In-situ sensor calibration for building HVAC systems with limited information using general regression improved Bayesian inference

Guannan Li[1,4,5,6], Jiahao Xiong[1], Rui Tang[2*], Shaobo Sun[3], Chongchong Wang[1]

[1] School of Urban Construction, Wuhan University of Science and Technology, Wuhan 430065, China
[2] Institute for Environmental Design and Engineering, The Bartlett, University College London, London, UK
[3] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong
[4] Key Laboratory of Low-grade Energy Utilization Technologies and Systems (Chongqing University), Ministry of Education of China, Chongqing University, Chongqing 400044, China
[5] State Key Laboratory of Green Building in Western China, Xian University of Architecture & Technology, Xian 710055, China
[6] Hubei Provincial Engineering Research Center of Urban Regeneration, Wuhan University of Science and Technology, Wuhan 430065, China

Abstract Sensors in building heating, ventilation and air-conditioning systems (HVACs) play important roles in maintaining indoor environmental quality and energy consumption. Owing to the repeatedly varied outdoor working environment and indoor users' demand, sensor faults could be inevitable in the lifespan. To allow HVACs worked at fault-tolerant way, previous studies developed the in-situ sensor calibration method via energy conservation equations and Bayesian inference (EC-BI). However, the practical application may encounter challenges like limited-variable information, low-quality data and increasing risks of calibration uncertainty by indirect information supplement. These cause increasing in-situ calibration complexity and modeling costs. To address these challenges, this study proposed a general regression improved Bayesian inference (BI) in-situ sensor calibration strategy. The multiple linear regression (MLR) was utilized as a typical example of general regression method to improve the BI method. The proposed MLR-BI method was validated using both simulated and practical data of two building HVAC systems in two case studies. The principle component analysis (PCA)-based sensor fault reconstruction method was used for comparison under five fault conditions covering both single and simultaneous faults. Five variable scenarios were considered to validate the effectiveness of MLR-BI on HVACs with the limited variable information. Results indicated that the calibration accuracy of MLR-BI is over 99% under four conditions of the simulated case 1, which is about 6% and 8% higher than PCA and EC-BI respectively. For all the three variable scenarios of the simulated case 1, the calibration accuracy of MLR-BI is 99.65% on average. Especially in the four-variable scenario with limited variable information, MLR-BI shows the average calibration accuracy of 99.75% while PCA obtains 79.46% and EC-BI fails to work because of variable limitation. For the fault condition of the limitted-variable practical case 2, MLR-BI still outperforms the other two and obtains 97.1% calibration accuracy in two practical scenarios.

Keywords: Building systems; Heating, ventilation and air-conditioning (HVAC); In-situ sensor calibration; Bayesian inference (BI); Multiple linear regression (MLR)

* Corresponding author: Rui Tang (e-mail: rui.tang@ucl.ac.uk).

Nomenclature	
AHU	Air handling unit
BES	Building energy system
BI	Bayesian inference
CAV	Constant-speed air volume
EC	Energy conservation
EC-BI	Calibration method of EC equations and Bayesian inference
EVIC	Extended virtual in-situ sensor calibration
GMM	Gaussian mixture model
HVACs	Heating, ventilation and air-conditioning systems
MAPE	Mean absolute percentage error
MAT	Mixed air inlet temperature sensor
MCMC	Markov chain Monte Carlo
MFR	Air mass flow rate sensor
MLR	Multiple linear regression
MLR-BI	Calibration method of multiple linear regression and Bayesian inference
PCA	Principle component analysis
PDF	Probability density function
RE	Relative error
VIC	Virtual in-situ sensor calibration
VVIC	Virtual sensor-assisted in situ sensor calibration
SAT	Air supply temperature sensor
α_0,β_0	Constant term of the MLR model
$\alpha_1 - \alpha_p$,$\beta_1 - \beta_q$	Coefficients of corresponding input variables
C_w	Specific heat capacity of water
d	Humidity ratio
$D()$	Distance function
f	Fault amplitude
$g_c()$	Compensation function
h	Enthalpy
M	Flow rate
O	Original measurement of the target sensor
$P(x\|Y)$	Posterior distribution
$P(Y)$	Normalization constant
$P(Y\|x)$	Likelihood function

P_{load}	Part load ratio
Q	Value of Q statistic
Q_{rec}	Value of Q statistic calculated by reconstructed data
T	Temperature
u_r	Value of unknown variables in the model after compensation
$V_1 - V_p$	Physical sensors except the target sensor to be calibrated
$V_1' - V_q'$	Target sensors to be calibrated
V_{ca}	Calibration value after compensating
W	The total power input
x	Compensation value
$\hat{x}$	Residual vector
X_{me}	Measurement data
$\hat{X}_{me}$	Residual of measured data
X_{rec}	Reconstructed data
$\hat{X}_{rec}$	Residual of reconstructed data
Y_c	Calibration value of the target sensor
$\hat{y}_{ca,i}$	Calibrated value of the target variable
Y_{Bse}	Benchmark of the sensor term
Y_{Bsy}	Benchmark of the system term
y_i	True value of the target variable
$y_{rel,n}$	Value of model related variables after compensation
Y_{Sme}	Measured value of the system model
ΔE	Heat exchange
$\pi(x)$	Prior distribution of x
$\vec{\mu}$	Unit vector
$\hat{\vec{\mu}}^+$	Moore–Penrose pseudo inverse of the unit vector
σ	Standard deviation of the prior distribution
ξ_{ca}	Calibration accuracy
φ	Relative humidity
a	Air side of cooling coil
ca	Calibrated value
co	Condenser
ev	Evaporator
rec	PCA reconstruction
i	ith sample of testing dataset

in	The inlet of evaporator and condensor
l	Number of system models
mix	The inlet of cooling coil air side
m	Number of sensors
n	Total number of testing data sample
out	The outlet of evaporator and condensor
p	Number of variables in the MI R model
q	Number of target variables in the MIR model
ret	The inlet of cooling coil water side
sup	Outlet including cooling coil air side and water side outlet
w	Water side of cooling coil

1 Introduction

1.1 Background

Surveys indicate that the building sector accounts for nearly 40% of the global energy consumption [1, 2], of which 60% is consumed by the heating, ventilation and air conditioning (HVAC) system in buildings [3]. HVACs are of importance to maintain not only indoor thermal comfort but also indoor environmental quality (i.e., indoor air temperature and humidity, CO_2 concentration) [4]. With the development of artificial intelligence and big data technologies, many data-driven techniques have been developed to achieve building energy-savings and improve building performance, such as in-situ modeling method [5], energy model calibration [6, 7], demand and load prediction [8, 9], system fault diagnosis [10, 11], control and operational optimization [12] for building HVAC systems. These techniques can decipher the information and establish the correlations between interested states from operational data [13], which can benefit HVAC operational management and performance optimization. Undoubtedly, the performance of these data-driven techniques depends heavily on the reliability and accuracy of the data used, which are measured by the sensors in HVACs.

1.2 Summary of sensor calibration studies in building HVAC systems

Owing to the repeatedly changed indoor environmental demands and outdoor weather conditions,

the building cooling and heating loads estimation could be difficult to achieve very high accuracy but with high forecast uncertainty. To match the time-varied building demands, the HVAC system may suffer great risks of abnormal operations, which causes significant energy waste in the long-term service [14]. Sensors in the system could also experience faults, i.e., error reading, bias, drifting and even complete failure [15]. These sensor faults could result in improper control strategies and the consequent system operation deviation from the expected. Additionally, the current data-driven methods for HVAC performance optimization may not work well if there is error measurements caused by various sensor faults. As a common type of HVAC system fault [16, 17], sensor fault includes two main categories: (1) hard fault that causes complete sensor failure due to structural damage and (2) soft fault that leads to sensor performance degradation due to improper installation and changed environment [18]. Many sensors are mounted in the HVAC system [19]. The conventional sensor calibration method [7, 20, 21], which compares the standard with the measured value, may only apply to the hard fault. As for the soft fault, advanced data-driven or statistical inference based in-situ sensor calibration methods can reduce time and cost while increase the efficiency [22].

Many researchers have dedicated their efforts for in-situ sensor calibration with soft faults in the building systems. Yu and Li [22] firstly proposed a virtual in-situ sensor calibration method (VIC), which developed in-situ benchmark sensors and evaluated the calibration results through statistical or model-based methods. Bayesian inference (BI), as a common data inference method based on statistics, has been widely used in various calibration problems of complex building energy systems because of its convenient calculation and few parameters to be determined. Yoon and Yu [23] developed an BI-based extended VIC method (EVIC) for the LiBr-H_2O absorption refrigeration system. It can process a large amount of variable information in the building system and greatly improve calibration efficiency. Yoon and Yu [24, 25] carried out quantitative comparison between BI and genetic algorithm for sensor calibration in the LiBr-H2O absorption refrigeration system. Also, the various hidden factors and corresponding complementing strategies [26-28] were seriously considered. Furthermore, Wang et al. [29] proposed a sensitivity coefficient optimization method to promote the development of an automated reviving calibration strategy. For the air handling unit (AHU), Wang et al. [30] verified the the VIC method under six normal and four extreme operating conditions. Zhao et al. [31] proposed a

Gaussian mixture model (GMM) to preprocess historical data and obtain steady-state measurement values under various operating conditions so that eliminate the impact of dynamic data on the calibration results of VIC method can be eliminated. Li et al. [19] investigated the influences of different calibration models on calibration accuracy. The models are constructed under different system regions. Choi and Yoon [32] proposed a virtual sensor-assisted in situ sensor calibration method (VVIC), which solved the problem of insufficient variable information when the VIC method was used to construct the calibration model. For the R-410A unitary air-conditioners, Yoon et al. [33] determined the balance error of refrigeration capacity between the air side and the refrigerant side caused by the measurement error with VIC, and improved the system performance by calibrating the measurement error.

In these studies, energy conservation (EC) is often used as the basis for modeling, which can be collectively referred as the EC-BI method. For example, the calibration model of the LiBr-H_2O absorption refrigeration system is usually constructed based on the conservation of the state enthalpy difference [23-29]. For the AHU [19, 30-32] and the R-410A unitary air-conditioners [33], the calibration model was usually constructed in accordance with the heat transfer balance between the air side and the water side of the heat exchanges. In addition, BI has been applied to calibrate sensors in other building energy systems. Mokhtari et al. [34] used Bayesian inference to calibrate the wind speed sensor in the cooling tower of a thermal power plant. The calibration model was constructed based on the equipment delivery instructions of the cooling tower. Sun et al. [35] used BI to quantify the flow uncertainty of a central cooling system with multiple chillers. Moreover, Sun et al. [36] proposed an online robust sequencing control based on the quantification results of flow uncertainty. The calibration model was construced based on the EC equations between the chillers and the cooling towers.

1.3 Challenges

Although the EC-BI method has been widely studied in the in-situ calibration of sensors in building HVAC systems, there are still some challenges for practical applications. First of all, since the construction of calibration model is usually based on EC equations, the information of key variables for developing calibration model are often relatively limited in practice due to sensor cost or technical

constraints. To overcome this problem, the VVIC method was proposed by Choi and Yoon [32]. The construction of virtual sensors can be regarded as an indirect information supplement method, which could not directly participate in the development of calibration model. More modeling deviations may be generated by more calculation steps. Secondly, due to the unfavorable working environment, signal transmission problem, and the dynamic characteristics of observation phenomena, the original sensor measured data tend to be noisy and incomplete [37] in practical HVACs. This may cause unreliable calibration models based on EC equations and further lead to deviation of the calibration result. Lastly, practical HVACs consist of complex and diverse components, such as chillers, cooling towers, fan coils, etc. For the sensors in different component, different EC equations should be employed to develop the BI-based in-situ calibration models. After sensor fault detection and diagnosis, it is time-consuming and laborious to find the suitable EC equations for calibrating the faulty sensors. If some components with long service life have lost the relevant manufacturing instructions, the EC-BI method may fail to work owing to the information loss. If only the EC equations can be used to establish the BI calibration model, the adverse influences of limited variable information or low-quality data can be significant. The main challenges of applying the in-situ EC-BI based sensor calibration method in practical HVAC systems can be summarized as follows:

(1) There could be information limitation constructing the in-situ EC-BI based sensor calibration model, and the propagation of estimation uncertainty of the virtual sensor assisted indirect information supplement method may enhance the risk of calibration uncertainty.

