This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Fault Diagnosis of Components and Sensors in HVAC Air Handling Systems with New Types of Faults

Ying Yan, Student Member, Peter B. Luh, Fellow, IEEE, and Krishna R. Pattipati, Fellow, IEEE.

Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA

Corresponding author: Ying Yan (e-mail: ying.yan@ uconn.edu).

The work was supported in part by the National Science Foundation under Grant CCF-1331850. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ABSTRACT Air handling systems are key sub-systems of Heating Ventilation and Air Conditioning (HVAC) systems. They condition and deliver air to satisfy human thermal comfort requirements and provide acceptable indoor air quality. Faults in their components and sensors may lead to high-energy consumption, poor thermal comfort and unacceptable indoor air quality. Additionally, new types of faults may falsely be identified as known types. Identifying failure modes and their severities with low false identification rates is thus critical to know what faults occur and how severe they are. However, this is challenging since (1) classifying both failure modes and fault severities generates many categories of failures, leading to high computational requirements; (2) updating model parameters to adapt to changing environments requires accurate recursive equations that are hard to obtain; and (3) model errors and measurement noise may cause high false identification rates in detecting new types of faults. In this paper, failure modes are identified by Hidden Markov Models (HMMs) and fault severities are estimated by filtering methods, leading to a decrease in the number of HMM states and low computational requirements. To adapt to changing environments, a new online learning algorithm is developed. In this algorithm, HMM parameters are obtained based on their posterior distributions given new observations, thereby avoiding the need for accurate recurrence equations. To identify new fault types with low false identification rates, a robust statistical method is developed to compare current HMM observations with those expected from existing states to obtain potential new types, and then confirm new types by checking whether observations have a significant change. Physical knowledge is then used to find the reason for the new fault type. Experimental results show that failure modes and fault severities of both known and new types of faults are identified with high accuracy.

INDEX TERMS Fault diagnosis, HVAC air handling system, online learning algorithm, hidden Markov model, new fault types

I. INTRODUCTION

Heating, Ventilation and Air-Conditioning (HVAC) systems account for 57% of energy used in the U.S. commercial and residential buildings [1]. Air handling systems are key sub-systems in HVACs. They condition and deliver air to rooms to satisfy human thermal comfort and provide acceptable indoor air quality. An air handling system consists of a mixing box, filters, cooling/heating coils, ducts and fans, as shown in Fig. 1. The mixing box consists of an Exhausted Air (EA) damper, a Return Air (RA) damper and an Outdoor Air (OA) damper; and controls the ratio of the return airflow to the outdoor airflow. Filters remove solid particulates such as dust from air, and cooling/heating coils condition the mixed air to satisfy human thermal comfort requirements. The supply fan delivers conditioned air to Variable Air Volume (VAV) boxes, which control temperatures and airflow rates delivered to rooms. The return fan then delivers return air from rooms to the outside and to the mixing box. In the air handling system, various sensors measure air/water flow rates, temperatures and humidity ratios, including the supply air temperature sensor, the supply air humidity ratio sensor, the supply air mass flow rate sensor, the return air temperature sensor, the return air humidity ratio sensor and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

the return air mass flow rate sensor, as shown in Fig .1. Given sensor readings, building management systems control HVACs to satisfy human thermal comfort requirements and provide acceptable indoor air quality. Faults in these components and sensors may result in high energy waste and poor thermal comfort and unacceptable indoor air quality. Fault diagnosis includes the identification of failure modes and estimation of their severities, thus is critical. Identification of failure modes helps to know which faults have occurred. Estimation of fault severities allows building management personnel to know how severe the faults are, and helps in scheduling and dispatching maintenance crews to repair or replace failed components/sensors.

Fault diagnosis, however, is challenging because (a) capturing both failure modes and fault severities may generate many categories, leading to high computational requirements; and (b) changing environments, e.g., weather and occupants, may cause sensor readings to change drastically even without faults, resulting in high false identification rates; and (c) it is hard to detect new fault types and find their reasons since new fault types are not captured in classifiers.

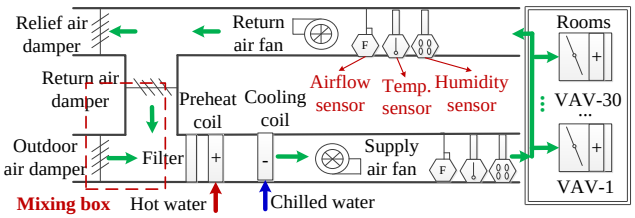

FIGURE 1. Schematic diagram of an air handling system and VAV boxes in HVAC.

This paper focuses on diagnosis of known and new types of faults in components and sensors of HVAC air handling systems. Based on practice in HVACs [2], [3], 17 faults of the OA damper, the EA damper, cooling coil, supply fan, return fan and ducts are considered. To test our method, two of these faults are considered as new fault types, and others are considered as known types. For sensors, if the output signal of a sensor differs from the correct value by a constant, the constant is called as the sensor bias. If the output signal slowly changes independent of the measured property, this is defined as the sensor drift. Based on practice in [4], [5], bias and drift of sensors mentioned before are considered. Some of these faults evolve slowly, such as a decrease in fan efficiency. Severities of such faults are estimated to schedule repair or replacement for the failed components or sensors. In Section 2, existing fault diagnosis methods are reviewed. In these methods, filtering methods, e.g., Kalman Filters (KFs) and Particle Filters (PFs), developed based on Bayes rule, and estimate fault-related parameters based on their previous estimates and new sensor readings, leading to accurate estimates. Failure modes are identified if they have different signatures in estimates. Additionally, fault severities are accurately identified based on continuous estimates. However, fault-related parameters may only reflect certain failure modes but not all of them, thus may not be enough to identify failure modes. Unlike filtering methods, Hidden Markov Models (HMMs) set combinations of failure modes and fault severities as discrete states, thus all faults and their severities can be identified by estimating states. However, the number of combinations increases drastically with increase in numbers of failure modes, resulting in many HMM states. Thus high computational effort is required. Additionally, estimates of states are discrete, and thus may not be accurate enough to identify fault severities. Most of existing methods do not adapt to changing environments, and rare papers investigate diagnosis of new fault types.

To identify both known types and new types of faults while adapting to changing environments, a systematic method is developed as shown in Fig. 2. In this method, to identify known types of failure modes and their severities with low computational requirements while adapting to changing environments, HMMs and filtering methods are used to identify failure modes and fault severities separately as shown in Section 3. Since only failure modes are captured in HMMs, a few states are involved, and the method is computationally efficient. To adapt to changing environments, existing methods require accurate recurrence equations to update HMM parameters. These equations are derived based on some assumptions (e.g., homogeneity over a time window and large number of samples) that may not always be satisfied. To address this issue, HMM parameters are generated based on their posterior distributions given new HMM observations, thereby obviating the need for accurate recurrence equations. To identify fault severities of components and sensors with high resolution, fault-related parameters and sensor bias/drift are estimated via filtering methods, e.g., KF or PF, given the identified failure mode.

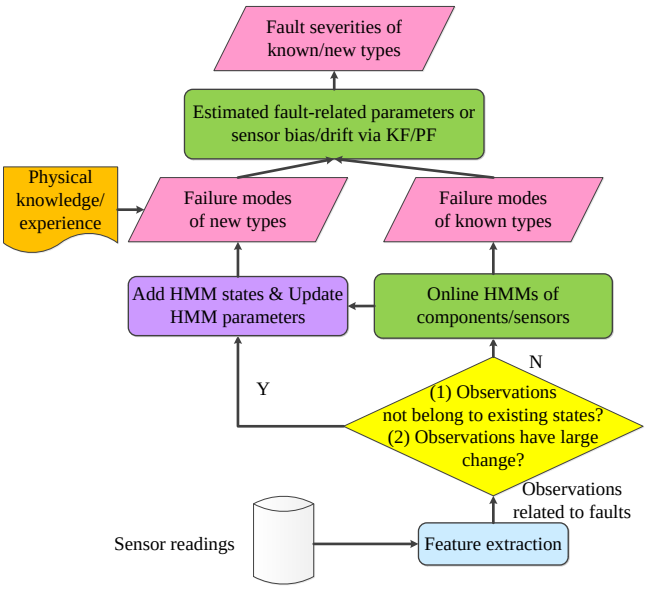

FIGURE 2. Flow chart of the fault diagnosis method.

Unlike known types of faults, new types of faults are not captured in HMMs, thus the method presented in Section 3 cannot detect and find reasons of new fault types. In Section

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

4, a method is developed to distinguish new fault types from existing ones, and then find reasons of detected faults and estimate their severities based on physical knowledge as shown in Fig. 2. To detect new fault types, the HMM-based method in Section 3 is improved by adding new HMM states dynamically to capture new fault types. Because of model errors and measurement noise, it is hard to detect new fault types with low false identification rates. Considering that occurrence of a new type is a state transition, it causes large changes in observations. Additionally, observations corresponding to new types deviate from those belonging to existing types. Thus a robust method is developed to distinguish new types from known types in these two perspectives. To check changes in observations, Kullback-Leibler (KL) divergence is used, since it measures how the distribution diverges from that represented by previous observations. To test whether current observations deviate from those of existing types, a robust Bayes-factor-based hypothesis testing is developed. To find the reason of the detected fault, changes of fault-related sensor readings are analyzed based on physical knowledge to find possible fault types as shown in Fig. 2. The new fault type is identified by eliminating known types from possible ones.

In Section 5, simulation data of a small building and ASHRAE-1312 data are used to test our method. Experimental results show that our method can identify known types and new types of failure modes and their severities with low false identification rates.

