

Unsupervised Learning for Detecting VAV Anomalies in Commercial Buildings

Hao Huang, and Craig Roussac, Buildings Alive

ABSTRACT

Machine learning-based fault detection and diagnosis methods have increased in popularity because of their flexibility and effectiveness in handling large datasets. When supervised learning is considered in anomaly detection, one needs to collect extensive historical data that contains both faulty and correct building behavior. However, it usually takes a long time to collect enough historical data to cover a building's range of operating conditions, and the process of labelling the collected data is laborious. This increases both the cost and time taken for deployment.

In this paper, we propose a scalable, unsupervised method for detecting defective variable air volume (VAV) components in commercial buildings. The methodology does not require extensive data history or manual data labelling, but instead exploits the relationships between near-identical VAV terminals within the same building. The proposed algorithm uses isolation forests to provide anomaly scores for multidimensional inputs related to VAVs. The anomalies associated with different inputs are then interpreted to identify root causes. The method has been deployed at a cloud server to perform real-time anomaly detection for hundreds of VAVs in Australian office buildings. The results show that, even with a short historical dataset, the unsupervised algorithm can accurately detect and diagnose control issues and mechanical faults of individual components. It was also found that fixing the identified faulty VAVs resulted in significant energy savings from associated air handling units (AHU) with improved thermal comfort.

Introduction

Fault detection and diagnosis (FDD) has proven to be an efficient tool in detecting faults swiftly and reducing the impact of them in building heating, ventilation and air-conditioning (HVAC) applications. The method of using historical data to detect and diagnose faults has gained more popularity in recent years due to its flexibility in dealing with large, complex datasets. History-based FDD methods are more suitable for complicated systems for which the theoretical system behavior of the model is insufficient to clarify the performance of the system (Kim and Katipamula 2017). While many machine-learning based FDD approaches have been investigated in the past, they can be grouped into three categories: supervised, semi-supervised, and unsupervised learning.

Supervised learning requires anomaly models to be trained using fully labelled data that capture both faulty and normal modes of building operations. The dataset is labeled either to be “faulty” or “normal” before it is fed into a supervised learning algorithm. The unseen data can then be classified based on the model learned from the historical training set. Using supervised learning, FDD classification accuracy can reach as high as 93% for typical AHU faults (Du, Fan, Chi and Jin 2014). Despite their good accuracy, supervised learning models can be hard to deploy in real-world applications because the process of labelling data is laborious, and the faults

associated with Building Management System (BMS) data are hard to validate without domain knowledge. Moreover, because faults do not occur very often, the size of ‘fault samples’ in the training set is usually too small to support the supervised learning method. The imbalanced dataset becomes the major obstacle for the successful deployment of supervised learning anomaly detection (Yan, Huang, Shen, and Ji 2020).

To solve the above-mentioned problem, semi-supervised learning has been proposed (Mirnaghi and Haghighat 2020). This approach also requires anomaly models to be trained based on historical data, but their training data only consists of normal data without any anomaly samples. The data points that deviate the most from the model prediction are identified as anomalies and labelled afterwards. Dey, Rana and Dudley (2018) developed a semi-supervised learning based FDD tool to detect faulty terminal units that have unlabeled and unstructured data. They tested five different machine learning models and found that the performance of a multiclass support vector machine was the best-fitted model for training their datasets. Despite this approach not requiring work on data labeling, it still requires the collection of fault free data, which are not always directly available in the real-world. Huang, Roach and Roussac (2020) presented a semi-unsupervised learning anomaly detection tool to detect building level energy anomalies and a top-down drill down approach to find the faulty equipment associated with the building level anomalies. The authors found that if historical data contains faulty operation, the resulting model would not be able to recognize the faults that occurred during the training period. Moreover, it usually takes at least several months to collect enough data to build a good semi-supervised learning model. This often delays the time at which the anomaly detection algorithm can be put into operation in the field.

Unsupervised learning algorithms do not need data labels, and instead discover hidden patterns or data groupings without the need for human intervention. Even with a limited amount of historical data, unsupervised learning can be deployed on sites quickly because it does not require models to be trained. The use of unsupervised learning has been investigated in the past. Roach (2018) applied several dimensionality reduction algorithms and clustering approaches and used them to detect unusual sensor behavior in multiple buildings. Sipple (2020) developed negative sampling with random forest or neural network classifiers. They demonstrated that the proposed anomaly detection method was capable of performing much better than state-of-the-art anomaly detection approaches. Yan, Huang, Shen and Ji (2020) introduced a framework that utilizes the generative adversarial network (GAN) to address the imbalanced data problem in FDD for air handling units (AHUs). The method can be used to perform FDD even with a limited number of faulty training samples.

