University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2016

A sensor fault detection strategy for air handling units using cluster analysis

Rui Yan

University of Wollongong, ry721@uowmail.edu.au

Zhenjun Ma

University of Wollongong, zhenjun@uow.edu.au

Georgios Kokogiannakis

University of Wollongong, gkg@uow.edu.au

Yang Zhao

Zhejiang University

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A sensor fault detection strategy for air handling units using cluster analysis

Abstract

Sensors are an essential component in the control systems of air handling units (AHUs). A biased sensor reading could result in inappropriate control and thereby increased energy consumption or unsatisfied indoor thermal comfort. This paper presents an unsupervised learning based strategy using cluster analysis for AHU sensor fault detection. The historical data recorded from sensors is first pre-processed to reduce the dimensions using principal component analysis (PCA). The clustering algorithm Ordering Points to Identify the Clustering Structure (OPTICS) is then employed to identify the spatial separated data groups (i.e. clusters), which possibly indicate the occurrence of sensor faults. The data points in different clusters are then checked for temporal separation in order to confirm the occurrence of sensor faults. The proposed sensor fault detection strategy is tested and evaluated with the data collected from a simulation system. The results showed that this strategy can detect single and non-simultaneously occurred multiple sensor faults in AHUs. The fault detection results were not strongly affected by the selection of the user defined input parameters required in OPTICS.

Keywords

fault, detection, strategy, air, handling, units, cluster, sensor, analysis

Disciplines

Engineering | Science and Technology Studies

Publication Details

Yan, R., Ma, Z., Kokogiannakis, G. & Zhao, Y. (2016). A sensor fault detection strategy for air handling units using cluster analysis. *Automation in Construction*, 70 77-88.

A Sensor Fault Detection Strategy for Air Handling Units Using Cluster Analysis

Rui Yan^a, Zhenjun Ma^{a,*}, Georgios Kokogiannakis^a, Yang Zhao^b

^aSustainable Buildings Research Centre (SBRC), University of Wollongong, New South
Wales (NSW), 2522, Australia

^bInstitute of Refrigeration and Cryogenics, Zhejiang University, China

*Phone: (61) 02 4221 4143, email: zhenjun@uow.edu.au

Abstract: Sensors are an essential component in the control systems of air handling units (AHUs). A biased sensor reading could result in inappropriate control and thereby increased energy consumption or unsatisfied indoor thermal comfort. This paper presents an unsupervised learning based strategy using cluster analysis for AHU sensor fault detection. The historical data recorded from sensors is first pre-processed to reduce the dimensions using principal component analysis (PCA). The clustering algorithm Ordering Points to Identify the Clustering Structure (OPTICS) is then employed to identify the spatial separated data groups (i.e. clusters), which possibly indicate the occurrence of sensor faults. The data points in different clusters are then checked for temporal separation in order to confirm the occurrence of sensor faults. The proposed sensor fault detection strategy is tested and evaluated with the data collected from a simulation system. The results showed that this strategy can detect single and non-simultaneously occurred multiple sensor faults in AHUs. The fault detection results were not strongly affected by the selection of the user defined input parameters required in OPTICS.

Keywords: Sensor fault detection; Air handling units; Principal component analysis; Cluster analysis; OPTICS

Nomenclature

Cov	covariance
Eps	a user defined radius
IQR	interquartile range
k	the k^{th} nearest data point
m	number of the sample data
$MinPts$	minimum number of data points
n	number of the variables
N_{Eps}	a set of data points within the Eps of a given point
o, p, q	data point
Q	quartile

26

27

28

29

30

31

32

33

34

35

36

37

1. Introduction

In commercial buildings, approximately 50% of energy is consumed by heating, ventilation and air conditioning (HVAC) systems in order to maintain satisfied indoor thermal comfort [1]. However, it was found that up to 15% of energy used in buildings is wasted due to improper selection of building systems and components, lack of appropriate maintenance, and the occurrence of various faults including control faults, sensor faults and component performance degradation, among which a significant proportion was attributed to HVAC systems [2, 3]. Energy efficiency in buildings can be improved if faults in HVAC systems can be timely detected, isolated and corrected with appropriate methods.

Over the last several decades, many efforts have been made on the development and application of various fault detection and diagnosis (FDD) strategies for individual components of HVAC systems [4-7]. FDD methods can be generally categorized into three groups, including physical model-based methods, rule-based methods, and data-driven methods [8]. Data-driven methods have attracted increasing attention due to the availability of abundant operational data from building management systems (BMSs) [8-14]. Data-driven methods do not require the high-level physical knowledge of the system and the computational costs are generally manageable [4]. However, they are only reliable for the operating conditions within the range covered by the training data, and extrapolation outside this range may lead to significant errors [4, 15].

An air handling unit (AHU) is one of the key components of HVAC systems and different data-driven methods have been developed and used to detect the AHU's sensor bias, control valve fault, coil fouling and inappropriate control [8, 11, 16-18]. Du *et al.* [8], for instance, developed a sensor fault detection strategy for AHUs, in which a basic neural network and an auxiliary neural network were used to detect the measuring biases of the sensors in the AHU supply air temperature control loop. Wang and Xiao [11] employed a principal component

analysis (PCA) model for AHU sensor FDD and concluded that the PCA-based method required less training data and was easier to be built as compared to artificial neural networks (ANN)-based methods. The proposed method successfully identified a biased outdoor air temperature sensor and a completely failed outdoor air humidity sensor based on the data collected from an AHU installed in a high-rise commercial building. Fan *et al.* [16] presented a hybrid FDD strategy for AHUs by using ANNs, in which wavelet analysis was used as a noise and outlier filter. Li and Wen [17] coupled a pattern matching method with PCA for AHU fault detection. The proposed pattern matching method was able to match the real-time operating condition with the historical data recorded under similar operating conditions. An Autoregressive Exogenous (ARX) model was utilized by Yoshida and Kumar [18] for offline detection of AHU faults. It was found that the ARX model is robust and capable of detecting most faults tested. Mulumba *et al.* [19] proposed an FDD strategy for AHU based on an ARX model and the support vector machine (SVM) method. The performance of this strategy was validated using the data collected from a real AHU system. Bruton *et al.* [20] developed a cloud-based system for AHU FDD, which can automatically retrieve data from BMS, run FDD application and present the results graphically. The developed system was tested with several AHU systems and successfully identified different faults and achieved €104,000 annual energy savings in total.

