Development and Field Evaluation of Data-driven Whole Building Fault Detection and Diagnosis Strategy

Yimin Chen[1], Jin Wen[2]

[1,2] *Department of Civil, Architecture and Environmental Engineering, Drexel University, Philadelphia, PA, 19104, USA*

yc596@drexel.edu
jw325@drexel.edu

[1] *School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 10044, China*

ABSTRACT

Faults, i.e., malfunctioned sensors, components, control, and systems, in a building have significantly adverse impacts on the building's energy consumption and indoor environment. To date, extensive research has been conducted on the development of component level fault detection and diagnosis (FDD) for building systems, especially the Heating, Ventilating, and Air Conditioning (HVAC) system. However, for faults that have multi-system impacts, component level FDD tools may encounter high false alarm rate due to the fact that HVAC subsystems are often tightly coupled together. Hence, the detection and diagnosis of whole building faults is the focus of this study. Here, a whole building fault refers to a fault that occurs in one subsystem but triggers abnormalities in other subsystems and have significant adverse whole building energy impact. The wide adoption of building automation systems (BAS) and the development of machine learning techniques make it possible and cost-efficient to detect and diagnose whole building faults using data-driven methods. In this study, a whole building FDD strategy which adopts weather and schedule information based pattern matching (WPM) method and feature based Principal Component Analysis (FPCA) for fault detection, as well as Bayesian Networks (BNs) based method for fault diagnosis is developed. Fault tests are implemented in a real campus building. The collected data are used to evaluate the performance of the proposed whole building FDD strategies.

1. INTRODUCTION

Buildings are responsible for almost 40% of primary energy consumed in the U.S., and 51% of the primary energy usage

in commercial buildings are consumed by heating, ventilation and air conditioning (HVAC) system (EIA, 2012). Malfunctioning sensors, components, and control systems, as well as degrading HVAC and lighting components are main reasons for energy waste and unsatisfactory indoor environment. Extensive research has been made on the development of component level (e.g. chiller, air handling unit (AHU) and variable air volume (VAV) terminal units) FDD tools. However, in building HVAC systems, some whole faults occur in one component but cause abnormality in other closed subsystems. For example, when the outdoor air damper of an AHU is stuck at a position that is much larger than normal in a cooling season, this fault could not only cause the AHU's cooling coil valve to be opened at higher than normal position, but also could cause abnormalities in the primary cooling system, such as causing the chilled water pump to run at a higher speed to provide extra cooling needed. A false alarm may be triggered by the primary cooling system FDD tool. In such cases, component level FDD tools may produce a higher false alarm rate which may hinder its wide adoption in the real practice (Katipamula & Brambley, 2005).

In recent years, some system level FDD methods have been developed to detect and diagnose whole building operation abnormalities. These method can be divided into energy model based methods and data-driven based methods. Bynum et al., (Bynum, Claridge, & Curtin, 2012) describes an Automated Building Commissioning Analysis Tool (ABCAT), which is a simulation tool combing expert system to detect the energy abnormality. A building energy model is firstly developed by using first principles. Actual energy consumption data collected from BAS combining with the weather data are used to calibrate the simulation model. Forecasted energy performances from the calibrated simulation model are then compared with the measured data to detect faults. O'Neill et al., (O'Neill, Pang, Shashanka, Haves, & Bailey, 2014) describes another energy model-

based FDD method. In their approach, energy consumption data from different subsystems including HVAC, lighting and plug equipment usage are compared with a reference EnergyPlus model which is used to calculate the annual energy consumption to determine if the system has operation faults. The advantage of model-based FDD methods is that they use detailed building energy simulation which can provide a relatively accurate energy baseline (if the simulation models are well calibrated) to analyze the building's energy performance. If the developed model accurately represents subsystems and whole building level operation, then fault diagnosis is not too difficult because both subsystem level and whole building level comparison (model vs. measurements) can be achieved to isolate faults. However, the main disadvantage of this approach is that in practice, developing such detailed and accurate building energy simulation model is very time-consuming and calibrating such models to achieve a subsystem level accuracy for fault isolation remains to be very challenging. A top-down strategy by incorporating temporal and spatial partition was developed to detect HVAC faults across different levels (Wu & Sun, 2011). In this study, building energy consumption flow feature was firstly extracted to system faults. Then, energy consumption data was partitioned and grouped according system temporal and spatial characteristics to locate fault source.