The challenges related to fault detection of VAVs have been discussed in the past, but there are still practical research questions yet to be answered. For example, how to identify root causes based on the fault detection results and how to prioritize the anomalies in the correct order so that facility operations teams can resolve them in the most cost- and time-efficient ways. In this paper, we present a case study where an unsupervised anomaly detection method for detecting defective variable air volume (VAV) terminal units was applied in a commercial building complex. The methodology does not require extensive data history or manual data labeling, but instead exploits the relationships between near-identical VAV terminals within the same building. We also score anomalies associated with multidimensional inputs and interpret them with statistical approaches to identify possible root causes. We demonstrate that the proposed method can quickly detect and diagnosis VAV terminal faults with minimum human intervention.

Data Description

We focus on VAV data in this study. VAVs are responsible for maintaining space temperatures and play an important role in maintaining occupant comfort and indoor air quality in commercial buildings. VAVs are also linked to the energy-efficiency control strategies commonly deployed in commercial buildings, such as static pressure reset, zone temperature setpoint reset, and supply air temperature reset strategies. The failure of a single VAV unit might result in inefficient setpoint schedules at the supervisory level, hence causing energy waste. There are usually hundreds or even thousands of near-identical terminal units in large commercial buildings which supply air flows to multiple zones. It can therefore be very time consuming and difficult to manually find VAV faults and prioritize rectification tasks appropriately.

The data used in this study were collected from an office park in Sydney, Australia. The office park is comprised of two office buildings: Buildings A and B. The building's ventilation system consists of twelve identical AHUs, each of them serving one corner of the building (north-east, south-east, south-west and north-west). The AHUs are controlled to supply 150 Pa~200 Pa air flow to the VAV terminals. There are 363 VAV terminal units in the buildings which serve 15 different floors. There are two feedback control loops within the VAV control loop, which are:

- The air flow rate setpoint is reset based on the deviation between space temperature and space temperature setpoint.
- The damper is modulated so that the actual air flow always tracks the air flow setpoint.

For each VAV terminal box, four time-series data streams are collected from the field for the purpose of analytics, including zone temperature (ZNT), zone temperature set-point (ZNTSP), measured air flow rate (FLR) and damper position (DPR). It should be mentioned that different VAV terminal systems may have other measured types, such flow rate setpoint, carbon dioxide level and heating coil percentage. The proposed method can easily incorporate these measure types into the anomaly detection framework once the data are available.

The BMS data are logged at 15-min intervals and uploaded every two hours to a cloud-server. The investigation of each VAV terminal was based on the trend data of a few months. Some example timeseries graphs of the VAV units' operation are illustrated in Figure 1. As we can see, the BMS opens and closes VAV dampers to maintain air flow rate (l/s) at the desired air flow rate setpoint. As the air flow rate increases, the zone temperature starts to drop to track the zone temperature setpoint of 22.5°C. When the VAV unit is working correctly, opening its damper position results in an increase of air flow rate, which reduces the zone temperature in the room during cooling mode. When the rooms are unoccupied, the damper position will be set to the default setting of 100%.

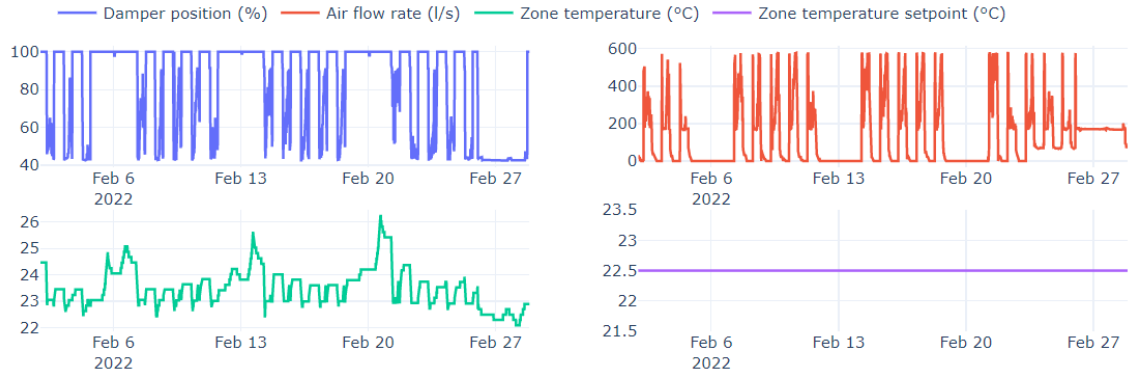


Figure 1. Sample time series data of a VAV terminal unit. Note that the damper position is override to 100% by default when the VAV is at the unoccupied mode.

Methodology

Feature Engineering

As the first step, we engineered timeseries features based off the data mentioned in the previous section. Two approaches were considered when determining what features were most appropriate. The first approach uses a set of statistical measures to describe the daily profile shape, for example, maximum, minimum, mean, standard deviation. The second approach uses the raw readings of each time interval (i.e., the profile itself) as the time series feature. We chose to proceed with the raw reading approach in the end because it was easier to calculate and gave similar performance to the statistical measures approach.