Data mining based methods have also been used for fault detection in buildings. Data mining based FDD methods extract the useful information directly from the whole data set and generally do not require a detailed training process as no any physical model is required. Yu *et al.* [21] proposed an association rule mining (ARM) based method for detecting the system faults and inefficient operation strategies. In the study carried out by Capozzoli *et al.* [22], a supervised learning method called Classification and Regression Tree (CART) was used to detect abnormal lighting energy usage patterns. The result was compared with that

using the k-means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) methods. It was concluded that CART is a more appropriate method for detecting abnormalities related to the use of lights and DBSCAN is useful for detecting outliers. However, DBSCAN can only identify the clusters with very similar densities and the fault detection results heavily rely on the user defined input parameters [22]. Fan *et al.* [23] proposed a generic framework to extract information from massive BMS data by using cluster analysis based data mining techniques to identify the typical operation patterns while the quantitative ARM method was used to extract the information of the system's faults and inefficient operation.

Cluster analysis is a data mining method that is able to directly extract the useful information from a given data set. It categorises the data into groups (i.e. clusters) that share common characteristics [24]. The use of cluster analysis methods for fault detection has been studied and reported in several research areas. Yiakopoulos *et al.* [25], for instance, demonstrated the benefits of using traditional k-means clustering for fault detection of rolling element bearing. A fuzzy c-means clustering based method was applied for sensor fault detection in nuclear steam turbines and mobile robots [26, 27]. However, fault detection methods based on cluster analysis have not been widely applied to HVAC systems mostly because it is often difficult to define the required input parameters in some cluster analysis methods such as DBSCAN and k-means [22, 24, 25].

Based on the above review, it can be seen that the research on fault detection of AHUs using data mining technologies is still not significantly progressed. In this study, a fault detection strategy using cluster analysis is developed to detect the possible faults in AHU sensors. The proposed method couples the clustering algorithm from Ordering Points to Identify the Clustering Structure (OPTICS) [28] with a PCA method to identify normal and

faulty operation data from the raw data set. The proposed sensor fault detection method was validated in a virtual environment through computer simulations.

2. Development of a cluster analysis based fault detection strategy

2.1 Outline of the fault detection strategy for air handling units

The proposed sensor fault detection strategy for AHUs is outlined in Fig. 1. It mainly consists of three steps; data pre-processing, data clustering and result evaluation. As sensor faulty data and normal data have different characteristics, the strategy is developed based on the assumption that the faulty data will be spatially and temporally separated from the normal data. The separation between the faulty data and normal data is detected using cluster analysis. In this fault detection strategy, a set of fault-free data is required as the reference to facilitate the fault detection, which can be generally obtained after a thorough commissioning of the HVAC system.

Based on the data retrieved from BMS and the recorded historical fault-free data, data categorization is first performed to separate the data into different subsets according to the operation modes of the HVAC systems (e.g. on/off status, heating operation, cooling operation, natural ventilation) because the data in different operation modes may result in the existence of different clusters. In this study, the data set with the HVAC mechanical cooling operation is used for subsequent analysis to demonstrate the effectiveness of the proposed fault detection strategy. The cooling energy required by the AHU is calculated using the measurements of the AHU chilled water supply and return temperatures and the water flow rate. The operation data with the calculated cooling energy within the interquartile range (IQR), which represents the range with the most concentrated observations, is then used for fault detection. The overall aim of this step is to filter out the data with strong dynamics and the data that is recorded under the extreme operational conditions. If the required measurements for cooling energy calculation are not available, other sensor measurements or

control signals such as the opening of cooling/heating coil control valves can be used for this purpose. However, if a significant sensor fault occurred and only a relatively short period (i.e. one or two days) of faulty operation data is in the data set used for fault detection, the whole faulty observations may be filtered out. To avoid missing any fault, the temporal distribution of the filtered data should be checked. If a period of the observations is missed, it can be concluded that a fault occurred and the fault detection is then terminated. Otherwise, PCA is used to reduce the dimensions of the raw data as the irrelevant dimensions may prevent the discovery of real clusters and the data will become sparse with the increasing dimensions, which could also result in the similarity measures (i.e. Euclidean distance) being meaningless [29].

Once the data pre-processing is completed, cluster analysis based on the OPTICS algorithm is then used for data analysis. How to use OPTICS for fault detection is described in Section 2.2. Through OPTICS analysis, the observations which meet the cluster criteria will be grouped into clusters while the rest of the data in less-dense areas will be grouped into a cluster and labelled as cluster 0, which is considered as noise and will be neglected in the later stages of the fault detection process.

Based on the results of the cluster analysis, the temporal distribution of the identified clusters with more than a minimum number of data points (*MinPts*, which will be introduced in Section 2.2) can be plotted using the box plot. The box plot is a commonly used statistical plotting method [30] and, in this strategy, is used to check whether the valid clusters exist within the identified clusters. An example of a box plot is illustrated in Fig. 2, in which the range between the first quartile (Q1) and the third quartile (Q3) is plotted as a box (also named as IQR), while the whiskers are extended from Q1 and Q3 to the minimum and maximum data points within $Q1-1.5 \times IQR$ and $Q3+1.5 \times IQR$ ranges, respectively. The data points fall outside this range is considered as outliers. Valid clusters are the clusters that are

temporally separated, which means that the temporal distribution of the observations in a cluster is not overlapped by the temporal distribution of at least another cluster. If the temporal distribution of the box and the whisker part of two clusters are overlapped and both clusters are temporally separated from another cluster, the cluster with fewer data points in the two overlapped clusters will be neglected in later analysis. However, for clusters with the box parts separated but the whisker parts overlapped, fault occurrence is also possible and such clusters are considered as valid clusters in this strategy. This is because the spatial separation may not be sufficiently good to detect the insignificant bias. If valid clusters are identified, the total fault number can be determined, which is the number of valid clusters minus 1. Otherwise, there is no sensor fault occurred in the AHU. It should be noted that outliers identified by the box plot are neglected in the analysis.

In order to facilitate the proposed fault detection process, an interface was developed to help users to select the required parameters and easily check the fault detection results by looking at the graph automatically generated, which will be introduced in Section 2.3.

