Y. Ye, W. Zuo, G. Wang 2019. “A Comprehensive Review of Energy-Related Data for U.S. Commercial Buildings.” Energy and Buildings, 186, pp. 126-137. DOI: 10.1016/j.enbuild.2019.01.020.

A Comprehensive Review of Energy-Related Data for U.S. Commercial Buildings

Yunyang Ye[a], Wangda Zuo[a, *], Gang Wang[b]

[a] Department of Civil, Environmental and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, U.S.A.

[b] Civil, Architectural and Environmental Engineering Department, University of Miami, Coral Gables, FL 33146, U.S.A

* Corresponding author. Email address: Wangda.Zuo@Colorado.edu (Wangda Zuo)

Abstract

U.S. commercial buildings consumed around 18% of total primary energy in 2017 and a 2.23 EJ increase is expected by 2050. Energy-related data for commercial buildings can be used for various applications, including benchmarking, building component analysis, market potential analysis, and policy making. Although there are plenty of data sources for energy usage in commercial buildings, they have not been thoroughly reviewed and summarized. As a result, users do not have comprehensive guidelines about selections of right data sources for specific application needs. To fill this gap, this paper conducts a comprehensive review to summarize data sources for energy usage in U.S. commercial buildings and discuss their usages for different applications. First, the paper summarizes the survey and simulation data sources for energy usage. The data sources are compared in terms of their data collection methods, released information, and relevant features. Second, this paper analyzes the applications for different survey and simulation data sources. This review categorizes the applications of data sources into five categories, including energy performance benchmarks, energy use forecasts and predictions, energy use contributions of building components, supports of building energy policies, and urban-scale energy use analysis. Moreover, the paper introduces several cases to demonstrate the usages of these data sources.

Key words: Commercial Buildings, Energy Consumption, Survey Data, Simulation Data, Applications

1. Introduction

The Energy Information Agency’s (EIA) Annual Energy Outlook 2018 estimates that the commercial building sector was responsible for around 18% of primary U.S. energy use in 2017 [1]. It also projected that the primary energy consumption of U.S. commercial buildings would increase 2.23 EJ by 2050. The comparison of energy use intensities (EUIs) between existing buildings and high-efficiency buildings showed a great potential to reduce the energy consumption in commercial buildings [2-7]. For example, Griffith, et al. [3] concluded that the site EUI with high-efficiency buildings, excluding photovoltaics (PV) panels, was 457.67 MJ/m^2-yr, which was only 45% of average site EUI of existing commercial buildings at the time of the study. Glazer [4] analyzed 272 building and climate combinations and stated that the high energy-efficient commercial buildings in the U.S. had the potential to consume only around 50% site energy compared to ASHRAE Standard 90.1-2013 [8]. To understand the broad picture of U.S. commercial building energy consumption, it is necessary to analyze the energy-related data for U.S. commercial buildings.

There are a rich set of data sources for energy-related data of U.S. commercial buildings. This paper focuses on the data sets where the data is easy to obtain and can be used for new research. This paper classifies these data sets into two categories – survey data sets and simulation data sets. Survey data sets in this paper collect the raw data mainly from surveys of building respondents and energy providers, on-site meters, utility bills, or other survey data sets. Then databases are created based on the data. Simulation data is generated by building energy simulation programs [9]. Model developers collect the building model inputs from survey data, energy standards, and expert knowledge. Then building energy simulation programs produce the results based on the building model inputs. This paper defines both building model inputs and simulation results as simulation data.

Survey data provides the possibility to overview building energy consumption in the commercial building sector. The federal and local governments and organizations summarized the energy-related characteristics of commercial buildings and energy usage data, and then built several databases [10-14]. Some data sources contain in-depth building characteristics and energy use for representative samples. For example, the Commercial Buildings Energy Consumption Survey (CBECS) is a nation's comprehensive survey that collects information on U.S. commercial buildings, including energy usage data and associated building characteristics [13]. Also, the California Commercial End-Use Survey (CEUS) is a popular state-comprehensive survey data source for energy consumption in California's commercial buildings [12]. The other data sources do not contain in-depth building characteristics and energy use for individual building samples. Instead, they pursue large-scale energy-related data like the Building Performance Database (BPD). BPD is a national database for the energy data of U.S. commercial and residential buildings based on multiple existing data sources [14, 15]. These data sources, including CBECS, CEUS, and BPD, focus on the overall commercial building sector. Meanwhile, some data sources only focus on specific types of commercial buildings. For instance, Labs 21 collects data only for laboratories across the country [16]. The New Building Institute (NBI) only collects the data from buildings certified by Leadership in Energy and Environmental Design (LEED) [17, 18].

Building energy simulation programs can generate energy-related simulation data for U.S. commercial buildings. Survey data, energy standards, existing building models, and expert knowledge are applied to select the model inputs for building characteristics and operations. Then based on the model inputs, the energy results are generated by using building energy simulation programs like EnergyPlus [19]. After that, depending on different building types, locations, and vintages, a rich set of building energy models can be created with different model input settings and the models generate different energy results. Finally, the simulation data sources collect the main model inputs and energy results. Nowadays, in the U.S., there are various existing simulation data sources and they are used for different objectives. For instance, Huang and Franconi [20] created 120 commercial building prototypes to quantify the contributions of components to heating and cooling loads in U.S. commercial buildings. Also, Griffith, et al. [21] created 4,820 building models to reflect the energy performance of commercial building samples in the 2003 CBECS. Based on these models, researchers assessed the technical potential for achieving net zero-energy (NZE) commercial buildings and energy impacts of various energy efficiency measures [3, 22]. To use a small set of building energy models to represent the majority of U.S. commercial buildings, the Department of Energy (DOE) created Commercial Reference Building Models by using EnergyPlus and aimed to evaluate the impact of various energy efficiency measures for commercial buildings [23]. The building models include three categories based on the vintages – pre-1980, post-1980, and new construction. After that, the DOE used those models to quantify the potential energy from the U.S. commercial buildings by improving the energy codes like American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 90.1 [8, 24-26] and International Energy Conservation Code (IECC) [27-30]. As a result, DOE has been continuously updated the building models based on different editions of ASHRAE Standard 90.1 and IECC. The updated building models are then called Commercial Prototype Building Models [31].

Recently, OpenStudio, an interface of EnergyPlus became popular [32, 33]. Thus, the National Renewable Energy Laboratory (NREL) developed a library called OpenStudio-Standards gem, which includes OpenStudio's version of Commercial Reference Building Models and Commercial Prototype Building Models [34, 35]. To conduct large scale analysis, NREL developed DEnCity to store a rich set of existing OpenStudio building models for future research [36-38].

Both survey and simulation data sources were used in various applications for different targeted users. First, the data sources can provide benchmarks for energy consumption in U.S. commercial buildings. For example, Energy Star Portfolio Manager and ASHRAE bEQ used the CBECS data as the foundation for its rating system targets [39, 40], and the LEED Rating System developed a rating system based on building models [41]. Also, the data sources are usually used for forecasting and predicting building energy consumption. The NZE potential study done by NREL provided a case study to analyze the energy saving potentials in the U.S. commercial building sector by using over 4,000 building models [3]. Furthermore, product developers are interested in energy impacts of building components like lights and systems. Thus, it is necessary to decompose impacts of components to building energy consumption. Huang and Franconi [20] analyzed the impacts of components to the heating and cooling loads in the U.S. commercial buildings. Both survey and simulation data sources can serve for the developments of building energy policies and standards. For instance, ASHRAE Standard 100 created EUI targets based on the energy data in CBECS [42, 43]. Thornton, et al. [44] used the Commercial Prototype Building Models to validate whether the new ASHRAE Standard can achieve the energy and cost saving goals by comparing to the old standard.

Recently, it became popular to use simulation data to conduct urban scale building energy analyses. Based on the geographic information system (GIS) and building energy modeling, it is able to create urban scale energy models. For example, Hong, et al. [45], Macumber, et al. [46], and Reinhart and Davila [47] created urban scale energy models and conducted analyses based on the models.

This paper is to summarize the existing data sources for energy-related data of U.S. commercial buildings and provide a guideline to select right data sources for specific applications. First, the paper summarizes the survey and simulation data sources for energy usage of U.S. commercial buildings and compares the data sources in terms of their data collection methods, released information, and features. Then this review analyzes the applications for different survey and simulation data sources. Based on the different features, this paper categorizes the applications of data sources into five categories, including energy performance benchmarks, energy usage forecasts and predictions, energy use contributions of building components, supports of energy policies and standards, and urban-scale energy use analysis, along with several cases to show how to use these data sources. Further, we introduce the Key Performance Indicators (KPIs) and guide users to select the databases for specific applications.

2. Existing Data Sources

This section introduces the main data sources in the U.S., including three survey data sources and six simulation data sources. It compares the features of each data source. The introduction and comparison will provide a broad picture of the data sources and preferences to select the data sources for different objectives.

