Building and Environment 157 (2019) 139–164

Contents lists available at ScienceDirect

Building and Environment

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

Understanding the adoption and usage of data analytics and simulation among building energy management professionals: A nationwide survey

Charu Srivastava, Zheng Yang, Rishee K. Jain*

Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Rm. 269A, Stanford, CA, 94305, USA

A R T I C L E I N F O

Keywords:
Building energy management
Energy efficiency
Survey
Decision-making
Data analytics
Simulation

A B S T R A C T

Improving the energy-efficiency of our building stock is critical to meeting our worldwide sustainability goals. In response to this need, two key tools have emerged to help engineers, building scientists and energy managers understand building energy usage and derive energy-efficiency solutions: data analytics and simulation. While both data analytics and simulation hold significant promise, we lack a clear understanding on the use, barriers and expectations of both in the building energy management decision-making process. This study conducts a nationwide survey of 448 building energy management professionals in the United States to help elucidate: 1) what impacts the adoption of data analytics and simulation among building energy management professionals; 2) in what phases of the building energy management decision-making process are data analytics and simulation currently used; and 3) what are the barriers of use for data analytics and simulation and how can they be improved to better support building energy management decision-making. Overall, results indicate that professional domain plays a large role in associating the uses, barriers and expectations for data analytics and simulation. Results also indicate that data analytics and simulation could be coupled to leverage functionality as they are used in similar phases of the decision-making process. Lastly, results point to opportunities for improving the applicability of data analytics and simulation tools as well as training for both. In the end, this study aims to provide a quantitative basis for improving the efficacy and integration of data analytics and simulation in the building energy management domain.

1. Introduction

Buildings consume 40% of the total U.S. energy use, more than any other sector, making enhancing the energy-efficiency of the built environment integral to our long-term sustainability goals. Engineers, building scientists and building energy managers have at their disposal two key tools to characterize building energy usage and derive energy-efficiency solutions: *data analytics* and *simulation*. *Data analytics* leverages building data that is becoming more and more widely available through "open data" and "smart city" initiatives to build analytical models that extract insights on the energy performance of a building. *Simulation* utilizes physics-based methods to model the complex thermodynamics of a building and enables effective and convenient assessment of how changes to building design and/or operations impact energy usage.

While both data analytics and simulation hold significant promise for the improvement of building energy management, widespread adoption and usage of both has been limited. Often key decision-makers in the building energy management domain lack the background and experience in data analytics and/or resources to create, interpret and translate results of complex simulations into actionable insights. As a result, it is of critical importance to gain a deeper and more comprehensive understanding of how building energy management professionals[1] utilize data analytics and simulation. We currently lack a comprehensive and in-depth analysis of the use, barriers and expectations of data analytics and simulation in the building energy management decision-making process and therefore are limited in our ability to design new tools that can achieve widespread adoption in practice.

This study aims to deepen our understanding on the use of data analytics and simulation in the building energy management domain through analysis of a nationwide survey of 448 building energy management professionals. Specifically, we aim to focus on: 1) what impacts the adoption of data analytics and simulation among building energy management professionals; 2) in what phases of the building energy management decision-making process are data analytics and simulation currently used; and 3) what are the barriers to use for data

* Corresponding author.
E-mail address: rishee.jain@stanford.edu (R.K. Jain).

[1] In this paper, we define building energy management professionals as anyone involved in the building energy management decision-making process.

https://doi.org/10.1016/j.buildenv.2019.04.016
Received 8 January 2019; Received in revised form 20 March 2019; Accepted 7 April 2019
Available online 13 April 2019

analytics and simulation and how can they be improved to better support building energy management decision-making. The rest of the paper is organized as follows: Section 2 reviews current literature and emphasizes the importance of the proposed survey and associated analysis. Section 3 introduces the survey design, administration and analysis procedure. Section 4 presents all the results of the survey analysis and discusses key findings. Section 5 concludes the paper, summarizing limitations and proposing future work.

2. Related work

Previous work has analyzed the usage and applications of both data analytics and simulation in the building energy management domain. In this paper, we define *simulation* as the process of virtually reproducing and representing building energy performance and behaviors over time and *data analytics* as the process for obtaining raw data from buildings and translating it into insights useful for building energy management decision-making. Simulation is currently used to model energy performance of buildings like lighting systems [1,2] HVAC (Heating, Cooling, and Air Conditioning) systems [3] and computational fluid dynamics [4]. Data analytics is mainly used to monitor and estimate building energy saving potentials through energy use pattern recognition [5], prediction and classification of energy consumption behaviors [6] and occupancy schedules [7], control optimization [8] and benchmarking for energy use efficiency [9]. Both simulation and data analytics are also applied to model energy use at the urban scale [10].

While both simulation and data analytics are utilized in the building energy management domain, previous research investigating their current usage and barriers to widespread adoption has been limited. A number of studies have used survey analyses to understand the usage and adoption trends of simulation in building energy management [11], the challenges that potentially block its application [12], user needs and function requirements for different use cases [13] and data exchange and interoperability [14]. Previous work has also sought to validate simulation methods and programs by testing and comparing their modeling results under variousconditions [15–18]. Other studies have focused on defining sustainable energy performance indicators in simulation [19], reviewing methods of modeling building energy systems [20] and using surveys to compare pros and cons of certain simulation programs [21,22]. However, many of the previous comparison studies have short-term implications due to the need for constantly improving simulation tools that meet the changing demands of building energy management [23]. Some studies highlighted the barriers to adopting simulation tools [24] and suggested requirements to enhance their future development [25]. Conversely, the application of data analytics to support building energy management is in its nascent stage of development and therefore little research has been undertaken to study current challenges of adoption and usage. One study proposed tools to better inform energy benchmarking models [26]. While another review paper discussed challenges and opportunities for increased adoption of data analytics in the design and optimization process [27], however, it drew conclusions primarily from interviews with a small number of optimization experts. Previous review work has focused on the summarization of technical advantages and trends of analytical techniques for building energy management but limited in its discussion as to how those techniques are being used by professionals [28–30]. This limited body of work further underscores the need for a deeper understanding of how data analytics is being utilized across multiple building energy management domains and what barriers are preventing more widespread adoption. As a result, the proposed study aims to focus on understanding the overarching usage trends, adoption barriers and potential areas of improvement for both data analytics and simulation in building energy management.

From a methodological perspective, previous works [31,32] have largely been interview based or case studies, which by design have a limited number of participants. Some studies divided participants into pre-defined professional groups (e.g. architect and engineer) [33,34], obscuring nuances between the many diverse professionals in the building energy management domain. Currently, the literature is lacking analysis of a diverse and representative sampling of building energy management professionals with varying positions, years of experiences and responsibilities. Also, while some studies explored general trends of information technologies (e.g. Building Information Modeling) [35–38], most of such studies only considered a single use of either data analytics or simulation. In many cases, data analytics and simulation are complementary in nature [39] and energy and environmental data is particularly useful for building energy model calibration [40,41]; therefore, their usage should be studied in tandem. Thus, this paper aims to contribute to current literature by studying the usage and barriers to adoption of data analytics and simulation in tandem through the use of an original, detailed survey of building energy management professionals. In doing so, we aim to identify areas of improvement for both approaches to maximize their adoption and impact on the area of building energy management.

3. Methodology

The methodology of this paper is divided into 3 sections: survey design, survey administration and survey analysis. The first two sections describe the iterative process by which the survey was developed and distributed to collect data from professionals in the building energy management domain. The last section describes how the computational bootstrap method was used to identify patterns and draw conclusions from the survey results.

3.1. Survey design

To gather data on professionals' opinions about current use, barriers and expectations of data analytics and simulation in building energy management, an online self–administered survey (Appendix A) was developed and distributed to building energy management professionals across the United States. Professionals were provided a link to the survey with a short description of the objective of the survey over email. The research team has gone through several iterations to test the validity of survey questions in order to minimize bias and misinterpretation of the survey questions.

The first part of the survey collects demographic information of participants, including organization, current position, years of experience, region, professional domain and decision-making team size. The quality of professionals' responses was controlled and examined to ensure the authenticity, diversity and representativeness of response data. The second part of the survey uses Likert scales and multiple-choice questions [42] to understand the status quo of data analytics for building energy management related decision-making. The collected information includes the level of use of data analytics in decision-making, the frequency of use of data analytics to support decision-making, the usefulness of data analytics in different phases of decision-making, the main barriers blocking the use of data analytics in decision-making and the expectations for further improvement. Similarly, the third part of the survey utilizes Likert scales and multiple-choice questions to understand the status quo of simulation for building energy management related decision-making. The collected information includes the level of use of simulation in decision-making, the frequency of use of simulation to support decision-making, the usefulness of simulation in different phases of decision-making, the main barriers blocking the use of simulation in decision-making and the expectations for further improvement. We also note that our survey does not specifically ask the types of buildings being managed by each participant. However, based on the organizational af- filiations of the participants we can infer that the majority of partici- pants manage commercial buildings and university buildings.

3.2. Survey administration

Energy professionals were randomly and conveniently selected from members of USGBC (U.S. Green Building Council), ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers), IFMA (International Facility Management Association), BCA (Building Commissioning Association) and AEE (Association of Energy Engineers). In total, 448 complete responses were received from two rounds (Round A and Round B) of survey distribution and response collection.

Round A used the original survey design with all 20 questions in 4 webpages for approximately 1/3 randomly selected building energy management professionals. The participants had to click three "next" buttons to finish all the questions. In this round, the complete response rate was only 5%. More than 50% of the participants who opened the survey quit after the second page, suggesting they may only stay focused for up to 2 pages of questions and lose interest after seeing the second "next" button. In general, we observed that the professionals lack incentives and patience for additional efforts (e.g., clicking several "next" buttons) to finish the survey.

To motivate the building energy management professionals to spend more time completing the survey, the survey distribution method was updated in Round B for the remaining 2/3 of professionals based on a previously developed survey design philosophy [43]. The 20 questions were divided into two parts and sent to participants in two stages. The first stage contained 11 questions in 1 page, asking for demographic information and opinions about the current practice and frequency of use for both data analytics and simulation to support building energy management decision-making. Some survey design features were also improved to help participants quickly and clearly finish the questions. For example, each line had 85 characters, using high contrast with no color (7%–8% males are color-blind) [44]. Helvetica font was used in order to keep consistency over different browsers. The website and logo of the research team were provided in case the participants had any questions or curiosity. After the first stage, we sent a follow-up email to those who completed all the questions to appreciate their contributions. We also requested them to complete the second stage containing the remaining 9 question in 1 page, including usefulness, main barriers and improvement criteria of both data analytics and simulation to support building energy management decision-making. Through this two-stage method, the research team successfully built more trust with participants and made them feel that their responses were indeed important to this research project. The building energy management professionals gained more incentives and interest for completing the survey. Some of them even helped distribute the survey to their networks and provided more contact information for potential participants. The response rate was improved to 18% in Round B.

3.3. Survey analysis

After collecting all responses, data cleaning was performed to remove responses with: 1) more than half questions unanswered, 2) no demographic information provided, 3) complete time less than 3 minutes, 4) poor verbatim "straightline" answers, and 5) obviously inconsistent answers. Afterwards, the survey analysis was organized to investigate three interconnected research questions: 1) what impacts the adoption of data analytics and simulation among building energy management professionals; 2) in what part of the building energy management decision-making process are data analytics and simulation currently used; and 3) what are the barriers to use for data analytics and simulation and how can they be improved to better support building energy management decision-making (Fig. 1). The first research question analyzes the impacts of demographics on the frequency and level of use for data analytics and simulation. The second research question investigates the impact of level of use for data analytics and simulation on their perceived importance during the different phases of building energy management decision-making. The third research question seeks to understand how different levels of use for data analytics and simulation impact the barriers to use and how the barriers to use affect the expectations for improvement (Fig. 2).