The above fault detection strategy was implemented in Rapidminer 6.0 [31]. The OPTICS algorithm was programmed in R [32] and linked with Rapidminer using R-extension. The figures with the analysis results were generated using the R package ggplot2 [33].

2.2 OPTICS cluster analysis

OPTICS is a density-based clustering method and is an extension of the DBSCAN algorithm. Generally, density-based clustering methods search for the connected dense spaces that are separated by low dense spaces[24]. Compared to some partitioning based algorithms like k-means, density-based methods can identify arbitrarily shaped clusters, are robust to noise, and do not require pre-determined cluster numbers [24]. The basic idea of the density-based clustering is that for each data point in a cluster the neighbourhood of a given radius (Eps) must contain at least a minimum number of data points ($MinPts$) [28].

In general, there is no generic way to determine *MinPts* [24, 34]. The low density regions with a small number of noise points will be labelled as clusters if a too small value was assigned to *MinPts*. In contrast, the meaningful clusters with the number of the data points fewer than *MinPts* are likely to be neglected if *MinPts* is too large. The results from Ankerst *et al.* [28] showed that *MinPts* in the range of 10-20 can mostly result in good clustering outcomes. In this study, *MinPts* was chosen as 15. For a given *MinPts*, *Eps* can be determined based on the k-distance graph, as suggested by Ester *et al.* [34]. The k-distance means the distance from a given point to its k^{th} nearest neighbour point, where k equals to *MinPts*. As shown in Fig. 3, k-distance graph plots the sorted k-distance of all data points of concern. The value of *Eps* can then be determined at the place where the k-distance starts to change dramatically. OPTICS is not sensitive to the *Eps* and *MinPts* selected, and this will be demonstrated in Section 5.

If a point p is directly density-reachable from another point q , it must satisfy the below criteria [34].

$$p \in N_{Eps}(q) \ \& \ |N_{Eps}(q)| \geq MinPts \quad (1)$$

where, $N_{Eps}(q)$ is the set of the data points within the *Eps* of the point q .

Unlike most other clustering algorithms that directly group the data points into clusters, the output from OPTICS algorithm is a list of the ordered data points with respect to the *reachability-distance* that allows for visualising and cluster identification. The *reachability-distance* of a point p with respect to another point o is defined in Eq. (2), in which the core-distance of the point p is defined in Eq. (3) [28]. Fig. 4 graphically demonstrated how the concepts are defined.

$$reachability-distance = \begin{cases} \text{UNDEFINED, if } |N_{Eps}(o)| < MinPts \\ \max(\text{core-distance}(o), \text{distance}(o, p)), \text{ otherwise} \end{cases} \quad (2)$$

$$\text{core-distance} = \begin{cases} \text{UNDEFINED, if } |N_{\text{Eps}}(p)| < \text{MinPts} \\ \text{MinPts-distance}(p), \text{ otherwise} \end{cases} \quad (3)$$

The OPTICS algorithm starts with a random point and expands the visiting to its *directly density-connected* neighbours, followed by sorting the points with the *reachability-distance* visited so far in an ascending order and then expanding again from the first unexpanded point in the order list. This process virtually walks through the Minimum Spanning Tree [35]. Fig. 5 presents an example of the OPTICS outputs, which are a cluster-order of the data points and the corresponding reachability-distance. In order to group the data into different clusters, a user defined reachability-distance threshold is required which can be intuitively selected based on the graph. It is worthwhile to note this cluster-order of the data points in Fig. 5 is sorted based on the sequence of the walking through the Minimum Spanning Tree, which is different from the order of the data points presented in Fig. 3.

The valleys under the threshold separated by the intersection of the threshold and the reachability-distance indicate the existence of clusters. For fault detection purposes, clusters that include a large amount of data points should be first investigated. Therefore, the reachability-distance threshold can be first set to 1.3 for this example as this threshold forms two distinct clusters. If the threshold is set as 0.52 (Fig. 5), three clusters can be identified indicating multiple faults may occur. The possibility of the existence of multiple faults should be confirmed by the temporal distribution of the identified clusters as mentioned previously. Details on OPTICS can be found in [28, 34].

2.3 User interface for Eps value and threshold selection

A user interface for assisting the selection of the k-distance (Eps) value, the reachability-distance threshold, and for visualization of the temporal separation of clusters was developed using Shiny [36]. The interface incorporates two slide bars for selecting the two values, as illustrated in Fig. 6. For Eps value selection, by dragging the slide bar, the selected Eps value

will be updated in the k-distance diagram and the reachability-distance graph will be updated as well based on the selected Eps value. Similarly, by dragging the reachability-distance threshold slide bar, the corresponding threshold line and the temporal distribution plot will be updated. With this user interface, the process of selecting the Eps value and the threshold for reachability-distance is streamlined.

2.4 PCA based dimensionality reduction

The raw data is often in high dimensions including several variables. Reducing the dimensions of the raw data will result in a better clustering result. Performing clustering on reduced dimensionality can also save computational cost [37].

There are two types of dimension reduction methods, i.e. feature selection and feature extraction. Feature selection based methods generate a subset of the raw data set through selecting the dimensions or features that contain irrelevant information (e.g. noise), while feature extraction based methods reduce the dimensions by extracting the major information from the raw data and representing it with the low dimensional data [38]. In this study, a PCA-based feature extraction method is used for the dimension reduction.

PCA is a multivariable data analysis method that is widely used for reducing high dimensions in the data sets with a large number of interrelated variables by transforming the existing data set into a new set of variables called principal components (PCs). For an n -dimensional data set, there are n corresponding PCs. For most cases, the first few PCs are enough to represent the most significant variations of the data set [39]. Mathematically, the PCs of a data set $X \in \Re^{m \times n}$, where m is the number of the sample data and n is the number of the variables, can be determined by calculating the eigenvectors and eigenvalues of its covariance matrix as defined in Eq. (4) [17].

$$\text{Cov}(X) = \frac{X^T X}{(m-1)} \quad (4)$$

The resulted first PC is the eigenvector associated with the largest eigenvalue, and the second PC is the eigenvector with the second largest eigenvalue and so on. A number of different methods can be used to determine the minimum number of PCs that should be retained. In this study, the intuitive scree diagram method as described by Wang and Xiao [11] was used. Once the number of PCs is determined, the original data will then be projected on the new PC space to get a new data set with fewer dimensions but retaining the most significant variations of the original data. It should be noted that as sensor readings are in different units and ranges, all variables should be standardised to zero mean and unit variance before conducting PCA. It is worthwhile to mention that unlike other PCA based fault detection strategies where PCA was used for both dimension reduction and residual prediction [11, 17, 40], the PCA is only used for dimension reduction in this proposed strategy. By doing this, the predictive model training and validation process required in PCA-based methods can be avoided. It is also to note that the threshold used in this study is different from the threshold used in PCA based FDD methods.