2.1. Survey Data

Based on the coverage area, survey data sources of energy consumption in commercial buildings can be classified into local and national sources. Based on the number of samples and information recorded from each sample, survey data sources can also be divided into in-depth and large-scale sources. The in-depth sources can provide detailed building characteristics and energy use data, like building geometry, building schedules, and end-use energy consumption, for each building sample. The large-scale sources usually only provide several key building characteristics and energy data for each building samples, but consist of rich sets of building samples. This section introduces three typical survey data sources. CEUS is one of the representative state's energy survey data sources for U.S. commercial buildings which can be considered as local and in-depth survey data sources [12]. CBECS is a nation's comprehensive survey belonging to national and in-depth survey data sources [13]. The last one, BPD, is one of the largest survey data sources for energy consumption in U.S. buildings, and it is a national and large-scale data source [14].

The three survey data sources were developed for different purposes. CEUS primarily aims to support the California Energy Commission's (CEC) energy demand forecasting activities. To gain the data for hourly end uses, CEUS involved building energy program to conduct post-procession for the survey data [12, 48, 49]. CBECS data is a national-level sample survey of commercial buildings [13]. Its target is to display the distribution of the energy performance of commercial building sector and, thus, it provides statistical design for sampling. The goal of BPD is not to achieve a representative national sample like CBECS. Instead, BPD is developed as a national-level decision-support platform. The BPD's purpose is to generate plenty of building samples, which makes it valid to conduct research at the local level [10]. BPD can be used to assess building energy efficiency, forecast building energy performance, and quantify uncertainty of building energy consumption. Due to different purposes, the three survey data sources provide different procedures to collect, process, and generate data. Figure 1 summarizes the procedures of developing the three survey data sources.

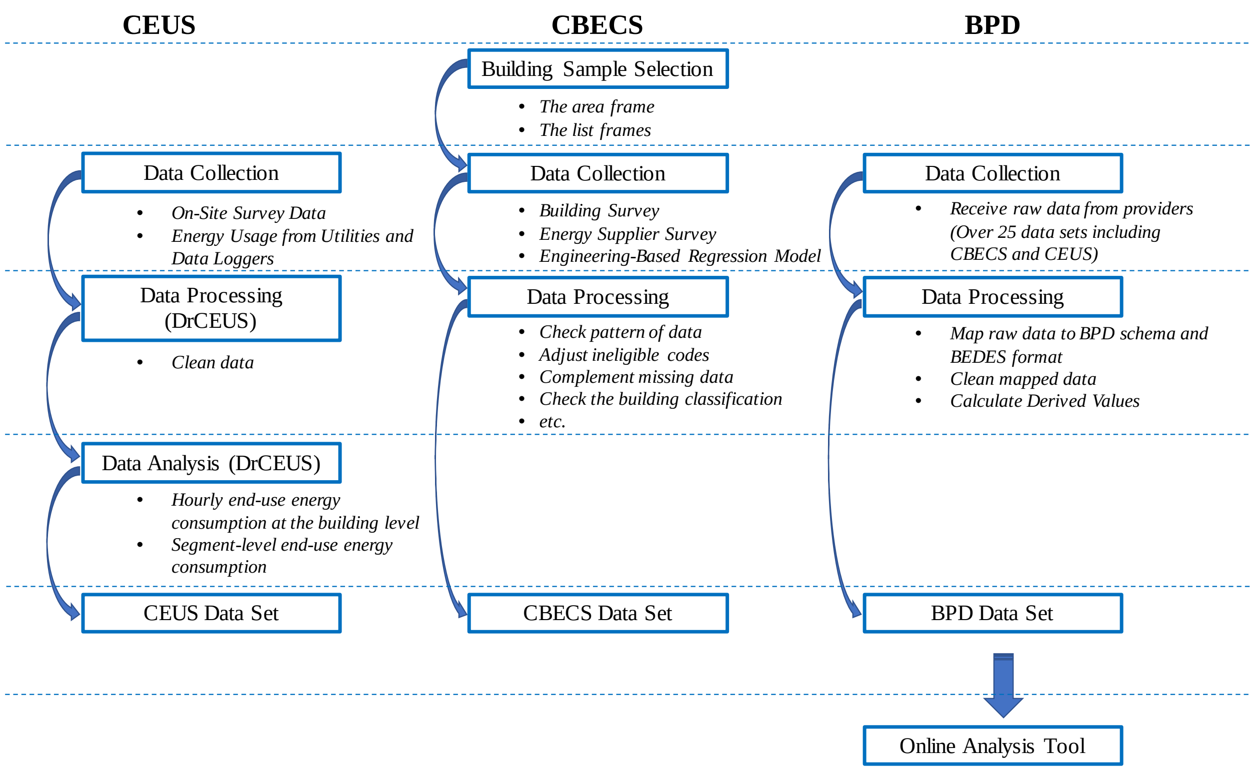

Figure 1. Procedures of developing the data sets

CEUS conducted a comprehensive on-site survey and collected the information of energy usage of samples from five utility companies [12]. Then DrCEUS (a simulation program developed based on SitePro and eQUEST) cleaned data and generated hourly end-use energy [50-52]. Finally, the CEUS data set was formed and developed based on the survey data and simulation end-use energy data.

As a national commercial building energy survey, the scope of CBECS is significantly larger than CEUS, which is focused on California. To reflect the change of the building energy performance over time, EIA updates the CBECS data continuously since 1979. The analysis of this paper focuses on the two latest and most used versions are the 2003 and the 2012 CBECS [53, 54]. The 2003 and 2012 CBECS adopted an area frame portion and an list frames portion to select building samples [55]. The area frame portion ensures that the distribution of building samples is representative of all the U.S. commercial buildings. The list frames portion optimizes the combination of large and small buildings to reflect their weighted impacts on the nation's energy consumption. Next, CBECS collected data of selected building samples from respondents and energy suppliers. In addition, regression models were used if the data could not be obtained [56-58]. After that, CBECS processed data to fix errors and fill missing data.

Another national survey data source, BPD, shows a different way to collect data. Instead of collecting data by on-site survey and from energy suppliers, BPD mapped over 25 source data sets including CBECS and CEUS, and collected data from these data sets. Then BPD adjusted the data formats in different data sets into the format following the criteria of the Building Energy Data Exchange Specification (BEDES). After that, BPD cleaned the data to fix errors. Finally, BPD derived its own data and formed the database [59]. Moreover, based on the data set, BPD created an online analysis tool which includes peer group analysis and performance comparison [60]. The online analysis tool makes it convenient for users to analyze the data and make decisions.

Besides the various data collection and procession methods adopted by the three survey data sources, they also publish different information about building energy consumption. Table 1 summarizes the available information provided by the three survey data sources. The data of individual building sample in CEUS is not available to the public. Instead, CEUS provides the analysis results of end-use energy consumption for the entire commercial building sector in California. Meanwhile, DrCEUS supports a variety of commercial end-use energy analyses [12]. On the contrary, CBECS provides the building characteristics and energy data of individual building sample. Users can estimate locations of building samples based on the climate zones and census divisions [13]. Although BPD does not provide the details of each building sample either, users can easily analyze and compare the distribution of building energy-related features like EUIs and total floor areas at the local levels via BPD's user interface [60].

Table 1. Information provided by the three survey data sources

	CEUS	**CBECS**	**BPD**
Year Published	2006	1979, 1983, 1986, 1989, 1992, 1995, 1999, 2003, 2012	Started from 2014
Number of Building Types	12	Principal Building Activity: 20 More Specific Building Activity: 51 *	Commercial Building Classification: 26 More Detailed Commercial Building Classification: 83
Location	California (Divided by climate zone)	U.S. (Divided by the CBECS climate zone and census division) *	U.S. (Divided by the ASHRAE climate zone, state, city, and zip code)
Number of Samples	2,800	Over 5,200 (2003 CBECS) Over 6,700 (2012 CBECS)	Over 75,000
Available Information	End-use energy consumption	Each sample building's characteristics, yearly	Distribution and comparison of yearly energy usages, building types, and ratings

	CEUS	CBECS	BPD
		expenditures, and end-use energy usages *	

* The information is only for the 2003 and 2012 CBECS.

Since survey data sources are created for specific purposes, they have their own features and are suitable for different applications. To help readers select the right data source which suits their application needs, this paragraph discusses and compares the features of the three typical survey data sources. Table 2 lists the comparisons of the three survey data sources' features. First, CEUS uses building energy simulation to complement the data collected by surveys. Thus, it is able to characterize the changes of energy end uses caused by occupants. It can also provide hourly data via DrCEUS. However, it is a state-level data source and lacks publicly available data, which limits the usefulness of the study [21]. Second, CBECS is a comprehensive energy related data source for U.S. commercial buildings that provides the details of building characteristics and energy data for each sample. Thanks to the adoption of the area frame and list frames portions to select building samples, CBECS is considered to be the best data source to reflect the distribution of the energy performance of the U.S. commercial building sector. However, CBECS cannot provide the hourly data. Due to the high cost of data collection, CBECS has a relatively small sample size and the data updates infrequently [61]. The sample set maybe insufficient when the analysis only focuses on a specific building type at a specific location. At the last, BPD is the largest energy related data source for U.S. commercial buildings. Its online analysis tool is a decision-support platform which provides probabilistic risk analysis [60]. However, it does not provide detailed information of each building sample [10]. Moreover, building samples' energy-related data is collected in different periods and one building in different periods could be recorded as different building samples.