Another type of methods is data-driven based methods. Data-driven based methods use system operation data and statistical analysis to find the system operation abnormality. For example, a statistical process control (SPC) and Kalman filter-based method was proposed for the system-level fault detection in HVAC systems (Sun, Luh, Jia, O'Neill, & Song, 2014). But the method was only evaluated through software simulation platform. Moreover, efficient whole building fault diagnosis method was not proposed to locate the fault root-cause in this research.

In this study, the aim of the whole building fault diagnostic strategy is to detect and diagnose the root-cause fault that happens in one subsystem but has impacts on different subsystems. This research proposes a hybrid strategy by integrating the developed data driven based method named as WPM and FPCA methods for fault detection (Chen & Wen, 2017) and newly proposed expert and data-driven based methods as BNs for fault diagnosis together. Hence, a complete framework for whole building FDD can be achieved. Building automation system operation data from a campus building during the cooling season are collected to evaluate the effectiveness of the proposed strategy.

2. FAULT DETECTION AND DIAGNOSIS STRATEGY

2.1. Overall Description of the Strategy

The overall fault detection and diagnosis strategy includes three procedures. Firstly, WPM is employed to generate baseline data from fault free dataset. Secondly, system feature based PCA is used to detect whole building level fault. In this step, a feature selection process is performed by using Partial Least Square Regression and Genetic Algorithm (PLSR-GA) method to handle the issue of high dimensionality of a whole building's dataset. Selected features are then used in a PCA modeling and fault detection process. Thirdly, when whole building fault is detected, BNs is used to diagnose and identify the fault root-cause. Figure 1 illustrate the structure of the proposed FDD strategy.
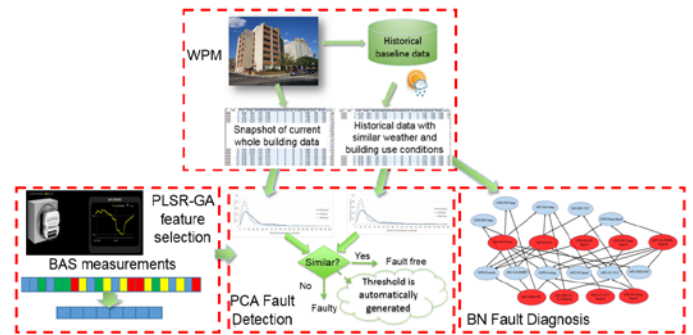

Figure 1 Structure of the proposed FDD strategy

2.2. Development of WPM Method

The operating condition of a HVAC system in buildings is a time-varying process, and is highly related to the change of external condition (e.g. weather) and internal condition (e.g. zone load) (Sun et al., 2014). A global data-driven model which reflects a long process operation may not function well when the system operation is changed under different conditions. Therefore, it is necessary to develop a local model for a small time period which may represents system behavior more accurately (Cheng & Chiu, 2004).

In order to develop a local model, the first step is to search the relevant data samples in the database. Here, a WPM method is developed to search and identify the "similar data" (i.e., data that have similar weather condition and time frame as the system snapshot data) from the fault free dataset. Symbolic Aggregate ApproXimation (SAX) (Keogh, Lin, & Fu, 2005) time series method is used to search similar weather time series data in historical fault free dataset. Firstly, a 30-minute snapshot window size in fault test dataset is decided after balancing the weather conditions change and computation efficiency. If a snapshot window size is too large, weather condition changes largely within a snapshot window and lead to a lower accuracy of Pattern Matching (PM). If the window size is too small, the computation burden will be increased as an increased number of snapshot window is generated. Each historical fault free day is divided with the same window size. Secondly, weather information (e.g. outdoor enthalpy) from test dataset and historical fault free dataset is combined to generate one time series dataset $X(t)$ with length n. Then dataset $X(t)$ can be represented in N space by a vector $\bar{X} = \bar{x}_1, \cdots, \bar{x}_N$. Here, $\bar{x}_i$ represents the ith element in the vector $\bar{X}$ and can be calculated by equation 1.

$$\bar{x}_i = \frac{N}{n} \sum_{j=\frac{n}{N(i-1)}+1}^{\left(\frac{n}{N}\right)i} x_j \qquad (1)$$

By this means, the time series data can be divided into N equal size data, and time series data from n dimensions can be reduced to N dimensions.

