Applied Energy 305 (2022) 117829

Contents lists available at ScienceDirect

Applied Energy

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

Status quo and opportunities for building energy prediction in limited data Context—Overview from a competition

Tong Xiao, Peng Xu *, Ruikai He, Huajing Sha

Department of Mechanical and Energy Engineering, Tongji University, Shanghai 201804, China

HIGHLIGHTS

- Discussed cross-building energy prediction in a limited-data context.
- Compared different methods in a same case via a competition.
- Identified hybrid strategies for hybrid building energy prediction models.
- Discussed the shortcomings and suggestions for the data preparation process.
- Highlighted the importance of data selection for accurate cross-building prediction.

ARTICLE INFO

Keywords:
Building energy
Energy prediction
Cross-building prediction
Hybrid model
Data-driven model
Data Preparation

ABSTRACT

With the evolution of new energy and carbon trading systems, it is important to accurately predict building energy consumption to help energy arrangements. Additionally, the widespread use of smart meters has introduced a new data context for building energy prediction. Building energy prediction techniques need improvement but the ideas of various new prediction methods are still on the way and have not yet been compared and tested side-by-side in the reported studies. Thus, we held a competition called 'Energy Detective'. To investigate the status quo of the current prediction techniques, we designed a representative prediction case: cross-building prediction with limited physical parameters and historical data. A total of 195 participants formed 89 teams to participate in the competition. This paper describes the models presented in the competition. By analysing the methods and results, we identified strategies for the future development of energy prediction in hybrid modelling and data-driven modelling. For hybrid modelling, we discuss the basic strategies for hybrid models and suggest that more hybrid models can be developed by combining a wide variety of individual models in sequence or parallel or via feedback methods to achieve accurate and interpretable models. For data-driven modelling, we analyse and discuss the areas of improvement for the current data-driven workflow and suggest that processes other than model application are also important and should be carefully considered. Considering the increasing amount of data available for prediction, we discuss the shortcomings and suggestions for improving the current data preparation process. We recommend comprehensive consideration of the anomaly types in data pre-processing and a focus on feature engineering for higher accuracy and model interpretability, while emphasising the vital role of data selection in cross-building energy prediction.

1. Introduction

1.1. Building energy prediction

Building operations represent the third-largest source of the world's energy consumption and carbon emissions [1]. With the increasing awareness of energy conservation, emission reduction, and ecological development, the accurate prediction of building energy and the reasonable allocation of energy have become important parts of energy-saving methods. Decision-making in the energy system must be based on accurate forecasts of energy demand and consumption [2]. Energy-consumption prediction is an important step in achieving demand-side management of an energy system. Accurate prediction helps decision makers—either people or machines—learn more about the future

* Corresponding author.
E-mail address: xupeng@tongji.edu.cn (P. Xu).

https://doi.org/10.1016/j.apenergy.2021.117829
Received 20 April 2021; Received in revised form 23 August 2021; Accepted 6 September 2021
Available online 21 September 2021

demand of the system and then decide what operations should be performed on the system to reduce waste. Typical operations, such as energy allocation and demand response strategies [3], can significantly improve the flexibility of the energy system.

Early building energy consumption forecasts are used for sizing air-conditioning systems with sufficient capacity; thus, the focus is on peak load demand forecasting, and highly accurate hourly prediction is not important [4,5]. Nowadays, forecasting focuses on the hourly consumption at the whole-building or even the whole-district level, and the results are used to enhance the operation of large-scale energy systems [6]. Moreover, with the development of low-energy-cost building technology and the need for better urban energy arrangements in carbon and energy trading systems, accurate energy consumption prediction is necessary as early as possible once the buildings are put into use.

In recent years, as smart electric meters have become less expensive and have been widely used in buildings, large amounts of historical energy consumption data have been collected and stored [7]. Increasing the amount of building energy data allows more accurate prediction. However, problems remain regarding how to optimally utilise the data. In the energy consumption prediction scheme, a large amount of data has been collected, but an accurate and flexible prediction model has not yet been developed. One new challenge is that more data makes the energy consumption model needlessly complex [8]. Knowledge has not yet been well learned from the data [7,9]. The vital differences in energy consumption among buildings and the necessary amount of data for data-driven models remain unclear. New forecasting methods need to be developed on the basis of the new data.

Meanwhile, data related to the building energy consumption, such as meteorological data, building operational data, and building physical parameters [9], are normally collected from different sources [10] and in different formats. Sometimes, certain parameters may not be available. Meanwhile, as mentioned above, accurate energy consumption prediction is needed once the buildings are put into use. In such task, historical electric meter data of the certain building is always not enough or unavailable. The existing forecasting model does not have sufficient flexibility in such limited data context. It appears that the preparation of the data is almost as important as the model development.

1.2. Research goal

The increasing interest in building energy and the wide usage of electric meters have introduced new opportunities and challenges to building energy prediction. Numerous review studies have been conducted to analyse existing energy consumption prediction methods. However, these methods were developed by different research groups under different project backgrounds, and the comparison of the methods is always based on the calibration score, which may be influenced by the task itself. Thus, these comparisons of the methods are not convincing. Furthermore, the current review studies focus on the prediction methods. As increasing amounts of data become available for prediction, the operations performed on the data may significantly affect the forecasting accuracy. However, there is still a lack of analysis of the operation of data.

Meanwhile, with the changes in the objectives and scenarios of energy consumption prediction, methods are being developed for new prediction cases. Energy prediction in limited data context is a kind of cases that widely be focused on in the recent studies. A typical case is to predict the energy consumption of a building without enough historical data of its own. Development ideas for prediction have been tested by researchers in the past few years but have not been summarised and discussed.

In view of the foregoing background, we completed a competition called ‘Energy Detective’ to analyse the development ideas and the status quo of energy consumption prediction methods in 2020. Reports from various teams have been used to analyse the methods and promote new ideas.

In contrast to a similar competition held previously that was based on existing cases with sufficient data [11,12,13], the prediction case given in our competition was designed with a limited data context, where the historical data of the target building were not provided. Participants were required to predict the hourly energy consumption of the target building within one year in three limited submissions. A dataset of the 3-year hourly energy consumption for 20 reference buildings of the same type was provided, as well as a physical description of the target buildings. Participants could not easily use a single existing prediction method to achieve good prediction accuracy.

We received 140 submissions from 56 teams associated with research institutions and universities worldwide. These submissions provide a picture of the current state of the prediction methods. Creative ideas have been proposed and tested. Because the limited data context of the given case is new for the existing prediction method, the accuracy difference among the submissions is not as important as the method used.

Although the analysis presented herein is based on one case, which is specially designed with a significant scene in current data context, the methods and ideas can aid technique development in wider scenes. If the method is capable of accurate cross-building prediction in a limited data context, it may be able to utilise the data for prediction in a more complete data context. Thus, we attempt to provide insight into the possible future development of the current techniques so that the special operation, which is only responsible for this case, will not be discussed.

In light of the above aims and the analysis that follows, we highlight the main contributions of this paper:

- Briefly review and discuss the ideas for cross-building prediction method development in a limited-data context.
- Compare different methods and their performance directly in a same case via a competition.
- Propose a generic hybrid strategy for hybrid models, which has been lacking in the use of hybrid models in previous studies.
- Discuss the methods used in data preparation process and their shortcomings in detail, which have received less attention in previous studies. And briefly demonstrate that the data preparation process has a status no less important than that of model application in data-driven workflows.
- Discuss and highlight the importance of data selection in cross-building prediction for the first time.

Meanwhile, we also highlight that the competition we held and the problem we provided are designed with a significant and unsolved prediction case in a limited data context. Thus, we could learn clearly about the current gap of the prediction method. We discuss the gap in both that of hybrid models and that of data-driven models and conclude the opportunities for the further studies. We believe that our studies will provide useful suggestions for further prediction method development.

The present paper presents a summary of these competitions. The remainder of the paper is organised as follows. Section 2 briefly reviews existing building energy consumption prediction methods and their development ideas in the new data context. Section 3 provides a brief introduction of Energy Detective, i.e., the competition that we held. The introduction includes a description of the proposed case, an overview of the participants, an overview of the results, and an overview of the methods used. Section 4 focuses on the prediction framework based on hybrid models, and Section 5 focuses on the methods used in a data-driven framework—particularly in the data preparation process. Discussions regarding the status quo and the potential of the methods are presented in Sections 4 and 5, and Section 6 concludes the paper.

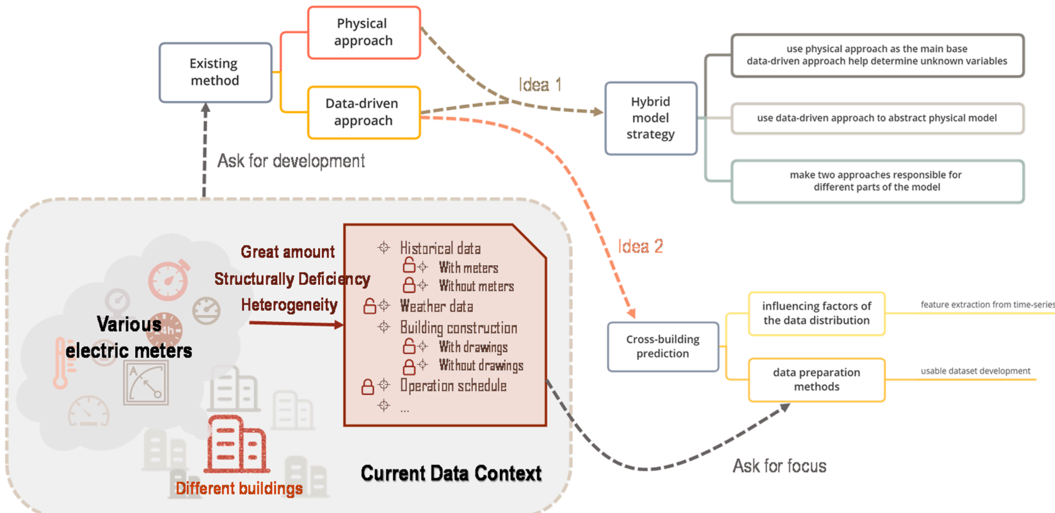

Fig. 1. Improvement ideas for building energy consumption prediction methods.

2. Literature review

2.1. Existing building energy prediction methods

Generally, existing methods for building energy consumption prediction can be divided into two approaches: the physical modelling approach and the data-driven approach [14].

Physical models, which are also called white-box models or forward models, mainly rely on detailed modelling or analysis of the thermodynamic rules of the building energy system. Building energy simulation software programs such as DOE-2 [15], EnergyPlus [16], e-QUEST [17], TRNSYS [18], ESP-r [19], and DeST [20] are widely used in physical modelling [21,22,23]. To construct a physical model, a large amount of information about the building must be collected, such as the building construction shape, thermal properties of the building envelope, operation schedules, and information on the equipped heating, ventilation, and air conditioning (HVAC) system [5,24,25]. With so many input details needed, the energy consumption model is highly explainable but often inaccurate because accurate input information, such as the usage pattern and the numerous building parameters, is difficult to obtain [26,27,28]. Moreover, physical models are always detailed and are based on detailed dynamics or static formulations. These models are always time-consuming to develop and solve [29].

Data-driven building energy consumption forecast models, which are also called black-box models, include statistical approaches and intelligent approaches [22]. Statistical approaches are mainly based on linear or multivariate regression methods, such as the autoregressive integrated moving average (ARIMA) [22,30,31]. Methods based on machine-learning algorithms are called intelligent approaches. Machine-learning algorithms used in energy prediction tasks include artificial neural networks and support vector machines [11,22,30,31]. In recent years, decision-tree algorithms [11], e.g. classification and regression tree (CART), random forest (RF), and boosting tree (BT), and deep-learning algorithms [32], e.g. long short-term memory (LSTM) networks, have been widely used in modellings. Data-driven models focus on the data relationship between inputs and outputs instead of the physical principle; thus, they are less explainable than physical models. These models are trained with historical/available energy consumption data and some necessary influence parameters [11]. Accurate model predictions can be achieved in some cases. However, building energy systems are highly physics-based systems, without sufficient explanation, and the prediction results may sometimes be unreliable [26]. Data-driven models are highly based on historical data. However, the historical data are sometimes insufficient, for example, in the initial stage of operation or before the building has been equipped with electric sensors [27]. Without sufficient historical data, the prediction accuracy of data-driven models is reduced [33,34].

