Evaluate the Impact of Sensor Accuracy on Model Performance in Data-Driven Building Fault Detection and Diagnostics Using Monte Carlo Simulation

Abstract

The performance of data-driven fault detection and diagnostics (FDD) is heavily dependent on sensors. However, sensor inaccuracy and sensor faults are pervasive in building operation: inaccurate and missing sensor readings deteriorate FDD performance; sensor inaccuracy will also affect the selection of sensor for data-driven FDD in the model training process, which is another key factor of data-driven FDD performance. Sensor accuracy and sensor selection individually are well-studied research topics in this field, but the impact of sensor accuracy on sensor selection and its further impact on FDD performance has not been evaluated and quantified. In this paper, we developed a novel analysis methodology that comprehensively evaluate sensor fault on sensor selection and FDD accuracy. Monte Carlo Simulation is applied to deal with multiple stochastic sensor inaccuracy and provide probabilistic analysis results of the impact of sensor inaccuracy on sensor selection and FDD accuracy. This methodology focuses on the net impact of fault states across a full sensor set. The developed methodology can be used for the early-stage sensor design and operation-stage sensor maintenance. A case study is conducted to demonstrate the analysis methodology using a commercial building model crated to Flexible Research Platform located at Oak Ridge National Laboratory.

Keywords

fault detection and diagnostics; sensor accuracy; sensor fault; sensor selection; data-driven modeling; Monte Carlo simulation

1. Introduction

In commercial buildings, it is estimated that faulted building energy and control systems wasted 5% to 30% of energy (Lin, Kramer, and Granderson 2020). Fault detection and diagnostics (FDD) is a technique to detect operational errors and faults (fault detection) and finds the fundamental causes (fault diagnostics) with minimal human intervention (Zhang, Frank, Kim, Jin, and Leach 2020). FDD technologies are beneficial to building energy efficiency, operation cost savings, maintenance, and energy management (Lin, Kramer, and Granderson 2020).

FDD methods can be categorized into quantitative, qualitative, and process history-based (Katipamula and Brambley 2005). “Black-box” model is a subset of process history-based model, which is also termed as purely data-driven model and is hereafter referred to as “data-driven” (Katipamula and Brambley 2005). Massive datasets are collected from sensors and meters, and the increasing data empower data-driven modeling method and data-driven FDD have been increasingly applied (Zhang et al. 2021). Since FDD methods are largely data-driven, sensors, which are the foundation and source of data, are increasingly studied in FDD research. According to the conclusions of a review study (Zhang, Frank, Kim, Jin, Leach, et al. 2020), sensor faults (33 papers were found, accounting for 37% of reviewed technical papers) and sensor selection (14 papers were found, accounting for 16% of reviewed technical papers) are the most widely studied sensor topics in FDD. This indicates that the selection and accuracy of sensors are extremely important sensor-related factors for FDD applications. Sensor selection is critical to early-stage sensor design and operation-stage model improvement in FDD applications, while sensor accuracy is important during the operational stage of FDD. The research topic of sensor accuracy and sensor selection have been widely studied for data-driven building FDD in existing literature.

Sensor Accuracy. Sensor inaccuracy and fault can be divided into four types: bias, drift, precision degradation, and complete failure (Chen and Lan 2010). Dai and Gao (Dai and Gao 2013) noted that detecting sensor/actuator faults via data-driven modeling may be complicated, because sensor/actuator faults may influence model input/output in the same way as process faults; when factoring in input and sensor noise, it is even more challenging to detect sensor faults. Du et al. (Du et al. 2014) explained that in real building and HVAC (heating, ventilation, and air-conditioning) systems, application of the proposed methodology relies on the quantity and quality of sensor data. The authors developed combined neural networks to detect sensor faults in air handling units; they utilized adaptive subtractive clustering analysis to diagnose fault sources. Wang et al. (Wang, Zhou, and Xiao 2010) presented a strategy for FDD of HVAC systems that leverages sensor fault detection at the system level. They applied principal component analysis, or PCA, to detect and diagnose sensor bias and correct it before the application of the system FDD scheme. The sensor FDD method is able to identify and correct sensor biases when system faults coexist, and the system FDD method is also effective to diagnose system-level faults using the measurements processed by sensor FDD scheme.

Sensor Selection. Sensor selection is the process of using feature selection techniques to reduce dimensionality and improve generalization of data-driven models, which is widely studied not only in building FDD studies, but also in building energy forecasting studies (Zhang, Wen, and Buildings 2019). The data collected for FDD purpose may include sensed data from various sensors within the system and feedback data from various components of the system; additional data from external sources such as weather data can also be collected (Yuwono et al. 2015). Consequently, the dimensionality and volume of these data can be enormous (Yuwono et al. 2015). Selecting an optimal sensor set is a crucial step in a data-driven FDD process. The benefits of selecting the minimal number of sensors include computational complexity reduction, correlation removal, and optimization of the number of to-be-installed sensors. Zhao et al. (Zhao, Wang, and Xiao 2013) evaluated fault detection performance using different feature selection techniques. The selection criteria in their study are: (1) to be capable of identifying unique operating conditions and distinguish between faulty data and normal data, (2) to provide informational redundancy to enhance robustness, and (3) to minimize the number of sensors to reduce computational complexity and to maintain sensitivity to faults at low severities. Han et al. (Han et al. 2011) employed mutual information-based filter and genetic algorithm-based wrapper to find the key sensors to improve chiller FDD performance and reduce initial sensor costs at the same time.

