CrossMark

Residential Building Energy Consumption: a Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods

Huyen Do[1,2] · Kristen S. Cetin[3]

Abstract

Purpose of Review Residential energy performance prediction has historically received less attention, as compared to commercial buildings. This likely is in part due to the limited availability of residential energy data, as well as the relative challenge of predicting energy consumption of buildings that are more highly dependent on occupant behavior. The purpose of this effort is to assess the types and characteristics of energy and non-energy data available for algorithm developed and methods that have been developed to predict residential consumption.

Recent Findings While there are several large residential building energy datasets, data availability is still generally very limited. Most energy prediction methods used recently include data-driven approaches, as well as combinations of multiple methods; however, many methods have not been tested for residential buildings, or at a range of energy data frequencies.

Summary The literature points to the need for the availability of more residential building data sources to be able to assess and improve models, and further testing is needed including those models that have not yet been significantly used for residential buildings.

Introduction

Energy consumption has significantly increased in recent years, particularly in buildings, growing at a rate of approximately 0.9% per year in the USA [1]. Consistently, the residential buildings consume approximately 38% of electricity and 21% of this energy [2]. Given buildings' overall significant contribution to energy use, as well as environmental

This article is part of the Topical Collection on *End-Use Efficiency*

✉ Huyen Do
huyendo@iastate.edu

Kristen S. Cetin
kcetin@iastate.edu

1 Iowa State University, 813 Bissell Road, 493 Town Engineering, Ames, IA 50011, USA

2 University of Danang, University of Science and Technology, Danang 550000, Vietnam

3 Iowa State University, 813 Bissell Road, 428 Town Engineering, Ames, IA 50011, USA

concerns and climate change, methods are needed to reduce this consumption. This is particularly the case for residential buildings, whose operation is highly dependent on occupants and their behavior [3–6]. There are many possible strategies to reduce energy use in residential buildings; the most common of which is through retrofitting an existing building with more energy efficient systems. What retrofits are completed is often a decision made by the homeowner, based on a variety of factors [7, 8]. While non-energy-related factors can be influential in making such decisions [9], the most strongly cited reason is costs, i.e., the economics of the upfront costs, rebates, or incentives provided, and the energy savings that the retrofit(s) will achieve over time. Another method to reduce consumption is through occupant energy behavior interventions, which aim to reduce consumption through altering the behavior of occupants, particularly how they use energy-consuming systems [10].

The ultimate decision of the homeowner to implement retrofits or change occupants' behaviors can depend on the information provided on quantification of the energy and costs savings achieved as a result of interventions, particularly if cost is the driving factor [11]. This includes (a) prediction of

consumption of the building in its existing state, (b) prediction of energy consumption after interventions, as well as (c) how their relative difference translates into energy and costs savings [12]. Therefore, building energy use prediction methods, used to determine (a), (b), and ultimately (c), play a highly important role in building energy conservation [13]. The methods proposed and used in recent literature to predict energy use of buildings range in complexity and the frequency and duration of input data needed. Some methods have also been developed and tested only for specific building types. A recent review of these methods and their use for different types of data and building applications is thus needed, particularly as the availability and range of types of data to develop these algorithms is highly limited.

This work reviews two critical topics in this research area. This includes, first, a review of known available data and published information, which is relevant for use in the development of methods to predict residential building consumption. This review includes the type(s), frequency, quality, and duration of data, as well as identifies the challenges and needs in the area of building energy datasets. Second is a review of recent published literature on the methods used to predict the energy consumption in residential buildings, as well as those developed for other building types that could be applied to residential buildings. This concludes with the limitations of existing data and methods, and future research needs in this area.

Residential Building Energy and Non-energy Data: Sources, Availability, and Characteristics

Critical to the ability to develop, test, and validate methods to predict building energy use is the availability of data for algorithm development. This includes real building energy data, as well as non-energy data, such as characteristics of the building(s) and their occupants, and/or weather data, all of which have demonstrated impacts on energy use. For residential buildings, much of this information can be challenging to obtain, particularly for a large number of buildings. This section is divided into two main sub-sections, including first, an overview of residential energy data and second, residential non-energy use data. Both these sub-sections review the sources of data, data availability, and characteristics of datasets, such as frequency, quality, and duration.

Residential Energy Data

Historical energy use data includes electricity use, gas use, and in some cases, other fuel use data collected at a regular frequency. This historical data is used in many cases, to train, test, and validate building energy use prediction

methodologies. One of the greatest sources of energy data is electric and gas utilities, which maintain large sets of energy use data from their residential customers. This is collected and stored at a minimum frequency of the monthly level for all residential buildings, with some locations having higher frequency data from utilizing AMR (automatic meter reading) and AMI (advanced metering infrastructure) technologies [14, 15]. However, the barriers associated with the use of this energy use data particularly for residential buildings are often privacy and law-related [16]. There are a small number of exceptions such as the city of Gainsville, Florida [17, 18], which provides public access to 6 years of monthly electricity and gas consumption data for all homes in the city; however, this type and availability of energy data is not common.

This means that in many cases, methods for predicting building energy use must often be developed and tested using limited data based on energy measurements from small number of occupied homes, energy measurements from real building(s) using simulated occupancy methods (e.g., using [19]), or energy use data based on simulated buildings resulting from a building energy modeling program such as EnergyPlus. While these real residential building data provide valuable information, larger datasets of real data can encompass energy use information for wider variety of home types, locations, occupant behaviors, and other natural variations in energy consumption that smaller datasets cannot. Given the significant variations in energy and occupant patterns that can occur in residential buildings, this can be beneficial to provide a more comprehensive understanding of how well a methodology works in comparison to others.