Table 2. Comparisons of the three survey data sources' features

Feature	CEUS	CBECS	BPD
Use the building energy programs to post-process the survey data	Yes	No	No
Provide the hourly end-use energy consumption	Yes	No	No
Reflect the distribution of the energy consumption of commercial building sector	Only California	U.S.	U.S.
Collect energy consumption of given periods	Yes	Yes	No
Provide the characteristics for each building sample	Limited	Detailed	Limited
Provide enough data to conduct large scale analysis or analyze energy performance at the local levels	Only California	Limited locations	Many locations
Provide probabilistic risk analysis and decision-support platforms	No	No	Yes

2.2. Simulation Data

Thanks to the fast computing speed, low cost, and easy modification, building energy simulation becomes more and more popular and provides rich energy-related data sets for broad applications. There are various building energy simulation programs, like DOE-2 and EnergyPlus, with different features for model inputs, calculation methods, and model outputs [19, 62, 63]. Several literatures summarized and compared the major building energy simulation programs [9, 33, 64-66]. This section introduces six commonly used simulation data sources generated by building energy simulation programs as shown in Figure 2.

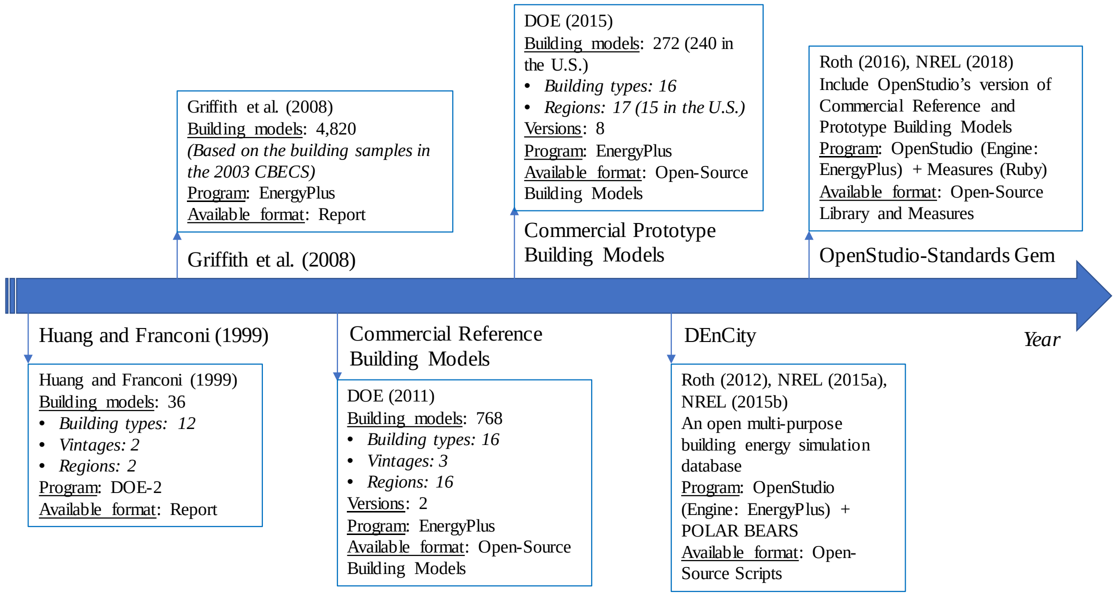

Figure 2. Summary of the six simulation data sources

First, Huang and Franconi [20] developed 36 commercial building models by using DOE-2 based on the previous researches [62, 63, 67-69]. Derived from the 1992 CBECS [70], the U.S. commercial building population were estimated by 36 building models which cover 12 main building types, two vintages, and two regions (six building types was only studied for one region). The building models were used to analyze the contributions of building components to heating and cooling loads in U.S. commercial building stocks and calculate the efficiencies of typical commercial heating and cooling systems to meet the loads.

Second, Griffith, et al. [21] created a set of building energy models to simulate existing commercial building sector. Based on the building samples in the 2003 CBECS, 4,820 building models were developed by using EnergyPlus [19]. The authors also introduced the procedure of creating the building models based on the limited survey data. By changing the model inputs of these building models, over 100,000 simulations were performed to analyze different scenarios of energy consumption in U.S. commercial buildings. For example, Griffith, et al. [3] used the simulation results to analyze the NZE potential of the U.S. commercial building sector. Benne, et al. [22] assessed the energy impacts of outside air in the commercial building sector.

In order to represent the U.S. commercial building stock with a small number of typical buildings, DOE created the Commercial Reference Building Models by using EnergyPlus [23, 71]. The inputs for the Commercial Reference Building Models came from several sources, such as ASHRAE Standard 90.1, ASHARE Standard 62.1, and CBECS data. There were 768 building models consisting of 16 building types, 3 vintages, and 16 locations, which represented nearly 70% of the U.S. commercial building floor area. The technical report (Deru et al. 2011) also discussed the methodology to create those building models and description of each model. The building models are open-source available at https://www.energy.gov/eere/buildings/commercial-reference-buildings.

Based on the Commercial Reference Building Models, the Commercial Prototype Building Models were developed by using EnergyPlus, which represented around 80% of the U.S. commercial building floor area and over 70% of the energy consumed in U.S. commercial buildings [31, 44]. The building models consist of 16 building types and 17 ASHRAE climate zones. The building models aimed to simulate the commercial buildings that meet the requirements of ASHRAE Standard 90.1 and IECC. So far, the Commercial

Prototype Building Models have eight versions to match the different versions of the ASHRAE Standard 90.1 and IECC. The building models are open-source available at https://www.energycodes.gov/development/commercial/prototype_models.

With the increased number of users for OpenStudio, which is a collection of software tools to support building energy modeling with EnergyPlus [32], there is an increasing need to enable the Commercial Reference Building Models and the Commercial Prototype Building Models in OpenStudio. However, the input file of EnergyPlus (.idf) cannot be directly transferred into the input file of OpenStudio (.osm), which limits users to use the Commercial Reference Building Models or the Commercial Prototype Building Models. Thus, NREL developed OpenStudio-Standards gem, a library created using the Ruby programming language [34, 35, 72]. By using the OpenStudio-Standards gem, users can automatically generate the OpenStudio versions of the Commercial Reference Building Models and the Commercial Prototype Building Models. In addition, NREL created a rich set of measures by using Ruby to enable the easy modification of model inputs [73]. The OpenStudio-Standards gem library is open-source available at https://github.com/NREL/openstudio-standards and measures are available at https://bcl.nrel.gov/.

Recently, DOE created an open-source multi-purpose building energy simulation database named as DEnCity [36-38]. The DEnCity database collects OpenStudio simulation inputs and results uploaded by users. Then using the large-scale simulation data stored in the database, DEnCity can provide users a quick analysis of building energy consumption for their simulation inputs without running the building energy simulation. The scripts of DEnCity are open source available at https://github.com/NREL/dencity-web and https://github.com/NREL/dencity-gem.

Each simulation data source has its own features which are suitable for specific applications. To guide users to select proper data sources for their applications, Table 3 compares three major features of six simulation data sources, including model availability, building status, and extendibility. In terms of model availability, Huang and Franconi [20] and Griffith, et al. [21] do not provide their models for individual buildings. Thus, users can not repeat or extend their work. However, users can still utilize their conclusions on the contributors to building heating and cooling loads, and their methodology to create building models based on the insufficient information from the survey data. On the other side, users can obtain all the inputs and outputs of the individual building models in the Commercial Reference Building Models, Commercial Prototype Building Models, and OpenStudio-Standards gem [23, 31, 35]. Thus, users can change the model inputs and generate new building models based on their specific requirements.

For building status, some sources cover only existing or new buildings and the others can be applied for both existing and new buildings. The building energy models for existing commercial buildings are usually created based on the survey and measured data while the models for new buildings are developed to meet the building energy standards. Based on the CBECS data, the work of Huang and Franconi [20] and Griffith, et al. [21] were created to estimate the energy consumption of existing U.S. commercial building stocks. The Commercial Prototype Building Models are used to evaluate the energy performance of the new commercial buildings complying with the ASHRAE Standard 90.1 and IECC. In addition, Commercial Reference Building Models aims to analyze the energy performance of both existing and new commercial buildings. Since OpenStudio-Standards gem consists of Commercial Prototype Building Models and Commercial Reference Building Models, it provides building models for both existing and new commercial buildings. DEnCity collects simulation inputs and results uploaded by users for both existing and new commercial buildings.

