Fault detection, diagnosis and data recovery for a real building heating/cooling billing system

Youming Chen,a,b,*, Lili Lan a,b

a College of Civil Engineering, Hunan University, Changsha, Hunan 410082, China
b Key Laboratory of Building Safety and Energy Efficiency, MOE, China

A R T I C L E   I N F O
Article history:
Received 24 March 2009
Received in revised form 16 November 2009
Accepted 8 December 2009
Available online 6 January 2010

Keywords:
Billing system
Sensor faults
FDD
Data recovery
PCA
Heating/cooling system

A B S T R A C T
A method of fault detection, diagnosis (FDD) and data recovery is proposed for building heating/cooling billing system in this paper. Principal component analysis (PCA) approach is used to extract the correlation of measured variables in heating/cooling billing system and reduce the dimension of measured data. The measured data of billing system under normal operating condition are used to build PCA model. Sensor faults of bias, drifting and complete failure are introduced to building heating/cooling billing system for detection and identification. Square prediction error (SPE) statistic is used to detect sensor faults in the system. Then, sensor validity index (SVI) was employed to identify faulty sensors. Finally, a reconstruction algorithm is presented to recover the correct data of faulty sensor in accordance with the correlations among system variables. A program for the FDD and data recovery method is developed and employed in the heating/cooling billing system of a real small-scale laboratory building to test its applicability and effectiveness. Validation results show that the proposed FDD and data recovery method is correct and effective for most faults in building heating/cooling billing system.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Billing system has become essential instrument to collect and manage the information of energy consumed by each terminal of a building heating/cooling system. Heating/cooling systems are installed to ensure indoor thermal comfort and air quality in residential, office and commercial buildings. These heating/cooling systems consume a large amount of energy, part of which is wasted due to improper setting of indoor temperature, unreasonable introducing fresh air, excessive opening of windows, extra operation time of terminals, etc. These buildings are often rented/used by various leaseholders/owners such as companies or stores. In order to reduce energy consumption and improve energy efficiency of building heating/cooling system, billing systems are installed in more and more office and commercial buildings to meter the heating/cooling energy consumed by each leaseholder/owner. Leaseholder/owner should pay heating/cooling charge according to the metering of billing system. This promotes the leaseholders/owners to take some energy-saving actions and measures to reduce the energy consumed by their heating/cooling terminals and their expenditure of heating/cooling. The correctness is the key of exact metering, fair and reasonable tolling for a billing system. Billing system consists of many sensors such as temperature sensors and flow sensors. Sensor faults bring on the incorrect metering of billing system. Therefore, it is necessary to develop a method to detect, identify and diagnose the faults occurring in the sensors of building billing system. When one sensor of billing system is faulty, the correct data of its corresponding measured variable should be recovered through the logged data by the other sensors in billing system.

Sensor faults are divided into four types: bias, drifting, precision degradation and complete failure. The bias, drifting, and precision degradation are soft failures, and the complete failure is hard failure. The types of sensor faults are shown in Fig. 1.

At present, there are two major methods for FDD: model-based methods and the pattern recognition-based methods. The model-based methods first obtain standard values of the system characteristics via system model, then judge whether the failures occur in the researched system through comparing the actual run-time characteristics with the standard characteristics, and according to the characteristic features of deviation. It is important to notice that the precondition of the model-based methods is the need for a relatively precise mathematical model. Many researchers have investigated in this field, like Ann [1], Dexter [2], Stylianou and Nikanour [3]. The pattern recognition-based methods first learn all operation conditions (regardless of whether there is a fault), and then use various heuristic reasoning for the existence of faults to make a judgment, aiming at some concrete operation condition. For example, Tzafestas [4] put forward expert systems,
Many advanced methods have been proposed for the detection and diagnosis of sensor faults. The model-based method [8] is most commonly used in modern FDD schemes. Usoro et al. [9] used this approach to detect an abrupt bias in a room temperature sensor. Wang [10] presented a conservation law based sensor fault detection, diagnosis and evaluation (FDD&E) strategy for typical building chilling system. Wavelet analysis was presented to isolate the fault of flow meters in typical building chilling system with several flow meters [11]. Fault detection using principal component analysis (PCA) has been recently investigated in some areas. PCA-based FDD scheme was presented for fault detection and identification and data recovery of flow meters and temperature sensors in typical building central air-conditioning systems [12,13]. PCA method was also employed in sensor fault detection and diagnosis for air-conditioning units [14,15]. Recently, a fault detection technique based on PCA was developed for air source heat pump water chiller/heaters [16].