(2) When the building HVAC system practical operational data are used, sensor inherent mearured errors and noise may cause unexpected deviations of EC equations [38], which may affect the model reliability of the in-situ EC-BI based calibration method.

(3) In the practical HVAC system, if there is lack of some critical manufacturing instructions and the detailed component descriptions for EC equations, the in-situ EC-BI model may be infeasible. Even if EC equations become realizable, the EC-BI calibration could be time-consuming and laborious and challenges (1) and (2) still remain to be solved.

1.4 Research contents and contributions of this study

Therefore, it is necessary to develop a pure data-driven general regression method to overcome the modeling requirements based on the laws of physics, for coping with the above challenges in practical HVACs. The method should effectively control the modeling cost and ensure calibration performance when calibrating key system sensors. As a typical regression method, multiple linear regression (MLR) is calculation-convenient, reliable and easy-to-realize. MLR has been used to solve the modeling problem of HVAC systems [39, 40]. It also satisfied the purposes of improving calibration while reducing calibration costs. Hence, this study introduced MLR as an example to develop the proposed general regression improved BI calibration method (MLR-BI) and further addressed the aforementioned challenges in three main aspects as follows:

(1) For both information-poor and information-rich HVACs in practice, MLR-BI can construct an in-situ sensor calibration model with high accuracy using only the built-in physical sensors measurement data of the practical HVACs. For variable-limitation scenarios, the increasing uncertainty risk by indirect information supplement can be alleviated since the calibration models are constructed no longer requiring the supplemental calculation processes.

(2) The MLR-BI method can obtain reliable calibration model as long as the general regression model (MLR in this study) achieves relatively high fitness and low-level regression residuals since the pure data-driven calibration model is constructed using the practical data of the HVAC systems.

(3) As the physical law of EC equations are no long the hard modeling requirement, the MLR-BI in-situ sensor calibration method can be constructed with sufficient operational data which greatly broaden the practical application by reducing the model complexity and cost.

To evaluate the effectiveness of the general regression improved Bayesian inference calibration method, this study validated the proposed MLR-BI methods in two case studies, i.e.,the simulated chiller-AHU and the practical chiller system. Both the single and simultaneous fault conditions were investigated. The other two commonly used calibration methods (PCA-based reconstruction [41] and EC-BI based calibration[32]) were selected for performance comparison. Besides, five variable scenarios were considered to simuate the information-rich simulated scenario, information-poor practical scenario and the extremely limited variable scenario of practical HVAC systems, to validate

that MLR-BI can still guarantee relatively good calibration performance with limited variable information.

2 Background knowledge

2.1 Principle of PCA-based reconstruction

PCA has been widely applied for sensor fault detection, diagnosis and reconstruction [42]. The Q statistic is usually used as the judgment standard for sensor fault detection and reconstruction. Q statistic represents the square of residual vector $\hat{x}$ projected on each dimension in the residual space, as shown in Eq. (1). According to the minimization of Q statistic principle [43], the faulty sensors measurement data can be reconstructed. The reconstructed data equal to the measured data minus the calculated fault amplitude, as shown in Eq. (2). To obtain the optimal reconstructed data, Q_{rec} ($Q \geq 0$) = 0 should be adopted as the optimal solution condition to solve the optimal fault amplitude, as shown in Eqs. (3) - (4). The optimal solution of the reconstructed data can be obtained by substituting Eq. (4) into Eq. (2), as shown in Eq. (5).

$$Q = \sum_{i=1}^{n} \hat{x}^2 \tag{1}$$

$$X_{rec} = X_{me} - \vec{\mu} f \tag{2}$$

$$Q_{rec} = \sum_{i}^{n} \hat{X}_{rec}{}^2 = \sum_{i}^{n} \left(\hat{X}_{me} - \hat{\vec{\mu}} f\right)^2 = 0 \tag{3}$$

$$f = \hat{\vec{\mu}}^{+} X_{me} \tag{4}$$

$$X_{rec} = (I - \vec{\mu}\hat{\vec{\mu}}^{+})\, X_{me} \tag{5}$$

where, Q represents Q statistic, $\hat{x}$ is the residual vector. n is the total number of samples, i is the number of data within the statistical range of samples. X_{rec} is the reconstructed data, X_{me} is the measurement data (usually the faulty data). $\vec{\mu}$ is the unit vector, f is the fault amplitude. Q_{rec} represents Q statistic calculated by reconstructed data, $\hat{X}_{rec}$ is the residual of reconstructed data, $\hat{X}_{me}$ is the residual of measured data. $\hat{\vec{\mu}}^{+}$ is the Moore–Penrose pseudo inverse of the unit vector. I is the identity matrix [43].

2.2 Principle of the calibration method based on Bayesian inference and energy conservation (EC-BI)

2.2.1 Principle of Bayesian inference (BI)

Bayesian inference has been successfully applied for sensor in-situ calibration [23]. Its basic mathematical descriptions are shown in Eqs. (6) - (8). BI updatas the prior probability with new information. Then prior is converted into posterior. The calibration problem is transformed into the optimization problem of minimizing the distance function, as shown in Eq. (8).

$$P(x|Y) = \frac{P(Y|x)\times\pi(x)}{P(Y)} \tag{6}$$

$$P(Y) = \int P(Y|x)\pi(x)dx \tag{7}$$

$$P(Y|x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}D(x)\right] \tag{8}$$

where, x is the compensation value that is eventually converted into the mean of posterior distribution. It is equivalent to the calibration result. $\pi(x)$ is the prior distribution of x. $P(x|Y)$ is the posterior distribution. $P(Y|x)$ is the likelihood function. $P(Y)$ is the normalization constant. σ is the standard deviation of the prior distribution.

It is very difficult to obtain the integral formula for $P(Y)$ directly. Often, the prior distribution is defined as normal distribution according to the central limit theorem [44], and the Markov chain Monte Carlo (MCMC) method is used to calculate the $P(Y)$ [45]. The Metropolitan Hastings algorithm is a widely used MCMC sampling method [46, 47], which can be used to obtain posterior distribution samples of x in the BI calibration model. The statistical characteristics of posterior distribution (mean, standard deviation, etc.) can further be obtained by samples.

2.2.2 Construction of EC-BI calibration model

From Section 2.2.1, it can found that the distance function plays an important role in the BI-based in-situ sensor calibration method. Previous studies [23, 32] employed the EC equations as bases to establish the distance function for calibration, which is the so-called EC-BI calibration method. The specific mathematical descriptions of EC-BI are shown in Eqs. (9) - (12). As shown in Eq. (9), the

calibration model includes two terms, the system term and the sensor term. The system term represents the difference between the measurements of the practical system where the target sensor is located and the benchmarks of the reliable system. Generally, the reliable system, which does not contain target sensor, is selected based on the EC equations. For example, the heat transfer between the air side and the water side of the heat exchanges. Given the target supply air temperature sensor on the air side cooling coils, the heat exchanges of water side is used as the benchmark of the reliable system. The sensor term represents the difference between the benchmark calculated by the single sensor model and the calibrated value calculated by the compensation function as shown in Eq. (12). The benchmark of the system model and the benchmark of the sensor model are shown in Eqs. (10) and (11), respectively. The related variable $y_{rel,n}$ and the unknown variable u_r in H_1 and H_2 need to be compensated through the compensation function $g_c(x)$.

$$D_{EC}(x) = \underbrace{\sum_l^L \left(Y_{Bsy,l} - Y_{Sme,l}\right)^2}_{\text{system term}} + \underbrace{\sum_m^M \left(Y_{Bse,m} - Y_{c,m}\right)^2}_{\text{sensor term}} \tag{9}$$

$$Y_{Bsy} = H_1\left(y_{rel,1}, y_{rel,1}, \dots, y_{rel,n}, u_1, u_2, \dots, u_r\right) \tag{10}$$

$$Y_{Bse} = H_2\left(y_{rel,1}, y_{rel,1}, \dots, y_{rel,n}, u_1, u_2, \dots, u_r\right) \tag{11}$$

$$Y_c = g_c(O, x) \tag{12}$$

where, $D_{EC}(x)$ is the distance function constructed based on EC equations. Y_{Bsy} is the benchmark of the system term, Y_{Bse} is the benchmark of the sensor term. l is the number of system terms, m is the number of sensors. Y_{Sme} is the measured value of the system term. $y_{rel,n}$ is the value of model related variables after compensation. u_r is the value of unknown variables in the model after compensation. Y_c is the calibration value of the target sensor, and O is the original measurement of the target sensor.

3 Proposed general regression improved Bayesian inference method for HVAC in-situ sensor calibration

Regression method is often used to solve building HVAC modeling problems including energy consumption prediction[39], virtual sensors based fault diagnosis[40]. The MLR-BI method can obtain reliable calibration model as long as the general regression model achieves relatively high fitness and low-level regression residuals. As described in the Section 1.3 challenges, EC equations may fail to

work for practical limited variable senarios for real-world HVAC systems. Owing to its high-precision data fitting capacity, regression methods can be utilized to replace the EC equations to establish the distance function. To address the challenges, this study proposed the general regression improved Bayesian inference method for HVAC in-situ sensor calibration. This section chooses the typical multiple linear regression (MLR) for validation. Figure 1 shows the flowchart of the MLR-BI method as described in Eq. (13). MLR is used to construct the system term, as shown in Eqs. (14) and (15). $f(V_q')$ represents the regression function that includes the target sensor to be calibrated, which is taken as the measurements of the practical HVAC system. $f(V_p)$ represents all variable information in the system except the target sensor to be calibrated, which is taken as the benchmark of the reliable system. The compensation function of the target sensor is shown in Eq. (16).

$$D_{MLR}(x) = \underbrace{\sum_p^P \sum_q^Q \left(f(V_p) - f(V_q')\right)^2}_{system\ term} + \underbrace{\sum_q^Q \left(V_{B,q} - V_{ca}\right)^2}_{sensor\ term} \quad (13)$$

$$f(V_p) = \alpha_1 V_1 + \alpha_2 V_2 + \cdots + \alpha_p V_p + \alpha_0 \quad (14)$$

$$f(V_q') = \beta_1 V_1' + \beta_2 V_2' + \cdots + \beta_q V_q' + \beta_0 \quad (15)$$

$$V_{ca} = V_q' + x \quad (16)$$

where $V_1' - V_q'$ are the target sensors to be calibrated, $V_1 - V_p$ are the physical sensors except the target sensor to be calibrated, α_0, β_0 are the constant terms of the MLR models, $\alpha_1 - \alpha_p$ and $\beta_1 - \beta_q$ are the coefficients of corresponding MLR models. V_{ca} is the calibrated value after compensating.

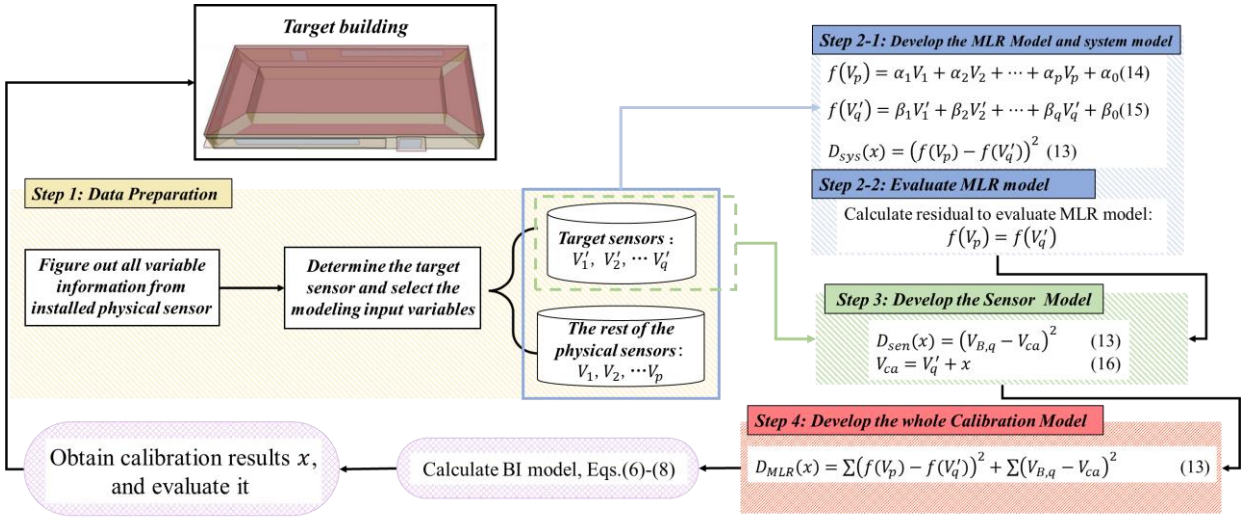

Figure 1 Flowchart of MLR-BI in-situ sensor calibration for target sensor in building HVAC systems.