The extendibility indicates the capacity of extending the existing data source for analysis outside the existing application domains. Since Huang and Franconi [20] and Griffith, et al. [21] do not provide the building models for user to repeat or extend their work, new users cannot conduct new simulations based on the existing building energy models. Thus, the two simulation data sources are not suitable to conduct analysis outside of their existing application domains. Because Commercial Reference Building Models, Commercial Prototype Building Models, and OpenStudio-Standards gem are open sources for building models, users can conduct analysis by changing model inputs and generating new simulations for different applications. Additionally, DEnCity is a large-scale database with a rich set of energy-related simulation data [36-38]. Based on the large-scale pre-simulated results, users can analyze building energy consumption for their similar applications quickly. However, the range of the suitable applications highly depend on the availability of the data in DEnCity and it is difficult to perform an analysis if there is lack of pre-simulated results.

Table 3. Comparisons of the six simulation data sources' features

Feature	**Huang and Franconi [20]**	**Griffith, et al. [21]**	**Commercial Reference Building Models**	**Commercial Prototype Building Models**	**OpenStudio-Standards Gem**	**DEnCity**
Model Availability	No	No	Yes	Yes	Yes	No
Building Status	Existing	Existing	Existing and New	New	Existing and New	Existing and New
Extendibility	No	No	Yes	Yes	Yes	Limited

2.3. Comparison between Survey Data and Simulation Data

There are discrepancies in energy data between survey and simulation data sources [74]. For example, Turner and Frankel [18] found that simulated energy use deviated from actual energy use by 25% or more in many buildings. Another example is the study of Griffith, et al. [21]. Although they provided great efforts to create the building energy models to simulate the energy consumption of building samples in the 2003 CBECS, the total site EUIs from the simulation and survey still have huge gaps in some building types. For instance, there is a nearly 40% average relative error for the site EUIs of food services between the simulation results and the 2003 CBECS data. There are four attributes to the differences between the survey data and simulation data: 1) difficulties to account for occupant behavior [75], 2) interactive effects between systems [76], 3) uncertainty in model inputs [77], and 4) inefficiencies in actual buildings [78, 79].

Despite their differences, the survey data and simulation data have their own advantages and disadvantages. On one hand, the survey data has shortages in aspects where the simulation data is excellent. First, the survey data does not contain details like system efficiency, insulation information, operating schedules, and hourly energy consumption while the simulation data sources usually include the information [19, 20, 63]. Also, it is difficult to predict the broad impacts or energy saving potentials of individual components (e.g. a new HVAC equipment or window) by using survey data, which can be easily done by using simulation data [3, 4, 20, 77]. Moreover, it is time consuming and costly to conduct the building survey, and it is difficult to perform the survey and update the survey data frequently. However, users can easily generate sample data under different situations by running simulations.

On the other hand, the survey data also has advantages over the simulation data and it has been used to improve the simulation data. The survey data provides the references to determine the model inputs, which ensures that the simulation data can reflect the actual building features [80-82]. Also, the energy data recorded in the surveys shows the actual energy consumption of existing buildings and are widely used to calibrate and validate the simulation data [83-85].

3. Application of Data Sources

Based on the previous discussion and comparison of survey and simulation data, this section summarizes the applications for the three survey data sources and six simulation data sources, and provides examples of the applications.

3.1. Applications of Survey Data

Energy-related survey data provides a rich set of useful information for different users [13]. This paper classifies the applications into four categories: 1) energy performance benchmarks, 2) energy usage forecasts and predictions, 3) recognition of building energy contributors, and 4) developments of energy policies and standards. Table 4 summarizes the recommended data sources for different applications and provides some cases to demonstrate the usages of the data sources.

First, building *energy performance benchmark* is to evaluate the energy performance of a single building by comparing with other similar buildings [86]. The survey data sources provide the representative databases of building samples for building energy performance benchmarking. CEUS and CBECS have a large number of representative sample data. Thus, they are suitable to create energy performance benchmarks for U.S. commercial buildings. For example, based on CEUS, the Lawrence Berkeley National Laboratory (LBNL) developed Cal-Arch, which is a California-based distributional benchmarking model [87, 88]. Moreover, Mathew, et al. [49] created a benchmarking tool that enabled users to perform the end-use and component-level benchmarking by using CEUS database. On the other hand, based on CBECS, the U.S. Environmental Protection Agency (EPA) developed Energy Star National Energy Performance Rating System, which is a national regression-based benchmarking model [39]. Also, Yalcintas and Aytun Ozturk [89] developed benchmarks by using Artificial Neural Network (ANN) method. Moreover, Sharp (1996, 1998) developed benchmarks for energy uses in offices and schools. To evaluate the performance of benchmarks created by CEUS and CBECS, Matson and Piette [90] compared the results from the Energy Star National Energy Performance Rating System and Cal-Arch. The results from the two sources validated each other, and both benchmarking tools had the excellent performance. Brown, et al. [60] pointed out that the objective of BPD is “not to achieve a representative national sample”. Thus, it is better to create the building energy performance benchmarks by using CBECS and CEUS.

Second, building *energy usage forecasts and predictions* display the future trend of building energy consumption. The survey data sources provide the building characteristics and energy use of existing commercial building samples. By using the regression models, the future building energy consumption can be estimated. All the three survey data sources can be used to forecast the trend of the commercial building sector and predict energy retrofit savings. With thousands of existing building samples in California, CEUS is a suitable source to predict the trend of energy consumption for the California’s commercial building sector. For example, Porras-Amores and Dutton [91] assessed the energy and indoor air quality potentials in office buildings based on the CEUS data. By using the CBECS data, many analyses were conducted. EIA [92], EIA [1], and Kelso [93] provided the prediction of energy consumption in the commercial building sectors in the next decades by using the data from CBECS. Also, based on the CBECS data, Robinson, et al. [94] estimated commercial building energy consumption with machine learning. On the other hand, BPD can perform both national scale analysis and small-scale analysis for a giving type of building. For example, Walter and Sohn [95] predicted energy retrofit savings of commercial buildings based on the large-scale building energy data from BPD. Since BPD contains enough data to approximate the distribution for specific location and building type [60], it can be also used to predict energy retrofit savings for narrowly defined sets.

The third application category is *recognition of building energy contributors* to quantify their contributions to building energy consumption and energy saving potentials. To do so, the data sources should provide enough sampling data. It should also be able to adjust one or several building characteristics and keep the others as default values. Survey data usually have limited details of building characteristics and energy data. Therefore, it is difficult to decompose the contributions of components to energy consumption only based on the survey data. However, CEUS provides hourly end-use information based on the on-site survey and modeling. Thus, CEUS is the best choice among the three survey data sources to identify the contributions of components to energy consumption in commercial buildings. For instance, Ma, et al. [96] and Stadler, et al. [97] introduced the possibility to analyze the decomposition of components to end-uses serving for the demand response researches.

The fourth application category is *developments of energy policies and standards*. To make an informed decision on energy policies and standards, policymakers need to understand the building characteristics and energy performance of existing building stocks in a given period as well as predict the potential impact of the new policy and standards. Thus, the survey data serving for policies and standards needs to be representative and be recorded in given periods. For the California's commercial building energy-related policies and standards, CEUS is one of the best choices. For the national energy-related policies and standards, CBECS is one of the survey data sources that are usually selected. For instance, ASHRAE Standard 100 used the CBECS data as reference to create EUI targets [42, 43]. Usually, survey data is not used for urban-scale energy analyses directly because it is difficult to increase a large number of building samples in a short period. Thus, the urban-scale energy analyses are usually conducted based on simulation data, which is introduced in Section 3.2.

Table 4. Recommended survey data sources for various applications

	Energy performance benchmarks	**Energy usage forecasts and predictions**	**Energy use contributions of building components**	**Supports of energy policies and standards**
CEUS	X	X	X	X
CBECS	X	X		X
BPD		X		
Case	Kinney and Piette [87], Mathew, et al. [49], EPA [39], Yalcintas and Aytun Ozturk [89], Sharp [98], Sharp [99], Matson and Piette [90]	Porras-Amores and Dutton [91], EIA [92], EIA [1], Kelso [93], Robinson, et al. [94], Walter and Sohn [95]	Ma, et al. [96], Stadler, et al. [97]	Sharp [43], ASHRAE [42]

3.2. Applications of Simulation Data

This paper introduces the five main applications of simulation data: 1) energy performance benchmarks, 2) energy usage forecasts and predictions, 3) recognition of building energy contributors, 4) developments of energy policies and standards, and 5) urban scale modeling. Table 5 summarizes the recommended data sources for different applications and provides some cases to demonstrate their usages.