Thirdly, different symbolic strings are assigned to the X(t) according to the normalized data as shown in Figure 2. In this study, a data search pool size S is defined to determine how many samples will be used to generate a baseline dataset. Although a smaller window size can accurately capture the weather change, less information is obtained to generate a valid baseline database if a small window size is used to partition the data. Therefore, an adjacent dataset to the snapshot window can be used as the sample pool because the building occupancy may remain steady in a relatively long period. Lastly, data windows in historical fault free dataset which has the same symbolic string with the snapshot window are combined together to generate the baseline dataset.
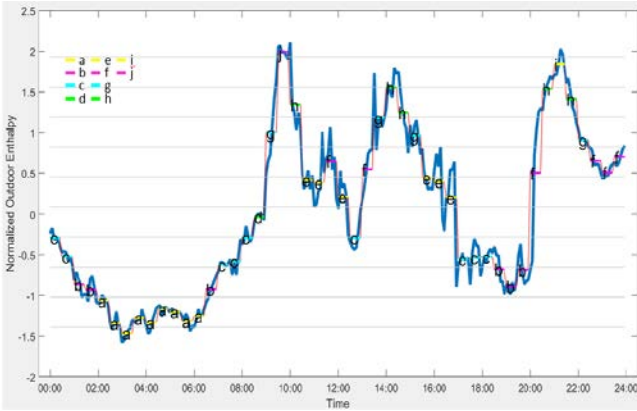

Figure 2 Illustration of SAX in PM

2.3. Feature based PCA for Fault Detection

2.3.1. Description of Feature Selection

In this research, a PLSR-GA based method is used to pre-select the candidate key variables which represent the system performance. This method employs PLSR to develop system performance model. Building energy consumption from fault free days is used as the target output variable. System operation variables are used as predictors. GA is used to search whole variable set and generate candidate variable subset for the PLSR. When the system performance standard is met, the variable subset is determined. By this process, the key informative variables can be selected and the variable dimensionality can be reduced. Detailed information about WPM method can be found in (Chen & Wen, 2017). Feature selection is employed by using off-line historical fault free

data. Whole building electrical energy consumption is used to develop the PLSR model. When implementing the GA method, the size of population is set to be 64, maximum generations are set to be 100, mutation rate is set to be 0.005. PLS toolbox 4.0 is used to implement the PLSR-GA method (Wise et al., 2007). Every five-day data from historical fault-free database were grouped together as one test case to implement one process of feature selection. 12 test cases from 2017 summer data were performed to determine which feature was selected to building PCA model and for fault detection. A total of 536 data measurements are obtained from the BAS in the selected building, the final selected variables are 73% of the whole building data measurements from operation data in the cooling season. The selected variables will be used in PCA fault detection.

2.3.2. Description of Feature Selection

PCA is one of the most popular data driven based methods employed in various industries to monitor system operation. PCA uses orthogonal transformation procedure to extract a set of linearly uncorrelated principal components (PCs) from the possibly correlated original variables (Lin, Keogh, Lonardi, & Chiu, 2003). In PCA, a data measurements matrix $X \in R^{n \times m}$ includes n samples and m process variables. X can be decomposed into a principal matrix $\widehat{X}$ plus a residual matrix E. The principal matrix $\widehat{X}$ is a product of a scores matrix T and a loadings matrix P, as given as below.

$$X = TP^T + E = \sum_{i=1}^{a} t_i p_i^T + E \qquad (2)$$

where t_i is a score vector (orthogonal) which contains information about relationship between samples and p_i is a loading vector (orthonormal) which contains information about relationship between variables. a is the number of PCs which are selected to retain in the model. Although, there is no best strategy to determine how many PCs are retained in the model, there are some other rules of thumb and user's knowledge used to determine the number of PCs (Wise et al., 2007). In this study, we used cumulative variance percentage contribution method. In this method, the smallest number of PCs is selected to capture a certain percentage of the cumulative variance which is pre-determined by the users. This number of PCs contributes the maximum cumulative variance (Zhu & Ghodsi, 2006).

$$\frac{d_1 + d_2 + \cdots + d_q}{d_1 + d_2 + \cdots + d_p} \geq \theta \qquad (3)$$

where θ is the pre-determined percentage, d is variance captured by each PC, p is the total number of PC, q is the retained number of PC. In this study, θ is set to 0.95, e.g. 95% cumulative variance will be remained in the PCA model.

In the study, T^2 statistic is used to obtain the detection result and can be calculated for each new observation by:

$$T^2 = x^T P \sum a\, P^T x \qquad (4)$$

where $\sum a$ contains the non-negative eigenvalues corresponding to the a principal components, x is a newly obtained process data vector.