4 Research framework

Figure 2 shows the research framework, which consists of three main aspects including research content, research purposes and evaluation indexes.

- ***Research content***

(1) The calibration performance was evaluted using both simulated and practical data of two building HVAC systems in two case studies. The effectiveness of in-situ sensor calibration methods was validated via five fault conditions covering both single and simultaneous faults.

(2) The proposed method uses the MLR to improve the BI method to simplify the calibration modeling process by reducing the hard requirements of indirect information supplement via additional virtual sensors, redundant sensors and the EC equations. EC-BI and PCA were used for comparison.

(3) Five variable scenarios were considered to simuate the information-rich simulated scenario, information-poor practical scenario and the extremely limited variable scenario of practical HVAC systems. Since the EC-BI fails to work in a limited variable information scenario, only the PCA was employed for comparison.

- ***Research purposes***

(1) To validate the applicability of MLR-BI under both single or simultaneous fault conditions.

(2) To evaluate the calibration performance of MLR-BI by comparison with PCA and EC-BI.

(3) To validate that MLR-BI can still guarantee relatively good calibration performance with limited variable information.

- ***Evaluation indexes***

This study evaluated the calibration performance using indexes including mean absolute percentage error (MAPE) [39], the calibration accuracy ξ_{ca} in Eq. (17), and relative error (RE) in Eq. (18).

$$\xi_{ca} = (1 - MAPE) \times 100\% = \left(1 - \frac{\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_{ca,i}}{y_i}\right|}{n}\right) \times 100\% \quad (17)$$

$$RE = \left|\frac{y_i - \hat{y}_{ca,i}}{y_i}\right| \times 100\% \quad (18)$$

where, y_i and $\hat{y}_{ca,i}$ represent the true value and calibrated value of the target sensor variable respectively. n represents the total number of testing data sample and i represents the ith sample of the testing dataset.

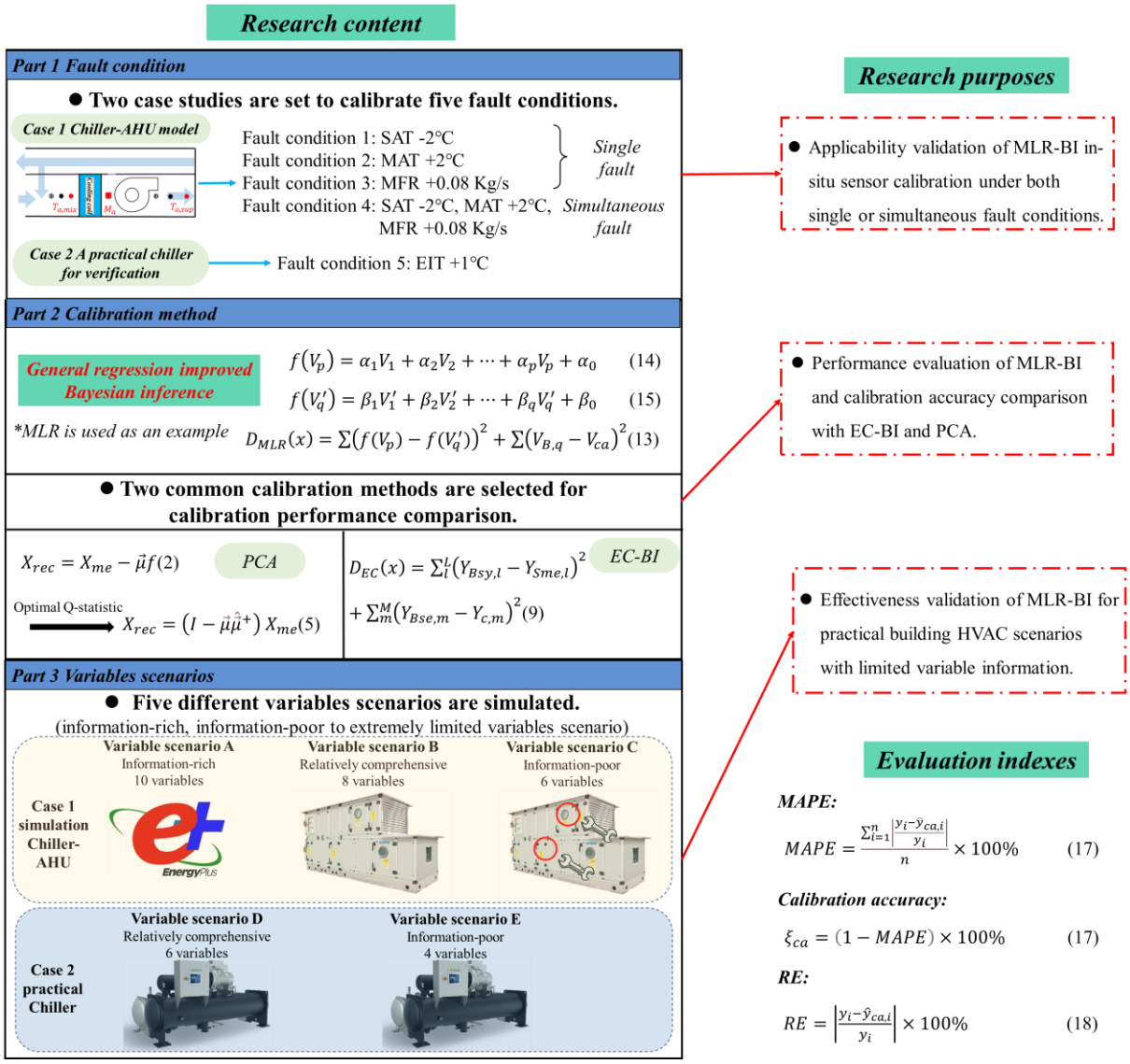

Figure 2 Research framework of this study.

5 Case studies

This study validated the proposed in-situ sensor calibration method using both simulated and practical data of two building HVAC systems in two case studies.

5.1 Case 1: The simulated chiller-AHU model

The first target HVAC system is a simulated chiller-AHU system using the EnergyPlus software with a 10-min data collection interval as shown in Figure 3. It is a constant-speed air volume system. In cooling season, the indoor temperature and the air supply temperature are set as 26 °C and 14 °C, respectively. As described in Table 1, the data collected in July 1st to 31st were used to train the calibration model while the data collected in August 1st to 29th are adopted for validation. Prior to model training and validation, the original data should be pre-processed in two steps, outlier removal and data screening. The obvious outliers like system power-off data at weekends should be removed to enhance the model performance. Moreover, the steady-state between 10:30 a.m. and 5:30 p.m. should be selected for modeling every day.

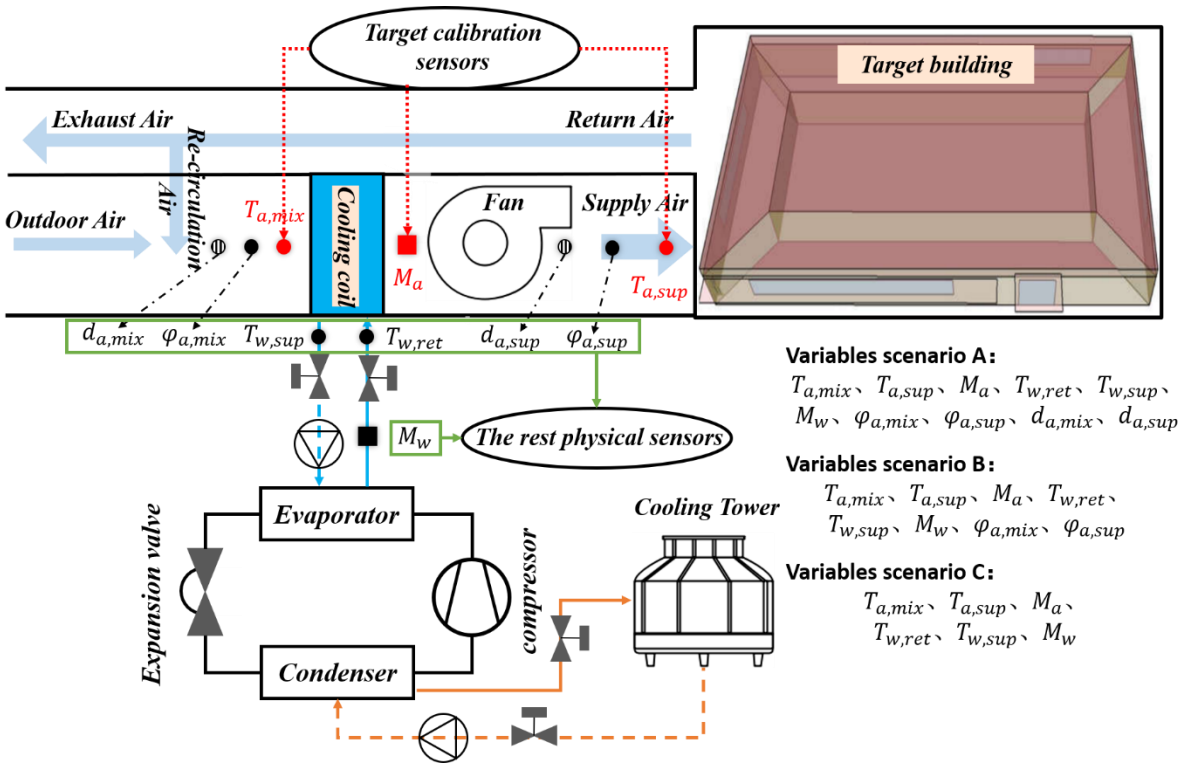

Figure 3 The simulated chiller-AHU system with sensor fault locations and variable scenarios in case 1.

Table 1 Description of the training and test datasets in case 1.

Time period		Training and test datasets
July 1st – July 31st	10:30 - 17:30	Calibration model construction training set (966 samples in total)
August 1st – August 29th	every day	Calibration test set (966 samples for each fault condition)

5.1.1 Setup of target sensor bias faults

The case 1 considered three target sensors including the coil mixed air inlet temperature sensor

(MAT), the coil air supply temperature sensor (SAT), and the coil air flow sensor (MFR), as shown in Figure 3. Table 2 shows the four fault conditions including three single bias fault conditions and a simultaneous fault condition. The sensor faults were introduced at the time period of August 1st and August 29th for each sensor.

Table 2 Paramters setting for the four fault conditions in simulated case 1.

Fault condition	MAT (°C)	SAT (°C)	MFR (kg/s)	Fault type
1	+2	0	0	Single fault
2	0	-2	0	
3	0	0	+0.08	
4	+2	-2	+0.08	Simultaneous fault

5.1.2 Setup of variable scenarios

In this section, the number of input variables of the calibration model is changed to simulate different variable scenarios in practical application. The details about the three variable scenarios are as follows.

(1) Variable scenario A uses the simulated model of a building HVAC system. All the simulated output variables are available for modeling. 10 variables are selected to construct the calibration model based on the principle of EC [32].

(2) Variable scenario B simulates a practical building HVAC system with relatively comprehensive data information. Only the measurable variables are available for modeling. The number of input variables is 8. Compared with scenario A, the humidity ratio d, which is a key variable for calculating the enthalpy value, is removed as it cannot be directly measured by physical sensors.

(3) Variable scenario C simulates a building HVAC system in practice with limited variable information. The number of input variables is 6. Only the temperature and flowrate variables are used in scenario C to simulate the limited information condition.

Table 3 presents the combinations of input variables used in each scenario. The variables include the supply air temperature, relative humidity and humidity ratio ($T_{a,sup}$, $\varphi_{a,sup}$, $d_{a,sup}$), the mixed air temperature, relative humidity and humidity ratio ($T_{a,mix}$, $\varphi_{a,mix}$, $d_{a,mix}$), the supply air flow rate (M_a), the chilled water inlet temperature ($T_{w,ret}$), outlet temperature ($T_{w,sup}$), and flow rate (M_w) of the cooling coil.

