BUILD SIMUL (2019) 12: 711–722
https://doi.org/10.1007/s12273-019-0540-6

Research Article

Building Systems and Components

Improvement of inverse change-point modeling of electricity consumption in residential buildings across multiple climate zones

Huyen Do[1], Kristen S Cetin[2] (✉)

1. Faculty of Project Management, The University of Danang – University of Science and Technology, Danang, Vietnam
2. Department of Civil Construction and Environmental Engineering, Iowa State University, Ames, IA, USA

Abstract

Inverse modeling is a common method to predict electricity consumption in buildings. Residential building electricity consumption patterns can vary significantly due to occupants and their sporadic energy-consuming behaviors, as well as due to variations in HVAC system types and characteristics across climate zones. However most data-driven methods in this area have been developed and evaluated using limited datasets. This points to the need for an understanding of how well data-driven models perform using residential energy consumption data from a range of locations and home types using a diverse dataset. Thus in this research, first, inverse change-point modeling methods are used to develop predictive models of monthly electricity use for 3,643 houses in four U.S. cities in three ASHRAE climate zones (2A, 4A, 5A), to evaluate the model performance. However, approximately 40% of homes did not fit within recommended criteria for change-point model development following a common model development sequence. Therefore, a modified version of the sequence, including a segmented change-point model, is then developed and evaluated. Change-point models with relaxed prerequisite criteria are also used to enable the fitting of models to a larger number of homes' data. As a result of these modifications, the number of homes with models increased from 60% to 71%, with a goodness-of-fit improvement of 13% (RMSE) and 8% (CV-RMSE) across the datasets evaluated. The results of this work enable improved prediction of energy use across a diversity of buildings and climate zones.

Keywords

inverse modeling,
change-point,
electricity use,
residential buildings

Article History

Received: 15 November 2018
Revised: 2 March 2019
Accepted: 12 March 2019

1 Introduction

In the recent past, the electricity consumption in residential buildings has increased, currently accounting for nearly 40% of total electricity consumption in the U.S. (US EIA 2016). This increasing electricity use, given the current mix of generation sources in the U.S., has an influence on the environment, including the production of greenhouse gas emissions which negatively impacts climate change (Pachauri and Meyer 2014). Therefore, it is of strong interest to reduce the electricity use in residential buildings. The ability to develop a model that predicts the electricity consumption of a building based on historical data is an important aspect of many of common methods used to assess the impacts of energy efficiency upgrades, and methods used for energy savings performance contracting (ESPC) (US DOE 2018). Such models are also used, for example, by utility companies to assess home's energy consuming behavior (Cetin et al. 2016), or by program such as the EnergyStar Portfolio Manager (US EPA 2019), to assess efficiency of buildings, in particular for benchmarking purposes. The more reliable these models are, and larger range of situations in which they can be used, the more impact they are likely to have on improving the efficiency of the building stock.

There are many types of models that have been developed in recent research to predict building electricity consumption, as summarized specifically for residential buildings in Do and Cetin (2018c). These data-driven or inverse models range significantly in complexity and input data requirements, including models such as change-point modeling (ASHRAE 2017a; Zhang et al. 2015; Kim et al. 2016, Kim and Haberl 2015a; Abushakra and Paulus 2016a,b,c; Chen et al. 2018),

E-mail: kcetin@iastate.edu

artificial neural networks (Biswas et al. 2016; Jovanović et al. 2015), genetic programming (Castelli et al. 2015; Jung et al. 2015), probabilistic graphic models (O'Neill and O'Neill 2016; Bassamzadeh and Ghanem 2017; Li et al. 2016), support vector machines (Ahmad et al. 2014; Jain et al. 2014), and occupant behavior models (Hong et al. 2015a,b) and Fourier series (Niu et al. 2018). Among these types of model, change-point models are a simpler method that is typically appropriate with the use of monthly level data (Do and Cetin 2018a; ASHRAE 2017a). For many buildings, particularly those with consistent energy consumption patterns, inverse change point modeling methods can provide sufficiently accurate predictions of energy consumption, particularly at the monthly energy consumption data level. In comparison with other machine learning methods, this method also has lower computational effort but also has been found to be able to achieve similar levels of accuracy of electricity use predictions compared to other more complex methods (Do and Cetin 2018a; Zhang et al. 2015).

For residential buildings, however, there are often homes that have irregular use patterns of the energy-consuming building systems that present challenges for predicting the energy consumption using such models due to the occupants and their energy-consuming behaviors, which are often more sporadic. For example, residential energy data collected and used in Do et al. (2018) shows monthly energy use ranged from 200 kWh to more than 1000 kWh for different homes under similar conditions. In addition, for similar weather months even where the heating, cooling, and ventilation (HVAC) energy use was similar, the total energy consumption of a home was found to vary by more than two times. However little efforts have been conducted to assess the ability of various data driven models to predict energy consumption across a variety of locations and homes.

This is in part due to the fact that currently, there is more publically available and/or accessible information and data on energy use in commercial buildings than residential buildings. The availability of more commercial building data is due, in part, to the increasing number of policies, laws and/or ordinances that support the public sharing of energy information in many large U.S. cities, such as Boston (2017), New York City (2018), and Washington D.C. (Department of Energy & Environment 2018). In addition, universities, where much research originates, typically have accessible data for commercial rather than residential buildings on their campuses. Commercial buildings are also much more likely to have more control systems and infrastructure, as well as budget, to monitor such data.

Energy information in residential buildings generally is associated with more privacy policies and laws on the sharing and use of this information (Do et al. 2018). This limited data has translated to a less comprehensive understanding of the ability of available modeling methods for use in predicting energy consumption across a broad range of residential buildings that make up the U.S. building stock. In addition, factors such as the type of HVAC system, which significantly impact energy consumption patterns, vary significantly across regions and climate zones. This points to the need for the use of residential energy data from a range of locations in the U.S. for model development.

For approximately 50% of residential buildings in the U.S., the only available collected energy use data is monthly energy consumption (Do and Cetin 2018a). As mentioned, for lower frequency energy consumption data, more simplified models are typically considered more appropriate for energy use prediction, one of the most common of which is change-point modeling. Change-point modeling methods are referenced in many energy performance standards (ASHRAE 2014, 2017b). These models are typically developed by using total energy consumption as dependent variable and outdoor weather data as predictor to decide the balance point in the type of five-, four-, three-, or two-parameter change-point models. To improve the ability of the existing models to predict electricity use of residential buildings, in this research a new sequence of change-point model development is proposed which includes modified versions of existing methods that enable the creation of change-point models for additional homes, with a similar or improved overall level of fit to in- and out-of-sample data. These are termed a "segmented" change-point model. This modified version of the initial inverse model development sequence improves the number of houses that have a model and enhancing the quality of prediction. In this research, a highly diverse dataset of electricity consumption data for a total of 3,643 residential buildings in four cities in the U.S. located in three ASHRAE climate zones is collected and analyzed, first, to determine how well existing models perform across a range of climate zones and buildings, and second, to determine the best sequence of change-point models to use to better improve the prediction of residential energy consumption.