First, simulation data can be used as building *energy performance benchmarks* [86]. The advantage of the simulation-based benchmarks is that the energy performance evaluation of target buildings will not be affected by the different operating assumptions. There are many successful examples to create building energy performance benchmarks by using the simulation data. For example, to estimate the potential cost saving of energy efficient commercial properties, Deru, et al. [100] developed the baseline building models based on the Commercial Reference Building Models. Similarly, Commercial Prototype Building Models and OpenStudio-Standards gem are also capable of quantifying the potential energy savings of different energy efficient measures. For larger scale analysis, DEnCity can perform the similar analysis with higher granularity as it can contain simulation results under higher granularity of inputs. Roth, et al. [36] provided a case study that illustrates the satisfying performance for energy performance benchmarks. Due to lack of available building energy models and the limited available simulation data in the studies of Huang and Franconi [20] and Griffith, et al. [21], these two data sources are not recommended for building energy benchmarks. The operations and weather conditions are different between survey-based benchmarks and target buildings while the simulation-based benchmarks can use the same operations and weather conditions as the target buildings. Thus, by comparing with survey-based benchmarks, simulation-based benchmarks can conduct a pure evaluation of the building energy performance.

Second, simulation data sources can be used for *energy usage forecasts and predictions*. By changing baseline model inputs based on new devices or improved technologies, users can predict their energy saving potentials. Griffith, et al. [3], Griffith, et al. [21], and Benne, et al. [22] predicted the technical potentials for achieving NZE buildings in the commercial building sector and calculated the impact factors by simulating thousands of building models. Also, Glazer [4] researched on the maximum of technically achievable energy targets for commercial buildings based on the Commercial Prototype Building Models. Moreover, EnCompass (Energy Impact Illinois) identified retrofit opportunities in Chicago office buildings by running 278,000 EnergyPlus simulations based on the Commercial Reference Building [101]. For the U.S. commercial buildings, DEnCity can do the same work as EnCompare with the similar procedure [36]. Survey data is usually used to forecast the energy consumption trend of the whole commercial building sector while simulation data has the ability to predict the energy consumption trend and energy saving potential of a typical building or building type.

The third application is *recognition of building energy contributors*. Because most of the popular building energy simulation programs are able to change the building model inputs and generate the hourly energy end-uses, it is easy to decompose impacts of components to building loads and energy consumptions. Field, et al. [102] mentioned the possibilities to analyze the impact of single building component on building energy consumption. Based on the decompositions of building energy models, the building components having great energy impacts can be identified [20, 102-105]. For example, Huang and Franconi [20] provided a case to identify the building components that have great impacts on the heating and cooling loads in the U.S. commercial buildings. DEnCity can help users identify the building energy contributors quickly by alternating the model inputs and showing their impacts on energy consumption. Generally, the simulation data is a better option to recognize the building energy contributors by comparing to survey data.

The simulation data can be applied for the *developments of energy policies and standards*. For instance, Field, et al. [102] introduced that NREL used Commercial Reference Building Models to estimate the aggregate savings of ASHRAE Standard 189.1 compared to 90.1-2004 and 90.1-2007. Recently, the Commercial Reference Building Models were not updated. Thus, it is recommended to develop energy policies and standards by referring to energy-related data from Commercial Prototype Building Models, which are updated to meet the different versions of ASHRAE Standard 90.1 and IECC. For instance, Thornton, et al. [44] used the Commercial Prototype Building Models to validate whether the new

ASHRAE Standard can achieve the energy and cost saving goals by comparing to the old standard. Moreover, Roth, et al. [36] stated that policymakers can investigate the effects of proposed changes on building stocks by running thousands of simulations, which is an application of DEnCity. It is worth to mention that survey data is more appropriate to make policymakers understand the energy performance of existing building stocks while simulation data is more proper to evaluate the building energy saving potentials by using the new standards.

The fifth application is *urban scale modeling*, which is a popular topic recently. Based on the GIS, database, and urban-scale energy calculator, many researches were conducted [45-47]. The urban-scale energy calculators are usually developed based on EnergyPlus or OpenStudio. It is necessary to create some baseline building energy models based on the survey data or building energy standards. For the U.S. commercial building sector, Commercial Reference Building Models, Commercical Prototype Building Models, and OpenStudio-Standards gem are the appropriate candidates.

Table 5. Recommended simulation data sources for various applications

	Energy performance benchmarks	**Energy usage forecasts and predictions**	**Energy use contributions of building components**	**Supports of energy policies and standards**	**Urban scale modeling**
Huang and Franconi [20]			X		
Griffith, et al. [21]		X			
Commercial Reference Building Models	X	X	X		X
Commercial Prototype Building Models	X	X	X	X	X
OpenStudio-Standards Gem	X	X	X	X	X
DEnCity	X	X	X	X	
Case	Deru, et al. [100], Roth, et al. [36]	Griffith, et al. [3], Griffith, et al. [21], Benne, et al. [22], Illinois [101], Glazer [4], Hooper, et al. [106]	Huang and Franconi [20], Field, et al. [102]	Thornton, et al. [44]	Hong, et al. [45] Macumber, et al. [46] Reinhart and Davila [47]

3.3. KPIs of Survey and Simulation Data

To guide users to select survey or simulation databases, Table 6 lists the KPIs for the five applications. Each KPI has the same weight to evaluate whether the database is suitable for a specific application. Users will provide scores for individual KPIs. The high score means that the database meets the description of

KPI. Then users are recommended to select the database with the highest aggregated score, which is the most suitable to be used for the specific application.

Table 6. KPIs for the five applications

Application	KPI
Energy Performance Benchmarks	Be representative
	Not be affected by the operations
	Provide building characteristics and energy use
Energy Usage Forecasts and Predictions	Have large sample size/good extendibility
	Provide building characteristics and energy use
Energy Use Contributions of Building Components	Be able to control variables
	Provide building characteristics and energy use
Supports of Energy Policies and Standards	Be representative
	Record data in given periods
	Follow the energy policies and standards
	Keep updating
Urban Scale Modeling	Be representative
	Have large sample size/good extendibility
	Be able to control variables
	Provide building characteristics and energy use
	Provide the whole building energy models

Figure 3 shows the case to select the survey and simulation databases for the five applications. As we analyzed the applications in Section 3.1 and 3.2, we need to determine the advantages and disadvantages for each candidate database based on the KPIs and then score the KPIs for each database. The subfigures (a) and (c) respectively show the scores for individual KPIs for each survey database and simulation database. The subfigures (b) and (d) respectively show the aggregated scores with the same weight. The vertical dot dash lines are the thresholds of the recommended databases. The databases with the higher score are recommended. For example, the CEUS and CBECS are recommended for the energy performance benchmarks. It is worth noting that we adjust the aggregated scores for Huang and Franconi [20] and Griffith, et al. [21] and recommended these two databases because their reports respectively provide data for energy use contributions of building components, and energy usage forecasts and predictions. This section provides a general guideline. Based on the KPIs and guideline, users still need to provide their own judgements to select the suitable databases.

