International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:14, No:11, 2020

Air Handling Units Power Consumption Using Generalized Additive Model for Anomaly Detection: A Case Study in a Singapore Campus

Ju Peng Poh, Jun Yu Charles Lee, Jonathan Chew Hoe Khoo

***Abstract*—**The emergence of digital twin technology, a digital replica of physical world, has improved the real-time access to data from sensors about the performance of buildings. This digital transformation has opened up many opportunities to improve the management of the building by using the data collected to help monitor consumption patterns and energy leakages. One example is the integration of predictive models for anomaly detection. In this paper, we use the GAM (Generalised Additive Model) for the anomaly detection of Air Handling Units (AHU) power consumption pattern. There is ample research work on the use of GAM for the prediction of power consumption at the office building and nation-wide level. However, there is limited illustration of its anomaly detection capabilities, prescriptive analytics case study, and its integration with the latest development of digital twin technology. In this paper, we applied the general GAM modelling framework on the historical data of the AHU power consumption and cooling load of the building between Jan 2018 to Aug 2019 from an education campus in Singapore to train prediction models that, in turn, yield predicted values and ranges. The historical data are seamlessly extracted from the digital twin for modelling purposes. We enhanced the utility of the GAM model by using it to power a real-time anomaly detection system based on the forward predicted ranges. The magnitude of deviation from the upper and lower bounds of the uncertainty intervals is used to inform and identify anomalous data points, all based on historical data, without explicit intervention from domain experts. Notwithstanding, the domain expert fits in through an optional feedback loop through which iterative data cleansing is performed. After an anomalously high or low level of power consumption detected, a set of rule-based conditions are evaluated in real-time to help determine the next course of action for the facilities manager. The performance of GAM is then compared with other approaches to evaluate its effectiveness. Lastly, we discuss the successfully deployment of this approach for the detection of anomalous power consumption pattern and illustrated with real-world use cases.

***Keywords*—**Anomaly detection, digital twin, Generalised Additive Model, Power Consumption Model.

I. Introduction

THIS emergence of digital twin technology has improved the real-time access to data from a collection of sensors about the performance of buildings. In Singapore, the digital transformation opened many opportunities to improve the management of the building by using the data collected to help monitor consumption patterns and energy wastage. The real-time access to data allows us to better manage our buildings. Firstly, it provides facilities' managers a better understanding of the overall energy footprint of the building through various performance metrics that is made available by the access of real-time data. Secondly, it allows for a more intelligent identification of possible equipment faults and inefficiencies [1]. Lastly, it enables building and business owners a more automated and guided system to determine the health of the building's energy system and react with a proactive approach in maintenance.

A central heating, ventilation and air-conditioning (HVAC) system is widely used in large buildings, such as academic buildings [2]. The HVAC systems are the largest energy end use in the non-residential sector [3] where inefficient operation and maintenance of the HVAC system can cause energy wastage, customer complaints, poor indoor air quality and even environmental damage [4]. In order to improve the facilities' management operations, the detection of anomalous power consumption pattern is particularly important for the efficient identification of the need for maintenance work. Thereby, in this paper, we illustrate the use of data collected from the building's measurement and verification system to help monitor consumption patterns and energy wastage in order to allow for a more efficient use of resources and the identification of the need for maintenance work. This is done through the integration of predictive models, such as GAM for anomaly detection. One of the key tenets of the GAM framework is that it fits a non-linear smooth function to each explanatory variable thus allowing for modelling non-linear relationships that the standard Ordinary Least Squares (OLS) type regressions will miss.

In this study, we trained the general GAM model using the historical power consumption data of AHU, an important component of HVAC between Jan 2018 to Aug 2019 from an education campus in Singapore to detect anomaly power increase or decrease among AHU power values. The identification of such anomalous data points is based entirely on historical pattern without explicit programmatic inputs from the domain expert. The maximum a posteriori (MAP) estimation method uses the maximisation of the posterior probability to determine an estimation of future observations. Anomalies can then be detected by contrasting with actual real-time AHU power consumption values when it is

Ju Peng Poh, Data Scientist, is with the Centre for Applied Intelligence, Certis Group, Singapore (phone: +65 94516192; e-mail: poh_ju_peng@certisgroup.com).
Jun Yu Charles Lee is with the Centre for Applied Intelligence, Certis Group, Singapore (phone: +65 84284380; e-mail: charlesjy_lee@certisgroup .com).
Jonathan Chew Hoe Khoo is a Co-founder and Data Scientist with Research Room, Singapore (e-mail: jon@researchroom.sg).

available.