The upper confidence threshold of T^2 can be calculated by using F-distribution:

$$T^2_{threshold} = \frac{a(n-1)}{n-a} F_{a,n-a,\alpha} \qquad (5)$$

where n is the number of samples in the data, α is the level of significance. In this study, α is set to 0.01.

In this study, a PCA model is not generated using the entire building measurements. Key variables selected from the feature selection process are employed for the PCA modeling to reduce computation burden, increase efficiency, and increase modeling accuracy.

2.4. Bayesian Networks (BNs) for Fault Diagnosis

After whole building fault is detected, it is critical to locate and isolate fault root-cause so building operator or maintenance personnel can correct those faults. Here in this study, a BNs based method is used to diagnose and isolate the fault root-cause as illustrated in this section.

2.4.1. Development of the BN Structure

BNs are a powerful tool to represent the knowledge and the inference under uncertainties. A probabilistic model which reveals the causal relations between faults and symptoms can be developed through BNs. The probabilities of relations in BNs can be updated after new observations (evidence) on the system are obtained (Lampis & Andrews, 2009). The process of developing BN structure is to generate a cause-and-effect inference and in this study expert knowledge is used to develop the system rules which will be mapped to the BNs. A two-layer of network which includes fault layer and fault evidence layer are employed to develop BNs for whole building fault diagnosis.

Network nodes, including fault nodes and fault evidence node need to be identified for whole building faults. Here, a fault node represents a whole building level fault. For example, an AHU outdoor air damper stuck fault (either stuck at a higher than normal or lower than normal) can be a fault node. The states of the fault node include faulty state and fault-free state. In this research, each fault node represents a specific fault type. For example, an AHU outdoor air damper stuck at a higher than normal position is assigned to one fault node. While, an AHU outdoor air damper stuck at a lower than normal position is assigned to another fault node.

Evidence nodes represent observable fault symptoms. Fault symptoms typically come from two sources:

1) Concurrent relationships among measurements. For example, when AHU outdoor air damper is stuck at a 100% open position, the mixed air temperature measurement has the same value as the outdoor air temperature measurement.

2) Historical relationship between the current value of a measurement and its historical baseline value. Using the same outdoor air damper stuck at a 100% open position fault example, an additional evidence (during a cooling season) could be that the value of the cooling coil valve position is higher than its historical baseline value under similar weather conditions. In this study, a pattern matching method, i.e., the WPM method is used to generate the baseline data.

Here, GeNIe and jSMILE (Drużdżel) BNs tools developed by Pittsburg University are used to generate the network. Six whole building fault root causes are considered and these root causes are set as fault nodes and ten evidence nodes are included. Figure 3 demonstrates the developed BN structure model for the whole building faults considered during cooling operation mode.
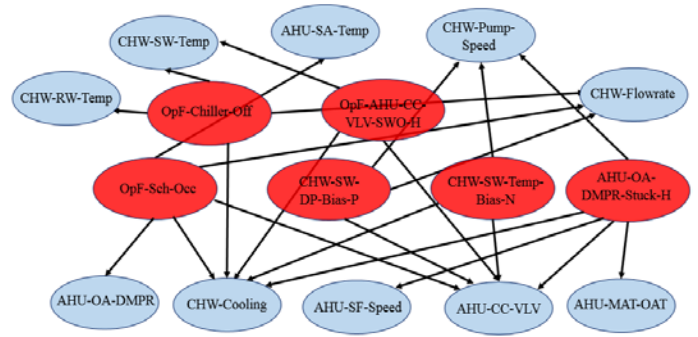

Figure 3 BN for nine whole building faults in the cooling operation mode

2.4.2. Determination of the BN Parameters

The second step in the development of BNs based fault diagnosis method is to identify values for the parameters (probabilities) in a BNs. The parameters of a BN reflect the quantitative relations among parent nodes and child nodes by using probability. Usually, three probabilities, i.e., *prior probabilities, conditional probabilities, and LEAK probabilities,* need to be determined when developing a BN model.

In this study, the values that have been reported for the component level fault diagnosis (Regnier & Wen, 2016) are adopted for prior probability. The prior probabilities should be updated when more system operation knowledge is obtained or statistic results can be found from the historical operation data.