In order to answer the above three research questions, we aim to analyze the association between a categorical answer for one question (the dependent variable) with a set of answers for other questions (independent variables) using logistic regression models, as done in previous work which analyzed survey results related to decision-making in building energy-efficiency retrofits [45]. The dependent variable can either have two values (e.g., selected or not) or multiple unique values (e.g., Likert scale selection). For example, if 1 represents the option is selected (with the probability of p) and 0 represents the option is not selected (with the probability of $1-p$), the ratio $p/(1-p)$ is the odds and the logit is the logarithm of the odds (log odds). The logit transformation can be mathematically formulated as:

$$logit\,(p) = ln\left(\frac{p}{1-p}\right)$$

Logistic regression models the linear relationships between log odds of answers for the dependent variable and answers for the independent variables, as follows:

$$ln\left(\frac{p}{1-p}\right) = \alpha + \beta x + e$$

Where $\mathbf{x}$ is a vector of answers for the independent variables that could have potential impacts on the answer for the dependent variable. $\boldsymbol{\beta}$ is the vector of coefficients and the exponentials of $\boldsymbol{\beta}$ indicate the effects of the independent variables on the odds ratio of dependent variable. e is the error term. When there are more than two unique values for the dependent variable, one-vs-rest (OvR) is applied. Maximum Likelihood Estimation (MLE) [46] is applied to estimate the coefficients.

To probabilistically analyze the regression results (e.g. what happens if we acquire the data again), this paper employs bootstrapping with logistic regression. Bootstrapping allows random resampling from a sample with replacement. It requires fewer assumptions of the data distribution, residuals and variance, sample size, etc., making it ideal for smaller and skewed samples. Sampling with replacement treats the data as a proxy for the true population. Each resampled data is called a bootstrap sample and a bootstrap replicate is the value of the statistic computed from the bootstrap sample, which is a simulated replica of the original data acquired by bootstrapping. It has been mathematically demonstrated that the empirical bootstrap analysis is more reliable and consistent than theoretical analysis [47].

Using bootstrapping method for logistic regression, this paper resamples responses from the response pool, runs logistic regression as described above and computes the statistics of interest (e.g. coefficients and statistical significance), in order to generalize calculations. Hypothesis testing is done by clearly stating the null hypothesis (the coefficient is equal to 0), generating many sets of simulated data assuming the null hypothesis is true, computing coefficients for each simulated data set and analyzing the percentage of coefficients that are either below or above 0 in order to compute the statistical significance. Consistent with convention related to survey analysis, we utilized a significance threshold of 0.05 (sig. $p < 0.05$).

4. Results and analysis

The survey responses reveal a wide distribution of organizations, regions, current positions, years of experience and professional domains (Fig. 3 and Fig. 4). Since building energy management domains are essentially interconnected, one professional might work on different domains simultaneously and is allowed to select more than one domain that she/he thinks is relevant to her/his daily work. N stands for the percentage of a specific selection among all responses. It can be seen

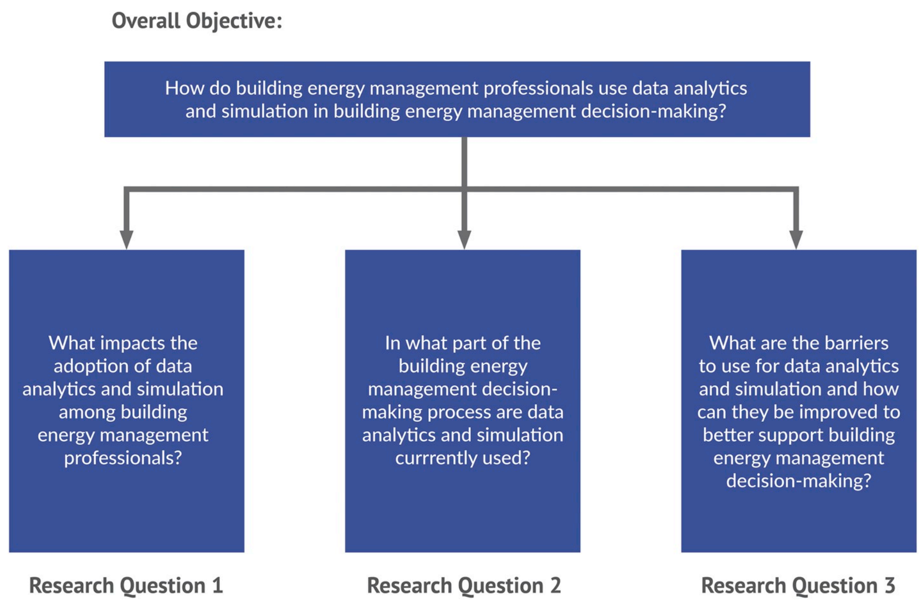

Fig. 1. The overall research objectives and research questions of this study.

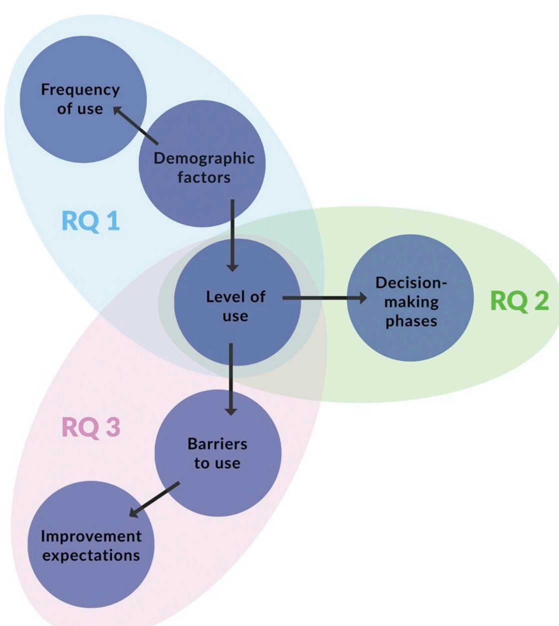

Fig. 2. The relationships between the three research questions investigated and the various components of the survey. The direction of the arrow indicates the relationship modeled by a logistic regression from the independent variable to the dependent variable.

that more participants come from *Private* organization (N = 71.21%) than *Public* organizations (N = 28.79%). The possible positions include *Engineers* (N = 49.33%), *Managers* (N = 26.34%), *Directors* (N = 17.41%), *Principals* (N = 0.89%) and *Presidents* (N = 6.03%). The years of experience range from 1 to 20 + years, grouped as follows: *1–5 years* (N = 12.50%), *5–10 years* (N = 24.77%), *10–20 years* (N = 25.45%) and *20 + years* (N = 37.28%). Participants come from all regions of the country, including *Northeast* (N = 29.69%), *West* (N = 28.35%), *Midwest* (N = 21.43%) and *South* (N = 20.54%). The professional domains cover fourteen building energy management related areas. Most participants work in the areas of *Building systems and equipment* (N = 94.41%), *Building operations and maintenance* (N = 81.70%) and *Building commissioning and energy auditing* (N = 79.46%), while fewer participants are involved in the areas of *Building occupancies* (N = 20.98%), *Grid optimization and demand response* (N = 21.88%) and *Community and neighborhood management* (N = 15.63%). From these statistics, it can be seen that the responses are diverse and encompass a wide range of organization types, current positions, years of experience, professional domains and regions.

4.1. RQ1: factors affecting the adoption of data analytics and simulation

In order to investigate what factors impact the adoption of data analytics and simulation in building energy management, we first compared the effect of various demographic factors, such as organization, years of experience, current position, region and professional domain on the frequency of use of both data analytics (Table 1) and simulation (Table 2). The exponentials of coefficients, which indicate the effects of the independent variable on the odds ratio of the dependent variable, were also calculated. A positive coefficient indicates an increase in the odds of selection for the dependent variable while a negative coefficient indicates a decrease in odds of selection for the dependent variable. Only coefficients with p-value < 0.05 were reported. Overall, the professional domain had the most significant impact on the adoption of data analytics and simulation.

The logistic regression tests reveal that amongst all positions, *Managers* have an increase of 68.2% increase in the odds to *Always* use data analytics (coef = 0.52, sig = 0.01). Participants working in the domain of *Building occupancy* are also likely (85.9% increase in odds) to *Always* use data analytics (coef = 0.62, sig = 0.02), while those working in *Building additions, alterations and retrofitting* (coef = 0.49, sig = 0.03, 63.2% increase in odds) and *Energy finance and market* (coef = 0.43, sig = 0.04, 53.7% increase in odds) are likely to *Sometimes* (*half of the time*) use data analytics. This could be because these domains have more data available for analysis and decision-making [48]. Participants in *Building operations and maintenance* are likely to *Never* adopt the use of data analytics (coef = 0.40, sig = 0.03, 49.2% increase in odds), as perhaps the development of building operations and maintenance is not very mature in the practice of collecting and analyzing data [49]. This presents an opportunity for increased adoption of data analytics as *Building operations and maintenance* could greatly benefit by using it to monitor energy usage during the life of the building, as has been done for many buildings after occupancy [50]. There is no significant positive impact of organization, years of experience or region on the frequency of use. No demographic factor significantly influences the odds of *Often* and