An alternative source of the utility energy use data collection is obtaining this information directly from homeowners, who have access to utility-collected monthly data and in some cases, 15 min or hourly data if a smart meter is installed in their home [20]. In rare cases, homeowners may have minute, sub-minute, or sub-metered data from a home energy monitoring system; however, these systems are not common currently. Thus, with homeowner consent, energy data can be obtained for algorithm development. However, large-scale collection of this information is time-consuming and costly. There are, however, some efforts towards more open access to energy use data, some available datasets, as well as broader platforms created to enable easier sharing of datasets.

Arguably, more information is currently available on commercial building energy use than for residential buildings. For commercial buildings, there are more policies supporting the public availability of energy information, particularly in large cities and for publically owned buildings. Large cities such as Boston [21], New York City [22], and Washington D.C. [23] among others have enacted laws and/or ordinances requiring energy benchmarking. Under these laws, buildings must report energy consumption on a regular basis, which is compiled into databases and often made publically available. In some

cases, such as Boston [21, 24], this data includes larger non-residential and multi-family residential buildings. However, the data in these datasets is also only reported at the annual level which has limited use for building energy prediction methods.

Similar benchmarking efforts could also be beneficial for residential buildings, particularly if the data was at an appropriate level of frequency. For example, the ECAD Ordinance [25] requires that all residential buildings bought and sold that are over 10 years of age to have an energy audit completed in Austin, TX; the results of which are compiled into a centralize database; the city of Chicago allows for disclosure of energy use and/or costs during the sale of a home [26]. These and other policy-enforced energy data sources could be highly valuable. Some local policy-enforced data sources are available, such as energy use by census block in Chicago [27], energy use by zipcode in New York [28–30], and aggregated annual energy use savings for homes in Austin [31]. However, these datasets are also aggregated and in most cases, only at the annual level.

Other efforts collect data from a variety of sources on commercial and/or residential energy use in a common location. The Building Information Database [32], supported by the US Department of Energy (DOE) consists of datasets of residential and commercial building energy use intensity on an annual basis, building characteristics and systems, and location. Similarly, the DOE-supported Building Dataset [33] contains information on energy use, building operations and analysis tools for buildings-related datasets, and the Energy Data Resources site [34] collects information on sources of energy data and tools from energy-related projects. The types of data vary, but do include datasets with energy consumption at varying levels of frequency.

There are a small number of datasets of residential energy use information that provide higher frequency and in some cases, disaggregated end use energy data for residential buildings. A large-scale study in the Pacific Northwest in the 1980s and 1990s collected whole-home and end-use data for residential buildings [35]. Many research papers were written based on this dataset, and the aggregated data is available online [36]. The results of this effort are also still used today in residential energy modeling programs [37, 38] for end-use modeling. The most recent US large-scale data collection effort for residential building data known to the authors is in Austin Texas [39••]. This database provides up to 1-min interval electricity and gas consumption for a large number of homes from 2012 to present and includes whole-home and end-use consumption. A number of recent research papers have used this to study residential energy use [40–42]. Given the current cost of equipment needed to obtain higher frequency and disaggregated data, it is unlikely that other efforts of this scale will occur frequently moving forward. However, given recent efforts to improve the ease of energy data equipment installation and collection, as well as improved abilities to disaggregate energy use data using higher-frequency whole-home energy data (e.g., [43]), lower-cost tools and/or equipment to obtain the frequency and quality of energy data for larger number of residential buildings may be more feasible moving forward.

Non-energy Data

Non-energy data, linked with the energy data, also has an important role in energy use prediction. Weather data is among the most critical non-energy factors impacting residential building energy use and particularly HVAC systems which are used in a high percentage of US residential buildings. Weather data is often available from public sources of ground-based weather station data, most commonly at airports [41, 44, 45]. However, as some recent research efforts have found, this weather data is not necessarily representative of the conditions where studied residential buildings are located. For example, recent efforts have found variations in localized wind speeds and temperatures (e.g., [46]). The state of the art in this general area has been summarized in several recent research articles (e.g., [47, 48]), and thus is not the focus of discussion herein. However, it is still important to note that while modeling methods and research efforts in building microclimates are significant, accessibility to raw weather data that well represents the actual conditions experienced by buildings is still a challenge. More recently, some fields of study have adopted the use of publically available satellite data-based weather data from MERRA [49], which is available worldwide on a regularly spaced grid. The use of this dataset reduces the dependency on ground-based weather stations.

Building characteristics, such as size, fuel type, HVAC system type, age, efficiency, appliances types, thermostat preferences, air exchange rate, and building envelope characteristics can also have a strong impact on energy consumption. Thus, while knowing this information can be highly beneficial, in many cases, this information is not available or linked with building-specific energy use data. The best publically available sources of building data originate from disparate sources, including assessors data, MLS data, cities' GIS databases, and LIDAR data. However, if energy use datasets are anonymized for privacy reasons, this makes linking energy and non-energy datasets very challenging.