Sensor faults (including various forms of inaccuracy) are common in building sensors and building automation systems; such faults can impact the data-driven FDD sensor selection process. While it is reasonable to expect that some effort would be made to ensure that a sensor set is well calibrated prior to FDD algorithm training, no real-world data set is perfect and data quality assessment for data-driven FDD modeling purposes is not generally straightforward. Additionally, for streaming or incremental machine learning applications, where algorithms are retrained on a rolling basis, it is unlikely that sensors would be recalibrated at every training interval. If a sensor typically selected as a feature by an FDD algorithm has failed (i.e., reads constant numeric values or reports a non-numeric error code), the data-driven FDD algorithm will not select it as a feature. In a less extreme case, where a sensor has some form and magnitude of inaccuracy but is not in a state of complete failure, the FDD algorithm may or may not choose to select it as a feature. And if it is selected, its inaccuracy may result in degradation of FDD detection and diagnosis performance. FDD algorithm performance may be very sensitive to the accuracy of certain sensors; in such cases, sensors should be carefully selected and calibrated when leveraged for FDD purposes. In other cases, sensor inaccuracy may not have much effect on feature selection or overall FDD performance. Quantifiable analysis is needed to evaluate these types of scenarios, which are common in the engineering practice of sensor selection, design, maintenance, and improvement. There are few studies that focus on the implications of sensor inaccuracy on FDD feature selection and performance.

Existing literature does not consider the interaction of sensor accuracy and sensor selection. However, sensor accuracy and sensor selection are not independent from each other. When training an FDD model, if the originally important sensor is not accurate, there is a tendency that this sensor will not be selected as the feature in the data-driven FDD model. Also, on incremental data-driven FDD modeling where input data is continuously collected to extend the existing model's knowledge or to further train the model, if an important sensor loses its accuracy, the data-driven method has a tendency of messing up with the prediction of model output since the inputs provide wrong information, resulting in the change of selected features in the incremental learning process. The impact of sensor inaccuracy on the sensor selection process and model performance in data-driven FDD has not been quantified or analyzed. Evaluating the impact of sensor accuracy on sensor selection and model performance is the next step to refine the existing work and fill the research gap.

In this paper, we explored the impact of sensor accuracy on sensor selection and model performance for data-driven building FDD model. We develop a Monte Carlo-based methodology to evaluate the impact of sensor accuracy on sensor selection and FDD accuracy. This method differentiates sensor types (temperature sensor, humidity sensor, electricity meter, etc.) and sensor fault types (bias, drifting, precision degradation, and complete failure). For each sensor type and sensor fault type, probabilities are assumed. For example, it can be assumed that the power meter fault rate is 10% and the probability of that sensor fault type being bias is 15% among all four sensor fault types mentioned earlier. All the assumed probabilities can be summarized via a probability table. Based on the probability table, Monte Carlo simulation is conducted to generate multiple decisive results. And the statistical probabilities of the decisive results are the final results of the evaluation workflow. This method focuses on a set of sensors as a whole, with a distribution of concurrent sensor faults across the set, instead of focusing on individual sensors. A case study is conducted to demonstrate these two methods using a commercial building model calibrated to Flexible Research Platform (FRP) located at Oak Ridge National Laboratory.

In terms of the organization of the paper, Section 2 introduces the Monte Carlo-based methodology of analyzing the impact of sensor inaccuracy on sensor selection and model performance in detail. Section 3 details the FRP case study of the developed workflow. Section 4 summarizes the results and findings of the case study. Section 5 draws conclusions and summarizes future work.

2. Monte Carlo-Based Analysis Methodology for Evaluating Sensor Accuracy on FDD Sensor Selection and Performance

As summarized from Section 1, sensor faults happen stochastically; there are hundreds of sensors in one building and chances are that multiple faults (different faulty sensors with different sensor fault types) are happening simultaneously. The stochastics among a large number of sensor faults make it even harder to quantify the uncertainty of the impact of sensor faults on sensor selection and FDD performance, and few papers are found in this research topic. In this section, a methodology to address this gap are presented: A Monte Carlo-based analysis methodology for evaluating sensor accuracy on FDD sensor selection and performance.

Building sensors malfunction and fail stochastically. Different sensor types have different robustness and malfunction rates; for example, enthalpy sensors and humidity sensors are less reliable than temperature sensors. Moreover, the probabilities of malfunction types vary from sensor to sensor. For example, complete failure of a temperature sensor may be less likely to happen than drifting or bias.

There are multiple potential sources of sensor accuracy uncertainty:

- Is the sensor faulty or operating normally?
- If the sensor is faulty, what is the type (bias, drifting, precision degradation, or complete failure, shown in Figure 1) of sensor fault?
- What is the severity, or magnitude, of the fault?

And each of these questions may depend on the sensor type (temperature, humidity, enthalpy, etc.). For each sensor type, the answers to these questions can be represented with probability distributions. We use the fault probability table structure presented in Table 1 to capture the likelihoods of the resulting combinations.

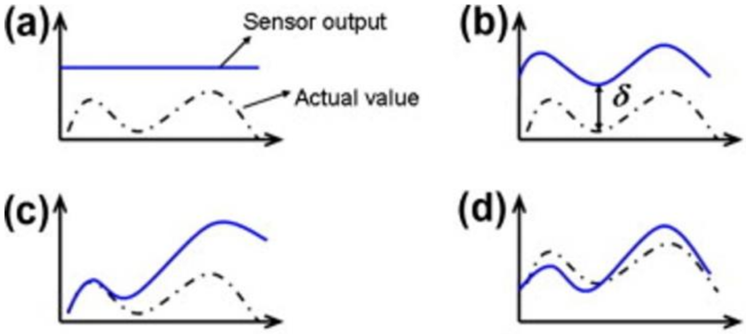


Figure 1. Diagram of (a) complete failure, (b) bias, (c) drifting, and (d) precision degradation

