can replicate the building stock observable parameters through weather data adjustment, statistical techniques integrating a physics-based energy modeling method. Fig. 1 illustrates the research flow of the proposed approach.

2. Modeling methodology

2.1. Weather adjustment of building energy consumption survey data

Building stock energy profile with inadequate survey data can be inferred by extrapolating the energy data of city or region with

* Corresponding author.
E-mail addresses: zhaof03@gmail.com (F. Zhao), sanghlee@lbl.gov (S.H. Lee), godfried.augenbroe@coa.gatech.edu (G. Augenbroe).

http://dx.doi.org/10.1016/j.enbuild.2015.10.001
0378-7788/

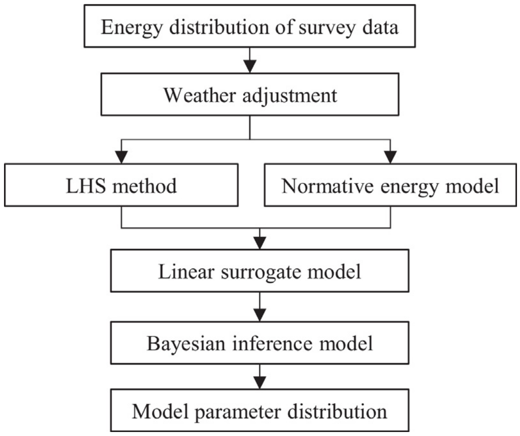

Fig. 1. Research flow of the proposed approach.

known energy data and weather information. This is built upon the assumption that the distribution of building design and operational characteristics in a city has no significant difference from the regional averaged distribution. The extrapolated probability distribution function (PDF) of the building energy consumption represents a scenario as "what if we move all buildings from the survey to the city with an interest". Given design and operational characteristics of individual buildings and their energy consumption, the "extrapolated" city is used for further analysis. When energy consumption survey results are available for a set of buildings, numbered as $i (i = 1, 2, \ldots, \mathbb{N})$, energy outputs and inputs of the dataset can be statistically formulated as:

$$PI_i = f(\boldsymbol{X}_{design,i}, \boldsymbol{X}_{operation,i}, \boldsymbol{X}_{climate,i}) + \varepsilon_i \quad (1)$$

where PI_i is an energy performance indicator of observation i, for instance, annual site and primary energy consumption, etc.; $\boldsymbol{X}_{design,i}$, $\boldsymbol{X}_{operation,i}$, and $\boldsymbol{X}_{climate,i}$ are vectors of design, operation, and climate parameters of observation i, respectively. f is a statistical function identical for all buildings in the data set, and ε_i is the error term. Given this function, building i can be relocated to another location to predict the adjusted energy performance indicator.

$$PI_{i,adj} = f(\boldsymbol{X}_{design,i}, \boldsymbol{X}_{operation,i}, \boldsymbol{X}_{climate,adj}) + \varepsilon'_i \quad (2)$$

Then, $PI_{i,adj}$ by substituting climate parameters $\boldsymbol{X}_{climate,i}$ with climate parameters from the new location is denoted as $\boldsymbol{X}_{climate,adj}$.

2.2. Generating a surrogate model for a building physical energy model

Input parameters for a dynamic simulation model are typically at the scale of hundreds or thousands. Scalability is a big problem preventing us from using such dynamic simulation models for large scale building stock analysis. Recent building stock modeling work conducted by National Renewable Energy Laboratory (NREL) [9,10] and Pacific Northwest National Laboratory (PNNL) [11] for office buildings have used EnergyPlus as an underlying engine. EnergyPlus is an advanced application that conducts a dynamic simulation to predict the building energy performance widely used for building energy research [12]. Use of the dynamic simulation may achieve deep details, yet requires high computing cost to perform simulation tasks. This is especially a bottleneck for tasks such as model calibration and stochastic inferences. Alternatively, the normative energy modeling method can be used for this purpose due to its scalability and transparency [13]. Lee et al. used the normative energy model for retrofit analysis of existing buildings and shared systems for large scale energy systems at a campus-scale [14]. However, a normative building energy model still cannot be directly expressed as a single explicit formula. Given feasible ranges of building design parameters, a set of inputs and the output of the normative model can be expressed as a linear regression model. Therefore, a linear regression surrogate model is developed to represent the normative model leveraging large input ranges and large amount of data samples.