One key point of this paper is the use of imputed AHU power consumption. There is no direct electrical power meter measuring the electrical consumption at the level of the AHUs due to legacy installation decisions made in the past. Using results from National Environmental Agency (NEA) [5], we used the speed of the VSD to derive the imputed electrical consumption.

Upon detecting anomalies in power consumption, we further proposed prescriptive analytics framework to include prediction of cooling load values in real-time by identifying data points that have deviation beyond a specified range of confidence interval. It is used to complement the power consumption prediction model in developing a semi-automated flow chart for investigation and decision making whereby it recommends course of action for facilities managers to take immediate remedial actions. The cooling load is the quantity of heat energy that needs to be extracted from a space to keep the temperature within an appropriate range [6]. Both the power consumption and cooling load prediction model determine whether a set of rule-based conditions are evaluated in real time to determine the next course of action to be taken to further investigate the root cause of the issue.

To assess its effectiveness, the performance of GAM will be compared to other relative approaches. Lastly, we discuss the deployment of this approach by illustrating the anomaly detection capabilities of the GAM for AHU power consumption and cooling load data. The results are promising as investigation of the anomalies identified issues such as a fault in the meter and equipment, and highlighted explanations for the deviations due to human activities.

There are plenty of research work on the application of GAM in predicting power consumption at office building and nation-wide level [7]. However, there is little to no information on its anomaly detection framework and capabilities, as well as the framework's integration with the digital twin technology. As such, our paper contributes to current literature by demonstrating the predictive model's ability to significantly improve the efficiency of the facilities management operations through an anomaly detection framework.

II. Related Work

Among the Artificial Intelligence (AI) methods, Artificial Neural Network (ANN) [7], [8] and Support Vector Machine (SVM) [6], [9] are the most widely used for power consumption modelling. They result in high accuracy but require knowledge about the number of neural layers, neuron nodes and the activation and kernel functions before training. Even with such knowledge, it is common that there are issues with overfitting or long training time. Methods using statistical models like time-series analysis or GAM [10] are also commonly used. ARMA (Autoregressive moving average) has been used to build power consumption model for households, while ARIMA and seasonal ARIMA were used to predict power consumption [11]. However, these methods are affected by the interrupted patterns because of holidays. As such, GAM [10] is used for our paper as it is an extended form of the of regression analysis that allows the analysis of individual factors. There is more flexibility as compared to regression model as these factors can be non-linear. For power consumption prediction using GAM models, the expected value and variance has been estimated using GAMARMA [12] and GAM-AR-ARCH (Autoregressive conditional heteroscedastic) [13]. They are then used for power anomaly detection in a research building. GAM's ability to decompose easily and accommodate to new components, such as a new source of seasonality, makes it a suitable model for the prediction of power consumption. Although current literature focused on the methodology of using GAM for predicting power consumption, there is limited illustration of its anomaly detection framework and capabilities, and the framework's integration with the latest development of digital twin technology.

III. Overview

A. Digital Twin

The Digital Twin for the framework is a digital representation of the academic building using BIM models. Leveraging on a digital platform in building our solution, it is to create a Digital Twin of the academic building using Data Management Systems, 3D BIM, Application Services and Analytics.

The Digital Twin is materialized through the creation and deployment of a solution stack consisting of connectivity to current selected building subsystems, ingesting, accumulating, and abstracting of data, to the use of customized applications and collaborative processes.

The existing systems linked to the Building Management System are linked into the Digital Twin. This platform can integrate with various subsystems in a building to consolidate data into a single platform. This allows us to eliminate building management using silo systems. The building integration is to collect energy related (ACMV) data which are required by analytical engine.

Having seamless integration to subsystems, it is widening the usage of digital twin by bringing alarms and events. It also provides 3D look and feel of the actual location of the alarm which helps the facilities management team to observe additional information such as other affected areas, critical conditions of other systems.

There are many facilities management application benefits for digital twin technology. This includes asset register and management, monitoring of assets in the Digital Twin, together with alarm and event management, detection of anomalies using analytic engine and asset maintenance history.