Some datasets, such as national-level datasets US Census data [50], RECS data [51], and American Community Survey [52] data, and localized datasets such as the Green Building Aggregate data in Austin, Texas [31], provide aggregate-level residential building and occupant characteristic data for enabling an understanding of building characteristics at a broader scale than the building level. The Better Building Neighborhood Program [53] provides

a large anonymized building-level dataset representing over 75,000 building energy-related characteristics, specified by region and zipcode information. This and the aggregated datasets can be useful to determine likely characteristics of a building in a specific area, or for use in community-scale energy use prediction methods (e.g., [54]), but is of limited benefit to building-level energy consumption prediction as they are not linked to specific residential building energy use data. The datasets mentioned in the previous section, including the Building Information Database [32], the Building Dataset [33], and the Energy Data Resources dataset [34], do contain some building energy use information linked to building characteristic data.

In summary, building energy data and non-energy datasets are available; the characteristics of which range significantly. There are some promising sources of quality and higher frequency data which can be valuable for residential energy consumption prediction methods. There are also promising methods to encouraging sharing of data that can be further explored. However, significant opportunities remain to improve data availability in this field, which if done, will be highly beneficial to improvements in the capabilities of energy performance prediction methods.

Building Energy Performance Prediction Methods

Using energy data and non-energy data sources, building energy performance prediction methods range significantly in complexity and required types and frequencies of input data. Most recent efforts have followed similar methodologies for model development, including, as discussed in Wang and Srinivasan [13], first, (a) the collection of data for model development, then (b) the raw data processing is completed to ensure the data is of sufficient quality and format. The third step (c) includes using historical data to train the model to follow the patterns of use associated with the training dataset, as well as determining what of the available input data is significant and ultimately used for the model. The final step is (d) model testing. The fit of the model to input data not included in the training dataset is determined and evaluated in this step. Common metrics and statistical indices utilized for evaluation include root mean square error, coefficient of determination, coefficient of variation of the root mean square error, sum of squares error, mean squared error, and normalized mean bias error. Energy use prediction methods can either be physics-based approaches, data-driven inverse modeling approaches, or a combination of the two [55•]. In this section, the most recent efforts in energy performance prediction methods are reviewed, most of which are data-driven methods.

Change-Point Modeling

Change-point modeling is among the more simple methods, which are typically single-variate models using dry-bulb temperature as the predictor. A balance point is determined which best fits the trends in the energy data, where building energy use switches between seasonal trends [55•]. Linear regression is then used to create a multi-parameter model based on the determined level of fit criteria [56••, 57]. Perez et al. [58] focused on its use to predict daily consumption of residential HVAC systems in Austin, TX, using data from [39••]. Kim and Haberl [59] used three-parameter change-point models to calibrate daily whole-building energy simulations for two single-family homes based on monthly billing data. Do et al. [40, 60] utilized large number of homes across multiple climate zones to study the use of change point models, demonstrating these methods can fit to a wide range of homes' use patterns. Zhang et al. [56••] used it to predict hourly and daily HVAC hot water energy and Abushakra and Paulus [61–63] used a hybrid inverse change-point model to predict consumption in simulated and actual buildings; however, both these efforts focused on commercial buildings.

The strength of the change-point models is the simpler development with lower computational effort in comparison to other methods [55•, 56••]. The accuracy of prediction in change-point models depends on the type and frequency of data available, but has been shown to demonstrate similar levels of accuracy to more complex models in some situations [56••]. Particularly for buildings with a limited number of data points, this method can be advantageous. However, as discussed in [40, 59], some data points can be considered outliers that may significantly impact the model fit, particularly for highly occupant-dependent residential buildings. With acceptable methods to assess what data is appropriate to use for residential building models as well model improvements such as those suggested by Abushakra and Paulus [61–63], this modeling method provides a simpler but often sufficiently accurate method.

Artificial Neural Networks

Artificial Neural Networks (ANN) consist of an input layer, one or more hidden layers, and an output layer, and have mostly been used for more frequent, hourly or sub-hourly building energy consumption prediction in recent literature [56••,64•]. Input variables typically include outdoor temperature, wind speed, solar radiation, and relative humidity. These methods have been used to predict whole-home HVAC, and appliance use in residential buildings [64•, 65], and hot water [56••], heating energy [66], total electricity [54, 67], and chilled water use [68] for commercial buildings. ANN has also been combined with other methods and/or enhancements, including feed forward backpropagation neural network,

radial basis function network, and adaptive neuro-fuzzy interference system [66], backpropagation algorithms [64•,69], particle swarm optimization and genetic algorithms [54], principal component analysis [54, 70], and hybrid lightning search algorithms [65] to improve and/or optimize performance.

ANN generally performs well with sufficient training data and can be advantageous particularly for non-linear electricity consumption [64•, 68]. Wang and Srinivasan [13] also found performance of ANN methods in short-term prediction is better than regression methods. Improvements made to ANN methods also further improve accuracy [54, 71] with lower error [70]. However, the complexity of the model also increases computational time [72] and has limited physical interpretation which limits applicability outside of the training data limits [13]. In some cases, ANN has also been found to perform worse than simpler models [56••]. ANN has only been used in recent literature to predict whole-home consumption of unoccupied rather than occupied residential buildings [64•].