Let $y_{m,u,t}$ be the observed value for input m of terminal unit u at time t . The time series feature of a full day can be displayed as $\mathbf{y}_{m,u} = \{y_{m,u,00:00}, y_{m,u,00:15}, \dots, y_{m,u,23:45}\}$. The feature matrix of a contiguous seven days can then be demonstrated as illustrated in Figure 2.

period	meter_type	measure_type	unit	date	1	...	92	93	94	95	96
0	VAV	DPR	VAV1A-CNW1	2022-02-02	100.000000	...	100.000000	100.000000	100.000000	100.000000	100.000000
1	VAV	DPR	VAV1A-CNW1	2022-02-03	100.000000	...	100.000000	100.000000	100.000000	100.000000	100.000000
2	VAV	DPR	VAV1A-CNW1	2022-02-04	100.000000	...	100.000000	100.000000	100.000000	100.000000	100.000000
3	VAV	DPR	VAV1A-CNW1	2022-02-05	100.000000	...	100.000000	100.000000	100.000000	100.000000	100.000000
4	VAV	DPR	VAV1A-CNW1	2022-02-06	100.000000	...	100.000000	100.000000	100.000000	100.000000	100.000000
...	...	...	...	...	...	...	...	...	...	...	...
7618	VAV	ZNT	VAVGB-SW2	2022-02-04	23.413017	...	23.011742	23.011742	23.011742	23.011742	23.011742
7619	VAV	ZNT	VAVGB-SW2	2022-02-05	23.011742	...	23.207766	23.410517	23.410517	23.410517	23.410517
7620	VAV	ZNT	VAVGB-SW2	2022-02-06	23.410517	...	23.216248	23.216248	23.216248	23.216248	23.216248
7621	VAV	ZNT	VAVGB-SW2	2022-02-07	23.216248	...	22.815369	22.815369	22.614338	22.614338	22.614338
7622	VAV	ZNT	VAVGB-SW2	2022-02-08	22.614338	...	23.022259	23.022259	23.022259	23.022259	23.022259

Figure 2. Data table illustrating the features of multidimensional inputs of different VAV terminal units.

Figure 2 shows that, for different measure types, each feature is just the recorded value at each 15-minute interval of the day. For example, the first feature of the dataset is the damper position of VAV1A-CNW1 and its measured value is 100% at 00:00 am. The last feature of the dataset is the zone temperature of VAVGM-SW2 and its measured value is 23°C at the last

interval of the day (#96 at 11:45 pm). It is not uncommon to observe missing values within the dataset, due to communication and connectivity issues. Therefore, it is important to remove them to maximize the overall accuracy of the anomaly detection algorithm. While there are many possible solutions for handling missing data points, here we fill the missing points with the mean of each row if fewer than 20% of a day's data is missing. A day's data will be completely dropped off if more than 80% is missing.

Unsupervised Learning Anomaly Detection

Many techniques for unsupervised learning are available within the Scikit-Learn toolbox, such as principal component analysis (PCA), fast multi-class scoring disagreement (MCSD), local outlier factor and isolation forest. In this study, we use isolation forests (Liu et al. 2008; Lie et al. 2012) to carry out anomaly detection because this approach is more suitable for handling multi-dimensional inputs and is fast to deploy. Another advantage of isolation forests stems from the fact that its detection performance converges quickly with a small number of trees, and it only requires a small sub-sampling size to achieve high detection accuracy (Liu, Ting, and Zhou 2008).

Figure 3 shows the structure of the isolation forests. It works by recursively partitioning data into subsets by randomly selecting splitting points. Points that are highly anomalous don't require many random splits for them to be partitioned into their own subsets, whereas those data points that are more "normal" will typically require many splits. The splits form a tree structure, and the number of splits can be thought of as the path length. Those points with long path lengths are "normal", whereas those points with short path lengths are considered "anomalous". An ensemble of trees is created in this way to create a "forest". The path length can be calculated by counting the number of edges an iTree traverses from the root node until the traversal is terminated at an external node. The final isolation score for each point is an average of the path lengths across each isolation tree in the forest. An anomaly score is required in order to rank the selected measure type. The anomaly score of an input sample is computed as the mean anomaly score of the trees in the forest.

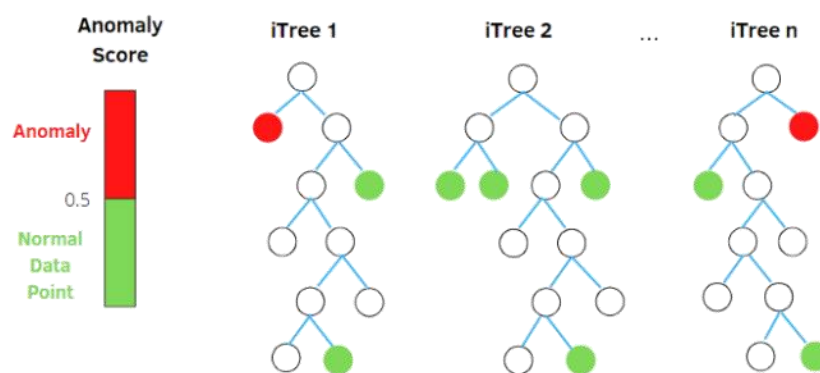


Figure 3. The concept of the isolation forest. The anomaly score is on a scale from 0 to 1.