II. LITERATURE REVIEW

To diagnose known types of faults in air handling systems, many methods were developed. However, papers that investigate diagnosing new fault types in air handling systems are rare.

Methods for Diagnosing Known Fault Types of Components and Sensors in Air Handling Systems. To diagnose known types of faults, multiple methods were developed, and are generally categorized into two types: model-free and model-based. Model-free methods, e.g., expert systems and decision trees, were developed based on physical knowledge and experience without establishing models. These methods infer faults by investigating cause-effect relationships between faults and their impacts. For example, expert systems employ physical knowledge and experience to generate if-then-else rules, and are widely used in HVACs, and in particular, air handling systems [6]-[8]. Decision trees established based on physical knowledge are also used. For instance, a decision tree was used to diagnose sudden and gradual faults of an Air Handling Unit (AHU) based on data from the ASHRAE project 1312-RP [9]. These methods have good explanatory capabilities for fault inference. However, developing rules for a specific system requires expertise and knowledge that may not be available. Additionally, rules are usually fixed and may not adapt to changing environments, resulting in false identifications.

In model-based methods, models are established to capture key features of systems, and can be generally categorized into three types, including (1) black-box models; (2) statistical models; and (3) physics-based models. Black-box models are established only based on data without considering physical knowledge. Black-model-based methods, e.g., Principal Component Analysis (PCA), Artificial Neural Networks (ANN), Support Vector Machine (SVM) and extreme learning machine, are widely used in fault diagnosis of air handling systems. For instance, PCA models are established for various sub-systems based on normal data. By comparing new data with these models, anomalies are detected and the fault source is identified according to models [4]. In [10], a wavelet-PCA method was developed to diagnose sudden and gradual faults of an AHU by removing influence of weather conditions. Compared with PCA, ANNs are good at classifying failure modes based on training data, and are widely used for HVACs [11]-[13]. Based on single hidden-layer feed-forward neural networks, extreme learning machine was developed by randomly selecting features for the hidden units. It has much faster learning speed compared with traditional ANNs, and was used to diagnose faults of AHUs [14]. Like ANNs, SVMs are also good at classification. They were applied to classify faults of AHUs based on estimates of fault-related model parameters [15]. These methods classify normal and failure modes only based on sensor readings or features extracted from them. State evolutions obtained from physical knowledge represent relationships among states, and thus help to identify current states. However, state evolutions are rarely considered in these methods. Moreover, structures of these classifiers are usually fixed and are not updated, and thus need to be retrained when new fault types occur. Thus, they may not adapt to changing environments, leading to false identifications. Most of existing methods focus on identifications of failure modes, but few of them consider estimating fault severities.

Compared to black-box models, HMMs capture state evolutions and distributions of sensor readings given states by statistical models. Conditions of components and sensors are considered as states of HMMs, and estimated to identify failure modes [16], [17]. However, if HMMs are used to identify both failure modes and fault severities, many HMM states are generated, leading to high computational requirements. Additionally, structures and parameters of HMMs used for fault diagnosis are usually fixed, and thus cannot adapt to changing environments [18]. To address this issue, several online algorithms were developed by updating HMM parameters based on the Baum-Welch (BW) algorithm [19], [20]. In these algorithms, recurrence equations are developed based on homogeneity over a time window and large sample assumptions. However, these assumptions may not always be satisfied.

Filtering methods, such as KFs and PFs capture both state evolutions, and relationships between states and sensor readings, by physics-based models. Fault-related parameters are modeled as states and are estimated to check whether they fall outside their control limits or not. If failure modes have different signatures, they are diagnosed [21]. However, existing models may not contain enough fault-related parameters to diagnose all faults considered. Additionally,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

estimates of parameters are continuous. They can reflect fault severities with higher resolution when compared to discrete state estimates obtained by HMMs.

Methods for Diagnosing New Fault Types. Diagnosing new fault types has two steps: detecting and finding reasons. To detect new fault types, certain statistical hypothesis testing methods can be used. For instance, detecting new types of faults is essentially equivalent to testing whether current sensor readings and previous ones are from different distributions. Bayes-factor-based hypothesis testing can be used to make such inferences [22]. In this method, marginal probabilities of belonging to a known distribution and that belonging to a different one are calculated. The ratio between the two probabilities is compared with one that is a constant threshold to check which case is more likely. Because of the constant threshold, model errors and measurement noise may cause high false identification rates. Additionally, analytical expressions of marginal probabilities for certain distributions are hard to obtain. To detect new fault types in high voltage electronic and power equipment, a novel clustering method, integrating Gaussian mixture models and k-means, was developed [23]. New fault types are detected based on confidence scores. In this method, no physical knowledge is considered for classification. Additionally, the classification is only based on sensor readings and state evolutions are not considered. Infinite HMMs, developed using Dirichlet processes, can also be used to detect new fault types [24]. In these HMMs, the number of states is determined based on data. They can capture new fault types by adding new states. However, they are purely data-driven, and no physical knowledge is used. In our problem, as in most engineering problems, physical knowledge and experience with failure modes are available. It is important to consider physical knowledge to improve robustness of the method under model errors and measurement noise. Papers that investigate identifying reasons of new fault types are rare.

III. IDENTIFICATION OF FAILURE MODES AND FAULT SEVERITIES FOR KNOWN FAULT TYPES

In this section, HMMs and filtering methods are developed to identify known types of failure modes and their severities, respectively. In subsection III-A, HMMs of components and sensors are established while capturing coupling among them. In subsection III-B, an online learning algorithm is developed to estimate HMM states to identify failure modes while adapting to changing environments. In subsection III-C, fault-related parameters and deviations of sensor readings from their normal values are estimated to identify fault severities of components and sensors, respectively.

A. HMMS TO IDENTIFY FAILURE MODES

To identify failure modes and fault severities of components and sensors, it is important to estimate their states, i.e., the normal condition or failure modes. States are estimated based on certain fault-related sensor readings including (a) temperatures, humidity ratios and mass flow rates of outdoor air, supply air and return air; (b) temperatures and mass flow rates of chilled water; and (c) power of supply and return fans. Additionally, some sensor readings cannot track set-points when faults occur. Consequently, residuals between these sensor readings and set-points are also considered. Relationships among sensor readings are represented by physics-based/gray-box models. A sensor reading can be estimated based on a model given other readings. Sensor readings and their estimates obtained from models are consistent under the normal condition, but not under faulty conditions. Residuals between sensor readings and their estimates represent the parity (consistency) relationships. Consequently, sensor readings and residuals mentioned above are considered as HMM observations. In this subsection, a cooling coil and sensors related to return air are used as examples to show how to establish HMMs for components and sensors.

HMM of Cooling Coils. A cooling coil is a coiled arrangement of tubes for heat transfer between chilled water and air as shown in Fig. 3. Fins are used to increase the heat transfer area. Valves are adjusted to control the chilled water mass flow rate. Chilled water flows through tubes, and air passes through fins. Air temperature is reduced through heat transfer between air and chilled water.

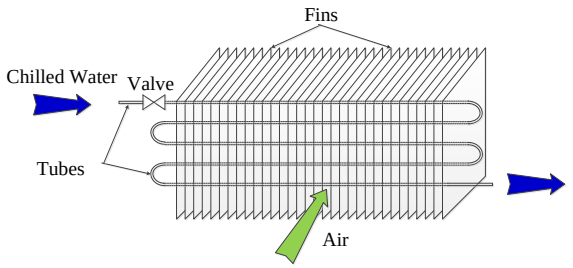

FIGURE 3. A typical cooling coil

To identify failure modes, an HMM of cooling coils is established as shown in Fig. 4.

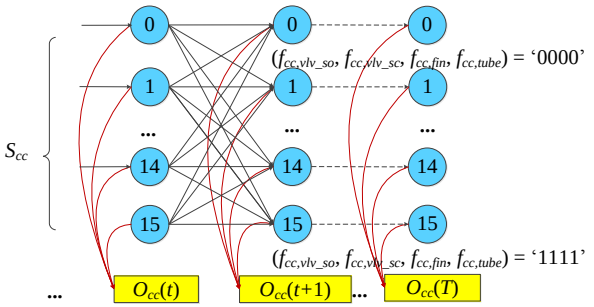

FIGURE 4. The HMM of a cooling coil

For the cooling coil, four failure modes are considered, including (a) tube fouling; (b) dust on fins; (c) valve stuck closed; and (d) valve stuck open. Since failure modes may be concurrent, the HMM state is set as ($f_{cc,tube}$, $f_{cc,fin}$, f_{cc,vlv_sc}, f_{cc,vlv_so}), where tube fouling is denoted by $f_{cc,tube}$; dust on fins is denoted by $f_{cc,fin}$; valve stuck closed is denoted by f_{cc,vlv_sc};