Genetic Programming

Genetic programming is an automated computational method based on the process of biological evolution [73] and has been used in combination with other methods to predict residential energy consumption. Castelli et al. [73] applied different genetic programming systems that use the genetics semantic operators to predict residential HVAC use. Jung et al. [74] used genetic programming with a hybrid of the direct search optimization algorithm and a conventional real-coded genetic algorithm, with least-squares support vector machine to predict daily commercial building energy. Genetic programming has been shown to be an effective method that produces lower errors than other methods [73] and to also provide an effective approach for parameter selection and better performance in terms of convergence time and iteration than conventional least-squares support vector machine methods. However, similar to ANN, genetic programming typically requires a larger set of input data. It also has only been used in limited studies for residential buildings.

Bayesian Networks

Bayesian Network models include nodes that represent random variables such as outdoor temperature and energy use with statistical and probabilistic dependencies between the cause nodes and the effect nodes with a probabilistic graphical model [76]. The parameters of such models are the conditional distributions at every node using Bayes' rule. This method has been used to predict appliance energy use in residential buildings [75] and hot water HVAC use in a commercial building [76]. Bassamzadeh and Ghanem [77] also used this model to forecast the aggregated electricity demand in smart grids. In

the limited number of studies that have used this method for building energy use prediction, the accuracy of the model predictions was within the recommended limits developed by ASHRAE for commercial buildings [76]. The uncertainties from input variables were also determined to be well represented using this type of method [77]. However, similar to the ANN and genetic algorithm methods discussed above, this method requires significant input data and can be highly complex to implement.

Gaussian Mixture Model

Gaussian mixture model (GMM) establishes a weighted sum of Gaussian component densities based on a parametric probability density function and multivariate non-linear regression function [56••]. This method has been used in a number of recent studies for a range of buildings. Li et al. [78] utilized GMM to design feasible time-of-use tariffs to minimize the electricity bills for residential customers. Also, in residential buildings, and Melzi et al. [79] used GMM to optimize smart meter electricity consumption, better understand consumer behavior and electricity use profiles. For other types of buildings, Zhang et al. [56••] applied GMM to predict daily and hourly commercial hot water energy and Carpenter et al. [80•] predicted supplied energy for a range of manufacturing processes in an industrial building. The advantage of this method found in [56••] was that it results in energy performance predictions that had the lowest error compared to change-point and ANN models, for commercial buildings. The GMM has also been found to capture non-linearity in simpler way than Bayesian or ANN methods [56••, 80•] for non-residential buildings. However, its performance in comparison to other methods for residential buildings is not well studied. Studies have also found that other statistical values of fitness are also higher for GMM than change-point modeling [80•].

Support Vector Machines

The final modeling method discussed is Support Vector Machines (SVM). This method has been shown to be effective in solving regression estimation problems and forecasting time series [72]. Jain et al. [81] used a version of SVM for regression estimation, Support Vector Regression, to evaluate the effect of temporal and spatial granularity of data on the prediction of energy in multi-family buildings. SVM has also been combined with genetic algorithms to predict energy use [74]. SVM has been assessed as a highly accurate and effective method for the energy prediction [72]. However, SVM requires multi-step forecasts, implemented using various features and selected techniques [81]; therefore, it is more complicated and requires more computational effort in comparison to other models discussed. Similar to other methods, it can

also benefit from additional evaluation for residential building energy performance prediction methods.

In summary, there are a number of different types of methods used in recent literature to predict energy consumption of residential buildings. Table 1 represents the summary of six main methods of building energy performance prediction. However, particularly for residential buildings, it is challenging to compare the capabilities and determine the overall "best" model for use for residential energy performance prediction, in part, due to the lack of studies that compare performance of the models using residential datasets. Many of the algorithms have been developed, utilized, and tested for commercial building applications, and may be well suited for residential buildings as well. Some residential building energy prediction studies have used larger datasets [58, 77]; however, the number of studies with this size dataset is limited, for both residential and commercial buildings. The type of energy data being predicted also varies. Some studies focus on the use of methods to predict whole-building consumption [54, 67], while others focus on HVAC [58], or other end uses [75]. Finally, the frequency of data and type of energy use data used to develop and test these models ranges significantly.

Conclusions

In summary, this review discusses both sources of energy and non-energy data, as well as methods that use these data to predict energy consumption. This review points to the need for the availability of more residential building energy and non-energy data sources to be able to improve energy performance prediction models, and the need to more comprehensively and comparatively study the accuracy of these models for residential buildings across a range of frequencies of data, and whole-home as well as end-use consumption. More specifically, the following conclusions can be drawn:

- Most available datasets provide energy or non-energy data; however, these are generally not linked together or do not have the ability to be linked as they are anonymized; this limits the usability of these datasets for energy use prediction methods. Datasets that link energy and non-energy data are needed and with higher frequencies and quantities of data
- Many available national-level and local-level datasets of energy use provide annual level data. Given that energy use prediction methods are often developed with the goal of

