

# Unsupervised Learning Techniques for HVAC Terminal Unit Behaviour Analysis

Maitreyee Dey

School of Engineering  
London South Bank University  
London, UK  
deym@lsbu.ac.uk

Manik Gupta

School of Engineering  
London South Bank University  
London, UK  
guptam@lsbu.ac.uk

Mikdam Turkey

School of Engineering  
London South Bank University  
London, UK  
turkeym@lsbu.ac.uk

Sandra Dudley

School of Engineering  
London South Bank University  
London, UK  
dudleym@lsbu.ac.uk

**Abstract**—In the pursuit of improved energy efficiency, older and new buildings are being fitted with Building Energy Management System (BEMS). BEMS can be used to extract valuable building data that can be further analysed to discover problems related to user comfort, building maintenance and energy wastage in buildings. The main focus of this paper is to demonstrate an effective method to remotely analyse and categorise the different Heating, Ventilation and Air-Conditioning (HVAC) Terminal Unit (TU) behaviours using BEMS data. Using a data-driven, unsupervised learning strategy to identify anomalous behaviours enabling notifications to the building manager regarding faulty TUs can go a long way in providing energy savings and improving building performance.

A novel feature extraction method based on event discovery from TU data is proposed and applied to multidimensional data streams retrieved from a building based in the city of London. Further, X-means clustering has been performed over the extracted features to group the different TU behaviours. The clustering results, validated through established statistical methods, successfully yield several distinct TU behaviour patterns in addition to the outliers. The clustering behaviour has been further verified across daily and weekly TUs.

**Index Terms**—Heating, Ventilation and Air-Conditioning (HVAC), Building Energy Management System (BEMS), Terminal Unit (TU), Feature Extraction, X-means Clustering, Fault Detection and Diagnosis (FDD).

## I. INTRODUCTION

Buildings are becoming more and more complex energy consuming systems consisting of several elements for example: HVAC, lighting, effort and control systems etc. Monitoring a building's energy consumption and concern over our environment has been the focus in recent years. Equipment failure and performance degradation of HVAC systems in commercial buildings often goes unnoticed until it negatively impacts occupant comfort, triggers an equipment-level alarm, reduces equipment lifespan or results in excessive energy consumption. Therefore buildings are installed with BEMS to pre-empt or promptly react to such issues via a dedicated staff member. BEMS has been defined as an IT-based solution proposed to extend the capabilities of sensing, control, and automation hardware to direct automated and/or manual improvements and system operations utilizing the data from multiple data streams.

The presented work focuses specifically on analysing the real BEMS data to discover anomalies that potentially lead to

high-energy consumption and poor performance in buildings. Existing building stock is specifically being retrofitted with BEMS and it is challenging to study the large amount of data being generated rapidly by different plant components. Therefore, dedicated research investigation is being carried out to analyse data retrieved from BEMS systems using big data and machine learning approaches.

The main focus of the presented research is on the Fault Detection and Diagnosis (FDD) of specific HVAC subcomponents known as TU. The primary goal of this research is to develop novel approaches based on statistical techniques and machine learning to understand different TU behaviours and identify faulty TUs. The algorithms employ a data driven approach harnessing the hidden information buried in the vast amount of historical building data. One of the expected outcomes of this research is to detect and/or predict faults remotely and provide real time basis for maintaining building's occupant comfort. This is done through automatic notifications generation on faulty behaviour of TUs, which will then in turn help building managers take appropriate corrective measures.

Terminal Unit, a specific subcomponent of HVAC system found in residential, commercial, and industrial buildings, is the research focus as it is responsible for the final delivery of comfort inside built environments. A TU is a common and simple device consisting of a heating and/or cooling heat exchanger or 'coil' and fan. It is generally ceiling-mounted and usually controlled by local thermostats. They may either primarily recirculate internal air, or can introduce proportion of 'fresh' air with the re-circulated air. Usually inside buildings, there is a central chiller plant that distributes cold water to all the cooling coils, and a central boiler that pumps out hot water to all the heating coils. The fan is operational at all the times. When the environment becomes too warm, the thermostat senses and signals the chilled water valve, and cold water is subsequently passed through the coil, thus extracting the heat from the air being blown by the fan. If it gets too cold depending on the local set point, the heating coil begins working in the same way. Fig. 1 shows the schematic of a TU.