This research is organized into three main sections, including the methodology, results, and conclusion and future work. The proposed method for the improvement and evaluation of inverse change-point of energy consumption in residential buildings. The results section shows the comparisons between initial and improved sequences among the different homes and regions.

2 Methodology

The proposed methodology for the improvement and evaluation of inverse change-point modeling of energy consumption in residential buildings is summarized in Fig. 1, including three major steps: (1) develop inverse

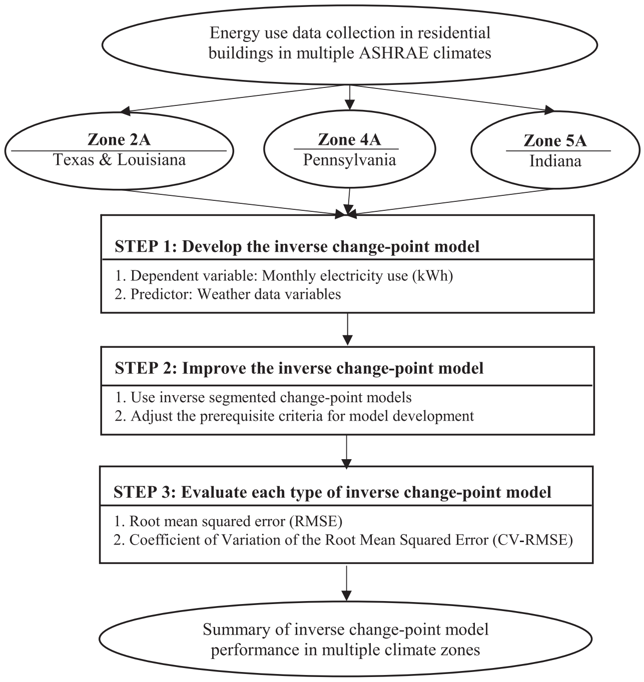

Fig. 1 Overview of methodology for improvement and evaluation of inverse modeling methods across multiple climate zones

change-point models using commonly-used methods (Paulus et al. 2015; Haberl et al. 2003; Kissock et al. 2003), (2) develop the new proposed inverse change-point model method and sequence, and (3) evaluate each inverse change-point model to determine and compare the overall quality of fit of the models. The major inputs into this evaluation include energy use data from residential buildings in multiple climate zones and the corresponding weather data, enabling the evaluation of the proposed improvements across different types of weather conditions, locations, and buildings. In many cases in other related studies, the evaluation of such methods is limited to a small number of buildings and/or climate regions (Zhang et al. 2015; Paulus et al. 2015). These types of input data are crucial to the assessment of the quality of fit of the methods as residential buildings are highly diverse (Do and Cetin 2018b). The details of each major step are analyzed in following sections.

2.1 Energy use data collection in residential buildings in multiple climate zones

The dataset of monthly electricity consumption was collected from residential buildings in four cities in the U.S., including New Orleans, Louisiana (1000 houses in ASHRAE hot-humid climate zone 2A), Houston and Austin, Texas (1,541 houses in ASHRAE hot-humid climate zone 2A), Philadelphia, Pennsylvania (102 houses in ASHRAE mixed-humid climate zone 4A), and northern Indiana (1,000 houses in ASHRAE cool-humid climate zone 5A). In general, characteristics of residential buildings in these locations are similar. An average 63%–66% of buildings are owner-occupied with an average of 2.5 people per home. The total annual household income in the four locations is also similar, including approximate 58%–61% of houses that have a household income under \$60,000; 19%–24% of houses have income from \$60,000 to \$100,000; and 18%–22% with an income over \$100,000.

Arguably, one of the most important characteristics of the studied homes that contribute to the electricity consumption in residential buildings is the type and characteristics of the HVAC system in each house. The distribution of HVAC system types in each climate region varies (Fig. 2). ASHRAE climate zone 2A has the highest percentage of homes using air conditioning (94%) in the summer, with 65% using electricity-based heat (e.g. heat pump, baseboard heat, etc.). With the mixed climate (ASHRAE climate zone 4A), 86% use air conditioning in the summer and a nearly an equal number use electric heat (42%) and gas heat (43%) in winter. For the cool climate

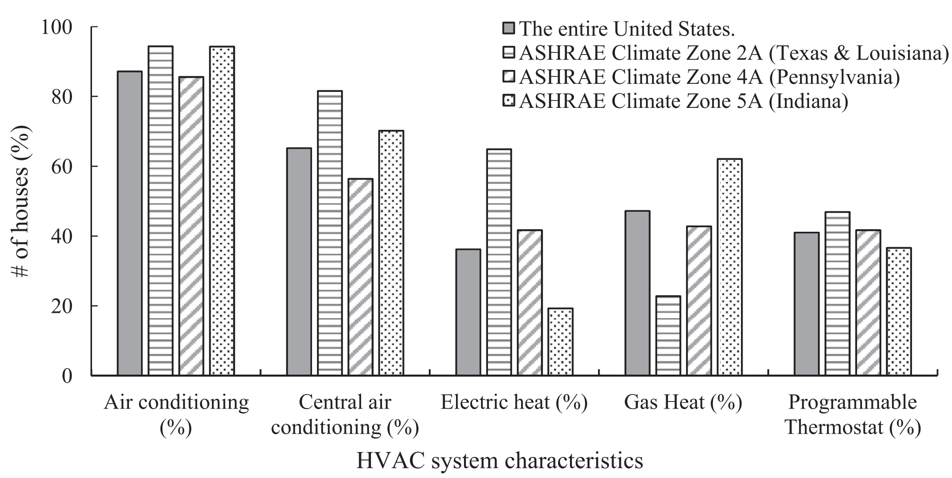

Fig. 2 HVAC system characteristics in residential buildings across the climate zones of studied homes

region (ASHRAE climate zone 5A), most of the houses use both gas heat and air conditioning in heating and cooling seasons due to the high demand in both seasons.