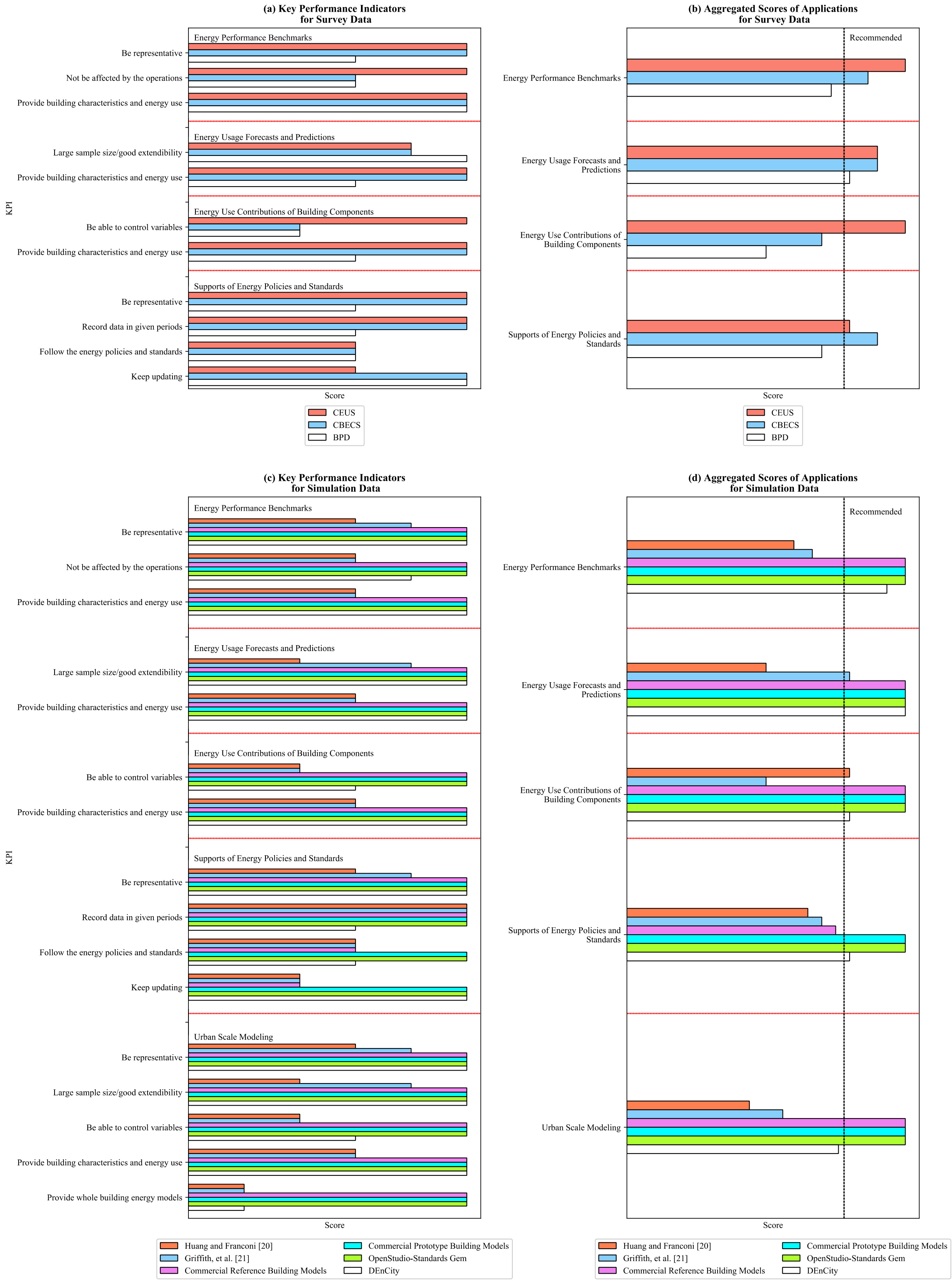

Figure 3. KPIs and aggregated score for survey and simulation data

3.4. Combined Applications of Survey and Simulation Data

Due to the complement of two data types, combined applications of survey and simulation data make the results more robust and accurate for special purposes [12, 21, 107, 108]. The building energy simulation can generate more samples based on the survey data. On the other hand, the survey data can be used to validate and calibrate the building energy simulation, and then improve the quality of the simulation results.

On one hand, the simulation data, fast and inexpensively generated by building energy simulation programs, complements the shortages of the survey data. First, the simulation data can normalize the weather conditions and building characteristics in survey databases, which can avoid the errors to evaluate energy performance caused by different conditions. For example, DrCEUS adjusted the on-site survey data in CEUS from actual historical weather data to normalized weather data [12, 48]. Second, the simulation data can generate more details of energy uses. For example, the CBECS data can only provide yearly energy data for the whole building. The simulation data can provide the details like hourly energy data for major components. Since CEUS combines the survey and modeling methods, hourly end-use energy consumptions can be identified [12]. Furthermore, the simulation data can consider various possible situations and broaden the sample sizes in survey data sources. Griffith, et al. [21] developed over 4,000 building energy models based on the 2003 CBECS. Then based on the models, Griffith, et al. [3] and Benne, et al. [22] changed the settings of the models to analyze the potential for achieving NZE buildings and assess the energy impacts of outside air. Finally, it is easy to conduct the uncertainty and sensitivity analyses by using simulation data [77, 109-111]. The energy data from surveys is affected by ambient conditions, building characteristics, and operating. Therefore, it is difficult for survey data to analyze the uncertainty of energy consumption due to the variances of some inputs and identify sensitive inputs to energy consumption. The simulation data can complement this shortage of the survey data. For instance, Eisenhower, et al. [77] provided a successful case to conduct uncertainty and sensitivity analyses, and decomposition of building energy models.

On the other hand, the survey data can also help to improve the quality of the simulation data. First, survey data provide some important model input data, although not all, for developing building energy models. For instance, several studies provided a methodology how to use the 2003 CBECS to model envelopes and system [80-82]. Griffith, et al. [21] provided a procedure to develop building energy models based on the 2003 CBECS data. Also, survey data can validate and calibrate the building models and improve the simulation results. As the discussion in Section 2, there are differences in energy data between survey and simulation data [18, 21]. Since survey data or measured data reflects the actual cases, they can be used to evaluate the quality of building energy models. Many researches provided methodologies to validate and calibrate the building energy models based on the survey data or measured data [83, 112-117]. After validation and calibration, the simulation data is more persuasive.

4. Conclusion

This paper conducts a critical review of energy-related data for U.S. commercial buildings along with their applications. Three typical survey data sources and six representative simulation data sources are summarized and compared.

Due to different objectives, the three survey data sources have different collecting methods and features. CEUS is a state-level in-depth survey data source, CBECS is a national in-depth source, and BPD is a national large-scale source. Thus, CEUS and CBECS provide detailed energy-related data. Even, by using

DrCEUS, the hourly end-use energy consumption can be obtained in CEUS. BPD has the largest sample size among the three survey data sources. Among simulation data sources, Huang and Franconi [20] and Griffith, et al. [21] do not provide their models for individual buildings, which limits users to repeat or extend their work. Commercial Reference Building Models, Commercial Prototype Building Models, and OpenStudio-Standards gem are open sources for their building energy models. Users can change model inputs and conduct new simulations. DEnCity, a multi-purpose building simulation database, provides users a quick analysis of building energy consumption without running the building energy simulation.

Both survey data and simulation data can be used for energy performance benchmarks, energy usage forecasts and predictions, recognition of building energy contributors, and developments of energy policies and standards. In addition, simulation data can be used to conduct urban scale modeling. Further, we provide the KPIs to guide users for selections of databases. Because of the advantages and disadvantages of the survey data and simulation data, the combined applications of both data types are also selected in some researches. By doing so, the two types of data can complement each other.

By using this paper, various users can use the paper for their own work. Building owners and managers can select data sources for benchmarking and use the existing cases for applications as references. Energy modelers can follow the research mentioned in this paper to forecast the energy consumption in U.S. commercial buildings. Product developers can use this guideline to gauge the market potential. Policy makers can update and validate the building energy standards by using data sources and the related applications mentioned by the paper as reference.

Acknowledgment

This paper is the outcome of the research project TRP-1771 sponsored by American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE).

Reference

[1] EIA, *Annual Energy Outlook 2018: With Projections to 2050*. Government Printing Office, 2018.

[2] P. Torcellini, S. Pless, M. Deru, and D. Crawley, "Zero energy buildings: a critical look at the definition," *National Renewable Energy Laboratory and Department of Energy, US,* 2006.

[3] B. Griffith, N. Long, P. Torcellini, R. Judkoff, D. Crawley, and J. Ryan, "Assessment of the technical potential for achieving net zero-energy buildings in the commercial sector," *National Renewable Energy Laboratory (NREL), Golden, CO.,* 2007.

[4] J. Glazer, "Development of Maximum Technically Achievable Energy Targets for Commercial Buildings," *ASHRAE,* 2016.

[5] D. H. Li, L. Yang, and J. C. Lam, "Zero energy buildings and sustainable development implications–A review," *Energy,* vol. 54, pp. 1-10, 2013.

[6] J. Kneifel, "Life-cycle carbon and cost analysis of energy efficiency measures in new commercial buildings," *Energy and Buildings,* vol. 42, no. 3, pp. 333-340, 2010.

[7] J. Kneifel, "Beyond the code: Energy, carbon, and cost savings using conventional technologies," *Energy and Buildings,* vol. 43, no. 4, pp. 951-959, 2011.

[8] ASHRAE, "Standard 90.1-2013, Energy standard for buildings except low rise residential buildings," 2013.

[9] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Building and environment,* vol. 43, no. 4, pp. 661-673, 2008.