Table 3 Combinations of input variables for different variable scenarios in simulated case 1 of this study.

variable scenarios	Input variables
Variable scenario A	$T_{a,mix}$、$T_{a,sup}$、M_a、$T_{w,ret}$、$T_{w,sup}$、M_w、$\varphi_{a,mix}$、$\varphi_{a,sup}$、$d_{a,mix}$、$d_{a,sup}$
Variable scenario B	$T_{a,mix}$、$T_{a,sup}$、M_a、$T_{w,ret}$、$T_{w,sup}$、M_w、$\varphi_{a,mix}$、$\varphi_{a,sup}$
Variable scenario C	$T_{a,mix}$、$T_{a,sup}$、M_a、$T_{w,ret}$、$T_{w,sup}$、M_w

5.2 Case 2: A practical chiller plant

In Figure 4, another target HVAC system is a practical chiller in an electronic factory building with a 1-hour data collection interval. For the cooling season, the evaporator water outlet temperature is set around 7 ºC. The summer operation data ranged from 2021/8/1 0:00 to 2021/9/31 11:00 were prepared. In this study, six sensor measured variables were selected since they have no missing values. The 6 variables are evaporator water inlet and outlet temperature ($T_{ev,in}$, $T_{ev,out}$), condenser water inlet and outlet temperature ($T_{co,in}$, $T_{co,out}$), part load ratio (P_{load}), and the total power input (W).

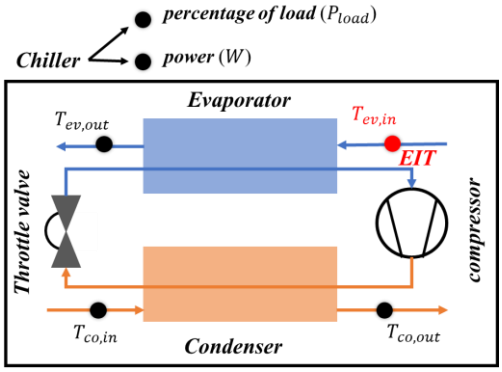

Figure 4 Illustration of the practical water-cooled chiller system with six measured variables in case 2.

5.2.1 Setup of target evaporator inlet temperature sensor with bias faults

In case 2, in order to validate the practical in-situ sensor calibration performance of MLR-BI, the +1 ºC bias fault was added to the evaporator inlet temperature (EIT) sensor from 2021/9/12 9:00 to 2021/9/31 11:00 in the practical fault-free chiller system. As shown in Table 4, the fault condtion 5 is a type of single bias fault on the EIT sensor. Prior to the model training and validation, the original data set was divided into the training set (600 samples) and the test set (400 samples).

Table 4 Paramters setting for the fault conditions in practical case 2.

Fault condition	EIT (°C)	Fault type
5	+1	Single fault

5.2.2 Setup of variable scenarios

Also, two variables scenarios D and F are considered in practical case 2 as shown in Table 5. The training set is used to develop regression models for variable scenarios D and F, respectively.

(1) Variable scenario D is the practical scenario with relatively comprehensive variable information. Nearly all the six measured variables are available for modeling. The number of input variables is 6.

(2) Variables scenario F is the practical scenario with limited variable information. The number of input variables is 4. Only the four temperature sensor measurements are remained for modeling.

Table 5 Combinations of input variables for different variables scenarios in practical case 2 of this study.

Variable scenarios	Input variables
Variable scenario D	$T_{ev,in}$, $T_{ev,out}$, $T_{co,in}$, $T_{co,out}$, P_{load}, W
Variable scenario F	$T_{ev,in}$, $T_{ev,out}$, $T_{co,in}$, $T_{co,out}$

5.3 Calibration of the target sensors using the three calibration methods

This section presents the calibration processes of four target sensors (three in the simulated case 1 and one from the practical case 2) using the three calibration methods, PCA, EC-BI and MLR-BI. For the five variable scenarios in this study, two of them are adopted for describing the calibration model construction processes in detail. The two scenarios are 10 variables from variable scenario A in simulated case 1 and 6 variables from variable scenario D in practical case 2. Especially for practical chiller system in case 2, the EC equations cannot be achieved owing to the lack of water flow variables. Hence, for practical case 2, only the two models PCA and MLR-BI without EC equations can be established for the two variable scenarios D and F.

5.3.1 Sensor calibration using PCA-based reconstruction

5.3.1.1 Calibration process of case 1

According to Section 2.1, Figure 5 shows the process of faulty sensor calibration using PCA-

based reconstruction. For simulated case 1, the PCA model is trained by the training dataset (input variables are: $T_{a,sup}$, $\varphi_{a,sup}$, $d_{a,sup}$, $T_{a,mix}$, $\varphi_{a,mix}$, $d_{a,mix}$, M_a, $T_{w,ret}$, $T_{w,sup}$, M_w). The test dataset is used for reconstruction calculation based on the optimal Q statistic.

5.3.1.2 Calibration process of case 2

For practical case 2, the PCA model is trained by the training dataset (input variables are: $T_{ev,in}$, $T_{ev,out}$, $T_{co,in}$, $T_{co,out}$, P_{load}, W).The test dataset is used for reconstruction calculation based on the optimal Q statistic.

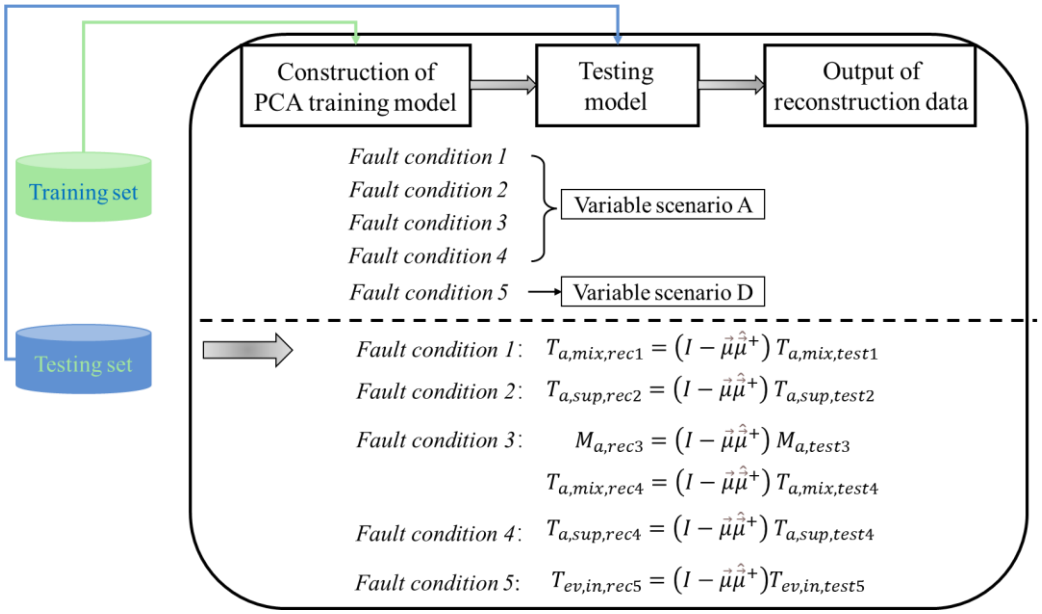

Figure 5 Sensor calibration using PCA-based reconstruction (Refering to Eq. (5)).

Note: The subscripts 'rec1-5' represent the variable reconstruction results under the fault conditions 1-5. The subscripts '$test$ 1-5' represent the four test datasets corresponding to the fault conditions 1-5.

5.3.2 Sensor calibration using EC-BI

The modeling process of EC-BI method is shown in Eqs. (19) - (25). The heat exchanges between the air side and the water side of cooling coil can be calculated using Eqs. (19) - (20), respectively. The calculation of the air side heat exchange involves the three target sensors, which is the practical measurements. The water side heat exchange is acted as the benchmark of the reliable system. Hence, the system term of the calibration model can be constructed using Eqs. (19) - (20). Since the enthalpy value cannot be obtained directly, a physical model is used to calculate it, as shown in Eq. (21). The compensation functions of the three target sensors are shown in Eqs. (22) - (24), which are used to

construct the sensor term of the calibration model. The calibration model is constructed as shown in Eq. (25). Figure 6 shows the sensor calibration flowchart using the EC-BI method considering the construction process of system terms only. This is because the construction process of sensor terms are much simpler.

$$\Delta E_a = \left(h_{a,sup} - h_{a,mix}\right) \times M_a \tag{19}$$

$$\Delta E_w = \left(T_{w,ret} - T_{w,sup}\right) \times M_w \times C_w \tag{20}$$

$$h = 1.01 \times T + d \times (2501 + 1.85 \times T) \tag{21}$$

$$T_{a,sup,ca} = T_{a,sup} + x_{a,sup,EC} \tag{22}$$

$$T_{a,mix,ca} = T_{a,mix} + x_{a,mix,EC} \tag{23}$$

$$M_{a,ca} = M_a + x_{Ma,EC} \tag{24}$$

$$D(x_{EC}) = \underbrace{(\Delta E_a - \Delta E_w)^2}_{system\ term} + \sum \underbrace{\left(V_{target} - V_{ca}\right)^2}_{sensor\ term} \tag{25}$$

where, ΔE_a represents the air side heat exchange, ΔE_w represents the water side heat exchange. $h_{a,sup}$ and $h_{a,mix}$ represent the enthalpies of supply air and mixed air, respectively. $T_{a,sup}$, $T_{a,mix}$ and M_a should be treated as V_{target}. C_w represents the specific heat capacity of water, i.e., 4.186J/(Kg.ºC). $T_{a,sup,ca}$, $T_{a,mix,ca}$, $M_{a,ca}$, represent the supply air temperature, mixed air temperature and supply air flowrate after correction by the compensation function, respectively. They should be employed as V_{ca}. $x_{a,sup,EC}$, $x_{a,mix,EC}$ and $x_{Ma,EC}$ represent the supply air temperature, and the compensation values of mixed air temperature and supply air flowrate, respectively. They should be considered as x_{EC}. (Finally, x_{EC} is equal to the mean of the posterior distribution as described in the principle of BI in Section 2.1).

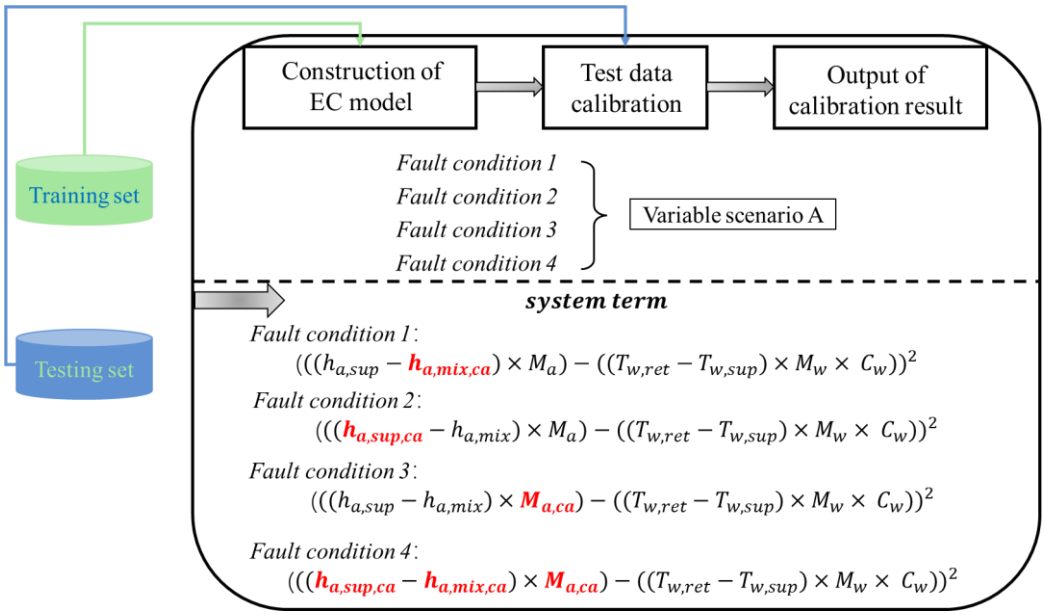

Figure 6 Sensor calibration using the EC-BI method (Referring to system term in Eq. (9) and Eqs. (19) - (20)).