For this project, the required historical data are extracted from the digital twin for modelling purposes. Although the data are efficiently collected, one of the major issues is unclean data. This limitation and our proposed solution are discussed further in future section.

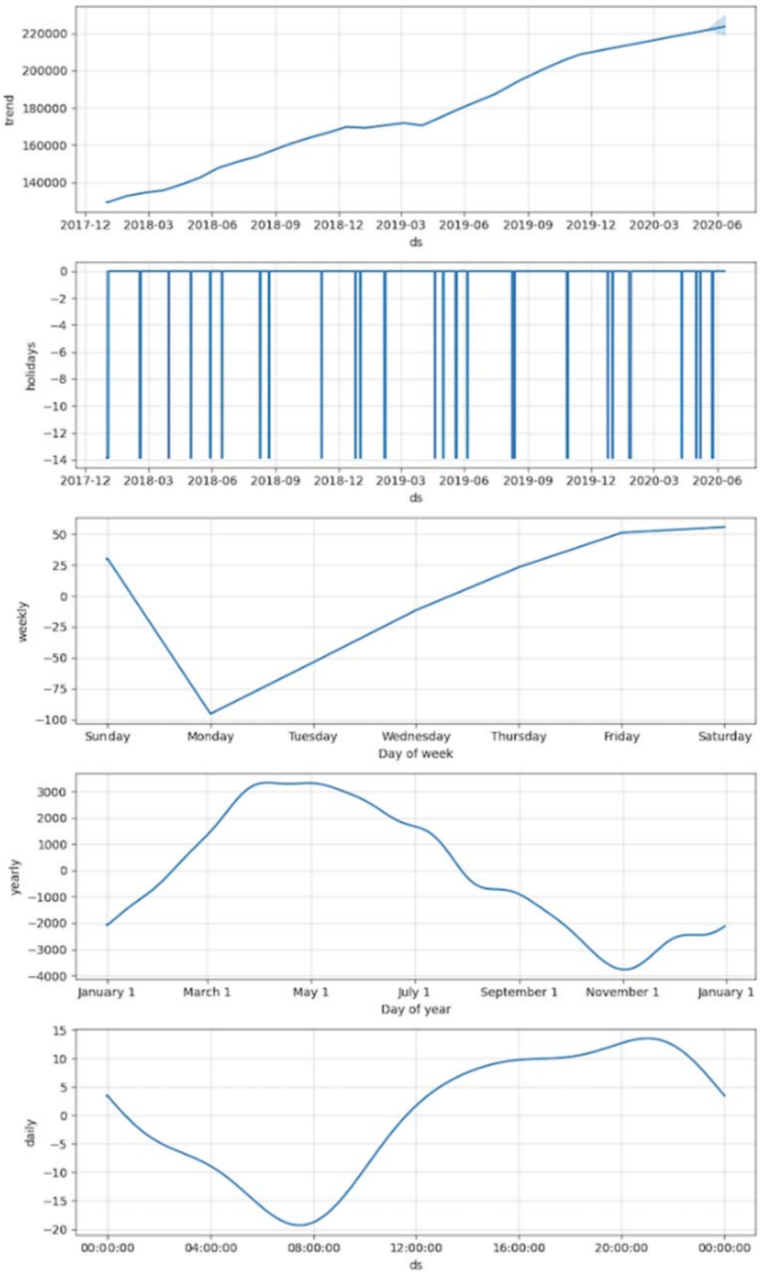

Fig. 1 Analysis of Components of the GAM model

B. GAM Model

Using the historical data extracted from the digital twin, GAM is used for modelling purposes. GAM [10] is an advanced regression model that explores the dependency of response variable (or dependent variable) $\boldsymbol{Y_t}$ to a set of explanatory variables $\boldsymbol{x_{t,1}}, \boldsymbol{x_{t,2}} \ldots \boldsymbol{x_{t,p}}$ in a flexible form, where the contributions from each explanatory variable $\boldsymbol{x_{t,i}}$ is summed up as terms of transfer functions (or smoothing functions) $\boldsymbol{f_i}(\boldsymbol{x_{t,i}})$, as shown in (1):

$$F\left(E\left(Y_t\right)\right) = f_i(x_{t,i}) + \cdots + f_p(x_{t,p}) + \varepsilon = \sum_{i=1}^{p} \boldsymbol{f_i}(\boldsymbol{x_{t,i}}) + \boldsymbol{\varepsilon_t} \quad (1)$$

The framework uses a predictive model using GAM for the power consumption and the cooling load of an academic building, taking into consideration the usage of resources, equipment and building operations. This is done by using the data from the Digital Twin, a digital representation of the academic building developed for this project.