[10] P. A. Mathew, L. N. Dunn, M. D. Sohn, A. Mercado, C. Custudio, and T. Walter, "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use," *Applied Energy*, vol. 140, pp. 85-93, 2015.
[11] Bonneville Power Administration. (2018). *End-Use Load and Consumer Assessment Program (ELCAP)*. Available: https://elcap.nwcouncil.org/
[12] CCEC, "California Commercial End-Use Survey," *Consultant Report. CEC-400-2006-005. Sacramento, California: California Energy Commission*, 2006.
[13] EIA. (2017). *Commercial Buildings Energy Consumption Survey (CBECS)*. Available: https://www.eia.gov/consumption/commercial/
[14] LBNL. (2014). *DOE Buildings Performance Database*. Available: http://energy.gov/eere/buildings/building-performance-database
[15] M. A. Berger, P. A. Mathew, and T. Walter, "Big data analytics in the building industry," *ASHRAE Journal*, vol. 58, no. LBNL-1005983, 2016.
[16] P. A. Mathew, "Advanced benchmarking for complex building types: laboratories as an exemplar," 2010.
[17] New Buildings Institute. (2008). *LEED case study database*. Available: https://newbuildings.org/
[18] C. Turner and M. Frankel, "Energy performance of LEED for new construction buildings," *New Buildings Institute*, vol. 4, pp. 1-42, 2008.
[19] DOE. (2017). *EnergyPlus*. Available: https://energyplus.net/
[20] J. Huang and E. Franconi, "Commercial heating and cooling loads component analysis," *Final Report. Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA*, 1999.
[21] B. Griffith, N. Long, P. Torcellini, R. Judkoff, D. Crawley, and J. Ryan, "Methodology for modeling building energy performance across the commercial sector," *National Renewable Energy Laboratory, Golden, CO*, 2008.
[22] K. Benne, B. Griffith, N. Long, P. Torcellini, D. Crawley, and T. Logee, "Assessment of the energy impacts of outside air in the commercial sector," *NREL/TP-550e41955. Golden, CO: National Renewable Energy Laboratory*, 2009.
[23] DOE. (2011). *Commercial Reference Building Models*. Available: https://energy.gov/eere/buildings/commercial-reference-buildings
[24] ASHRAE, "Standard 90.1-2004, Energy standard for buildings except low rise residential buildings," 2004.
[25] ASHRAE, "Standard 90.1-2007, Energy standard for buildings except low rise residential buildings," 2007.
[26] ASHRAE, "Standard 90.1-2010, Energy standard for buildings except low rise residential buildings," 2010.
[27] IECC, "International Energy Conservation Code 2006," 2006.
[28] IECC, "International Energy Conservation Code 2009," 2009.
[29] IECC, "International Energy Conservation Code 2012," 2012.
[30] IECC, "International Energy Conservation Code 2015," 2015.
[31] DOE. (2015). *Commercial Prototype Building Models*. Available: https://www.energycodes.gov/commercial-prototype-building-models
[32] R. Guglielmetti, D. Macumber, and N. Long, "OpenStudio: an open source integrated analysis platform," in *Proceedings of the 12th conference of international building performance simulation association*, 2011.
[33] L. Brackney, A. Parker, D. Macumber, and K. Benne, "Building Energy Modeling with OpenStudio," 2018.
[34] A. Roth. (2016). *New OpenStudio-Standards Gem Delivers One Two Punch*. Available: https://energy.gov/eere/buildings/articles/new-openstudio-standards-gem-delivers-one-two-punch
[35] NREL. (2018). *openstudio-standards gem*. Available: https://github.com/NREL/openstudio-standards

[36] A. Roth, M. Brook, E. T. Hale, B. L. Ball, K. Fleming, and N. Long, "DEnCity: An Open Multi-Purpose Building Energy Simulation Database," presented at the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, 2012.
[37] NREL. (2015). *DEnCity Gem*. Available: https://github.com/NREL/dencity-gem
[38] NREL. (2015). *DEnCity Rails Application*. Available: https://github.com/NREL/dencity-web
[39] EPA, "Energy Star Portfolio Manager," *Energy Star, Washington, DC*, vol. 2013, 2013.
[40] R. E. Jarnagin, "ASHRAE Building EQ," *ASHRAE Journal*, vol. 51, no. 12, p. 18, 2009.
[41] USGBC. (2014). *Leadership in Energy and Environmental Design (LEED) Rating System*. Available: www.usgbc.org/leed
[42] A. ASHRAE, "Standard 100-2006," *Energy Conservation in Existing Building*, 2006.
[43] T. R. Sharp, "Derivation of Building Energy Use Intensity Targets for ASHRAE Standard 100," Oak Ridge National Laboratory (ORNL), Oak Ridge, TN2014.
[44] B. Thornton *et al.*, "Achieving the 30% goal: Energy and cost savings analysis of ASHRAE Standard 90.1-2010," *Pacific Northwest National Laboratory (PNNL), Richland, WA (US)*, 2011.
[45] T. Hong, Y. Chen, S. H. Lee, and M. A. Piette, "CityBES: A web-based platform to support city-scale building energy efficiency," *Urban Computing*, 2016.
[46] D. Macumber *et al.*, "City Scale Modeling with OpenStudio," *Proceedings of SimBuild*, vol. 6, no. 1, 2016.
[47] C. F. Reinhart and C. C. Davila, "Urban building energy modeling–A review of a nascent field," *Building and Environment*, vol. 97, pp. 196-202, 2016.
[48] R. Ramirez, F. Sebold, T. Mayer, M. Ciminelli, and M. Abrishami, "A building simulation palooza: the California CEUS project and DrCEUS," *Proceedings of Building Simulation 2005*, 2005.
[49] P. Mathew, E. Mills, N. Bourassa, and M. Brook, "Action-oriented benchmarking: Using the CEUS database to benchmark commercial buildings in California," *Energy Engineering*, vol. 105, no. 5, pp. 6-18, 2008.
[50] T. Mayer, F. Sebold, A. Fields, R. Ramirez, B. Souza, and M. Ciminelli, "DrCEUS: Energy and demand usage from commercial on-site survey data," in *Proc. of the International Energy program Evaluation Conference*, 2003, pp. 19-22.
[51] DOE. (2016). *eQUEST: the QUick Energy Simulation Tool*. Available: http://www.doe2.com/equest/
[52] Itron. (2018). *Load Research Services*. Available: https://www.itron.com/sitecore/content/Sites/Itron/Home/Technology/product-services-catalog/Products/1/D/8/load-research-services
[53] EIA. (2006). *2003 Commercial Buildings Energy Consumption Survey*. Available: https://www.eia.gov/consumption/commercial/data/2003/
[54] EIA. (2017). *2012 Commercial Buildings Energy Consumption Survey*. Available: https://www.eia.gov/consumption/commercial/data/2012/
[55] EIA. (2012). *How Were Buildings Selected for the 2012 CBECS?* Available: https://www.eia.gov/consumption/commercial/reports/2012/methodology/sampling.php
[56] EIA. (2016). *How Was Energy Usage Information Collected in the 2012 CBECS?* Available: https://www.eia.gov/consumption/commercial/reports/2012/methodology/usage.php
[57] EIA. (2015). *How Was the 2012 CBECS Buildings Survey Conducted?* Available: https://www.eia.gov/consumption/commercial/reports/2012/methodology/conducted.php
[58] EIA. (2003). *2003 CBECS Survey Data - Methodology & Development*. Available: https://www.eia.gov/consumption/commercial/data/2003/index.php?view=methodology
[59] DOE. (2013). *Building Energy Data Exchange Specification (BEDES)*. Available: https://www.energy.gov/eere/buildings/building-energy-data-exchange-specification-bedes
[60] R. E. Brown, T. Walter, L. N. Dunn, C. Y. Custodio, P. A. Mathew, and L. Berkeley, "Getting real with energy data: Using the buildings performance database to support data-driven analyses and decision-making," in *Proceedings of the ACEEE summer study on energy efficiency in buildings*, 2014, pp. 11-49.

[61] National Research Council, *Effective tracking of building energy use: improving the commercial buildings and residential energy consumption surveys*. National Academies Press, 2012.

[62] R. B. Curtis *et al.*, "The DOE-2 building energy analysis program," *Lawrence Berkeley Laboratory Report# LBL-18046*, 1984.

[63] F. Winkelmann *et al.*, "DOE-2 supplement: version 2.1 E," Lawrence Berkeley Lab., CA (United States); Hirsch (James J.) and Associates, Camarillo, CA (United States)1993.

[64] D. B. Crawley *et al.*, "EnergyPlus: creating a new-generation building energy simulation program," *Energy and buildings*, vol. 33, no. 4, pp. 319-331, 2001.

[65] M. S. Al-Homoud, "Computer-aided building energy analysis techniques," *Building and Environment*, vol. 36, no. 4, pp. 421-433, 2001.

[66] H. Rpllapalli, "A comparison of EnergyPlus and eQUEST: whole building energy simulation results for a medium sized office building. 2010. 84 p," Thesis (Master Program)–Arizona State University, Arizona, 2010.

[67] J. Huang, H. Akbari, L. Rainer, and R. Ritschard, "481 prototypical commercial buildings for 20 urban market areas," 1991.

[68] A. Tuluca and J. Huang, "Market-Drive and Engineering Factors for Cogeneration in Commercial Buildings (LBL-36021)," *Lawrence Berkeley National Laboratory, Berkeley, CA.*, 1995.

[69] O. Sezgen, E. M. Franconi, J. G. Koomey, S. E. Greenberg, A. Afzal, and L. J. Shown, "Technology data characterizing space conditioning in commercial buildings: Application to end-use forecasting with COMMEND 4.0," 1995.