Table 1. Probability Table of Sensor Fault Types

Sensor Type/Sensor Fault Type	**Failure P(B1)**	**Bias P(B2)**	**Drifting P(B3)**	**Precision Degradation P(B4)**
Power Meter, P(A1)	P(B1|A1)	P(B2|A1)	P(B3|A1)	P(B4|A1)
Flow Meter, P(A2)	P(B1|A2)	P(B2|A2)	P(B3|A2)	P(B4|A2)
Thermometer, P(A3)	P(B1|A3)	P(B2|A3)	P(B3|A3)	P(B4|A3)
Differential pressure sensor, P(A4)	P(B1|A4)	P(B2|A4)	P(B3|A4)	P(B4|A4)
Enthalpy Sensor, P(A5)	P(B1|A5)	P(B2|A5)	P(B3|A5)	P(B4|A5)
…	…	…	…	…

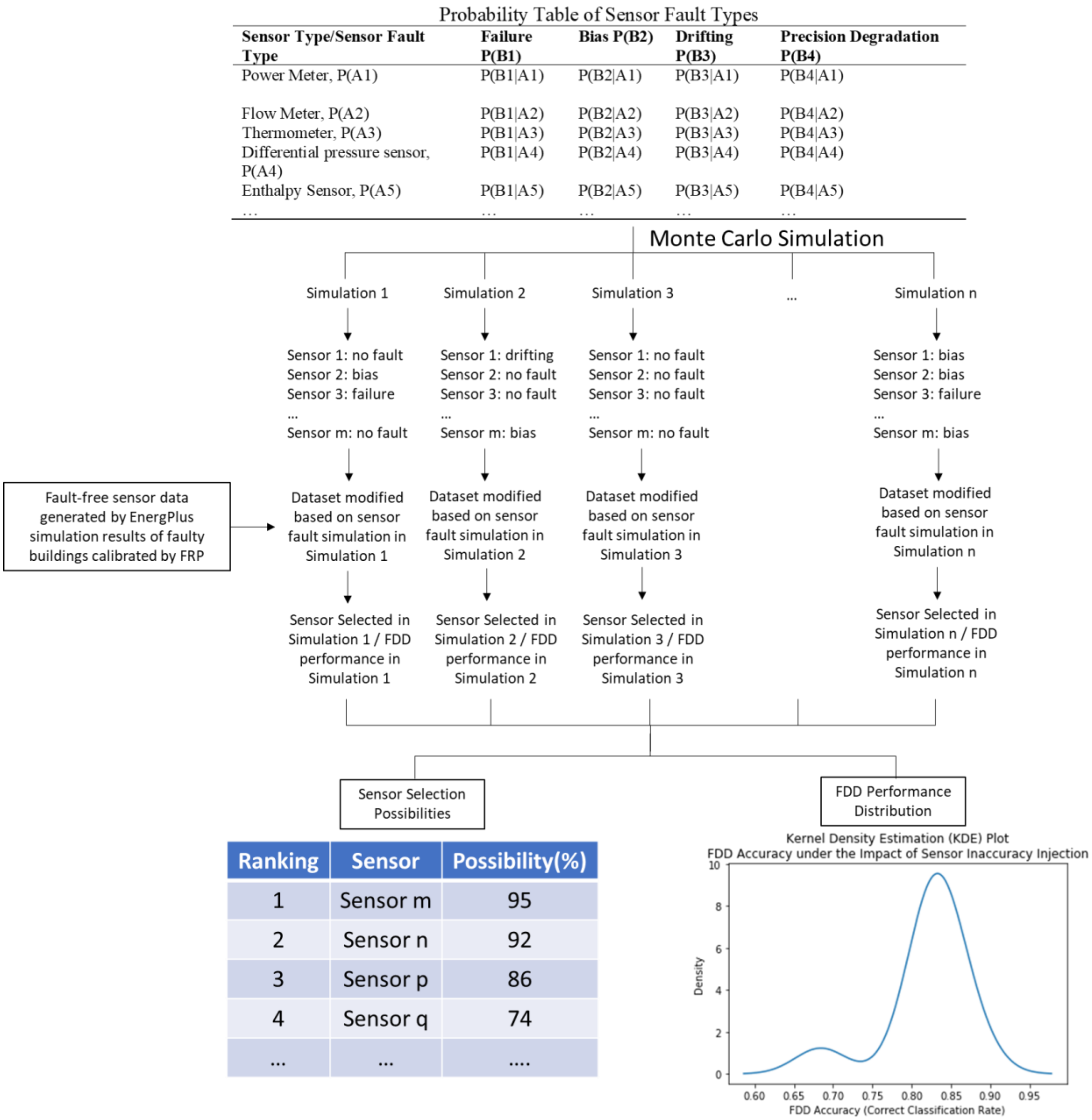
The fault rates P(A1), P(A2), etc. represent and quantify how robust each sensor type is. If P(A1) equals 0.1, that means that a power meter has a 10% chance to be faulty (for purposes of the sensor selection and FDD modeling process). More robust sensor types will have lower values; those prone to faulty operation will have higher values. For example, enthalpy sensors are generally less reliable, so P(A5) will have a relatively larger value than that for other sensor types. The probabilities P(B1), P(B2), etc. capture how likely certain sensor fault types are to occur. For example, the probability of complete failure P(B1) should be less than the probability of sensor bias P(B2), which is a very common sensor fault. Finally, conditional probabilities, in the form of P(Bn|Am), capture the likelihood of a specific fault type (bias, drifting, precision degradation, or complete failure) when a fault is present for a certain sensor type. For example, P(B1|A1) is the probability that a faulty power meter sensor is experiencing complete failure.