The following are the types of variables that are included in the model:

1) Trend variables: the non-stationary component of the time-series.
2) Continuous timing variables: hour of the day: day in the year:

3) Categorical timing variables: day type which contains seven days in a week and holidays

The GAM formulation has the benefit of effortlessly decomposing and accommodating new components when required, such as when a new source of seasonality is discovered [14]. GAMs also fit very quickly, so that the user can interactively change the model parameters

The analysis of the components of the GAM model is shown in Fig. 1. It illustrates the decomposition of the GAM model into "trend" variable, categorical timing variable as "holiday" and days-in-a-week variable as "weekly" variable and continuous timing variables as day-in-the-year and hour-of-the-day as "yearly" and "daily".

IV. Experiment Setup

In this paper, we trained the GAM model using the historical AHU power consumption and cooling load data between Jan 2018 to Aug 2019 from an education campus in Singapore to derive the prediction models. It has a gross floor area of 23.6 thousand sqm and air conditioned areas of 8.4 thousand sqm.

A. Power Consumption Pattern

Figs. 2 and 3 illustrate the power consumption pattern of the entire building. Fig. 2 shows the total electricity consumption of the campus building for each month between Jan 2018 to Mar 2019. Fig. 3 shows the total electricity consumption of the campus building for different periods of the campus' academic calendar between Jan 2018 to Mar 2019. The results are consistent with the operations of the building for the different periods of the campus calendar. During the study weeks, there would be a higher level of electricity consumption since more academic activities take place during this period. The electricity consumption during vacations is only slightly lower as the campus still opens several classes to the public during this period. Among the various sources of energy consumption within a building, the airside of the air conditioning system contributes to around 10% of total consumption on average.

To project the predictive range of total AHU power consumption and cooling load, we created a forecasting model using a GAM framework. In the machine learning model, the modelling seasonality as an additive component is the same approach taken by exponential smoothing. This includes hour of the day, day in the year, day of the week, and capturing the seasonality effect of the campus schedule which includes vacation week, school week and examination period. It dynamically captures the change based on past power consumption behaviour and reflect in the forward prediction range.

B. Predictive Analytics of Power Consumption

Using the forecasting model of machine learning model GAM, any power consumption that is beyond the upper and lower level of the predictive range may be identified as anomalies. The magnitude of the deviation coupled with the persistence of the anomaly corresponds directly to seriousness of the anomaly. To facilitate swift interpretation, anomalies have been classified using the familiar traffic light colour coding system reflecting different levels or seriousness of risks, with red reflect the highest, amber in the middle, and green as lowest as shown in Figs. 4 and 5.

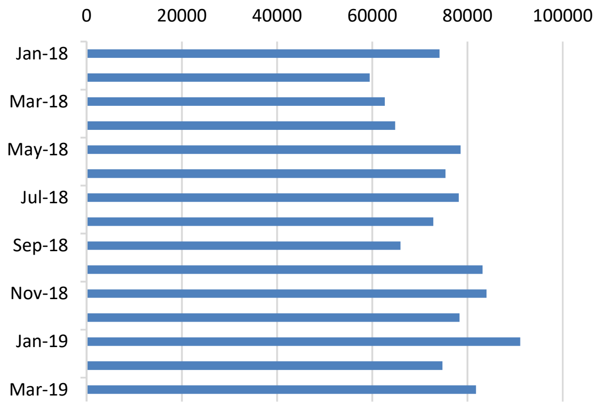

Fig. 2 Total Electricity Consumption by Month (kWh)

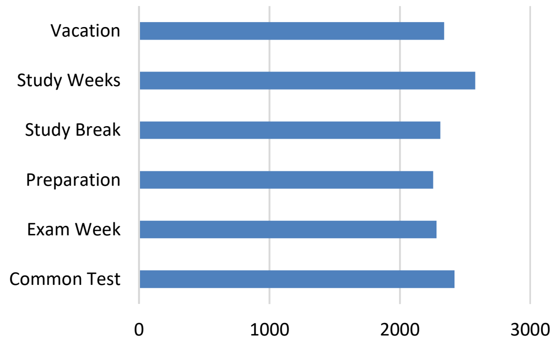

Fig. 3 Average Electricity Consumption by Schedule (kWh)