2.3. Solving a linear inverse problem to reconstruct the building stock

Once a linear regression of the building energy model is identified, it needs to derive design and operational parameters of buildings to the replicate the building stock. Different from typical energy modeling process, this problem is defined as: "knowing the outputs of a model, how do we derive the inputs?" This type of problem is defined as an *inverse problem*. The approach solves a linear inverse problem to generate distributions of the building energy model input variables, which can replicate the building stock primary EUI distribution. This can be expressed as the following model (vectors and matrices are in bold; scalars in normal font, and vectors are indicated with a small letter; matrices with capital letter):

$$\boldsymbol{y} = f(\boldsymbol{\beta}, \boldsymbol{X}) \quad (3)$$

where $\boldsymbol{\beta}$ is the model parameter vector, $\boldsymbol{X}$ the input variable matrix, and $\boldsymbol{y}$ the output variable vector. A *forward problem* is defined as: Given the parameter matrix $\boldsymbol{\beta}$, what are the values of $\boldsymbol{y}$ for $\boldsymbol{X}$? On the contrary, an *inverse problem* is defined as: Having data ($\boldsymbol{X}$, $\boldsymbol{y}$), how to calculate or estimate the parameter vector $\boldsymbol{\beta}$? Another form is that having data ($\boldsymbol{\beta}$, $\boldsymbol{y}$), how to calculate or estimate the variable matrix $\boldsymbol{X}$? As a special inverse problem, if the function f is a linear function so that there is no interaction between elements of $\boldsymbol{x}$, the inverse problem is a *linear inverse problem*. A linear model is typically written in matrix notation as $\boldsymbol{Ax} = \boldsymbol{b} + \epsilon$, where $\boldsymbol{x}$ a vector of variables, and ϵ an error vector. A general formulation of a linear model considering additional equality and inequality constraints can be expressed as:

$$\begin{cases} \text{Inequality constraints}: \ \boldsymbol{Ax} = \boldsymbol{b} + \epsilon \\ \text{Equality constraints}: \ \boldsymbol{Ex} = \boldsymbol{f} \\ \text{Inequality constraints}: \ \boldsymbol{Gx} \geq \boldsymbol{h} \end{cases} \quad (4)$$

An inverse problem is usually under-determined or over-determined. In Eq. (4), $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{x}$ is an $n \times 1$ vector. If $m < n$, meaning that there are more unknown variables than equations, the system is *underdetermined* and usually has infinite solutions. Monte Carlo sampling methods can be used to sample the feasible region of an underdetermined linear problem in a uniform way. The term ϵ can then be considered as the uncertainties in the data. On the contrary, if $m > n$, meaning that there are more equations than unknown variables, the system is *over-determined* and usually there is no solution for which $\epsilon = 0$. An over-determined linear model can be solved by minimizing a norm of the error term $\epsilon = \boldsymbol{Ax} - \boldsymbol{b}$, for example the sum of squares $\sum \epsilon^2$. In this case, the term ϵ represents a model error term rather than uncertainties in the data.

The study uses an algorithm proposed by Meersche et al. to solve the over-determined linear inverse problem [15]. The algorithm contains two steps: (1) eliminate the equality constraints $\boldsymbol{Ex} = \boldsymbol{f}$ and (2) perform a random walk on the reduced problem. In the equality elimination step, $\boldsymbol{x}$ elements in the exact equality $\boldsymbol{Ex} = \boldsymbol{f}$ are linearly transformed to a vector $\boldsymbol{q}$ so that all elements are linearly independent. This linear transformation merges