The electricity consumption data for residential buildings in these climate zones was quality controlled following the methods suggested by Cetin and Novoselac (2015). Additionally, given that typical electricity monthly billing start and end dates are not consistent across all homes and are also often a slightly different number of days in each billing cycle, the electricity data for each billing cycle was normalized to 30 days per month across the entire dataset. The distribution of the normalized monthly electricity usage of residential buildings across multiple climate zones, including Louisiana, Texas, Pennsylvania, and Indiana is shown in Fig. 3. Overall, the utilized data ranges from 2014 to 2016. Data is from November 2014 to December 2015 in Louisiana (14 months), from May 2014 to April 2015 in Texas (12 months), from April 2015 to January 2016 in Pennsylvania (10 months), and from January 2015 to April 2016 in Indiana (16 months) respectively. The average monthly electricity use in the studied locations varies (Fig. 3). Pennsylvania has the highest average of 1,134 kWh/month; Louisiana and Texas have a similar distribution of monthly electricity use with an average of 1,037 kWh/month and 893 kWh/month respectively. Indiana has the lowest average of monthly electricity use among four studied locations (448 kWh/month), likely associated with the low use of electricity-based heating. Weather data utilized in this research is collected from the local airports within the vicinity of the studied residential buildings.

2.2 Step 1—Develop inverse change-point model

To develop inverse change-point models for each residential building, an inverse single-variate model (ASHRAE 2017a; Zhang et al. 2015) is used with monthly electricity consumption as the dependent variable and outdoor temperature data as the predictor. Compared with an inverse multi-variate model or other forms of inverse models such as artificial neural networks (ANN) (Biswas et al. 2016; Jovanović et al. 2015), genetic programming (Castelli et al. 2015; Jung et al. 2015), Bayesian networks (O'Neill and O'Neill 2016; Bassamzadeh and Ghanem 2017; Li et al. 2016), or support vector machine (Ahmad et al. 2014; Jain et al. 2014), etc., an inverse single-variate model in the form of a change-point model is more statically appropriate (Do and Cetin 2018a), given the monthly-level frequency of data and the number of data points per home (from 10 to 16 data points for in-sample data). Previous studies' comparisons of the change-point modeling method to other models

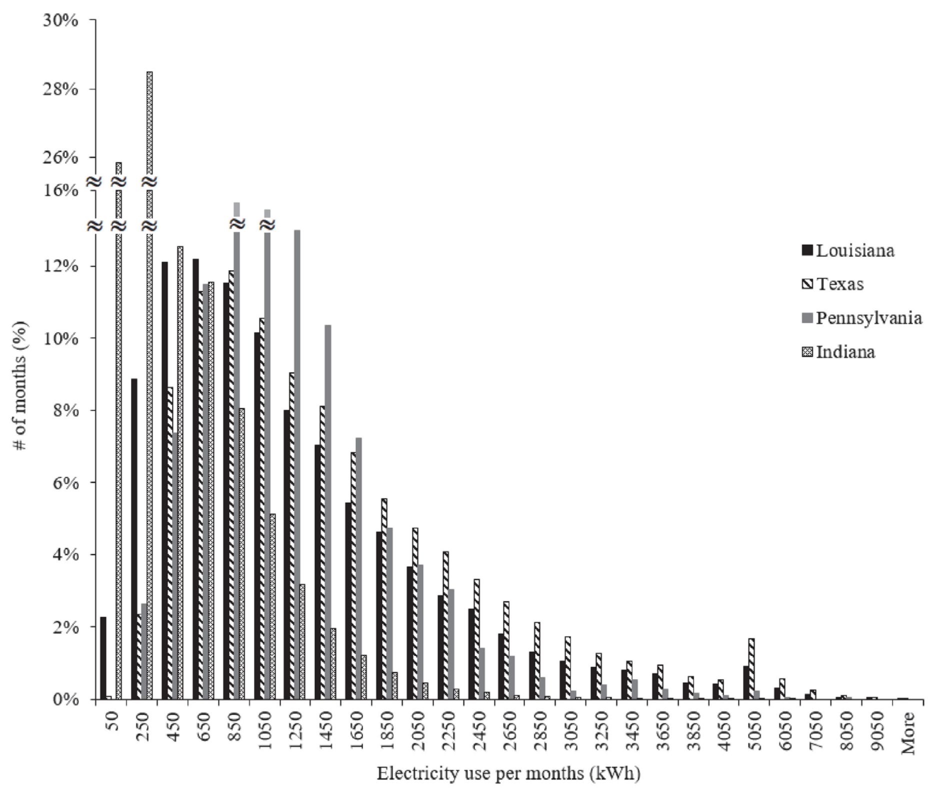

Fig. 3 Distribution of monthly energy usage data for residential buildings in Louisiana, Texas, Pennsylvania, and Indiana (note: The bin sizes utilized for electricity consumption of 5050 kWh/month and higher are larger; this is done to better show the tail of the distribution)

(Zhang et al. 2015) have demonstrated the statistical accuracy of this model for this level frequency of data. For the predictor of the model, the outdoor temperature variable is the most commonly used and typically more statistically significant, in comparison other common weather variable such as solar radiation, wind speed, and relative humidity (Do and Cetin 2018a; ASHRAE 2017a).

Inverse change-point models take different forms based on the number of parameters utilized each model, including the five-, four-, three-, and two-parameter change-point cooling and/or heating models (ASHRAE 2017a; Zhang et al. 2015). The balance point in this model is represented by the base temperature that represents the transition between the heating and cooling seasons. The base temperature of inverse change-point model in each residential buildings was automatically chosen by custom-developed algorithm in MATLAB (Paulus et al. 2015; Do and Cetin 2018a). To choose the best fit of inverse model type in each house, the developed algorithm applies four prerequisite criteria that all must be passed for the final chosen model type, including a shape test, significance test, R^2 test, and data population test (Paulus et al. 2015; Do and Cetin 2018a). The shape test requires the appropriate slopes of the regression lines in the inverse change-point model (Paulus et al. 2015). A p-value of 0.05 is the required threshold for the significance test (Paulus et al. 2015). For the R^2 test, a coefficient of determination of 0.5 is required as an acceptable threshold to check the fit of model (Paulus et al. 2015). Finally, for the data population test, at least three data points are required in each portion of the regression line. A final type of inverse change-point model is chosen if these four prerequisite criteria are all passed. The sequence utilized to choose the best type of inverse change-point model from 5-parameter to 2-parameter is demonstrated in some previous studies (Paulus et al. 2015; Do and Cetin 2018a). If any of the prerequisite tests are failed for all of the change-point model types, no model is developed for the studied home and the home is considered to have no model.