Table 1 Summary of the building energy performance prediction methods

No.	Method	Most common data frequency	Advantages	Disadvantages	References
1	*Change-Point Models*: supervised machine learning; single-variate or in some cases multi-variate steady-state model including a balance point and weather variable(s) as the predictor(s)	Monthly, daily	Simpler; lower computational effort; easy to interpret and explain results	Outlier(s) may impact the model fit	[39••, 40, 55•, 56••, 57–63]
2	*Artificial Neural Networks*: supervised machine learning; consists an input layer, one or more hidden layers, and an output layer; can be combined with other methods and/or enhancements	Daily, hourly, sub-hourly	Works well for non-linear consumption data	Higher computational demand, limited physical interpretation; requires significant input training data; over or under fitting	[13, 54, 55•, 56••, 64•, 65–72]
3	*Genetic Programming*: evolutionary algorithm; an automated computational method based on the process of biological evolution	Daily, hourly	Effective in parameter selection, convergence time, and iteration	Requires significant input training data, higher computational demand	[73, 74]
4	*Bayesian Networks*: probabilistic graphical model; includes cause nodes and effect node with a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph	Hourly, sub-hourly	Ability to assess uncertainties	Requires significant input training data	[75–77]
5	*Gaussian Mixture Models*: probabilistic model; unsupervised learning; a weighted sum of Gaussian component densities based on parametric probability density function and multivariate non-linear regression function	Monthly, daily, hourly	Captures non-linearity in simpler way than Bayesian Networks or Artificial Neural Networks	Statistical values of fitness have been found to be better than change-point modeling in some studies	[56••, 78, 79, 80•]
6	*Support Vector Machines*: supervised machine learning; solves classification and regression estimation problems	Daily, hourly	Less prone to overfitting than some other supervised methods	Higher computational effort with multi-step forecasts	[72, 74, 81]

predicting energy use at higher frequencies, this limits the data usability. There are some recent efforts to make large-scale studies' data and law-mandated data available; however, more efforts are needed in this area, including those datasets associated with publications in this area, almost none of which are available for broader use. Recent efforts to improve the infrastructure, ease and motivation for energy data sharing [82, 83], may help to improve this moving forward

- Further and more comprehensive testing is needed to assess the different energy prediction methods at different data frequencies; this will help to assess which models are most appropriate and best able to predict consumption for each frequency level, as this is currently not well established

- Similarly, many of the prediction methods discussed have been tested for commercial buildings more than for residential, and in many cases, only tested for specific end uses; testing of the possible methods across larger sets of diverse residential buildings could provide a more comprehensive picture of capabilities of these methods

- The complexity of prediction models ranges significantly, as well as the amount of input data needed. Further clarity is needed as to the positives and negatives associated with more complex versus less computationally complex methods

As more technologies become available that connect to the internet and are able to collect energy and non-energy data, such as through the internet of things, there is a significant opportunity to improve energy prediction methods. As energy efficiency continues to be a priority, improved data combined with improvements in prediction algorithms using this data will help to improve the accuracy and reliability of such models, and as a result, likely drive efficiency improvements as well.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. United States Energy Information Administration (US EIA). Total energy. Annual Energy Review. U.S. Department of Energy (DOE). 2017. https://www.eia.gov/totalenergy/data/annual/index.php. Accessed 10 Nov 2017.

2. United States Energy Information Administration (US EIA). How much energy is consumed in residential and commercial buildings in the United States? U.S. Department of Energy (DOE). 2016. https://www.eia.gov/tools/faqs/faq.php?id=86&t=1. Accessed 10 Nov 2017.

3. Hong T, Taylor-Lange SC, D'Oca S, Yan D, Corgnati SP. Advances in research and applications of energy-related occupant behavior in buildings. Energ Building. 2016;116:694–702. https://doi.org/10.1016/j.enbuild.2015.11.052.

4. Gaetani I, Hoes PJ, Hensen J. Occupant behavior in building energy simulation: towards a fit-for-purpose modeling strategy. Energ Building. 2016;121:188–204. https://doi.org/10.1016/j.enbuild.2016.03.038.

5. Aksanli B, Akyurek AS, Rosing TS. User behavior modeling for estimating residential energy consumption. Smart City 360°. Springer International Publishing. 2015. p. 348–361. https://doi.org/10.1007/978-3-319-33681-7_29

6. Yan D, O'Brien W, Hong T, Feng X, Gunay HB, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: current state and future challenges. Energ Building. 2015;107:264–78. https://doi.org/10.1016/j.enbuild.2015.08.032.

7. Langheim R, Arreola G, Reese C. Energy efficiency motivations and actions of California solar homeowners. 2014 ACEEE Summer Study on Energy Efficient in Buildings. Pacific Grove, CA 2014. p. 1–13.

8. Udalov V, Perret J, Vasseur V. Environmental motivations behind individuals' energy efficiency behavior: evidence from Germany, the Netherlands and Belgium. 40th Annual IAEE International Conference. 2017. http://www.iaee.org/iaee2017/submissions/Presentations/udalov_presentation.pdf.

9. Im J, Seo Y, Cetin KS, Singh J. Energy efficiency in U.S. residential rental housing: adoption rates and impact on rent. Appl Energ. 2017;205:1021–33. https://doi.org/10.1016/j.apenergy.2017.08.047.

10. D'Oca A, Fabi V, Corgnati SP, Andersen RK. Effect of thermostat and window opening occupant behavior models on energy use in homes. Build Simul. 2014;7(6):683–94. https://doi.org/10.1007/s12273-014-0191-6.

11. Cetin KS, Siemann M, Sloop C. Disaggregation and future prediction of monthly residential building energy use data using localized weather data network. 2016 ACEEE Summer Study on Energy Efficient in Buildings. Pacific Grove, CA. 2016 p. 12:1–12.

12. ASHRAE Guideline 14-2014. Measurement of energy, demand, and water savings. Standard by ASHRAE. 2014.

13. Wang Z, Srinivasan RS. A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. Proceedings of the 2015 Winter Simulation Conference of the IEEE 2015. p. 3438–3448.