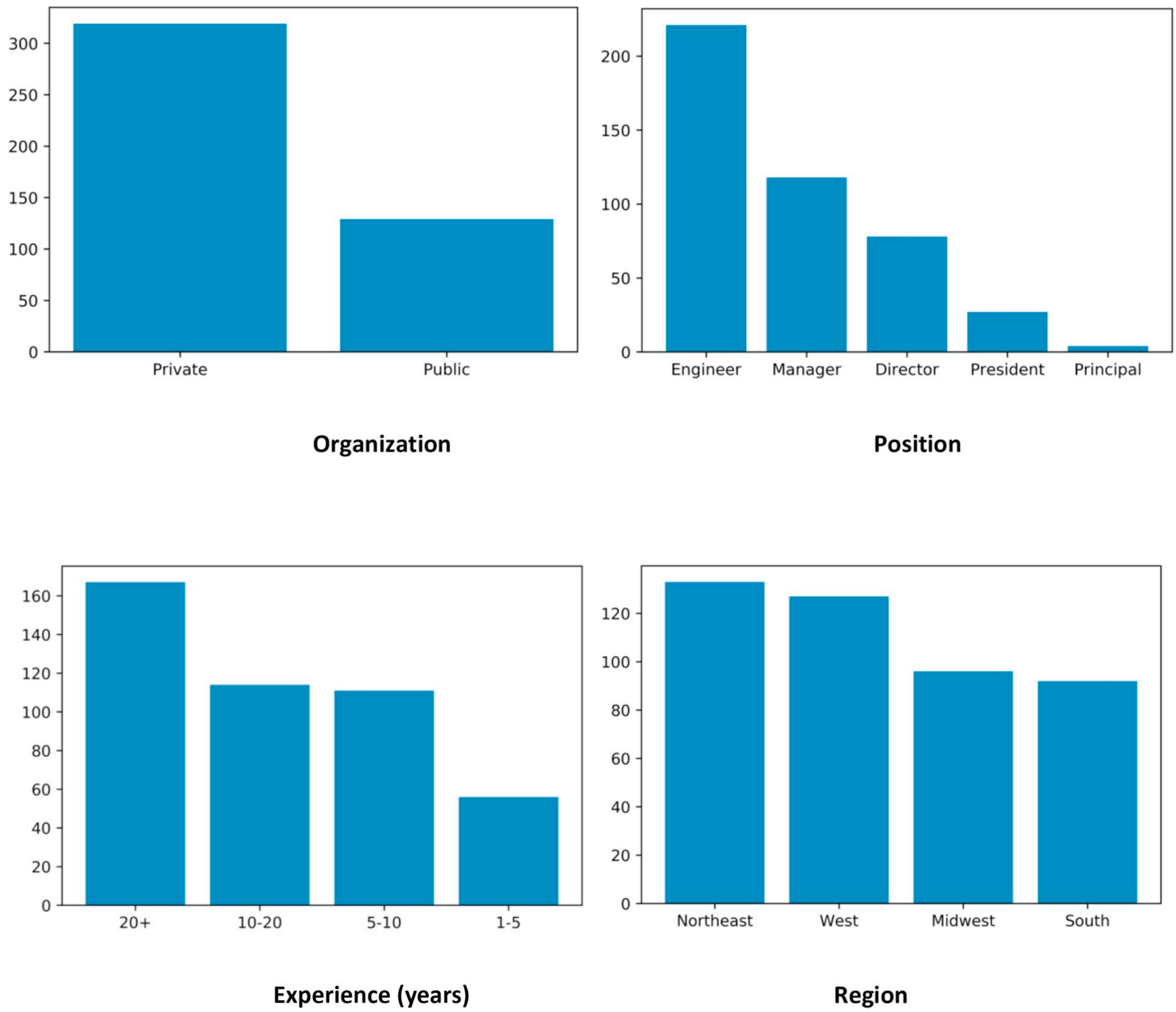

Fig. 3. Number of participants by organization, current position, years of experience and region.

Seldom use of data analytics.

Participants from several domains are likely to *Sometimes (half of the time)* use simulation: *Building occupancies* (coef = 0.65, sig = 0.01, 91.5% increase in odds), *Energy finance and market* (coef = 0.40, sig = 0.05, 49.2% increase in odds) *and Grid optimization and demand response* (coef = 0.62, sig = 0.01, 85.9% increase in odds). *Building systems and equipment* (coef = 0.55, sig = 0.02, 73.3% increase in odds) and *Building commissioning and auditing* (coef = 0.68, sig = 0.01, 97.4% increase in odds) are likely to *Never* or *Seldom* (coef = 0.57, sig = 0.02, 76.8% increase in odds) use simulation. Participants working in the *Building operations and* maintenance domain are likely to *Often (more than half of the time)* use simulation (coef = 0.63, sig = 0.01, 87.8% increase in odds). They are more likely to use simulation than data analytics (data analytics: *Never*). There is no significant impact on frequency of use by organization, current position, years of experience or region. The results show that no demographic factor significantly influences the odds of use *Always* regarding simulation.

Next, we aim to deepen our understanding of what impacts the adoption of data analytics and simulation in building energy management by analyzing their level of use. The different levels of use are classifications of technologies based on their different requirements of specialty, knowledge and tools for implementation in building energy

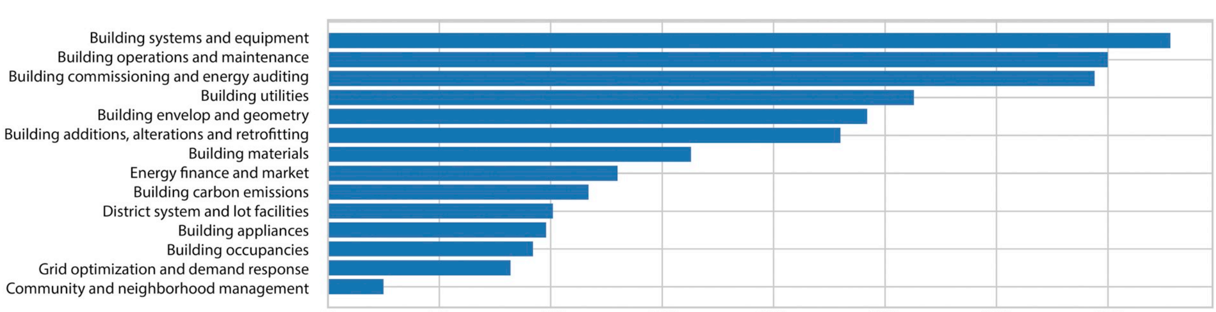

Fig. 4. Number of participants by professional domain.

Table 1
Coefficients of factors that may affect adoption of data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

Demographic	Frequency of use	Always	Often	Sometimes	Seldom	Never
Position	Manager	**0.52 (68.2%)**				
Professional Domain	Building operations and maintenance					**0.40 (49.2%)**
	Building additions, alterations, and retrofitting			**0.49 (63.2%)**		
	Building occupancies	**0.62 (85.9%)**				
	Energy finance and market			**0.43 (53.7%)**		

Table 2
Coefficients of factors that may affect adoption of simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

Demographic	Frequency of use	Always	Often	Sometimes	Seldom	Never
Professional Domain	Building systems and equipment				**0.57 (76.8%)**	**0.55 (73.3%)**
	Building operations and maintenance		**0.63 (87.8%)**			
	Building commissioning and energy auditing					**0.68 (97.4%)**
	Building occupancies			0.65 (91.5%)		
	Energy finance and market			0.40 (49.2%)		
	Grid optimization and demand response			0.62 (85.9%)		

management. From a descriptive analysis perspective, the bootstrap-calculated confidence intervals were equivalent to the equation-based confidence intervals. Generally, data analytics is often used for computationally light analysis. The most common use for data analytics is *Manual or spreadsheet oriented calculations* (N = 83.93%), followed by *Qualitative and descriptive interpretations* (N = 69.64%) and *Data processing tools and simple models* (N = 66.5%). Data analytics is used less for computationally heavy processes such as *Statistical analysis, data mining and learning, complex datasets* (N = 25.67%) or *Real-time analysis, deep learning*[2] *and predictive models, limited datasets* (N = 26.56%). While the level of use of data analytics is more distributed, simulation is clearly used most commonly to *Perform quick energy estimate and analysis* (N = 88.62%) and second to *Create virtual building components and compare design alternatives* (N = 52.90%). It is used least to *Analyze impacts of the model* (N = 29.24%). After the descriptive analysis, we then compared the effect of various demographic factors on the level of use of both data analytics and simulation (Table 3 and Table 4).

Participants from *Private* organizations are more likely to use data analytics for *Manual or spreadsheet oriented calculations* (coef = 0.71, sig = 0.00, 103.4% increase in odds), while participants from *Public* organizations are more likely to use it for *Qualitative and descriptive interpretations* (coef = 0.22, sig = 0.05, 24.6% increase in odds) or *Data processing tools and simple models* (coef = 0.31, sig = 0.04, 36.3% increase in odds). This could be due to the fact that professionals from *Public* organizations are encouraged to explore more advanced technologies [51]. *Principals* are more likely to use data analytics for less computationally complex uses such as *Qualitative and descriptive interpretations* (coef = 0.70, sig = 0.0, 101.4% increase in odds), *Manual or spreadsheet oriented calculations* (coef = 0.42, sig = 0.00, 52.2% increase in odds) and *Data processing and simple models* (coef = 0.67, sig = 0.03, 95.4% increase in odds), since the responsibilities of *Principals* primarily consist of strategic tasks, like many executive roles in industry settings [52]. Participants with *10-20 years* of experience in the building industry are more likely to use data analytics for simpler tasks such as *Manual or spreadsheet oriented calculations* (coef = 0.45, sig = 0.02, 56.8% increase in odds), while participants with 5-10 years of experience also use it for that (coef = 0.43, sig = 0.04, 53.7% increase in odds) as well as for *Data processing tools and simple models* (coef = 0.38, sig = 0.02, 46.2% increase in odds). This could be due to the fact that the new generation of building energy professionals are more exposed and trained on advanced technologies and tools [53].

There is a bifurcation for participants in *Building operations and maintenance* who use data analytics either for *Qualitative and descriptive interpretations* (coef = 0.52, sig = 0.01, 68.2% increase in odds) or *Real-time analysis, deep learning and predictive models, limited datasets* (coef = 0.54, sig = 0.02, 71.6% increase in odds); there are fewer in-between significant uses. This is an interesting finding which reflects how the work of energy professionals in this domain have diverged [54]. Participants in *Community and neighborhood management* are more likely to use data analytics for *Manual or spreadsheet oriented calculations* (coef = 0.58, sig = 0.05, 78.6% increase in odds) and *Data processing tools and simple models* (coef = 0.95, sig = 0.01, 158.6% increase in

[2] Deep learning is more computationally intensive and thus has higher hardware requirements (e.g. GPU) compared to traditional machine learning, representing a different level of use of data analytics in building energy management.

Table 3
Coefficients of factors that may affect level of use of data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

Demographic	Use	Qualitative and descriptive interpretations	Manual or spreadsheet oriented calculations	Data processing tools and simple models	Statistical analysis, data mining and learning, complex datasets	Real-time analysis, deep learning, and predictive models, limited datasets
Organization	Private		**0.71 (103.4%)**			
	Public	**0.22 (24.6%)**		**0.31 (36.3%)**		
Position	Principal	**0.70 (101.4%)**	**0.42 (52.2%)**	**0.67 (95.4%)**		
	10 to 20		**0.45 (56.8%)**			
Experience	5 to 10		**0.43 (53.7%)**	**0.38 (46.2%)**		
Professional Domain	Building operations and maintenance	**0.52 (68.2%)**				**0.54 (71.6%)**
	Community and neighborhood management		**0.58 (78.6%)**	**0.95 (158.6%)**		
	Grid optimization and demand response					**0.55 (73.3%)**

Table 4
Coefficients of factors that may affect level of use of simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

Demographic	Use	Create virtual building components and compare design alternatives	Test assumptions and hypothesis	Perform quick energy estimate and analysis	Examine sensitivity and effects of key parameters	Analyze climatic and neighboring impacts
Organization	Private			**0.52 (68.2%)**		
	Public			**0.45 (56.8%)**		
Position	Principal	**0.98 (166.4%)**		**0.21 (23.4%)**		
Experience	1 to 5				**0.46 (58.4%)**	
	10 to 20			**0.50 (64.9%)**		
Region	Midwest			**0.49 (63.2%)**		
	South			**0.49 (63.2%)**		
Professional Domain	Building materials					**0.40 (49.2%)**
	Building additions, alterations, and retrofitting			**0.78 (118.1%)**		
	Community and neighborhood management	**0.51 (66.5%)**		**0.77 (116.0%)**		
	District systems and lot facilities				**0.42 (52.2%)**	-

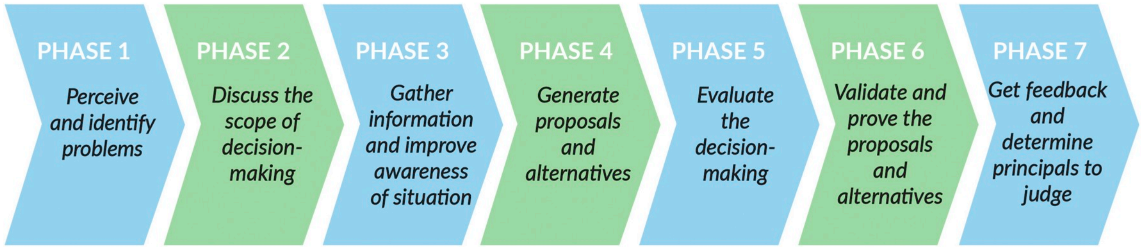

Fig. 5. The seven phases of the building energy management decision-making process.

odds), which makes sense as data analytics is often used for larger, urban scale modeling [55]. Participants in *Grid optimization and demand response* are also more likely to use data analytics for *Real-time analysis, deep learning and predictive models, limited datasets* (coef = 0.55, sig = 0.04, 73.3% increase in odds), conforming the cutting-edge nature of their work [56,57].