When the score of the observation is close to 1, the path length is very small, and the data point is easily isolated: we have an anomaly. When the score of that observation is smaller than

0.5, the path length is large and then we have a normal data point. When training the isolation forest, the two most important hyper-parameters are number of trees T and max sampling size ψ . In this work, we set the number of trees to be 100 because the path lengths usually converge well before $t = 100$. The maximum sample size is set to 256.

Let $y_{m,u,t}$ be the observed value for measure m in unit u at time t . Then a full day's historical data can be presented as: $\mathbf{y}_{m,u} = \{y_{m,u,00:00}, y_{m,u,00:15}, \dots, y_{m,u,23:45}\}$. We train the isolation forest, $f_m(\{\mathbf{y}_{m,u}\}_u)$ for individual measure types using a full day's data so that each measure type has an anomaly score generated (ranging from 0 to 1). We run the training algorithm continuously for multiple days to obtain aggregated anomaly scores of each measured type for any given time period.

These anomaly scores allow us to rank each VAV unit based on how anomalous its readings are for measure m . If a certain measure was anomalous, then it is likely that every relationship involving that measure would also be classified as anomalous. For example, a faulty air flow sensor might cause zone temperature deviation to occur at the building. It is therefore more reasonable to rank the anomaly level of individual VAV terminal units based on multidimensional measures rather than a single measure.

Figure 4 shows the anomaly scores of four measure types for VAV1B-NE1, including ZNT, ZNTSP, FLR, and DPR. We can quickly see that this unit has unusual ZNT readings, as the anomaly score of the measured type ZNT is 0.95, which indicates this VAV unit has more anomalous zone temperature readings than the other terminal units. The damper position (DPR) looks normal when compared with other units in the group.

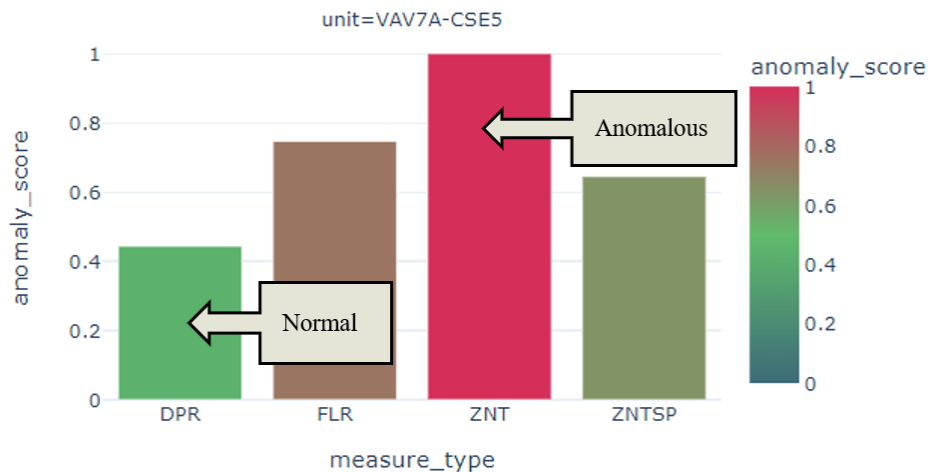


Figure 4. Anomaly scores of different measure types for a VAV terminal unit.

Anomaly Interpretation

To aid diagnosis and defect remediation, we are able to provide further analysis to help users understanding the anomaly along with the anomaly scores. Here we derive the statistical features of the anomalous data points to get insights into why a VAV terminal is relatively anomalous compared to the others. The example anomaly interpretation is listed in Table 1. Once

a measured type is identified as an anomaly, a separate algorithm will be executed to compare the statistic features of the anomalous measured type against the other units. These features include:

- Mean
- Standard deviation
- Maximum
- Minimum
- Absolute difference zone temperature value and zone temperature setpoint
- Coefficients of a quadratic curve fit to a particular relationship (e.g. VAV_FLR and VAV_DPR)

The data used to calculate daily anomaly scores for each measured type are also used to calculate statistical features. If the anomaly scores are calculated upon multiple days, then the statistical features of multiple days will be calculated instead. We compare the static feature: if a certain statistical feature of the anomalous measured types is significantly higher relative to the same features of the other units, then we conclude that this feature is contributing to the high anomaly score of the unit.

