The Innovation Hub

for Affordable Heating and Cooling

Knowledge Sharing Report #002
PV Analytics and Energy Analytics Development

Improving the accuracy of PV analytics and energy analytics in buildings using open asset standards and data platform integration

19 November 2021

VBIS, CSIRO, PrediQ

About i-Hub

The Innovation Hub for Affordable Heating and Cooling (i-Hub) is an initiative led by the Australian Institute of Refrigeration, Air Conditioning and Heating (AIRAH) in conjunction with CSIRO, Queensland University of Technology (QUT), the University of Melbourne and the University of Wollongong and supported by Australian Renewable Energy Agency (ARENA) to facilitate the heating, ventilation, air conditioning and refrigeration (HVAC&R) industry's transition to a low emissions future, stimulate jobs growth, and showcase HVAC&R innovation in buildings.

The objective of i-Hub is to support the broader HVAC&R industry with knowledge dissemination, skills-development and capacity-building. By facilitating a collaborative approach to innovation, i-Hub brings together leading universities, researchers, consultants, building owners and equipment manufacturers to create a connected research and development community in Australia.

This Project received funding from ARENA as part of ARENA's Advancing Renewables Program.
The views expressed herein are not necessarily the views of the Australian Government, and the Australian Government does not accept responsibility for any information or advice contained herein.

Primary Project Partner

The information or advice contained in this document is intended for use only by persons who have had adequate technical training in the field to which the Report relates. The information or advice should be verified before it is put to use by any person. Reasonable efforts have been taken to ensure that the information or advice is accurate, reliable and accords with current standards as at the date of publication. To maximum extent permitted by law, the Australian Institute of Refrigeration, Air Conditioning and Heating Inc. (AIRAH), its officers, employees and agents:

a) disclaim all responsibility and all liability (including without limitation, liability in negligence) for all expenses, losses, damages and costs, whether direct, indirect, consequential or special you might incur as a result of the information in this publication being inaccurate or incomplete in any way, and for any reason; and

b) exclude any warranty, condition, guarantee, description or representation in relation to this publication, whether express or implied.

In all cases, the user should be able to establish the accuracy, currency and applicability of the information or advice in relation to any specific circumstances and must rely on his or her professional judgment at all times.

Improving the accuracy of PV analytics and energy analytics in buildings using open asset standards and data platform integration

Large scale adoption of technology in buildings has resulted in availability of large volume of data on the performance of HVAC and renewable generation assets. This data can be used for predicting availability of onsite generation, effectively manage HVAC operations and support decarbonisation of buildings.

While operational data exists, application of data driven analytics for accurate predictions will still require customisation of methods/models to specific systems and assets.

This project proposes the utilisation of the VBIS asset classification open standard and link them with Data Clearing House based semantic models to demonstrate the benefits of integrated asset and building tagging systems to be used by PrediQ's analytics solutions.

Lead organisation

VBIS

Project commencement date

06 May, 2021

Completion date

30 June, 2021

Date published

19 November 2021

Contact name

T.K Wang

Email

tkwang@vbis.com.au

Project website

https://www.ihub.org.au/dch7-improving-the-accuracy-of-pv-analytics-and-energy-analytics-in-buildings-using-open-asset-standards-and-data-platform-integration/