Poorly controlled or faulty TUs such as, fan coil units can be responsible for significant energy wastage and user discomfort. For example, a faulty fan coil unit can signal a false heating demand to the boiler, causing the boiler and ancillary equipment to activate and begin distributing hot water, leading to energy

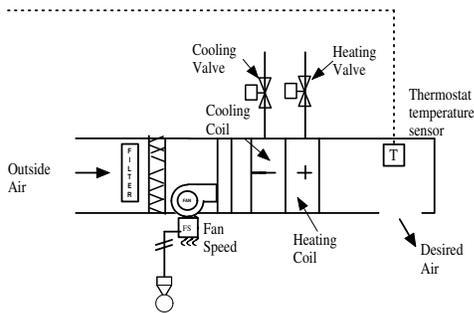


Fig. 1: TU schematic diagram

wastage and user discomfort. There are a multitude of issues that can lead to faulty TU behaviour and need expert building engineer knowledge to identify each one of these issues. The manual TU data investigation can be extremely tedious and impossible with the ever-increasing amount of building data. Hence this process needs to be automated using data mining and machine learning approaches.

### A. Background

FDD in HVAC systems has evolved tremendously and additionally, significant development has been made in terms of data mining techniques. Active research and exploration in this area began in the 1980s [1], [2], but from the onset, there have been practical limitations such as scalability and complexity issues related to HVAC systems that have made FDD extremely challenging. Based on our literature survey, the FDD techniques have been classified into three categories: model based, rule based and data-driven based approaches.

The data generated from plants and industries is considered in model based approaches. Shaw et.al. [3] further evolved a modified physical model to reduce the computational cost. In [4], [5], the model proposed by Shaw et al. has been augmented to include real time HVAC application.

Rule-based or Knowledge-based perspective for FDD in HVAC systems is applicable when there is insufficient data available for building the mathematical models. This type of method is applicable for large scale, complex HVAC system, and requires a large amount of expert knowledge of the faults of such systems [6]. Expert knowledge of a system can be implemented using some rules and can generally be divided into three types: (a) system with partial knowledge by framing IF-THEN rules, (b) system with deep-expert knowledge involving functional reasoning, and (c) system with machine learning capabilities. There are three different techniques: casual analysis, pattern classification, and expert systems that have been already developed for Air Handling Unit (AHU).

Signed Directed Graph (SDG) has been proposed for casual analysis by employing the fault symptoms without first principles [7]. Schein et. al., [8] developed a FDD system for operational control with partial or limited knowledge by setting up IF-THEN rules for an AHU. Schein and Bush by [9] expanded the concept by proposing hierarchical rules for FDD. Yet, expert knowledge based techniques are limited

when real data is unavailable. Pattern classification based algorithms can be used to build the non-linear correlations between data patterns and fault classes in the absence of clear model structures. Some popular pattern classification based methods are Bayes Classifier [10], Artificial Neural Networks (ANN) [11], Support Vector Machine (SVM) Classifier [12], Fuzzy Logic [13].

Data-driven based techniques build relationships between data patterns and fault classes for a system [14]. These approaches extract the key data components and transform the dimension of the entire data. Then, these key components are used instead of the whole dataset for FDD. This approach is fit for modern HVAC systems being used in huge commercial buildings. There are two types of categories for this approach: Signal based FDD and Multi Variable Statistics (MVS) based FDD. Some existing signal based procedures are wavelet transformation and short-time Fourier analysis [15]. A combined method using wavelet transformation and PCA is proposed to diagnose faults for AHU system in [16]. It detects faults in large systems using dimensionality reduction technique that maps the data to a lower dimensional space.

All of these above stated approaches are limited to particular fault categories (e.g. fan failure, valve stuck) or fixed data. However, in the recent times due to the installation of a huge number of sensors, there is a growing amount of data that needs to be stored and processed. The BEMs are trying to solve this problem and in that context, the BEMS data analysis is fairly novel. The data gathered from the terminal units inside building amass a wealth of key hidden information related to building behaviour and in fact, the characterisation of the TU behavioural patterns using real BEMS data has not been given much attention in the current research, which is indeed the focus of this current work.

The outline of the remaining paper is as follows. Section 2 focuses on the proposed methodology and provides details about the feature extraction technique proposed as a part of this work. Section 3 gives the experimental result analysis and discussion of the different TU behaviour patterns obtained as a result of the clustering. Section 4 concludes the paper and provides the future research directions of this work.