PCA method [17,18] is one of the popular statistical process control (SPC) methods, in which a number of related variables are transformed to a smaller set of uncorrelated variables. PCA is a useful multivariate analysis technology, which can be used for data compression, reduction of the data dimension, feature extraction, and image compression. PCA produces a lower dimensional representation in a way that preserves the correlation structure among the process variables, and is optimal in terms of capturing the variability in the data [19].

All of these FDD strategies were applied to monitor the health of sensors or components during the operation of HVAC or other systems. These investigations were carried out on simulation platforms or experimental setups. A few of investigations are based on real systems or the field data from real systems. By now, no investigation has addressed fault detection and diagnosis for a billing system of building heating/cooling system, especially for a real billing system. It is worthy of being noted that no investigation is particularly conducted on the recovering algorithm for the false data logged by billing system. This paper presents a PCA-based FDD and data recovery method for real building heating/cooling billing systems. After energy meter faults are detected in billing system, fault identification and reconstruction are employed to locate the faulty sensor, and reconstruct the correct data for faulty sensor in accordance with the correlations among all sensors, assuming that other sensors are fault free. The FDD and reconstruction focuses on identifying faulty sensors and reconstructing sensors based on the measurement correlation in billing system. Sensor reconstruction is both used to validate the original measurement by the model built by PCA and to recover the correct data of faulty sensor. A parameter called unreconstructed variance (URV) is calculated to estimate the goodness of the reconstruction and determine the optimal number of principal components. A sensor validity index (SVI) based on the most sensitive residual to sensor faults is employed for sensor fault identification. Three types of sensor faults are considered in this paper: bias, drifting and complete failure. This method of FDD and data recovery for building heating/cooling billing system is implemented and validated in a real small-scale laboratory building.

2. Methodology for FDD and data recovery

2.1. PCA and its modeling process

PCA method constructs new integrated variables (namely, principal components) consisted by the linear combination of the original variables, primarily through eigenvalue decomposition of the system variables covariance matrix. Then, a certain number of principal components are chose to approximately represent the researched system, in guarantee of as less loss of system information as possible. In this way, the basic correlation among the original variables is extracted and the dimension of system is reduced.

Suppose \( x \in \mathbb{R}^m \) \((x = [x_1, x_2, \ldots, x_m]^T)\) is a vector of \( m \) original variables, and \( X \in \mathbb{R}^{n \times m} \) is the matrix of \( n \) samples (rows) of the \( m \) original variables (columns). Using PCA, the data matrix \( X \) can be decomposed as follows:

\[
\begin{align*}
X & = \bar{X} + E \\
\bar{X} & = TP^T \\
E & = TP^T
\end{align*}
\]

where, \( \bar{X} \) is principal component subspace (PCS), it represents the correct direction of the measured vectors. \( E \) presents residual subspace (RS), it is the direction of faulty measurements. When the measurements are fault free, \( E \) is noise or uncertain disturbance mostly. \( \bar{T} \) is score matrices, \( \bar{T} \in \mathbb{R}^{n \times l} \), \( T = XP \). \( P \) is loading matrices, \( P \in \mathbb{R}^{m \times l} \). \( l \) is principal components (PCs) number of the model. The columns of \( P \) are eigenvectors of the correlation matrix associated with the \( l \) largest eigenvalues \( \lambda_i \), and the columns of \( P \) are the remaining \( m - l \) eigenvectors. So the matrix \( [PP] \) is orthonormal, and \([TT]\) is orthogonal as well. The PCA used in this paper only adopts the loading matrices \( P \).

The modeling process using PCA is shown in Fig. 2 and summarized as follows [16,20,21]:

1. The filtration and normalization for the data of original measured variables.
2. Calculation of the covariance matrix \( \Sigma \).

![Diagram](image-url)
In PCA, covariance matrix $\Sigma$ must be obtained first. But the covariance matrix $\Sigma$ is always unknown in practice; it is usually estimated from the samples of variables under normal condition.