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

and valve stuck open is denoted by f_{cc,vlv_so}. State S_{cc} = 0 is equivalent to ($f_{cc,tube}$, $f_{cc,fin}$, f_{cc,vlv_sc}, f_{cc,vlv_so}) = '0000' which means that no faults present in the cooling coil. ($f_{cc,tube}$, $f_{cc,fin}$, f_{cc,vlv_sc}, f_{cc,vlv_so}) = '1111' means that all faults have occurred. The HMM thus has 2^4 = 16 states denoted by s_{cc} = 0, 1, …, 15 as shown in Fig. 4, and the number of states is denoted by N_{cc} = 16. To estimate these states, certain fault-related sensor readings are used. For instance, the tube fault or the valve fault may cause a sudden or gradual change in the chilled water mass flow rate $\dot{m}_{chw}$. Similarly, dust on fins may cause a change in the supply air mass flow rate $\dot{m}_{a,\sup}$. Moreover, faults may cause a difference between the supply air temperature $T_{a,sup}$ and its set-point. Thus, the residual $\Delta T_{spt,sup}$ between $T_{a,sup}$ and its set-point is used as an HMM observation. Additionally, residuals representing parity relationships are considered. For instance, a cooling coil model was developed in [24] as

$$\frac{\dot{m}_{a,mix}\cdot(E_{a,mix}-E_{a,dis})}{LMTD}=\left[\frac{1}{\alpha_{a,e}\tau}+\frac{\delta_{tube}}{\lambda_{tube}}+R_f+\frac{1}{\alpha_w}\right]^{-1}(A_{tube,out}+A_{fin}), \text{ and} \quad (1)$$

$$UA_{cc}=\dot{m}_{a,\sup}\cdot(E_{a,mix}-E_{a,dis})/LMTD, \quad (2)$$

$$\text{with } LMTD=(\Delta T_{\sup}-\Delta T_{rn})/(\ln\Delta T_{\sup}-\ln\Delta T_{rn}), \text{ [25]} \quad (3)$$

where A_{fin} and $A_{tube,out}$ are the fin surface area and the tube outside surface area, respectively; $A_{tube,in}$ is the tube's inside surface area; δ_{tube} is the tube thickness; λ_{tube} is the tube's thermal conductivity; η_{fin} is the fin efficiency; $d_{tube,in}$ is the tube inside diameter; and variable $\dot{v}_{chw}$, is the chilled water volumetric flow rate. Here, $\Delta T_{sup} = T_{a,dis} - T_{chw,sup}$ and $\Delta T_{rn} = T_{a,mix} - T_{chw,rn}$. Variables $T_{a,mix}$ and $T_{a,dis}$ are the mixed air and the discharge air temperatures, respectively; variables $T_{chw,sup}$ and $T_{chw,rn}$ are the supply and the return chilled water temperatures, respectively. This model is validated by comparing it with a validated detailed physics-based model of cooling coils [26]. For validation, 10620 simulation data points from 7/18 to 9/25 are used. Variables, e.g., mixed air temperature, air mass flow rate and chilled water mass flow rate, representing conditions of a cooling coil, are input in the two models. Cooling capacities are calculated as model output since they represent performances of cooling coils. Residuals between cooling capacities of these two models are calculated, and the relative error of residuals is 0.041. The right-side of (1) depends on geometric parameters, e.g., $d_{tube,in}$ and A_{fin}. If these parameters are set to their normal values, (1) will not be valid when faults occur. Thus, the residual R_{gray} between the left-side and the right-side of (1) is used as an HMM observation. Similarly, the difference between the enthalpy of mixed air $E_{a,mix}$ and the enthalpy of discharge air $E_{a,dis}$ depends on the amount of heat exchange, and thus is determined based on geometric parameters of the cooling coil. Component faults may cause changes in the geometric parameters, leading to increase in residuals. To represent the relationship between the residual and sensor readings, a NN is established. The inputs of this NN are (a) $\dot{m}_{a,\sup}$; (b) $\dot{m}_{chw}$; (c) the difference between the mixture air temperature $T_{a,mix}$ and the discharge air temperature $T_{a,dis}$; and (d) the difference between the mixture air humidity ratio $W_{a,mix}$ and the discharge air humidity ratio $W_{a,dis}$. The output is $E_{a,mix}$ - $E_{a,dis}$. The residual $R_{E,NN}$ between $E_{a,mix}$ - $E_{a,dis}$ and its estimate obtained from the NN is considered as an HMM observation. The matrix consisting of fault-related variables mentioned above is,

$$X_{cc}=\begin{bmatrix} m^1_{a,\sup} & \dot{m}^1_{chw} & R^1_{gray} & R^1_{E,NN} & \Delta T^1_{spt,\sup} \\ . & . & . & . & \\ m^K_{a,\sup} & \dot{m}^K_{chw} & R^K_{gray} & R^K_{E,NN} & \Delta T^1_{spt,\sup} \end{bmatrix}. \quad (4)$$

Since these variables have different units, and differences among them are large, they are normalized to avoid X_{cc} becoming an ill-conditioned matrix. Additionally, these variables may be correlated, and contain redundant information. To remove redundancy, PCA is used to project the matrix into a reduced space. If m vectors represent more than 95% of the correlated information, they are considered as adequate set of "principal components" to reflect the original space [27]. The first three principal components capture 96.349% of variability in data, and thus are denoted by O_{cc} and used to estimate HMM states in Fig. 4. The HMM has three types of parameters, including (a) initial state probability distribution; (b) state transition matrix; and (c) emission probabilities. The initial state probability distribution determines the likelihood of belonging to each state at the initial time; state transition matrices represent probabilities of transition among states; and emission probabilities govern distributions of observation sets which are observations corresponding to each state. Since the HMM has 16 states, the HMM has a 16×16 state transition matrix and 16 observation sets. Similarly, HMMs of other components are established.

HMMs of Sensors Related to Return Air. For return fan, the temperature sensor, the humidity ratio sensor and the airflow sensor are considered. Since the return fan is linked with VAV boxes, the return air is the mixture of airflows of all VAV boxes. The return airflow rate is equal to the sum of those of VAV boxes,

$$\hat{m}_{a,rn}=\sum_{i=1}^{M}\dot{m}_{a,vav\_i}. \quad (5)$$

where $\hat{m}_{a,rn}$ is the estimate of the air mass flow rate of return air; variable m_{a,vav_i} is the air mass flow rate of the i^{th} VAV box; and M is the number of VAV boxes. The estimate will deviate from $m_{a,rn}$ if a sensor fault occurs. Thus the residual $\hat{m}_{a,rn}-m_{a,rn}$ is used in the HMM. Similarly, estimate of the return air temperature $T_{a,rn}$ can also be calculated based on zone temperatures of VAV boxes as,

$$\hat{T}_{a,rn}=(\sum_{i=1}^{M}\dot{m}_{a,vav\_i}\cdot T_{zone,i})/\sum_{i=1}^{M}\dot{m}_{a,vav\_i}, \quad (6)$$

where $T_{zone,i}$ is the temperature of the i^{th} zone. Residuals between estimates and measurements are considered as HMM observations. For the return air humidity ratio sensor, the humidity ratio is estimated as

$$\hat{W}_{a,rn}=(\sum_{i=1}^{M}\dot{m}_{a,vav\_i}\cdot W_{zone,i})/\sum_{i=1}^{M}\dot{m}_{a,vav\_i}, \quad (7)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

where $W_{zone,i}$ is the humidity ratio of the i^{th} zone. The residual between the estimate and the measurement is considered. For each sensor, drift and bias are considered as failure modes, and thus the HMM of a sensor has $2^2 = 4$ states. Similarly, HMMs of other sensors are established. Since each component or sensor has its own HMM, concurrent faults of different components and sensors can be identified separately based on corresponding HMMs.

Coupling among Components and Sensors. Components are linked through air or chilled water flows. A fault in a component may cause a load increase in other components, which will then be more likely to break down. For instance, valve stuck closed in a cooling coil results in decreased chilled water. The supply fan needs to provide more air to rooms, and is more likely to wear, leading to a decrease in fan efficiency. Similarly, sensor faults may cause components to work under extreme conditions. Thus, coupling among adjacent components, and coupling between components and related sensors are considered. This is done by using a coupled HMM algorithm developed in [18] with state transition matrices that are dependent on other components. For instance, to capture the coupling between the cooling coil and the supply fan, the HMM of the supply fan have different state transition matrices corresponding to various states of the cooling coil.

B. AN ONLINE LEARNING ALGORITHM TO IDENTIFY FAILURE MODES OF KNOWN FAULT TYPES

Conditions of components and sensors change with operating environments, e.g., weather and occupants. To adapt to changing environments, HMM parameters need to be updated based on new observations. Existing Baum-Welch (BW)-based methods require accurate state recurrence equations which are derived based on assumptions that may not always be satisfied [19], [20]. Unlike them, the Gibbs-sampler-based method does not require these equations. Thus a Gibbs-sampler-based online learning algorithm is derived to update HMM parameters. In this method, observation sets are updated by adding new observations and removing oldest ones. Given prior distribution, Bayesian inference is used to deduce posterior distribution of HMM parameters based on updated observation sets and state estimates. HMM parameters are then drawn from their posterior distributions to replace their previous values. The method of updating HMM parameters is described below.

Update Observation Sets of HMMs. Assume that the current time is t+1, observation sets for different states are available as well as state estimates from the beginning to time t. To update observation sets at time t+1, the new observation O_{t+1} should be added in the corresponding observation set. Thus it is important to know the state to which O_{t+1} belongs. To estimate the state, the Viterbi algorithm is usually used since it gives the probability of the most likely sequence of states. However, it is difficult to use the algorithm in this case since HMM parameters at time t+1 are yet to be derived. Unlike using the Viterbi algorithm, using the forward variable $\alpha_{t+1}(i)$ provides us with a probability of the partial observation sequence until time t+1, with State i at time t+1,

$$\alpha_{t+1}(i) = [\sum_{j=1}^{N} \alpha_t(j) a_{ji,t}] b_i(O_{t+1}). \quad (8)$$

This variable depends on state transition probabilities $a_{ji,t}$ at time t and the observation likelihood $b_i(O_{t+1})$, and it gives the marginal probability for each state. Thus the state having the forward variable with the maximum value is considered as the rough state estimate at time t+1. To show the procedure in detail, updating HMM parameters of a cooling coil coupled with the supply fan at time t+1 is presented as an example and is shown in Fig. 5.