3. Simulation of AHU and sensor faults

In this study, the effectiveness of the proposed sensor fault detection strategy was validated through simulations using TRNSYS [41]. The AHUs simulated were installed in the Sustainable Buildings Research Centre at the University of Wollongong, Australia. There are two AHUs, named as AHU1 and AHU2, used to provide the heating and cooling to the ground floor and the first floor of the building, respectively. In this study, only AHU2 under the cooling operation was considered for the sensor fault detection. The HVAC plant operated from 7:00am to 6:00pm during the weekdays and the data sampling interval was 10 minutes. In the simulation, the cooling coil was simulated using TRNSYS component model Type 52b. An air cooled chiller modelled using Type 655 was used to supply the chilled water at 7.0°C.

The supply air temperature was maintained at 14.0°C through controlling the opening of the control valve to regulate the required chilled water flow rate.

As shown in Fig. 7, the major measurement instruments related to the operation of AHUs include the supply air temperature sensor, supply and return chilled water temperature sensors, supply air flow rate meter and supply water flow rate meter. Any fault in these sensors may result in inefficient operation of the HVAC system or poor indoor thermal comfort. As certain types of correlations exist among these measured variables, the readings from a biased sensor may not comply with the existing correlations and the fault can then be detected by identifying the variations of the correlations among the measurements. In the simulation, a sensor reading bias generator was used to introduce different sensor biases into the sensor measurements.

The simulation was carried out for three months from January to March (i.e. approximately 2100 hours) under Sydney weather conditions. The international weather for energy calculation (IWEC) weather file was used in the simulation.

Table 1 summarises different test cases used to validate the effectiveness of the proposed fault detection strategy. The first test case was mainly used to demonstrate how to detect AHU sensor faults using cluster analysis and the outputs from each step of the proposed strategy. The test cases 2-7 and test cases 8-10 were used to validate the performance of the proposed fault detection strategy with single sensor fault and multiple sensor faults in the AHU, respectively. If the measurement of an AHU sensor is used for system control, the bias introduced will influence the system energy consumption. For instance, supply air temperature is often used to control the opening of the control valves to regulate the chilled water flow rate. A negative bias in the supply air temperature sensor will result in a larger chilled water flow rate than that of normal condition (i.e. no bias), which will increase the

power consumption of the water pumps. The minimum fault severity introduced in these cases are in line with the severity tested in previous studies [8, 11, 16].

4. Performance test and evaluation of the fault detection strategy

4.1 Demonstration of the fault detection process of the proposed strategy

In test case 1, a -2.0K bias was introduced to the AHU supply air temperature sensor at the 1421st hour. The operation data was first categorized according to the on-off status and only the cooling operation data was used for fault detection. The cooling operation data that was not within the IQR in terms of the cooling energy required by the AHU was filtered out. Fig. 8 presents the temporal spread of the post-filtered observations within the IQR as histogram and the density profile of the remaining data points, where each bin in Fig. 8(a) represents the total number of the data points remained in each test day. It can be seen that the post-filtered observations distributed over the whole test period without significant skewness, which demonstrates that the post-filtered observations can represent the observations in the whole test period. It is worthwhile to note that, in some test days, no data points remained due to the small cooling energy required by the AHU and the data points in these test days were therefore filtered out. The next step determined the number of PCs based on the scree diagram and 4 PCs were selected.

Once the PC number has been determined, the next step is to determine the input parameter *Eps* for OPTICS by using the *k-distance* diagram. Fig. 9 illustrates the *k-distance* of all data points remained for fault detection. As the *k-distance* started to change dramatically when *Eps* was 1.2, this value was therefore selected for *Eps* in OPTICS analysis.

Fig. 10 shows the reachability-distance of the cluster-order of the data points. It can be observed that there were two distinct valleys when Threshold 1 was used. The two distinct valleys resulted in two significant clusters, indicating that a sensor fault could have occurred. However, the occurrence of the sensor fault needs to be further verified by analysing the

temporal distribution of the clusters as illustrated in Fig. 11, in which the number in the box indicated the total number of the data points in a cluster. As Fig. 11(a) shows, there were three clusters formed under Threshold 1, among which the two major clusters (i.e. clusters 1 and 2) were completely temporally separated while the cluster 3 was overlapped with the cluster 1. This confirms that a sensor fault occurred during the test period. The data points in the cluster 2 started at the 1421.17 hour, which shows that the fault was detected at this time.

Less significant clusters should also be checked to confirm whether more faults occurred during the test period. As shown in Fig. 10, there were five clusters if, for example, the Threshold 2 is chosen. However, from Fig. 11(b), it can be observed that clusters 1-3 are overlapped with each other but separated from the overlapped clusters 4 and 5, indicating that there were only two valid clusters and one fault occurred during the whole test period. The fault was detected at the 1423.17 hour.

The above results demonstrated that the fault can be detected by using the proposed method. Different thresholds for the reachability-distance may result in different numbers of clusters. However, the proposed strategy is capable of determining whether the identified clusters can represent the occurrence of the sensor fault.

4.2 Validation of the proposed strategy for single sensor fault detection

In this section, the performance of the proposed fault detection strategy was tested and evaluated with no AHU sensor fault (Case 2) and single sensor fault in different sensors (Cases 3-7). Fig. 12 shows the reachability-distance of the cluster-order of the data points when the AHU operated under the fault-free conditions. Ideally, there will be only one cluster during the whole test period since there was no fault introduced to the sensors. However, Fig. 12 shows that there were two distinct clusters together with another three insignificant clusters that can be observed when a small threshold of 0.34 for the reachability-distance was used. As the data points in the two insignificant clusters were fewer than the *MinPts* and both

clusters were therefore not further considered. Fig. 13 shows the temporal distribution of the data points in the identified clusters. It is clearly illustrated that the three clusters identified were temporally overlapped, which means that there was no fault occurred in the AHU sensors.