We can sample the probability distribution space represented in the probability table structure to simulate potential combinations of sensor set inaccuracy. Since the sampling is random, the exact combination of fault states varies for each simulation. In this methodology, Monte Carlo simulation is conducted to sample the probability distribution space sufficiently to comprehensively assess the impact of sensors faults on FDD performance. Monte Carlo method (or experiment) is a category of computational algorithms: it obtains numerical results by repeated random sampling. The underlying mechanism of Monte Carlo method is to use randomness to solve deterministic problems. It is often applied in mathematical and physical problems. It is widely applied in numerical integration, optimization, and generating draws from a probability distribution. It also helps people understand the impact of uncertainty and risk in forecasting.

Figure 2 shows how Monte Carlo simulation driven by a fault probability table captures the uncertainty associated with sensor selection and predicts the corresponding impact on FDD performance. For each iteration of the Monte Carlo simulation, a decisive sensor fault scenario is generated from the probability table and we select the most important sensors (e.g., top ten) selected as features by the FDD algorithm. The process is repeated n times, and after summarizing all n simulations, we can generate a probability table that summarizes the likelihood of each sensor being selected as a feature, and also a corresponding probability distribution of FDD performance.

It is worth mentioning that the key inputs of this methodology are the probability table of sensor fault happening rate on different sensor types and sensor fault types, and this probability table can be decided by two methods depending on how this methodology is used. If this methodology is used for conceptual and preliminary analysis, the probability table can just be assumed to any numbers that favor the analysis. For example, people can assume the probability in two experiment and compare the difference of the results. If this methodology is used for more detailed and realistic analysis, the fault rate for different sensor types and sensor fault types can be found in the manufacture manuals for each sensor.

This developed analysis methodology focuses on the net impact of fault states across a full sensor set. This method can provide a more comprehensive understanding of the impact of sensor inaccuracy on sensor selection. The next section details a case study that demonstrates the application of these two methods using a commercial building model calibrated to ORNL's FRP experimental facility



Probability Table of Sensor Fault Types

Sensor Type/Sensor Fault Type	Failure P(B1)	Bias P(B2)	Drifting P(B3)	Precision Degradation P(B4)
Power Meter, P(A1)	P(B1|A1)	P(B2|A1)	P(B3|A1)	P(B4|A1)
Flow Meter, P(A2)	P(B1|A2)	P(B2|A2)	P(B3|A2)	P(B4|A2)
Thermometer, P(A3)	P(B1|A3)	P(B2|A3)	P(B3|A3)	P(B4|A3)
Differential pressure sensor, P(A4)	P(B1|A4)	P(B2|A4)	P(B3|A4)	P(B4|A4)
Enthalpy Sensor, P(A5)	P(B1|A5)	P(B2|A5)	P(B3|A5)	P(B4|A5)
…	…	…	…	…

Ranking	Sensor	Possibility(%)
1	Sensor m	95
2	Sensor n	92
3	Sensor p	86
4	Sensor q	74
…	…	….

Figure 2. Diagram of how Monte Carlo simulation driven by a fault probability table predicts sensor selection and resulting FDD performance

3. Case Study: Flexible Research Platform (FRP)

This section details a case study conducted to demonstrate the analysis methodology introduced in Section 2. The virtual testbed that simulates faulty buildings and generates simulation data is introduced in Section 3.1. The detailed simulated faults are introduced in Section 3.2. The data generated by the faulty building simulations are introduced in Section 3.3. Details of the data-driven FDD modeling workflow are introduced in Section 3.4. The detailed settings for the Monte Carlo-based analyses are introduced in Section 3.5.

3.1. Virtual Testbed

In this case study, a virtual testbed simulates faulty building and generates simulation data. The virtual testbed is a EnergyPlus (Department_of_Energy 2013) model that is calibrated on Flexible Research Platform (FRP) located at Oak Ridge National Laboratory (Goldwasser et al. 2018; Im et al. 2016). Figure 3 shows an exterior view of the facility and a rendering of the virtual testbed. Figure 4 shows the building floorplan (left) and HVAC system layout (right). Details on the building constructions and characteristics of the testbed can be found in Table 1 of the paper (Im et al. 2016).

Figure 3. EnergyPlus virtual testbed (right), calibrated to FRP at Oak Ridge National Laboratory (left) (Im et al. 2016)