For conditional probability, due to the limitation of testing faults in a real building, obtaining condition probability from fault data is unrealistic. Obtaining accurate values for

condition probabilities from expert knowledge is very difficult as well due to the fact that the same fault could behave slightly different in different buildings. However, expert knowledge could provide a range of condition probability for a fault and its associated fault evidences. For example, when AHU outdoor air damper is stuck at a 100% open position, the fault symptoms including 1) mixed air temperature measurement has the same value as the outdoor air temperature measurement; and 2) AHU cooling coil valve has a position that is higher than normal (baseline) position). Both of these two symptoms are strong symptoms, i.e., they would occur whenever the fault occurs. In this study, a fault evidence is firstly judged by whether it is a strong evidence or not. Three association levels, namely, strong evidence, medium evidence, and weak evidence, are used. If a fault evidence is a strong evidence, i.e., when a fault occurs, this evidence will most likely to occur, we consider that the condition probability of this evidence when a fault occurs is 90%, out of which, 45% is considered to have very sever fault symptom. In the example above, fault evidence node (difference between mixed air and outdoor air temperatures) is a strong evidence and a 0.45 conditional probability is assigned to this node for a very severe fault symptom (very abnormal), a 0.45 conditional probability is assigned to this node for a sever fault symptom, and a 0.1 conditional probability is assigned to this node as a low severe fault symptom. Similar treatment is used for the other two association levels, i.e., medium association nodes and weak association nodes.

LEAK probability represents the probability of an evidence node to be abnormal when all of the parent fault nodes are absent (when no fault occurs). *LEAK* probabilities for each evidence node are obtained by considering the outliers in the baseline data. After comparing with the baselined data, outliers in system operation data can be categorized into three trends as positive, negative and normal. Accordingly, an outlier is defined as:

$$|x(i) - \bar{x}| > t \cdot \sigma \qquad (6)$$

where $\bar{x}$ is the mean of the data sequence, σ is the standard deviation and t is the threshold.

In this study, two classes of threshold, i.e., 2σ is set for "very high/very low" and 1σ is set for "high/low" are used. Therefore, *LEAK* probability can be calculated as:

$$LEAK\ Probability = \frac{Number\ of\ outlier\ data\ sample}{Total\ number\ of\ baseline\ data\ sample} \qquad (7)$$

WPM method (Chen & Wen, 2017) is firstly used to identify baseline data that has a similar weather condition as the incoming snapshot data. The outliers in the baseline data are then counted by using the pre-defined thresholds. *LEAK* probability distribution can be obtained through Equation 5. An overall *LEAK* probability for a fault evidence is eventually calculated by averaging *LEAK* probability from the entire baseline database.

2.4.3. Fault Isolation

The posterior probability is used to isolate fault root cause. In this study, fault root cause is isolated by comparing the fault cause posterior probability. A fault is isolated, i.e., identified as the root-cause for an abnormality, by the following two isolation rules: 1) the posterior probability of this fault node is higher than 15%; and 2) the posterior probability of this fault node is the highest among all fault nodes and is 10% higher than the second-highest one.

3. FAULT DETECTION AND DIAGNOSIS STRATEGY

3.1. Building System Description

In this study, one Drexel campus building – Nesbitt Hall was selected to evaluate the proposed method. Nesbitt Hall is a seven-story, 78,000 square-foot mixed use building that houses offices, classrooms, laboratories, and an auditorium. This building includes a typical HVAC system that is commonly seen in medium-sized commercial buildings. A water cooled chiller subsystem provides the cooling need for the building. One air distribution subsystem which includes three air handling units (AHUs) and eighty-eight variable air volume (VAV) terminal units is used to serve all seven floors. One steam-to-hot-water heat exchanger subsystem located in the basement is used to provide domestic hot water and space heating needs. BAS is connected to the Nesbitt Hall to monitor and control the HVAC system. Figure 4 illustrates the HVAC system configuration, chiller plant and AHU configuration in the BAS in the Nesbitt Hall.
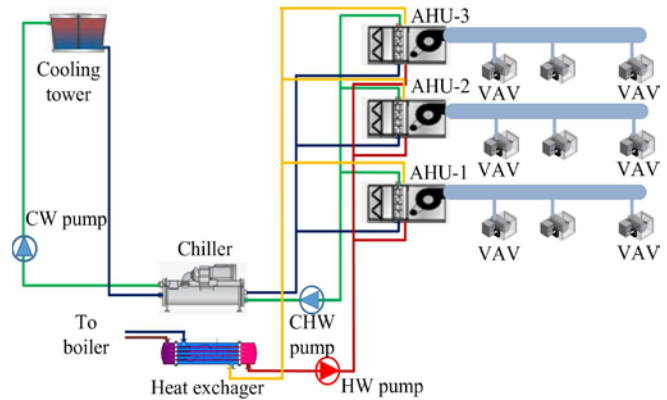