Physical models and data-driven models with a good capacity for many energy consumption prediction tasks have been developed. However, in some cases, prediction tasks in the real world are complex, and the information needed for the developed models is unavailable. For example, building owners are eager to know the future energy consumption of a newly constructed building, but new buildings lack historical datasets to train blackbox models [34]. Models must evolve for new prediction scenarios.

2.2. Ideas for improving building energy prediction methods

There are two main ideas for improving the energy prediction method, as shown in Fig. 1. One is to develop a hybrid model to complement physical and data-driven approaches. The other is to improve the capability of the data-driven model so that it can be used in a wider range of scenarios, such as cross-building prediction.

2.2.1. Hybrid model

One idea is the hybrid model. Hybrid models (also called grey-box models) are based on a combination of the two aforementioned approaches. Hybrid models work under a limited dataset, and the input data remain physically interpreted [26]. Researchers believe that they can provide more accurate predictions than the individual models in some cases, for example, when the detailed physical model is not entirely known, the data are insufficient for a single data-driven model, or the physical model is too complex to solve [35]. There are three basic strategies for the hybrid models [26].

(1) The model is mainly based on a physical approach and uses a data-driven approach to determine the unknown variables. Reynders et al. [36] built a grey-box model whose structure relies on physical knowledge about the system dynamics and then estimated the unknown parameters using a statistical method in day-ahead predictions and simulations of the thermal response of a dwelling.
(2) A data-driven approach is used to abstract physical models to reduce the complexity of the physical approach. Afshari et al. [35] applied the strategy by identifying an average grey-box model, inverting the measured hourly system load and weather data of a physical model in a statistical way. Elbeltagi et al. [37] abstracted the simulation model into a multi-parameter statistical

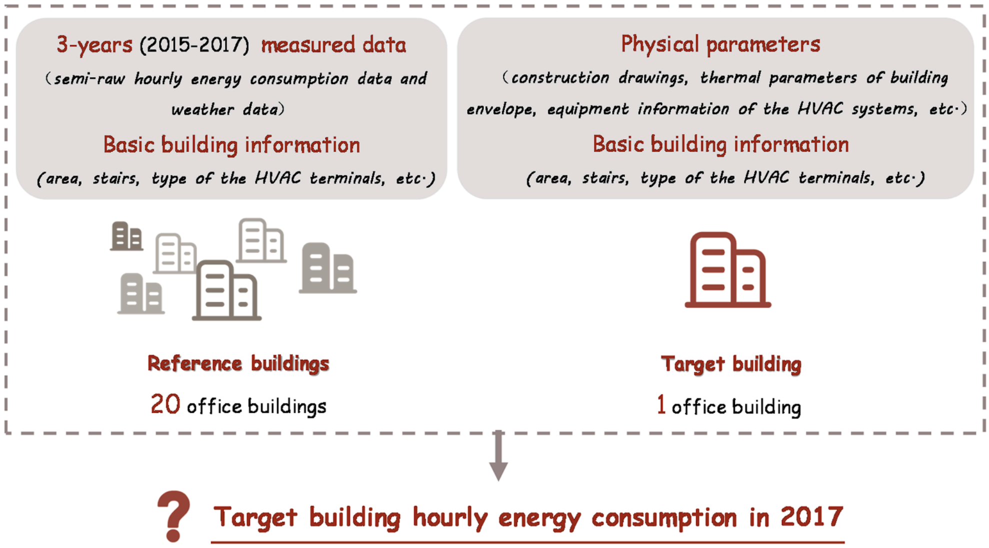

Fig. 2. Prediction case of the competition.

model so that designers could determine the energy consumption at early design stages even without experience using simulation tools.

(3) The two approaches are responsible for different parts of the model. This is the most flexible strategy for hybrid models. Fumo et al. [38] proposed a simple approach to enrich historical data using an EnergyPlus benchmark model. Vaghefi et al. [39] combined a data-driven energy forecasting model, which was based on simulation data created by EnergyPlus, and a heating/cooling physical model to construct an energy forecasting model for the optimal controller of a building.

Although there are strategies for hybrid models, models capable of accurate and widely usable energy consumption forecasting have not yet been proposed. Reliable hybrid modelling approaches are still under development.

2.2.2. Data-driven model improvement

Improving the data-driven model is also a feasible approach for developing the models. Currently, data-driven models work well for one specified building. The input of historical data to the data-driven model can ensure to the maximum extent possible that the training data (the input data) and the test data (the prediction result) are drawn from the same feature space and the same distribution, which is needed for effective data-driven methods [40]. To make a data-driven model applicable to more scenes, the capability for cross-building forecasting is required. Some researchers believe that buildings of roughly similar sizes and uses will have similar energy consumption distributions [38], which is also a basic concept for building benchmarking. Although building consumption data are gathered in diverse contexts [34], which may not be within the same distribution, it may be possible to improve the scalability of the data-driven model by constructing a model with a group of buildings having similar distributions. Similar work in other fields, such as the medical domain, relies on extracting features from the dataset [41].

In previous studies, features that influence building energy consumption underlying building physics and experiences were often selected [42]. The chosen features have been described in physical models and are usually the building construction, equipment, and operation schedule. Researchers believe that such features diversify the distribution of the energy consumption. If they can be added to the data-driven model, the model may be able to perform cross-building forecasting. Such features are not always available directly, but they may be mined from the energy consumption time series. Researchers have performed work on knowledge discovery of energy consumption data using data-mining techniques [9,43]. Xiao et al. [7] employed clustering and association rule mining techniques to determine the typical operation modes among buildings. Chen et al. [44] and Qiu et al. [45] mined an HVAC system control strategy hidden in energy consumption data.

Moreover, some researchers believe that the current building energy description features are insufficient. To better describe the energy consumption distribution differences among buildings, more physics-based features need to be developed. Pacheco-Torres et al. [46] considered a deeper sub-categorisation of activities within buildings after their analysis of different distributions among university buildings. Additionally, temporal features, which describe the statistical difference among series and cannot be connected with single existing physical factors such as trend and seasonality [47], make sense in building energy description [41]. Although considerable work has been done, the current work for feature extraction remains insufficient, and there is a lack of analysis on how the features impact the forecasting model.

Moreover, a usable dataset is an important part of this study. With the increase in the amount of available building energy data, large-scale empirical building energy databases have become available in the past few years. The DOE-funded Building Performance Database (BPD) [48] contains monthly building energy consumption data from > 740,000 buildings in the USA. Because there are no hourly data in BPD, it is useful for medium-term energy allocation but not for enriching the hourly forecasting model. The Building Data Genome Project (BDG) [49] and Building Data Genome Project 2 (BDG2) [50] provide hourly non-residential meter data of 507 buildings and 1636 buildings. These two datasets were developed with a meter-rich background in recent years. The hourly data and metadata make them useful for wide-energy analysis scenes. Several techniques for data pre-processing are needed for dataset development to ensure data quality. Miller et al. [51] reported the techniques they used in BDG2 development, which included normalisation and anomaly detection. Fan et al. [52] reviewed the data pre-processing techniques using in current building operational data analysis. However, as increasing amounts of raw data on building energy consumption become available, there is still not enough research focusing on data preparation methods that can handle a variety of raw data. Researchers should pay more attention to the data preparation methods.

Table 1
Data provided in the competition.

Data Type	File Name	File Variables and Short Descriptions
Physical parameters and basic building information of the target building	test_building.zip	construction drawings of the target buildingincluding: • 7 typical floor plans in jpg format • Brief introduction of the floor plans
	test_building_info.docx	Basic building information and some physical parameters of the target buildingIncluding: • Area (m^2) • Stairs (numbers aboveground and underground) • Type of HVAC terminals • Thermal parameters of building envelope and windows (basement, podium, and tower) • Description of some energy-saving measures on building envelope • Equipment information of the HVAC system (numbers and capacities of chillers and pumps)
Three-year measured data	weather.zip	Hourly weather data from 2015 to 2017 in epw format and xlsx formatIncluding variables: • Time: timestamp • Temperature (°C) • Dew point temperature (°C) • Relative humidity (%) • Atmospheric pressure (Pa) • Wind speed (m/s)
	train.csv	Hourly energy consumption data of the reference buildings from 2015 to 2017 in csv formatIncluding variables: • Time: timestamp • BuildingID: serial number of the reference building • Type: meter type ('Q' for light and plug and 'W' for HVAC system) • Record: hourly electric consumption (kWh)
Basic building information of the reference buildings	train_building_info.xlsx	Basic building information of the reference buildingIncluding: • BuildingID: serial number of the reference building • Stair1: number of stairs up ground • Stair2: number of stairs underground • Area (m^2) • HVACType: Type of HVAC terminals

3. Energy Detective building energy forecasting competition

3.1. Prediction case

As mentioned above, prediction case in limited data context that available physical parameters are limited and historical data are insufficient or unavailable is common in current studies. However, existing methods cannot provide accurate predictions in such situations. Thus, the prediction case provided in the competition is based on this representative situation.

Participants were asked to predict the hourly energy consumption of two meter types during 2017 for a target building located in Shanghai, China. The energy consumption cost of lights and plugs was one of the meter types (marked as 'Q'), and that of the HVAC system was the other (marked as 'W'). In the case presented to the participants, hourly historical data of two meter types from 2015 to 2017 for 20 reference buildings and some physical parameters of the target building were given. The reference buildings were located at the same location as the target building and were of the same usage type. Hourly weather data from 2015 to 2017 and basic information of the 21 aforementioned buildings were also provided. The prediction case is shown in Fig. 2. A detailed description of the data is presented in Table 1.

The actual records of the hourly historical data of the two meter types in 2017 for the target building were used to evaluate the participants' results. The basic evaluation metric selected was the coefficient of variation of the root-mean-square error (CV-RMSE). It is an energy measurement statistical model that is widely used for energy consumption prediction. Global organisations such as ASHRAE, IPMVP, and FEMP have set their own standards for the baseline model [53,54,55]. It is intuitive to see how accurately the participants completed their predictions.

The CV-RMSE is calculated using Eq. (1).

$$\mathrm{CV-RMSE} = \sqrt{\frac{\sum_{k=1}^{n}(y_k - \widehat{y_k})^2}{n}} \bigg/ \frac{\sum_{k=1}^{n} y_k}{n} \tag{1}$$

Here,

- y_k represents an actual record,
- $\widehat{y_k}$ represents a prediction result for the record, and
- n represents the total number of records.

Because there were two meter types in this case, we calculated the CV-RMSE score for each meter type and then combined them with weights. These weights were chosen because the change of the 'W' meter type among a year of the target building and that of the 'Q' meter type is about 7:3. The total evaluation metric was calculated using Eq. (2).

$$\varepsilon = \mathrm{CV-RMSE(Q)} \times 0.3 + \mathrm{CV-RMSE(W)} \times 0.7 \tag{2}$$

Here,

- ε represents the evaluation score;
- $\mathrm{CV-RMSE(Q)}$ represents the $\mathrm{CV-RMSE}$ score of meter 'Q', and $\mathrm{CV-RMSE(W)}$ represents that of meter 'W'.

The hourly historical data used in the competition of the two meter types are gathered by raw metadata, which come from a private dataset built by an energy management company. Abnormal metadata, in which the standard deviation during one year is $> 10{,}000$, are deleted during gathering. Records in meter type 'Q' are summed by the metadata of the electric meters' measuring lights and plug loads. Records in meter type 'W' are contributed by the measured electric load for heating/cooling sources, fans, pumps, and other related equipment in the HVAC system. It is stated that some of the records in meter type 'Q' of the reference buildings may include the electricity cost of the HVAC terminals, owing to possible electric meter misconnection of the HVAC terminals occurring in reality. However, such situations did not fit the records of the target building.