Suppose $\mathbf{X} \in \mathbb{R}^{n \times m}$ is the matrix of $n$ samples (rows) of the $m$ original variables (columns) under normal operation, as shown by Eq. (4):

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]^T = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\
x_{21} & x_{22} & \ldots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix}$$

(4)

According to the statistics, an unbiased estimation of $\Sigma$ is calculated as:

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{0})(\mathbf{x}_i - \mathbf{0})^T$$

(5)

For the convenience of calculation, the columns of $\mathbf{X}$ are normalized to zero mean and unit variance. In this way, the average value of $\mathbf{x}$ is zero ($\mathbf{x} = \mathbf{0}$), and Eq. (5) can be written as follow:

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$$

(6)

As long as a certain number of samples under normal operation are collected, covariance matrix $\Sigma$ can be estimated using Eq. (6).

1. Eigenvalue decomposition of covariance matrix $\Sigma$, to obtain the $m$ eigenvalue $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ and the corresponding eigenvector matrix $\mathbf{P}$.
2. Determine an optimal number $I$ of PCs.
3. Choose loading matrices $\mathbf{C}$ according to $I$.
4. The projection matrices $\mathbf{C}$ and $\mathbf{C}$ loading are calculated by $\mathbf{C} \mathbf{X} = \mathbf{P}\Sigma^H$. As shown in Eqs. (7) and (8). The original $m$ data space is substituted by the $I$ PCs and $m - I$ RS, the correlations of variables are removed.

$$\mathbf{C} = \mathbf{P}^T \Sigma^H$$

(7)

$$\mathbf{C} = \mathbf{P}^T \Sigma^H = (\mathbf{I} - \mathbf{C})$$

(8)

After the PCA model has been built, when new measurements are collected, the PCA model can be used for fault detection.

2.2. Fault detection

Using PCA, a new sample vector can be decomposed into two parts:

$$\mathbf{x} = \mathbf{\bar{x}} + \mathbf{x}$$

(9)

$$\mathbf{\bar{x}} = \mathbf{C} \mathbf{x}$$

(10)

$$\mathbf{x} = \mathbf{C} \mathbf{x}$$

(11)

where $\mathbf{\bar{x}}$ is the projection of the sample vector $\mathbf{x}$ on the PCS, and $\mathbf{x}$ is the projection of the sample vector $\mathbf{x}$ on the RS. In normal operation conditions, normal data variation occurs in $\mathbf{\bar{x}}$, and noise mainly occurs in $\mathbf{x}$. But in faulty operation conditions, abnormal variation may occurs in $\mathbf{x}$, which increases the projection on the RS significantly. Therefore, the magnitude of the projection of a measurement $\mathbf{x}$ on the RS can be used for detecting abnormal conditions.

Squared Prediction Error (SPE) represents the sum of squares of the distance of $\mathbf{x}$ from the PCs that the PCA model defines. SPE is called Q-statistic as well. It can be defined as Eq. (12):

$$\text{SPE}(\mathbf{x}) = ||\mathbf{\bar{x}}||^2 = (\mathbf{x} - \mathbf{C}\mathbf{x})^T(\mathbf{I} - \mathbf{C})\mathbf{x}$$

(12)

As shown in Eq. (12), the SPE is mainly used to detect the RS, which makes the SPE be calculated directly by the PCA model and the measurement vector $\mathbf{x}$.

$\delta^2$ denotes a confidence limit or threshold for the SPE, it can be defined as Eq. (13) [17]:

$$\delta^2 = \hat{\theta}_1 \left[ \frac{2\hat{h}_2 \hat{h}_2^2}{\hat{\theta}_1} + 1 + \frac{\hat{\theta}_2 (\hat{h}_0 - 1)}{\hat{\theta}_1} \right]$$

(13)

$$\hat{\theta}_1 = \sum_{j=1}^{m} \hat{\lambda}_j \quad \hat{\theta}_2 = \sum_{j=1}^{m} \hat{\lambda}_j^2 \quad \hat{\theta}_3 = \sum_{j=1}^{m} \hat{\lambda}_j^3$$

(14)

$$\hat{h}_0 = 1 - \frac{2\hat{h}_1 \hat{\theta}_1}{3\delta^2}$$

(15)

where $c_i$ is the confidence limit for the $(1 - \alpha)$ percentile in a normal distribution, $\hat{\lambda}_i$ is eigenvalues of the covariance matrix $\Sigma$.