## II. PROPOSED METHODOLOGY

A novel feature extraction technique is proposed and experimentally evaluated in this paper in order to project the high dimensional TU data into a lower dimensional space followed by an unsupervised learning algorithm to characterize different TU behaviours. Fig. 2 shows the different steps involved as a part of this work: raw data gathering from buildings, data pre-processing, feature extraction, feature clustering and a detailed behaviour analysis to produce final insights on TU behaviours. The following subsections describe each of these steps on more details.

### A. TU Data Selection and Pre-processing

The present case study is based on a building located in the city of London. The building has 17 floors and 731 terminal



Fig. 2: Steps involved in proposed methodology

units spread across the different floors. The data collected from a terminal unit is shown in Fig. 3 and various data streams that are examined for a single terminal unit are as follows:

- Control temperature [ $^{\circ}C$ ] (reported space temperature), frequently measured by each terminal unit or in some cases a zone space temperature is used.
- Set point temperature and deadband [ $^{\circ}C$ ].
- Heating and cooling (power/effort) valve or damper actuator control and feedback signals.
- Enabled signal to indicate the hours of operation.

For a TU, each datum is sampled at a 30-minute time interval have been extracted from the BEMS data and used for further investigation.

Data pre-processing plays a vital role in selecting the appropriate data from the raw data. The pre-processing helps deal with problems like different sampling rate, missing values etc. Therefore, the TUs with the missing data are filtered out in the analysis phase and linear interpolation has been used to resample data at every 10-minute time interval. For further analysis, the proposed feature extraction has been implemented using the multiple data streams.

### B. Feature Extraction Using TU Data

The novel feature extraction method generates events (E) which is divided into three different stages as described in the following subsections:

- 1) Event Discovery
- 2) Event Area Calculation
- 3) Event Aggregation

1) *Event Discovery*: In the event discovery stage, the events are marked using pre-processed temperature and effort data streams. On a given day, when the heating and cooling units inside a building are enabled during the operational hours, the temperature starts changing depending upon the environmental demand. Now depending on the temperature variations, the data stream is sliced into different time periods and used to designate different events. Whenever the temperature values change with respect to the set point value, an event is considered to happen.

Four different types of events are identified in this event characterization stage as shown in the Fig. 3.

- Event Start (ES) is assumed to occur when BMS starts on a given day (enabled signal gets switched on) and the time instant when the temperature starts to change.
- Due to the process variable delay during the dead time, the temperature starts to respond only after a certain delay from the previous point when the BMS start happens and this event is termed as Response Delay (RD).

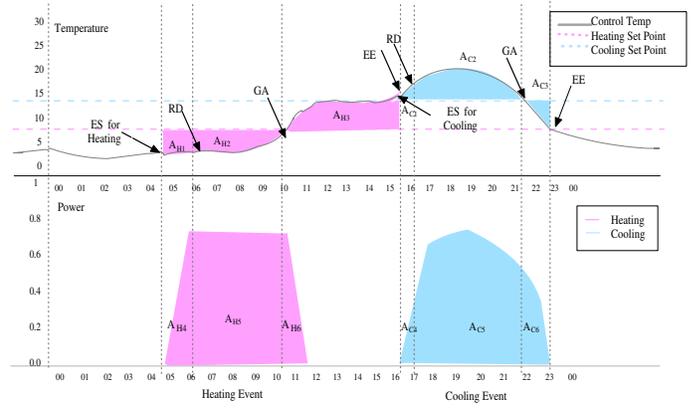


Fig. 3: Event discovery process for daily TU data

- When the temperature reaches to the desired set point, it is assumed as Goal Achieved (GA). GA can be considered as the time instant when the control temperature reaches the steady-state, or final value.
- The temperature may either continue to be within the dead band till it exceeds the dead band, and then it is assumed to be an Event End (EE).