3.2. Overview of participants

The competition involved communication over the Internet. The signup was available online on 3 May 2020, and the prediction case was made public on 6 May 2020. The related files were open to participants via email attachments and could also be downloaded at Baidu to file addresses. Because the online publicity of the competition was mostly in Chinese, the participants were mainly people who could read Chinese

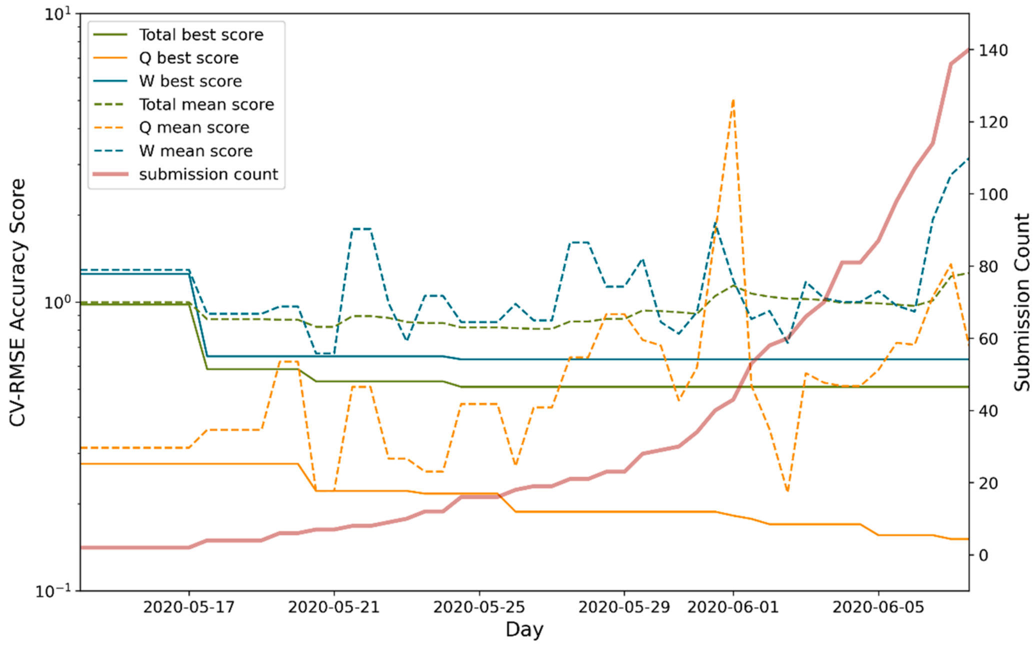

Fig. 3. Changes in best and mean evaluation scores during the competition.

but from different education backgrounds.

Participants were allowed to sign up with a team (with a limit of three people per team) in the competition. There were 195 participants from six countries comprising 89 teams in the competition. Most of the participants worked or studied in China, and 10 of them came from the United States, Germany, Singapore, Australia, and the United Kingdom.

3.3. Overview of prediction results

The competition lasted one month—from 6 May 2020 to 6 June 2020 (the case was made available on 6 May 2020, and 6 June 2020 was the final submission date). Each team could submit up to three times during the competition. The submission limit was based on the assumption that the accuracy of a method should not be totally based on parameter adjustment. Because the prediction case involved a building without its historical consumption data, we preferred a method that could achieve good accuracy in a limited number of attempts.

At the end of the competition, there were 140 submissions from 56 teams. We calculated the evaluation score for each submission and informed the participants the score via e-mail once they submitted. The best and mean evaluation score changes are shown in Fig. 3. Over time, some teams found a good way to improve their models so that they achieved a better score. However, the change in the mean scores indicated that the prediction case certainly presented challenges for some of the current methods. Methods may perform poorly in such prediction cases, and some adjustments of the methods may not make sense.

Fig. 4 shows the monthly evaluation score distribution for all the submissions. To view the distribution more clearly, we performed logarithmic operations on all the scores to scatter the distribution. Thus, the lower scores indicate the better accuracy and the 0 on the x-axis equals to the 1 in CV-RMSE scoring. As shown in (a), which presents the distribution of the total evaluation score, the distribution of the accuracy behaved differently among months. The prediction of the energy consumption of lights and plugs achieved similar accuracy among months but a scattered distribution for each month, as shown in (c), which presents the distribution for the CV-RMSE score of the ‘Q’ meter type. Most of the submissions achieved sufficient accuracy because the light and plug loads were closely related to the building area and workday and changed little over the months. The key difference over the months was caused by the energy prediction for the HVAC system, as shown in (b), which presents the distribution of the CV-RMSE score of the ‘W’ meter type.

In this case, the participants’ predictions were more accurate during the cooling season (summer) than other seasons. A bimodal distribution appeared in June, July, and August, as shown in (a), indicating that the different methods exhibited different behaviours for summer prediction. Compared with the summer and transition seasons, the evaluation scores in winter had a more scattered distribution and lower accuracy. This may have been due to the large difference in the HVAC load demand, which was due to different building construction and users’ habits, and the heating source, which may be a heat pump or boiler but was not given in this case. Such differences generally occurred among office buildings in Shanghai and maybe in other cities with similar climates.

As mentioned previously, we focused on the methods and ideas that can provide insight into the possible future development of the current techniques—not special operations applicable only to the present case. According to the monthly evaluation score distribution, we conclude that it is more meaningful to discuss the accuracy in the summer and transition seasons. The participants mainly differed in their predictions for the ‘W’ meter type, which represents the energy consumption of the HVAC system. Thus, we focused on the methods that predict HVAC systems.

It should also be stated out that accurate prediction in a limited-data context remains a difficult task in previous studies, for example, Nutkiewicz achieved an hourly CV-RMSE of 46.0% and a monthly CV-RMSE of 27.9% in a limited-data context in 2018 [56]. Although the hourly CV-RMSE scores achieved by the participants, with a best value of 63.37%, do not indicate that they completed accurate hourly forecasts, the monthly CV-RMSE scores they achieved, with a best value of 11.88%, somehow indicate that the monthly forecasts were accurate. We discuss the methods used in the competition not only because they were creative, but also because they achieved a more accurate monthly prediction in a limited-data context.

3.4. Overview of prediction framework

After the submission deadline on 6 June 2020, we asked the 31 top-ranked teams to provide descriptions of their prediction frameworks so that we could learn more about what the participants did. Two ideas for

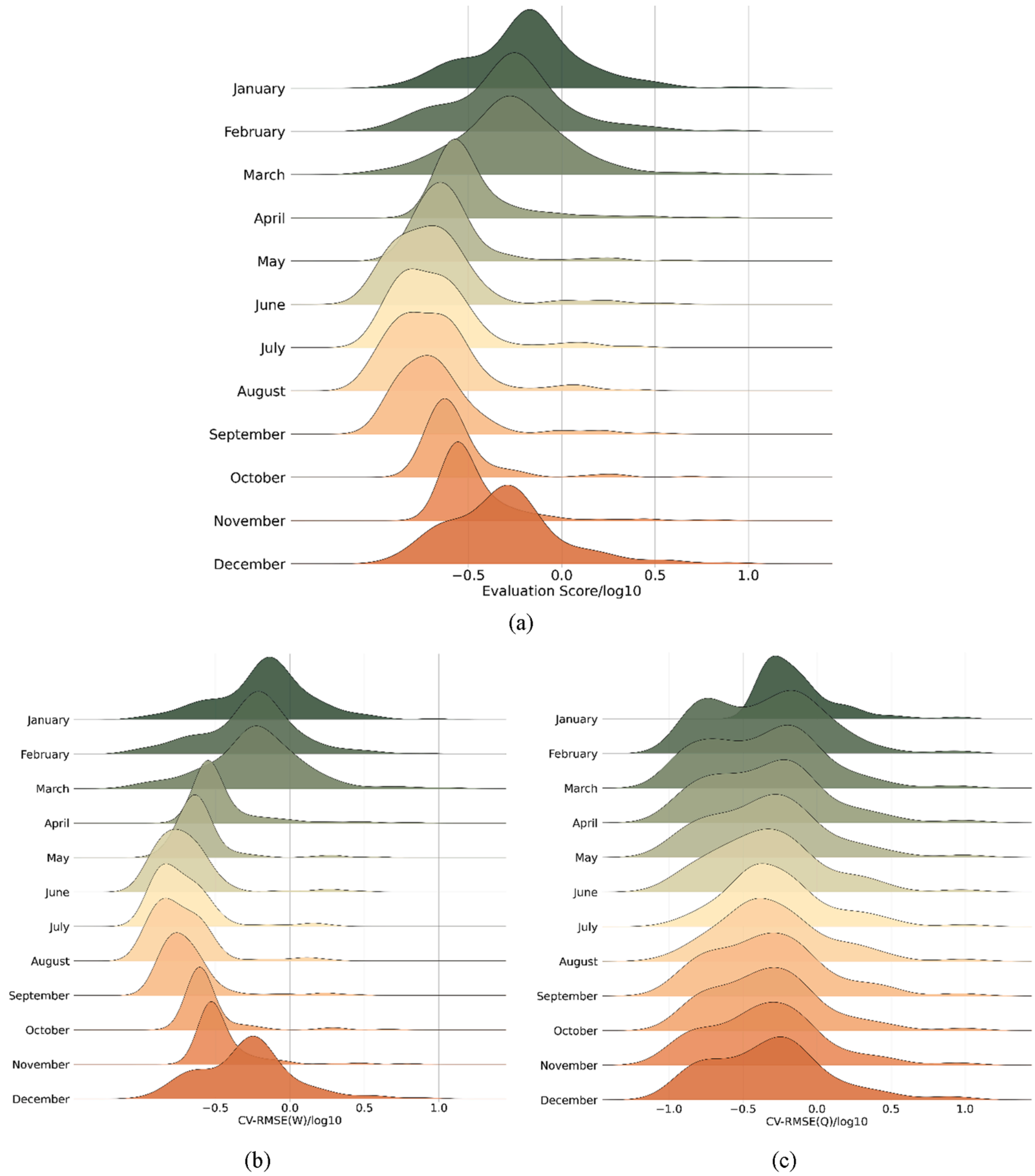

Fig. 4. Monthly evaluation score distribution of all submissions.

the improvement of prediction in a new data context, as mentioned in Section 2.2, were tested by the participants.

It appears that pure physical simulations were not trusted by the participants, as no team selected a physical model for prediction. All the teams used the historical data given. The methods used in the competition are shown in Fig. 5. Five of the 31 teams selected a grey-box model to build their prediction framework, and the other 26 used a data-driven method as the base of their framework. Grey-box model users applied creative ideas to build their prediction frameworks. Some of the participants using data-driven models modified the methods. Because the case is different from the cases data-driven methods deal with before, that is, predicting a building without the historical data, as mentioned previously, methods based on basic statistical operations are used frequently, unlike that in the related competition [11].

4. Prediction framework based on hybrid model

4.1. Strategy for hybrid model

As stated in Section 2.2.1, there are several basic strategies for hybrid models. All hybrid models used in the competition are based on the most flexible strategy, which involves making two approaches responsible for different parts of the model. Several variants of this strategy were used in the competition. They can be classified into three types: sequence, parallel, and feedback.

Fig. 6 shows a simple sketch of the three variants. Here, a circle labelled A or B represents the single method, which is physical or data-driven. The order of the methods can be changed. The sequence strategy involves solving the two models in order. The solution of the lower-ranked model must depend on the solution of the previous model. Only the lower-ranked model was directly responsible for the final result. For example, in Fig. 6(a), model B cannot be solved before model A gives a result, and the final result will be certained once model B is solved. In the parallel strategy, the solving order of the two models is not

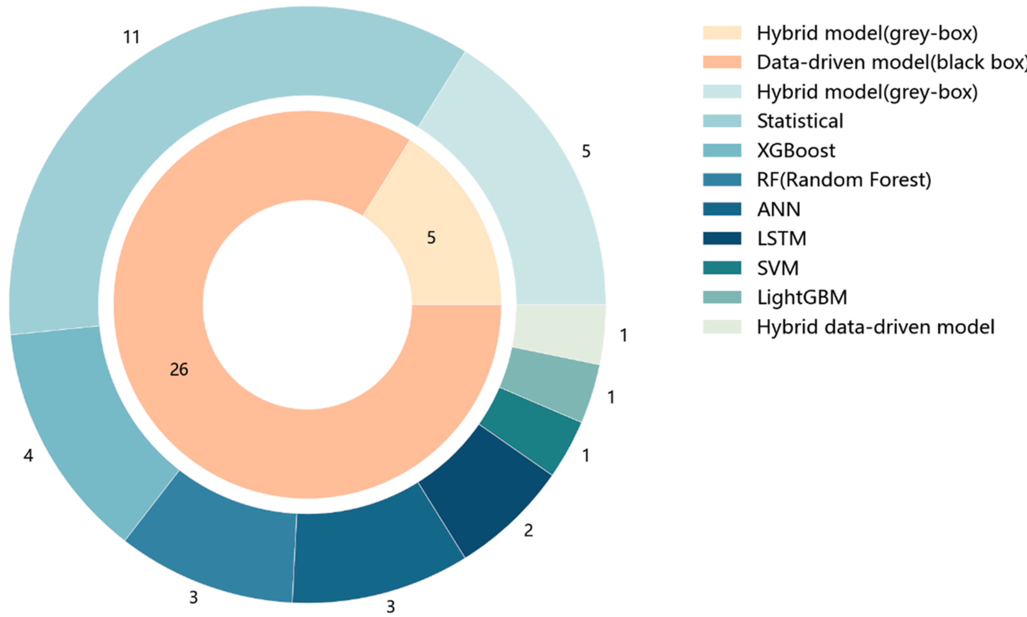

Fig. 5. Prediction methods used in the competition.