C. Predictive Analytics of Cooling Load

In our prescriptive analytics framework, the anomaly detection of cooling load using the GAM prediction model is used to complement the power consumption prediction model. The cooling load is defined as the amount of heat energy that needs to be removed from a space to ensure the temperature is within a reasonable range [6]. Similar to the forecast model for power consumption, any cooling load data point that is beyond the upper and lower range may be identified as anomalies as shown in Fig. 5.

If an anomalously high or low level of power consumption is met with a corresponding anomalously high or low level of cooling load, the cause of this event is attributed to an unusual change in load. This is expected as an increase in cooling load results in greater power consumption needed to remove the heat from the space. A further investigation is not warranted in this case. However, when there is an absence of a corresponding anomalous change in cooling load, a set of rule-based conditions are evaluated in real time to determine the next course of action to be taken to further investigate the root cause of the issue. An alert with the recommendations is sent to the facilities manager for follow up action.

D.Predictive Models in Prescriptive Analytics Framework

The prediction models for both power consumption and cooling load are used in our framework to determine the recommended next course of action. The benefit of the prediction models is the ability to capture the consumption behaviour and cooling load in near real-time with the dynamic predictive range. These dynamic ranges help new facilities manager with little prior knowledge to immediately know with a glance whether a particular power consumption for a particular PAHU/AHU (or any other equipment) is excessive on a real-time basis, something that is not possible in the past even with a knowledge management in place. Even for the seasoned facilities manager, this immediate measure of abnormality allows the manager to skip the strenuous mental process of comparing to his/her experience against expected values of particular sensors at that specific times and decide if they are showing abnormal values, and instead spend his/her time to decide whether those flagged anomalies deserves further investigations and take immediate remedial actions according to the framework.

V. Applications

In this section, the application of our framework for anomaly detection is illustrated.

For modelling purposes, historical data are extracted from the digital twin. We then improved the functionality of the machine learning model, GAM, by integrating it into a real-time anomaly detection system that projects predicted ranges based on historical data. The degree of variance from the upper and lower bounds of the preceding uncertainty interval is used to determine the anomalous data points. A set of rule-based conditions is then used to help to determine the course of action next. This is done in real time with the course of action being recommended through an alert message to the facility manager.

Figs. 6-8 illustrate the overall framework of the capabilities of the machine learning model, GAM, in identifying anomalously high or low power consumption and identify the next course of action for each set of rule-based conditions in real time. The semi-automated flow chart is developed in consultation with subject matter experts based on the operations of the building's facilities management.

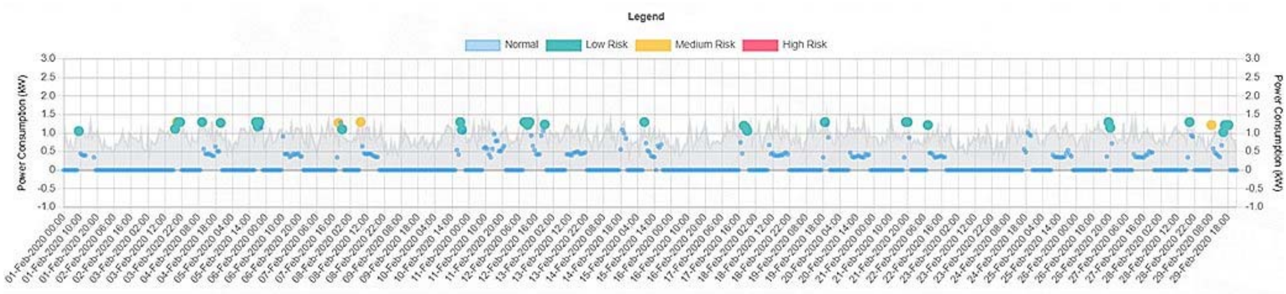

Fig. 4 Power Consumption with Predictive Range and Identified Anomalies in 10-hourly intervals