Several demographic types use simulation significantly to perform *Quick energy estimate and analysis*, including: *Private* organizations (coef = 0.52, sig = 0.02, 68.2% increase in odds), *Public* organizations (coef = 0.45, sig = 0.02, 56.8% increase in odds), *Principals* (coef = 0.21, sig = 0.02, 23.4% increase in odds), *10-20 years* experience (coef = 0.50, sig = 0.00, 64.9% increase in odds), *Midwest* (coef = 0.49, sig = 0.03, 63.2% increase in odds) and *South* regions (coef = 0.49, sig = 0.05, 63.2% increase in odds), *Building additions, alterations, and retrofitting* (coef = 0.78, sig = 0.00, 118.1% increase in odds) and *Community and neighborhood management* professionals (coef = 0.77, sig = 0.05, 116.0% increase in odds). *Principals* are also likely to use simulation to *Create virtual building components and compare design alternatives* (coef = 0.98, sig = 0.00, 166.4% increase in odds), a job that people in that position are likely to perform for higher level decision-making, similar to strategy roles for company board directors [58]. Participants with *1 to 5 years of* experience are most likely to use simulation to *Examine sensitivity and effects of key parameters* (coef = 0.46, sig = 0.05, 58.4% increase in odds). Sensitivity analysis is a more complicated simulation tool and it makes sense that younger employees trained with more cutting-edge technologies might have more experience with newer, more technical processes. This trend is representative of a larger notion of adolescents becoming proficient with technology at a young age [59].

Professionals in *Building materials* use simulation to *Analyze climatic and neighboring impacts* (coef = 0.40, sig = 0.03, 49.2% increase in odds), which makes sense as materials can affect urban microclimate and building energy use [60]. *Community and neighborhood management* are likely to use it to *Create virtual building components* (coef = 0.51, sig = 0.05, 66.5% increase in odds), while *District systems and lot facilities* mostly use simulation to *Examine sensitivity and effects of key parameters* (coef = 0.42, sig = 0.03, 52.2% increase in odds). We postulate that this is the case as models have emerged that now enable neighborhood and urban scale simulation and analysis [61,62].

Overall, our results indicate several trends in respect to factors influencing the adoption of data analytics and simulation for building energy management:

- Professional domain plays a significant effect on adoption of data analytics and simulation, but in general, position, years of experience and region do not.
- Most demographic factors (with the exception of region) have an effect on the adoption of data analytics for less computationally complex uses. Professional domain affects the adoption of data analytics at all levels of use.
- The use of simulation is more consistent than data analytics across demographic factors. It is common for professionals to use simulation sometimes and its primary use is *quick energy estimates and analysis*.

4.2. RQ2: utilization of data analytics and simulation in the decision-making process

We aim to understand the utilization of data analytics and simulation during the seven phases of the building energy management decision-making process (Fig. 5). Survey participants indicated the usefulness on a scale of 1–5 (1 = least useful, 5 = most useful). Both data analytics and simulation are considered more useful for Phase 2: *Discuss the scope of decision-making*, Phase 5: *Get feedback and determine principles to judge* and Phase 4: *Generate proposals and alternatives*. Data analytics is less useful for Phase 6: *Validate and prove the proposals and alternatives* and Phase 3: *Gather information and improve awareness of situation*, while simulation is less useful for Phase 7: *Evaluate the decision-making* and Phase 6: *Validate and prove the proposals and alternatives* (in order of perceived importance).

We used logistic regression tests to analyze in which phases of the decision-making process data analytics and simulation are most useful (Table 5 and Table 6), based on the level of use. Overall, data analytics and simulation are both most significant for Phase 1: *Perceive and identify problems*, Phase 4: *Generate proposals and alternatives* and Phase 5: *Get feedback and determine principles to judge*. Simulation is also significant for Phase 3: *Gather information and improve awareness of situation*.

For *Qualitative and descriptive interpretations*, data analytics is *very useful (4)* for Phase 1: *Perceive and identify problems* (coef = 0.39, sig = 0.04, 47.7% increase in odds), *most useful (5)* for Phase 4: *Generate proposals and alternatives* (coef = 0.53, sig = 0.0, 69.9% increase in odds) and *very useful (4)* for Phase 5: *Get feedback and determine principles to judge* (coef = 0.38, sig = 0.04, 46.2% increase in odds). For *Statistical analysis, data mining and learning, complex datasets*, it is *most useful (5)* for Phase 1: *Perceive and identify problems* (coef = 0.44, sig = 0.0, 55.3% increase in odds) and Phase 4: *Generate proposals and alternatives* (coef = 0.41, sig = 0.02, 50.7% increase in odds) while *Real-time analysis, deep learning and predictive models, limited datasets* is *very useful* (4) in Phase 5: *Get feedback and determine principles to judge* (coef = 0.39, sig = 0.05, 47.7% increase in odds). For participants who use data analytics for *Data processing tools and simple models*, it is only *somewhat useful* (2) for Phase 5: *Get feedback and determine principles to judge* (coef = 0.36, sig = 0.05, 43.3% increase in odds).

We then investigated in which phases of the decision-making process simulation is most useful for (Table 6), based on the level of use. For participants that use simulation to *Create virtual building components and compare design alternatives*, it is *very useful (4)* for Phase 3: *Gather information and improve awareness of situation* (coef = 0.46, sig = 0.04, 58.4% increase in odds). *Testing hypothesis and assumptions* is *most useful*

Table 5
Usefulness of data analytics in each phase of decision-making process by level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Phase 1: Perceive and identify problems				
	1	2	3	4	5
Qualitative and descriptive interpretations				**0.39 (47.7%)**	
Statistical analysis, data mining and learning, complex datasets					**0.44 (55.3%)**
	Phase 4: Generate proposals and alternatives				
	1	2	3	4	5
Qualitative and descriptive interpretations					**0.53 (69.9%)**
Statistical analysis, data mining and learning, complex datasets					**0.41 (50.7%)**
	Phase 5: Get feedback and determine principles to judge				
	1	2	3	4	5
Qualitative and descriptive interpretations				**0.38 (46.2%)**	
Data processing tools and simple models		**0.36 (43.3%)**			
Real-time analysis, deep learning, and predictive models, limited datasets				**0.39 (47.7%)**	

Table 6
Usefulness of simulation in phases of decision-making process based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Phase 1: Perceive and identify problems				
	1	2	3	4	5
Test assumptions and hypothesis		**0.57 (76.8%)**			
	Phase 3: Gather information and improve awareness of situations				
	1	2	3	4	5
Create virtual building components and compare design alternatives				**0.46 (58.4%)**	
Analyze climatic and neighboring impacts				**0.31 (36.3%)**	
	Phase 4: Generate proposals and alternatives				
	1	2	3	4	5
Analyze climatic and neighboring impacts		**0.51 (66.5%)**			
	Phase 5: Get feedback and determine principles to judge				
	1	2	3	4	5
Test assumptions and hypothesis					**0.59 (80.4%)**
Examine sensitivity and effects of key parameters				**0.43 (53.7%)**	

(5) for Phase 5: *Get feedback and determine principles to judge* (coef = 0.59, sig = 0.0, 80.4% increase in odds) and only *somewhat useful (2)* for Phase 1: *Perceive and identify problems* (coef = 0.57, sig = 0.00, 76.8% increase in odds). For the use *Examine sensitivity and effects of key parameters*, simulation is *very useful (4)* in Phase 5: *Get feedback and determine principles to judge* (coef = 0.43, sig = 0.05, 53.7% increase in odds), while *Analyzing climatic and neighboring impacts* is *very useful (4)* for Phase 3: *Gather information and improve awareness of situation* (coef = 0.31, sig = 0.05, 36.3% increase in odds) and only *somewhat useful (2)* for Phase 4: *Generate proposals and alternatives* (coef = 0.51, sig = 0.03, 66.5% increase in odds). The association of usefulness of data analytics and simulation across project design phases suggests there might be opportunities to further improve the integration of technologies in the building design domain [63].

Overall, our analysis indicates that:

- When data analytics is used for qualitative or statistical analysis, it is *most useful* to *Perceive and identify problems identification* (Phase 1) and *Generate proposals and alternatives* (Phase 4). When doing both qualitative and real-time analysis, data analytics is *most useful* to *Get feedback and determine principles to judge* (Phase 5).
- Simulation is *most useful* to *Gather information and improve awareness of situations* (Phase 3) when creating alternatives or analyzing impacts. It is also *very useful* to *Get feedback and determine principles to judge* (Phase 5) when testing assumptions or doing sensitivity analysis.
- Simulation and data analytics were found to be useful for similar phases in the decision-making process and therefore opportunities may exist for integrated tools that are capable of leveraging both methods.

4.3. RQ3: barriers and suggestions for improvement of data analytics and simulation

The next set of questions identify the barriers associated with using data analytics (Table 7) and simulation (Table 8) in building energy management decision-making. The most common barrier for data analytics is *Low data quality, inaccurate and missing data* (N = 68.08%), while the least common barrier is *Unexpected volume, velocity, and variety of data* (N = 38.39%). These barriers are consistent across current position, years of experience and professional domain. The most common barrier for simulation is *Effort and time required to build models* (N = 79.69%) and the least common barrier is *Diverse simulation programs and algorithms, inconsistent performances* (N = 37.05%).

Logistic regression tests were also conducted to examine how the perceived barriers are related to the different levels of use. Regarding data analytics, the tests reveal that for the use of *Qualitative and descriptive interpretations*, the biggest barrier is *Low data quality*, *inaccurate and missing data* (coef = 0.32, sig = 0.05, 37.7% increase in odds), which is intuitive given the nature of qualitative analysis. For *Manual or spreadsheet oriented calculations*, the main barriers are *Lack of timely and relevant data* (coef = 0.51, sig = 0.01, 66.5% increase in odds) and *Unexpected volume, velocity and variety of data* (coef = 0.46, sig = 0.04, 58.4% increase in odds). These are understandable barriers for manual calculations and spreadsheet-based analyses, which often only capture one moment in time, unlike real-time data [64] and can also become cumbersome to deal with for large datasets [65]. There were no significant barriers for the computationally complex levels of use (i.e., *Data processing tools and simple models*; *Statistical analysis, data mining and learning, complex datasets; and Real-time analysis, deep learning and predictive models, limited datasets*).

Table 7
Barriers for data analytics based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Lack of timely and relevant data	Unexpected volume, velocity, and variety of data	Low data quality, inaccurate and missing data	Insufficient expertise and inexperience to analyze data
Qualitative and descriptive interpretations			**0.32 (37.7%)**	
Manual or spreadsheet oriented calculations	**0.51 (66.5%)**	**0.46 (58.4%)**		

Table 8
Barriers for simulation based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Inaccurate and unreliable simulation results	Diverse simulation programs and algorithms, inconsistent performances	Efforts and time required to build models	Lack of expertise to analyze input-output relationships
Create virtual building components and compare design alternatives	**0.32 (37.7%)**			
Examine sensitivity and effects of key parameters				**0.40 (49.2%)**

Table 9
Criteria for improvement based on barrier with data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Criteria 8: Graphing and visualization				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Unexpected volume, velocity, and variety of data	**0.34 (40.5%)**				
	Criteria 10: Transparency of analytics process (e.g. assumptions, limitations, risks)				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Unexpected volume, velocity, and variety of data		**0.71 (103.4%)**			

Table 10
Criteria for improvement based on barrier with simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level, the first number is the value of coefficient, and the number in parenthesis is the percentage increase in odds, calculated by the exponential of coefficient).