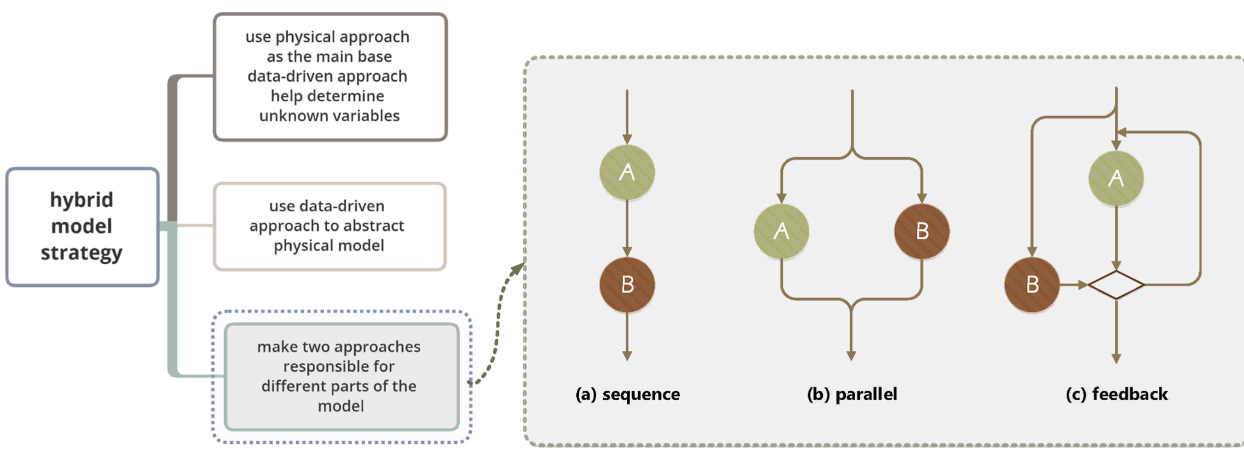

Fig. 6. Three strategy variants for the hybrid (grey-box) model (circles labelled A or B in the figure represent single approaches, physical approaches, or data-driven approaches).

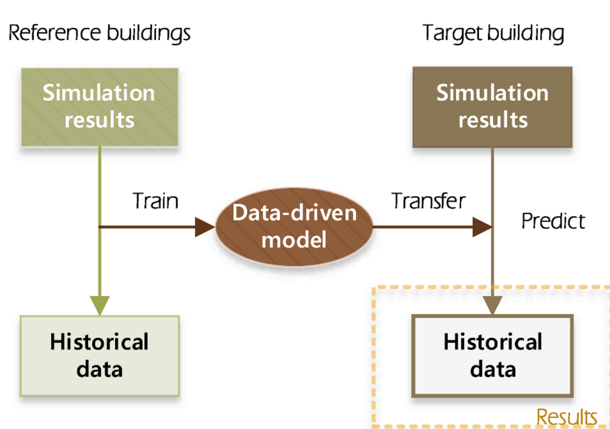

Fig. 7. Role of the data-driven model in the hybrid model used by Team 2.

important. Both models were directly responsible for the final result. While the feedback strategy does not have a strict order for the model solution, one model plays the main role (model A in Fig. 6(c)), and the other helps to modify it (model B in Fig. 6(c)). Both are directly responsible for the final result.

4.2. Overview of hybrid model framework

Hybrid models with sequence strategies are the most widely used in the competition. They can be concluded by using the simulation result as the input of the data-driven model in the competition.

Among the three teams that used the sequence strategy, Team 2 raised the most creative ideas. They first performed a physical simulation with urban building energy modelling (UBEM) and then used a data-driven model to improve the accuracy. The UBEM does not require manual physical modelling of the building structure, which significantly reduces the time required for physical model development, making this method attractive, particularly for district energy modelling [57]. Related research combining UBEM and data-driven models was conducted by Nutkiewicz et al. [56] in 2018. They proposed a framework called data-driven urban energy simulation (DUE-S), which combines a

Table 2
Overview of the hybrid models used in the competition.

Team No.	Monthly CV-RMSE(W) (from April to November)	Hourly CV-RMSE(W) (from April to November)	Physical model base	Data-driven model base	Hybrid strategy	Strategy description
2	0.4561	0.8920	Modelica UBEM	LigntGBM	sequence	Use simulation results as the input of the data-driven model
8	0.2337	0.6901	Design Builder	Periodic decomposition and superposition	parallel	Use a physical model to obtain energy consumption index and use a data-driven model to obtain unit energy consumption; then, multiply them
23	0.2951	0.7804	EnergyPlus	Basic statistical operation	feedback	Use simulation results for evaluation
34	0.1286	0.7383	EnergyPlus vba	BPNN	sequence	Use simulation results as the input of the data-driven model
75	0.2690	1.0297	EnergyPlus	Hephaestus	sequence	Use simulation results as the input of the data-driven model

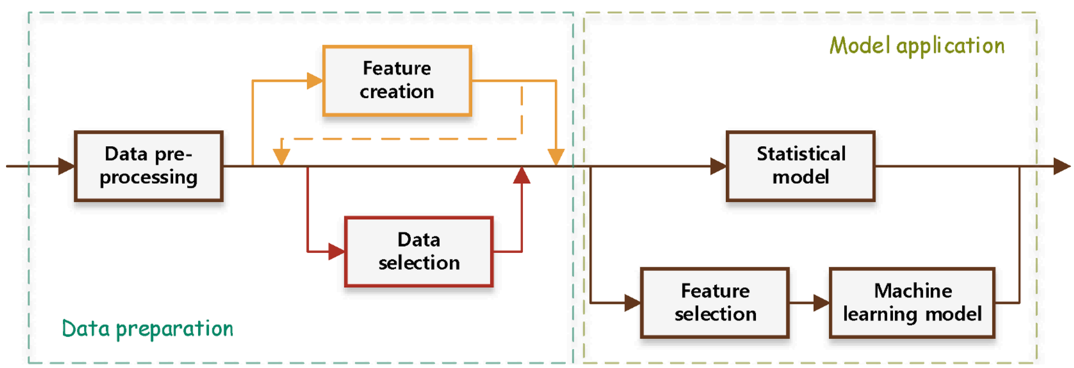

Fig. 8. Data-driven workflow used in the competition.

data-driven model ResNet with UBEM. The case in the competition differs from the circumstance that DUE-S is designed for, because the detailed location of the buildings and the historical data of the target building are unavailable. Team 2 used AixLib to construct building models in Modelica [58,59] and then used the LightGBM algorithm to learn the difference between the simulation results and the historical data. They selected the input parameters of the physical model with key variables determined via building clustering [60]. With insufficient knowledge about the historical energy consumption of the target building, they applied the model training using reference buildings to the target building to modify the simulation results of the target building, as shown in Fig. 7.

Team 75 also used a hybrid model with a sequential strategy. In their work, an existing data-driven model for cross-building prediction called Hephaestus [34] was tested. Hephaestus, which was proposed by Ribeiro et al. [34] in 2018, is a cross-building prediction method based on transfer learning. In contrast to other transfer learning methods, Hephaestus considers seasonal and trend adjustments, which are based on the time-series regression model [47], in the transfer learning framework. With seasonal and trend adjustment, Hephaestus can use limited historical data of the target building to predict its future consumption via transfer learning from other historical data. Hephaestus is designed for cases where there is insufficient historical data of the target building, which differs from the case of the competition. Team 75 employed simulation results calculated using EnergyPlus to replace the historical data of the target building in Hephaestus.

A hybrid model with a parallel strategy was used by Team 8. They considered that different buildings of the same use type and location will have similar energy consumption trends in the same year, and the differences arose from multiplication by indices that differed among the buildings. Therefore, they used a data-driven approach to determine the unit energy consumption trend and a physical approach to determine the energy consumption index. With the concern that building energy consumption has strong intra-daily and seasonal cycles [61], they decomposed the daily and weekly periodicity from the historical data of the reference buildings and then performed superposition to obtain the unit energy consumption trend. Then, they performed a simulation using EnergyPlus to determine the energy consumption index of the target building. The final results were calculated by multiplying the energy consumption index by the unit energy consumption trend.

Team 23 constructed a hybrid model using a feedback strategy. They employed simulation results calculated using EnergyPlus as the evaluation data to evaluate the data-driven model.

Table 2 presents a brief overview of the hybrid models used in the competition, including the models that are not mentioned above. The frameworks based on the hybrid model did not achieve good accuracy in the competition because the hybrid model attempt is still at the early stage. However, researchers believe that hybrid models which combine the advantages both of the physical model and the data-driven model may achieve better prediction accuracy than single model recently. And many hybrid model attempts are on the way. In our competition, some ideas of the hybrid models were tested, providing insights for hybrid-model development.

4.3. Opportunity on hybrid model framework

As mentioned in Section 2.2.1, hybrid models are considered to be effective for combining the advantages of data-driven and physical approaches while avoiding the disadvantages. The strategies for hybrid models proposed in previous research may not be sufficiently detailed for model development, and models that are capable of accurate and widely usable energy consumption forecasting have not yet been established. Although the models used in the competition did not achieve excellent results, they offer strategies and new ideas that may provide guidance for future hybrid-model development.

As shown in Fig. 6, the strategies for the hybrid models are enriched

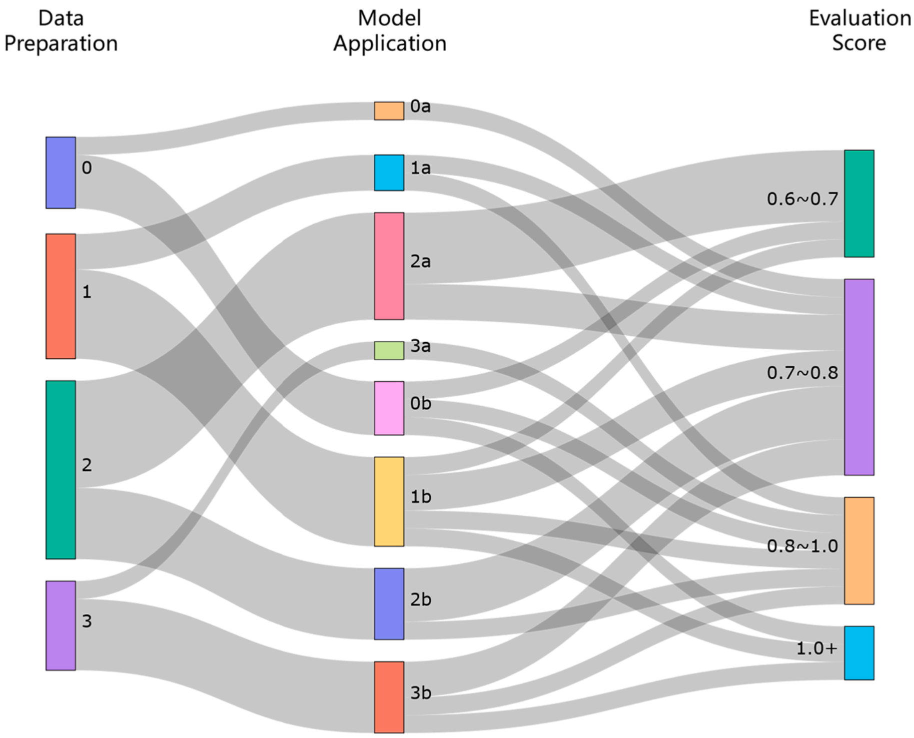

Fig. 9. Data-driven model workflows used by the teams and the best accuracies they achieved (The evaluation score indicates the hourly CV-RMSE(W) (from April to November); the labels in the figure indicate the different method types: '0′ indicates data prepared only with data pre-processing; '1′ indicates data prepared with data pre-processing, feature creation, and data selection; '2′ indicates data prepared with data pre-processing and data selection; '3′ indicates data prepared with data pre-processing and feature creation; 'a' indicates the use of a statistical model, and 'b' indicates the use of a machine-learning model).