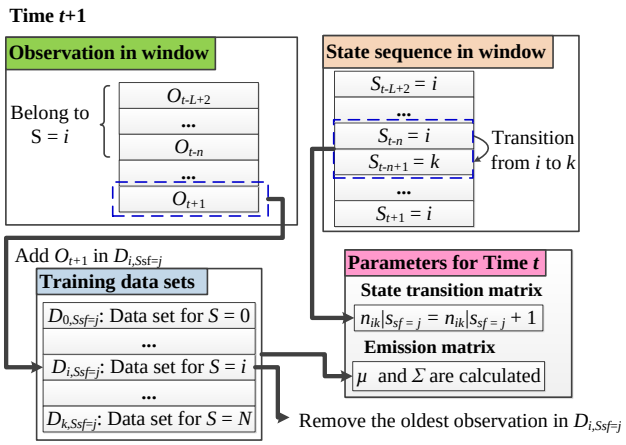

FIGURE 5. **The procedure of updating HMM parameters at time t+1.**

At time t+1, the supply fan is in State j. By calculating forward variables, the rough state estimate of the cooling coil is i. The observation O_{t+1} is added in the observation set $D_{i,s_{sf}=j}$ corresponding to State i of the cooling coil and State j of the supply fan, and oldest observations are removed.

Update Parameters of HMMs and Estimate States. After updating observation sets, Bayesian rule is used to infer posterior distributions of HMM parameters given their prior distributions. As discussed in [18], the prior probability of the mean of observations corresponding to $S_{cc} = i$ and $S_{sf} = j$ is normally distributed

$$\mu_i|_{s_{sf}=j} \ldots \sim N(\mu_{0,i}, \Sigma_{0,i}^{-1}). \quad (9)$$

Based on the Bayes rule, the posterior distribution is derived as

$$\mu_i|_{s_{sf}=j} O_{cc} \ldots \sim N(\bar{\mu}_i|_{s_{sf}=j}, \bar{\Sigma}_i|_{s_{sf}=j}), \quad (10)$$

with

$$\bar{\mu}_i|_{s_{sf}=j} = (n_i \bar{\Sigma}_i|_{s_{sf}=j} + \Sigma_{0i}^{-1})^{-1} (n_i|_{s_{sf}=j} \Sigma_i^{-1} \bar{o}_i|_{s_{sf}=j} + \Sigma_{0i}^{-1} \cdot \mu_{0,i}), \quad (11)$$

and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

$$\bar{o}_i\,|_{s_{sf}=j} = \left(\sum_{k=1}^{n_i} o_k\,|_{s_{sf}=j}\right) \Big/ n_i\,|_{s_{sf}=j}\,, \tag{12}$$

where $n_i\,|_{s_{sf}=j}$ is the number of observations corresponding to State i of the cooling coil and State j of the supply fan; and $\bar{o}_i\,|_{s_{sf}=j}$ is the average value of the observations. The prior density of the covariance matrix of observations corresponding to $S_{cc} = i$ and $S_{sf} = j$ are assumed to follow an Inverse Wishart (IW) distribution. The prior distribution is

$$\Sigma_i\,|_{s_{sf}=j} \sim IW(\Psi_i, \upsilon_i). \tag{13}$$

Then, the posterior distribution is derived as,

$$\Sigma_i\,|_{s_{sf}=j}\, O_{cc} \sim IW\Big(\Psi + \sum_{k=1}^{n_i} (o_k\,|_{s_{sf}=j} - \mu_i\,|_{s_{sf}=j})(o_k\,|_{s_{sf}=j} - \mu_i\,|_{s_{sf}=j})^T \,, \upsilon_i + n_i\,|_{s_{sf}=j}\Big). \tag{14}$$

As justified in [18], probabilities of state transitions follow a Dirichlet distribution. The prior distribution is

$$(a_{i,0},\ldots,a_{i,N_{cc}-1})\,|_{s_{sf}=j} \sim Dir(1,\ldots,1). \tag{15}$$

Given the number of visits to each state, the posterior distribution is also Dirichlet and is given by:

$$(a_{i,0},\ldots,a_{i,N_{cc}-1})\,|_{s_{sf}=j} \sim Dir(n_{i0}\,|_{s_{sf}=j}+1,\ldots,n_{i,N_{cc}-1}\,|_{s_{sf}=j}+1), \tag{16}$$

where $n_{i0}\,|_{s_{sf}=j}$ is the number of transitions from State i to State 0 (normal condition) when the supply fan is in State j. These values are obtained by counting visits to each state given state estimates before time t+1. Given updated parameters, the Viterbi algorithm is used to estimate states in the current moving window consisting of O_{t-L+2}, …, O_{t+1}, where L is the window length. The computational complexity of the Viterbi algorithm is $O(N^2T)$, where N is the number of states and T is the length of the state sequence. In our method, for each t, the state sequence in the moving window needs to be estimated, thus the complexity is $O(N^2TL)$. The state estimate at the end of the sequence is considered as the estimate at time t+1.

Evaluate State Estimates. To measure fault diagnosis accuracy, F-measure is used since it reflects the precision and recall (correct/false identification rates) [15]. To generate F-measures, four statistical measures related to identification rates are considered, including (a) true positive; (b) false positive; (c) true negative; and (d) false negative. True positive means that the normal condition is correctly estimated; true negative means that the failure mode is correctly estimated; false positive means that the normal condition is falsely estimated as a failure mode; and false negative means that the failure mode is falsely estimated as the normal condition or other failure modes. The larger the F-measure is, the better the performance of a diagnosis method is.

C. METHOD OF IDENTIFYING SEVERITIES OF KNOWN FAULT TYPES

For components, deviations of fault-related parameters from their normal values represent severities of fault impacts. For sensors, deviations of sensor readings from their actual values reflect fault severities. Filter-based methods including KF and PF are developed to estimate fault severities of components and sensors. For linear models, KF is used since it is an optimal linear filter. For the nonlinear case, PF is used since it is good at dealing with nonlinearities. Our methods are discussed below.

Identify Fault Severities of Components. As presented before, four faults in cooling coils are considered, including (a) tube fouling; (b) dust on fins; (c) valve stuck closed; and (d) valve stuck open. Tube fouling causes a decrease in the tube inside diameter $d_{tube,in}$. Similarly, dust accumulation leads to a decrease in the fin outside surface area A_{fin}. Severities of these two faults are reflected as the amount of decease in the two parameters. To estimate the parameters, they are considered as parametric states $x_{cc} = [d_{tube,in}\ A_{fin}]^T$. The state evolution is

$$x_{cc}(t+1) = x_{cc}(t) + v_{cc}(t), \tag{17}$$

where process noise is denoted by $v_{cc}(t) = [v_{tube}(t)\ v_{fin}(t)]^T$. For simplification, process noise is assumed white, zero mean and normally distributed, and $v_{tube}(t)$ and $v_{fin}(t)$ are uncorrelated. To represent the relationship between the two parametric states and sensor readings, the measurement equation is derived from the physics-based model (1):

$$z_{cc}(t) = h_{cc}(x_{cc}(t)) + w_{cc}(t), \text{ with} \tag{18}$$

$$\text{where } h_{cc}(x_{cc}(t)) = \left[\frac{1}{\alpha_{a,e}\tau} + R_f + \frac{\delta_{tube}}{\lambda_{tube}} + \frac{1}{\alpha_w(d_{tube,in})}\right]^{-1} (A_{tube,out} + A_{fin}), \tag{19}$$

and $w_{cc}(t)$ is the measurement noise and is normally distributed. To estimate states of this nonlinear model, a PF is used. PF uses a set of particles to represent the posterior distribution given noisy observations. For PF, particles can be input in nonlinear models for computation directly. There is no big difference between linear models and nonlinear models for PF, and thus it is good at dealing with nonlinearities. Unlike tube fouling and dust on fins, valve stuck closed occurs suddenly, and thus the chilled water mass flow rate becomes almost 0. The cooling coil cannot reduce the air temperature. In this case, the failure mode should be resolved immediately. Similarly, valve stuck open also occurs suddenly. The supply air temperature is too low to track its set-point, and this fault should also be resolved as soon as possible.

Identify Fault Severities of Sensors. If the output signal of a sensor differs from the correct value by a constant, the constant is called as the sensor bias. If the output signal slowly changes independent of the measured property, this is defined as the sensor drift. The two kinds of sensor faults are considered. To identify fault severities of sensors, estimates of sensor bias and drift are required. To represent the two kinds of faults, two models are developed separately. For instance, state equations capturing a drift in the supply air mass flow rate sensor are,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

$$\begin{cases} k_{M\sup}(t+1) = k_{M\sup}(t) + w_{k,M\sup}(t), \\ x_{M\sup}(t+1) = x_{M\sup}(t)\cdot[1+k_{M\sup}(t)] + w_{M\sup}(t), \end{cases} \quad (20)$$

where $k_{Msup}(t)$ is the drift rate of the sensor at time t; the term $x_{Msup}(t)$ represents the sensor drift at time t; For simplification, process noises $w_{k,Msup}(t)$ and $w_{Msup}(t)$ are assumed to be normally distributed, and uncorrelated. Based on the fan model in [28], the measurement equation is obtained as,

$$z_{sf}(t) = h_{sf}(x_{M\sup,drift}(t)) + v_{sf}(t), \text{ with} \quad (21)$$

$$z_{sf}(t) = (Q_{sf}\cdot\rho_{air})/(\dot{m}_{\sup,des}\cdot\Delta P_{sf}), \text{ and} \quad (22)$$

$$h_{sf}(x_{M\sup drift}(t)) = (c_1 + c_2 f_{f,sf} + c_3 f_{f,sf}^2 + c_4 f_{f,sf}^3 + c_5 f_{f,sf}^4)/e_{sf}(t), \text{ with} \quad (23)$$

$$f_{f,sf} = [\dot{m}_{a,\sup}(t) - x_{M\sup,drift}(t)]/\dot{m}_{\sup,des}. \quad (24)$$

The measurement noise $v_{sf}(t)$ is normally distributed. Similarly, for the bias, the state equation is developed as,

$$x_{M\sup bias}(t+1) = x_{M\sup bias}(t) + w_{M\sup bias}(t), \quad (25)$$

since the sensor bias is assumed to be a constant. The measurement equation is similar to that of drift, except replacing the drift term $x_{Msup,drift}$ by the bias term $x_{Msup,bias}$ in (21), (23) and (24). Given the inference of failure modes from the HMM, occurrence of a bias or a drift in sensors is known, and the appropriate model is selected. KF or PF are then used to estimate sensor bias/drift based on the models as fault severities. Unlike the supply air mass flow rate, other sensor readings, such as the return air temperature, are not contained in existing physics-based/gray-box models. To identify their fault severities, residuals between sensor readings and their estimates are considered. For instance, as shown in (6), the return air temperature can be estimated based on zone temperatures. The residual between the return air temperature and its estimate is used to represent the fault severity.