Fig. 14 shows the fault detection results in terms of the temporal distribution of the data points in the clusters when different sensor faults were introduced individually. When a -0.5K supply air temperature sensor bias was introduced (Case 3), three clusters were identified, among which only the clusters 2 and 3 were valid clusters as the cluster 1 was overlapped with both clusters 2 and 3. This indicates that one sensor fault was detected and the fault was detected at the 726.5 hour, which was close to the actual time when the fault was introduced. When a -0.5K return water temperature sensor bias was introduced at the 271 hour (Case 4), three clusters were also identified. Similar to Case 3, there were only two valid clusters (i.e. clusters 1 and 2). The fault was detected at the 271.17 hour.

In the test cases 5 and 7, there were only two clusters and both of them were temporally separated. Therefore, only one fault was identified in each case. The fault was detected at the 1421.67 hour for the case 5 and the 896.33 hour for the case 7. For the test case 6, the insignificant cluster 3 was overlapped with the significant cluster 1. Therefore, the cluster 3 was not further considered. The two significant clusters 1 and 2 had the box parts separated but the whisker parts partially overlapped. As mentioned in Section 2, both clusters were considered as the valid clusters, which means that a fault has occurred. The fault detected time can be considered as the starting time of the cluster 2, i.e. the 1302.33 hour, which was slightly different from the actual time when the fault was introduced.

The above results showed that the proposed fault detection strategy is capable of detecting single AHU sensor fault and can relatively reliably determine the fault occurring time.

4.3 Validation of the proposed strategy for multiple sensor faults detection

The performance of the proposed strategy was also validated for multiple AHU sensor faults. Three different cases were tested. In the test case 8, a -0.5K bias was introduced to the supply air temperature sensor and return water temperature sensor at the 568 and 1615 hour, respectively. Fig. 15 shows the OPTICS clustering output. It can be seen that there were three significant clusters formed together with two small clusters if the threshold was set to 0.4. However, the temporal distribution of the identified clusters indicated that only two valid clusters existed out of the five identified clusters (see Fig. 16). The second valid cluster started at the 1641.33 hour, which was close to the return water temperature sensor fault occurring time. It was also tested that new valid clusters cannot be formed if the reachability-distance threshold is further reduced. Therefore, only the return water temperature sensor fault was identified in this test case.

To further test whether the proposed strategy can detect the multiple sensor faults if the severity of the fault increases, a -1.0K bias was introduced to the supply air temperature sensor and return water temperature sensor at the 568 and 1615 hour, respectively. The reachability-distance of the cluster-order of the data points for this case is shown in Fig. 17. It can be seen that there were two distinct clusters formed under the Threshold 1. The resulted temporal distribution of the data points in the clusters are shown in Fig. 18(a). It is clearly shown that there were two valid clusters, indicating the existence of a sensor fault, which was detected at the 1618.83 hour. If further decreasing the threshold (i.e. Threshold 2), 10 clusters were formed. It should be noted that only the clusters with the data points more than the *MinPts* were illustrated in Fig. 18 (b). From the temporal distribution of the data points in the clusters shown in Fig. 18 (b), it is found that there were three valid clusters among the 10 identified clusters. The three valid clusters mean the existence of two sensor faults and the faults were detected at the 610.83 and 1783.83 hour. The second fault detected at the 1783.83

hour represents the same fault as detected with the Threshold 1. The fault detected time by using the Threshold 1 was closer to the actual time when the fault was introduced. Therefore, when determining the fault occurring time in multiple sensor faults scenarios and the same fault is detected with different thresholds, the fault detected time with the larger threshold should be used.

In order to further evaluate the performance of the proposed strategy, a test case was prepared with three sensor faults, including the return water temperature sensor fault, water flow rate meter fault and supply air temperature sensor fault. The above three faults were introduced at the 584, 774 and 1810 hour, respectively. The fault severities for the three sensors were -1.0K, -0.254 L/s and -1.0K, respectively. Fig. 19 shows the reachability-distance of the cluster-order of the data points when the three different faults were introduced. When the Threshold 1 was applied, two valid clusters were formed and a sensor fault was detected at the 774.33 hour (see Fig. 20(a)). Three valid clusters were formed when the Threshold 2 was used (see Fig. 20(b)) and two sensor faults were detected at the 587.17 hour and 774.33 hour, respectively. The detected fault occurring time matched well with the time when the actual faults were introduced. Furthermore, a new valid cluster was identified if the Threshold 3 was employed, as shown in Fig. 20(c) where only the valid clusters were plotted. A new fault was detected at the 1784 hour and it is associated with the -1.0K biased supply air temperature sensor fault. This demonstrates that the spatial separation formed due to the -1.0K biased supply air temperature sensor fault was less significant when comparing to the spatial separation formed by the other two faults as detecting this sensor fault requires a small threshold of the reachability-distance.

The above results showed that the proposed strategy is able to detect multiple sensor faults of AHUs. However, the detection of different types of sensor faults may require different

thresholds. It is essential to check smaller valleys in the reachability-distance of the cluster order of the data points to avoid failing to detect all possible faults.

5. Sensitivity analysis of OPTICS user defined parameters on the fault detection results

As mentioned earlier, there are two user defined parameters in OPTICS algorithm (i.e. *Eps* and *MinPts*). A sensitivity analysis for the selection of the two parameters on the overall fault detection results is therefore undertaken in this section. The analysis was carried out based on the conditions specified in Case 1 with a -2.0K supply air temperature sensor bias. The *Eps* and *MinPts* varied from the default settings determined in Section 4.1 (i.e. $Eps = 1.2$ and $MinPts = 15$) by ± 0.5 and ± 5 , respectively. Fig. 21 shows the reachability-distance of the cluster-order of the data points and the temporal distributions of the data points in the identified clusters when different values of *Eps* and *MinPts* were used. It can be seen that two distinct clusters can be successfully identified from the reachability-distance diagrams under different combinations of *Eps* and *MinPts* values. The temporal distribution of the data points in the identified clusters also showed that there were only two valid clusters and the fault detected time was at the actual time when the fault was introduced. The result is also summarized in Table 2. This demonstrates that the OPTICS based fault detection strategy is not sensitive to the user defined input parameters.