Table 3
Overview of the data-driven workflow used by the top five teams for W-meter-type prediction.

Team No.	Monthly CV-RMSE(W) (from April to November)	Hourly CV-RMSE (W) (from April to November)	Data pre-processing	Feature creation	Data selection	Model chosen and application
18	0.1420	0.6337	Delete the point anomalies (global outliers) and collective anomalies (stable values)	Create day type features; create three temporal features from time series (detailed discussion in Section 5.3.2)	Select the most similar buildings using three temporal features created	XGBoost(with feature chosen, parameter adjustment and 5-fold cross-validation)
65	0.1188	0.6467	Delete the point anomalies (global outliers)	–	Select the building with the same type of HVAC system as the target one and a similar number of stairs	Statistical method (area and stair-related factor multiplied by the energy intensity calculated from the dataset)
29	0.1774	0.6491	Delete the point anomalies (global outliers)	–	Delete the building with bad data quality	Statistical method (area multiplied by the average energy intensity at each moment in the dataset)
72	0.1830	0.6493	Delete the point anomalies (global outliers)	–	Delete the building that behaves differently among the reference buildings	Statistical method (area multiplied by the average energy intensity at each moment in the dataset)
6	0.2293	0.6568	Delete the point anomalies (global outliers) and collective anomalies (stable values)	–	Select the data at each moment based on the nearest neighbourhood searching	Statistical method (area multiplied by the energy intensity calculated from the dataset)

in competition. The data-driven approach and physical approach can be combined in sequence, parallel, or feedback. The choice of individual models for the hybrid model can vary. The physical approach used can be single detailed building physical modelling or urban building energy modelling. The proposed data-driven approach can be applied to hybrid-model development, e.g. for the transfer learning prediction model, Hephaestus. More hybrid-model attempts can be performed using the strategies obtained from the competition to construct a better hybrid model.

5. Prediction framework based on data-driven model

5.1. Data-driven workflow

The data-driven workflow used by the participants is shown in Fig. 8. The workflow can be divided into two processes: data preparation and model application. Each process has several variants.

In the data preparation process, after data pre-processing, some of the teams selected the proper data for modelling with the concern that not all the data would improve the model prediction accuracy. Some features were also created from the energy consumption time series as

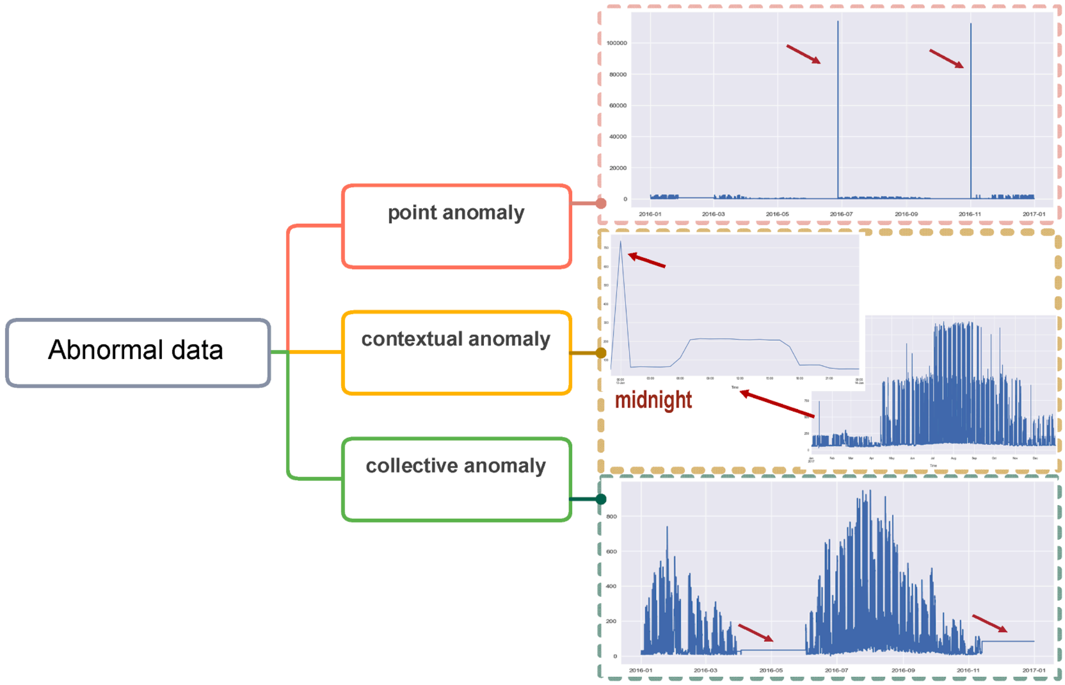

Fig. 10. Incorrect data type in energy consumption data.

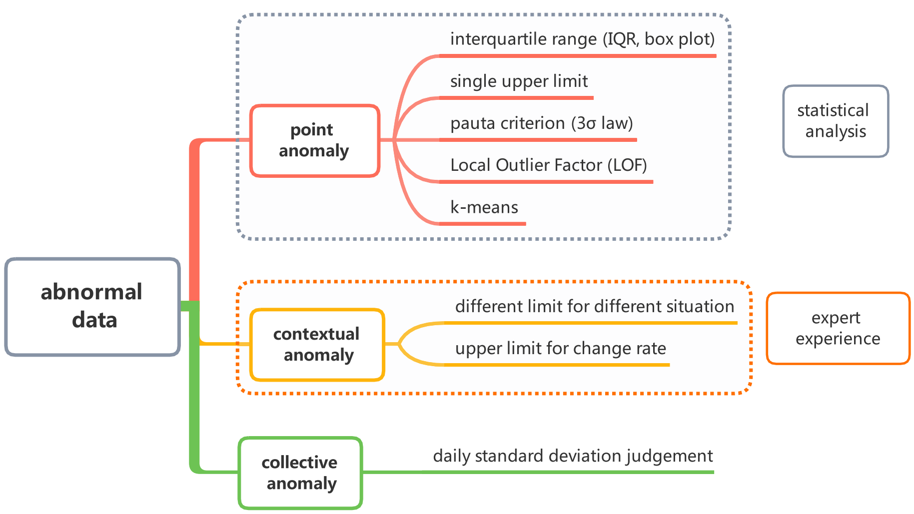

Fig. 11. Methods used to recognise abnormal data in the competition.

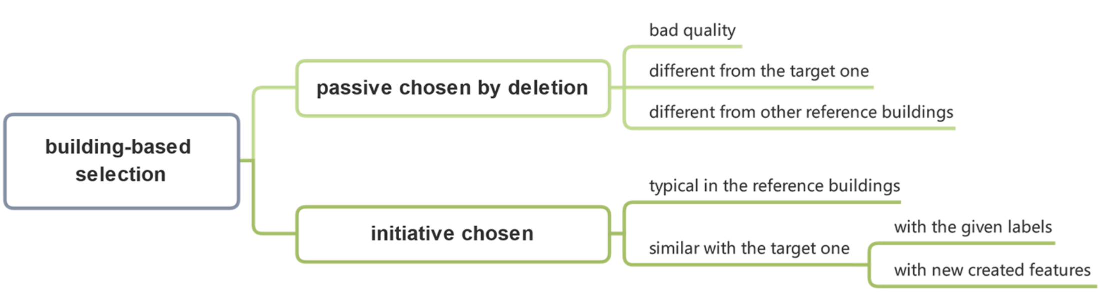

Fig. 12. Subdivision of the building-based data selection strategy.

the input of the model or as the basis for model selection.

Two types of models were used in the model application process: statistical models and machine-learning models. As mentioned in Sections 2.4, statistical methods are used often in this case, such as seasonal-trend decomposition [62,63].

Fig. 9 presents a summary of the data-driven workflow used by the teams and the best accuracy achieved during the competition. The labels in the figure indicate the different method types, which are listed

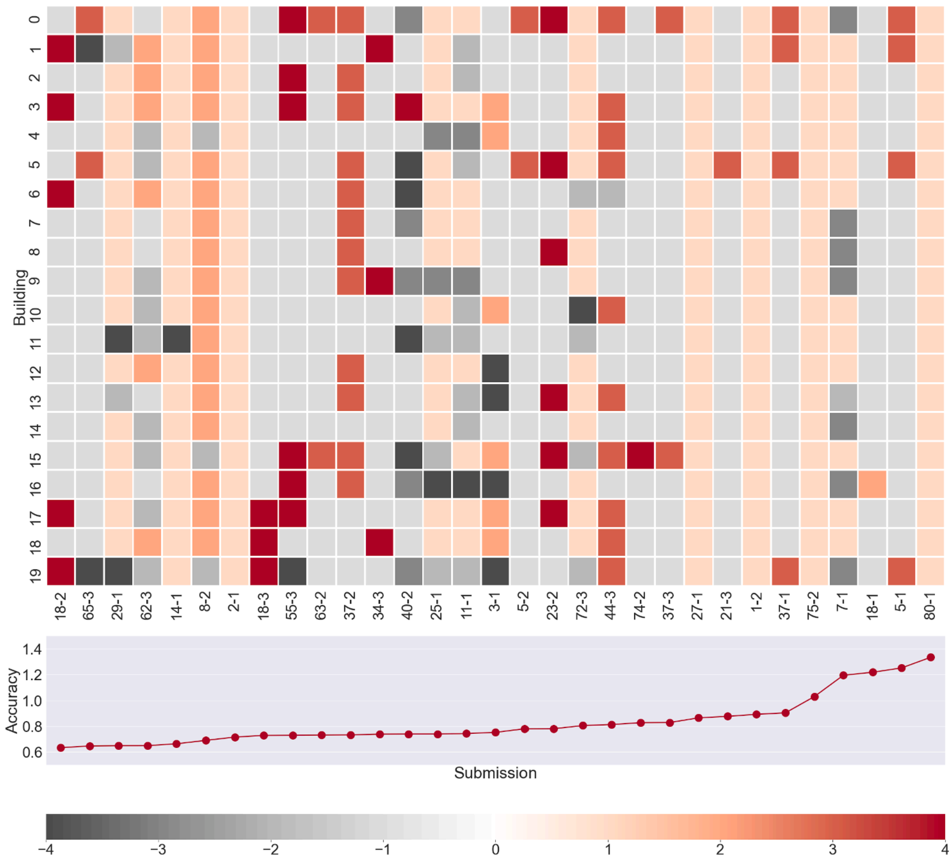

Fig. 13. Submissions with data selection based on building selection and their accuracy (evaluation score) (the colours in the upper subplot marked with numbers in the colour bar indicate whether the building was selected and the selection strategy, the negative values indicate not chosen, and the positive values indicate chosen; –4 indicates that the building was deleted owing to poor data quality; –3 indicates that the building was deleted because it had a different energy consumption distribution from the target building; –2 indicates that the building was deleted because it behaved differently from the other reference buildings; –1 indicates that the building was simply not chosen; 1 indicates that the building was simply chosen; 2 indicates that the building was chosen because it behaved typically among the reference buildings; 3 indicates that the building was chosen because it had the same given label as the target building; 4 indicates that the building was chosen because it had the same energy consumption distribution as the target building).

beneath the figure. For a clear comparison, the best evaluation scores of each team were divided into four groups: 0.6–0.7, 0.7–0.8, 0.8–1.0, and 1.0+, corresponding to great, nice, good, and fair, respectively. As shown in the figure, multiple workflows achieved good accuracy, and predictions based on the same workflow did not necessarily achieve similar accuracy. The prediction workflow responsible for good accuracy had not yet occurred in this case. Different operations in every process led to different results. Moreover, the use of machine-learning models did not lead to better results. The current machine-learning model may not be sufficient to solve such a problem. Further work is needed to determine the limitations of the current machine-learning models for problem scenarios in which cross-building prediction is required without sufficient historical data.