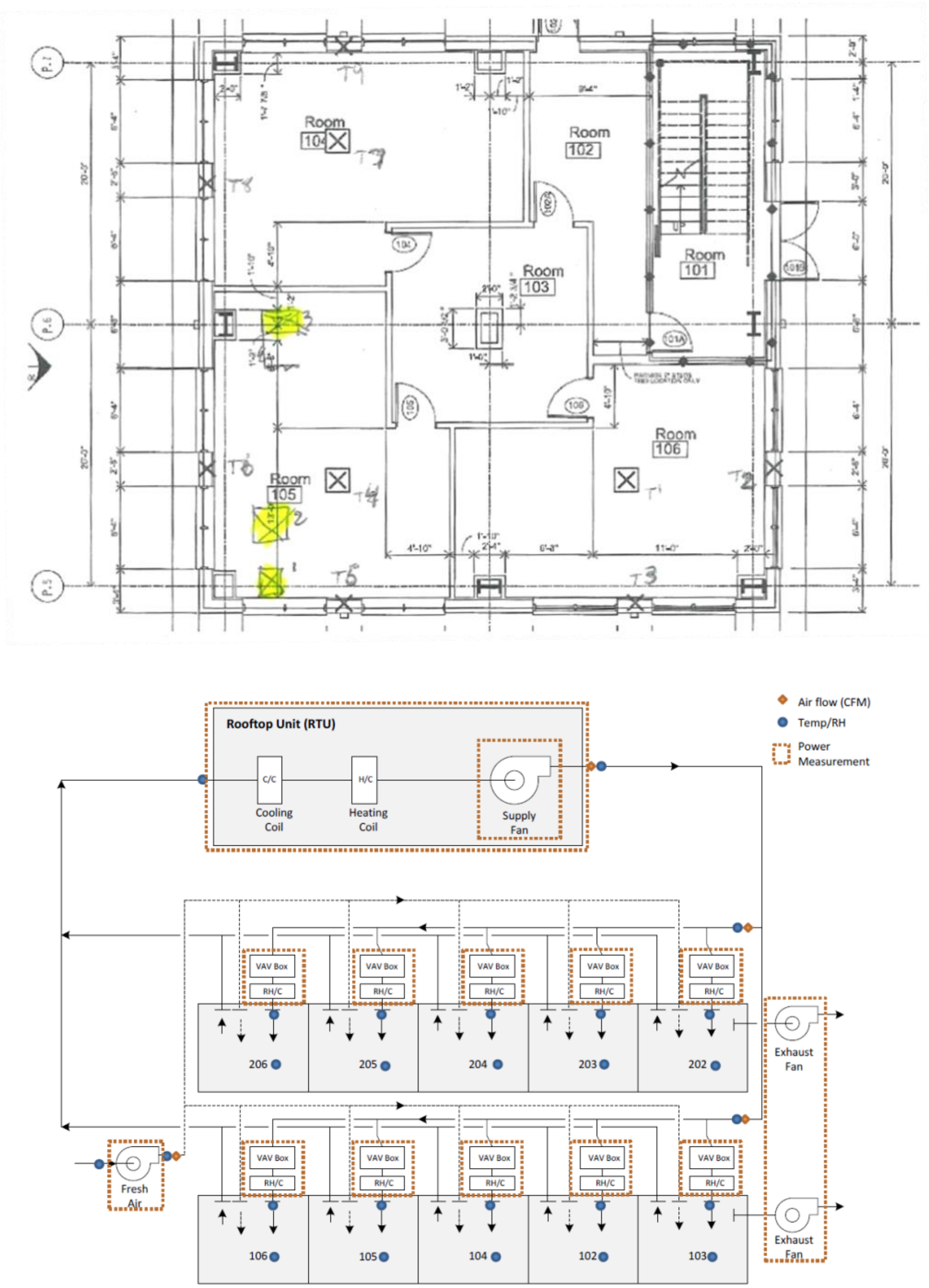


Figure 4. Floorplan and HVAC system layout of FRP (Source: (Im, Bhandari, and New 2016))

3.2 Building Faults Simulation

In order to generate data for data-driven FDD modeling and Monte Carlo simulation analysis, fourteen faults are modeled and simulated in the EnergyPlus model introduced in Section 3.1. The simulated faults include: (1) AHU fan motor degradation, (2) biased sensor: mixed air temperature, (3) Duct fouling, (4) HVAC setback error: delayed onset, (5) HVAC setback error: early termination, (6) HVAC setback error: no overnight setback, (7) Improper time delay setting in occupancy sensors, (8) Lighting setback error: delayed onset, (9) Lighting setback error: early termination, (10) Lighting setback error: no overnight setback, (11) Outdoor air damper stuck at certain position, (12) Return air duct leakages, (13) Supply air duct leakages, and (14) Thermostat measurement bias. Specifically, each fault is modeled as OpenStudio measure (https://github.com/NREL/OpenStudio-fault-models), and they are independently applied to the fault-free EnergyPlus model introduced in Section 3.1, generating altogether fourteen faulty building models. All the fourteen faulty building model and the fault-free model are later simulated under the typical meteorological year (TMY) weather file of Knoxville, Tennessee where FRP locates, to generate the data used for data-driven FDD modeling. More details about fault definitions, faulty model development, and EnergyPlus models are introduced in the study of (Kim, Frank, Braun, et al. 2019; Kim, Frank, Im, et al. 2019).

3.3. Training/Testing Data for Data-Driven FDD Model

The data-driven FDD model is trained with the virtual sensor data generated from the simulation of the 14 faulty building models and the fault-free building model introduced in Section 3.1 and Section 3.2. The data resolution of all simulations is 60 minutes. A total of 70 virtual sensors are considered in this study. These sensors have high likelihood of being incorporated into typical building automation systems. The sensors are listed in Table 2.

Table 2. Sensor list for the analysis in the case study

No.	Sensor Name
1	Electricity facility (W)
2	Cooling electricity (W)
3	Heating electricity (W)
4	Whole building facility total HVAC electric demand power (W)
5	Fan electricity (W)
6	Interior equipment electricity (W)
7	Interior lights electricity (W)
8	Gas facility (W)
9	Heating gas (W)
10	Diffuse solar radiation rate (W/m2)
11	Direct solar radiation rate (W/m2)
12	Outdoor air barometric pressure (Pa)
13	Outdoor air dry-bulb temperature (°C)
14	Outdoor air relative humidity (%)
15	Outdoor air wet-bulb temperature (°C)
16	Rain status
17	Rooftop cooling coil air outlet temperature (°C)
18	Rooftop discharge air temperature (°C)
19	Rooftop mixed air temperature (°C)
20	Rooftop heating coil air outlet temperature (°C)
21	Rooftop supply fan electricity (W)
22	Rooftop heating coil energy (W)
23-32	VAV reheat damper discharge air temperature (°C) (room 102, 103, 104, 105, 106, 202, 203, 204, 205, 206)
33-42	Supply reheat coil outlet air temperature (°C) (room 102, 103, 104, 105, 106, 202, 203, 204, 205, 206)
43-56	Zone mean air temperature (°C) (1F plenum, 2F plenum, room 101, 102, 103, 104, 105, 106, 201, 202, 203, 204, 205, 206)
57-70	Zone air relative humidity (%) (1F plenum, 2F plenum, room 101, 102, 103, 104, 105, 106, 201, 202, 203, 204, 205, 206)

To study the quantitative impact of the sensor fault types on FDD and building performance, quantification of sensor severity is needed. The error injected to the simulation data to emulate the happening of the sensor inaccuracy follows the parameters shown in Table 3. It is worth mentioning that the assumption and decision of the sensor fault intensities used in this study (e.g., 5% for bias and drifting) are based on common values of sensor bias and drifting in many other studies (Amritha and Banu).