Figure 4 HVAC system configuration (Chen & Wen, 2017)

3.2. Whole Building Fault Description

Three types of faults including operator fault, primary cooling subsystem fault and supply air subsystem faults were implemented during summer 2017. The operator faults include faults such as 1) "chiller is off while under normal operation, it should be on", 2) "AHU cooling coil valve control override at a higher than normal position", and 3) "system is occupied while under normal operation, it should be unoccupied". Primary cooling subsystem faults include

faults such as 1) "chilled water supply temperature sensor negative bias (screen reading higher than real value)", and 2) "chilled water differential pressure sensor positive bias (screen reading higher than real value)". Supply air subsystem faults include faults such as "AHU outdoor air damper stuck at higher than normal position". These faults were selected because they are considered to have impacts on different subsystems or have significant impact on energy consumption. The faults were implemented through BAS connected to the HVAC system in the Nesbitt Hall or directly adjusting the equipment controller (e.g. chiller control panel) as shown in Figure 5.

Figure 5 Chiller and BAS connected to the HVAC system

4. RESULTS AND DISCUSSION

Eight fault cases which include seven artificially implemented faults and one naturally occurred fault from summer 2017 are employed to evaluate the developed BN strategy. Out of the eight fault cases, all fault are successfully detected and seven fault cases are successfully diagnosed. The detection and diagnosis results are listed in Table 1.

Table 1 Fault detection and diagnosis result

Date	Fault Description	Detection and Diagnosis Result
07/09	Chiller is off while under normal operation, it should be on	Detected/Diagnosed
07/11	AHU-2 OA damper stuck on 90% open (higher than normal)	Detected/Diagnosed
07/18	AHU-2 OA damper stuck on 100% open (higher than normal)	Detected/Diagnosed
07/22	Chiller DP sensor positive bias 1.4 kPa	Detected/Diagnosed
08/03	Chiller CHWS temperature negative bias 4°C	Detected/Mis-diagnosed
08/05	System is occupied while under normal operation, it should be unoccupied	Detected/Diagnosed
08/11	AHU-2 cooling coil valve position software override at 100% open (higher than normal)	Detected/Diagnosed
09/15	Chiller DP sensor positive bias 0.7 kPa	Detected/Diagnosed

4.1. Successfully Detected and Diagnosed Case Example

On July 11th, 2017, a damper stuck fault (stuck at a higher than normal position) was implemented on AHU-2. The outdoor air damper at AHU-2 was artificially stuck at 90% and 100% open positions by overriding the corresponding control signal in the BAS from 10:00PM to 08:01PM. The stuck positions (90% open) was higher than the damper's normal position (15% open) under similar weather conditions. In the summer season when HVAC operates under cooling mode, the AHU outdoor air damper is usually controlled to be at 15% opening to maintain the minimum requirement of fresh air flowrate. Under such circumstances, if the outdoor air damper is stuck at a position that is higher than normal (15% during cooling mode), cooling coil valve position will be increased to ensure the supply air temperature meet the setpoint requirement. However, in the early morning and evening hours, when outdoor enthalpy is lower than the return air enthalpy, the outdoor air damper is controlled under economizer mode to save energy. In these situations, the damper stuck fault would not yield a strong fault impact, as the stuck position is very similar to the normal damper position (under economizer mode). Figure 6 shows the fault detection results. It can be see that both of the faults are successfully detected by the proposed method. However, in some time periods (especially early morning and evening hours), T^2 statistic value does not overpass the threshold due to the lack of fault impacts. The BN method reports that there is (on average) 40% posterior probability for the "AHU-2 outdoor air damper stuck at too high fault". This posterior probability ranks No.1 compared with the posterior probabilities of other causes as shown in Figure 7. Therefore, this fault root-cause is successfully diagnosed.
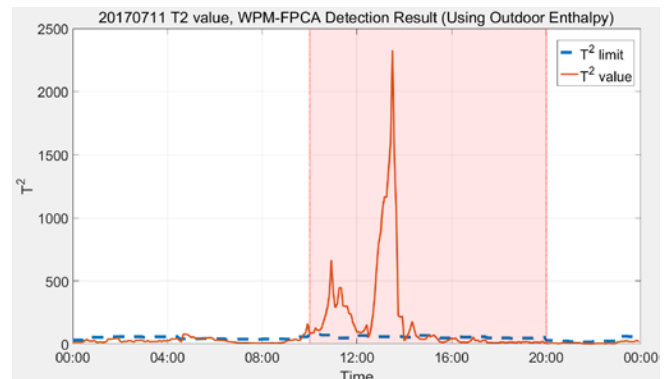