2) *Event Area Calculation*: Once the appropriate heating and cooling events have been discovered, next the approximate area under the curve for each event needs to be calculated. Depending on the event type, whether it is a heating (H) or a cooling (C) event, the area calculations respectively for both the temperature and effort curves are carried out. So, in effect, six areas (three from temperature and three from effort curve) are calculated for each heating event and similarly, six areas are calculated for each cooling event. Finally, twelve different areas are derived using the daily data for every TU. Eq. (1) shows the area ( $A_E$ ) computation under the curve  $f(x)$  at every time interval  $\Delta x$ .

$$A_E = \sum_{i=0}^n f(x_i) \Delta x \quad (1)$$

For a heating event, the calculated areas for temperature (T) are denoted by  $A_{H1} - A_{H3}$  and the areas for effort (P) are denoted by  $A_{H4} - A_{H6}$ . Similarly for a cooling event, the calculated areas for temperature (T) are denoted by  $A_{C1} - A_{C3}$  and the areas for effort (P) are denoted by  $A_{C4} - A_{C6}$ . Once these areas have been computed, they are normalized to get the features as  $F_{H1} - F_{H6}$  and  $F_{C1} - F_{C6}$  as represented by (2) - (5).

Eq. (2) and (3) show the area calculations for a heating event.

$$F_{H1} = \frac{A_{H1}}{T_{H1}}, \quad F_{H2} = \frac{A_{H2}}{T_{H1}}, \quad F_{H3} = \frac{A_{H3}}{T_{H2}} \quad (2)$$

where,  $T_{H1} = \max(A_{H1} + A_{H2})$   
and,  $T_{H2} = \max(A_{H3})$

$$F_{H_4} = \frac{A_{H_4}}{P_{H_1}}, \quad F_{H_5} = \frac{A_{H_5}}{P_{H_1}}, \quad F_{H_6} = \frac{A_{H_6}}{P_{H_2}} \quad (3)$$

where,  $P_{H_1} = \max(A_{H_4} + A_{H_5})$   
and,  $P_{H_2} = \max(A_{H_6})$

Eq. (4) and (5) show the area calculations for a cooling event.

$$F_{C_1} = \frac{A_{C_1}}{T_{C_1}}, \quad F_{C_2} = \frac{A_{C_2}}{T_{C_1}}, \quad F_{C_3} = \frac{A_{C_3}}{T_{C_2}} \quad (4)$$

where,  $T_{C_1} = \max(A_{C_1} + A_{C_2})$   
and,  $T_{C_2} = \max(A_{C_3})$

$$F_{C_4} = \frac{A_{C_4}}{P_{C_1}}, \quad F_{C_5} = \frac{A_{C_5}}{P_{C_1}}, \quad F_{C_6} = \frac{A_{C_6}}{P_{C_2}} \quad (5)$$

where,  $P_{C_1} = \max(A_{C_4} + A_{C_5})$   
and,  $P_{C_2} = \max(A_{C_6})$

3) *Event Aggregation*: Event Aggregation is the next step in the feature extraction process since multiple heating and cooling events can occur during the duration of a single day. Therefore, all the events of a given type need to be aggregated to represent the averaged values.

The final aggregated features can be represented using equations where,  $k$  denotes the event number and  $n$  denotes total no of occurrences for event of each type. Thus, the daily data for a TU can be represented using twelve features representing both the heating and cooling events as shown in (6) and (7).

$$F_{H_k} = \frac{1}{n} \sum_{i=1}^n (F_{H_{k_i}}) \quad (6)$$

$$F_{C_k} = \frac{1}{n} \sum_{i=1}^n (F_{C_{k_i}}) \quad (7)$$

### C. X-means Clustering

Clustering is carried out on the twelve extracted features of the TU daily data. This is performed in order to discover a set of TU behaviours into meaningful groups that are similar to each other. This unsupervised technique is applied since there is very limited knowledge about the datasets.

Here, an extended  $K$ -Means ( $X$ -means) [17] is used to avoid the constraint of conventional  $K$ -means clustering, which tries to automatically determine the number of clusters based on Bayesian Information Criterion (BIC) scores.  $X$ -means consists of the following two operations repeated until completion:

- 1) *Improve-params* - This step consists of running the conventional  $K$ -means to convergence. The average square distance is used as a similarity measure to determine cluster membership at each iteration of the  $K$ -Means.
- 2) *Improve-structure* - This step consists of picking one centroid and producing a new centroid nearby, run  $K$ -means to completion and see if the resulting model

scores better. If it does, new centroid is accepted, else the previous structure is retained.  $X$ -means makes local decisions about which subset of the current centroids should split themselves in order to provide a better fitting of the data. The splitting decision is determined by computing the BIC criterion.

- 3) If  $K > K_{max}$  stop and report the best scoring model found during the search, else goto 1.