This paper suggests: when SPE($\mathbf{x}$) < $\delta^2$, the system is under normal operation condition, when SPE($\mathbf{x}$) > $\delta^2$, an abnormal condition exists.

2.3. Data recovery of faulty sensor

Assume that the $i$th variable of sample $\mathbf{x}$ is faulty, and $\mathbf{x}_i$ is calculated by Eq. (10). $\mathbf{x}_i$ is an estimation of the correct variable value $\mathbf{x}_i$. $\mathbf{x}_i$ also contains fault, but comparing to $\mathbf{x}_i$, the failure of $\mathbf{x}_i$ is smaller. Therefore, $\mathbf{x}_i$ is closer to $\mathbf{x}_i$ than $\mathbf{x}_i$. Replacing $\mathbf{x}_i$ with $\mathbf{x}_i$, Eq. (10) is used to calculate the estimation of $\mathbf{x}_i$ continually, then the new estimation is closer to $\mathbf{x}_i$. After repeated many times iteration, the estimated value obtained will approximates to $\mathbf{x}_i$.

The iterative process can be written as:

$$\mathbf{x}_i^{new} = [\mathbf{c}_i^T \ 0 \ \mathbf{c}_i^T] \mathbf{x} + c_i \mathbf{x}_i^{old}$$

(16)

where, $[\mathbf{c}_i^T \ 0 \ \mathbf{c}_i^T]$ is the vector that after the $i$th column of matrix $\mathbf{c}$ is replaced with $\mathbf{0}$. It can be proved that the iterative always converges to [22,23]:

$$\mathbf{x}_i = [\mathbf{c}_i^T \ 0 \ \mathbf{c}_i^T] \mathbf{x} \frac{1}{1 - c_i}$$

(17)

where, $c_i \neq 1$, if $c_i = 1$, it shows that this variable and other variables are not relevant. This variable is isolated variable, and cannot be reconstructed by other variables.

2.4. Fault identification

When fault occurs, the vector samples can be expressed as:

$$\mathbf{x} = \mathbf{x}^* + f \hat{z}_i$$

(18)

where, $\mathbf{x}^*$ represents the part of the normal measurements, $f$ is fault size, $\hat{z}_i$ is fault direction which is denoted with a unit vector.

Changes of reconstructed SPE ($\mathbf{x}_i^*$) can be used to identify faults [23]. To the measured value $\mathbf{x}$, the SPE ($\mathbf{x}$) will increase significantly when the fault occurs. Fault reconstruction is the process gradually moving towards PCS along the fault direction. Therefore, if the direction of fault reconstruction is precisely the fault occurred direction, the reconstructed SPE ($\mathbf{x}_i^*$) will be significantly reduced; if the direction of fault reconstruction is not the direction of fault, the SPE ($\mathbf{x}_i^*$) will not significantly change. This paper assumes there is only one fault occurred, Sensor Validity Index (SVI) can be used for fault identification:

$$\text{SVI}_j = \frac{\text{SPE}(\mathbf{x}_j^*)}{\text{SPE}(\mathbf{x})}$$

(19)

where, $\mathbf{x}_i^*$ is the vector of measurement $\mathbf{x}$ along the direction of the $j$th data vector after reconstruction.
Generally speaking, if the value of $SVI_i$ is less than 0.5, $\zeta_i$ is considered to be the direction of fault; otherwise, if the value of $SVI_i$ is greater than 0.5, $\zeta_i$ is considered not to be the fault occurred direction.

2.5. Determining optimal number of principal component

The determination of principal component number is the most important step in building PCA model. The number of PCs directly influences the result of PCA-based fault detection method [24–26]. If the less PCs are retained, the more variances on the RS are retained, which lead to the faults' effect on RS is small, and faults are difficult to be detected. Also, more PCs means heavier computation burden in modeling.

This paper adopts unreconstructed variance (URV) to determine the number of PCs that will generate the best reconstruction [12,26].

$$u_j = \text{Var}(\zeta^2_j(x - x_j) = \frac{\zeta^2_j(I - C)^s(I - C)\zeta_j}{[\zeta^2_j(I - C)\zeta_j]^2})$$  \hspace{1cm} (20)

where, $\zeta_j$ is fault direction vector, $x$ is measurement vector, $x_j$ is the reconstructed value of $x$ in the fault direction $\zeta_j$. And $u_j$ is the URV in the fault direction $\zeta_j$. The smaller the $u_j$ is, the better the reconstruction. To consider all fault directions simultaneously, the $u_j$ should be minimized to determine an optimal number of PCs.

$$\min \left( \sum_{j=1}^{m} u_j \right)$$  \hspace{1cm} (21)

where, $m$ is the number of sample variables. Calculate $\sum u_j$ separately by choosing different number of $l$, finally the number of $l$ corresponding to the minimum $\sum u_j$ is adopted as the optimal PCs.