2.3 Step 2—Improve the inverse change-point model

To improve the inverse change-point model, a new sequence of inverse change-point model development is applied, including a new type of change-point model called as a "segmented" change-point model. The "segmented" model does not require the different segments of the change-point model to intersect. The improved sequence for development of inverse change-point models is represented in Fig. 4.

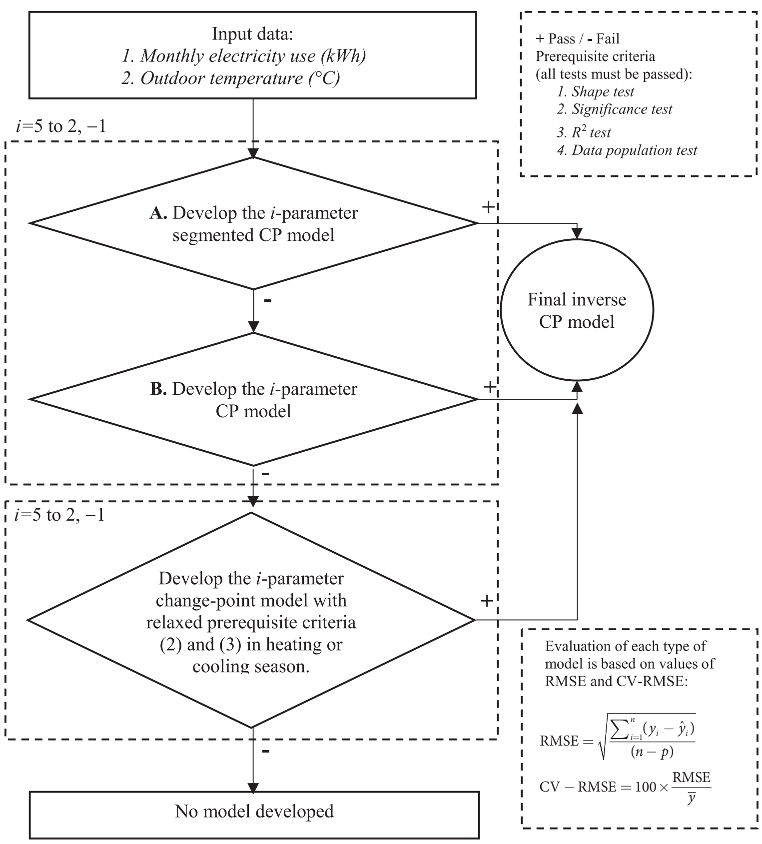

Fig. 4 Improved sequence for development of inverse change-point (CP) models (note: the "i = 5 to 2, −1" indicates the number of parameters of the CP model is iterated in a loop starting with 5-parameter and decreasing sequentially to 2-parameter CP model)

First is the development of inverse segmented/normal change-point models with 5-, 4-, 3-, and 2-parameter with all four tests checked. This step is similar to the common sequence in previous studies (ASHRAE 2017a; Paulus et al. 2015; Do and Cetin 2018a). In this sequence, the homes with the 5-parameter and 4-parameter change-point models represent the best fit of electricity consumption data. Otherwise, the 2-parameter change-point model is appropriate for the homes that have higher variation of electricity consumption in the heating or cooling seasons. Therefore, the 5-parameter and 4-parameter change-point models are considered first, then the 3-parameter and 2-parameter change-point models. Several other sequences were considered, however the proposed sequence was found to provide the most improvement in overall performance as compared to other model sequence orders. It is anticipated that the non-contiguous nature of the proposed piecewise model in the transition seasons may be appropriate for some homes due to variations in occupants' thermostat settings, and the presence, non-presence, or switching between dual-setpoint and single setpoint thermostat settings over a month-long period. As each data point represents a month of data, there can also be significant variations in weather and in HVAC use over this.

For increasing the number of homes where a model can be developed, this new sequence also considers relaxed criteria of two of the four tests used when developing change-point models, including increasing the acceptable *p*-value and/or threshold of R^2. If the house does not satisfy the criteria in the first step, and thus no model is developed, relaxed criteria are considered for the significance and/or R^2 test in the heating or cooling seasons. The evaluation of the inclusion of each of these changes, including the segmented model and the relaxed criteria is evaluated.

2.4 Step 3—Evaluate each type of inverse change-point model

The accuracy of each model is evaluated with the values of root mean squared error (RMSE), and coefficient of variation of the root mean squared error (CV-RMSE) with in-sample data and out-of-sample data. These values are common metrics used to assess the level of fit in prediction models, where RMSE evaluates the residual variance in the prediction model (Grace-Martin 2012) and CV-RMSE evaluates the variability of the error between measured and model-predicted values, indicating the model's ability to predict the overall load reflected in the data (Zhang et al. 2015). CV-RMSE is also used in commonly referenced guidelines for building energy use predictions such as ASHRAE Guideline 14 (ASHRAE 2014), International Performance Measurement & Verification Protocol (IPMVP) (US DOE 2002). The values of RMSE and CV-RMSE are computed as the below equations (Zhang et al. 2015):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{(n-p)}} \tag{1}$$

$$\mathrm{CV-RMSE} = 100 \times \frac{\mathrm{RMSE}}{\bar{y}} \tag{2}$$

where y, $\bar{y}$, and $\hat{y}_i$ respectively represent the actual, average, and predicted electricity consumption.

Based on these methods, the best-fit inverse change-point model is determined to be the model with the lowest values of RMSE and CV-RMSE using in-sample data. The out-of-sample data is used to evaluate the performance of prediction for each house. With the relaxation of the criteria associated with the significance test and R^2 test, the quality of model fitness could be impacted. Based on the result of in- and out-of-sample data, the overall resulting RMSE and CV-RMSE values across the studied homes using the original model criteria with no relaxed criteria were at appropriate or similar levels to those of the models with the relaxed criteria.

2.5 Summary of inverse change-point model performance in multiple climate zones

The results are then compared among the climate zones, including the inverse model performance with the initial sequence and improved sequence.

3 Results and discussion

Of the available data in each location of study, energy data from a total of 681 houses in Louisiana, 1133 in Texas, 68 in Pennsylvania and 301 in Indiana were used to develop the inverse change-point models using the commonly-used method (Paulus et al. 2015; Haberl et al. 2003; Kissock et al. 2003) that includes outdoor temperature as predictor and monthly electricity consumption data as dependent variable (Table 1). The inverse models developed include the five-, four-, three-, and two-parameter change-point models. Examples of each type of inverse change-point model using the common sequence are demonstrated in Fig. 5. Other houses from these locations have a high variation in monthly electricity consumption, therefore, these houses did not pass the prerequisite tests in the model development sequence. In other words, there is no model developed for these homes.