### D. Internal Evaluation of Clustering

To assess an unsupervised learning technique where the external benchmarks or known class labels are not present, the use of internal validation criteria is appropriate. The internal criteria are built using statistical methods to evaluate the effectiveness of a clustering algorithm. This evaluation measures the degree of intra-cluster cohesion and inter-cluster separation. Here, Silhouette (SI) criterion has been used for the evaluation of the clustering technique. The silhouette indexing is measured by the similarity of each point with other points in its own cluster, when compared to points in other clusters. The silhouette value is defined in (8). Here  $a_i$  is the average dissimilarity of a data point  $i$  with all other data points within the same cluster, and  $b_i$  is the lowest average dissimilarity of  $i$  to any other cluster, of which  $i$  is not a member.

$$SI = \frac{1}{nk} \sum_{i \in k} \frac{b_i - a_i}{\max(a_i, b_i)} \quad (8)$$

The silhouette value ranges from -1 to +1. A high value indicates that  $i$  is well-matched to its own cluster. The clustering solution is considered appropriate if most points have a high silhouette value [18].

## III. RESULTS

The results as obtained after performing the proposed feature extraction and  $X$ -means clustering on the pre-processed TU data are presented. Given that the main motive of this research is to identify the significant TU patterns and the faulty TUs so that appropriate action can be taken and subsequently maintenance of the essential building services can be carried out, an explanation of the TU behaviour patterns and outlier analysis both are given in details in the subsections below.

### A. Experimental Data

The experiments have been conducted on daily and weekly selected data for the month of July, 2015 for all the HVAC TUs located inside the building chosen in the city of London. The day analysis has been performed on 17th July 2015 TUs while the weekly analysis has been performed from 17th July to 23rd July for 5 days since weekends have been excluded from the analysis shown in Table I. Operating TUs represent the TUs after data pre-processing and filtering TUs with missing data.

No of Days	Total no of TUs	Operating TUs	Description
1 Day	731	723	17th July, 2015
1 Week	3655	3615	17th-23rd July, 2015

TABLE I: Experimental data details

### B. Cluster Analysis

In this section, the clustering results obtained after performing X-means clustering algorithm are presented. Based on BIC for the best model selection, the number of clusters were varied from 2 to 10 and 6 has been obtained as the number of optimal clusters to describe the different TU behaviours. Each cluster captures discrete TU behaviour depending on the extracted feature data.

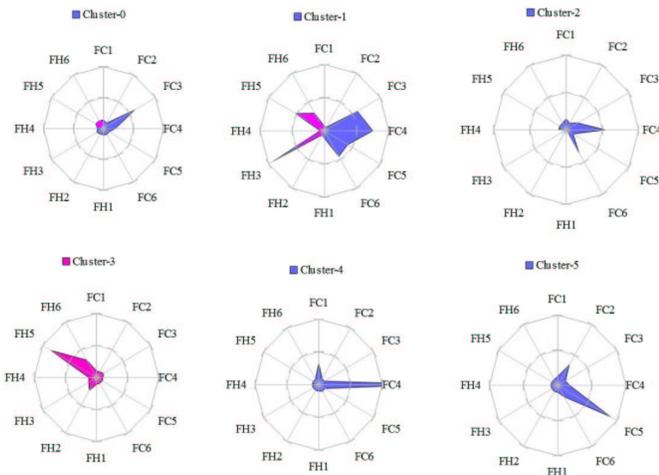


Fig. 4: Cluster wise feature distribution for daily TU analysis

Fig. 4 shows cluster wise distribution of the twelve features with respect to both temperature and corresponding effort for the daily TU data. Each radar graph represents an individual cluster and each of its axes represents a separate feature. Each graph represents a particular TU behaviour based on the variations of the extracted features. The six cooling features ( $F_{C_1} - F_{C_6}$ ) and six heating features ( $F_{H_1} - F_{H_6}$ ) are plotted in blue and in pink colour respectively.

It can be observed that cluster 2, 4 and 5 shows cooling, cluster 3 shows heating and clusters 0 and 1 shows both heating and cooling trend. It is to be further noted that more area under RD to EE area is acceptable than the area under ES-RD because this implies that TU achieves the set point (goal) and spends more time within the dead band. When the area under ES-RD is more, it denotes that the temperature is not within the dead band or takes a long time to reach the set point.