5.3.3 Sensor calibration using MLR-BI method

5.3.3.1 Calibration process of case 1

For MLR-BI calibration process of the simulated case 1, Figure 7 shows the system terms under the four fault conditions. When developing MLR models, one variable is selected as the dependent variable while the other variables are taken as the independent variables. The regression functions are shown in Eqs. (26) - (27), theoretically $f(V_1) = f(V_1')$ (Fault condition 1 is taken as an example). Compensation function is also required for target sensors under other fault conditions, as shown in Eqs. (28) - (30).

$$f(V_1) = \underbrace{\alpha_1 T_{a,mix} + \alpha_2 M_a + \cdots + \alpha_9 \varphi_{a,mix}}_{other\ 9\ physical\ sensor\ information} + \alpha_0 \tag{26}$$

$$f(V_1') = T_{a,sup} \tag{27}$$

$$T_{a,sup,ca} = T_{a,sup} + x_{a,sup,MLR} \tag{28}$$

$$T_{a,mix,ca} = T_{a,mix} + x_{a,mix,MLR} \tag{29}$$

$$M_{a,ca} = M_a + x_{Ma,MLR} \tag{30}$$

where, α_0 is the constant term of the MLR model, $\alpha_1 - \alpha_9$ are the coefficients corresponding to the input variables. $x_{a,sup,MLR}$, $x_{a,mix,MLR}$, and $x_{Ma,MLR}$ represent the compensation values of the supply

air temperature, the mixed air temperature, and the supply air flowrate, respectively.

5.3.3.1 Calibration process of case 2

For MLR-BI calibration process of the practical case 2, the regression functions are shown in Eqs. (31) - (32). Compensation function is shown in Eq. (33). The calibration model of case 2 is shown in Eq. (34).

$$f(V_2) = \underbrace{\beta_1 T_{ev,out} + \beta_2 T_{co,in} + \cdots + \beta_5 T_{co,out}}_{other\ 5\ physical\ sensor\ information} + \beta_0 \tag{31}$$

$$f(V_2') = T_{ev,in} \tag{32}$$

$$T_{ev,in,ca} = T_{ev,in} + x_{ev,in,MLR} \tag{33}$$

$$D(x_{MLR}) = \underbrace{\left(f(V_p) - f(V_q')\right)^2}_{system\ term} + \sum \underbrace{\left(V_{target} - V_{ca}\right)^2}_{sensor\ term} \tag{34}$$

where, β_0 is the constant term of the MLR model, $\beta_1 - \beta_9$ are the coefficients corresponding to the input variables. $x_{ev,in,MLR}$ represent the compensation values of the evaporator water inlet temperature. $x_{a,sup,MLR}$, $x_{a,mix,MLR}$, $x_{Ma,MLR}$ and $x_{ev,in,MLR}$ are referred to as x_{MLR}, collectively. Finally, x_{MLR} is equal to the mean of the posterior distribution as described in the principle of BI in Section 2.1.

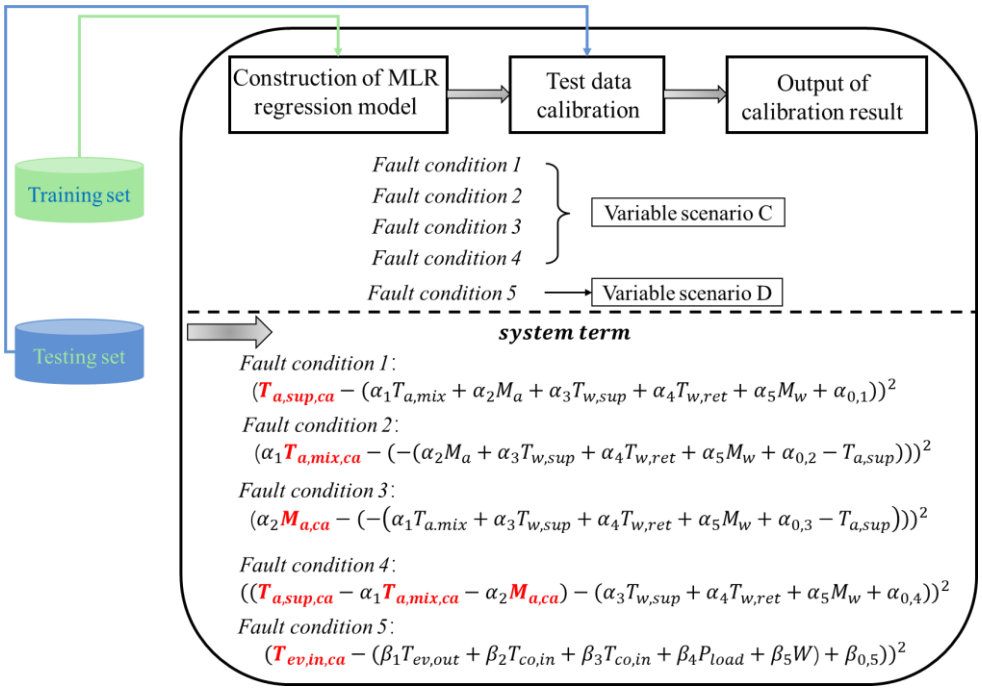

Figure 7 Sensor calibration using MLR-BI method (Referring to system term in Eq. (13), Eqs. (26) - (27) and Eqs. (31) - (32)).

6 Results and discussion

6.1 Calibration results under four fault conditions for the simulated case 1

Figure 8 shows the calibration accuracy of the three methods in calibrating single sensor faults. The MLR-BI method shows good performance with calibration accuracies of 99.9%, 99.9% and 98.0% for fault conditions 1-3, respectively. The calibration accuracy under fault conditions 1-3 using the PCA-based reconstruction method are 96.9%, 95.9% and 89.8%, respectively. Figure 9 shows the calibration accuracies of the three methods in calibrating simultaneous sensor faults. MLR-BI shows good calibration performance on the simultaneous sensor faults with an average calibration accuracy of 99.45%. For the MAT simultaneous sensor fault condition, all the three methods obtain accuracies over 98%. The calibration accuracy even reaches 99.99% when using MLR-BI. For SAT simultaneous sensor calibration, PCA shows calibration accuracy of 91.57% less than the other two methods. EC-BI obtains similar calibration accuracy as MLR-BI. For the MFR simultaneous sensor calibration, EC-BI shows lower calibration accuracy of only 62.85% while MLR-BI achieves calibration accuracy of 98.49%.

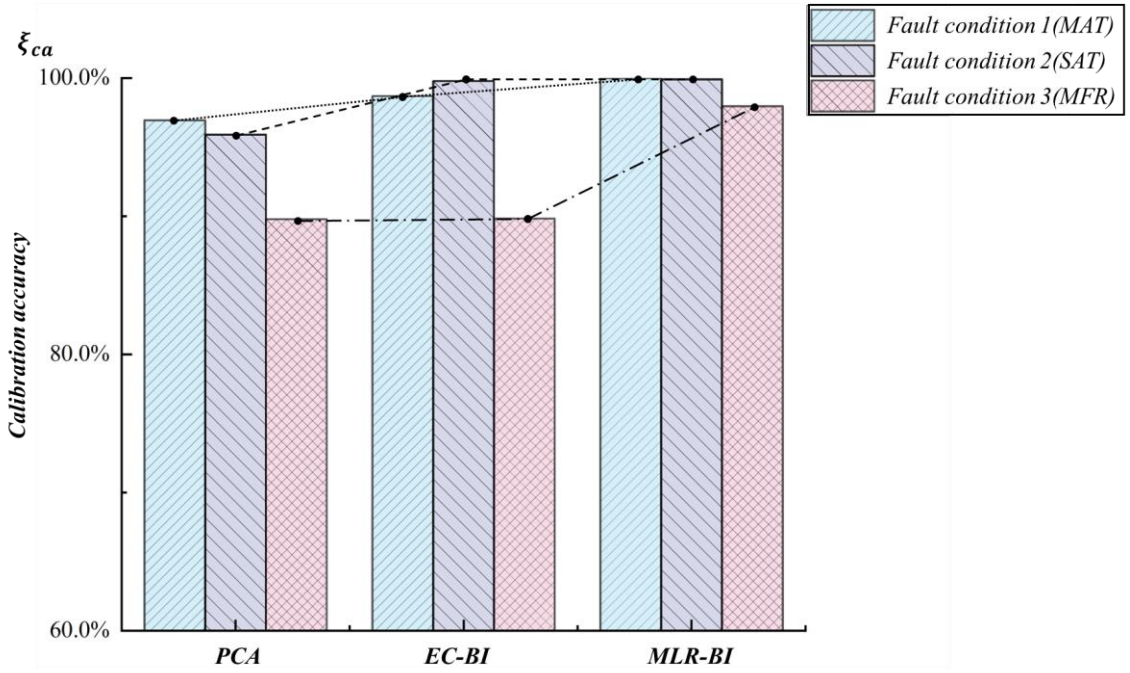

Figure 8 Sensor calibration accuracy under fault conditions 1-3 (single fault) for simulated case 1.

Note: All the calibration accuracy ξ_{ca} in section 6 can be calculated based on Eq. (17).

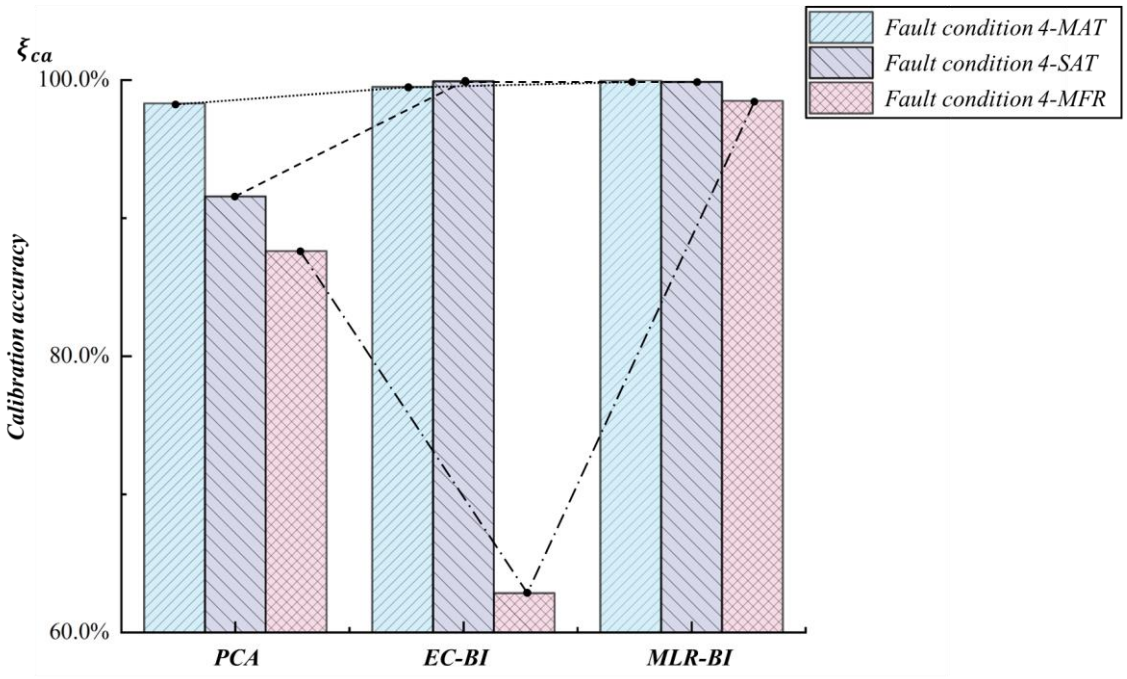

Figure 9 Sensor calibration accuracy under fault condition 4 (simultaneous fault) for simulated case 1.

Figures 10-15 show in-situ sensors calibration results using the three calibration methods under the four fault conditions. Calibration results of sensor fault conditions 1-3 are described Figures 10-12 while the sensor fault condition 4 calibration results are shown in Figures 13-15. Figures 10 and 13, show the calibration results of PCA for the three target sensors under single fault and simultaneous fault conditions, respectively. Unlike the two BI methods, PCA is capable of reconstructing the faulty data to the final calibrated data via Eqs. (1) - (5) without the statistical deduction on the posterior distribution results. From Section 2.1, PCA actually has no resampling step which means it may not need the posterior distribution result. But for EC-BI and MLR-BI calibration results in Figures 11, 12, 14 and 15, the posterior distributions of the sensor faults are also presented. In terms of probability density function (PDF) curves. For example, Figure 11 (a2) illustrates the probability density function of the posterior distribution for the MAT sensor. Savitzky Golay (SG) [48] is used to smooth the line curves, and the number of window points is 150. MLR-BI shows better performance on sensor calibration than the other two methods. EC-BI performs well in calibrating SAT and MAT sensors.