Based on these model designations, the types of HVAC systems in the residential buildings studied generally appear to be similar to that of the homes in the studied ASHRAE

Table 1 Percentage of homes with different types of change-point (CP) model using the common sequence of inverse CP model development

Types of models	Louisiana	Texas	Pennsylvania	Indiana
5-parameter CP	8.4%	10.1%	14.7%	0.5%
4-parameter CP	28.9%	10.5%	19.6%	7.9%
3-parameter CP cooling	19.6%	48.2%	10.8%	7.9%
3-parameter CP heating	4.4%	1.0%	18.6%	4.9%
2-parameter CP cooling	6.2%	3.7%	2.9%	6.8%
2-parameter CP heating	0.6%	0.1%	0.0%	2.1%

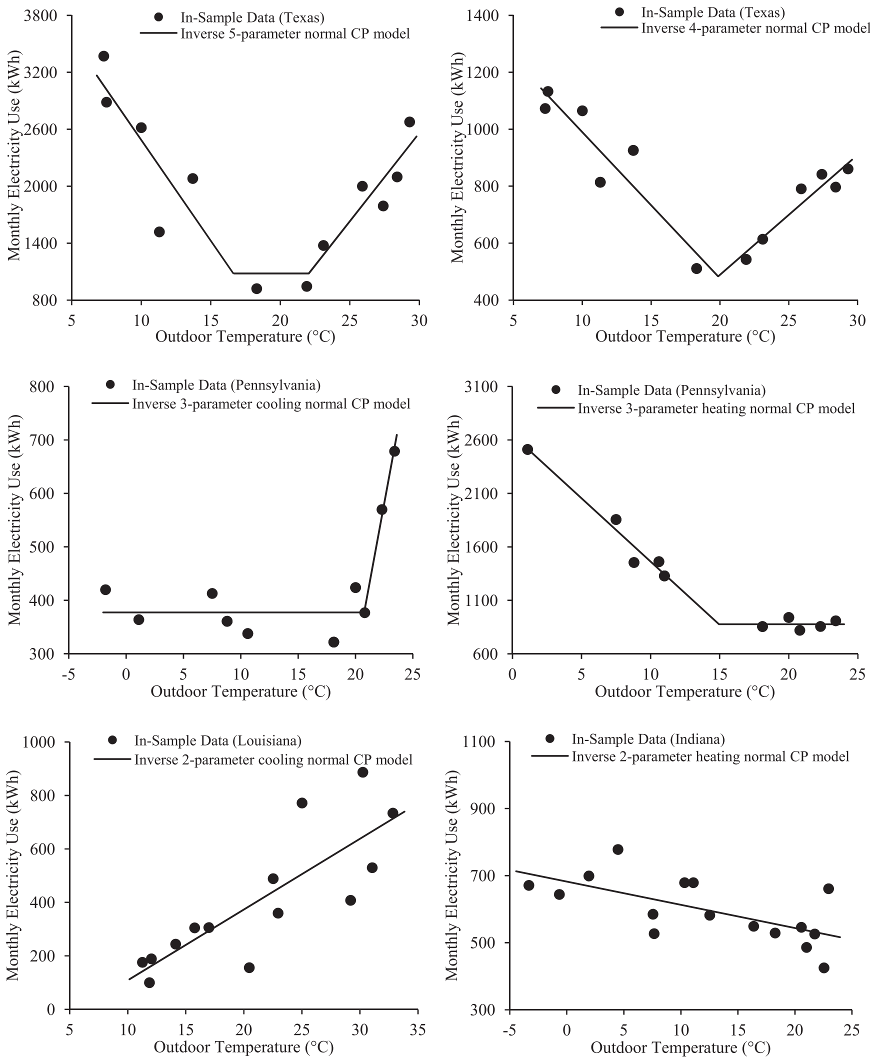

Fig. 5 Examples of inverse change-point models developed in each residential building with the common sequence in four locations in three ASHRAE climate zones

climate zones. Typically homes with 3-parameter heating, 4-parameter, or 5-parameter models represent homes with electricity-based heat, and 3-parameter cooling represents homes with gas-based heat. 62% of homes in the climate zone where the Indiana homes are located (ASHRAE climate zone 5A) use gas-based heat and 19% use electricity for heat, compared to 8% and 13% of inverse models developed respectively. Similarly, in Pennsylvania (ASHRAE climate zone 4A), the number of homes that have electric heat and gas heat is approximately 42% and 43% respectively, compared to 53% and 11% of the corresponding types of inverse models developed, respectively, in climate zone 4A.

Approximately 48% of total homes in Texas and 20% in Louisiana have 3-parameter change-point cooling models (Table 1). These values are slightly different from HVAC system characteristics in ASHRAE climate zone 2A where about 23% of houses using electric heat.

Examples of inverse change-point models developed for each residential building with the proposed improved sequence in the four locations in three ASHRAE climate zones are included in Fig. 6. In these examples, Figs. 6(a)–(d) show the inverse change-point segmented model; the inverse change-point model with relaxed prerequisite criteria in the cooling and heating season is shown in Figs. 6(e) and (f) respectively. Overall, with the common sequence of model development, approximately 68%, 74%, 67%, and 30% of total of houses respectively across Louisiana, Texas, Pennsylvania, and Indiana have an inverse model developed (Table 1). However, with the improved sequence of inverse change-point model development, the number of houses that have models are significantly increased, including 71%, 84%, 69% and 60% respectively in each of the climate zones (2A, 4A, and 5A) (Table 2), or in total 71% of homes on average. Of particular improvement is Indiana (ASHRAE climate zone 5A), in which the number of homes with models increased from 30% to 60%. Using the improved sequence of inverse change-point model development, in all ASHRAE climate zones (2A, 4A, and 5A), an average of 35% of houses utilized the inverse change-point segmented model, 23% used the inverse change-point model, and 13% of houses used the inverse change-point models with relaxed prerequisite criteria in cooling and/or heating seasons. The improved