Table 3 presents the data-driven workflow used by the top five teams for W-meter type prediction. As shown in Fig. 9 and Table 3, simple statistical methods were frequently used. And Fig. 9 indicates that the use of simple methods did not certainly lead to poor prediction accuracy. Meanwhile, the use of the machine learning model did not promise good accuracy. It seems that the model selection is not the only influencing factor of the prediction accuracy. The data preparation process differed among the participants, which is worthy of further discussion. As mentioned in Section 2.2.2, although the amount of data available has been increasing, little attention has been paid to data preparation techniques in review studies. Thus, a detailed discussion of the data preparation methods is presented below.

5.2. Data pre-processing methods

5.2.1. Overview of data pre-processing methods

Data pre-processing plays an important role in Big Data analysis. Data collected using IoT smart meters may have problems such as outliers, which may hinder the analysis based on the data [64].

Basic abnormal data types can be classified as point anomalies, contextual anomalies, and collective anomalies [65]. Point anomalies are single-point outliers that are considered anomalous with the rest of the data [65]. As shown in Fig. 10, for example, negative and extra-large values are typical examples of point anomalies. Contextual anomalies are observations or sequences that deviate from the expected patterns within the time series [66], such as sudden increases/decreases in consumption and large consumption at night. This type of anomaly should be certain with the specific context in the time series; otherwise, it may be ignored. Collective anomalies [65] refer to a collection of related data instances that are anomalous when they occur together. In energy consumption data, a stable series is a typical example of a collective anomaly.

a. Point anomalies

Point anomalies are the most basic anomalies and have attracted attention of most of the teams. Although some pre-processing work was done before the contest, as mentioned in Section 2.1, there were negative outliers. Most of the teams simply found them by setting the lower limit of the data to zero. Participants applied different procedures to identify positive outliers. The method of Tukey's fences [67], which is based on the interquartile range (IQR) [68], was widely used among the participants for detecting outliers. The first quartile was subtracted from the third quartile to determine the IQR and then multiplying it with a constant such as 1.5, so that a certain value can be determined whether it is an outlier.

However, because a zero value makes sense in the HVAC system but may be determined as an outlier with the IQR, the IQR should be modified in the present case. Team 6 suggests setting the 99th percentile of the dataset as the upper limit of the data.

The PauTa criterion (3σ law) [69] was also used extensively for outlier determination. Team 54 first converted all the data to a 0–1 normal distribution and then used the PauTa criterion to identify outliers.

Sometimes, data collected from different buildings may have different distributions, and it may be a good idea to determine the outlier with separated groups. Team 14 used k-means to detect outliers in each type of meter data for each building, and Team 27 used the local outlier factor, which is widely used for identifying density-based outliers [70].

b. Contextual anomaly

In this case, contextual anomaly recognition requires combining more physical analysis with the data. Because energy consumption data have strong physical significance, some expert experience is useful. For example, with different load commands in winter and summer, the upper limit for judging outliers should be different. Moreover, the load command on the weekend and at night is seldom large. Team 85 sets a rule based on physical meaning to calibrate the data. They considered that the continuous peak load occurring at night was abnormal. They also set different upper limits for light consumption data at the weekend and HVAC consumption data at night on the weekend.

Furthermore, as the energy consumption data are time-series data, a single point outlier judgement is not sufficient to identify all the abnormal values. Values that have a large change rate in the time series may be incorrect. Team 80 calculated the first-order difference of the energy consumption series and then used the PauTa criterion to determine the abnormal values.

c. Collective anomaly

As for collective anomalies, stable values were regular collective anomalies in energy consumption data, which may have been caused by sensor errors, sensors being offline, etc. Many participants did not deal with them, because these types of values may be nonzero and are easy to overlook. Some teams identified stable values by observing the energy consumption series plots. Team 18 calculated the standard deviation of the daily energy consumption data for each building and then considered the data with a daily standard deviation below a specified threshold as stable values. Team 85 determined data that remained unchanged for one week to be stable values. However, the collective anomalies with overall offsets in energy consumption did not attract the attention of the participants.

The methods used for abnormal data recognition in the competition are shown in Fig. 11, including the methods for detecting points, context, and collective anomalies.

5.2.2. Opportunity for improving data pre-processing methods

As shown in Fig. 11, the participants paid more attention to point anomaly detection and tested more methods, while only a few teams dealt with the contextual and collective anomalies in the data pre-processing, and the methods they used were not enough automatically. A similar situation occurred in the ASHRAE Great Energy Predictor III competition, in which most of the winners only paid attention to point anomalies [11]. This is typical in today's data-driven energy consumption prediction cases, particularly when an increasing number of data science engineers attend, because point anomaly detection requires little physical background knowledge, while the other two require.

Methods for context and collective anomaly detection in energy consumption data have been proposed in several papers. Forecasting the energy consumption baseline and then detecting the anomalies with the baseline is the most widely used strategy. Ploennigs et al. [71] proposed an anomaly detection method based on a baseline consumption prediction model, generalised additive models (GAMs), and Chou et al. [72] predicted daily consumption baseline with the ARIMA and then identified anomalies with the baseline by applying the two-sigma rule. Additionally, typical daily patterns extracted from energy consumption data will help to identify context and collective anomalies. Piscitelli et al. [73] proposed an anomaly detection method based on the recognition of anomalous electrical daily energy patterns.

Context and collective anomaly detection methods perform well when the data is as clean as possible; thus, a front point anomaly detection process is helpful. This was ignored by the participants who focused on the context and collective anomalies in the competition. Further work on the data pre-processing workflow can combine the point anomaly detection method with context and collective anomaly detection methods to achieve better automatic pre-processing.

5.3. Feature creation and data selection

5.3.1. Need for feature creation and data selection

In the competition, participants do not know the historical energy consumption data of the target building but instead have some knowledge about the historical energy consumption of several buildings of the same type. However, not all buildings marked as the same use type have the same distribution of energy consumption. For example, buildings sometimes behave differently from their labels [74], and existing buildings may have served different or mixed purposes over time. Meanwhile, the current research shows that even buildings with the same primary use type can have different time-series features [75], which indicates their different distributions. Thus, not all the reference buildings behaved similarly to the target building at any time. Proper data selection, such as clustering, can improve the prediction accuracy [76]. Moreover, the features offered to describe the differences among buildings, as stated in Table 1, are not sufficient to describe the differences among the energy consumption distributions. Pattern mining from the time series may help us obtain hidden information [77]. Thus, extra feature creation or data selection is needed to reduce the amount of misleading information for prediction.

As shown in Figs. 8 and 9, some participants enriched their data preparation processes with feature creation and data selection. Some teams created features to improve their models, while others focused on existing information and chose a method for selecting data. Some teams considered that the combination of feature creation and data selection may help improve the prediction accuracy; thus, they created features to analyse the differences in the building energy consumption distribution among buildings and then select data.

5.3.2. Overview of feature creation methods

Except for the features given, the features that described the day type were important features that significantly affected the energy consumption. Thus, many teams created day-type features, such as nominal attributes about the holiday, as the input variables of their models.

Additionally, some features were created according to the given energy consumption time series in the competition. All the temporal features in the competition were based on statistics, which are easy to obtain but may include rich meaning to help describe the differences among buildings.

Team 18 built three statistical features from the energy consumption data to further understand the buildings. They defined the energy consumption intensity of working time, the relative energy consumption level of holidays and weekends compared with weekdays, and the relative energy consumption level of after-work time on weekdays to describe buildings, considering that office buildings differ from each other because of different working schedules and additional commercial usage. The three features they created were ratio-based features, which had a normalising effect, making it easy to compare buildings [41].

Team 23 focused on the control strategy and equipment behaviour hidden in the time-series data. They built a feature list that included six time-series features with physical meanings. Five of the six time-series features were related to the cooling plant in the building. They considered whether there were stages in the daily consumption plot, and information about the cooling plants, such as the number of chillers equipped and their capacity, could be inferred. Thus, they obtained these features through manual analysis of the stages of the daily consumption plot. The other feature was whether the HVAC system is equipped with variable-frequency equipment. It was inferred from the existence of stages in the consumption plot of the peak-load day. All the temporal features that they created were based on manual observation of the consumption plot; nonetheless, it was a good attempt at feature creation.

5.3.3. Overview of data selection methods

Several data selection strategies were used in the competition. They can be mainly classified into two types: moment-based and building-based. Moment-based strategies involve selecting suitable data at each moment, regardless of the specific building. Building-based strategies involve determining the building set in use at the beginning and then using the data obtained from these buildings for prediction at every moment.

Only one team used a moment-based strategy. It was based on the consideration that buildings of the same type may have some moments that behave differently from each other, and those in different types may also have some moments that behave in the same way. Thus, selecting the data of a certain moment may increase the utilisation of the data compared with the selection of the building. Team 6 proposed a method to automatically select data for each day via a nearest-neighbour search in a large-scale dataset [78,79]. Considering that different types of days, which are the different days of the week in different months, may have different consumption features, they first described the typical daily energy consumption time series for each type of day for each building. They then abstracted a time series to three features, i.e. the mean consumption before, during, and after the working time. With the features extracted from each time series, the difference between them was measured using the Euclidean distance. Team 6 then deleted the data that behaved differently from others in each day type.

The building-based strategy was a widely used data selection strategy in the competition. It can be subdivided into several types, as shown in Fig. 12. For some teams, buildings were chosen passively by deleting the improper buildings, whereas other teams chose them with consideration of their suitability for prediction.

Analysing a typical building was one of the bases for data selection. Buildings that behaved differently from other buildings were not selected by some teams, considering that the target building should be a group of buildings with traditional behaviour among office buildings. Some teams analysed the hourly consumption of a typical day for each building and then deleted the buildings with untraditional consuming behaviour from the dataset. Team 3 used hierarchical clustering to identify the group of buildings with the most traditional behaviour. They used the average hourly energy consumption data for one year as the characteristic to cluster for each year and each meter type and finally chose the buildings in the largest cluster for each type of meter.

Consideration and attempt to find a building similar to the target building among the reference buildings is another basis for data selection. Although the energy consumption data of the target building are not available, some information about the building is provided. Identifying buildings similar to the target building among the 20 buildings given was an achievable task. The easiest way to do this (as many of the teams did) was to select the buildings that had the same type of HVAC system as the target building, i.e. a central plant all-air system. However, the consumption difference between system types may not have been large, but the occupancy may have had a significant influence. Thus, mining the consumption behaviour hidden in the data and inferring that of the target building were vital.

Energy consumption patterns are important characteristics in that energy consumption data may have many differences from each other. Team 80 used k-medoids to identify typical buildings among all the buildings. Then, they analysed which of the typical buildings was closer to the target building. With the typical building chosen, they used k-means to perform pattern searching and then built the energy consumption data of the target building with every pattern that they found.

Sometimes, the features of the time series may help us obtain hidden information. With three features created, as mentioned in Section 4.3.2, team 18 chose buildings with different values of the features as historical data to predict the target building and finally selected those that could obtain the best accuracy. With the feature list mentioned in Section 4.3.2, team 23 also easily identified the most similar buildings and used them to predict the target building.

5.3.4. Opportunity on feature creation and data selection of energy consumption data

The feature creation process was applied for two functions in the competition: for enriching input variables of the model and for selecting data. As previous research has indicated, the model input will significantly influence the prediction result; thus, feature engineering has gained increasing attention [80]. In the competition, another role of feature engineering in building energy consumption prediction calls for concern. The role that helps data selection in feature engineering is inseparable from knowledge discovery on time-series data. Future data-driven prediction frameworks can combine more data-mining techniques with feature creation in the data preparation process to achieve higher interpretation and accuracy.

Fig. 13 presents the building selected and the corresponding accuracy for some of the submissions in the competition. The colour in the upper subplot, which is marked with a number in the colour bar, indicates whether the building was selected and the selection strategy. As shown in the figure, the data selection significantly affected the accuracy. Adding building to the training set did not necessarily improve the accuracy, and the choice of the building also had a significant influence. Because not all the buildings that marked as the same use type have similar hourly energy time series. Adding data from buildings that have significantly differet time series from the target building will not help accurate prediction while selecting data that have similar features from that of the target building will be helpful. What's more, the chosen building sets differed among different data selection strategies, as shown in Fig. 13. The sample reported in the competition is insufficient to draw a conclusion, but the strategies and methods used in the competition may aid further analysis. Further research is needed to determine the types of buildings that will help predict.

6. Conclusions

With the emergence of new energy systems and the widespread use of smart meters, accurate load prediction for buildings is demanded. A new data context for building energy consumption prediction has emerged.