Table 3. Parameters of sensor fault types to be considered in the case study

Sensor Fault Type	Parameter Definition	Values
Complete failure	Constant sensor reading	average of annual actual value
Bias	Absolute sensor value error	±5%
Drifting	Sensor value change rate per year	±5%/year
Precision Degradation	Random bias range	-5% to 5%

In the case study, correct classification rate is used to quantify the FDD performance. Correct classification rate (Frank et al. 2019) computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels.

3.4. Data-Driven FDD Model Settings

The data-driven FDD model is formulated as a multi-class classification problem to classify 15 classes (14 faulty classes and 1 fault-free class). The inputs of the classifier are the values of sensors that are listed in Table 2. The output of the classifier is the class label indicating the faulty type or faulty-free class. The data resolution for inputs and output is one hour.

We applied random forest as the data-driven algorithm for FDD modeling. The random forest algorithm is brought up by L. Breiman in 2001; it has been successful as a general-purpose classification and regression algorithm. Random forest algorithm combines several randomized decision trees and aggregates their predictions by averaging (Biau and Scornet 2016) and has shown excellent performance for FDD research (Frank et al. 2016; Li et al. 2016). In this study, we use scikit-learn (a Python package for machine learning (Pedregosa et al. 2011), and the version number is 0.24.2) to realize random forest modeling. Except for the number of trees, the random forest parameter values are set to default in the hyperparameter setting. The number of trees is set to 250, which is decided by a test that calculate the threshold which will no longer improve model accuracy when increasing the number of trees.

3.5 Settings of the Monte Carlo-Based Analysis Methodology

The probability table that captures the probability of each combination of sensor and fault type represents the key input to the Monte Carlo-based analysis. Specifying this set of inputs is critical to mapping out the resulting selection probabilities and corresponding FDD performance distribution. Extensive fault prevalence data would be required to accurately specify these inputs. For now, we applied the example set of inputs presented in Table 4 to evaluate the feasibility and potential of the analysis methodology introduced in Section 2. Our example set of inputs is simply defined as follows: (1) each sensor type has a 10% chance of faulted behavior, or the fault rate is 10%; and (2) faulted behavior is evenly attributed to the possible fault types (sensor failure, sensor bias, sensor drifting and sensor precision degradation). The reason we make the simple assumptions on the probability table is that we do not have enough practical information and domain knowledge on the actual probability of sensor fault. The information can be got from the user manual and detailed specifications of different sensors from different manufacturers which is related to more case-specific and time-consuming surveys and search: we want this paper to focus on developing this analysis methodology, instead of making the focus of this study to collect the sensor manufacture information.

Table 4. Presumed probability table of sensor inaccuracy

Sensor Type/ Sensor Fault Type	**Failure P(B1)**	**Bias P(B2)**	**Drift P(B3)**	**Precision Degradation P(B4)**
Electricity meter, P(A1) = 0.1	P(B1|A1) = 0.25	P(B2|A1) = 0.25	P(B3|A1) = 0.25	P(B4|A1) = 0.25
System node temperature sensor, P(A2) = 0.1	P(B1|A2) = 0.25	P(B2|A2) = 0.25	P(B3|A2) = 0.25	P(B4|A2) = 0.25

Room temperature sensor, P(A3) = 0.1	P(B1|A3) = 0.25	P(B2|A3) = 0.25	P(B3|A3) = 0.25	P(B4|A3) = 0.25
Energy meter, P(A4) = 0.1	P(B1|A4) = 0.25	P(B2|A4) = 0.25	P(B3|A4) = 0.25	P(B4|A4) = 0.25
Weather meter, P(A5) = 0.1	P(B1|A5) = 0.25	P(B2|A5) = 0.25	P(B3|A5) = 0.25	P(B4|A5) = 0.25
Room humidity sensor, P(A6) = 0.1	P(B1|A6) = 0.25	P(B2|A6) = 0.25	P(B3|A6) = 0.25	P(B4|A6) = 0.25
Gas meter, P(A7) = 0.1	P(B1|A7) = 0.25	P(B2|A7) = 0.25	P(B3|A7) = 0.25	P(B4|A7) = 0.25

The number of runs for Monte Carlo simulation is set to 1000. For each run, a new machine learning model needs to be retrained; the resulting computational cost is high. With more computing power, we can increase the number of runs to pursue a more accurate result, but for now, the results are based on 1000 runs.

4. Results and Discussion

This section presents the results of the case study demonstration (introduced in Section 3) of the Monte Carlo-based analyses (introduced in Section 2). Final results for sensor selection possibilities and FDD performance distribution are presented and summarized in this section.