IV. IDENTIFICATION OF FAILURE MODES AND FAULT SEVERITIES FOR NEW FAULT TYPES

In this section, a statistical method is developed to identify new failure modes and their severities. In subsection IV-A, a robust Bayes-factor-based method is developed to find potential new fault types. In subsection IV-B, a KL-divergence-based method is developed to confirm potential types as true ones. In subsection IV-C, physical knowledge is used to find the cause of the new fault type and estimate its severities.

A. FINDING POTENTIAL NEW FAULT TYPES

To detect new types of faults, current observations are compared with the expected observations of all existing states. If current observations do not belong to all existing states, a potential new fault type is declared. To test whether two groups of observations are from different distributions, Bayes-factor-based testing is usually used. In the testing, the ratio of the probability of belonging to the same distribution to that of belonging to different distributions is calculated. If the ratio is smaller than one, which is a constant threshold, a new fault type is detected. Because of measurement noise and modeling errors, false declaration of a new fault type may result. To distinguish new fault types from noise and model errors, the constant threshold is replaced by control limits on the ratio. Consequently, the comparison is not sensitive to measurement noise and model errors. Additionally, in the testing, an analytical expression for the Bayes-factor in the multivariate case is derived.

In our Bayes-factor based testing, to test whether current observations and observations of an existing state i are from different distributions, there are two hypotheses which are H_0: $\mu_{new} = \mu_i$ versus H_1: $\mu_{new} \neq \mu_i$, where μ_i is the mean of the observation distribution corresponding to State i and μ_{new} is that of observations in the current moving window. The mean μ_i is unknown. To simplify the problem, μ_i is approximated by the average value of observations corresponding to State i. Thus the hypotheses are converted as $H_{0,i}: \mu_{new} = \bar{\mu}_i$ versus $H_{1,i}: \mu_{new} \neq \bar{\mu}_i$. Based on Bayes rule, the probability that current observations belong to State i is obtained

$$P(H_{0,i} \mid o_{new}) = \frac{P(o_{new} \mid H_{0,i})P(H_{0,i})}{P(o_{new})}. \quad (26)$$

Similarly, the probability that current observations do not belong to State i, $P(H_{1,i}|o_{new})$, is obtained. Since most of failure modes are considered and covered by our HMMs, occurrence of a new fault type is a small probability event. Prior probabilities for the two hypotheses are set as

$$\begin{cases} P(H_{0,i}) = 0.99 \\ P(H_{1,i}) = 0.01 \end{cases}. \quad (27)$$

Based on Bayes rule, the ratio of the two posterior probabilities is

$$\frac{P(H_{0,i} \mid o_{new})}{P(H_{1,i} \mid o_{new})} = \frac{P(o_{new} \mid H_{0,i})P(H_0)}{P(o_{new} \mid H_{1,i})P(H_1)} = \frac{m_0}{m_1}\cdot\frac{P(H_0)}{P(H_1)} = B_f \cdot \frac{P(H_0)}{P(H_1)}, \quad (28)$$

where $B_f = m_0/m_1$ is the Bayes factor. To detect new fault types, control limits of this ratio are calculated. If the ratio is lower than the lower bound of the control limits, it means that the probability of belonging to the existing State i is much lower than that of belonging to other states. Considering that the ratio may be too large or too small, log of the ratio is considered. By assuming that log of the ratio follows a normal distribution, the control limits are obtained based on training data as

$$[\bar{\mu}_{\log\_rto} - 2\bar{\sigma}_{\log\_rto}, \bar{\mu}_{\log\_rto} + 2\bar{\sigma}_{\log\_rto}], \quad (29)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

where $\bar{\mu}_{\log\_rto}$ is the average value of log ratios, and $\bar{\sigma}_{\log\_rto}$ is the standard deviation of log ratios. In (28), the posterior probability $m_{0,i}$ corresponding to State i is

$$\begin{aligned} m_{0,i} &= \int L(\Sigma)\pi_a(\Sigma)d\Sigma \\ &= \int |\Sigma|^{-n/2} \exp\left\{-\frac{n}{2}(\bar{X}-\bar{\mu}_i)'\Sigma^{-1}(\bar{X}-\bar{\mu}_i)-\frac{1}{2}tr\left(S\Sigma^{-1}\right)\right\}|\Sigma|^{-1}d\Sigma \\ &= \int |\Sigma|^{-1-n/2} \exp\left\{-\frac{1}{2}tr\left((U+S)\Sigma^{-1}\right)\right\}d\Sigma. \end{aligned} \quad (30)$$

where $U = n(\bar{X}-\bar{\mu}_i)'(\bar{X}-\bar{\mu}_i)$, $S = \sum_{j=1}^{L}(X_j-\bar{X})'(X_j-\bar{X})$; the parameter L is the moving window length; X_j is the jth observation in the current moving window; $\bar{X}$ is the average value of observations; and Σ is the covariance matrix of observations. The prior $\pi_a = |\Sigma|^{-1}$ was considered [29]. This is an integration with respect to a matrix Σ, and is defined as the iterated integral of a function with respect to each element of Σ [30]. Elements in Σ are denoted by σ_{jk}, j = 1, …, d and k = 1, …, d, where d is the observation dimension. Then, (30) is converted into

$$m_{0,i} = \int\cdots\int\left\{|\Sigma|^{-1-n/2}\exp\left[-\frac{1}{2}tr\left((U+S)\Sigma^{-1}\right)\right]\right\}\prod_{\substack{j=1,\ldots,d\\k=1,\ldots,d}} d\sigma_{jk}. \quad (31)$$

If the dimension of Σ is large, it is complex to calculate this integration. Since the probability density function of the inverse matrix gamma distribution is

$$\begin{aligned} &\int_0^\infty f(X,\alpha,\beta,\Psi)dX \\ &= \int_{-\infty}^{\infty}\frac{|\Psi|^{\alpha}}{\beta^{d\alpha}\Gamma_d(\alpha)}|X|^{-\alpha-(d+1)/2}\exp\left(tr\left(-\frac{1}{\beta}\Psi X^{-1}\right)\right)dX = 1, \end{aligned} \quad (32)$$

it follows,

$$\int_0^\infty |X|^{-\alpha-(d+1)/2}\exp\left(-\frac{1}{\beta}tr\left(\Psi X^{-1}\right)\right)dX = \frac{\beta^{d\alpha}\Gamma_d(\alpha)}{|\Psi|^{\alpha}}. \quad (33)$$

It is evident that (33) is like (30). Letting α = -(d+1)/2 + n/2 +1, β = 2 and $\Psi = U + S$, (30) is converted to

$$m_{0,i} = \frac{2^{d[-(d+1)/2+n/2+1]}\Gamma_d(-(d+1)/2+n/2+1)}{|U+S|^{-(d+1)/2+n/2+1}}, \quad (34)$$

which can be easily calculated. The multivariate gamma function is

$$\Gamma_d(a) = \pi^{d(d-1)/4}\prod_{j=1}^{d}\Gamma(a+(1-j)/2). \quad (35)$$

(34) is converted into

$$m_{0,i} = \frac{2^{d[-(d+1)/2+n/2+1]}\prod_{i=1}^{d}\Gamma(n/2-d/2+0.5+(1-i)/2)}{\pi^{-d(d-1)/4}|U+S|^{-(d+1)/2+n/2+1}}. \quad (36)$$

The analytical expression for $m_{0,i}$ is obtained. Compared with $m_{0,i}$, $m_{1,i}$ is difficult to calculate. In [29], a recursive formula was derived to calculate $m_{1,i}$ as

$$\begin{aligned} m_{1,i} &= \frac{\prod_{j=1}^{d}\left\{\Gamma((n-1)/2)\cdot 2^{(n-1)/2}\right\}}{2^d(2\pi)^{(n-d)d/2}\cdot n^{d/2}}\prod_{j=1}^{d-1}\left\{|S_j|^{-1/2}\right\} \\ &\cdot\frac{1}{s_{11}^{(n-1)/2}}\prod_{j=2}^{d}\left(\frac{|S_{j-1}|}{|S_j|}\right)^{(n-1)/2}. \end{aligned} \quad (37)$$

Thus, the Bayes factor BF is obtained.