	Criteria 2: Simple input methods for review and modification				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results	**0.33 (39.1%)**				
	Criteria 5: High model resolution				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results			**0.33 (39.1%)**		

In the use of simulation to *Create virtual building components and compare design alternatives*, the main barrier is *Inaccurate and unreliable simulation results* (coef = 0.32, sig = 0.05, 37.7% increase in odds), which conforms to a previous finding that accuracy is one of the primary issues with energy simulation [66]. The use of simulation to *Examine sensitivity and effects of key parameters* often faces a *Lack of expertise to analyze input-output relationship* (coef = 0.40, sig = 0.01, 49.2% increase in odds), a common barrier in sensitivity analysis for building energy analysis [67].

Finally, we analyzed how the use of data analytics and simulation can be improved to better support decision making. Participants rated various criteria on a scale of five-level importance for the different areas of improvement in data analytics (Table 9). From the descriptive results, it can be seen the most important expectations are *Ease of interpretation and follow-up*, *Informative conclusions extracted* and *Integration of data from different sources*. Improvement related to *Flexible targeting and navigation* and *Quick and easy evaluations of alternatives* are relatively less important. Participants also rated the importance of various criteria on a scale of five for the different areas of improvement in simulation (Table 10). The most important expectations are *Accuracy and robustness*, *Reduced uncertainty* and *Simple input methods for review and modification*. Less important for decision-making are *High model resolution* and *Results analysis* (such as statistical analysis or summaries).

Logistic regression tests analyzing the expectations for improvement based on the perceived barrier for data analytics reveal that for the barrier of *Unexpected volume, velocity and variety of data*, it is *Extremely important* to improve Criteria 8: *Graphing and visualization* (coef = 0.34, sig = 0.03, 40.5% increase in odds) and *Very important* to improve Criteria 10: *Transparency of analytics process*, e.g. *assumptions, limitations, risks* (coef = 0.71, sig = 0.0, 103.4% increase in odds). The results indicate the perceived importance of graphing and visualization in data analytics for building energy management, which conform to some currently ongoing initiatives to mitigate the lack of strong visual representation for big data in general [68–70]. None of the other barriers have significant impacts on criteria for improvement.

Logistic regression tests analyzing the expectations for improvement based on the perceived barrier for simulation reveal that for the barrier of *Inaccurate and unreliable simulation results*, it is *Extremely important* to improve Criteria 2: *Simple input methods for review and modification* (coef = 0.33, sig = 0.04, 39.1% increase in odds) and *Moderately important* to improve Criteria 5: *High model resolution* (coef = 0.33, sig = 0.04, 39.1% increase in odds). This finding supports previous works that demonstrated it can be difficult to rerun simulations in the case of inaccurate results if the input or modification methods are complicated [71] and that unreliable simulation results can be caused by low resolution models [72]. None of the other barriers have significant impacts on criteria for improvement.

Overall, our analysis indicates several key trends related to barriers and areas of improvements for data analytics and simulation:

- Professionals using data analytics for less computationally complex uses (qualitative and descriptive interpretations, Manual or spreadsheet oriented calculations) are more likely to face barriers with the software related to *Inaccurate, unexpected or irrelevant data.*
- Professionals using simulation for computationally complex uses such as *Examine sensitivity and effects of key parameters* are more likely to face barriers of *Insufficient expertise and inexperience to analyze data,* while professionals using it for more straightforward uses such as *Create virtual building components and compare design alternatives* are likely to face barriers of *Inaccurate and unreliable simulation results.*
- For data analytics, the barrier of *Unexpected volume, velocity and variety of data* is most strongly correlated with a desire to improve the *Graphing and visualization* and *Transparency of analytical process.*
- For simulation, the barrier of *Inaccurate and unreliable simulation results* is most strongly correlated with a desire to improve *Simple input methods for review and modification* and *High model resolution.*

5. Limitations and future work

While this study aims to take a first-step in deepening our understanding of how data analytics and simulation are used in the building energy management decision-making process, several limitations exist. First, while the survey population is large and diverse enough to yield statistically significant and meaningful results, future work that increases the sample size and a more even distribution across various demographic factors could enable more detailed and less biased insights. For example, the percentage of engineers is much higher than other positions and more participants have 10 + years of experience. A larger and more diverse sample of survey participants would also enable the use of various analytical methods that require more data (e.g., Ridge regression, Lasso regression, association rule mining) and capable of uncovering more nuanced insights. In addition, in this paper only the statistically significant coefficients were presented and discussed, detailed and deep analysis of the insignificant coefficients is not included. Further study is thus necessary to dive into each logistic regression and uncover new insights on the multi-correlations of the answers to the questions. Second, only building energy management professionals focused on the O&M phase in the United States were surveyed, thus the results might be biased to the special conditions in the American building energy management domains represented in the O&M phase. More worldwide responses could significantly improve the representativeness of the analysis for the whole life-cycle of building energy management and could be taken up in future work. Additionally, the effects and implications of the proposed analysis could be further enhanced by coupling the survey data with other data from interviews or focus groups from both the U.S. and other countries around the world. Third, while this paper compares the same phases of the building energy management decision-making process for data analytics and simulation, the uses, barriers and expectations for improvement are not exactly equivalent, demonstrating the existence of different requirements and demands for data analytics and simulation as the decision-making process proceeds. As a result, the process of integrating data analytics and simulation to complement each other and improve actual decision-marking regarding building energy management still remains unanswered and could be the scope of future research.

6. Conclusions and implications

In order to establish a comprehensive and clear understanding about the use, barriers and expectations of data analytics and simulation in each phase of the building energy management decision-making process, this paper analyzes a nationwide survey completed on 448 building management professionals; the responses are diverse and representative. The three main foci of the survey analysis include: 1) what impacts the adoption of data analytics and simulation in building energy management; 2) in what phases of the energy management decision-making process are data analytics and simulation currently used; and 3) what are the barriers to use for data analytics and simulation and how can they be improved. The survey analysis employs bootstrapped logistic regression to improve the reliability and consistency of analysis results. Overall, our study reveals several key insights:

- Professional domain plays a large role in driving the uses, barriers and expectations for data analytics and simulation tools.
- Data analytics and simulation are most used in similar phases of the decision-making process and can be coupled to leverage their functions.
- The accuracy of results needs to be improved for both data analytics and simulation tools.
- Professionals need more and improved training, especially for simulation tools.

This study represents one of the first works to survey building energy management professionals across the United States about their use, barriers and expectations of both data analytics and simulation. The results provide a quantitative basis for both academia and industry to improve the efficacy of these tools and integrate their functions such that the adoption of data analytics and simulation is increased within the building energy management domain. Both data analytics and simulation are tools that will undoubtedly play a crucial role in improving the energy-efficiency of our building stock and thus will have a substantial impact on our ability to meet the world's sustainable energy goals.

Declarations of interest

None.

Acknowledgements

This material is based upon the work supported in part by the Stanford School of Engineering under a Terman Faculty Fellowship and the National Science Foundation under Grants No. 1461549 and 1642315. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We would also like to thank the USGBC (U.S. Green Building Council), ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers), IFMA (International Facility Management Association), BCA (Building Commissioning Association), and AEE (Association of Energy Engineers) for their assistance in distributing our survey to their members.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.buildenv.2019.04.016.

Appendix B

Table 1
Coefficients of factors that may affect adoption of data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

Demographic	Frequency of use	Always	Often	Sometimes	Seldom	Never
Organization	Private	-0.38	-0.26	-0.21	-0.21	-0.68
	Public	-0.39	-0.33	-0.13	-0.18	-0.59
Position	Director	0.27	0.07	-0.13	-0.35	-1.00
	Engineer	0.00	-0.05	-0.16	-0.24	-0.11
	Manager	**0.52**	-0.22	-0.36	-0.03	-0.50
	President	-1.13	-0.09	-0.01	-0.03	0.43
	Principal	-0.43	-0.31	0.32	0.25	-0.10
Experience	1 to 5	0.01	-0.29	-0.23	0.22	-0.41
	5 to 10	-0.11	-0.31	-0.13	0.07	-0.41
	10 to 20	-0.20	-0.07	0.09	-0.58	0.34
	20+	-0.47	0.07	-0.07	-0.10	-0.80
Region	Midwest	0.09	-0.33	-0.10	-0.10	-0.23
	Northeast	-0.40	-0.04	-0.03	-0.15	-0.30
	South	-0.35	0.06	-0.34	0.02	-0.66
	West	-0.12	-0.29	0.12	-0.16	-0.10
Professional Domain	Building envelop and geometry	-0.17	0.26	0.08	-0.41	-0.08
	Building materials	-0.21	0.06	0.10	-0.13	-0.31
	Building systems and equipment	-0.56	0.33	-0.44	0.38	-0.48
	Building appliances	0.23	0.03	0.01	-0.15	-0.56
	Building operations and maintenance	0.03	0.10	-0.26	-0.25	**0.40**
	Building commissioning and energy auditing	0.02	0.19	-0.29	-0.04	-0.18
	Building additions, alterations, and retrofitting	0.20	-0.30	**0.49**	-0.28	-0.15
	Building occupancies	**0.62**	-0.05	-0.05	-0.70	-0.46
	Building utilities	-0.41	0.04	0.09	0.20	-0.61
	Energy finance and market	0.37	-0.15	**0.43**	-0.34	-0.73
	Building carbon emissions	0.13	0.02	-0.60	0.31	0.32
	Community and neighborhood management	-0.26	0.07	0.09	0.12	-0.49
	Grid optimization and demand response	-0.10	-0.26	0.19	0.34	-0.83
	District systems and lot facilities	0.12	0.39	-0.13	-0.51	-0.65

Table 2
Coefficients of factors that may affect adoption of simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

Demographic	Frequency of use	Always	Often	Sometimes	Seldom	Never
Organization	Private	-0.43	-0.16	-0.05	-0.36	-0.92
	Public	-0.70	-0.22	-0.04	-0.29	-0.49
Position	Director	-0.01	0.17	-0.18	-0.36	-0.60
	Engineer	-0.24	0.01	-0.09	-0.21	-0.33
	Manager	-0.34	-0.67	0.20	0.18	-0.30
	President	-0.25	0.18	-0.39	-0.14	-0.02
	Principal	-0.29	-0.08	0.38	-0.12	-0.16
Experience	1 to 5	0.07	-0.31	0.38	-0.33	-0.87
	5 to 10	-0.40	0.00	-0.16	0.01	-0.09
	10 to 20	-0.20	0.08	-0.01	-0.34	-0.03
	20+	-0.60	-0.14	-0.30	0.01	-0.42
Region	Midwest	-0.33	-0.14	0.03	-0.18	-0.09
	Northeast	0.10	-0.19	-0.10	-0.27	0.03
	South	-0.67	0.09	-0.10	-0.09	-0.46
	West	-0.23	-0.14	0.09	-0.12	-0.89
Professional Domain	Building envelop and geometry	0.08	-0.09	-0.37	0.28	0.12
	Building materials	0.10	-0.02	0.04	-0.01	-0.59
	Building systems and equipment	0.13	-0.33	-0.71	**0.57**	**0.55**
	Building appliances	0.44	0.04	0.27	-0.32	-1.26
	Building operations and maintenance	0.13	**0.63**	-0.59	0.06	-0.91
	Building commissioning and energy auditing	-0.59	0.04	0.01	-0.05	**0.68**
	Building additions, alterations, and retrofitting	0.34	-0.03	-0.22	0.03	-0.05
	Building occupancies	-0.38	-0.25	0.65	-0.22	-0.34
	Building utilities	0.05	-0.32	0.07	0.24	-0.54
	Energy finance and market	-0.37	-0.16	0.40	-0.18	0.32
	Building carbon emissions	0.15	0.13	0.17	-0.28	-0.49
	Community and neighborhood management	-0.05	-1.52	0.33	0.51	0.46
	Grid optimization and demand response	-0.21	-0.04	0.62	-0.33	-1.02
	District systems and lot facilities	0.37	-0.20	-0.52	0.26	0.23