[70] EIA. (1995). *1992 Commercial Buildings Energy Consumption Survey*. Available: https://www.eia.gov/consumption/commercial/data/1992/

[71] M. Deru *et al.*, "US Department of Energy commercial reference building models of the national building stock," *National Renewable Energy Laboratory (NREL), Golden, CO.*, 2011.

[72] Ruby community. (2018). *Ruby: A Programmer's Best Friend*. Available: https://www.ruby-lang.org/en/

[73] NREL. (2017). *Building Component Library*. Available: https://bcl.nrel.gov/

[74] N. Li, Z. Yang, B. Becerik-Gerber, C. Tang, and N. Chen, "Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?," *Applied energy*, vol. 159, pp. 196-205, 2015.

[75] E. M. Ryan and T. F. Sanquist, "Validation of building energy modeling tools under idealized and realistic conditions," *Energy and Buildings*, vol. 47, pp. 375-382, 2012.

[76] S. Chidiac, E. Catania, E. Morofsky, and S. Foo, "Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings," *Energy*, vol. 36, no. 8, pp. 5037-5052, 2011.

[77] B. Eisenhower, Z. O'Neill, V. Fonoberov, and I. Mezić, "Uncertainty and sensitivity decomposition of building energy models," *Journal of Building Performance Simulation*, vol. 5, no. 3, pp. 171-184, 2012.

[78] Z. O'Neill, M. Shashanka, X. Pang, P. Bhattacharya, T. Bailey, and P. Haves, "Real time model-based energy diagnostics in buildings," *Proc. Building Simulation'11*, 2011.

[79] E. Mills, "Building commissioning: A golden opportunity for reducing energy costs and greenhouse-gas emissions," 2009.

[80] D. Winiarski, M. Halverson, and W. Jiang, "Analysis of building envelope construction in 2003 CBECS," ed: Richland, WA: PNNL Press, 2007.

[81] D. Winiarski, M. Halverson, and W. Jiang, "DOE's Commercial Building Benchmarks-Development of Typical Construction Practices for Building Envelope and Mechanical Systems from the 2003 CBECS," in *2008 ACEEE Summer Study on Energy Efficiency in Buildings*, 2008.

[82] D. Winiarski, W. Jiang, and M. Halverson, "Review of Pre-and Post-1980 Buildings in CBECS-HVAC Equipment," Pacific Northwest National Laboratory (PNNL), Richland, WA (US)2006.

[83] Z. O'Neill, B. Eisenhower, S. Yuan, T. Bailey, S. Narayanan, and V. Fonoberov, "Modeling and Calibration of Energy Models for a DoD Building," *ASHRAE Transactions*, vol. 117, no. 2, 2011.

[84] B. Eisenhower, Z. O'Neill, S. Narayanan, V. Fonoberov, and I. Mezić, "A methodology for meta-model based optimization in building energy models," *Energy and Buildings*, vol. 47, pp. 292-301, 2012.

[85] D. Coakley, P. Raftery, and M. Keane, "A review of methods to match building energy simulation models to measured data," *Renewable and sustainable energy reviews*, vol. 37, pp. 123-141, 2014.

[86] DOE. (2018). *Building Energy Use Benchmarking*. Available: https://www.energy.gov/eere/slsc/building-energy-use-benchmarking

[87] S. Kinney and M. A. Piette, "Development of a California commercial building benchmarking database," 2002.

[88] S. Kinney and M. A. Piette, "California Commercial Building Energy Benchmarking: Final Project Report," *Lawrence Berkeley National Laboratory, High Performance Commercial Building Systems Report*, no. E2P2, p. 1, 2003.

[89] M. Yalcintas and U. Aytun Ozturk, "An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database," *International Journal of Energy Research*, vol. 31, no. 4, pp. 412-421, 2007.

[90] N. E. Matson and M. A. Piette, "Review of California and national methods for energy performance benchmarking of commercial buildings," 2005.

[91] C. Porras-Amores and S. Dutton, "Assessing the energy and IAQ potential of dynamic minimum ventilation rate strategies in offices," in *Proceedings of the Symposium on Simulation for Architecture & Urban Design*, 2015, pp. 172-181: Society for Computer Simulation International.

[92] EIA, *Annual Energy Outlook 2017: With Projections to 2050*. Government Printing Office, 2017.

[93] J. D. Kelso, "Buildings energy data book," *US Dept. of Energy*, 2011.

[94] C. Robinson *et al.*, "Machine learning approaches for estimating commercial building energy consumption," *Applied Energy*, vol. 208, pp. 889-904, 2017.

[95] T. Walter and M. D. Sohn, "A regression-based approach to estimating retrofit savings using the Building Performance Database," *Applied energy*, vol. 179, pp. 996-1005, 2016.

[96] O. Ma *et al.*, "Demand Response for Ancillary Services," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1988-1995, 2013.

[97] M. Stadler *et al.*, "Electric storage in California's commercial buildings," *Applied Energy*, vol. 104, pp. 711-722, 2013.

[98] T. R. Sharp, "Benchmarking energy use in schools," in *Proceedings of the ACEEE 1998 Summer Study on Energy Efficiency in Buildings*, 1998: Citeseer.

[99] T. R. Sharp, "Energy benchmarking in commercial office buildings," in *Proceedings of the ACEEE*, 1996, vol. 7996, pp. 321-29.

[100] M. Deru, B. Griffith, M. Leach, E. Bonnema, and E. Hale, "179D Easy Calculator Technical Support Document," *National Renewable Energy Laboratory (NREL), Golden, CO.*, 2012.

[101] E. I. Illinois, "EnCompass," *Retrieved from: encompass. energyimpactillinois. org [accessed on 08.12. 14]*, 2013.

[102] K. Field, M. Deru, and D. Studer, "Using DOE commercial reference buildings for simulation studies," *Proceedings of SimBuild*, vol. 4, no. 1, pp. 85-93, 2010.

[103] N. Wang, S. Goel, and A. Makhmalbaf, "Commercial Building Energy Asset Score System: Program Overview and Technical Protocol (Version 1.1)," *Pacific Northwest National Laboratory (PNNL), Richland, WA*, 2013.

[104] N. Wang, S. Goel, A. Makhmalbaf, and N. Long, "Development of building energy asset rating using stock modelling in the USA," *Journal of Building Performance Simulation*, pp. 1-15, 2016.

[105] N. Wang, S. Goel, V. Srivastava, and A. Makhmalbaf, "Building Energy Asset Score Program Overview and Technical Protocol (Version 1.2)," *Pacific Northwest National Laboratory (PNNL), Richland, WA*, 2015.

[106] B. Hooper *et al.*, "The BayREN Integrated Commercial Retrofits (BRICR) Project: An Introduction and Preliminary Results," 2018.

[107] T. P. McDowell, S. Emmerich, J. W. Thornton, and G. Walton, "Integration of airflow and energy simulation using CONTAM and TRNSYS," *Transactions-american society of heating refrigerating and air conditioning engineers*, vol. 109, no. 2, pp. 757-770, 2003.

[108] D. Kosterev *et al.*, "Load modeling in power system studies: WECC progress update," in *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008, pp. 1-8: IEEE.

[109] W. Tian, "A review of sensitivity analysis methods in building energy analysis," *Renewable and Sustainable Energy Reviews*, vol. 20, pp. 411-419, 2013.

[110] J. Sanyal and J. New, "Building Simulation Modelers–Are we big data ready?," in *Proceedings of the ASHRAE/IBPSA-USA Building Simulation Conference*, 2014, pp. 449-456.

[111] J. Sanyal, J. New, R. Edwards, and L. Parker, "Calibrating building energy models using supercomputer trained machine learning agents," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 13, pp. 2122-2133, 2014.

[112] J. R. New, J. Sanyal, M. Bhandari, and S. Shrestha, "Autotune e+ building energy models," *IBPSA-USA Journal*, vol. 5, no. 1, pp. 270-278, 2012.

[113] Y. Heo, R. Choudhary, and G. Augenbroe, "Calibration of building energy models for retrofit analysis under uncertainty," *Energy and Buildings*, vol. 47, pp. 550-560, 2012.

[114] T. A. Reddy, "Literature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty, and Tools," *ASHRAE transactions*, vol. 112, no. 1, 2006.

[115] T. A. Reddy, I. Maor, and C. Panjapornpon, "Calibrating detailed building energy simulation programs with measured data—Part I: General methodology (RP-1051)," *Hvac&R Research*, vol. 13, no. 2, pp. 221-241, 2007.

[116] T. A. Reddy, I. Maor, and C. Panjapornpon, "Calibrating detailed building energy simulation programs with measured data—part II: application to three case study office buildings (RP-1051)," *Hvac&r Research*, vol. 13, no. 2, pp. 243-265, 2007.

[117] J. Sun and T. A. Reddy, "Calibration of building energy simulation programs using the analytic optimization approach (RP-1051)," *HVAC&R Research*, vol. 12, no. 1, pp. 177-196, 2006.