Especially, MLR-BI shows better calibration results than EC-BI for the MFR sensor when dealing with simultaneous sensor faults. This is mainly because the data of other fault sensors have a certain impact on the calibration model based on EC equations. In addition, for the simulated constant-speed air volume system, the MFR sensor measurement data are fixed with less fluctuations, which may

further amplify the sensor fault impact compared with other fluctuated data. After performance comparison, it can be found that MLR-BI shows good calibration performance for both single and simultaneous faults for the three target sensors.

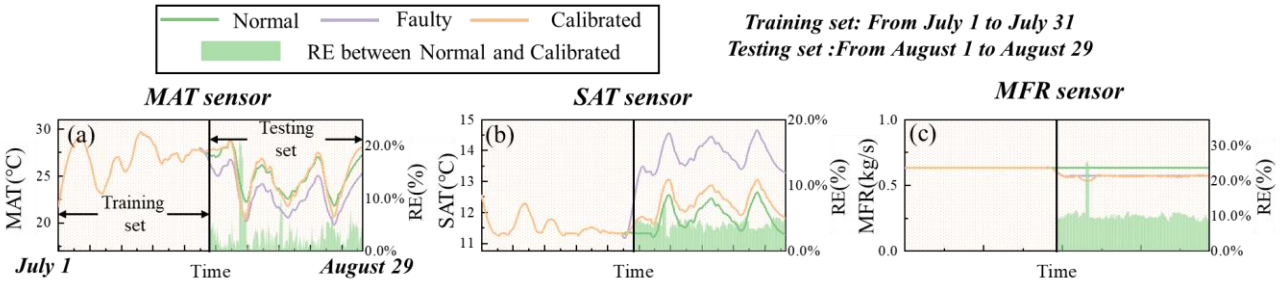

Figure 10 Sensor calibration results using the PCA-based reconstruction method for three single fault conditions 1-3: (a) MAT, (b) SAT and (c) MFR.

Note:

"Normal" represents the system normal operational data which means that the building HVAC system works at fault-free condition. The Normal data are directly outputted by the EnergyPlus simulated model.

"Faulty" represents the system sensor fault data which means that the building HVAC system works at sensor fault condition. The Faulty data are outputted by the EnergyPlus simulated model after introducing the sensor bias fault but before the senor fault in-situ calibration process.

"Calibrated" represents the calibrated system sensor fault data which means that the building HVAC system works at sensor fault condition but the sensor error readings are calibrated. At this time, the building HVAC system works at the sensor fault-tolerant condition. The Calibrated data are obtained via Eqs. (1) - (5) using the Python PCA-based reconstruction model and used to replace the sensor fault data for fault-tolerant operation.

The left axes show the real-time values of the three target sensors (MAT, SAT and MFR) in the building HVAC system. Each curve represents the real-time developing trend of the given target sensor.

The right axes show the values of RE between Normal and Calibrated for the testing set. "RE between Normal and Calibrated" represents the relative error change before and after calibration. The RE can be obtained using Eq. (18) in Section 4.

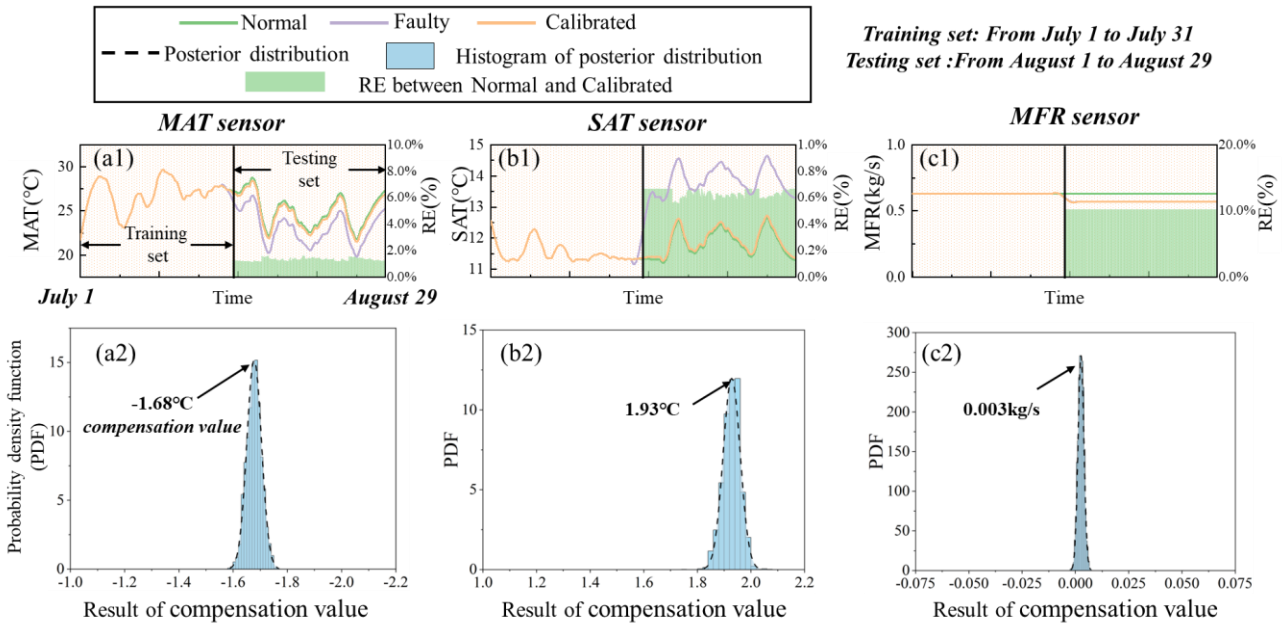

Figure 11 Sensor calibration results using the EC-BI method for three single fault conditions 1-3: (a1) MAT, (b1) SAT, (c1) MFR; The Posterior distribution of (a2) MAT, (b2) SAT and (c2) MFR.

Note:

"Calibrated": The Calibrated data are obtained via Eqs. (6) - (12) using the EC-BI model and used to replace the sensor fault data for fault-tolerant operation.

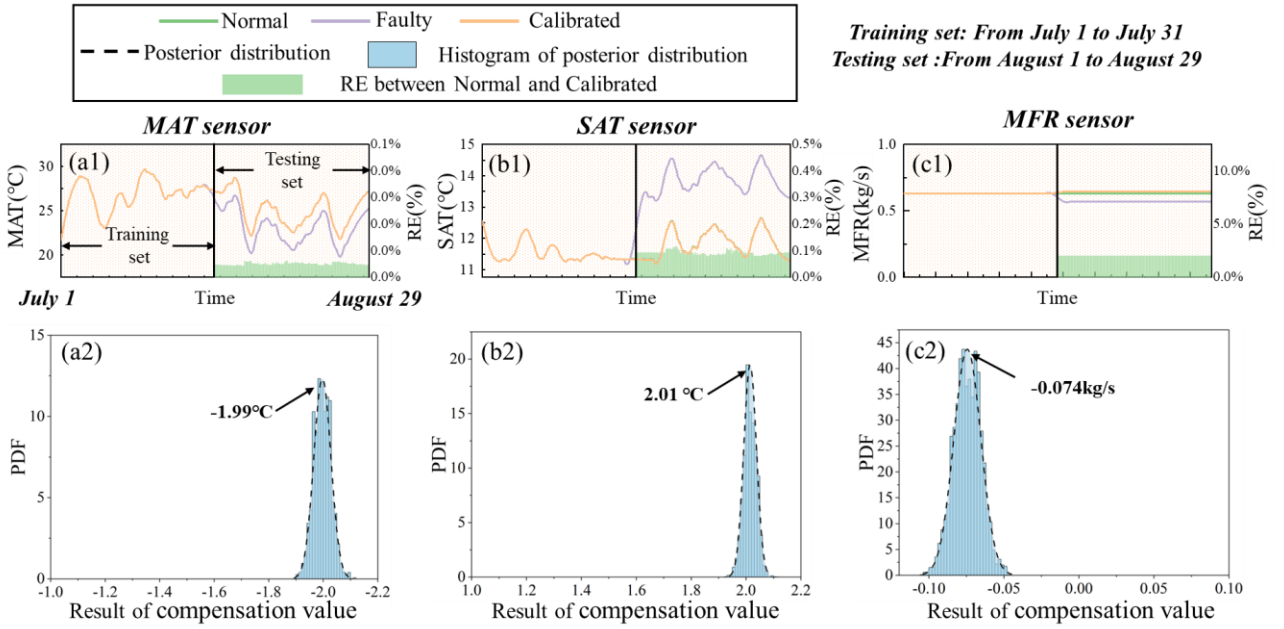

Figure 12 Sensor calibration results using the MLR-BI method for three single fault conditions 1-3: (a1) MAT, (b1) SAT, (c1) MFR; The Posterior distribution of (a2) MAT, (b2) SAT and (c2) MFR.

Note:

"Calibrated": The Calibrated data are obtained via Eqs. (6) - (8) and Eqs. (13) - (16) using the MLR-BI model and used to replace the sensor fault data for fault-tolerant operation.

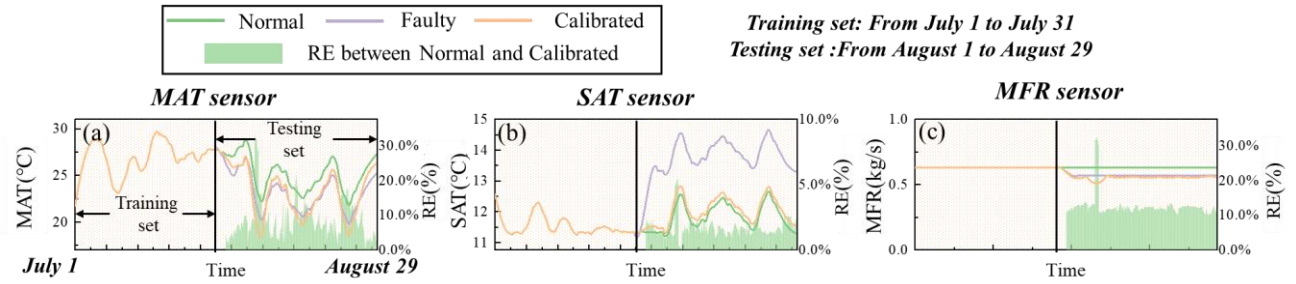

Figure 13 Sensor calibration results using the PCA-based reconstruction method for simultaneous fault conditions4: (a) MAT, (b) SAT and (c) MFR.

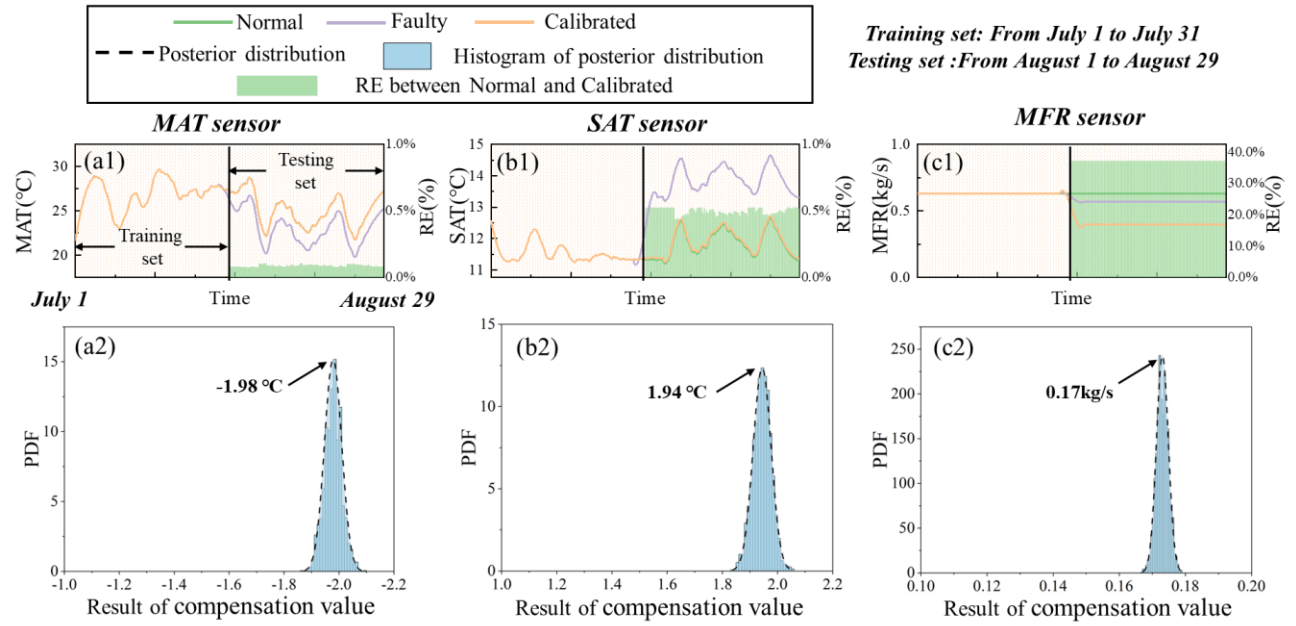

Figure 14 Sensor calibration results using the EC-BI method for simultaneous fault conditions 4: (a1) MAT, (b1) SAT, (c1) MFR; The Posterior distribution of (a2) MAT, (b2) SAT and (c2) MFR.