Table 1 shows how the statistical measure can be linked to possible root causes. For example, if a unit has a high anomaly score, and if its average zone temperature during occupied hours of the selected period is significantly lower than the average zone temperature of all the VAVs, then we attribute the “average feature” as the reason of the high anomaly score. The root cause explanation might be an over-cooling, or faulty temperature sensor. If the standard deviation of a flow sensor reading is zero, it would suggest that the airflow sensor reading is flat and does not change at all, hence it is faulty.

It is worth mentioning that in most cases, faults and their symptoms are easily mixed up. One fault may have different symptoms on multiple features, and one symptom may be associated with multiple faults (Qin and Wang, 2005). To identify the interactions between symptoms and faults, usually one needs to set up more complicated engineering rules to identify the root causes. Interaction between faults is not the focus of this study and will be the subject of future research.

Table 1. Anomaly interpretation

Measure types	Feature description	Possible faults
ZNT FLR	Mean value is too high/low	Senor reading is too high Senor reading is too low Zone temperature setpoint was inappropriately assigned
ZNT FLR	Maximum value is too high	Sensor reading goes beyond the upper range
ZNT FLR	Minimum value is too low	Sensor reading goes beyond the lower range
ZNT FLR DPR	Standard deviation is too small	Stale temperature or air flow sensor reading VAV damper stuck

ZNT ZNTSP	Absolute difference between zone temperature and zone temperature setpoint is too big	Zone temperature does not track setpoint
FLR DPR	Coefficients of a quadratic curve fit to a particular relationship	VAV damper sticking PID control error

Once anomalies are identified, it is important to rank them so that building operators can prioritize remediation tasks. Here we ranked the anomalies based on two different approaches:

1. Obtain top-N anomalies of individual measure types and rank them based on anomaly scores.
2. Rank the units based on the total anomaly scores of multidimensional inputs.

Results and Discussion

We extracted data from 363 VAV terminal units from 2/2/2022 to 2/25/2022 to test the proposed anomaly detection algorithm. Since the data set is relatively short and there are already many known faults in existence within the period, it would be hard to achieve satisfactory results using supervised or semi-supervised learning algorithms.

For each VAV unit, the above mentioned four measure types were considered. No data labeling work was done before we fed the raw data into the unsupervised learning algorithm. For each day, our algorithm returned an anomaly score for each measure type, based on its profile shape versus other similar units. This process repeats from 2/2/2022 to 2/25/2022 so that one could assess the results over an extended period. The daily anomaly scores of each measured type are added up to obtain an aggregated anomaly score for the entire assessment period. The VAVs with the highest anomaly scores were deemed to have the most abnormal behavior compared to other units.

Anomaly Detection Results Based on Separate Measures

First, we investigated the anomaly detection results by looking at separate measure types. We ranked the units by their anomaly scores and selected the top five most anomalous units to perform analysis. The timeseries graphs of the top five most anomalous units were also plotted to allow users (mainly building engineers or BMS technicians) to the anomalies. Figure 5 shows the five most anomalous zone temperature readings returned from our anomaly detection algorithm. It clearly shows that the top five most anomalous units either have stale sensor readings or abnormally high values throughout the experimental period. Figure 6 shows the most ordinary zone temperature readings: the temperatures were fluctuating around the temperature setpoint for most of the time, which is as expected. Figure 7 shows the five most anomalous air flow rate readings: it appears that the flow rates of the VAV units never dropped to zero at all, even though they are not supposed to run ‘after-hours’. Moreover, the air flow rates of the first anomaly VAV (VAV2B-NW3) seem to have had relatively high readings compared to other units. Figure 8 shows the normal flow rate readings which have some obvious setback patterns. After checking with the on-site engineering operations team, we confirmed that those

temperature and flow sensors were either offline or needed to be replaced. This points to the effectiveness of the implemented algorithm.



Figure 5. Time series plot showing top five most anomalous zone temperature readings.

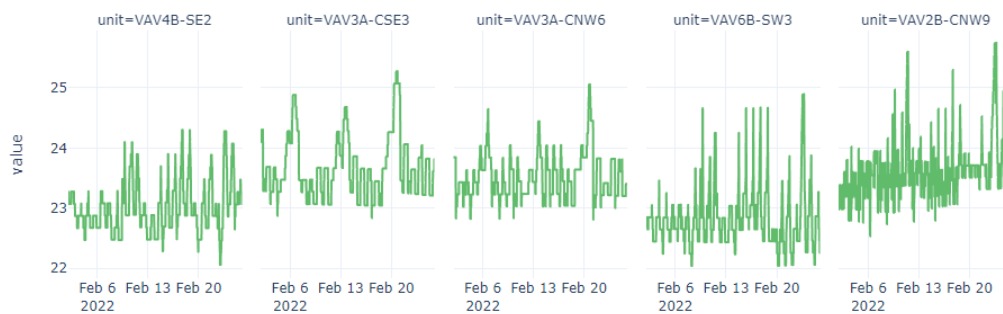


Figure 6. Time series plot showing top five normal zone temperature readings.