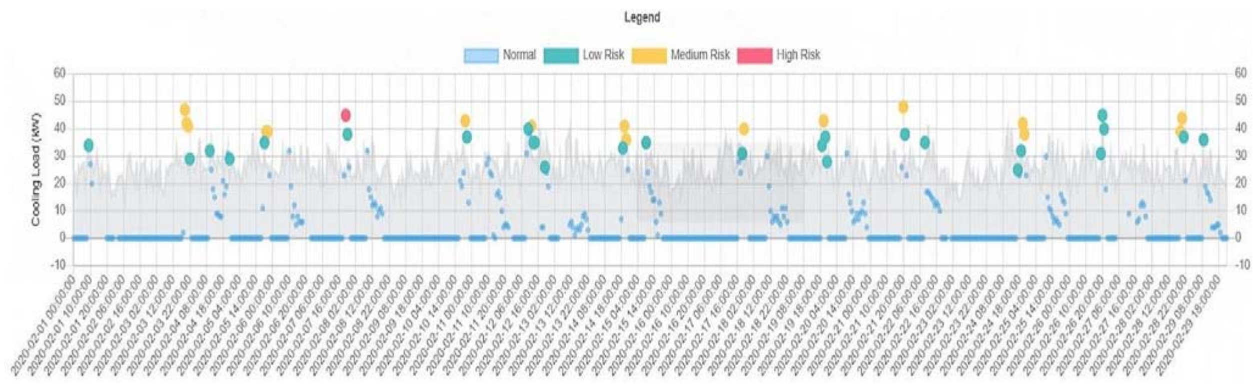

Fig. 5 Cooling Load with Predictive Range and Identified Anomalies in 10-hourly intervals

If an anomalously high or low level of power consumption is met with a corresponding anomalously high or low level of cooling load, the cause of this event is attributed to an unusual change in load. This is expected as an increase in cooling load results in greater power consumption needed to remove the heat from the space. A further investigation is not warranted in this case. However, when there is an absence of a corresponding anomalous change in cooling load, a set of rule-based conditions are evaluated in real time to determine the next course of action to be taken to further investigate the root cause of the issue. An alert with the recommendations is sent to the facilities manager for follow up action.

For a case whereby an anomalously high or low level of power consumption coincides with a corresponding

International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:14, No:11, 2020

anomalously high or low level of cooling load, an investigation is not required. However, for a case whereby there is no corresponding anomalously high or low level of cooling load, various rule-based conditions are evaluated.

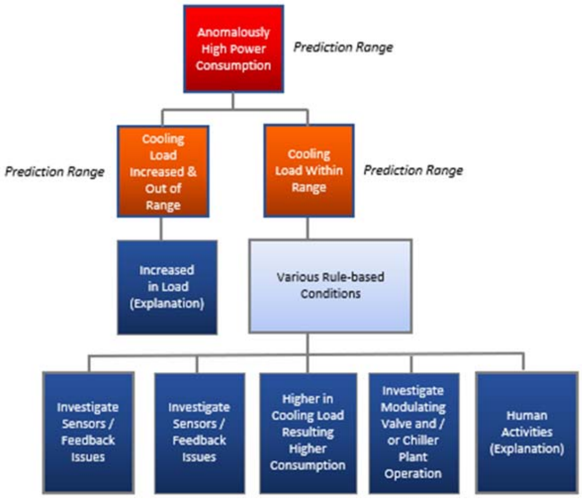

Fig. 6 General Framework for an Anomalously High Level of Power Consumption

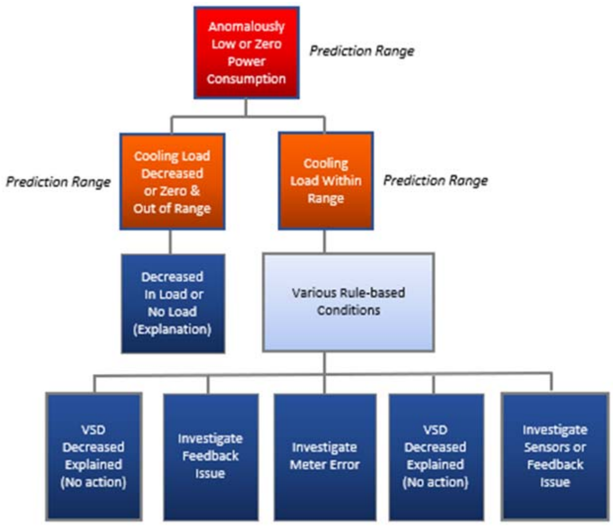

Fig. 7 General Framework for an Anomalously Low or Zero Level of Power Consumption Detected