14. Wilde P. The gap between predicted and measured energy performance of buildings: a framework for investigation. Autom Constr. 2014;41:40–9. https://doi.org/10.1016/j.autcon.2014.02.009.

15. Hammon R, Narayanamurthy R, Clarin B, VonKorf H, Herro CR, Dock A. Field evaluation of long-term performance of energy-efficient homes. 2016 ACEEE Summer Study on Energy Efficient in Buildings. Pacific Grove, CA; 2016. p. 2:1–11.

16. Abrams Environmental Law Clinic. Freeing energy data—a guide for regulators to reduce one barrier to residential energy efficiency. University of Chicago Law School 2016. https://www.law.uchicago.edu/files/file/freeing_energy_data_report_abrams_environmental_clinic_june_2016.pdf. Accessed 10 Nov 2017.

17. City of Gainesville, FL. GRU customer electric consumption. https://data.cityofgainesville.org/Better-Future/GRU-Customer-Electric-Consumption/gk3k-9435. Updated 03 Nov 2017. Accessed 10 Nov 2017.

18. City of Gainesville, FL. GRU customer natural gas consumption. https://data.cityofgainesville.org/Better-Future/GRU-Customer-

Natural-Gas-Consumption/dbcj-nniz. Updated 03 Aug 2017. Accessed 10 Nov 2017.

19. Spam B, Hudon K, Earle L, Booten C, Tabares-Velasco PC, Barker G, Hancock CE. Greenbuilt retrofit test house final report. National Renewable Energy Laboratory (NREL). 2014. https://www.nrel.gov/docs/fy14osti/54009.pdf.

20. Cooper A. Electric company smart meter deployments: Foundation for a smart grid. IEI report. The Edison Foundation—Institute for Electric Innovation. 2016. http://www.edisonfoundation.net/iei/publications/Documents/Final%20Electric%20Company%20Smart%20Meter%20Deployments-%20Foundation%20for%20A%20Smart%20Energy%20Grid.pdf.

21. The city of Boston. Building energy reporting and disclosure ordinance (BERDO). https://www.boston.gov/environment-and-energy/building-energy-reporting-and-disclosure-ordinance. Updated 3 Nov 2017. Accessed 10 Nov 2017.

22. The city of New York. NYC Resource. Office of Sustainability. http://www.nyc.gov/html/gbee/html/home/home.shtml. Accessed 10 Nov 2017.

23. Department of Energy & Environment. Washington D.C. https://doee.dc.gov/energybenchmarking. Accessed 10 Nov 2017.

24. The City of Boston. Building energy reporting and disclosure ordinance (BERDO). https://data.boston.gov/dataset/building-energy-reporting-and-disclosure-ordinance. Accessed 10 Nov 2017.

25. Austin Energy. Energy conservation audit and disclosure (ECAD) ordinance. The city of Austin https://austinenergy.com/ae/energy-efficiency/ecad-ordinance/energy-conservation-audit-and-disclosure-ordinance. Accessed 10 Nov 2017.

26. Philbrick D, Scheu R, Blaser J. Moving the market: energy cost disclosure in residential real estate listings. ACEEE Summer Study on Energy Efficiency in Buildings. 2016. p. 7:1–12.

27. City of Chicago. Energy usage 2010. https://catalog.data.gov/dataset/energy-usage-2010-24a67. Updated 16 June 2017. Accessed 10 Nov 2017.

28. City of New York, NY. Natural gas consumption by zip code—2010. https://catalog.data.gov/dataset/natural-gas-consumption-by-zip-code-2010-0329b. Updated 23 Sep 2017. Accessed 10 Nov 2017.

29. City of New York, NY. Heating gas consumption and cost (2010–2016). https://data.cityofnewyork.us/Housing-Development/Heating-Gas-Consumption-And-Cost-2010-2016-/it56-eyq4. Updated 22 Feb 2017. Accessed 10 Nov 2017.

30. City of New York, NY. Electric consumption and cost (2010–2016). https://data.cityofnewyork.us/Housing-Development/Electric-Consumption-And-Cost-2010-2016-/jr24-e7cr. Updated 22 Feb 2017. Accessed 10 Nov 2017.

31. City of Austin, TX. Green building aggregate data. 2017. https://catalog.data.gov/dataset/green-building-aggregate-data. Updated 23 Sep 2017. Accessed 10 Nov 2017.

32. U.S. Department of Energy (DOE). Lawrence Berkeley National Laboratory. https://bpd.lbl.gov/#explore. Accessed 10 Nov 2017.

33. U.S. Department of Energy (DOE). Building Technologies Office. Building operations data. https://trynthink.github.io/buildingsdatasets/. Accessed 10 Nov 2017.

34. Duke University, Durham, NC. Energy data resources. https://energy.duke.edu/research/energy-data/resources. Accessed 10 Nov 2017.

35. Pratt RG, Conner CC, Cooke BA, Richman E. Metered end-use consumption and load shapes from the ELCAP residential sample of existing homes in the Pacific Northwest. Energ Building. 1993;19(3):179–293. https://doi.org/10.1016/0378-7788(93)90026-Q.

36. The Regional Technical Forum (RTF), the Cadmus Group, Inc. The ELCAP (End Use Load and Consumer Assessment Program) database. https://elcap.nwcouncil.org/. Accessed 10 Nov 2017.