For example: cluster C1 captures the TU behaviour where the areas are mostly in GA-EE state and the TU makes adequate effort to achieve the set point and therefore has more area under ES-RD for the effort curve. This behaviour pattern is indicative of well behaving cooling TU patterns. This good

behaviour holds for cluster C0 as well given that most of them too achieve the goal, given that there is more area under RD-GA and they have more area under ES-RD for the effort curve. Whereas in case of cluster C3, the area under RD-GA is more which indicates higher heating effort are required, though the goal is achieved, hence indicating good behaviour. Cluster C5 is similar to cluster C3 in terms of more areas under the RD-GA, but captures the cooling TU behaviour and also the associated effort levels are higher. Cluster C2 is good with respect to temperature behaviour because of more area under GA-EE and more area under the ES-RD for the effort curve, it implies that the TU exerts initial effort to achieve the goal. In case of cluster C4, more time is spent in the ES-RD state that indicates that the TUs take longer to reach the set point and hence struggle to achieve the goal. Thus, it is seen that based on the cluster analysis, distinct TU patterns can be captured and distinguished from one another.

The above experiment is further conducted on weekly TU data to capture the behaviour patterns. Table II shows the summary of the clustering behaviour across the different experiments with both daily and weekly selected TUs. It has been observed that the clustering yields similar TU behaviour patterns across these experiments too.

Cluster No	Day	Cluster No	Week
	No of TU		No of TU
0	41	0	218
1	481	2	2142
2	60	5	359
3	58	1	254
4	42	3	259
5	41	4	383

TABLE II: Number of TUs per cluster and mapping of cluster behaviours for both daily and weekly TU analysis

### C. Cluster Validation

In order to validate the clustering results, internal cluster evaluation criteria has been used since the method is completely unsupervised and there is no prior label information available for the data. SI has been used to validate the clustering results. Fig. 5 shows the summarized cluster validation results across daily and weekly selected TUs. A high SI index value indicates good clustering. It can thus be seen that the clustering results are appropriate since highest index values are achieved at the number of selected clusters i.e. 6.

### D. Outlier Analysis

The purpose of outlier analysis is to detect the faulty TUs from the well behaving TUs within each cluster. Therefore, based on the intra cluster distance measurements obtained from the clustering analysis, it has been experimentally tested that when the distance value of a data point from the centre of a cluster is greater than a particular threshold ( $2 \times \text{standarddeviation} + \text{mean}$ ), it is considered as a faulty TU.

This experiment has been done for daily and weekly, selected TU's, though the results are presented only for the

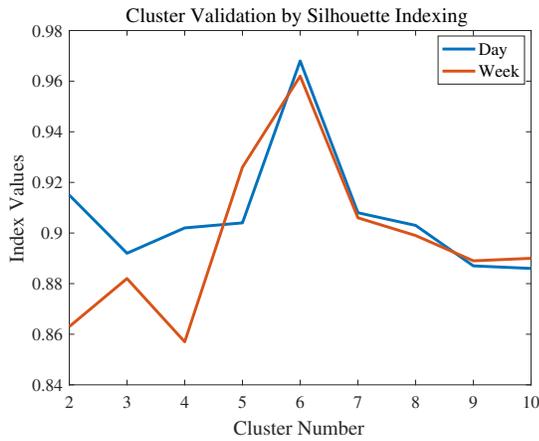


Fig. 5: Cluster validation using SI

day TU data. Fig. 6 shows the data plots for different clusters, where the blue dots represent the normal behaving TUs, whereas the red ones depict the outliers for a particular cluster. These outliers have been further verified with the help of expert building engineers as fault occurrences across the HVAC TU's.