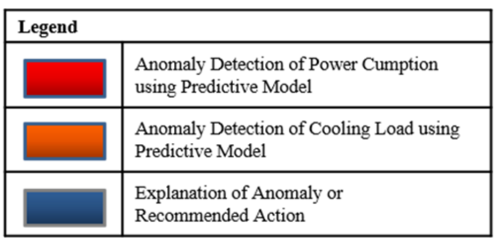

Fig. 8 Colour Code Legend for the General Framework

The PAHU and AHU Variable speed drives (VSDs) and other sets of variables are used to determine the next course of action based on a set of rule-based conditions. The several metrics that could be used for the framework include CO_2 level, Static Pressure, Return Air Temperature (RAT), VAV Box. A corresponding alert would be sent to a facility manager in real time to prompt for an investigation of the issue. This may include the investigation of feedback issues, modulating value or chiller plant operation.

Table I lists several metrics used for our framework.

TABLE I
METRICS USED FOR ANOMALY FRAMEWORK

	Meet Set Point	Does not Meet Set Point
CO_2	Equal or lower than CO_2 set point	Higher than CO_2 set point
Static Pressure	Equal or higher than SP set point	Lower than SP set point
RAT	Equal or lower than RAT set point	Higher than RAT set point
VAV Box	Equal or lower than VAV Box Temperature Set Point	Higher than VAV Box Temperature Set Point

The use of the machine learning model to identify anomalously high or low level of power consumption enables for a more efficient operation and the identification of the maintenance of the HVAC system.

VI. DEPLOYMENT

Before the deployment, the identification of anomalies is based on various individual metrics using pre-defined setpoints. These metrics include CO_2 level, Static Pressure, RAT, VAV Box. As they are identified individually, the existing system does not have sufficient information to make actionable recommendation or suggest possible explanations. As a result, the anomaly alerts are often unmeaningful and require time-consuming investigation to understand the best course of action to be taken. On average, 25 alerts were received by the facilities manager per month. As these alerts are often false alarms and unmeaningful without a recommended course of action or explanation of the anomalously high or low level of power consumption, they are often neglected.

By implementing our framework with the anomaly detection capabilities of the machine learning model, GAM, for both the power consumption and cooling load, it helps us to identify anomalies more meaningfully. This is because the framework takes into consideration the relationship between the power consumption pattern and all the various metrics together. After the implementation of our framework, the number of alerts fell to 5 per month on average. This has also resulted in a more efficient operation for the facilities managers as the course of actions are recommended and explanation is suggested in real time based on the semi-automated flow chart.

At the operational level, the framework reduced the time and effort required by the facilities' manager to identify these anomalously high or low level of power consumption and determine the next course of investigative actions to be taken. The facilities' manager would receive only a smaller subset of these alerts. Using the framework, the alert would also reflect the recommended course of action that should be taken by the facilities manager. This allows for a more targeted investigation over fewer number of alerts with higher level of accuracy. However, there is still a need for testing for the next several months to further validate the results.

VI. Comparison of Models

In the field of power consumption forecasting, the following are models that have been applied for anomaly detection.

TABLE II
COMPARISON BETWEEN DIFFERENT METHODS FOR POWER CONSUMPTION FORECASTING

Methods	Strength	Weakness
ANN, SVM	- Good prediction results	- Requires best combination of neural layer number, neuron node number, activation function, and kernel functions - Over-fitting or of long training time are often encountered.
ARMA and ARIMA		- Unable to accommodate interrupted patterns caused by holidays. - Not trivial to accommodate interrupted patterns caused by holidays.
GAM	- Missing values do not need to be interpolated e.g. from removing outliers - Easily accommodate seasonality with multiple periods - Fitting is very fast - Forecasting model has easily interpretable parameters	

This paper does not purport to advocate that GAM modelling is the best or only approach to power the predictive system. However, it does have certain features that we have chosen it. The GAM formulation has the capability to decompose efficiently and cater for new components when needed, for example when a new source of seasonality is detected [14]. GAMs also fit very quickly, so that the user can interactively change the model parameters. In addition, missing values do not need to be interpolated. Thus, this model is easy to be implemented at scale across other sensors of same type and extends to sensors of other types.