37. Wilson E, Metzger CE, Horowitz S, Hendron R. 2014 Building America House Simulation Protocols. Technical Report. National Renewable Energy Laboratory (NREL). 2014. https://energy.gov/sites/prod/files/2014/03/f13/house_simulation_protocols_2014.pdf. Accessed 10 Nov 2017.

38. Tabares-Velasco PC, Maguire J, Horowitz S, Christensen C. Using the beopt automated residential simulation test suite to enable comparative analysis between energy simulation engines. The 2014 ASHRAE/IBPSA-USA Building Simulation Conference. Atlanta, Georgia 2014. https://www.nrel.gov/docs/fy14osti/62273.pdf. Accessed 10 Nov 2017.

39.•• The Pecan Street Research Institute. The University of Texas at Austin, TX. The Dataport database. http://www.pecanstreet.org/category/dataport/. Accessed 10 Nov 2017. **Provides a large dataset of whole-home and end-use energy consumption in different frequencies in residential buildings.**

40. Do H, Cetin KS, Andersen T. Characteristics and causes of outliers in inverse modeling of residential building energy use data. ASHRAE Winter Conference, Chicago, IL. 2018. In press.

41. Cetin KS, Manuel L, Novoselac A. Effect of technology-enabled time-of-use energy pricing on thermal comfort and energy use in mechanically-conditioned residential buildings in cooling dominated climates. Build Environ. 2016;96:118–30. https://doi.org/10.1016/j.buildenv.2015.11.012.

42. Parson O, Fisher G, Hersey A, Batra N, Kelly J, Singh A, Knottenbelt W, Rogers A. Dataport and NILMTK: a building data set designed for non-intrusive load monitoring. 2015 I.E. Global Conference on Signal and Information Processing (GlobalSIP). Orlando, FL 2015. https://doi.org/10.1109/GlobalSIP.2015.7418187.

43. John J. Startup goes public with its energy disaggregation results. Software & Analytics. GTM Research. 2015. https://www.greentechmedia.com/articles/read/eeme-goes-public-with-energy-disaggregation-test-results#gs.0zUJRcI. Updated 13 Mar 2015. Accessed 10 Nov 2017.

44. Cetin KS, Novoselac A. Single and multi-family residential central all-air HVAC system operational characteristics in cooling-dominated climate. Energ Building. 2015;96:210–20. https://doi.org/10.1016/j.enbuild.2015.03.039.

45. Cetin KS, Valesco P, Novoselac A. Appliance daily energy use in residential buildings: use profiles and variation in time-of-use. Energ Building. 2014;84:716–26. https://doi.org/10.1016/j.enbuild.2014.07.045.

46. Srebric J, Heidarinejad M, Liu J. Building neighborhood emerging properties and their impacts on multi-scale modeling of building energy and airflows. Build Environ. 2015;91:246–62. https://doi.org/10.1016/j.buildenv.2015.02.031.

47. Santamouris M, Cartalis C, Synnefa A, Kolokotsa D. On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review. Energ Building. 2015;98:119–24. https://doi.org/10.1016/j.enbuild.2014.09.052.

48. Akbari H, Cartalis C, Kolokotsa D, Muscio A, Pisello AL, Rossi F, et al. Local climate change and urban heat island mitigation techniques—the state of the art. J Civ Eng Manage. 2016, 22;(1):1–16.

49. Goddard Space Flight Center. Modern-era retrospective analysis for research and applications, version 2 (MERRA2). Global Modelling and Assimilation Office. National Aeronautics and Space Administration. https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. Updated 10 May 2017. Accessed 10 Nov 2017.

50. The U.S. Census Bureau. U.S. Department of Commerce https://www.census.gov/data.html. Accessed 10 Nov 2017.

51. United States Energy Information Administration (US EIA). Residential energy consumption survey (RECS). 2015. https://www.eia.gov/consumption/residential/data/2015/. Accessed 10 Nov 2017.

52. The American Community Survey (ACS). The U.S. Census Bureau. U.S. Department of Commerce. https://www.census.gov/programs-surveys/acs/data.html. Accessed 10 Nov 2017.

53. U.S. Department of Energy (DOE). Office of Energy Efficiency & Renewable Energy (EERE). Better buildings neighborhood program single-family home upgrade project dataset. https://openei.org/datasets/dataset/better-buildings-neighborhood-program-single-family-home-upgrade-project-dataset. Accessed 10 Nov 2017.

54. Li K, Hu C, Liu G, Xue W. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energ Building. 2015;108:106–13. https://doi.org/10.1016/j.enbuild.2015.09.002.

55.• The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). ASHRAE handbook—fundamentals. 2017. **Provide a good review of the fundamentals of inverse modelling techniques of energy consumption in buildings.**

56.•• Zhang Y, O'Neill Z, Dong B, Augenbroe G. Comparisons of inverse modeling approaches for predicting building energy performance. Build Environ. 2015;86:177–90. **Strong comparisons of energy prediction methods, including change-point models, Gaussian mixture model, and artificial neural networks.**

57. Paulus MT, Claridge DE, Culp C. Algorithm for automating the selection of a temperature dependent change point model. Energ Building. 2015;87:95–104. https://doi.org/10.1016/j.enbuild.2014.11.033.

58. Perez KX, Cetin K, Baldea M, Edgar TF. Development and analysis of residential change-point models from smart meter data. Energ Building. 2017;139:351–9. https://doi.org/10.1016/j.enbuild.2016.12.084.