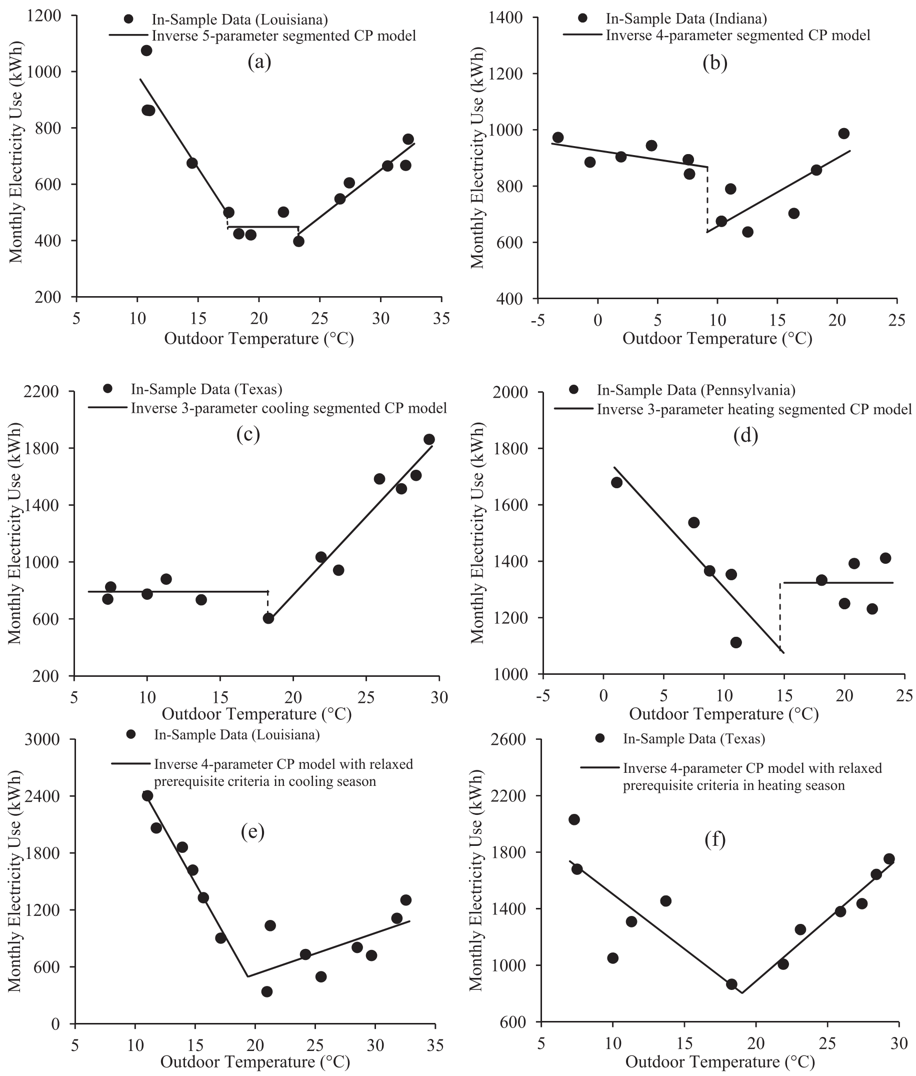

Fig. 6 Examples of inverse change-point models developed in each residential building with the improved sequence in four locations in three ASHRAE climate zones (note: Figs. 6(a)–(d): inverse change-point segmented model; Fig. 6(e): inverse change-point model with relaxed prerequisite criteria (2) and (3) in cooling season; Fig. 6(f): inverse change-point model with relaxed prerequisite criteria (2) and (3) in heating season)

sequence of inverse change-point model development enables an additional 11% of total houses to have a model with an acceptable level of fit that did not have a model with the initial sequence (Table 3).

To compare the model fitness using the initial sequence and the improved sequence, both the in-sample electricity consumption data in the three ASHRAE Climate Zones and out-of-sample electricity consumption data, where available, are used (Table 4). As mentioned in methodology section, the values of model fitness coefficients such as RMSE and CV-RMSE are applied to compare the accuracy and performance of inverse models developed. For both RMSE and CV-RMSE, the lower the values, the better model performance. Therefore, any sequence that has the better model fitness with in-sample data and better prediction with out-of-sample data is considered to overall be the better model development sequence to forecast the monthly electricity consumption in residential buildings. The results of evaluation of model fitness quality of each type of inverse change-point model using both initial and improved sequences are shown in Table 4.

First, for in-sample data, with the commonly used sequence of inverse change-point model development, the values of RMSE and CV-RMSE are 161.2 and 15.6%, on average, respectively among the three climate zones. However, with the new sequence of inverse change-point model development, the RMSE and CV-RMSE are lower, at 141.2 and 14.3% respectively, representing an approximately 12% and 8% improvement in these values. In addition, the RMSE values for the change-point models, segmented change-point models, and change-point models with relaxed prerequisite criteria in the cooling and/or heating seasons in the three ASHRAE climate zones are 128, 149.7 and 170.5 respectively. Similarly, the CV-RMSE values for each type of inverse models using the improved sequence are, on average, 12.4%, 15.3%, and 16.0% across the dataset. From these results of RMSE and CV-RMSE using in-sample data, it is clearly seen that the improved sequence of inverse change-point model development enhanced the quality of model fitness in four locations in three ASHRAE climate zones.

Table 2 Percentage of homes with different types of change-point (CP) model using the improved inverse CP model sequence

Types of models	Louisiana	Texas	Pennsylvania	Indiana
A. Segmented CP models				
5-parameter segmented CP	21.6%	9.9%	19.6%	22.1%
4-parameter segmented CP	3.1%	1.0%	4.9%	1.5%
3-parameter segmented CP cooling	14.8%	6.9%	7.8%	13.6%
3-parameter segmented CP heating	1.5%	0.8%	4.9%	4.6%
B. CP models				
5-parameter CP	0.8%	3.6%	3.9%	0.2%
4-parameter CP	5.6%	2.1%	4.9%	2.5%
3-parameter CP cooling	10.9%	44.4%	2.0%	4.7%
3-parameter CP heating	0.4%	0.0%	1.0%	0.1%
2-parameter CP cooling	0.7%	2.1%	1.0%	0.1%
2-parameter CP heating	0.0%	0.1%	0.0%	0.2%
C. Relaxed prerequisite criteria (2) and (3) in cooling season				
CP segmented model	1.6%	0.3%	2.0%	1.9%
CP model	1.1%	0.2%	2.9%	3.1%
D. Relaxed prerequisite criteria (2) and (3) in heating season				
CP segmented model	3.7%	2.6%	6.9%	2.3%
CP model	4.7%	9.9%	6.9%	3.5%

Note: Prerequisite criteria (2) is significance test, and prerequisite criteria (3) is R^2 test.