3. System description

A heating/cooling system of air-source heat pump is installed in a laboratory office building of Hunan University. The laboratory building adopts building energy management and control system (EMCS) to monitor and bill the air-conditioning system, shown in Figs. 3 and 4. A heating/chilling plant located outdoor is an air-source heat pump water chiller/heater with a design cooling capacity of 63,000 W, a centrifugal water pump of single stage type, and an auxiliary electric heater to ensure the requirements of heating in winter, shown in Fig. 5. The air-conditioning system is an air-water system, which used for cooling in summer and heating in winter. The indoor areas use fan coils, and the system does not design a separate fresh air system, the fresh air is supplied by outdoor air infiltration instead. The heated/chilled water system adopts a closed mechanical circulation. The EMCS controls the start/stop of heat pump unit and supply water pump automatically, according to building return water temperature and supply/return water differential pressure. This heating/cooling system is both used for experiment research and air-conditioning for the laboratory office building.

A billing system embedded in the EMCS is used to meter and log the heat/cooling amount consumed by each user of the air-conditioning system. The users in the building should monthly pay their heating/cooling charge according to the heat/cooling amount logged by the billing system. The billing system includes four ultrasonic energy meters, which consist of ultrasonic flow sensor and supply/return water temperature sensor; five electronic energy meters, which consist of electronic flow sensor and supply/return...
To building

From building

Outdoor air temperature

Auxiliary electric heater

Air-source heat pump unit

Heat pump

Energy meter

Flow sensor

Temperature sensor

Differential pressure sensor

Fig. 5. Schematic of air-source heat pump heating/cooling system.

water temperature sensor. Energy meters are installed on the following locations: building supply and return water main lines, return water tanks on the first, second and third floor, return water pipes of the hall and laboratory room on the first floor, and return water pipes of three laboratory rooms on the third floor. The signal of supply water temperature sensor in main line is utilized by all energy meters because supply water pipe is not long for the small-scale laboratory building. The return water temperature sensor for each energy meter is installed at the same position as the corresponding energy meter. Besides, a differential pressure sensor is installed between the building water supply and return main lines. And an outdoor air temperature sensor was installed at about 4 m away from the air-source heat pump water chiller/heater, shown in Fig. 5. Also, indoor air temperature sensors are respectively installed on the return air intakes of room fan coil units. The computer for EMCS is installed in the control room on the second-floor.

In order to monitor the health of sensors in the billing system and ensure the fair and exact billing, the FDD and data recovery method described in Section 2 is proposed. A program based on the method is developed and online connected to the billing system to detect and diagnose the sensor faults in the billing system and recover the correct data of faulty sensors. The program consists of five modules of data filtering, PCA modeling, fault detection, fault identification and reconstruction (data recovery). The following sensor signals in the billing system are adopted by the program to test and validate the proposed FDD and data recovery method: ultrasonic flow sensors and supply/return water temperature sensors of four ultrasonic energy meters, differential pressure sensor, outdoor air temperature sensor and indoor air temperature sensor in Room 206. The sampling interval for the measured data was 30 s, and the logged data were filtered by Exponential Weighted Moving Average (EWMA) method. The modules for PCA modeling, fault detection, fault identification and data recovery are described mathematically in Section 2.

4. Test and validation

As one kind of black-box models, PCA model depends on the data of system process. The quality of the data samples directly determines the sensitivity and accuracy of PCA model used for FDD and data recovery. It is necessary to have enough training data to capture the correlations among system variables and to meet the needs of SPE statistic calculation. However, considering the operation conditions of heating/cooling system, the coverage range of training data cannot be too wide. If there are great changes in operating conditions, the sensitivity of PCA model may be degraded because the model error is much larger compared with the data change caused by faults. In order to improve the quality of training data and the sensitivity of the PCA model, it is necessary to choose training data carefully. Generally, steady-state data is chosen for PCA modeling, because the consistency among the system parameters can be reflected correctly under the steady-state condition.