Previous studies indicate that the new data context has introduced not only new opportunities for the development of prediction technology but also challenges. According to a review of existing research, there are two main ideas for prediction-method improvement. One is to develop a hybrid model, and the other is to improve the capability of the data-driven model for general scenarios. However, the current methods based on these two ideas are still in the early development phase, and new methods need to be developed and tested.

With the analysis of the methods and results from 'Energy Detective', i.e. a competition that we designed and held with a typical data context today, we gained knowledge regarding the status quo and the opportunity for future prediction-method improvement.

In general, current techniques have limited accuracy because of the prediction with limited physical parameters and historical data of a building. Thus, improvement of the methods is needed. In the competition, ideas were tested, providing insight into possible future developments.

Several teams used hybrid models for prediction. We discuss how their models were constructed and whether there are basic strategies for hybrid-model development. After the discussion, hybrid strategies for hybrid models were identified.

- Hybrid models can be built by combining a data-driven approach and a physical approach in sequence, parallel, and feedback methods.
- The choice of a single physical approach for a hybrid can be varied. In addition to the single detailed physical modelling of buildings, urban building energy modelling can be used for hybrids.
- The choice of a single data-driven approach for hybrids can also vary. Some new data-driven approaches, such as the transfer learning prediction model Hephaestus, can be used in hybrid-model development.

Data-driven approaches were popular among the participants. Analysing the methods used revealed that the choice and application of the models were not the only vital parts of the workflow. Thus, we propose suggestions for the workflow. Additionally, according to an analysis of the data preparation processes used in the competition, including data pre-processing, feature creation, and data selection, we drew the following conclusions regarding their current states and future development ideas.

- Every process of the data-driven workflow affected the prediction accuracy. Different operations in every process may lead to different accuracy levels.
- The data preparation process has not received sufficient attention in the current state. As increasing amounts of data become available, data preparation will become increasingly important and require more attention.
- Contextual and collective anomalies are as important as point anomalies in data pre-processing, but researchers currently pay little attention to them. Combining point anomaly detection methods with context and collective anomaly detection methods may help achieve better automatic pre-processing.
- Feature engineering, including feature creation, extraction, and selection, plays an important role in improving model interpretability and accuracy. It is responsible for not only enriching the input variables of the model but also selecting data in building energy consumption prediction. More data-mining techniques can be applied to feature creation in the data preparation process to achieve higher interpretability and accuracy.
- Data selection is vital for the cross-building energy prediction. Adding building data to the training set did not lead to a higher prediction accuracy. The choice of the building significantly affected the results. Several data selection strategies were identified from the competition. Further research is needed to determine how to categorise various types of buildings for energy prediction purposes.

CRediT authorship contribution statement

Tong Xiao: Conceptualization, Methodology, Writing – original draft. **Peng Xu:** Supervision. **Ruikai He:** Writing – review & editing. **Huajing Sha:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research presented in this paper is supported by the Science and Technology Project of China Southern Power Grid (No. GDKJXM20212099).

Appendix. A

A total of 159 participants were recruited from 45 universities and scientific research institutions. 36 of the participants worked in 20 enterprises. Information on the teams that achieved top rankings in the competition is presented in Table 4.

Table 4
Information on the top 15 teams in the competition.

Team No.	Team Type	Organisations the members worked in or studied at
18	Researchers or students	Chongqing Univ.
29	Researchers or students	Shenzhen Univ.
6	Researchers or students	Shenzhen Univ., Hong Kong Polytechnic Univ.
2	Researchers or students	RWTH Aachen Univ. in Germany, Technical Univ. of Berlin
72	Researchers or students	Southwest Jiaotong Univ.
14	Researchers or students	Shenzhen Univ.
40	Researchers or students	Huazhong Univ. of Science and Technology
55	Engineers	SRIBS (Shanghai Research Institute of Building Scientific)
37	Researchers or students	Huazhong Univ. of Science and Technology
3	Researchers and engineers	Tongji Univ., Jiangsu Enn Energy Development Co., Ltd.
63	Researchers or students	Huazhong Univ. of Science and Technology
23	Engineers	Shanghai East Low Carbon Industry Co., Ltd.
5	Engineers	Shanghai DeAng Technology Co., Ltd.
85	Engineers	Shanghai Inwhile Intelligent Technology Co., Ltd.
34	Researchers or students	Huazhong Univ. of Science and Technology

References

[1] Pan W, Garmston H. Compliance with Building Energy Regulations for New-Build Dwellings. Energy 2012;48(1):11–22. https://doi.org/10.1016/j.energy.2012.06.048.
[2] Hahn H, Meyer-Nieberg S, Pickl S. Electric Load Forecasting Methods: Tools for Decision Making. Eur J Oper Res 2009;199(3):902–7. https://doi.org/10.1016/j.ejor.2009.01.062.
[3] Chen, Yongbao, Peng Xu, Yiyi Chu, Weilin Li, Yuntao Wu, Lizhou Ni, and et al., 'Short-Term Electrical Load Forecasting Using the Support Vector Regression (SVR) Model to Calculate the Demand Response Baseline for Office Buildings', Applied Energy, 195 (2017), 659–70 DOI: 10.1016/j.apenergy.2017.03.034.
[4] White JA, and Reichmuth R. 'Simplified Method for Predicting Building Energy Consumption Using Average Monthly Temperatures', in IECEC 96. Proceedings of the 31st Intersociety Energy Conversion Engineering Conference, 1996, iii, 1834–39 vol.3 DOI: 10.1109/IECEC.1996.553381.
[5] Al-Homoud MS. Computer-Aided Building Energy Analysis Techniques. Build Environ 2001;36(4):421–33. https://doi.org/10.1016/S0360-1323(00)00026-3.
[6] Kaur, Amanpreet, Lukas Nonnenmacher, and Carlos F. M. Coimbra, 'Net Load Forecasting for High Renewable Energy Penetration Grids', Energy, 114 (2016), 1073–84 DOI: 10.1016/j.energy.2016.08.067.
[7] Xiao Fu, Fan C. Data Mining in Building Automation System for Improving Building Operational Performance. Energy Build 2014;75:109–18. https://doi.org/10.1016/j.enbuild.2014.02.005.
[8] Ding Y, Zhang Q, Yuan T, Yang F. Effect of Input Variables on Cooling Load Prediction Accuracy of an Office Building. Appl Therm Eng 2018;128:225–34. https://doi.org/10.1016/j.applthermaleng.2017.09.007.
[9] Yu, Zhun, Benjamin C. M. Fung, and Fariborz Haghighat, 'Extracting Knowledge from Building-Related Data - A Data Mining Framework', Building Simulation, 6.2 (2013), 207–22 DOI: 10.1007/s12273-013-0117-8.
[10] Zhiguo C, Yong C, Jingshu W, Xiaofeng M, Ran Li, Yannan T. Research and Application Progress of Data Mining Technology in the Field of Building HVAC. Build Sci 2018;34(04):85–97.
[11] Miller, Clayton, Pandarasamy Arjunan, Anjukan Kathirgamanathan, Chun Fu, Jonathan Roth, June Young Park, and et al., 'The ASHRAE Great Energy Predictor III Competition: Overview and Results', Science and Technology for the Built Environment, 26.10 (2020), 1427–47 https://doi.org/10.1080/23744731.2020.1795514.