Table 3 Improvements in the percentage of homes with change-point (CP) models assigned using improved sequence

% of houses	Louisiana	Texas	Pennsylvania	Indiana
% of houses with no model developed from initial sequence	31.9%	26.4%	33.3%	69.9%
% of houses with no model developed from improved sequence (Fig. 4)	29.5%	16.2%	31.4%	39.6%
% of houses that have CP model using improved sequence that did not with the initial sequence	**2.4%**	**10.2%**	**1.9%**	**30.3%**

Table 4 Evaluate the quality of model fitness of each type of inverse change-point model using both initial and improved sequences

Average values of coefficients of model fitness	In-sample data				Out-of-sample data			
	Louisiana	Texas	Pennsylvania	Indiana	Louisiana	Texas	Pennsylvania	Indiana
A. Initial sequence	***n* = 681**	***n* = 1133**	***n* = 68**	***n* = 301**	***n* = 681**	***n* = 1133**	—	—
RMSE	176.0	168.3	146.7	153.7	268.8	389.3	—	—
CV—RMSE	16.4%	12.8%	13.5%	19.5%	27.1%	22.7%	—	—
B. Improved sequence	***n* = 705**	***n* = 1291**	***n* = 70**	***n* = 604**	***n* = 705**	***n* = 1291**		
RMSE	157.4	153.5	126.4	127.6	261.9	303.3	—	—
CV—RMSE	15.4%	11.7%	12.3%	18.0%	26.1%	21.1%	—	—
1. CP models	*n* = 184	*n* = 804	*n* = 13	*n* = 78	*n* = 184	*n* = 804		
RMSE	128.6	142.3	140.0	101.1	236.4	276.2	—	—
CV—RMSE	13.2%	10.8%	10.7%	14.8%	21.8%	19.1%	—	—
2. Segmented CP models	*n* = 410	*n* = 288	*n* =38	*n* = 418	*n* = 410	*n* = 288		
RMSE	170.3	167.8	125.5	135.2	267.3	360.1	—	—
CV—RMSE	16.0%	12.8%	13.2%	19.0%	26.8%	21.4%	—	—
3. CP models with relaxed prerequisite criteria in cooling or heating season	*n* = 111	*n* = 199	*n* = 19	*n* = 108	*n* = 111	*n* = 199		
RMSE	187.4	189.1	148.7	156.7	284.2	374.3	—	—
CV—RMSE	16.9%	14.3%	13.6%	19.1%	30.7%	26.6%	—	—

Note: RMSE = root mean squared error, CV-RMSE = coefficient of variation of the root mean squared error.

The accuracy of prediction from inverse models using both the initial and improved sequence with out-of-sample data in two locations (Louisiana and Texas) with sufficiently longer periods of data collection is next evaluated. From the initial sequence of inverse change-point model development, the average values of RMSE and CV-RMSE are 329.1 and 24.9% respectively, while these values are lower with the improved sequence, at 282.6 and 23.6% respectively. More specifically the RMSE are 256.3, 313.7, 329.3 and the CV-RMSE are 20.4%, 24.1%, and 28.6% respectively, on average, for the change-point model, segmented change-point models, and change-point models with relaxed prerequisite criteria. Overall, improved sequence of inverse change-point model development performs better with both in-sample data and out-of-sample data.

The evaluation of the values of RMSE and CV-RMSE in this research follows industry guidelines for electricity use in buildings including ASHRAE Guideline 14 (ASHRAE 2014) and the International Performance Measurement & Verification Protocol (IPMVP) (US DOE 2002). The average value of CV-RMSE under the improved sequence of inverse change-point model development is 14.3% and thus below the industry guidelines threshold of 15% for monthly data. A recent previous study of Do and Cetin (2018a) which applied the original sequence of inverse change-point model development for electricity consumption in residential buildings had average values of RMSE and CV-RMSE for the developed change-point models of 160.4 and 17.2% respectively.

In this research, by using the improved sequence of inverse change-point model development, improved both the RMSE and CV-RMSE values across the datasets (141.2 and 14.3% respectively) are lower and thus demonstrate better fit on average. In addition, these values are also significantly lower than the machine learning methods assessed in the previous study (Do and Cetin 2018a) using monthly electricity use data; the ANN (artificial neural network) that has average value of 386.8 for RMSE and 44% for CV-RMSE. The values of RMSE and CV-RMSE using the improved sequence of model development in this study are also lower than values in other studies such as 14.96% for CV-RMSE in research of Zhang et al. (2015) with hourly energy use data, and 17.2% for CV-RMSE in the study of Kim and Haberl (2015). Therefore, in other words, the model fitness for electricity use prediction are significantly improved, even with a highly diverse dataset of data, with the proposed improved sequence of inverse change-point model development.

4 Conclusions

With the high dependency of residential energy consumption on the potentially significant variations in occupant behavior and use of occupant-dependent loads, variations in electricity consumption patterns in residential buildings occur and strongly affect the inverse model development. Particularly across the diverse dataset utilized in this work, over 40% of studied houses in the four cities in three ASHRAE climate zones were found to not fit the inverse change-point model development criteria following the four prerequisite tests often utilized for model development. The modified version, including an inverse segmented change-point model technique and an inverse change-point model with relaxed prerequisite criteria in cooling or heating season, improved the RMSE values by 13%, and CV-RMSE values by 8% across the studied dataset. This improved sequence of inverse change-point model development also reduced the number of houses in the three ASHRAE climate zones that do not have a model developed from 40.4% to 29.2%. These results demonstrate that this improved sequence works well and enables better quality prediction of energy consumption of residential buildings. There are a range of applications of this effort, such as in energy performance contracting, and energy efficiency evaluation of residential buildings. This will help both owners of residential buildings and energy contractors be able to evaluate building performance.

The characteristics of the homes in these dataset and occupants who are living in these homes vary, including the characteristics of the HVAC systems, therefore the types of inverse change-point models also vary. While it has proven to be challenging to gather residential energy consumption data in the U.S. due to privacy issues, further studies in this area would help to assess residential energy consumption patterns and change point model performance across a broader range of homes, HVAC systems and climate zones.

Acknowledgements

This study was supported, in part, by The Vietnam Ministry of Education and Training, and Earth Networks, provided data. However, any results, opinions, and conclusions in this paper do not reflect the views of these organizations and are strictly those of the authors.

References

Abushakra B, Paulus MT (2016a). An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part I: Background (1404-RP). *Science and Technology for the Built Environment*, 22: 976–983.

Abushakra B, Paulus MT (2016b). An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part II: Methodology (1404-RP). *Science and Technology for the Built Environment*, 22: 984–995.