B. CONFIRMING POTENTIAL NEW FAULT TYPES AS TRUE ONES

New fault types cause large changes in observations. To further reduce false identification rates, deviations of current observations from previous ones are also considered. The potential new fault type is confirmed as the true one if observations have significant changes. To identify changes in observations, statistical process controls are usually used. However, observations may have multiple dimensions. Statistical process controls are usually applied to the univariate case rather than the multivariate case. It is difficult to combine analysis for each dimension together. To address this issue, a KL-divergence-based method is developed since it measures how the distribution represented by all dimensions diverges from that represented by previous observations. In our method, potential new fault types are confirmed as true ones if KL-divergence falls outside its control limits determined based on a small false identification rate. To derive the control limits, distributions of KL divergences are required. As discussed in [31], the KL divergence between two univariate random variables follows a non-central chi-square distribution with one degree of freedom. However, the distribution of KL divergence for the multivariate case is not available, and is derived. The method is presented below.

Confirm Potential New Fault Types as True Ones based on KL Divergence. The KL divergence between observations o_{new} in the current moving window and observations o_{pre} in the previous moving window is denoted by $D_{KL}(o_{new}||o_{pre})$ and is calculated as

$$D_{KL}(o_{new}||o_{pre}) = \int_{-\infty}^{\infty} p_{new}(x)\log\frac{p_{new}(x)}{p_{pre}(x)}dx, \quad (38)$$

where $p_{new}(x)$ and $p_{pre}(x)$ are probability density functions of o_{new} and o_{pre}. By assuming that observations follow normal distributions, the KL divergence between current observations and previous observations is [31]

$$\begin{aligned} D_{KL}(o_{new}||o_{pre}) = \frac{1}{2}\Bigg[&\log\frac{|\Sigma_{pre}|}{|\Sigma_{new}|}+Tr[\Sigma_{pre}^{-1}\Sigma_{new}]-d \\ &+(\mu_{new}-\mu_{pre})^T\Sigma_{pre}^{-1}(\mu_{new}-\mu_{pre})\Bigg], \end{aligned} \quad (39)$$

where μ_{new} and μ_{pre} are means of distributions represented by o_{new} and o_{pre}, respectively; and Σ_{new} and Σ_{pre} are covariance matrices of o_{new} and o_{pre}, respectively. Since observations are extracted via PCA, they are uncorrelated. It is therefore reasonable to assume that Σ_{new} and Σ_{pre} are almost diagonal.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

Since Σ_{new} and Σ_{pre} are different, it can be found that $\Sigma_{new} = \phi\Sigma_{pre}$ where ϕ is a diagonal matrix. Then

$$D_{KL}(o_{new}||o_{pre}) = \frac{1}{2}\left[\log\frac{1}{|\phi|} + Tr[\phi] - d + (\mu_{new}-\mu_{pre})^T\Sigma_{pre}^{-1}(\mu_{new}-\mu_{pre})\right]$$
$$= \frac{1}{2}[(\mu_{new}-\mu_{pre})^T\Sigma_{pre}^{-1}(\mu_{new}-\mu_{pre})] + b, \quad (40)$$

where $b = \frac{1}{2}\left[\log\frac{1}{|\phi|} + Tr(\phi) - d\right]$.

Since Σ_{pre}^{-1} is diagonal, the first term of the KL divergence shown in (40) is

$$X = \frac{1}{2}(\mu_{new}-\mu_{pre})^T\Sigma_{pre}^{-1}(\mu_{new}-\mu_{pre}) = \frac{1}{2}\sum_{j=1}^{d}\frac{(\mu_{new,j}-\mu_{pre,j})^2}{\sigma_{pre,j}^2}$$
$$= \sum_{j=1}^{d}\frac{\mathrm{var}(D_j)}{2s_j}\frac{D_j^2}{\mathrm{var}(D_j)} = \sum_{j=1}^{d}\frac{1}{\omega_j}\chi_1^2(\lambda_j) = \sum_{j=1}^{d}c_j\chi_1^2(\lambda_j), \quad (41)$$

with

$$\Sigma_{pre}^{-1} = \begin{bmatrix}\sigma_{pre,1}^{-2} & & 0\\ & \cdots & \\ 0 & & \sigma_{pre,d}^{-2}\end{bmatrix}, \quad (42)$$

$$\mu_{new} = \begin{bmatrix}\mu_{new,1}\\ \vdots \\ \mu_{new,d}\end{bmatrix} \text{ and } \mu_{pre} = \begin{bmatrix}\mu_{pre,1}\\ \vdots \\ \mu_{pre,d}\end{bmatrix}, \quad (43)$$

where $D_j = \mu_{new,j} - \mu_{pre,j}$; λ_j is the non-centrality parameter of the non-central chi-square distribution $\chi(\lambda_j)$ corresponding to the jth dimension of observations. It can be found that X is a linear combination of non-central chi-square distributions as shown in (41). In [32], the cumulative density function of a linear combination of non-central chi-square distributions was developed. Given the false identification rate $P_f = 0.01$, the control limits are calculated by using the cumulative density function as:

$$P_f = P(D_{KL} > h \mid H_0) \quad (44)$$

where h is the upper limit of D_{KL} since D_{KL} is definitely positive. If D_{KL} is larger than h, it means that there is a large change in observations, and the potential new fault type is confirmed as a true one.

Add New States and Update HMM Parameters to Capture The Detected New Fault Type. To capture the detected new fault type, new states are added in the HMM, and HMM parameters are updated accordingly. For instance, a new fault type of the cooling coil is detected. The number of failure modes is increased from four to five. Thus, the number of states N_{cc} is increased from 2^4 to 2^5. Accordingly, each $2^4\times2^4$ state transition matrix is replaced by a $2^5\times2^5$ state transition matrix. New state transition matrices are generated by following (16). To generate new observation sets, it is reasonable to assume that observations in the current moving window belong to the detected new fault type. Considering the coupling between the supply fan and the cooling coil, observation sets corresponding to the new state of the cooling coil and existing states of the supply fan are generated. Means and covariance matrices of these observation sets are obtained based on (10) and (14) to represent emission matrices.

C. IDENTIFYING THE DETECTED NEW FAULT TYPE AND ITS SEVERITIES VIA PHYSICAL KNOWLEDGE

After detecting the new fault type, it is important to find the cause of the fault based on physical knowledge. For instance, pump leakage is considered as a new fault type. A fault tree of the cooling coil is established as shown in Fig. 6. In this figure, known types of faults are represented by solid circles and new fault types are represented by dashed circles. Both known types and new types of faults are reflected by certain fault-related sensor readings, such as the chilled water mass flow rate marked by a black dot.

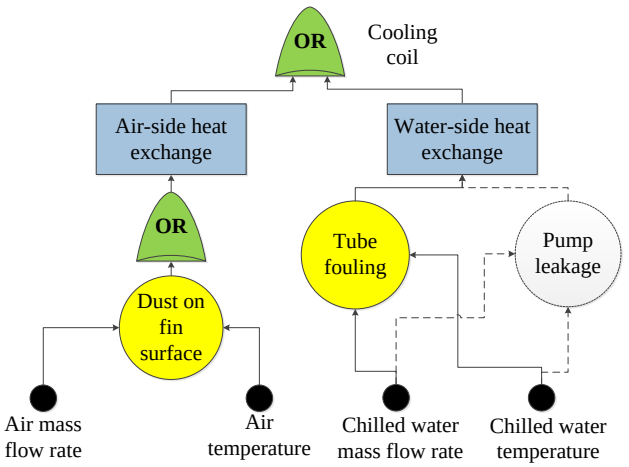

FIGURE 6. The fault tree of the cooling coil.

After detecting a new fault type, by checking sensor readings used by the HMM of the cooling coil, it can be found that the chilled water mass flow rate gradually decreases. Based on the fault tree, either tube fouling or pump leakage occurs. Considering that tube fouling is an existing fault, the detected fault should be pump leakage. Since pump leakage causes a decrease in the chilled water mass flow rate passing through the pump, the amount of decrease is used to identify the fault severity.

V. EXPERIMENTAL RESULTS

Our fault diagnosis method was implemented by using MATLAB 2014a and was run on a laptop with Intel Core i7-6920HQ 2.9GHz processor and 32GB of memory. The method is tested using simulation data and real data. In Example 1, a small building is simulated by using two packages: DesignBuilder [33] and EnergyPlus [26]. Results show that (a) known types of failure modes in components

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

and sensors are identified by coupled online HMMs with low false identification rates; (b) fault severities of existing faults are estimated accurately given inferences of failure modes; and (c) our statistical method of integrating Bayes-factor testing and KL divergence detects new fault types with low false identification rates. In Example 2, our method is tested using real data from the ASHRAE project 1312-RP [3]. Results illustrate that both known types and new types of failure modes are diagnosed with low false identification rates, and their severities are estimated accurately.

Example 1: The simple building has two 95.517 m^3 rooms. In the building, tube diameter $d_{tube,in}$ is set to be 0.01445 m; the outside surface area of fins is 43.59555 m^2; the fan efficiency is 0.7. For the cooling coil, tube fouling is simulated from 7/18 to 7/23; dust on fins is simulated from 7/24 to 7/29; and pipe leakage is simulated from 7/30 to 8/3. For the supply fan, a decrease in supply fan efficiency is simulated from 8/2 to 8/4. Drift of the return air temperature sensor is simulated from 9/23 to 9/25, and the bias is simulated on 9/30. Other sensor faults are also simulated. 2/3 of the data are used for training and the rest are used for testing.

Identify Known Types of Failure Modes of Components. States of the cooling coil are estimated by using the coupled online HMM. Actual states and state estimates are represented by black dashed lines and blue stars, respectively, as shown in Fig. 7. The x-axis is the time, and the y-axis is the state of the cooling coil. States corresponding to the normal condition, the tube fouling, the dust on fins and valve stuck closed are denoted by '0,' '8,' '4' and '2.' In the figure, most actual state points and their estimates are the same and are overlapped. Also, there are some false identifications. For instance, certain points belonging to the normal condition are falsely estimated as tube fouling on 8/17. Similarly, certain points belonging to the normal condition are falsely estimated as dust on fins around 10/4. *F*-measures of the tube fouling, dust on fins and valve stuck closed are 0.989, 0.928 and 0.994. The false alarm rate is 0.6%. The coupled online HMM captures coupling between the cooling coil and the supply fan and adapts to changing environments. States of the cooling coil are therefore estimated with low false identification rates.