6. Conclusions

This paper presented an offline sensor fault detection strategy for air handling units (AHUs) using clustering algorithm Ordering Points to Identify the Clustering Structure (OPTICS). The fault(s) can be detected through identifying the spatial and temporal separation of the monitored data. The effectiveness of the proposed strategy was tested and evaluated through computer simulations. The results showed that the overall sensor fault detection results were not sensitive to the selection of the user defined parameters (i.e. the given radius *Eps* and the minimum number of data points *MinPts*) in OPTICS. The proposed

strategy is capable of detecting single sensor fault and multiple sensor faults in AHUs. In practice, the proposed method can be used as a standalone strategy for AHU sensor fault detection by analysing the raw data collected from building management systems and the results can be used to support the commissioning of AHU sensors.

It is worthwhile to mention that the threshold of the reachability-distance to identify the clusters in the proposed strategy should be carefully selected. If the threshold is not appropriately selected, some faults might not be detected. To make sure all possible faults can be detected, small valleys in the reachability-distance of the cluster order of the data points should be checked. The proposed sensor fault detection strategy might not be able to detect insignificant sensor faults if multiple faults occur in AHUs. The effect of mechanical faults on the effectiveness of the proposed strategy for AHU sensor fault detection was not considered, and this should be a topic of interest in future studies. It is also worthwhile to mention that a method which can automatically select the reachability-distance threshold and determine the valid clusters is desired, which can allow the proposed strategy to be used for online applications.

References

- [1] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, *Energy and Buildings*, 40 (3) (2008), pp. 394-398, <http://dx.doi.org/10.1016/j.enbuild.2007.03.007>.
- [2] F. Xiao, S. Wang, Progress and methodologies of lifecycle commissioning of HVAC systems to enhance building sustainability, *Renewable and Sustainable Energy Reviews*, 13 (5) (2009), pp. 1144-1149, <http://dx.doi.org/10.1016/j.rser.2008.03.006>.
- [3] E. Mills, N. Bourassa, M.A. Piette, H. Friedman, T. Haasl, T. Powell, D. Claridge, The cost-effectiveness of commissioning new and existing commercial buildings: Lessons from

478 224 buildings, Proceedings of the 2005 National Conference on Building Commissioning,
 479 2005, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.396.4806>.

480 [4] S. Katipamula, M.R. Brambley, Methods for fault detection, diagnostics, and prognostics
 481 for building systems—a review, Part I, HVAC&R Research, 11 (1) (2005), pp. 3-25,
 482 <http://dx.doi.org/10.1080/10789669.2005.10391123>.

483 [5] S. Katipamula, M.R. Brambley, Methods for Fault Detection, Diagnostics, and
 484 Prognostics for Building Systems—A Review, Part II, HVAC&R Research, 11 (2) (2005), pp.
 485 169-187, <http://dx.doi.org/10.1080/10789669.2005.10391133>.

486 [6] Z. Ma, S. Wang, Fault-tolerant supervisory control of building condenser cooling water
 487 systems for energy efficiency, HVAC&R Research, 18 (1-2) (2011), pp. 126-146,
 488 <http://dx.doi.org/10.1080/10789669.2011.568320>.

489 [7] Z. Ma, S. Wang, Online fault detection and robust control of condenser cooling water
 490 systems in building central chiller plants, Energy and Buildings, 43 (1) (2011), pp. 153-165,
 491 <http://dx.doi.org/10.1016/j.enbuild.2010.09.003>.

492 [8] Z. Du, B. Fan, J. Chi, X. Jin, Sensor fault detection and its efficiency analysis in air
 493 handling unit using the combined neural networks, Energy and Buildings, 72 (2014), pp. 157-
 494 166, <http://dx.doi.org/10.1016/j.enbuild.2013.12.038>.

495 [9] Z. Hou, Z. Lian, Y. Yao, X. Yuan, Data mining based sensor fault diagnosis and
 496 validation for building air conditioning system, Energy Conversion and Management, 47 (15–
 497 16) (2006), pp. 2479-2490, <http://dx.doi.org/10.1016/j.enconman.2005.11.010>.

498 [10] Z. Du, X. Jin, Y. Yang, Fault diagnosis for temperature, flow rate and pressure sensors in
 499 VAV systems using wavelet neural network, Applied Energy, 86 (9) (2009), pp. 1624-1631,
 500 <http://dx.doi.org/10.1016/j.apenergy.2009.01.015>.

501 [11] S. Wang, F. Xiao, AHU sensor fault diagnosis using principal component analysis
 502 method, *Energy and Buildings*, 36 (2) (2004), pp. 147-160,
 503 <http://dx.doi.org/10.1016/j.enbuild.2003.10.002>.

504 [12] H. Han, B. Gu, J. Kang, Z. Li, Study on a hybrid SVM model for chiller FDD
 505 applications, *Applied Thermal Engineering*, 31 (4) (2011), pp. 582-592,
 506 <http://dx.doi.org/10.1016/j.applthermaleng.2010.10.021>.

507 [13] Y. Zhao, F. Xiao, J. Wen, Y. Lu, S. Wang, A robust pattern recognition-based fault
 508 detection and diagnosis (FDD) method for chillers, *HVAC&R Research*, 20 (7) (2014), pp.
 509 798-809, <http://dx.doi.org/10.1080/10789669.2014.938006>.

510 [14] Y. Zhao, S. Wang, F. Xiao, A statistical fault detection and diagnosis method for
 511 centrifugal chillers based on exponentially-weighted moving average control charts and
 512 support vector regression, *Applied Thermal Engineering*, 51 (1–2) (2013), pp. 560-572,
 513 <http://dx.doi.org/10.1016/j.applthermaleng.2012.09.030>.

514 [15] J. Cui, A robust fault detection and diagnosis strategy for centrifugal chillers,
 515 [dissertation], Department of Building Services Engineering, The Hong Kong Polytechnic
 516 University, 2005.

517 [16] B. Fan, Z. Du, X. Jin, X. Yang, Y. Guo, A hybrid FDD strategy for local system of AHU
 518 based on artificial neural network and wavelet analysis, *Building and Environment*, 45 (12)
 519 (2010), pp. 2698-2708, <http://dx.doi.org/10.1016/j.buildenv.2010.05.031>.