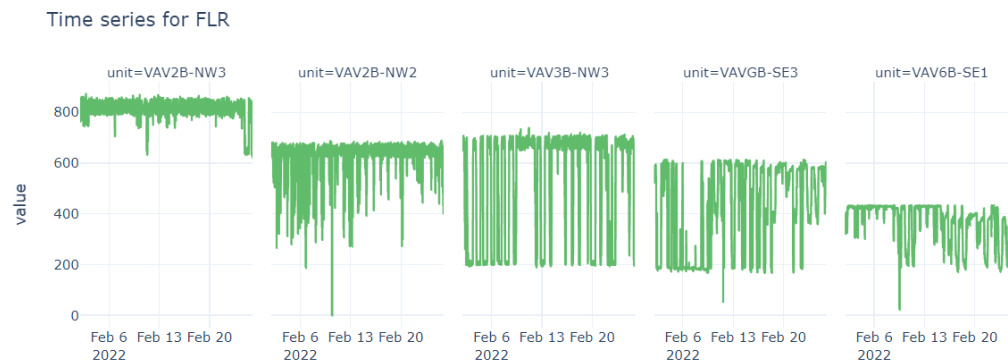


Figure 7. Time series plot showing top five most anomalous flow rate readings.

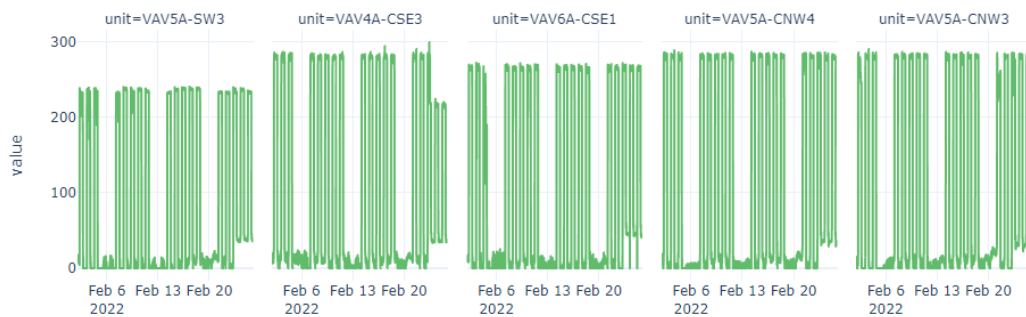


Figure 8. Time series plot showing top five normal flow rate readings.

The same practice was repeated for two other measure types, FLR and DPR. Table 2 summarizes all the anomalies that require investigation. The issues included disconnected sensors, faulty flow sensor readings, and wrongly located zone temperature sensors. It was found that, for each unit, even if one of the four measures has a high anomaly score, it does not always lead to high anomaly scores of other measures from the same unit. This is because the failure of a single measure does not always lead to operational issues.

Table 2. Top anomalies for separate measure types

Faulty terminal unit	Measure type	Fault Symptoms
VAV1A-NW1	ZNTSP	Setpoint is set to be 16°C
VAV2A-SE1	ZNTSP	Setpoint is set to be 24°C
VAV2A-SW2	ZNTSP	Setpoint is set to be 24°C
VAV2B-CNW3	DPR	Incorrect damper position data
VAV2B-NW1	DPR	Setpoint is set to be 16°C
VAV3B-CNW2	ZNT	Zone temperature reading out of range
VAV3B-CNW2	FLR	Air flow sensor reading is too high
VAV3B-NW1	DPR	Stuck damper
VAV4A-CSE6	ZNT	Zone temperature reading out of range
VAV4A-CSE6	FLR	Air flow sensor reading is too high
VAV4B-CNW3	DPR	Stuck damper
VAV5A-CSE5	ZNTSP	Setpoint is set to be 16°C
VAV5A-NE3	ZNT	Zone temperature reading out of range
VAV5A-NE3	FLR	Air flow sensor reading is too high
VAV7A-CSE5	ZNT	Zone temperature reading out of range
VAV7A-CSE5	FLR	Air flow rate is lower than threshold
VAVGA-SE1	DPR	Stuck damper
VAVGB-CNW2	ZNT	Zero reading
VAVGB-CNW2	FLR	Zero reading
VAVGB-CNW2	ZNTSP	Zero reading

Anomaly Detection Based on Multiple Measures

In the second experiment, we ranked the units based on the overall anomaly scores of the four measure types for individual VAV terminal units. For better visualization we use stacked bar charts to indicate anomaly scores associated with different measure types, on different days. Each unit has four bars, and each of them represents a specific measure type. The daily anomaly score is stacked on top of another to form the total anomaly score for an entire period. This allows user to easily see the anomaly score and understand which inputs contributed the most to the total anomaly score. Figure 9 ranks the top ten most anomalous VAVs. The units that have the highest overall anomaly scores over the entire period are the most anomalous.