Table of contents

1. EXECUTIVE SUMMARY 8

2. DATA INGESTION 9

2.1 Buildings 9

2.2 Network Structure 13

2.3 Data Collection Completion 13

3. APPLICATION 1 – PV ANALYTICS 15

3.1 Data Query 15

3.2 PV System Overview 15

3.3 Data Analytics 16

3.3.1 API/ Data Source & Quality Overview 16

3.3.2 Microinverter Data and Groups 16

3.3.3 Comparing Performance 20

3.3.4 PV Forecasting Model 25

3.3.5 Reproducibility of PV Analytics & Visualisations 26

4. APPLICATION 2 – ENERGY ANALYTICS 28

4.1 Data Query 28

4.2 Energy Profile - Clustering 28

4.2.1 Introduction 28

4.2.2 Dataset 28

4.2.3 Feature 29

4.2.4 Experiments & Mythologies 29

4.2.5 Optimal Results 29

4.2.6 Observations 31

4.2.7 Reproducibility 31

4.2.8 Library at the Dock System Clustering 31

4.3 Energy Consumption Forecasting Model 38

4.3.1 Introduction 38

4.3.2 Dataset 38

4.3.3 Feature Engineering 38

4.3.4 Experiment 39

4.3.5 Observations 41

4.3.6 Notes and Future Improvement 41

4.3.7 Reproducibility 41

4.4 Thermal Models from BIM Models 42

4.5 Energy Optimisation Models 44

4.6 Carbon Neutral Models 47

5. APPLICATION 3 – BUILDING PERFORMANCE ANALYTICS 49

5.1 Data Query 49

5.2 Building Performance Models 49

5.3 BIM 53

6. APPLICATIONS USER INTERFACE OVERVIEW 54

7. CONCLUSION 58

7.1 Use of BRICK Model and VBIS Tags 58

7.2 Analytical Summary 58

7.3 Uncertainties 58

List of tables and figures

Figure 2.2.1: DCH7 network topology 13

Figure 2.3.1: Data Collection Points of Boyd School 14

Figure 2.3.2: Data Collection Points of Fitzroy Garden Service Centre 14

Figure 2.3.3: Data Collection Points of East Melbourne Library 14

Figure 2.3.4: Data Collection Points of Lady Huntingfield 15

Figure 3.3.2.1: Microinverter Group - East Melbourne Library 17

Figure 3.3.2.2: Missing values - East Melbourne Library 17

Figure 3.3.2.3: Distribution of Daily Energy Production - East Melbourne Library 18

Figure 3.3.2.4: Monthly Maximum Generations - East Melbourne Library 19

Figure 3.3.2.5: Microinverter/ Solar Panel Groupings for Fitzroy Gardens and Lady Huntingfield 20

Figure 3.3.2.6: Carlton Baths Solar Panel Grouping/ Clustering Performance 20

Figure 3.3.3.1. Comparing Actual and Designated Electricity Generation of East Melbourne Library 21

Figure 3.3.3.2. The gap between Actual and Designated Electricity Generation of East Melbourne Library 21

Figure 3.3.3.3: Warranty proposed performance 22

Figure 3.3.4.4: Data Quality & Generations Visual 22

Figure 3.3.3.5: Monthly total performance vs Warranty 23

Figure 3.3.3.6: Comparing PV Energy Generation in Similar Suburbs 24

Figure 3.3.4.1. The framework of the prediction model 25

Figure 3.3.4.2: Prediction error in the training and testing dataset 25

Figure 3.3.4.3: Measured and predicted power in the testing dataset 26

Figure 3.3.4.4: Data Quality & Generations Visual 27

Figure 3.3.4.5: Low-Performance Identification & Loss Assessment 27

Figure 4.4.1: Detailing of thermal models 42

Figure 4.4.2: BIM Creation – Lady Huntingfield 42

Figure 4.4.3: Enclosed structures setting 43
Figure 4.4.4: Internal gains setting 43
Figure 4.4.5: Weather data setting 44

1. EXECUTIVE SUMMARY

This knowledge share report details the development of applications deployed in the i-Hub sub project DCH7.

The applications developed for this project focuses on PV analytics and Energy analytics for use across 5 City of Melbourne facilities connected to the i-HUB Data Clearing House (DCH). The 5 sites are Boyd School Community Hub, Lady Huntingfield Children's Centre, Fitzroy Gardens Visitors Centre, East Melbourne Library and Library at the Dock. Each site was selected due to the availability of solar onsite generation as well as having a deployed building management system (BMS).

The following sections provide details on the methodology used to develop the applications, data points which will be retrieved from DCH along with the data techniques used to model each building.

2. DATA INGESTION

PrediQ carries out data collection services on the 5 project buildings - Boyd Community Hub, Lady Huntingfield Children's centre, Fitzroy Gardens Visitor Centre, East Melbourne Library, and Library at the Dock. PrediQ accesses data from building BMS systems, and forward the live data to DCH through the MQTT gateway. PV system data is obtained through third-party APIs and uploaded to DCH via MQTT in the same manner.