520 [17] S. Li, J. Wen, Application of pattern matching method for detecting faults in air handling
 521 unit system, *Automation in Construction*, 43 (2014), pp. 49-58,
 522 <http://dx.doi.org/10.1016/j.autcon.2014.03.002>.

523 [18] H. Yoshida, S. Kumar, Development of ARX model based off-line FDD technique for
 524 energy efficient buildings, *Renewable Energy*, 22 (1–3) (2001), pp. 53-59,
 525 [http://dx.doi.org/10.1016/S0960-1481\(00\)00033-1](http://dx.doi.org/10.1016/S0960-1481(00)00033-1).

526 [19] T. Mulumba, A. Afshari, K. Yan, W. Shen, L.K. Norford, Robust model-based fault
527 diagnosis for air handling units, *Energy and Buildings*, 86 (2015), pp. 698-707,
528 <http://dx.doi.org/10.1016/j.enbuild.2014.10.069>.

529 [20] K. Bruton, P. Raftery, P. O'Donovan, N. Aughney, M.M. Keane, D.T.J. O'Sullivan,
530 Development and alpha testing of a cloud based automated fault detection and diagnosis tool
531 for Air Handling Units, *Automation in Construction*, 39 (2014), pp. 70-83,
532 <http://dx.doi.org/10.1016/j.autcon.2013.12.006>.

533 [21] Z. Yu, F. Haghghat, B.C.M. Fung, L. Zhou, A novel methodology for knowledge
534 discovery through mining associations between building operational data, *Energy and*
535 *Buildings*, 47 (2012), pp. 430-440, <http://dx.doi.org/10.1016/j.enbuild.2011.12.018>.

536 [22] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques
537 for a cluster of smart office buildings, *Expert Systems with Applications*, 42 (9) (2015), pp.
538 4324-4338, <http://dx.doi.org/10.1016/j.eswa.2015.01.010>.

539 [23] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building
540 automation data and its application in building diagnostics, *Automation in Construction*, 50
541 (2015), pp. 81-90, <http://dx.doi.org/10.1016/j.autcon.2014.12.006>.

542 [24] P.N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, Pearson Addison
543 Wesley, USA, 2006 (ISBN 9780321321367).

544 [25] C.T. Yiakopoulos, K.C. Gryllias, I.A. Antoniadis, Rolling element bearing fault detection
545 in industrial environments based on a K-means clustering approach, *Expert Systems with*
546 *Applications*, 38 (3) (2011), pp. 2888-2911, <http://dx.doi.org/10.1016/j.eswa.2010.08.083>.

547 [26] M.A.K. Jaradat, R. Langari, A hybrid intelligent system for fault detection and sensor
548 fusion, *Applied Soft Computing*, 9 (1) (2009), pp. 415-422,
549 <http://dx.doi.org/10.1016/j.asoc.2008.05.001>.

550 [27] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, R. Seraoui, Clustering for unsupervised
551 fault diagnosis in nuclear turbine shut-down transients, *Mechanical Systems and Signal*
552 *Processing*, 58–59 (2015), pp. 160-178, <http://dx.doi.org/10.1016/j.ymssp.2014.12.018>.

553 [28] M. Ankerst, M.M. Breunig, H.-P. Kriegel, J. Sander, OPTICS: ordering points to identify
554 the clustering structure, *ACM Sigmod Record*, 28 (2) (1999), pp. 49-60,
555 <http://dx.doi.org/10.1145/304182.304187>.

556 [29] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, Elsevier Science,
557 USA, 2006 (ISBN 9780080475585).

558 [30] K. Potter, H. Hagen, A. Kerren, P. Dannenmann, *Methods for presenting statistical*
559 *information: The box plot*, *Visualization of Large and Unstructured Data Sets*, 4 (2006), pp.
560 97-106, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.66>.

561 [31] Rapidminer[Software], Rapid-I, 2015 (Available from: <https://rapidminer.com/>).

562 [32] R Development Core Team, *R: A language and environment for statistical*
563 *computing*[Software], R Foundation for Statistical Computing, Vienna, Austria, 2008
564 (Available from: <https://www.r-project.org/>).

565 [33] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York, 2009
566 (ISBN 9780387981413).

567 [34] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering
568 clusters in large spatial databases with noise, *2nd International Conference on Knowledge*
569 *Discovery and Data Mining*, 1996, pp. 226-231,
570 <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.9220>.

571 [35] H.-P. Kriegel, P. Kröger, J. Sander, A. Zimek, *Density-based clustering*, Wiley
572 *Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1 (3) (2011), pp. 231-240,
573 <http://dx.doi.org/10.1002/widm.30>.

574 [36] W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, shiny: Web Application
575 Framework for R[Software], 2015 (Available from: <http://shiny.rstudio.com/>).

576 [37] R.W. Sembiring, J.M. Zain, A. Embong, Dimension Reduction of Health Data Clustering,
577 International Journal on New Computer Architectures and Their Applications, 1 (4) (2011),
578 pp. 1041-1050, <http://arxiv.org/abs/1110.3569>.

579 [38] S. Alelyani, J. Tang, H. Liu, Feature Selection for Clustering A Review, in: Charu C.
580 Aggarwal, C.K. Reddy (Eds.) Data Clustering: Algorithms and Applications, Vol. 29, CRC
581 Press, 2013 (ISBN 9781466558212).

582 [39] I.T. Jolliffe, Principal component analysis, Springer-Verlag, New York, 1986 (ISBN
583 9780387962696).

584 [40] X. Xu, F. Xiao, S. Wang, Enhanced chiller sensor fault detection, diagnosis and
585 estimation using wavelet analysis and principal component analysis methods, Applied
586 Thermal Engineering, 28 (2) (2008), pp. 226-237,
587 <http://dx.doi.org/10.1016/j.applthermaleng.2007.03.021>.

588 [41] Trnsys - a transient system simulation program[Software], University of Wisconsin--
589 Madison, 2015 (Available from: <http://sel.me.wisc.edu/trnsys/>).