Table 3
Coefficients of factors that may affect level of use of data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

Demographic	Use	Qualitative and descriptive interpretations	Manual or spreadsheet oriented calculations	Data processing tools and simple models	Statistical analysis, data mining and learning, complex datasets	Real-time analysis, deep learning, and predictive models, limited datasets
Organization	Private	0.10	**0.71**	-0.15	-0.18	-0.29
	Public	**0.22**	0.13	**0.31**	-0.23	-0.38
Position	Director	0.26	0.17	-0.27	-0.03	-0.16
	Engineer	-0.12	0.06	-0.26	-0.05	-0.33
	Manager	0.17	-0.22	-0.05	-0.02	-0.59
	President	-0.70	0.42	0.07	-0.70	0.44
	Principal	**0.70**	**0.42**	**0.67**	0.40	-0.02
Experience	1 to 5	-0.05	0.01	0.03	0.28	0.15
	10 to 20	0.16	**0.45**	-0.11	-0.15	-0.39
	20+	0.05	-0.06	-0.15	-0.18	-0.15
	5 to 10	0.15	**0.43**	**0.38**	-0.36	-0.28
Region	Midwest	-0.10	0.11	0.03	-0.27	-0.35
	Northeast	0.28	0.27	-0.06	-0.18	-0.16
	South	0.09	0.41	-0.02	-0.06	-0.09
	West	0.05	0.05	0.21	0.11	-0.07
Professional Domain	Building envelop and geometry	-0.23	-0.16	-0.13	0.19	0.29
	Building materials	-0.25	-0.01	0.31	-0.04	0.06
	Building systems and equipment	0.20	0.37	0.29	0.08	-0.26
	Building appliances	0.27	-0.07	0.17	-0.25	-0.08
	Building operations and maintenance	**0.52**	-0.24	0.16	-0.21	**0.54**
	Building commissioning and energy auditing	-0.07	0.12	0.33	-0.10	0.30
	Building additions, alterations, and retrofitting	0.02	-0.25	0.22	-0.22	-0.20
	Building occupancies	0.30	0.04	-0.48	-0.07	-0.34
	Building utilities	0.00	0.08	0.18	-0.02	0.08
	Energy finance and market	-0.30	0.41	0.16	0.06	-0.23
	Building carbon emissions	-0.27	-0.60	-0.13	0.26	0.18
	Community and neighborhood management	0.23	**0.58**	**0.95**	-0.10	-0.43
	Grid optimization and demand response	-0.29	-0.21	0.00	-0.23	**0.55**
	District systems and lot facilities	0.10	0.06	-0.64	0.12	0.04

Table 4
Coefficients of factors that may affect level of use of simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

Demographic	Use	Create virtual building components and compare design alternatives	Test assumptions and hypothesis	Perform quick energy estimate and analysis	Examine sensitivity and effects of key parameters	Analyze climatic and neighboring impacts
Organization	Private	-0.07	-0.09	**0.52**	0.08	-0.43
	Public	-0.15	-0.17	**0.45**	-0.36	-0.10
Position	Director	-0.14	-0.17	0.56	-0.29	-0.27
	Engineer	-0.22	0.23	0.18	-0.52	-0.19
	Manager	-0.62	0.11	0.10	-0.19	0.00
	President	-0.22	0.08	-0.08	0.15	0.06
	Principal	**0.98**	-0.49	**0.21**	0.57	-0.12
Experience	1 to 5	0.22	-0.27	0.04	**0.46**	0.18
	10 to 20	0.04	0.07	**0.50**	-0.56	-0.44
	20+	-0.25	-0.07	0.35	-0.08	-0.23
	5 to 10	-0.23	0.02	0.09	-0.10	-0.04
Region	Midwest	-0.29	0.26	**0.49**	0.05	-0.06
	Northeast	0.08	-0.06	0.16	-0.52	-0.07
	South	-0.15	-0.22	**0.49**	-0.05	0.01
	West	0.14	-0.22	-0.17	0.24	-0.41
Professional Domain	Building envelop and geometry	0.02	0.01	-0.03	0.01	0.15
	Building materials	-0.25	0.26	0.11	-0.23	**0.40**
	Building systems and equipment	0.33	0.31	-0.31	0.46	0.08
	Building appliances	0.04	-0.07	0.34	0.03	-0.05
	Building operations and maintenance	0.31	-0.50	0.00	-0.17	0.15
	Building commissioning and energy auditing	0.35	0.07	0.15	0.18	-0.24
	Building additions, alterations, and retrofitting	0.19	0.23	**0.78**	0.23	0.18
	Building occupancies	-0.15	0.04	0.21	-0.12	-0.15
	Building utilities	-0.03	-0.18	-0.02	-0.19	0.09
	Energy finance and market	0.21	-0.40	-0.31	-0.52	0.08
	Building carbon emissions	0.22	0.13	-0.51	0.05	0.04
	Community and neighborhood management	**0.51**	0.10	**0.77**	0.58	-0.07
	Grid optimization and demand response	-0.08	-0.10	-0.43	0.03	0.05
	District systems and lot facilities	0.00	-0.26	0.24	**0.42**	-0.02

Table 5
Usefulness of data analytics in each phase of decision-making process by level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Step 1: Perceive and identify problems				
	1	2	3	4	5
Qualitative and descriptive interpretations	0.00	0.06	-0.70	**0.39**	-0.24
Manual or spreadsheet oriented calculations	-0.03	0.21	-0.25	-0.05	-0.31
Data processing tools and simple models	-0.41	0.07	0.36	-0.24	-0.07
Statistical analysis, data mining and learning, complex datasets	-0.39	-0.36	0.12	-0.38	**0.44**
Real-time analysis, deep learning, and predictive models, limited datasets	-0.25	0.18	-0.06	-0.13	-0.01
	Step 2: Discuss the scope of decision making				
	1	2	3	4	5
Qualitative and descriptive interpretations	-0.03	-0.25	0.08	-0.10	-0.01
Manual or spreadsheet oriented calculations	0.17	0.00	-0.19	-0.75	0.25
Data processing tools and simple models	0.24	-0.27	-0.14	0.03	-0.17
Statistical analysis, data mining and learning, complex datasets	0.18	-0.19	-0.01	0.05	-0.29
Real-time analysis, deep learning, and predictive models, limited datasets	-0.12	-0.30	-0.02	0.46	-0.11
	Step 3: Gather information and improve awareness of situations				
	1	2	3	4	5
Qualitative and descriptive interpretations	-0.28	-0.37	0.23	-0.08	-0.15
Manual or spreadsheet oriented calculations	-0.91	-0.34	-0.16	0.21	-0.17
Data processing tools and simple models	-0.47	0.09	0.12	-0.13	-0.11
Statistical analysis, data mining and learning, complex datasets	0.04	-0.38	0.02	0.19	-0.20
Real-time analysis, deep learning, and predictive models, limited datasets	-0.30	0.07	-0.19	-0.24	0.23
	Step 4: Generate proposals and alternatives				
	1	2	3	4	5
Qualitative and descriptive interpretations	-0.24	0.18	-0.15	-0.48	**0.53**
Manual or spreadsheet oriented calculations	-0.05	-0.24	0.03	-0.31	0.15
Data processing tools and simple models	0.07	0.02	-0.32	0.04	-0.04
Statistical analysis, data mining and learning, complex datasets	-0.13	0.17	-0.25	-0.40	**0.41**
Real-time analysis, deep learning, and predictive models, limited datasets	0.38	-0.73	-0.03	0.00	0.10

	Step 5: Get feedback and determine principles to judge				
	1	2	3	4	5
Qualitative and descriptive interpretations	-0.09	-0.44	-0.10	**0.38**	0.06
Manual or spreadsheet oriented calculations	-0.32	-0.22	-0.17	-0.16	0.37
Data processing tools and simple models	-0.47	**0.36**	-0.23	0.07	0.01
Statistical analysis, data mining and learning, complex datasets	-0.23	-0.41	0.25	0.09	-0.05
Real-time analysis, deep learning, and predictive models, limited datasets	0.01	-0.23	-0.40	**0.39**	0.14
	Step 6: Validate and prove the proposals and alternatives				
	1	2	3	4	5
Qualitative and descriptive interpretations	-0.18	-0.37	-0.21	0.07	0.09
Manual or spreadsheet oriented calculations	-0.89	0.05	0.17	-0.40	0.18
Data processing tools and simple models	-0.45	-0.03	-0.28	0.07	0.04
Statistical analysis, data mining and learning, complex datasets	-0.01	0.34	-0.09	-0.42	0.17
Real-time analysis, deep learning, and predictive models, limited datasets	0.25	-0.41	-0.21	-0.07	0.21
	Step 7: Evaluate the decision making				
	1	2	3	4	5
Qualitative and descriptive interpretations	0.01	0.01	-0.05	-0.08	-0.09
Manual or spreadsheet oriented calculations	-0.92	0.23	-0.17	0.22	-0.21
Data processing tools and simple models	-0.19	-0.08	0.06	-0.06	-0.05
Statistical analysis, data mining and learning, complex datasets	0.06	-0.10	-0.26	0.34	-0.18
Real-time analysis, deep learning, and predictive models, limited datasets	-0.56	0.24	0.35	-0.28	-0.13

Table 6
Usefulness of simulation in phases of decision-making process based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Step 1: Perceive and identify problems				
	1	2	3	4	5
Create virtual building components and compare design alternatives	-0.12	-0.09	0.26	-0.14	-0.01
Test assumptions and hypothesis	-0.16	**0.57**	-0.96	-0.02	-0.19
Perform quick energy estimate and analysis	0.01	-0.24	-0.64	-0.26	0.34
Examine sensitivity and effects of key parameters	0.03	0.23	-0.09	-0.60	0.11
Analyze climatic and neighboring impacts	0.31	-0.01	-0.72	-0.05	0.04
	Step 2: Discuss the scope of decision making				
	1	2	3	4	5
Create virtual building components and compare design alternatives	-0.18	0.01	-0.09	0.29	-0.19
Test assumptions and hypothesis	0.08	-0.04	-0.18	0.20	-0.47
Perform quick energy estimate and analysis	-0.07	0.00	0.08	-0.46	-0.97
Examine sensitivity and effects of key parameters	-0.17	-0.02	0.26	-0.32	-0.04
Analyze climatic and neighboring impacts	-0.12	0.32	-0.04	-0.26	-1.06
	Step 3: Gather information and improve awareness of situations				
	1	2	3	4	5
Create virtual building components and compare design alternatives	0.16	0.08	-0.67	**0.46**	-0.09
Test assumptions and hypothesis	-0.15	-0.45	0.04	-0.06	0.22
Perform quick energy estimate and analysis	-0.15	-0.03	-0.51	0.09	0.03
Examine sensitivity and effects of key parameters	-0.14	0.12	-0.01	0.01	-0.05
Analyze climatic and neighboring impacts	0.16	0.26	-0.58	**0.31**	-0.21
	Step 4: Generate proposals and alternatives				
	1	2	3	4	5
Create virtual building components and compare design alternatives	0.09	0.02	-0.12	-0.30	0.07
Test assumptions and hypothesis	0.32	0.01	-0.31	-0.04	-0.06
Perform quick energy estimate and analysis	0.23	-0.37	-0.08	-0.09	-0.31
Examine sensitivity and effects of key parameters	-0.25	-0.04	0.03	-0.16	0.20
Analyze climatic and neighboring impacts	-0.03	**0.51**	0.07	-0.21	-0.46