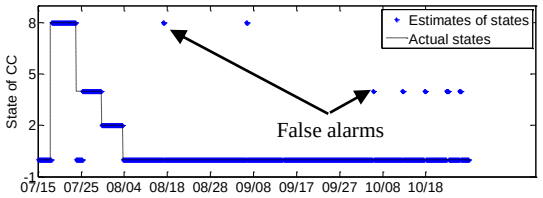

FIGURE 7. States of the cooling coil estimated by coupled online HMM.

If the coupled HMM with fixed parameters is used, the changing environments cannot be tracked. Thus, more false identifications occur when compared to the coupled online HMM as shown in Fig, 8. *F*-measures of the three failure modes are 0.981, 0.930 and 0.992. It can be concluded that most of the *F*-measures are worse than those of using the coupled online HMM. Additionally, the false alarm rate is 0.7%, which is slightly larger than that of using the coupled online HMM.

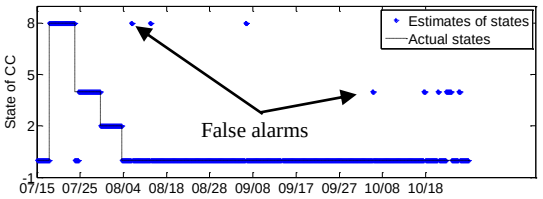

FIGURE 8. States of the cooling coil estimated by coupled HMM with fixed parameters.

If the online HMM is used, states are estimated as shown in Fig. 9. Since the online HMM does not capture coupling among components, and thus it performs worse than the coupled online HMM. *F*-measures of the three failure modes are 0.982, 0.926 and 0.938, and the false alarm rate is 0.9%, considerably higher.

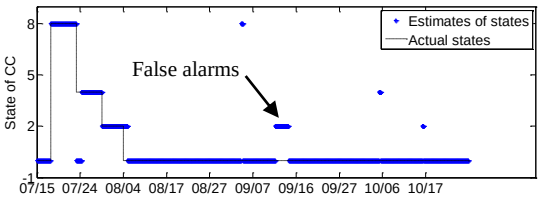

FIGURE 9. States of the cooling coil estimated by online HMM.

Estimate Fault Severities of Components. To illustrate the estimation of component fault severities, the supply fan is used as an example. Failure modes of the supply fan are estimated by using the coupled online HMM as shown in Fig. 10, where '0' means the normal condition, and '4' means a decrease in fan efficiency. Decrease in fan efficiency is diagnosed on 8/2. Given the state estimates, KF is used to estimate supply fan efficiency based on the fan model containing fan efficiency. The normal value of fan efficiency is 0.7. Residuals between the estimates of fan efficiency and its normal value represent severities as shown in Fig. 11. It can be found that residuals increase gradually with decrease in fan efficiency, and have three segments with different increased speed. This is because the fan efficiency cannot be decreased constantly due to limitations of the simulation packages used.

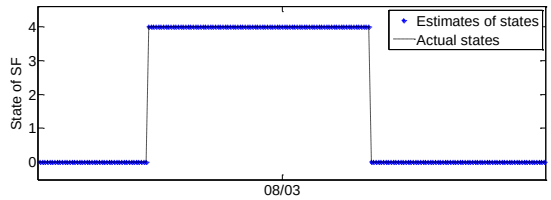

FIGURE 10. Estimates of failure modes of the supply fan.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

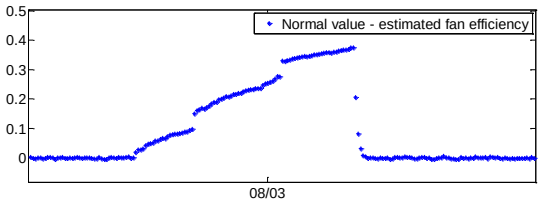

FIGURE 11. Residuals between estimates of the supply fan efficiency and its normal value.

Identify Known Types of Failure Modes of Sensors. To show the process of identifying failure modes of sensors, the return air temperature sensor is considered as an example. By using the coupled online HMM, states of the sensor are inferred as shown in Fig, 12, where '1' and '2' mean 'bias' and 'drift,' respectively. Their *F*-measures are 0.998 and 0.996.

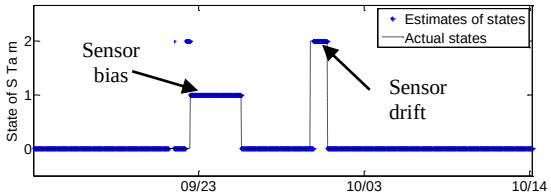

FIGURE 12. State estimates of the return air temperature sensor obtained by coupled online HMM.

Estimate Fault Severities of Sensors. Return air temperature can be estimated based on zone temperatures corresponding to all VAV boxes. Residuals between sensor readings and the estimates are calculated and shown in Fig. 13. It can be found that residuals in Fig. 13 correspond to state estimates in Fig. 12. The residual has a sudden change due to the sensor bias and a gradually increase caused by the sensor drift, and represents fault severities.

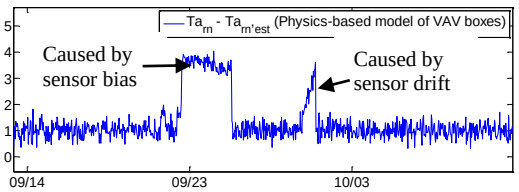

FIGURE 13. Residuals between sensor readings and actual values of the return air temperature sensor.

Detect and Identify New Types of Faults. To test our method, pump leakage is considered as a new fault type. It occurs on 8/9 and is detected by our statistical method after 86 hours. As discussed before, compared to known fault types, new fault types have limited corresponding observations compared to known ones. Their observation information may not be enough for distinguishing. Additionally, our method is conservative to achieve a low false alarm rate. Thus more time is required to detect the new fault types compared with that of known fault types. The detected new fault type is then diagnosed by the coupled online HMM as shown in Fig. 14. A low false alarm rate 0.5% is achieved.

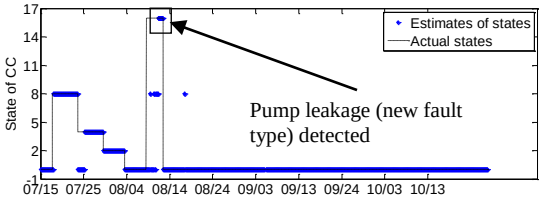

FIGURE 14. Pump leakage (new fault type) detected by our statistical method based on Coupled online HMM.

By observing sensor readings of the cooling coil, it can be seen that the chilled water mass flow rate is reduced. Based on physical knowledge, it is known that both tube fouling and pump leakage can cause a decrease in the chilled water flow rate. Tube fouling is an existing fault, thus the new fault type must be the pump leakage.

Example 2: In ASHRAE project 1312-RP, there are two AHUs, AHU-A and AHU-B, which were calibrated to be identical [3]. AHU-B is fault-free, and multiple faults were implemented in AHU-A during spring, summer and winter. This paper focuses on the cooling mode, thus summer data from 8/19 to 9/8 are used. Detailed description of data can be found in [10]. In the data, faults of the EA damper, the OA damper, ducts, the return fan were implemented, but sensor faults are not implemented.

Identify Known Types of Failure Modes in Components. To show the process of identifying the failure modes of components, the EA damper and the OA damper are used as examples. By using the coupled online HMMs, failure modes of the EA damper are estimated as shown in Fig. 15. In the figure, '1' means that the damper is stuck closed. Its *F*-measure is 0.966. Similarly, states of the OA damper are estimated as shown in Fig. 16. '1' denotes the damper leakage, and '2' means that the damper is stuck closed. *F*-measures of the two failure modes are 1 and 1, respectively.

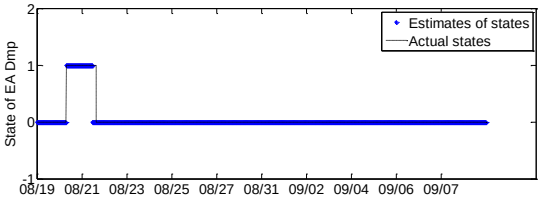

FIGURE 15. State estimates of the EA damper.

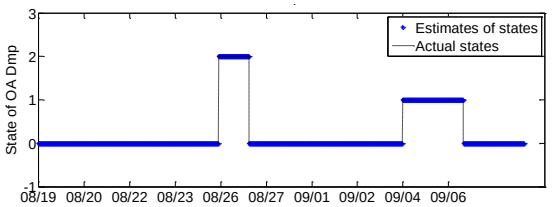

FIGURE 16. State estimates of the OA damper.

Estimate Severities of Failure Modes. OA damper leakage is used as an example to show the process of estimating fault severities. As mentioned in [3], the normal value of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2806373, IEEE Access

damper opening is 40%. To realize the damper leakage, the damper opening is changed to 45% on 9/5 and 55% on 9/6. Residuals between the damper opening and its normal value are calculated to represent the fault severities as shown in Fig. 17.

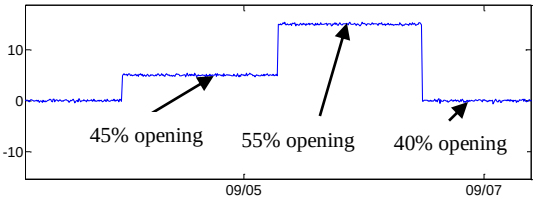

FIGURE 17. Residuals between the OA damper opening and its normal value (40% opening is normal).