590

591

592

Table 1 Summary of different test cases

Case no	Sensor fault(s)	Fault severity (relative percentage)	Occurring time (hour)*	Duration of fault lasted (hours)	Purpose
1	Supply air temperature	-2.0K (14.3%)**	1421 st	679	Detailed demonstration of sensor fault detection process
2	Fault-free	-	-	-	
3	Supply air temperature	-0.5K (3.6%)**	707 th	1393	
4	Return water temperature	-0.5K (4.2%)**	271 st	1829	Validate the effectiveness of the proposed strategy for single sensor fault detection
5	Supply water temperature	-0.5K (7.1%)**	1421 st	679	
6	Air flow rate	-0.583m ³ /s (10%)*	1254 th	846	
7	Water flow rate	-0.254L/s (10%)*	896 th	1204	
8	Supply air temperature/return water temperature	-0.5K/-0.5K	568 th / 1615 th	1532/ 485	Validate the effectiveness of the proposed strategy for multiple sensor faults detection
9	Supply air temperature/return water temperature	-1.0K/-1.0K (7.1%/8.3%)*	568 th / 1615 th	1532/ 485	
10	Return water temperature/ water flow rate/supply air temperature	-1.0K/ -0.254L/s/ -1.0K	584 th / 774 th / 1810 th	1516/ 1326/ 290	

* The fault was randomly introduced.

** The percentages indicate the proportion of the bias to the typical measured values (14°C, 7°C and 12°C for supply air temperature, supply chilled water temperature and return chilled water temperature, respectively)

***The percentages indicate the proportion of the bias to the maximum flow rate during the simulated period.

Table 2 Summary of the sensitivity analysis results

Case	<i>Eps</i>	<i>MinPts</i>	Detected Fault occurring time (hour)
Original	1.2	15	1421.17
a	1.7	15	1421.33
b	0.7	15	1421.17
c	1.2	10	1421.17
d	1.2	20	1421.17

Figure Captions

Fig. 1 Outline of the proposed fault detection strategy

Fig. 2 An illustration of a box plot.

Fig. 3 Illustration of the order of the data points in the k-distance graph

Fig. 4 Illustration of how the three types of distances are defined. c-d and r-d are abbreviations core-distance and reachability-distance. The point p3 is directly density reachable from the point o [28].

Fig. 5 Illustration of the OPTICS outputs and how clusters are identified with the user-defined threshold

Fig. 6 User interface for the two thresholds selection. Fig. 7 Illustration of the measurement instruments of AHUs

Fig. 8 Temporal spread of the post-filtered observations (the system started to operate at the 53 hour) – Case 1

Fig. 9 k-distance of the data points remained for fault detection – Case 1

Fig. 10 The reachability-distance of the cluster-order of the data point - Case 1

Fig. 11 Temporal distribution of the data points in clusters in Case 1 under (a) Threshold 1; (b) Threshold 2

Fig. 12 The reachability-distance of the cluster-order of the data point – Case 2 with fault-free operation

Fig. 13 Temporal distribution of the data points in the identified clusters – Case 2

Fig. 14 Temporal distribution of the data points in the identified clusters with single sensor fault – Cases 3-7

Fig. 15 The reachability-distance of the cluster-order of the data points - Case 8

Fig. 16 Temporal distribution of the data points in clusters – Case 8

Fig. 17 The reachability-distance of the cluster-order of the data points - Case 9

Fig. 18 The temporal distribution of the data points in the identified clusters - Case 9

Fig. 19 The reachability-distance of the cluster-order of the data points – Case 10

Fig. 20 Temporal distribution of the data points in the identified clusters - Case 10.

Fig. 21 The reachability-distance of the cluster-order of the data points of different input parameter combinations and the corresponding temporal distribution of the identified clusters.

Fig. 1 Outline of the proposed fault detection strategy.

Fig. 2 An illustration of a box plot.

Fig. 3 Illustration of the order of the data points in the k-distance graph.

Fig. 4 Illustration of how the three types of distances are defined, where c-d and r-d are abbreviations of core-distance and reachability-distance. The point p_3 is directly density reachable from the point o [28].

Fig. 5 Illustration of the OPTICS outputs and how clusters are identified with the user-defined threshold.

Fig. 6 User interface for the two thresholds selection.

654

655

656

657

Fig. 7 Illustration of the measurement instruments of AHUs.

658

659

(a) Distribution of the remaining data points

(b) Density profile of the remaining data points

Fig. 8 Temporal spread of the post-filtered observations (the system started to operate at the 53 hour) – Case 1.

665

Fig. 9 k-distance of the data points remained for fault detection – Case 1.

666

667

Fig. 10 The reachability-distance of the cluster-order of the data point - Case 1.

(a) Temporal distribution of the clusters under the Threshold 1

(b) Temporal distribution of the clusters under the Threshold 2

668

669

670

Fig. 11 Temporal distribution of the data points in clusters in Case 1 under (a) Threshold 1; (b) Threshold 2.

Fig. 12 The reachability-distance of the cluster-order of the data point – Case 2 with fault-free operation.

Fig. 13 Temporal distribution of the data points in the identified clusters – Case 2.

(a) Biased supply air temperature sensor (case 3)

(b) Biased return water temperature sensor (case 4)

(c) Biased supply water temperature sensor (case 5)

(d) Biased air flow rate meter (case 6)

(e) Biased water flow rate meter (case 7)

Fig. 14 Temporal distribution of the data points in the identified clusters with single sensor fault – Cases 3-7.

Fig. 15 The reachability-distance of the cluster-order of the data points - Case 8.

Fig. 16 Temporal distribution of the data points in clusters – Case 8.

Fig. 17 The reachability-distance of the cluster-order of the data points - Case 9.

(a) Temporal distribution of the clusters under the Threshold 1

(b) Temporal distribution of the clusters under the Threshold 2

Fig. 18 Temporal distribution of the data points in the identified clusters - Case 9.

Fig. 19 The reachability-distance of the cluster-order of the data points – Case 10

(a) Temporal distribution of the clusters under the Threshold 1

(b) Temporal distribution of the clusters under the Threshold 2

(c) Temporal distribution of the clusters under the Threshold 3

Fig. 20 Temporal distribution of the data points in the identified clusters - Case 10.

Fig. 21 The reachability-distance of the cluster-order of the data points of different input parameter combinations and the corresponding temporal distribution of the identified clusters.