Figure 6 Fault detection result (July 11th, 2017)

4.2. Mis-diagnosed Case Example

On August 3th, 2017, a chilled water supply temperature sensor negative bias of 4 °F fault was implemented by adjusting the chilled water outlet temperature setpoint on the chiller control panel. The fault test period was from 10:00AM to 09:27PM. This fault had impacts on the downstream subsystems. During the test, we observed that only cooling coil valves in AHU-1 had higher than normal positions and

chilled water return temperature had higher than baseline value.

In this case, the root cause of fault was not successfully diagnosed by the BNs method. In the diagnosis result, posterior probabilities for "AHU-2 cooling coil valve open too high" fault and "AHU-2 outdoor air damper stuck at too high" fault reach higher levels compared with "chilled water supply temperature negative bias" fault. The peak posterior probability for these fault cause are: 0.78 for "AHU-2 cooling coil valve operate at too high " fault, 0.24 for "AHU2 outdoor air damper stuck at too high position" fault, posterior probability, and 0.04 for "Chilled water temperature negative bias" fault as shown in Figure 8.
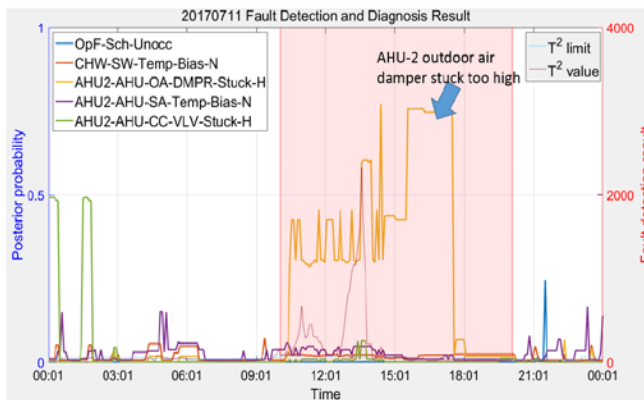

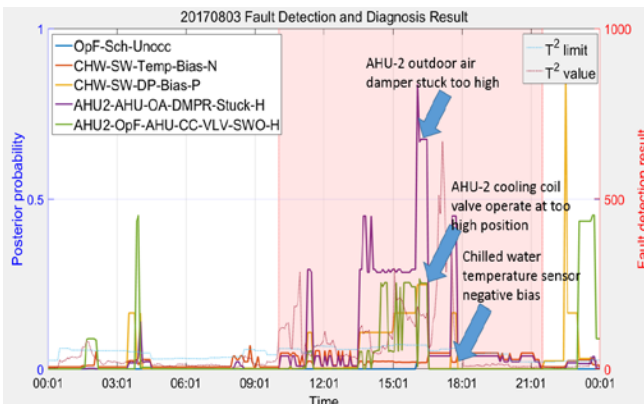Figure 7 Fault diagnosis result (July 11th, 2017)

Figure 8 Fault diagnosis result (August 3th, 2017)

There are two reasons may cause this mis-diagnosed case. One reason is that not too many evidences are found in chiller, AHUs during the fault test period. In this case, only chilled water return temperature and AHU-1 cooling coil valve position are found to be abnormal compared with the baseline data. As more evidences can strengthen BN inference and hence generate a higher posterior probability, it is therefore BN based fault diagnosis method does not diagnose the fault in this case. Another potential reason could be that the conditional probability for each root fault cause is set to be equal when these faults have the same evidences.

Since the "chilled water temperature negative bias" fault would cause similar fault evidences as other faults such as "cooling coil valve open too high" and "outdoor air damper stuck at too high" faults, without additional conditional probability information, the BN would not be able to differentiate the faults.