Abushakra B, Paulus MT (2016c). An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part III: Results and analysis (1404-RP). *Science and Technology for the Built Environment*, 22: 996–1009.

Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, Saidur R (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33: 102–109.

ASHRAE (2014). ASHRAE Guideline 14 - 2014. Measurement of Energy and Demand Savings. Atlanta, GA, USA: The American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASHRAE (2017a). ASHRAE Handbook—Fundamentals. Atlanta, GA, USA: The American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASHRAE (2017b). ASHRAE Standard 140 - 2017. Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs. Atlanta, GA, USA: The American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Bassamzadeh N, Ghanem R (2017). Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. *Applied Energy*, 193: 369–380.

Biswas MAR, Robinson MD, Fumo N (2016). Prediction of residential building energy consumption: A neural network approach. *Energy*, 117: 84–92.

Boston (2017). Building energy reporting and disclosure ordinance (BERDO). The city of Boston. Available at https://www.boston.gov/environment-and-energy/building-energy-reporting-and-disclosure-ordinance. Accessed Jun 2018.

Castelli M, Trujillo L, Vanneschi L, Popovič A (2015). Prediction of energy performance of residential buildings: A genetic programming approach. *Energy and Buildings*, 102: 67–74.

Cetin KS, Novoselac A (2015). Single and multi-family residential central all-air HVAC system operational characteristics in cooling-dominated climate. *Energy and Buildings*, 96: 210–220.

Cetin KS, Siemann M, Sloop C (2016). Disaggregation and future prediction of monthly residential building energy use data using localized weather data network. In: Proceedings of ACEEE Summer Study on Energy Efficient Buildings, Pacific Grove, CA, USA.

Chen Z, Freihaut J, Lin B, Wang CD (2018). Inverse energy model development via high-dimensional data analysis and sub-metering priority in building data monitoring. *Energy and Buildings*, 172: 116–124.

Department of Energy & Environment (2018). Washington D.C. Available at https://doee.dc.gov/energybenchmarking. Accessed Jun 2018.

Do H, Cetin KS (2018a). Evaluation of the causes and impact of outliers on residential building energy use prediction using inverse modeling. *Building and Environment*, 138: 194–206.

Do H, Cetin KS (2018b). Impact of occupant behavior in data-driven energy use modeling in diverse residential buildings across multiple climates. In: Proceedings of the 4th Residential Building Design & Construction Conference, State College, PA, USA.

Do H, Cetin KS (2018c). Residential building energy consumption: A review of energy data availability, characteristics, and energy performance prediction methods. *Current Sustainable/Renewable Energy Reports*, 5: 76–85.

Do H, Cetin K, Andersen T (2018). Characteristics and causes of outliers in inverse modeling of residential building energy use data. In: Proceedings of the ASHRAE Winter Conference, Chicago, USA.

Grace-Martin K (2012). Assessing the fit of regression models. Cornell University. Cornell Statistical Consulting Unit. Available at https://www.cscu.cornell.edu/news/statnews/stnews68.pdf. Accessed Jun 2018.

Haberl JS, Claridge DE, Sreshthaputra A, Kissock JK (2003). Inverse modeling toolkit: Application and testing. *ASHRAE Transactions*, 109(2): 435–448.

Hong T, D'Oca S, Turner WJN, Taylor-Lange SC (2015a). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. *Building and Environment*, 92: 764–777.

Hong T, D'Oca S, Taylor-Lange SC, Turner WJN, Chen Y, Corgnati SP (2015b). An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema. *Building and Environment*, 94: 196–205.

Jain RK, Smith KM, Culligan PJ, Taylor JE (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123: 168–178.

Jovanović RŽ, Sretenović AA, Živković BD (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, 94: 189–199.

Jung HC, Kim JS, Heo H (2015). Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach. *Energy and Buildings*, 90: 76–84.

Kim KH, Haberl JS (2015a). Development of methodology for calibrated simulation in single-family residential buildings using three-parameter change-point regression model. *Energy and Buildings*, 99: 140–152.

Kim KH, Haberl JS (2015b). Development of a home energy audit methodology for determining energy-efficient, cost-effective measures in existing single-family houses using an easy-to-use simulation. *Building Simulation*, 8: 515–528.

Kim KH, Haberl JS (2016). Development of a home energy audit methodology for determining energy and cost efficient measures using an easy-to-use simulation: Test results from single-family houses in Texas, USA. *Building Simulation*, 9: 617–628.

Kissock JK, Haberl JS, Claridge DE (2003). Inverse modeling toolkit: Numerical algorithms. *ASHRAE Transactions*, 109(2): 438–448.

Li R, Wang Z, Gu C, Li F, Wu H (2016). A novel time-of-use tariff design based on Gaussian Mixture Model. *Applied Energy*, 162: 1530–1536.

New York (2018). NYC Resource. Office of Sustainability. The city of New York. Available at http://www.nyc.gov/html/gbee/html/home/home.shtml. Accessed Jun 2018.

Niu F, O'Neill Z, O'Neill C (2018). Data-driven based estimation of HVAC energy consumption using an improved Fourier series decomposition in buildings. *Building Simulation*, 11: 633–645.

O'Neill Z, O'Neill C (2016). Development of a probabilistic graphical model for predicting building energy performance. *Applied Energy*, 164: 650–658.

Pachauri RK, Meyer LA (2014). Climate change 2014. Synthesis report. Intergovernmental Panel on Climate Change and World Meteorological Organization, Geneva, Switzerland.

Paulus MT, Claridge DE, Culp C (2015). Algorithm for automating the selection of a temperature dependent change point model. *Energy and Buildings*, 87: 95–104.

US DOE (2002). International performance measurement & verification protocol (IPMVP). Concepts and options for determining energy and water savings. United States Department of Energy. Available at https://www.nrel.gov/docs/fy02osti/31505.pdf. Accessed Jun 2018.

US DOE (2018). Energy savings performance contracting. Office of Energy Efficiency & Renewable Energy. United States Department of Energy. Available at https://www.energy.gov/eere/femp/energy-savings-performance-contracts-federal-agencies. Accessed Jun 2018.

US EIA (2016). How much energy is consumed in residential and commercial buildings in the United States? United States Energy Information Administration, U.S. Department of Energy. Available at https://www.eia.gov/tools/faqs/faq.php?id=86&t=1. Accessed Jun 2018.

US EPA (2019). Energy Star. Use Portfolio Manager. Available at https://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager. Accessed Jan 2019.

Zhang Y, O'Neill Z, Dong B, Augenbroe G (2015). Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment*, 86: 177–190.