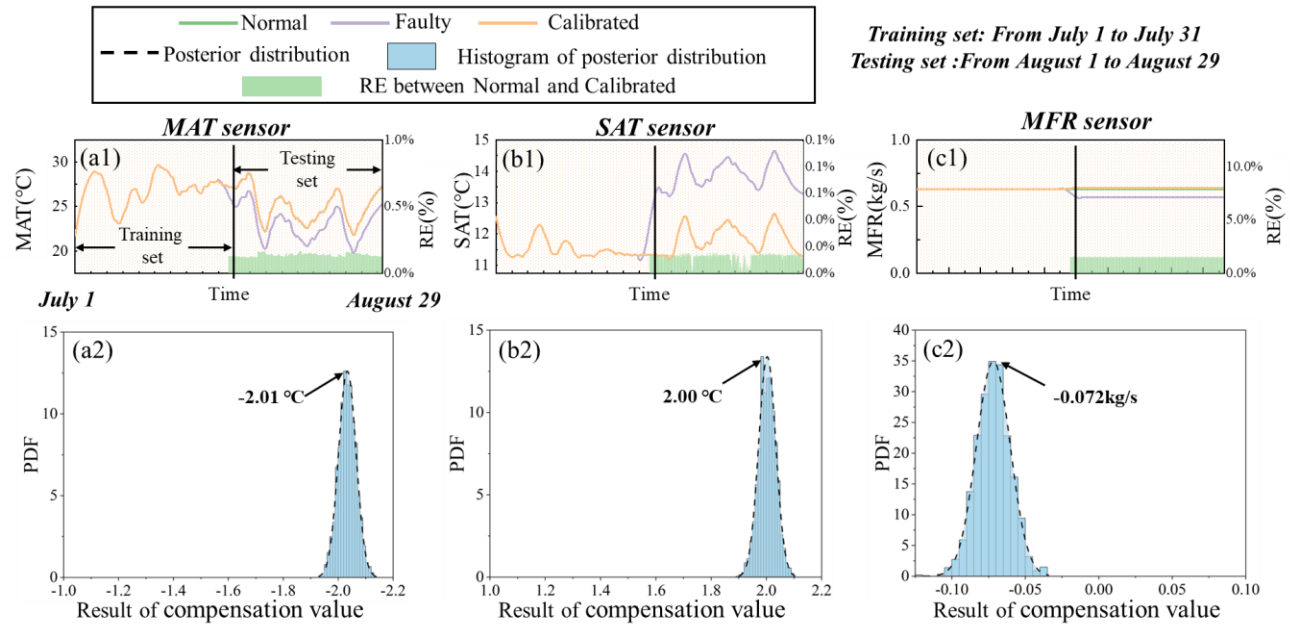

Figure 15 Sensor calibration results using the MLR-BI method for simultaneous fault conditions 4: (a1) MAT, (b1)

SAT, (c1) MFR; The Posterior distribution of (a2) MAT, (b2) SAT and (c2) MFR.

6.2 Calibration results of the three variable scenarios for the simulated case 1

Figure 16 depicts the residual distributions of the regression models developed based on the variables in the three variable scenarios A, B and C, respectively. The regression residual sums of squares are 0.0064, 0.0064 and 0.0073, respectively. This means the MLR models show very high regression precision for the three variable scenarios in this study.

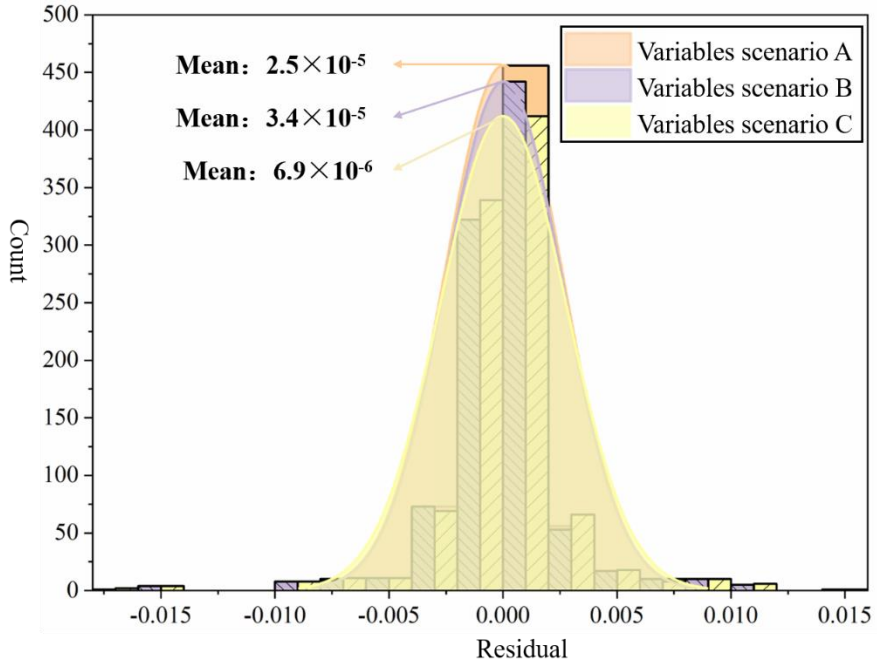

Figure 16 Residual distributions of the developed MLR models in the three variable scenarios A, B and C for simulated case 1.

Figure 17 compares the in-situ sensor calibration performance of MLR-BI and PCA under four fault conditions in three different scenarios. Figure 17 (a) shows the results under fault conditions 1-3. MLR-BI obtains very high calibration accuracy of 99.3% on average for single sensor faults for all the three variable scenarios. By contrast, PCA shows relatively lower calibration accuracies than MLR-BI in all the three variable scenarios. The average calibration accuracy is about 82.87% since the average calibration accuracy of MAT sensor is only 63.27%. For the simultaneous sensor faults calibration results in Figure 17 (b), MLR-BI shows high calibration accuracy of 99.5% on average for all the three variable scenarios. When the PCA is used, the calibration accuracy of SAT sensor is as high as 98.8% on average, but MFR sensor is slightly lower 87.71%, and the MAT sensor is extremely low 0.91%

which means PCA can hardly calibrate the MAT sensor in variable scenario B. Such significant calibration accuracy drop of the MAT sensor could be caused by the removal of two variables relative humidity and humidity ratio. The two humidity variables could have certain impact on the PCA-based sensor reconstruction process for the MAT sensor. But MLR-BI can quickly establish the in-situ sensor calibration model based on the available built-in sensors. Compared with PCA, it is obvious that MLR-BI obtains much more stable calibration results.

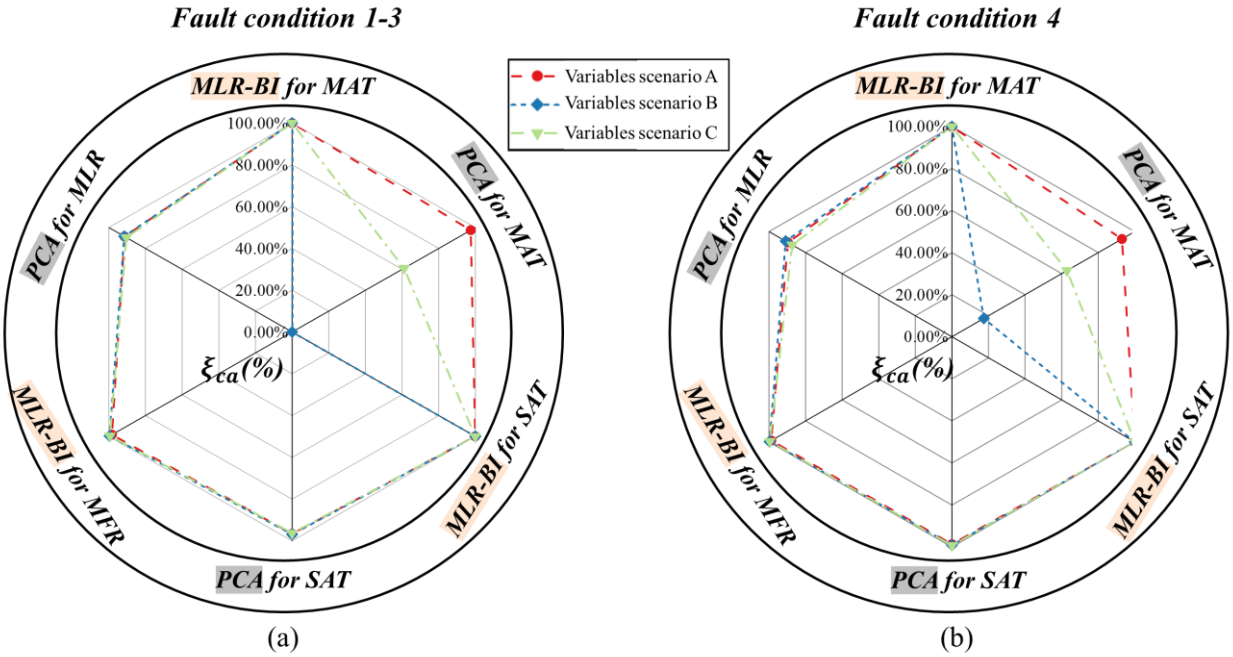

Figure 17 Calibration accuracies of the three in-situ sensor calibration methods under different variable scenarios (a) single fault conditions 1-3 and(b) simultaneous fault condition 4.

6.3 Calibration results of the two variable scenarios for the practical case 2

For the practical chiller system in case 2, only MLR-BI and PCA still work for the two variable scenarios D and F. For the MLR-BI method, the residual sums of squares of MLR models are no more than 0.1, which means the MLR models are reliable with high-precision to replace the EC equations and establish the distance function for improving the BI-based in-situ sensor calibration performance. Figure 18 shows the sensor calibration accuracies of two methods. MLR-BI obtains the calibration accuracies of 97.1% in both variables scenario D and F. By contrast, the sensor calibration accuracies of PCA are 92.9% and 95.6% for variables scenario D and F, which are much lower than MLR-BI. In general, the preliminary application validation results show that the proposed MLR-BI method

performs well using operational data in the practical building HVAC system.