Detect and Identify New Types of Faults. To test our method, stuck open of the EA damper is considered as a new fault type. The new fault type is detected by using our statistical method, and new states are estimated by our coupled online HMM as shown in Fig. 18. Since the new fault type is detected by the HMM of the EA damper, the new fault type should be related to the airflow passing through the EA damper. By observing the airflow, it can be found that the airflow rate is always maximum. Thus the fault is identified as damper stuck open.

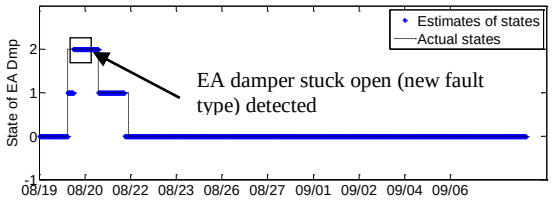

FIGURE 18. EA damper stuck open (new fault type) detected by our statistical method.

Comparison between our Method and Others. Many methods such as SVM, decision tree, Bayesian network and ANN have been developed to diagnose faults in air handling systems. To evaluate our method, *F*-measures of known fault types are compared with those obtained by using other methods also based on the ASHRAE project 1312-RP data as shown in Table. 1.

TABLE I
F-MEASURES OF FAILURE MODE OBTAINED BY OUR METHOD AND OTHERS

	Our method	SVM [15]	Decision tree [9]
$f_{ea\_dmp,sc}$	0.966	0.923	0.9
$f_{oa\_dmp,sc}$	1	NA	NA
$f_{oa\_dmp,leak}$	1	0.994	NA
f_{cc,vlv_sc}	1	0.928	1
f_{cc,vlv_so}	1	0.785	0.98
$f_{duct,lb}$	0.998	0.981	1
$f_{duct,la}$	0.821	1	NA
$f_{rf,cf}$	0.999	0.887	1
$f_{rf,fs}$	1	0.795	1
Average	0.976	0.923	0.97

In this table, it can be seen that the average *F*-measure of our method is better than that of the SVM [15] and the decision tree [9]. Additionally, as mentioned in [15], the average *F*-measure of their method is 0.923 that is significantly better than other methods, e.g., LibSVM, Naïve Bayes, radial basis function network, Bayesian network, NN and random forest decision tree. Therefore, the performance of our method is also better than these methods.

VI. CONCLUSION

In this paper, a systematic method is developed to identify known and new types of failure modes and their severities in air handling systems with low false identification rates. In this method, to identify known types of faults, an online learning algorithm is developed to estimate the states of components and sensors, while updating HMM parameters to adapt to changing environments. To identify new types of faults with low false identification rates, a robust statistical method is developed to detect the new fault type, and the new fault type is labeled based on physical knowledge (e.g., a fault tree or a human-in-the-loop). By adapting to changing environments and capturing coupling among components, our method performs better than others.

REFERENCES

[1] U.S. Department of Energy, *Building Energy Data Book*. Available: http://buildingsdatabook.eren.doe.gov, 2009.

[2] Building optimization and fault diagnosis source book, IEA ANNEX 25, 1996.

[3] S. Li and J. Wen, Description of fault test in Summer of 2007, in: ASHRAE 1312 Report, 2007.

[4] S. W. Wang, F. Xiao, "AHU sensor fault diagnosis using principal-component analysis method," *Energy Buildings*, Vol. 36, No. 2, 2004, pp. 147–60.

[5] S. W. Wang and J. B. Wang, "Robust sensor fault diagnosis and validation in HVAC systems," Transactions of the Institute of Measurement and Control, Vol. 24, No. 3, 2002, pp. 231-262.

[6] S. Katipamula, R.G. Pratt, D.P. Chassin, Z.T. Taylor, K. Gowri, and M.R. Brambley, "Automated fault detection and diagnostics for outdoor-air ventilation systems and economizers: Methodology and results from field testing," *ASHRAE Transactions*, Vol. 105, No. 1, 1999, pp. 1-13.

[7] J. M. House, Vaezi-Nejad and J.M. Whitcomb, "An expert rule set for fault detection in air handling units," *ASHRAE Transactions*, Vol. 107, No. 1, 2001, pp. 858-871.

[8] J. Schein, S. T. Bushby, N. S. Castro and J. M. House, "A rule-based fault detection method for air handling units," *Energy and Buildings*, Vol. 38, No. 12, pp. 1485-1492, 2006.

[9] R. Yan, Z. J. Ma, T. Zhao and G. Kokogiannakis, "A decision tree based data-driven diagnostic strategy for air handling units," DOI: http://dx.doi.org/doi:10.1016/j.enbuild.2016.09.039.

[10] S. Li and J. Wen, "A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform", *Energy and Buildings*, Vol. 68, 2014, pp. 63-71.

[11] W. Y. Lee, J. M. House, C. Park and G. E. Kelly, "Fault diagnosis of an air-handling unit using artificial neural networks," *ASHRAE Transactions*, Vol. 102, No. 1, 1996, pp. 540-549.

[12] Z. Du, B. Fan, X. Jin and J. Chi, "Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and

subtractive clustering analysis," *Building and Environment*, Vol. 73, 2014, pp. 1-11.

[13] S. B. Sun, G. N. Li, H. X. Chen, Q. Y. Huang, S. B. Shi and W. J Hu, "A hybrid ICA-BPNN-based FDD strategy for refrigerant charge faults in variable refrigerant flow system," Applied Thermal Engineering, DOI: http://dx.doi.org/10.1016/j.applthermaleng.2017.08.047, 2017

[14] K. Yan, Z. W. Ji, H. J. Lu, J. Huang, W. Shen and Y. Xue, "Fast and Accurate Classification of Time Series Data Using Extended ELM: Application in Fault Diagnosis of Air Handling Units," IEEE Transactions on Systems, Man, And Cybernetics: Systems, 2017.

[15] T. Mulumba, A. Afshari, K. Yan, W. Shen and L.K. Norford, "Robust model-based fault diagnosis for air handling units," *Energy and Buildings*, Vol. 86, 2015, pp. 698-707.

[16] H. M. Ertunc, K. A. Loparo and H. Ocak, "Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs)," *International Journal of Machine Tools and Manufacture*, Vol. 41, 2001, pp. 1363–1384.

[17] A. Kodali and K. Pattipati, "Coupled factorial hidden Markov models (CFHMM) for diagnosing multiple and coupled faults," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 43, No. 3, 2013, pp. 522-534.

[18] Y. Yan, P. B. Luh and K. R. Pattipati, "Fault diagnosis of HVAC air handling systems considering fault propagation impacts among components," *IEEE Transactions on Automation Science and Engineering*, Vol. 14, No. 2, 2017, pp. 705-717.

[19] G. Mongillo and S. Deneve, "Online learning with hidden Markov models," Neural Computation, Vol. 20, 2008, pp. 1706-1716.

[20] T. Chis and P. G. Harrison, "Adapting hidden Markov models for online learning," *Electronic Notes in Theoretical Computer Science*, Vol. 318, 2015, pp. 109-127.

[21] H, Yoshida, T. Iwami, H. Yuzawa and M. Suzuki, "Typical faults of air-conditioning systems, and fault detection by ARX Model and extended Kalman filter," *ASHRAE Transactions*, Vol. 102, No. 1, 1996, pp. 557-564.

[22] S. Goodman, "Toward evidence-based medical statistics. 1: The P value fallacy," *Annals of Internal Medicine*, Vol. 130, No. 12, 1999, pp. 995–1004.

[23] H. C. Yan, J. H. Zhou and C. K. Pang, "New types of faults detection and diagnosis using a mixed soft & hard clustering framework," 2016.

[24] M. J. Beal, Z. Ghahramani, Z and C. E. Rasmussen, "The infinite hidden Markov model," In Proceedings of Neural Information Processing Systems 14, MIT Press, Cambridge, MA, 2002.

[25] Q. S. Yan, W. X. Shi and C. Q. Tian, Refrigeration Technology for Air Conditioning, China, Construction Industry Press, Beijing, 2010, in Chinese.

[26] EnergyPlus Engineering Reference, http://apps1.eere.energy.gov/buildings/energyplus/pdfs/engineeringreference.pdf.

[27] F. P. Incropera and D. P. De Witt. Fundamentals of Heat and Mass Transfer. John Wiley and Sons, New York, 4th edition, 1996.

[28] Y. Yan, P. B. Luh, K.R. Pattipati, "Fault Diagnosis Framework for Air Handling Units based on the Integration of Dependency Matrices and PCA," in Proceedings of 2014 IEEE Conference on Automation Science and Engineering, Taipei, Taiwan, August 2014.

[29] D. Sun, "Objective priors for the multivariate normal model," in Proceedings of Valencia/ISBA 8th World Meeting on Bayesian Statistics Benidorm, Alicante, Spain, June, 2006.

[30] A. K. Gupta and D. K. Nagar, "Matrix variate distributions," in *Monographs and Surveys in pure and applied mathematics*, Boca Raton, FL: Chapman and Hall/CRC, 2000.

[31] D. I. Belov and R. D. Armstrong, "Distributions of Kullback-Leibler divergence and its application for the LSAT," Law School Admission Council, Research Report 09-02, October 2009.

[32] S. J. Press, "Linear combinations of non-central chi-square variates," The Annual of Mathematical Statistics, Vol. 37, No. 2, 1966, pp. 480-487.

[33] DesignBuilder 2.1 User Manual, http://www.designbuildersoftware.com/docs/designbuilder/DesignBuilder_2.1_Users-Manual_Ltr.pdf.