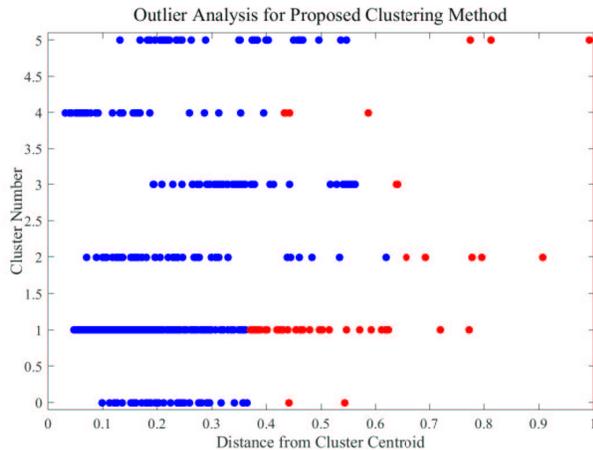


Fig. 6: Outlier Analysis for different clusters

#### IV. CONCLUSIONS

In this paper, a novel feature extraction technique has been proposed to generate twelve features from a huge amount of TU data. These features provide a good approximation for the HVAC TU data. Further on application of unsupervised learning technique, a number of different cluster patterns have been identified that help in the identification of different TU behaviours and can be used to further identify the faulty and abnormal TU's. Currently the work is limited to a particular type of TU's (Fan coil units); but it is going to be extended to study different types of terminal units like VAVs and chilled beams etc. Additionally based on the obtained clustering behaviours, categorical label assignment is going to be done to build an automated TU classification system.

#### V. ACKNOWLEDGEMENTS

This work has been carried out as a part of the Innovate UK funded research project "Energy Management & Analysis Exploiting Existing BMS Infrastructure & Data" and the authors are grateful to the Demand Logic team for providing us with the BEMS data and invaluable information on building services and related performance issues.

#### REFERENCES

- [1] P. Usoro, I. Schick, and S. Negahdaripour, "An innovation-based methodology for hvac system fault detection," *Journal of dynamic systems, measurement, and control*, vol. 107, no. 4, pp. 284–289, 1985.
- [2] D. Anderson, L. Graves, W. Reinert, J. Kreider, J. Dow, and H. Wubbena, *A quasi-real-time expert system for commercial building HVAC diagnostics*, Jan 1989, vol. 95, Part 2.
- [3] S. Shaw, L. Norford, D. Luo, and S. Leeb, "Detection and diagnosis of hvac faults via electrical load monitoring," *HVAC&R Research*, vol. 8, no. 1, pp. 13–40, 2002.
- [4] T. Salsbury, "A temperature controller for vav air-handling units based on simplified physical models," *HVAC&R Research*, vol. 4, no. 3, pp. 265–279, 1998.
- [5] T. Salsbury and R. Diamond, "Fault detection in hvac systems using model-based feedforward control," *Energy and Buildings*, vol. 33, no. 4, pp. 403–415, 2001.
- [6] L. H. Chiang, R. D. Braatz, and E. L. Russell, *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media, 2001.
- [7] J. Shiozaki and F. Miyasaka, "A fault diagnosis tool for hvac systems using qualitative reasoning algorithms," in *Proceedings of the Building Simulation*, vol. 99, 1999.
- [8] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "A rule-based fault detection method for air handling units," *Energy and buildings*, vol. 38, no. 12, pp. 1485–1492, 2006.
- [9] J. Schein and S. T. Bushby, "A hierarchical rule-based fault detection and diagnostic method for hvac systems," *HVAC&R Research*, vol. 12, no. 1, pp. 111–125, 2006.
- [10] C.-F. Chien, S.-L. Chen, and Y.-S. Lin, "Using bayesian network for fault location on distribution feeder," *IEEE Transactions on Power Delivery*, vol. 17, no. 3, pp. 785–793, 2002.
- [11] W.-Y. Lee, J. M. House, C. Park, and G. E. Kelly, "Fault diagnosis of an air-handling unit using artificial neural networks," *Transactions American Society of Heating Refrigerating and Air Conditioning Engineers*, vol. 102, pp. 540–549, 1996.
- [12] J. Liang and R. Du, "Model-based fault detection and diagnosis of hvac systems using support vector machine method," *International Journal of refrigeration*, vol. 30, no. 6, pp. 1104–1114, 2007.
- [13] A. L. Dexter and D. Ngo, "Fault diagnosis in air-conditioning systems: a multi-step fuzzy model-based approach," *HVAC&R Research*, vol. 7, no. 1, pp. 83–102, 2001.
- [14] D. Li, G. Hu, and C. J. Spanos, "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis," *Energy and Buildings*, vol. 128, pp. 519–529, 2016.
- [15] R. Isermann, *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer Science & Business Media, 2011.
- [16] Y. Zhao, S. Wang, and F. Xiao, "Pattern recognition-based chillers fault detection method using support vector data description (svdd)," *Applied Energy*, vol. 112, pp. 1041–1048, 2013.
- [17] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.
- [18] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.