	Step 5: Get feedback and determine principles to judge				
	1	2	3	4	5
Create virtual building components and compare design alternatives	0.02	0.20	-0.29	-0.01	-0.02
Test assumptions and hypothesis	0.06	-0.26	-0.16	-0.25	**0.59**
Perform quick energy estimate and analysis	-0.01	-0.49	0.10	0.04	-0.35
Examine sensitivity and effects of key parameters	-0.18	-0.11	-0.14	**0.43**	-0.26
Analyze climatic and neighboring impacts	0.23	0.03	0.19	-0.29	-0.45
	Step 6: Validate and prove the proposals and alternatives				
	1	2	3	4	5
Create virtual building components and compare design alternatives	-0.05	-0.33	0.31	0.31	0.31
Test assumptions and hypothesis	0.23	0.19	-0.22	-0.22	-0.22
Perform quick energy estimate and analysis	-0.62	0.07	-0.30	-0.30	-0.30
Examine sensitivity and effects of key parameters	-0.45	0.01	0.20	0.20	0.20
Analyze climatic and neighboring impacts	0.19	0.12	-0.16	-0.16	-0.16
	Step 7: Evaluate the decision making				
	1	2	3	4	5
Create virtual building components and compare design alternatives	-0.74	0.09	0.02	-0.13	0.32
Test assumptions and hypothesis	0.01	-0.46	0.05	-0.18	0.34
Perform quick energy estimate and analysis	0.06	-0.14	-0.22	-0.48	0.32
Examine sensitivity and effects of key parameters	0.13	-0.34	0.22	-0.04	-0.17
Analyze climatic and neighboring impacts	0.08	-0.01	-0.31	-0.05	0.18

Table 7
Barriers to adopt data analytics based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Lack of timely and relevant data	Unexpected volume, velocity, and variety of data	Low data quality, inaccurate and missing data	Insufficient expertise and inexperience to analyze data
Qualitative and descriptive interpretations	0.08	-0.17	**0.32**	-0.18
Manual or spreadsheet oriented calculations	**0.51**	**0.46**	-0.05	0.01
Data processing tools and simple models	-0.09	-0.11	-0.27	-0.02
Statistical analysis, data mining and learning, complex datasets	0.03	0.29	-0.24	-0.11
Real-time analysis, deep learning, and predictive models, limited datasets	-0.08	-0.05	-0.06	0.00

Table 8
Barriers to adopt simulation based on level of use (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Inaccurate and unreliable simulation results	Diverse simulation programs and algorithms, inconsistent performances	Efforts and time required to build models	Lack of expertise to analyze input-output relationships
Create virtual building components and compare design alternatives	**0.32**	-0.14	-0.13	-0.43
Test assumptions and hypothesis	-0.21	0.15	0.27	-0.24
Perform quick energy estimate and analysis	0.28	-0.34	-0.13	0.15
Examine sensitivity and effects of key parameters	0.00	0.17	-0.12	**0.40**
Analyze climatic and neighboring impacts	-0.07	0.03	0.05	-0.09

Table 9
Criteria for improvement based on barriers to adopt data analytics (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Criteria 1: Ease of interpretation and follow-up				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	-0.24	0.18	0.25	-0.85	0.35
Unexpected volume, velocity, and variety of data	0.09	-0.16	0.14	-0.64	-0.11
Low data quality, inaccurate and missing data	-0.31	0.14	0.10	-0.68	0.24
Insufficient expertise and inexperience to analyze data	-0.10	-0.03	0.18	-0.14	-0.09
	Criteria 2: Informative conclusions extracted				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	-0.24	0.18	0.25	-0.85	-1.57
Unexpected volume, velocity, and variety of data	0.09	-0.16	0.14	-0.64	-0.29
Low data quality, inaccurate and missing data	-0.31	0.14	0.10	-0.68	-0.38
Insufficient expertise and inexperience to analyze data	-0.10	-0.03	0.18	-0.14	-0.79
	Criteria 3: Interactions with analytics				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	0.16	-0.05	-0.11	-0.86	0.26
Unexpected volume, velocity, and variety of data	-0.97	0.14	-0.03	0.24	0.27
Low data quality, inaccurate and missing data	-0.01	-0.04	0.00	-0.81	-0.50
Insufficient expertise and inexperience to analyze data	0.36	-0.13	0.01	-0.49	-0.56
	Criteria 4: Integration of data from different sources				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	-0.07	0.07	-0.08	-0.20	0.05
Unexpected volume, velocity, and variety of data	-0.24	0.15	-0.07	-0.13	-0.11
Low data quality, inaccurate and missing data	-0.15	0.11	-0.08	-0.28	-0.19
Insufficient expertise and inexperience to analyze data	0.02	0.04	0.09	-0.75	0.02
	Criteria 5: Easy to learn and flat learning curve				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	-0.40	0.04	0.03	0.20	0.25
Unexpected volume, velocity, and variety of data	-0.26	-0.11	0.12	0.20	0.15
Low data quality, inaccurate and missing data	0.27	-0.09	-0.21	0.02	0.08
Insufficient expertise and inexperience to analyze data	-0.30	0.29	-0.20	0.06	0.21
	Criteria 6: Flexible targeting and navigation				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	0.14	0.23	-0.22	-0.27	-0.24
Unexpected volume, velocity, and variety of data	-0.73	0.19	-0.01	-0.11	-0.15
Low data quality, inaccurate and missing data	-0.25	-0.01	0.08	-0.28	-0.16
Insufficient expertise and inexperience	0.47	-0.09	-0.09	-0.08	-0.12

to analyze data					
	Criteria 7: Comparison to ideal indices				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	0.06	0.14	-0.11	-0.30	0.05
Unexpected volume, velocity, and variety of data	0.13	-0.49	0.28	0.01	0.21
Low data quality, inaccurate and missing data	0.12	-0.07	-0.08	-0.11	-0.09
Insufficient expertise and inexperience to analyze data	-0.32	0.04	-0.06	0.30	0.15
	Criteria 8: Graphing and visualization				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	0.07	-0.22	0.10	0.13	0.08
Unexpected volume, velocity, and variety of data	**0.34**	-0.39	-0.04	0.12	0.24
Low data quality, inaccurate and missing data	-0.09	0.15	-0.12	-0.15	-0.08
Insufficient expertise and inexperience to analyze data	-0.16	-0.27	0.44	0.31	0.22
	Criteria 9: Quick and easy evaluations of alternatives				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	0.24	-0.32	0.17	-0.52	-0.31
Unexpected volume, velocity, and variety of data	-0.13	0.17	-0.09	-0.32	-0.11
Low data quality, inaccurate and missing data	-0.12	0.06	-0.10	0.02	0.12
Insufficient expertise and inexperience to analyze data	-0.04	-0.03	0.12	-0.46	-0.15
	Criteria 10: Transparency of analytics process (e.g. assumptions, limitations, risks)				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Lack of timely and relevant data	-0.06	-0.04	0.24	-0.67	0.05
Unexpected volume, velocity, and variety of data	-0.41	**0.71**	-0.19	-0.69	-0.24
Low data quality, inaccurate and missing data	0.28	-0.19	-0.03	-0.59	-0.27
Insufficient expertise and inexperience to analyze data	0.11	-0.19	0.00	-0.04	-0.08

Table 10
Criteria for improvement based on barriers to adopt simulation (**bold** with gray background highlight indicates statistically significant results at the 0.05 level).

	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Criteria 1: Graphic representation					
Inaccurate and unreliable simulation results	-0.13	-0.14	0.02	0.07	0.15
Diverse simulation programs and algorithms, inconsistent performances	0.14	-0.19	0.03	0.03	-0.71
Efforts and time required to build models	-0.05	-0.20	0.00	0.04	-0.69
Lack of expertise to analyze input-output relationships	0.11	-0.04	-0.10	-0.48	0.24
Criteria 2: Simple input methods for review and modification					
Inaccurate and unreliable simulation results	**0.33**	-0.12	-0.39	0.12	0.21
Diverse simulation programs and algorithms, inconsistent performances	-0.05	0.00	-0.10	0.23	0.18
Efforts and time required to build models	-0.28	0.16	-0.06	-0.46	-0.26
Lack of expertise to analyze input-output relationships	-0.17	0.04	0.15	-0.58	-0.23
Criteria 3: Accuracy and robustness					
Inaccurate and unreliable simulation results	0.00	0.08	-0.69	0.05	0.17
Diverse simulation programs and algorithms, inconsistent performances	0.13	-0.06	-0.44	-0.32	-0.19
Efforts and time required to build models	-0.10	-0.06	0.03	0.08	0.11
Lack of expertise to analyze input-output relationships	0.11	-0.03	-0.23	0.21	0.09
Criteria 4: Reduced uncertainty					
Inaccurate and unreliable simulation results	-0.02	0.04	-0.20	-0.07	-0.04
Diverse simulation programs and algorithms, inconsistent performances	-0.12	0.15	0.01	-0.81	0.32
Efforts and time required to build models	-0.11	-0.07	0.31	-1.02	0.41
Lack of expertise to analyze input-output relationships	0.11	-0.08	-0.13	0.10	-0.21
Criteria 5: High model resolution					
Inaccurate and unreliable simulation results	-0.07	-0.26	**0.33**	-0.33	-0.26
Diverse simulation programs and algorithms, inconsistent performances	-0.81	-0.32	-0.28	0.35	0.05
Efforts and time required to build models	-1.02	-0.39	0.00	-0.26	-0.18
Lack of expertise to analyze input-output relationships	0.10	0.13	-0.15	-0.46	0.13
Criteria 6: Awareness of simulation algorithmic process, assumptions, and limitations					
Inaccurate and unreliable simulation results	0.07	-0.02	-0.07	-0.27	-0.51
Diverse simulation programs and algorithms, inconsistent performances	-0.13	0.03	0.06	-0.01	-0.94
Efforts and time required to build models	0.36	-0.01	-0.42	-0.32	-0.69
Lack of expertise to analyze input-output relationships	-0.12	0.02	0.08	-0.30	-0.26