VII. Limitations

As stated, historical data are collected from the digital twin to construct the model. One of the major problems we faced is the integrity of this data. Even though data are efficiently collected, it is not clean. It is also not cost effective to employ manpower to clean the data, because this person needs to be meticulous and a domain expert. On the other hand, if we proceed with unclean data, we will be faced with dealing garbage-in-garbage-out.

We propose a compromised solution to deal with this problem. We still proceed with the extant uncleaned dataset and proceed with a clean-as-you-go system. Value from the system is derived from day one of implementation, and it improves as bad data is cleaned with feedback of the facilities managers who investigates the alarms and monitors the data. We propose simple data cleaning process, to exclude data-points that are known to the expert as erroneous, and in the next training cycle the errors would not propagate, and predictions will improve.

VIII. Conclusion

In our paper, we proposed an anomaly detection framework to monitor the AHU power consumption data in real-time by identifying data points that have deviation beyond a specified range of confidence interval from the GAM forecasting model. Based on a semi-automated flow chart, recommended course of action or an explanation for the anomalous level of power consumption is sent for facilities managers as an alert to take immediate remedial actions. The results are promising as investigation of the anomalies identified issues such as fault in the meter and equipment failure, and highlighted explanations for the deviations with higher level of accuracy. This paper illustrated the deployment of our approach and demonstrated the predictive model's ability to significantly improve the efficiency of the facilities management operations through an anomaly detection framework. This approach can be also be easily adapted for the application to other equipment such as water meter and chiller plant and other related equipment.

Acknowledgment

This research is supported by the National Research Foundation, Singapore and Building and Construction Authority (BCA) under its Green Buildings Innovation Cluster (GBIC) Program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and BCA.

The authors would like to thank Philip Wong from BCA Braddell Campus and Chow Neng Choong from G-Energy Singapore for sharing expert knowledge and information throughout this project. This project is made possible from the effort put in by both.

References

[1] J. Luther and T. Reindl, Solar Photovoltaic (PV) Roadmap for Singapore (A Summary). Singapore: Solar Energy Research Institute of Singapore, 2013, pp. 29-31
[2] C.P. Au-Yong, A.S. Ali, F. Ahmad, Improving occupants' satisfaction with effective maintenance management of HVAC system in office buildings, Autom. Constr. 43 (2014) 31–37.
[3] L. Perez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (2008) 394–398.
[4] S. Wu, K. Neale, M. Williamson, M. Hornby, Research opportunities in maintenance of office building services systems, J. Qual. Maint. Eng. 16 (1) (2010) 23–33.
[5] National Environmental Agency, Energy Efficiency Opportunities Assessment and Energy Performance Measurement Guidelines for New Ventures (2018)
[6] J.-S. Chou and D.-K. Bui, "Modelling heating and cooling loads by artificial intelligence for energy-efficient building design," Energy and Buildings, vol. 82, pp. 437–446, 2014.
[7] A. I. Dounis, "Artificial intelligence for energy conservation in

buildings," Advances in Building Energy Research, vol. 4, no. 1, pp. 267–299, 2010.

[8] M. Kim, S. Park, J. Lee, Y. Joo, and J. Choi, "Learning-based adaptive imputation method with KNN algorithm for missing power data," Energies, vol. 10, no. 10, 2017.

[9] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ann and svm for building electrical energy consumption forecasting," Renewable

[10] T. Hastie and R. Tibshirani, Generalized Additive Models. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1990.

[11] K. K. Sumer, O. Goktas, and A. Hepsag, "The application of seasonal latent variable in forecasting electricity demand as an alternative method," Energy policy, vol. 37, no. 4, pp. 1317–1322, 2009.

[12] J. Ploennigs, B. Chen, A. Schumann, and N. Brady, "Exploiting generalized additive models for diagnosing abnormal energy use in buildings," in Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, pp. 1–8, ACM, 2013.

[13] B. Chen, M. Sinn, J. Ploennigs, and A. Schumann, "Statistical anomaly detection in mean and variation of energy consumption," in Pattern Recognition (ICPR), 2014 22nd International Conference on, pp. 3570–3575, IEEE, 2014.

[14] S. J. Taylor and B. Letham, "Forecasting at Scale," The American Statistician, vol. 72 (1), pp. 37-45, 2017.