As detailed in Section 2.1, the output of the Monte Carlo-based analysis consists of two parts: the sensor selection possibilities and the corresponding FDD performance distribution. Table 5 shows the 20 sensors with the highest probability of selection across the fault probability space mapped out by the Monte Carlo-based analysis. We can see from the results that: (1) sub-system energy meters, including interior lights electricity, rooftop heating coil heating energy, gas facility, heating gas, fans electricity, heating electricity, and cooling electricity, are very important features that have high possibility to be selected under faulty sensor scenarios; (2) HVAC equipment outlet temperatures, including room VAV reheat damper outlet system node temperature, rooftop heating coil outlet system node temperature, and rooftop cooling coil outlet system node temperature, are also important; (3) room temperature is relative less important features in this FDD model, and (4) relative humidity sensors are least important sensors in the sensor selection result. It is worth mentioning that the sensor selection result shown in Table 5 is only applicable to the case study. Basically, we provide a generic methodology that can be applicable to different building types, FDD data-driven algorithms, sensor types, etc., and the feature selection results are different from case to case. To sum up, the table with possibility of each sensor being selected (Table 5) can reflect the uncertainty of feature selection process considering sensor inaccuracy. Instead of showing the decisive sensor selection results, the output of this methodology indicates the probabilities that can help people better understand the sensor inaccuracy impact and decide the sensors for FDD purpose.

Table 5. Sensors with highest probability to be selected

Ranking	**Sensor**	**Selected Possibility (%)**
1	Interior lights electricity [W]	82.9
2	Rooftop heating coil heating energy [W]	80.1
3	Outdoor air barometric pressure [Pa]	79.2
4	Gas facility [W]	78.4
5	Heating gas [W]	78.4
6	Room 106 vav reheat damper outlet system node temperature [C]	77.6
7	Rooftop supply fan electric energy [W]	77.6
8	Fans electricity [W]	76.8
9	Room 203 zone mean air temperature [C]	76.0
10	Heating electricity [W]	75.2
11	Rooftop heating coil outlet system node temperature [C]	75.2
12	Cooling electricity [W]	72.8
13	Rooftop cooling coil outlet system node temperature [C]	72.0
14	Room 203 vav reheat damper outlet system node temperature [C]	71.2
15	Room 206 vav reheat damper outlet system node temperature [C]	71.2
16	Room 103 vav reheat damper outlet system node temperature [C]	71.2
17	Room 104 zone mean air temperature [C]	71.2
18	Room 202 zone mean air temperature [C]	68.0
19	Room 202 vav reheat damper outlet system node temperature [C]	67.2
20	Room 105 vav reheat damper outlet system node temperature [C]	66.4

Figure 5 shows the kernel density estimation (KDE) plot of FDD accuracy. The KDE plot represents the probability distribution of FDD accuracies for the fault space mapped out by the Monte Carlo-based analysis. We can see from this figure that with the uncertainty of sensor inaccuracy, FDD accuracy decreases from 0.950 (original FDD model without sensor faults) to an average of 0.905, and the accuracy can be further decreased below 0.850 in some cases. This plot is intuitive and quantitative to reflect how sensor inaccuracy can affect the FDD performance uncertainty.

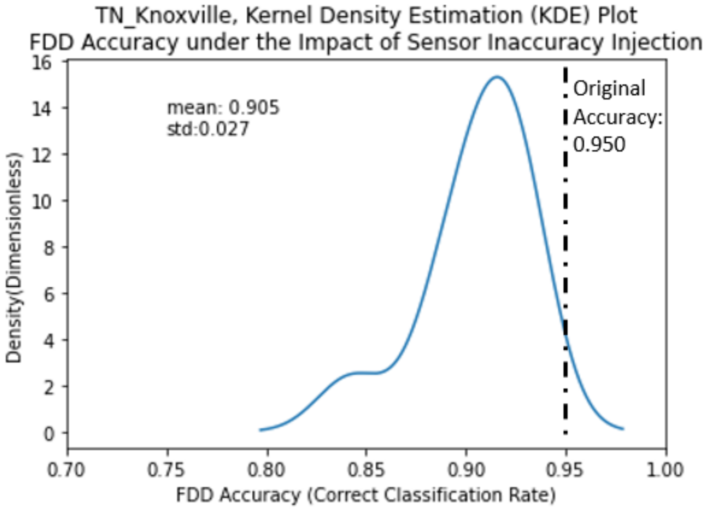


Figure 5. FDD performance distribution for the Monte Carlo-based analysis

5. Conclusions

In this study, we explore the impact of sensor accuracy on sensor selection and model accuracy for data-driven building energy system FDD. Monte Carlo-based analysis methodology is developed and applied to evaluate this impact. This methodology is based on a probability table for each sensor type and sensor fault type that defines the fault input space. Based on the table, Monte Carlo simulation are executed to get a large number of decisive results and calculate selection probabilities for the most impactful sensors and generate a corresponding FDD performance probability distribution. This analysis methodology provides a sense of how robust a collective sensor set is to stochastic sensor error, as well as how sensitive overall FDD performance is to faulty or poorly calibrated sensor data.

The developed methodology can be applied to evaluate the interaction of sensor accuracy and sensor selection in both early-stage sensor design and operation-stage sensor improvement and maintenance. The methodology defines the sensor fault uncertainty as a whole to comprehensively evaluate the sensor inaccuracy. The case study in this paper demonstrates the effectiveness of the developed methodology in a commercial building model calibrated to the FRP experimental test facility. Results demonstrate the feasibility of the method and highlight its advantages: (1) it provides quantitative analysis on the uncertainty of multiple sensors inaccuracy on FDD performance, (2) it provides insights on the sensor selection decisions in sensor design and sensor improvement for FDD purpose, and (3) it is useful in the applications where the interactions among multiple sensor faults cannot be ignored or need to be quantified. It is even applicable when the sensor inaccuracies are unknown and needs assumptions.