The heating/cooling system of the laboratory building operates from 9:00 am to 7:00 pm. Since tests were performed at the site of a real building system, the driving conditions were uncontrollable. In order to get steady-state data, typically, the tests and validations were conducted from around 11:00 am to 4:00 pm over 30 days from July to August, 2007. The data in transient conditions, such as the startup and closedown periods, was not used. The data in the first 3 days under normal condition was used to train PCA model. The data in the 4th day (The first 300 samples were under normal condition and the 301st–600th samples were introduced 5 °C bias to supply water temperature sensor) was used to test the fault detection index – SPE. The data in the rest days was used to test the capacity of the PCA model for detection, identification and reconstruction (data recovery) of various faults. The test results for typical faults are given in this section. The data in the 5th, 6th and 7th days was respectively introduced with the bias of 3 °C, 0.5 °C and 1 °C to building return water flow sensor. The data of the 8th and 9th days was added with a drifting of about 400 l/h to second-floor flow sensor. The data in the 10th day was introduced complete failure (0.0 Pa) to the differential pressure sensor between building supply and return water main lines. The case for one faulty sensor was considered in all tests.

4.1. Training of PCA model

Covariance matrix $\Sigma$ was calculated firstly using the logged data in the first 3 days, then eigenvalue decomposition of $\Sigma$ was carried out to obtain the eigenvector matrix $P$. In order to determine the optimal number of PCs, the total URV was calculated by reconstruction based on the measurements. Fig. 6 shows the variation of the total URV versus the number of PCs. The total URV is minimal when the number of PCs is 1 and 3. Three principal components are reasonable and retained in the PCA model here. After the optimal number of PCs was determined, the loading matrices $P$, and the projecting matrices $C$ and $C'$ were obtained accordingly. In this way, the PCA model of the billing system was built up. The threshold $d^2$ of the SPE statistic was obtained using the PCA model. The fault detection threshold SPE at 95% confidence level was calculated to be 21.2680.

The data logged in the 4th day was used to test the SPE index of fault detection for the FDD method. The result is shown in Fig. 7. At first, no fault was added to any sensors, the SPE of the first 300 samples was below its threshold $d^2$. It is obvious that the system was under normal operation condition. But after a bias of 5 °C was added to building supply water temperature sensor, the SPE increased significantly and exceeded its threshold from the 301st
sample. It can be found that the there is fault occurring and the fault was detected soon after it occurs. Besides, it also indicates that the PCA model excellently captures the major correlation and variances among the variables in the billing system as well.

4.2. Tests for FDD and data recovery

In order to verify the performance of the program to find faulty sensors and recover their correct data, a large number of sensor faults were introduced to the billing system and the field data logged by the billing system to test and validate the effectiveness and robustness of the strategy of FDD and data recovery method and its program. Some typical study cases are presented as follows.

4.2.1. Test I – bias

In the 5th day, a bias of 3 °C was introduced to building return water temperature sensor from the 301st sample. The fault detection result is shown in Fig. 8. After the bias was added, the SPE increased significantly and exceeded its threshold $d^2$ from the 301st sample. It is quite obvious that there was fault occurred in the system. In order to identify the faulty sensor, the SVI of each sensor was also calculated by reconstructing its measuring signal using the measurements of the remaining sensors. After the fault was added, the SVI of building return water flow sensor was close to 0, shown in Fig. 8b. At the same time, the SVI of other sensors, such as the flow sensor in supply main line, was still close to 1, shown in Fig. 9. The SVI of other sensors are not shown. The SVI shows that the fault occurred on the building return water flow sensor, and the other sensors were fault free. Fig. 8c shows the reconstructed SPE were significantly below its threshold. This shows that the recovered data did not contain fault any more. This result is as expected, which also demonstrates the proposed FDD and data recovery method.
In the 6th day, a bias of 0.5°C was introduced to the building return water flow sensor from the 301st sample as well. The fault detection result is shown in Fig. 10. However, SPE were still below its threshold $\delta^2$ after the bias was introduced. The fault cannot be detected. This shows that the built PCA model is not able to detect small sensor faults.