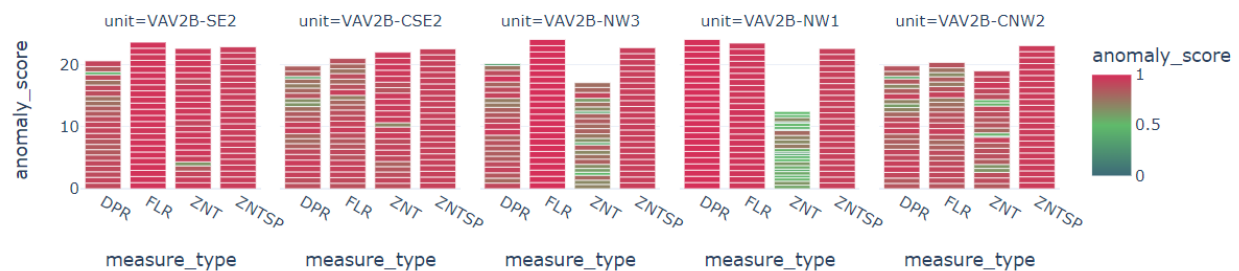


Figure 9. VAV terminal units that have the highest anomaly scores throughout the month.

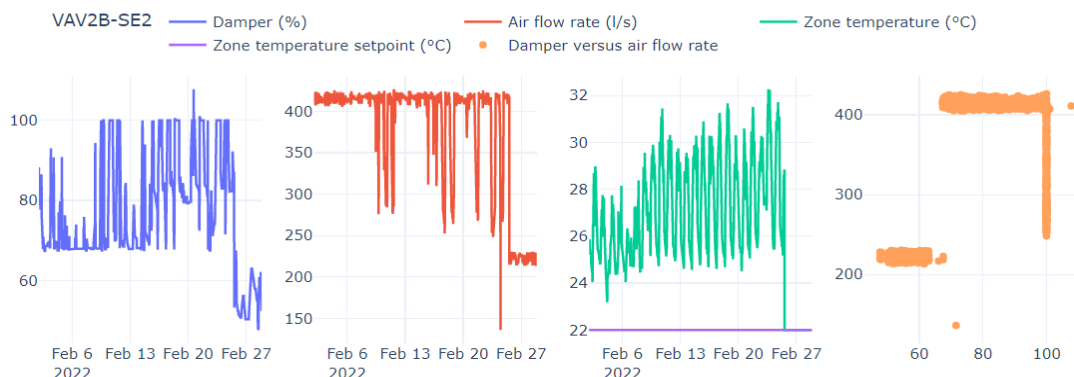


Figure 10. Time series graph of the most anomalous VAV (VAV2B-SE2).

Figure 10 shows the time series plots of the four measure types associated with the most anomalous VAV (VAV2B-SE2). The graph indicates that the zone temperature was not able to reach the desired temperature setpoint of 22°C, even though the air flow rate was high enough to supply enough cool air into the space. The scatter plot shows there was a step change in air flow rate when the damper position changed from 60% to 70%, while under normal circumstances the damper position should be linearly correlated with the rate of air flow.

Figure 11 shows the time series plot of a normal VAV to illustrate the difference. This figure shows that the air flow rate is zero during after-hours operation but increases during the operating hours, which fits the normal operational pattern of air flow rates. The scatter plot also shows there is an obvious linear correlation between the damper position and air flow rate, which indicates the damper is functioning properly. The zone temperature seems slightly abnormal as it was drifting away from the setpoint from 14th February onwards but was within the 1.5°C deadband for most of the time so the anomaly score was still acceptable.

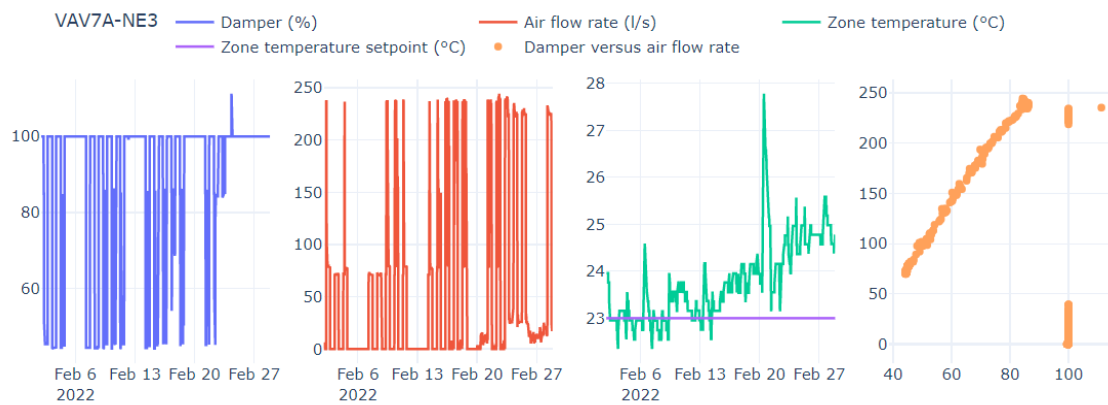


Figure 11 Time series graph of a normal VAV (VAV2B-SE).