In the future, more detailed content related to the decision of the probability table which is the input of this method can be more discussed. Second, the faulty intensity of sensors is not considered as a variable in this study; in the future, the fault intensity can also follow some probability distribution to reflect more detailed uncertainty scenarios. Third, the number of runs of Monte Carlo simulation is currently decided by the maximum computing power of the computer we use for the simulation: a more systematic way to decide the minimum number of runs to ensure robust results is a future research direction.

Acknowledgments

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Reference

Amritha, C, and U Sabura Banu. 'Detection of Bias, Drift, Freeze and Abrupt Sensor Failure using Intelligent Dedicated Observer Based Fault Detection and Isolation for Three Interacting Tank Process'.

Biau, Gérard, and Erwan Scornet. 2016. 'A random forest guided tour', Test, 25: 197-227.

Chen, Youming, and Lili Lan. 2010. 'Fault detection, diagnosis and data recovery for a real building heating/cooling billing system', Energy Conversion and Management, 51: 1015-24.

Dai, Xuewu, and Zhiwei Gao. 2013. 'From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis', IEEE Transactions on Industrial Informatics, 9: 2226-38.

Department_of_Energy, U.S. 2013. 'EnergyPlus Energy Simulation Software, Version 8.1. ', Washington, DC: U.S. Department of Energy. https://energyplus.net/.

Du, Zhimin, Bo Fan, Xinqiao Jin, and Jinlei Chi. 2014. 'Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis', Building and Environment, 73: 1-11.

Frank, Stephen, Michael Heaney, Xin Jin, Joseph Robertson, Howard Cheung, Ryan Elmore, and Gregor Henze. 2016. 'Hybrid model-based and data-driven fault detection and diagnostics for commercial buildings', Proceedings of the 2016 ACEEE Summer Study on Energy Efficiency in Building: 12-1.

Frank, Stephen M, Guanjing Lin, Xin Jin, Rupam Singla, Amanda Farthing, Liang Zhang, and Jessica Granderson. 2019. "Metrics and methods to assess building fault detection and diagnosis tools." In.: National Renewable Energy Lab.(NREL), Golden, CO (United States).

Goldwasser, David, Brian L Ball, Amanda D Farthing, Stephen M Frank, and Piljae Im. 2018. "Advances in Calibration of Building Energy Models to Time Series Data." In.: National Renewable Energy Lab.(NREL), Golden, CO (United States).

Han, H, B Gu, T Wang, and ZR Li. 2011. 'Important sensors for chiller fault detection and diagnosis (FDD) from the perspective of feature selection and machine learning', International Journal of Refrigeration, 34: 586-99.

Im, Piljae, Mahabir Bhandari, and Joshua %J Oak Ridge New, TN. 2016. 'Multi-Year Plan for Validation of EnergyPlus Multi-Zone HVAC System modeling using ORNL’s Flexible Research Platform'.

Im, Piljae, Mahabir S Bhandari, Joshua Ryan %J Oak Ridge National Lab New, Building Technologies Research, and TN Integration Center : Oak Ridge, USA. 2016. 'Multiyear Plan for

Validation of EnergyPlus Multi-Zone HVAC System Modeling using ORNL’s Flexible Research Platform'.

Katipamula, Srinivas, and Michael R Brambley. 2005. 'Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I', Hvac&R Research, 11: 3-25.

Kim, Janghyun, Stephen Frank, James E Braun, and David Goldwasser. 2019. 'Representing small commercial building faults in energyplus, Part I: Model development', Buildings, 9: 233.

Kim, Janghyun, Stephen Frank, Piljae Im, James E Braun, David Goldwasser, and Matt Leach. 2019. 'Representing Small Commercial Building Faults in EnergyPlus, Part II: Model Validation', Buildings, 9: 239.

Li, Dan, Yuxun Zhou, Guoqiang Hu, and Costas J Spanos. 2016. 'Fault detection and diagnosis for building cooling system with a tree-structured learning method', Energy and buildings, 127: 540-51.

Lin, Guanjing, Hannah Kramer, and Jessica Granderson. 2020. 'Building fault detection and diagnostics: Achieved savings, and methods to evaluate algorithm performance', Building and Environment, 168: 106505.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. 2011. 'Scikit-learn: Machine learning in Python', Journal of machine learning research, 12: 2825-30.

Wang, Shengwei, Qiang Zhou, and Fu Xiao. 2010. 'A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults', Energy and Buildings, 42: 477-90.

Yuwono, Mitchell, Ying Guo, Josh Wall, Jiaming Li, Sam West, Glenn Platt, and Steven W Su. 2015. 'Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems', Applied Soft Computing, 34: 402-25.

Zhang, Liang, Stephen Frank, Janghyun Kim, Xin Jin, and Matt Leach. 2020. 'A systematic feature extraction and selection framework for data-driven whole-building automated fault detection and diagnostics in commercial buildings', Building and Environment, 186: 107338.

Zhang, Liang, Stephen Frank, Janghyun Kim, Xin Jin, Matt %J Building Leach, and Environment. 2020. 'A systematic feature extraction and selection framework for data-driven whole-building automated fault detection and diagnostics in commercial buildings', 186: 107338.

Zhang, Liang, Jin %J Energy Wen, and Buildings. 2019. 'A systematic feature selection procedure for short-term data-driven building energy forecasting model development', 183: 428-42.

Zhang, Liang, Jin Wen, Yanfei Li, Jianli Chen, Yunyang Ye, Yangyang Fu, and William %J Applied Energy Livingood. 2021. 'A review of machine learning in building load prediction', 285: 116452.

Zhao, Yang, Shengwei Wang, and Fu Xiao. 2013. 'Pattern recognition-based chillers fault detection method using support vector data description (SVDD)', Applied Energy, 112: 1041-48.