	Criteria 7: Interoperability with other models (e.g. CAD)				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results	-0.40	0.26	-0.09	0.19	-0.09
Diverse simulation programs and algorithms, inconsistent performances	-0.07	0.00	0.14	-0.10	-0.61
Efforts and time required to build models	0.09	-0.31	-0.25	0.48	-0.29
Lack of expertise to analyze input-output relationships	0.07	-0.08	-0.08	-0.35	0.44
	Criteria 8: Results analysis (e.g. statistical analysis, summary)				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results	0.09	-0.07	-0.12	0.10	-0.87
Diverse simulation programs and algorithms, inconsistent performances	-0.34	0.16	0.11	-0.37	-0.23
Efforts and time required to build models	0.07	-0.10	-0.20	-0.03	-0.70
Lack of expertise to analyze input-output relationships	0.03	-0.16	0.05	0.20	-0.55
	Criteria 9: Quick and easy evaluations of alternatives				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results	-0.04	0.01	-0.12	0.09	-0.56
Diverse simulation programs and algorithms, inconsistent performances	-0.26	0.14	0.10	-0.65	-0.21
Efforts and time required to build models	0.06	0.33	-0.40	-0.60	-0.96
Lack of expertise to analyze input-output relationships	-0.05	-0.15	0.12	-0.03	-0.18
	Criteria 10: Transparency of analytics process (e.g. assumptions, limitations, risks)				
	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Inaccurate and unreliable simulation results	0.30	-0.18	-0.16	-0.67	-0.02
Diverse simulation programs and algorithms, inconsistent performances	-0.18	-0.02	-0.31	-0.69	0.32
Efforts and time required to build models	-0.34	0.24	0.24	-0.59	-0.29
Lack of expertise to analyze input-output relationships	-0.29	-0.41	0.48	-0.04	-0.36

References

[1] C.E. Ochoa, et al., State of the art in lighting simulation for building science: a literature review, J. Build. Perf. Simulat. 5 (4) (2012) 209–233.
[2] M. Inanici, A. Hashemloo, An investigation of the daylighting simulation techniques and sky modeling practices for occupant centric evaluations, Build. Environ. 113 (2017) 220–231.
[3] D. Coakley, et al., A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141.
[4] A. Foucquier, et al., State of the art in building modeling and energy performances prediction: a review, Renew. Sustain. Energy Rev. 23 (2013) 272–288.
[5] R. Kumar, et al., Aggarwal R, Sharma J. Energy analysis of a building using artificial neural network: a review, Energy Build. 65 (2013) 352–358.
[6] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592.
[7] X. Liang, et al., Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192.
[8] A. Nguyen, et al., A review on simulation-based optimization methods applied to building performance analysis, Appl. Energy 113 (2014) 1043–1058.
[9] Z. Li, et al., Methods for benchmarking building energy consumption against its past or intended performance: an overview, Appl. Energy 124 (2014) 325–334.
[10] J. Schiefelbein, et al., Automated Urban Energy System Modeling and Thermal Building Simulation Based on OpenStreetMap Data Sets, Build. Environ. 149 (2019) 630–639.
[11] C. Reinhart, A. Fitz, Findings from a survey on the current use of daylight simulations in building design, Energy Build. 38 (7) (2006) 824–835.
[12] S. Attia, et al., Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs, J. Build. Perf. Simulat. 5 (2012) 155–169.
[13] P. Bynum, et al., Building information modeling in support of sustainable design and construction, J. Constr. Eng. Manag. 139 (1) (2012) 24–34.
[14] A. Elmualim, et al., Application of computer-aided facilities management (CAFM) for intelligent buildings operation, Facilities 27 (11/12) (2009) 421–428.
[15] K.J. Lomas, H. Eppel, C.J. Martin, D.P. Bloomfield, Empirical validation of building energy simulation programs, Energy Build. 26 (3) (1997) 253–275.
[16] E.M. Ryan, T.F. Sanquist, Validation of building energy modeling tools under idealized and realistic conditions, Energy Build. 47 (2012) 375–382.
[17] P. de Wilde, The gap between predicted and measured energy performance of buildings: a framework for investigation, Autom. ConStruct. 41 (2014) 40–49.
[18] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141.
[19] A. Mwasha, R.G. Williams, J. Iwaro, Modeling the performance of residential building envelope: the role of sustainable energy performance indicators, Energy Build. 43 (9) (2011) 2108–2117.
[20] V.S.K.V. Harish, Arun Kumar, A review on modeling and simulation of building energy systems, Renew. Sustain. Energy Rev. 56 (2016) 1272–1292.
[21] D.B. Crawley, J.W. Hand, M. Kummert, B.T. Griffith, Contrasting the capabilities of building energy performance simulation programs, Build. Environ. 43 (4) (2008) 661–673.
[22] B. Becerik-Gerber, K. Kensek, Building information modeling in architecture, engineering, and construction: emerging research directions and trends, J. Prof. Barriers Eng. Educ. Pract. 136 (3) (2010) 139–147.
[23] S. Attia, J.L.M. Hensen, L. Beltrán, A. De Herde, Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs, J. Build. Perf. Simulat. 5 (3) (2012) 155–169.
[24] C. Bleil de Souza, A critical and theoretical analysis of current proposals for integrating building thermal simulation tools into the building design process, J. Build. Perf. Simulat. 2 (4) (2009) 283–297.
[25] J.A. Clarke, J.L.M. Hensen, Integrated building performance simulation: progress, prospects and requirements, Build. Environ. 91 (2015) 294–306.
[26] A.C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy

performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap, Appl. Energy 97 (2012) 355–364.
[27] S. Attia, M. Hamdy, W. O'Brien, S. Carlucci, Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design, Energy Build. 60 (2013) 110–124.
[28] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205.
[29] J. Wang, et al., Review on multi-criteria decision analysis aid in sustainable energy decision-making, Renew. Sustain. Energy Rev. 13 (9) (2009) 2263–2278.
[30] M. Molina-Solana, et al., Data science for building energy management: a review, Renew. Sustain. Energy Rev. 70 (2017) 598–609.
[31] D.B. Crawley, et al., Contrasting the capabilities of building energy performance simulation programs, Build. Environ. 43 (4) (2008) 661–673.
[32] S. Attia, et al., Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs, J. Build. Perform. Simulat. 5 (2012) 155–169.
[33] M. Pilgrim, N. Bouchlaghem, D. Loveday, M. Holmes, Towards the efficient use of simulation in building performance analysis: a user survey, Build. Serv. Eng. Technol. 24 (3) (2003) 149–162.
[34] S. Attia, et al., Selection criteria for building performance simulation tools: contrasting architects' and engineers' needs, J. Build. Perf. Simulat. 5 (2012) 155–169.
[35] S. Azhar, Building information modeling (BIM): trends, benefits, risks, and challenges for the AEC industry, Leader. Manag. Eng. 11 (3) (2011) 241–252 7.
[36] B. Becerik-Gerber, K. Kensek, Building information modeling in architecture, engineering, and construction: emerging research directions and trends, J. Prof. Barriers Eng. Educ. Pract. 136 (3) (2010) 139–147.
[37] P. Bynum, et al., Building information modeling in support of sustainable design and construction, J. Constr. Eng. Manag. 139 (1) (2012) 24–34.
[38] T. Hong, et al., Advances in research and applications of energy-related occupant behavior in buildings, Energy Build. 116 (2016) 694–702.
[39] B.S. Kim, et al., Data modeling versus simulation modeling in the big data era: case study of a greenhouse control system, Simulation 93 (7) (2017) 579–594.
[40] M. Royapoor, T. Roskilly, Building model calibration using energy and environmental data, Energy Build. 94 (2015) 109–120.
[41] P. Raftery, M. Keane, A. Costa, Calibrating whole building energy models: detailed case study using hourly measured data, Energy Build. 43 (12) (2011) 3666–3679.
[42] Priscillia A. Glasow, Fundamentals of Survey Research Methodology, (2013) Retrieved January 18 (2005).
[43] T. Segaran, T. Hammerbacher, Beautiful Data: the Stories behind Elegant Data Solutions, O'Reilly Media, Inc, 2009.
[44] Colour Blind Awareness. Colour Blindness.
[45] C. Kontokosta, Modeling the energy retrofit decision in commercial office buildings, Energy Build. 131 (2016) 1–20.
[46] R.H. Myers, Classical and Modern Regression with Applications vol. 2, Duxbury Press, Belmont, CA, 1990.
[47] An Introduction to the Bootstrap (Bradley Efron).
[48] A. Carol, Market Models: A Guide to Financial Data Analysis, John Wiley & Sons, 2001.
[49] D. Yan, et al., Occupant behanvior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278.
[50] C. Scheuer, G.A. Keoleian, P. Reppe, Life cycle energy and environmental performance of a new university building: modeling challenges and design implications, Energy Build. 35 (10) (2003) 1049–1064.
[51] Buildsmart NY. Executive Order 88. Better Buildings. Department of Energy. USA.
[52] S.L. Hart, R.E. Quinn, Roles executives play: CEOs, behavioral complexity, and firm performance, Hum. Relat. 46 (5) (1993) 543–574.
[53] C. Juma, How Young Engineers Will Mold the Future. Harvard Kennedy School for Science and International Affairs, Belfer Center, 2013.
[54] America's Energy Future: Technology and Transformation. Chapter 4: Energy Efficiency, National Academies Press, 2009.
[55] A. Nutkiewicz, Z. Yang, R.K. Jain, Data-driven Urban Energy Simulation (DUE-S): integrating machine learning into an urban building energy simulation workflow, Energy Procedia 142 (2017) 2114–2119.
[56] P. Siano, Demand response and smart grids – a survey, Renew. Sustain. Energy Rev. 30 (2014) 461–478.
[57] V.C. Gungor, et al., Smart grid technologies: communication technologies and standards, IEEE Trans. Ind. Inf. 7 (4) (2011) 529–539.
[58] D. Forbes, F.J. Milliken, Cognition and corporate governance: understanding boards of directors as strategic decision-making groups, Acad. Manag. Rev. 24 (3) (1999).
[59] D.W. Shaffer, How Computer Games Help Children Learn, Palgrave Macmillian., New York, USA, 2006.
[60] R. Pacheco, J. Ordóñez, G. Martínez, Energy efficient design of building: a review, Renew. Sustain. Energy Rev. 16 (6) (2012) 3559–3573.
[61] A. Nutkiewicz, Z. Yang, R.K. Jain, Data-driven Urban Energy Simulation (DUE-S): integrating machine learning into an urban building energy simulation workflow, Energy Procedia 142 (2017) 2114–2119.
[62] Y. Chen, T. Hong, M.A. Piette, Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, Appl. Energy 205 (2017) 323–335.
[63] H.C. Howard, R.E. Levitt, et al., Computer Integration: Reducing Fragmentation in AEC Industry vol. 3, (1989) (1).
[64] A. McAfee, E. Brynjolfsson, Big data: the management revolution, Harv. Bus. Rev. 90 (10) (2012) 60–68.
[65] A. Katal, M. Wazid, R.H. Goudar, Big Data: issues, challenges, tools and good practices, Sixth International Conference on Contemporary Computing (IC3), IEEE, Noida, India, 2013.
[66] A. Mushtaq, C.H. Culp, Uncalibrated building energy simulation modeling results, HVAC R Res. 12 (4) (2006) 1141–1155.
[67] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renew. Sustain. Energy Rev. 20 (2013) 411–419.
[68] S.A. Kim, et al., Integrated energy monitoring and visualization system for Smart Green City development: designing a spatial information integrated energy monitoring model in the context of massive data management on a web based platform, Autom. ConStruct. 22 (2012) 51–59.
[69] D.A. Keim, Information visualization and visual data mining, IEEE Trans. Vis. Comput. Graph. 8 (1) (2002) 1–8.
[70] M.O. Ward, G. Grinstein, D.A. Keim, Interactive Data Visualization: Foundations, Techniques, and Applications, AK Peters/CRC Press, 2015.
[71] Z. Yang, B. Becerik-Gerber, A model calibration framework for simultaneous multi-level building energy simulation, Appl. Energy 149 (2015) 415–431.
[72] N. Li, et al., Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? Appl. Energy 159 (2015) 196–205.