Rectifying The Anomalies

The facilities management team reviewed the anomalies generated, most of which were addressed by the on-site technicians. The action items included rewiring the temperature sensors, fixing the stuck dampers, resetting flow rate controls, replacing flow sensors, and relocating temperature sensors. There were also some non-actionable anomalies, such as unoccupied zones, over-sized equipment, and even wrongly mapped data points. Some anomalies are obvious and can be easily picked up by simple engineering rules. We would anticipate that filtering out this dataset by using the proposed fault detection tool may also further improve accuracy.

Conclusions And Future Work

In this paper, we present the use of an unsupervised learning anomaly detection method for detecting faulty VAV units in commercial buildings. This method does not require building engineers to manually label data, or spend time collecting fault-free training data, so the deployment cost is much lower than other machine learning based anomaly detection methods. With this method, we can run the FDD analysis immediately after data collectors are installed. While most data-drive fault detection approaches do not contain any diagnosis, this paper also demonstrates a simple method of using a list of statistical measures to interpret anomalies, which provides a good foundation for root cause analysis.

We have tested two different approaches to rank the collected anomalies. Interestingly, we noticed that the two ranking algorithms were returning different groups of faults. The first approach prioritizes the units that have the highest anomaly scores on single measure types. The anomalies picked up by this approach have the highest anomaly scores, so they tend to be more anomalous than others. The second approach prioritizes the units that have multiple symptoms/faults happened at the same time, which indicates the symptoms of one fault are likely to propagated to other measure types. During the application process, we found that the most accurate outcome can be achieved by investigating the top anomalies from both groups. Although this work focuses exclusively on VAVs, the same approach can be easily extended to other HVAC sub-systems. Since our approach finds anomalous data patterns by comparing time series profiles among many identical or similar systems within buildings, it can easily be applied to any major system components that have similar specification and operation, such as AHUs.

The diagnostic method presented in this study is relatively simple but may still require expert knowledge to perform accurate root cause diagnostics. Our future work will focus on improving the interpretability of the anomalies and fully automate the diagnostics part.

References

- Dey, M., S. Rana, and S. Dudley. 2018. "Semi-Supervised Learning Techniques for Automated Fault Detection and Diagnosis of HVAC Systems". In *proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. doi.org/10.1109/ICTAI.2018.00136
- Du, Z., B. Fan., J. Chi., and X. Jin. 2014. "Sensor fault detection and its efficiency analysis in air handling unit using the combined neural networks". *Energy and Buildings*, 72: 157-166. doi.org/10.1016/j.enbuild.2013.12.038
- Huang, H., C. Roach, and A.C. Roussac. 2020. "From local optima to overall optimum: a field study of model-based anomaly detection in commercial buildings". In *proceedings of the 2020 ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, California: ACEEE
- Hyndman, R. J., E. Wang, and N. Laptev. 2015. "Large-Scale Unusual Time Series Detection". In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* 1616–9. dx.doi.org/10.1109/ICDMW.2015.104.
- Kim, W., and S. Katipamula. 2017. "A review of fault detection and diagnostics methods for building systems". *Science and Technology for the Built Environment* 24(1): 3-21. doi.org/10.1080/23744731.2017.1318008
- Liu, T.F., K.M. Ting, and Z. Zhou. 2008. "Isolation Forest". In *proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pp. 413-422, [doi:10.1109/ICDM.2008.17](https://doi.org/10.1109/ICDM.2008.17).
- Mattera, C., H.R. Shaker, and M. Jradi. 2019. "Consensus-Based Method for Anomaly Detection in VAV Units". *Energies* 12(3): 468. doi.org/10.3390/en12030468
- Mirnaghi, M., and F. Haghighat. 2020. "Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review". *Energy and Buildings* 229: 110492. doi.org/10.1016/j.enbuild.2020.110492
- Roach, C. 2018. "Exploring unusual sensor behaviour in buildings using BMS data and unsupervised learning techniques". In *Proceedings of the 2018 ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, California: ACEEE
- Sipple, J., 2020. "Interpretable, Multidimensional, Multimodal Anomaly Detection with Negative Sampling for Detection of Device Failure". *arXiv:2007.10088* [eess.SP] doi.org/10.48550/arXiv.2007.10088

Qin, J., and S. Wang. 2005. “A fault detection and diagnosis strategy of VAV air-conditioning systems for improved energy and control performances”. *Energy and Buildings* 37(10): 1035-1048. doi.org/10.1016/j.enbuild.2004.12.011

Yan, K., J. Huang., W. Shen., and Z. Ji. 2020. “Unsupervised learning for fault detection and diagnosis of air handling units”. *Energy and Buildings* 210: 109689. doi.org/10.1016/j.enbuild.2019.109689