[12] Hong T, Pinson P, Fan Su. Global Energy Forecasting Competition 2012. Int J Forecast 2014;30(2):357–63. https://doi.org/10.1016/j.ijforecast.2013.07.001.
[13] Hong T, Xie J, Black J. Global Energy Forecasting Competition 2017: Hierarchical Probabilistic Load Forecasting. Int J Forecast 2019;35(4):1389–99. https://doi.org/10.1016/j.ijforecast.2019.02.006.
[14] Amasyali K, El-Gohary NM. A Review of Data-Driven Building Energy Consumption Prediction Studies. Renew Sustain Energy Rev 2018;81:1192–205. https://doi.org/10.1016/j.rser.2017.04.095.
[15] 'DOE-2' <http://doe2.com/DOE2/index.html> [accessed 26 November 2020].
[16] 'EnergyPlus | EnergyPlus' <https://energyplus.net/> [accessed 26 November 2020].
[17] 'EQUEST' <http://www.doe2.com/equest/> [accessed 26 November 2020].
[18] 'Welcome | TRNSYS : Transient System Simulation Tool' <http://www.trnsys.com/> [accessed 26 November 2020].
[19] 'Welcome to ESP-r' <http://www.esru.strath.ac.uk/Courseware/ESP-r/tour/> [accessed 26 November 2020].
[20] 'DeST download station' https://update.dest.com.cn/> [accessed 26 November 2020].
[21] Wei, Yixuan, Xingxing Zhang, Yong Shi, Liang Xia, Song Pan, Jinshun Wu, and et al., 'A Review of Data-Driven Approaches for Prediction and Classification of Building Energy Consumption', Renewable and Sustainable Energy Reviews, 82 (2018), 1027–47 DOI: 10.1016/j.rser.2017.09.108.
[22] Fumo N. A Review on the Basics of Building Energy Estimation. Renew Sustain Energy Rev 2014;31:53–60. https://doi.org/10.1016/j.rser.2013.11.040.
[23] Wenjie, Wu, 'An Overview of Building Energy Consumption Analysis Methods in China', in Sustainable Cities Development and Environment Protection, Pts 1-3, ed. by N. Xu, L. Tian, and F. Dai (Stafa-Zurich: Trans Tech Publications Ltd, 2013), ccclxi–ccclxiii, 366–69 DOI: 10.4028/www.scientific.net/AMM.361-363.366.
[24] Woloszyn M, Rode C. Tools for Performance Simulation of Heat, Air and Moisture Conditions of Whole Buildings. Build Simul 2008;1(1):5–24. https://doi.org/10.1007/s12273-008-8106-z.
[25] Crawley DB, Hand JW, Kurnmert M, Griffith BT. Contrasting the Capabilities of Building Energy Performance Simulation Programs. Build Environ 2008;43(4):661–73. https://doi.org/10.1016/j.buildenv.2006.10.027.
[26] Foucquier A, Robert S, Suard F, Stephan L, Jay A. State of the Art in Building Modelling and Energy Performances Prediction: A Review. Renew Sustain Energy Rev 2013;23:272–88. https://doi.org/10.1016/j.rser.2013.03.004.
[27] Gao Y, Ruan Y, Fang C, Yin S. Deep Learning and Transfer Learning Models of Energy Consumption Forecasting for a Building with Poor Information Data. Energy Build 2020;223:110156. https://doi.org/10.1016/j.enbuild.2020.110156.
[28] Smith A, Fumo N, Luck R, Mago PJ. Robustness of a Methodology for Estimating Hourly Energy Consumption of Buildings Using Monthly Utility Bills. Energy Build 2011;43(4):779–86. https://doi.org/10.1016/j.enbuild.2010.11.012.
[29] Li X, Wen J. Review of Building Energy Modeling for Control and Operation. Renew Sustain Energy Rev 2014;37:517–37. https://doi.org/10.1016/j.rser.2014.05.056.
[30] Zhao H-X, Magoulès F. A Review on the Prediction of Building Energy Consumption. Renew Sustain Energy Rev 2012;16(6):3586–92. https://doi.org/10.1016/j.rser.2012.02.049.
[31] Deb, Chirag, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah, 'A Review on Time Series Forecasting Techniques for Building Energy Consumption', Renewable & Sustainable Energy Reviews, 74 (2017), 902–24 DOI: 10.1016/j.rser.2017.02.085.
[32] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep Learning for Estimating Building Energy Consumption. Sustainable Energy Grids Networks 2016;6:91–9. https://doi.org/10.1016/j.segan.2016.02.005.
[33] Qian F, Gao W, Yang Y, Dan Yu. Potential Analysis of the Transfer Learning Model in Short and Medium-Term Forecasting of Building HVAC Energy Consumption. Energy 2020;193:315–24. https://doi.org/10.1016/j.energy.2019.116724.
[34] Ribeiro, Mauro, Katarina Grolinger, Hany F. ElYamany, Wilson A. Higashino, and Miriam A. M. Capretz, 'Transfer Learning with Seasonal and Trend Adjustment for Cross-Building Energy Forecasting', Energy and Buildings, 165 (2018), 352–63 DOI: DOI: 10.1016/j.enbuild.2018.01.034.
[35] Afshari A, Liu N. Inverse Modeling of the Urban Energy System Using Hourly Electricity Demand and Weather Measurements, Part 2: Gray-Box Model. Energy Build 2017;157:139–56. https://doi.org/10.1016/j.enbuild.2017.01.052.
[36] Reynders G, Diriken J, Saelens D. Quality of Grey-Box Models and Identified Parameters as Function of the Accuracy of Input and Observation Signals. Energy Build 2014;82:263–74. https://doi.org/10.1016/j.enbuild.2014.07.025.
[37] Elbeltagi E, Wefki H, Abdrabou Sa, Dawood M, Ramzy A. Visualized Strategy for Predicting Buildings Energy Consumption during Early Design Stage Using Parametric Analysis. Journal of Building Engineering 2017;13:127–36. https://doi.org/10.1016/j.jobe.2017.07.012.
[38] Fumo N, Mago P, Luck R. Methodology to Estimate Building Energy Consumption Using EnergyPlus Benchmark Models. Energy Build 2010;42(12):2331–7. https://doi.org/10.1016/j.enbuild.2010.07.027.
[39] Vaghefi SA, Jafari MA, Zhu J, Brouwer J, Lu Y. A Hybrid Physics-Based and Data Driven Approach to Optimal Control of Building Cooling/Heating Systems. IEEE Trans Autom Sci Eng 2016;13(2):600–10. https://doi.org/10.1109/TASE.2014.2356337.
[40] Pan, Sinno Jialin, and Qiang Yang, 'A Survey on Transfer Learning', Ieee Transactions on Knowledge and Data Engineering, 22.10 (2010), 1345–59 DOI: 10.1109/TKDE.2009.191.
[41] Miller C, Meggers F. Mining Electrical Meter Data to Predict Principal Building Use, Performance Class, and Operations Strategy for Hundreds of Non-Residential Buildings. Energy Build 2017;156:360–73. https://doi.org/10.1016/j.enbuild.2017.09.056.
[42] Zhang L, Wen J. A Systematic Feature Selection Procedure for Short-Term Data-Driven Building Energy Forecasting Model Development. Energy Build 2019;183:428–42. https://doi.org/10.1016/j.enbuild.2018.11.010.
[43] Miller C, Nagy Z, Schlueter A. A Review of Unsupervised Statistical Learning and Visual Analytics Techniques Applied to Performance Analysis of Non-Residential Buildings. Renew Sustain Energy Rev 2018;81:1365–77. https://doi.org/10.1016/j.rser.2017.05.124.
[44] Chen, Zhe, Peng Xu, Fan Feng, Yifan Qiao, and Wei Luo, 'Data Mining Algorithm and Framework for Identifying HVAC Control Strategies in Large Commercial Buildings', Building Simulation DOI: 10.1007/s12273-019-0599-0.
[45] Qiu S, Feng F, Li Z, Yang G, Peng Xu, Li Z. Data Mining Based Framework to Identify Rule Based Operation Strategies for Buildings with Power Metering System. Build Simul 2019;12(2):195–205. https://doi.org/10.1007/s12273-018-0472-6.
[46] Pacheco-Torres R, Heo Y, Choudhary R. Efficient Energy Modelling of Heterogeneous Building Portfolios. Sustainable Cities and Society 2016;27:49–64. https://doi.org/10.1016/j.scs.2016.08.001.
[47] Hylleberg, Svend, 'CHAPTER 2 - The Definition of Seasonality', in Seasonality in Regression, ed. by Svend Hylleberg (Academic Press, 1986), pp. 15–35 DOI: 10.1016/B978-0-12-363455-9.50007-2.
[48] 'Building Performance Database (BPD)', Energy.Gov <https://www.energy.gov/eere/buildings/building-performance-database-bpd> [accessed 22 November 2020].
[49] Miller, Clayton, and Forrest Meggers, 'The Building Data Genome Project: An Open, Public Data Set from Non-Residential Building Electrical Meters', Energy Procedia, CISBAT 2017 International ConferenceFuture Buildings & Districts – Energy Efficiency from Nano to Urban Scale, 122 (2017), 439–44 DOI: 10.1016/j.egypro.2017.07.400.
[50] Miller, Clayton, Anjukan Kathirgamanathan, Bianca Picchetti, Pandarasamy Arjunan, June Young Park, Zoltan Nagy, and others, 'Buds-Lab/Building-Data-Genome-Project-2: V1.0' (Zenodo, 2020) DOI: 10.5281/zenodo.3887306.
[51] Miller, Clayton, Anjukan Kathirgamanathan, Bianca Picchetti, Pandarasamy Arjunan, June Young Park, Zoltan Nagy, and et al., 'The Building Data Genome Project 2, Energy Meter Data from the ASHRAE Great Energy Predictor III Competition', Scientific Data, 7.1 (2020), 368 DOI: 10.1038/s41597-020-00712-x.
[52] Fan C, Chen M, Wang X, Wang J, Huang B. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. Front Energy Res 2021;9:652801. https://doi.org/10.3389/fenrg.2021.652801.
[53] Hong, Taehoon, Chan-Joong Kim, Jaemin Jeong, Jimin Kim, Choongwan Koo, Kwangbok Jeong, and et al., 'Framework for Approaching the Minimum CV(RMSE) Using Energy Simulation and Optimization Tool', Energy Procedia, CUE 2015 - Applied Energy Symposium and Summit 2015: Low carbon cities and urban energy systems, 88 (2016), 265–70 DOI: 10.1016/j.egypro.2016.06.157.
[54] ASHRAE, Guideline 14-2014: Measurement of Energy and Demand Savings, ASHRAE, Atlanta, Georgia, (2014).
[55] Pan Yiqun, Huang Zhizhong, Gang Wu. Calibrated Building Energy Simulation and Its Application in a High-Rise Commercial Building in Shanghai. Energy Build 2007;39(6):651–7. https://doi.org/10.1016/j.enbuild.2006.09.013.
[56] Nutkiewicz Alex, Yang Zheng, Jain Rishee K. Data-Driven Urban Energy Simulation (DUE-S): A Framework for Integrating Engineering Simulation and Machine Learning Methods in a Multi-Scale Urban Energy Modeling Workflow. Appl Energy 2018;225:1176–89. https://doi.org/10.1016/j.apenergy.2018.05.023.
[57] Reinhart Christoph F. and Carlos Cerezo Davila, 'Urban Building Energy Modeling – A Review of a Nascent Field'. Build Environ 2016;97:196–202. https://doi.org/10.1016/j.buildenv.2015.12.001.
[58] Mueller, Dirk, Moritz Lauster, Ana Constantin, Marcus Fuchs, and Peter Remmen, 'AixLib – An Open-Source Modelica Library within the IEA-EBC Annex 60 Framework', 2016.
[59] 'AixLib' <https://build.openmodelica.org/Documentation/AixLib.html> [accessed 29 November 2020].
[60] De Jaeger, Ina, Glenn Reynders, Chadija Callebaut, and Dirk Saelens, 'A Building Clustering Approach for Urban Energy Simulations', Energy and Buildings, 2020; 208; 109671 DOI: 10.1016/j.enbuild.2019.109671.
[61] Chen Bei, Sinn Mathieu, Ploennigs Joern, Schumann Anika. Statistical Anomaly Detection in Mean and Variation of Energy Consumption. In: in *2014 22nd International Conference on Pattern Recognition (Icpr)*; 2014. p. 3570–5. https://doi.org/10.1109/ICPR.2014.614.
[62] Hyndman, Rob J, and George Athanasopoulos, Forecasting: Principles and Practice (Melbourne, Australia: OTexts) https://Otexts.com/fpp2/> [accessed 30 November 2020].
[63] CLEVELAND, R. B., 'STL : A Seasonal-Trend Decomposition Procedure Based on Loess', Journal of Office Statistics, 6.1 (1990), 3–73.
[64] Fan Cheng, Xiao Fu, Yan Chengchu. A Framework for Knowledge Discovery in Massive Building Automation Data and Its Application in Building Diagnostics. Autom Constr 2015;50:81–90. https://doi.org/10.1016/j.autcon.2014.12.006.
[65] Chandola, Varun, Arindam Banerjee, and Vipin Kumar, 'Anomaly Detection: A Survey', ACM Computing Surveys, 41.3 (2009), 15:1–15:58 DOI: 10.1145/1541880.1541882.
[66] Cook AA, Mısırlı G, Fan Z. Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet Things J 2020;7(7):6481–94. https://doi.org/10.1109/JIOT.2019.2958185.

[67] Tukey, John Wilder, Exploratory Data Analysis (Reading, Mass. : Addison-Wesley Pub. Co., 1977) <http://archive.org/details/exploratorydataa00tuke_0> [accessed 30 November 2020].

[68] Upton, Graham, and Ian Cook, Understanding Statistics (OUP Oxford, 1996).

[69] Li, Limin, Zongzhou Wen, and Zhongsheng Wang, 'Outlier Detection and Correction During the Process of Groundwater Lever Monitoring Base on Pauta Criterion with Self-Learning and Smooth Processing', in Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems, ed. by Lin Zhang, Xiao Song, and Yunjie Wu, Communications in Computer and Information Science (Singapore: Springer, 2016), pp. 497–503 DOI: DOI: 10.1007/978-981-10-2663-8_51.

[70] Breunig Markus M, Kriegel Hans-Peter, Ng Raymond T, Sander Jörg. LOF: Identifying Density-Based Local Outliers. ACM SIGMOD Record 2000;29(2): 93–104. https://doi.org/10.1145/335191.335388.

[71] Ploennigs, Joern, Bei Chen, Anika Schumann, and Niall Brady, 'Exploiting Generalized Additive Models for Diagnosing Abnormal Energy Use in Buildings', in Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, BuildSys'13 (New York, NY, USA: Association for Computing Machinery, 2013), pp. 1–8 DOI: 10.1145/2528282.2528291.

[72] Chou Jui-Sheng. and Abdi Suryadinata Telaga, 'Real-Time Detection of Anomalous Power Consumption'. Renew Sustain Energy Rev 2014;33:400–11. https://doi.org/10.1016/j.rser.2014.01.088.

[73] Piscitelli MS, Brandi S, Capozzoli A, Xiao F. A Data Analytics-Based Tool for the Detection and Diagnosis of Anomalous Daily Energy Patterns in Buildings. Build Simul 2021;14(1):131–47. https://doi.org/10.1007/s12273-020-0650-1.

[74] Quintana Matias, Arjunan Pandarasamy, Miller Clayton. Islands of Misfit Buildings: Detecting Uncharacteristic Electricity Use Behavior Using Load Shape Clustering. Build Simul 2021;14(1):119–30. https://doi.org/10.1007/s12273-020-0626-1.

[75] Miller Clayton. What's in the Box?! Towards Explainable Machine Learning Applied to Non-Residential Building Smart Meter Classification. Energy Build 2019;199:523–36. https://doi.org/10.1016/j.enbuild.2019.07.019.

[76] Shahzadeh A, Khosravi A, Nahavandi S. 'Improving Load Forecast Accuracy by Clustering Consumers Using Smart Meter Data'. International Joint Conference on Neural Networks (IJCNN) 2015;2015:1–7. https://doi.org/10.1109/IJCNN.2015.7280393.

[77] Ploennigs J, Chen B, Palmes P, Lloyd R. 'E2-Diagnoser: A System for Monitoring, Forecasting and Diagnosing Energy Usage'. IEEE International Conference on Data Mining Workshop 2014;2014:1231–4. https://doi.org/10.1109/ICDMW.2014.56.

[78] Beyer, Kevin, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft, 'When Is "Nearest Neighbor" Meaningful?', in Database Theory — ICDT'99, ed. by Catriel Beeri and Peter Buneman, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer, 1999), pp. 217–35 DOI: 10.1007/3-540-49257-7_15.

[79] Ferhatosmanoglu, Hakan, Ertem Tuncel, Divyakant Agrawal, and Amr El Abbadi, 'High Dimensional Nearest Neighbor Searching', Information Systems, 2006; 31.6: 512–40 https://doi.org/10.1016/j.is.2005.01.001.

[80] Fan, Cheng, Fu Xiao, and Yang Zhao, 'A Short-Term Building Cooling Load Prediction Method Using Deep Learning Algorithms', Applied Energy 2017; 195: 222–33 DOI: 10.1016/j.apenergy.2017.03.064.