In the 7th day, a bias of 1°C was introduced to building return water flow sensor from the 301st sample as well. The fault detection result is shown in Fig. 11. After the bias was introduced, SPE was around its threshold $\delta^2$. Fig. 11b shows SPE was significantly below its threshold after reconstructed by the remaining sensors. The fault detection ability of the built PCA model is a bias of 1°C, which is the detectable fault margin of temperature sensor in the billing system.

4.2.2. Test II – drifting

In the 8th day, a drifting of 10% (about 400 l/h) was introduced to the second-floor flow sensor from the 480th sample. The fault detection result is shown in Fig. 12. At first, SPE increased gradually, but was below its threshold $\delta^2$. Finally, it exceeded its threshold $\delta^2$ in the 9th day. It shows that there was fault occurred in the system. It also indicates that the built PCA model is not able to detect small sensor faults. With a long-term accumulation of drifting in a certain period of time, it reaches a certain quantity viewed as bias fault. The SVI of each sensor was also calculated by reconstructing using the measurements of the remaining sensors. From the 481th sample, the SVI of all flow sensors decreased significantly (as shown in Fig. 12b), which was below 0.5 and close to 0. At the same time, the SVI of other sensors was still close to 1 (SVI of other sensors are not shown). The SVI shows that the fault occurred among flow sensors, and the other sensors were fault free. Thus, other approaches such as wavelet analysis are needed to isolate faulty flow sensors [11]. Fig. 12c shows the reconstructed SPE from the measurement of the remaining sensors was below its threshold as expected. This indicates that the recovered data did not contain fault any more.

4.2.3. Test III – complete failure

In the 10th day, complete failure was introduced to the differential pressure sensor between building supply and return water main line from the 401st sample. The fault detection result is shown in Fig. 13. Immediately after fault presented, SPE soared and exceeded its threshold $\delta^2$ from 401st sample. It is quite obvious that there was fault occurred in the system. The SVI of each sensor was also calculated. Although the SVI of differential pressure sensor was below 0.5 before the fault was added, it decreased significantly after the fault presented, and it was close to 0. At the same time, some SVI values of indoor air temperature sensor in
Room 206 were around 0.5. The SVI of other sensors, such as outdoor air temperature sensor and flow sensor in main line, was still close to 1, shown in Fig. 14. The SVI shows that the fault occurred on the differential pressure sensor, and other sensors were fault free. Fig. 13b shows that the SPE was below its threshold after reconstructed from the remaining sensors. This indicates that the recovered data did not contain fault any more.

5. Conclusions

Building heating/cooling billing system can promote the leaseholders/owners to take some energy-saving measures to reduce their charge for heating/cooling. Billing system consists of many sensors such as temperature sensors and flow sensors. Sensor faults bring on the incorrect metering. For a billing system, sensor’s correctness is the key of exact metering and reasonable tolling. It is significant work to develop a method and program for detecting and identifying faulty sensors in building billing system, and recovering the correct data from the measurements of other sensors in billing system when one sensor is faulty in billing system.

A PCA-based FDD and data recovery method is proposed for building heating/cooling billing system to detect sensor fault, identify faulty sensor and recover the data of faulty sensor. A large amount of data is collected under normal and steady-state operation condition of heating/cooling system through a real billing system of EMCS to train a PCA model of system. SPE statistic based on the PCA model is used to detect fault occurring in billing system by comparing SPE with its threshold. After a fault is detected, SVI of each sensor is calculated to identify which sensor is faulty. When its SVI is close to 1, the sensor is normal. On the contrary, when its SVI is close to 0, the sensor is faulty. Then, the faulty sensor is reconstructed from the measurements of the remaining normal sensors and its correct data is recovered.
A program was developed and employed in a real heating/cooling billing system of laboratory office building to test and validate the proposed method of FDD and data recovery. Various faults with different levels are introduced to the field data logged by the billing system. The program successfully detected most sensor faults, identified temperature sensor faults and correctly recovered the data of faulty sensors. The faulty flow sensors need other approaches such as wavelet analysis to isolate. The FDD method cannot detect very small fault in billing system. A number of tests and validations show that the proposed method of FDD and data recovery is effective for most faults and that it is able to recover the correct data of faulty sensor by reconstructing from the measurements of remaining sensors.

Acknowledgments

The research work presented in this paper is financially supported by a grant of Building Intelligent Control Research Fund (No. JRP0901). The authors would like to thank Mr. W. H. Chan and Dr. J. Y. Qin of Swire Properties Management Limited for the essential support to the research work.

References