59. Kim KH, Haberl JS. Development of methodology for calibrated simulation in single-family residential buildings using three-parameter change-point regression model. Energ Building. 2015;99:140–52. https://doi.org/10.1016/j.enbuild.2015.04.032.

60. Do H, Cetin KS. Impact of occupants behavior in data-driven energy use modelling in diverse residential buildings across multiple climates. The 4th Residential Building Design & Construction Conference. The Pennsylvania State University, PA. 2018. In press.

61. Abushakra B, Paulus MT. 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). Sci Technol Built Environ. 2016;22(7):977–83. https://doi.org/10.1080/23744731.2016.1215222.

62. Abushakra B, Paulus MT. 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). Sci Technol Built Environ. 2016;22(7):984–95. https://doi.org/10.1080/23744731.2016.1215199.

63. Abushakra B, Paulus MT. 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). Sci Technol Built Environ. 2016;22(7):996–1009. https://doi.org/10.1080/23744731.2016.1215659.

64.• Biswas M, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. Energy. 2016;117:84–92. **Introduces an approach of neural network inverse modelling for energy use prediction in residential buildings.**

65. Ahmed MS, Mohamed A, Homod RZ, Shareef H. Hybrid LSA-ANN based home energy management scheduling controller for residential demand response strategy. Energies. 2016;9(9):716. https://doi.org/10.3390/en9090716.

66. Jovanovic RZ, Sretenovic AA, Zivkovic BD. Ensemble of various neural networks for prediction of heating energy consumption. Energ Building. 2015;94:189–99. https://doi.org/10.1016/j.enbuild.2015.02.052.

67. Roldan-Blay C, Escriva-Escriva G, Alvarez-Be C, Roldan-Porta C, Rodriguez-Garcia J. Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model. Energ Building. 2013;60:38–46. https://doi.org/10.1016/j.enbuild.2012.12.009.

68. Deb C, Eang LS, Yang J, Santamouris M. Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. Energ Building. 2016;121:284–97. https://doi.org/10.1016/j.enbuild.2015.12.050.

69. Bocheng Z, Kuo L, Dinghao L, Jing L, Xuan F. Short-term prediction of building energy consumption based on GALM neural network. International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII). 2015. p. 867–71.

70. Plation R, Dehkordi VR, Martel J. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. Energ Building. 2015;92:10–8. https://doi.org/10.1016/j.enbuild.2015.01.047.

71. Kumar R, Aggarwal RK, Sharma JD. Energy analysis of a building using artificial neural network: a review. Energ Building. 2013;65:352–8. https://doi.org/10.1016/j.enbuild.2013.06.007.

72. Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sust Energ Rev. 2014;33:102–9. https://doi.org/10.1016/j.rser.2014.01.069.

73. Castelli M, Trujillo L, Vanneschi L, Popovic A. Prediction of energy performance of residential buildings: a genetic programming approach. Energ Building. 2015;102:67–74. https://doi.org/10.1016/j.enbuild.2015.05.013.

74. Jung HC, Kim JS, Heo H. Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach. Energ Building. 2015;90:76–84. https://doi.org/10.1016/j.enbuild.2014.12.029.

75. Basu K, Hawarah L, Arghira N, Joumaa H, Ploix S. A prediction system for home appliance usage. Energ Building. 2013;67:668–79. https://doi.org/10.1016/j.enbuild.2013.02.008.

76. O'Neill Z, O'Neill C. Development of a probabilistic graphical model for predicting building energy performance. Appl Energ. 2016;164:650–8. https://doi.org/10.1016/j.apenergy.2015.12.015.

77. Bassamzadeh N, Ghanem R. Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. Appl Energ. 2017;193:369–80. https://doi.org/10.1016/j.apenergy.2017.01.017.

78. Li R, Wang Z, Gu C, Li F, Wu H. A novel time-of-use tariff design based on Gaussian mixture model. Appl Energ. 2016;162:1530–6. https://doi.org/10.1016/j.apenergy.2015.02.063.

79. Melzi FN, Same A, Zayani MH, Oukhellou L. A dedicated mixture model for clustering smart meter data: identification and analysis of electricity consumption behaviors. Energies. 2017;10(10):1446. https://doi.org/10.3390/en10101446.

80.• Carpenter J, Woodbury K, O'Neill Z. A comparison of Gaussian process regression and change-point regression for the baseline model in industrial facilities. ASHRAE and IBPSA-USA SimBuild. Building Performance Modeling Conference, Salt Lake City, UT. 2016. p. 79–86. **A good comparison of inverse modeling methods Gaussian process regression and change-point.**

81. Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance

accuracy. Appl Energ. 2014;123:168–78. https://doi.org/10.1016/j.apenergy.2014.02.057.

82. Tolone W, Talukder A, Djorgovski S, Hadzikadic M, Tao Y, Al-Shaer E. CIF21 DIBBS: EI: VIFI: virtual information-fabric infra-structure (VIFI) for data-driven decisions from distributed data. National Science Foundation (NSF). Award number 1640818. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1640818. Accessed 10 Nov 2017.

83. Antin P, Lyons E, Merchant N, Micklos D, Vaughn M, Ware D. CyVerse. The project is funded by National Science Foundation (NSF) http://www.cyverse.org/about. Accessed 10 Nov 2017.