5. CONCLUSION

In this research, a fault detection and diagnosis strategy which incorporates data-driven methods and expert knowledge based method is proposed for detecting and diagnosing whole building level faults. A WPM method is employed to find similar system operation data and generate baseline data. FPCA method are proposed to detect whole building fault detection. BN is developed based on the expert knowledge for BN structure model generation. *LEAK* probability distribution and online system evidence generation are obtained by data-driven method. System operation data collected from one Drexel campus building is employed to evaluate the proposed strategy. The evaluation from the cooling season demonstrates that the proposed strategy is very efficient to detect and diagnose whole building level faults with strong symptoms in different sub-systems.

ACKNOWLEDGEMENT

REFERENCES

Bynum, J. D., Claridge, D. E., & Curtin, J. M. (2012). Development and testing of an Automated Building Commissioning Analysis Tool (ABCAT). *Energy and Buildings, 55*, 607-617. doi: 10.1016/j.enbuild.2012.08.038

Chen, Y., & Wen, J. (2017). *A whole building fault detection using weather based pattern matching and feature based PCA method.* Paper presented at the 2017 IEEE International Conference on Big Data, Boston, Massachusetts, U.S.

Cheng, C., & Chiu, M.-S. (2004). A new data-based methodology for nonlinear process modeling. *Chemical Engineering Science, 59*(13), 2801-2810.

Drużdżel, M. J. Intelligent Decision Support Systems Based on SMILE.

EIA. (2012). How much energy is consumed in residential and commercial buildings in the United States? Retrieved January 30, 2018, from https://www.eia.gov/tools/faqs/faq.php?id=86&t=1

Katipamula, S., & Brambley, M. R. (2005). Methods for fault detection, diagnostics, and prognostics for building

systems—a review, part I. *HVAC&R Research, 11*(1), 3-25.

Keogh, E., Lin, J., & Fu, A. (2005). *Hot sax: Efficiently finding the most unusual time series subsequence.* Paper presented at the Data mining, fifth IEEE international conference on data mining.

Lampis, M., & Andrews, J. (2009). Bayesian belief networks for system fault diagnostics. *Quality and Reliability Engineering International, 25*(4), 409-426.

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). *A symbolic representation of time series, with implications for streaming algorithms.* Paper presented at the Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery.

O'Neill, Z., Pang, X., Shashanka, M., Haves, P., & Bailey, T. (2014). Model-based real-time whole building energy performance monitoring and diagnostics. *Journal of Building Performance Simulation, 7*(2), 83-99. Bynum, J. D., Claridge, D. E., & Curtin, J. M. (2012). Development and testing of an Automated Building Commissioning Analysis Tool (ABCAT). *Energy and Buildings, 55*, 607-617. doi: 10.1016/j.enbuild.2012.08.038

Regnier, A., & Wen, J. (2016). *Automated Fault Diagnostics for AHU-VAV Systems: A Bayesian Network Approach*. Paper presented at the 4th High Performance Buildings, West Lafayette, Indiana, U.S.

Sun, B., Luh, P. B., Jia, Q.-S., O'Neill, Z., & Song, F. (2014). Building energy doctors: An SPC and Kalman filter-based method for system-level fault detection in HVAC systems. *IEEE transactions on automation science and engineering, 11*(1), 215-229.

Wise, B. M., Gallagher, N., Bro, R., Shaver, J., Windig, W., & Koch, R. S. (2007). PLS Toolbox 4.0. *Eigenvector Research Incorporated, 3905*.

Wu, S., & Sun, J. Q. (2011). A top-down strategy with temporal and spatial partition for fault detection and diagnosis of building HVAC systems. *Energy and Buildings, 43*(9), 2134-2139.

Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis, 51*(2), 918-930.

Biographies

Yimin Chen is currently pursuing Ph.D. degree in the Department of Civil, Architectural and Environmental Engineering at Drexel University. He research focus is development of fault detection and diagnosis methods for whole building system funded by the Department of Energy, U.S. He received his Master Degree in Electrical Engineering at Beijing Jiaotong University in 2009, research focus was monitor and detection of terminal temperature rise molded case circuit breaker in buildings.

Jin Wen (Ph.D. in Architectural Engineering). She is a professor in the Department of Civil, Architectural and Environmental Engineering at Drexel University. She received her Ph.D. degree in the Department of Mechanical and Industrial Engineering, University of Iowa in 2003. Prof. Wen has long been engaged in the field of HVAC system FDD and building energy efficiency. She served as a past Chair for the Fault Detection and Diagnosis Sub-committee (under Smart Building Systems Committee) in ASHRAE.