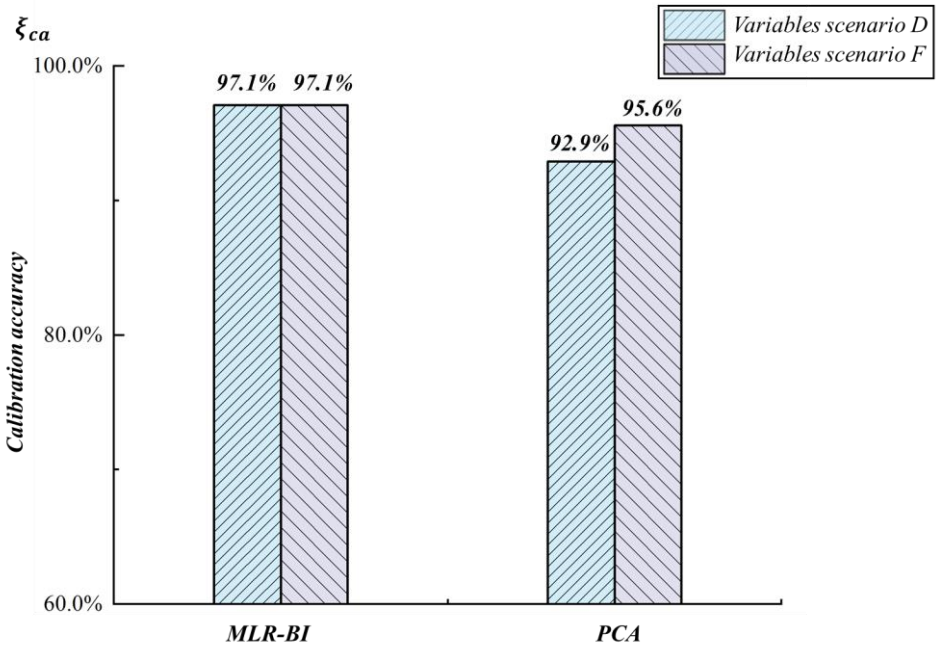

Figure 18 Calibration accuracies of evaporator water inlet temperature sensor for the two variable scenarios D and F in the practical case 2.

7 Conclusions

This paper proposed and validated a general regression improved Bayesian inference in-situ sensor calibration method for building HVAC systems with the consideration of limited variable information senarios in practice. Especially, this study developed the MLR-BI method to improve the BI method via multiple linear regression. MLR-BI can simplify the calibration modeling process by reducing the hard requirements of indirect information supplement via additional virtual sensors, redundant sensors and the EC equations. Two case studies of two building HVAC systems was adopted to evaluate the calibration performance considering both simulated and practical data. The effectiveness of MLR-BI was validated via five fault conditions covering both single and simultaneous faults and five variable scenarios taking account for the limited variable information. Results indicated that MLR-BI obtains high-level in-situ sensor calibration accuracy for the two HVAC systems. Main conclusions are as follows:

(1) For the simulated case 1 HVAC system, compared with PCA and EC-BI, MLR-BI increases the average calibration accuracy by 6.01% and 8.44% under four fault conditions at the variable

information-rich senario, respectively. For the three studied variable scenarios of the simulated case 1, the calibration accuracy of MLR-BI is 99.65% on average. Especially in the four-variable scenario with limited variable information, MLR-BI shows the average calibration accuracy of 99.75% while PCA obtains 79.46% and EC-BI fails to work because of variable limitation.

(2) For the fault condition of the limitted-variable practical case 2, MLR-BI still outperforms the other two and obtains 97.1% calibration accuracy in two practical scenarios. Compared with PCA, MLR-BI improves the calibration accuracy by 2.8% on average.

Although the general regression improved Bayesian inference in-situ sensor calibration method has been successfully applied in two building HVAC systems and obtains good calibration performance for five fault conditions and five variable scenarios. There is still a lack of validation study on more HVAC systems, sensor type and practical application scenarios with considerations of more real-world complex fault conditions including bias, drift, precision degradation, etc. Meanwhile, calibration models can be further extended, such as nonlinear regression model, high-precision regression model, etc. Apart from the sensor calibration process, future work should further focus on the whole sensor fault-tolerant control strategy for the practical building HVAC system.

Acknowledgements

This work was jointly supported by the Opening Fund of Key Laboratory of Low-grade Energy Utilization Technologiesand Systems of Ministry of Education of China (Chongqing University) (LLEUTS-202305), the National Natural Science Foundation of China (51906181), the Opening Fund of State Key Laboratory of Green Building in Western China (LSKF202316), 2021 Construction Technology Plan Project of Hubei Province (2021-83), the Excellent Young and Middle-aged Talent in Universities of Hubei, China (Q20181110).

References

[1] A. Taal, L. Itard, W. Zeiler, A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems, Energy and Buildings 179 (2018) 144-155.

[2] M. Horrigan, W.J.N. Turner, J. O'Donnell, A statistically-based fault detection approach for environmental and energy management in buildings, Energy and Buildings 158 (2018) 1499-1509.

[3] C. Fan, F. Xiao, Z. Li, J. Wang, Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review, Energy and Buildings 159 (2018) 296-308.
[4] S. Yoon, In-situ sensor calibration in an operational air-handling unit coupling autoencoder and Bayesian inference, Energy and Buildings 221 (2020) 110026.
[5] S. Yoon, In situ modeling methodologies in building operation: A review, Building and Environment 230 (2023) 109982.
[6] A. Chong, K. Menberg, Guidelines for the Bayesian calibration of building energy models, Energy and Buildings 174 (2018) 527-547.
[7] D. Coakley, P. Raftery, P. Molloy, G. White, Calibration of a Detailed BES Model to Measured Data Using an Evidence-Based Analytical Optimisation Approach, International Ibpsa Conference International Ibpsa Conference, 2011.
[8] R. Sendra-Arranz, A. Gutiérrez, A long short-term memory artificial neural network to predict daily HVAC consumption in buildings, Energy and Buildings 216 (2020) 109952.
[9] C. Xu, H. Chen, J. Wang, Y. Guo, Y. Yuan, Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method, Building and Environment 148 (2019) 128-135.
[10] G. Li, Y. Zheng, J. Liu, Z. Zhou, C. Xu, X. Fang, Q. Yao, An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information, Journal of Building Engineering 43 (2021) 102812.
[11] G. Li, Y. Hu, An enhanced PCA-based chiller sensor fault detection method using ensemble empirical mode decomposition based denoising, Energy and Buildings 183 (2019) 311-324.
[12] Y. Guo, J. Wall, J. Li, S. West, Real-Time HVAC Sensor Monitoring and Automatic Fault Detection System, Springer International Publishing2017.
[13] T. Li, Y. Zhao, C. Zhang, J. Luo, X. Zhang, A knowledge-guided and data-driven method for building HVAC systems fault diagnosis, Building and Environment 198 (2021) 107850.
[14] G. Li, Q. Yao, C. Fan, C. Zhou, G. Wu, Z. Zhou, X. Fang, An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems, Building and Environment 203 (2021) 108057.
[15] N. Torabi, H.B. Gunay, W. O'Brien, T. Barton, Common human errors in design, installation, and operation of VAV AHU control systems – A review and a practitioner interview, Building and Environment 221 (2022) 109333.
[16] Z. Du, X. Jin, L. Wu, Fault detection and diagnosis based on improved PCA with JAA method in VAV systems, Building and Environment 42(9) (2007) 3221-3232.
[17] Z. Du, X. Jin, Multiple faults diagnosis for sensors in air handling unit using Fisher discriminant analysis, Energy Conversion and Management 49(12) (2008) 3654-3665.
[18] Y. Yu, W. Li, D. Sheng, J. Chen, A novel sensor fault diagnosis method based on Modified Ensemble Empirical Mode Decomposition and Probabilistic Neural Network, Measurement 68 (2015) 328-336.
[19] J. Li, T. Zhao, P. Wang, S. Yoon, Y. Yu, Effects of various partitions on the accuracy of virtual in-situ calibration in building energy systems, Journal of Building Engineering 32 (2020) 101538.
[20] Z. Li, G. Huang, Preventive approach to determine sensor importance and maintenance requirements, Automation in Construction 31 (2013) 307-312.

[21] Y. Pan, Z. Huang, G. Wu, Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai, Energy and Buildings 39(6) (2007) 651-657.
[22] Y. Yu, H. Li, Virtual in-situ calibration method in building systems, Automation in Construction 59 (2015) 59-67.
[23] S. Yoon, Y. Yu, Extended virtual in-situ calibration method in building systems using Bayesian inference, Automation in Construction 73 (2017) 20-30.
[24] S. Yoon, Y. Yu, A quantitative comparison of statistical and deterministic methods on virtual in-situ calibration in building systems, Building and Environment 115 (2017) 54-66.
[25] S. Yoon, Y. Yu, A Comparison of Stochastic and Deterministic Optimization Algorithms on Virtual In-situ Calibration in Building Systems, 2017 ASHRAE Winter Conference, 2017.
[26] S. Yoon, Y. Yu, Hidden factors and handling strategy for accuracy of virtual in-situ sensor calibration in building energy systems: Sensitivity effect and reviving calibration, Energy and Buildings 170 (2018) 217-228.
[27] S. Yoon, Y. Yu, Hidden factors and handling strategies on virtual in-situ sensor calibration in building energy systems: Prior information and cancellation effect, Applied Energy 212 (2018) 1069-1082.
[28] S. Yoon, Y. Yu, Strategies for virtual in-situ sensor calibration in building energy systems, Energy and Buildings 172 (2018) 22-34.
[29] P. Wang, S. Yoon, J. Wang, Y. Yu, Automated reviving calibration strategy for virtual in-situ sensor calibration in building energy systems: Sensitivity coefficient optimization, Energy and Buildings 198 (2019) 291-304.
[30] P. Wang, K. Han, R. Liang, L. Ma, S. Yoon, The Virtual In-Situ Calibration of Various Physical Sensors in Air Handling Units, Science and Technology for the Built Environment (2020) 1-25.
[31] T. Zhao, J. Li, P. Wang, S. Yoon, J. Wang, Improvement of virtual in-situ calibration in air handling unit using data preprocessing based on Gaussian mixture model, Energy and Buildings 256 (2022) 111735.
[32] Y. Choi, S. Yoon, Virtual sensor-assisted in situ sensor calibration in operational HVAC systems, Building and Environment 181 (2020) 107079.
[33] S. Yoon, Y. Yu, H. Li, Y. Choi, Y. Hong, Improved energy balance calculation of unitary air conditioners via virtual in-situ calibration, Journal of Building Engineering 45 (2022) 103464.
[34] A. Mokhtari, M. Ghodrat, P. Javadpoor Langroodi, A. Shahrian, Wind speed sensor calibration in thermal power plant using Bayesian inference, Case Studies in Thermal Engineering 19 (2020) 100621.
[35] S. Sun, S. Wang, K. Shan, Flow measurement uncertainty quantification for building central cooling systems with multiple water-cooled chillers using a Bayesian approach, Applied Thermal Engineering 202 (2022) 117857.
[36] S. Sun, K. Shan, S. Wang, An online robust sequencing control strategy for identical chillers using a probabilistic approach concerning flow measurement uncertainties, Applied Energy 317 (2022) 119198.
[37] M. Lillstrang, M. Harju, G. del Campo, G. Calderon, J. Röning, S. Tamminen, Implications of properties and quality of indoor sensor data for building machine learning applications: Two case studies in smart campuses, Building and Environment 207 (2022) 108529.
[38] G. Li, J. Xiong, S. Sun, J. Chen, Validation of virtual sensor-assisted Bayesian inference-based

in-situ sensor calibration strategy for building HVAC systems, Building Simulation (2022) 1-19.
[39] S. Chen, X. Zhou, G. Zhou, C. Fan, P. Ding, Q. Chen, An online physical-based multiple linear regression model for building's hourly cooling load prediction, Energy and Buildings 254 (2022) 111574.
[40] N. Torabi, H. Burak Gunay, W. O'Brien, R. Moromisato, Inverse model-based virtual sensors for detection of hard faults in air handling units, Energy and Buildings 253 (2021) 111493.
[41] Y. Hu, H. Chen, G. Li, H. Li, R. Xu, J. Li, A statistical training data cleaning strategy for the PCA-based chiller sensor fault detection, diagnosis and data reconstruction method, Energy and Buildings 112 (2016) 270-278.
[42] H. Zhang, H. Chen, Y. Guo, J. Wang, G. Li, L. Shen, Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on PCA and preprocessed by combined clustering, Applied Thermal Engineering 160 (2019) 114098.
[43] R. Dunia, S.J. Qin, Joint diagnosis of process and sensor faults using principal component analysis, Control Engineering Practice 6(4) (1998) 457–469.
[44] R.M. Dudley, Central Limit Theorems for Empirical Measures, Annals of Probability 6(6) (1978) 899-929.
[45] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Introducing Markov Chain Monte Carlo, Markov Chain Monte Carlo in Practice (1996).
[46] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika (1) (1970) 97-109.
[47] P. Huang, G. Huang, Investigation of maximum cooling loss uncertainty in piping network using Bayesian Markov Chain Monte Carlo method, Energy Procedia 143 (2017) 258-263.
[48] A. Candan, H. Inan, A unified framework for derivation and implementation of Savitzky-Golay filters, Signal Processing (2014).