2.1 Buildings

Building 1: Boyd School Community Hub	
Location:	207-227 City Road, Southbank
Purpose:	School
Size:	950m2
PV System:	20kWp

Building 2: Lady Huntingfield Children's Centre	
Location:	89-97 Haines St, North Melbourne VIC 3051
Purpose:	School
Size:	Unknown
PV System:	40.26kWp

Building 3: Fitzroy Gardens Visitor Centre	
Location:	Fitzroy Gardens, Wellington Parade, East Melbourne VIC 3002
Purpose:	Info Centre
Size:	Unknown
PV System:	27.04kW

Building 4: East Melbourne Library	
Location:	122-126 George St, East Melbourne VIC 3002
Purpose:	Library
Size:	715m2
PV System:	10.4kWp

Building 5: Library at the Dock	
Location:	107 Victoria Harbour Promenade, Docklands VIC 3008
Purpose:	Library
Size:	2389m2
PV System:	85kWp

2.2 Network Structure

Below is the diagram of the data collection topology. After DCH receives the data, query language provided by CSIRO can be used to query PrediQ's application cloud (using AWS) and perform modelling processes.

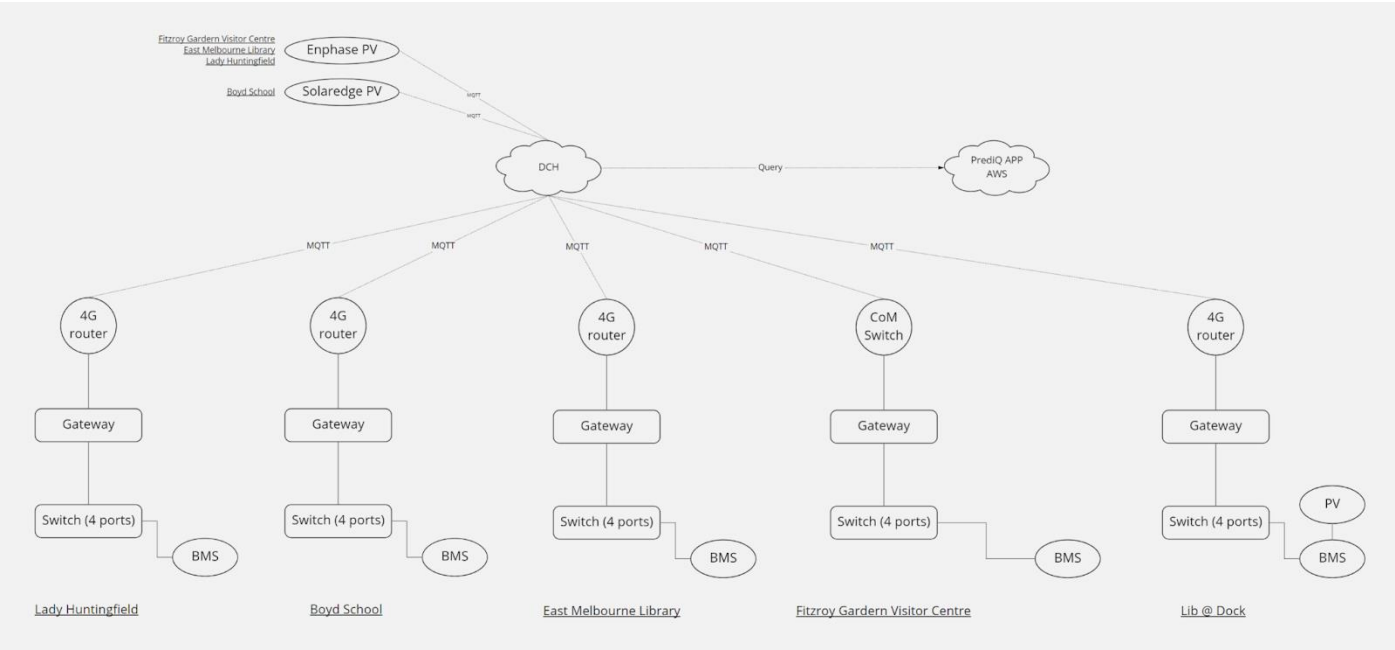


Figure 2.2.1: DCH7 network topology

Gateway with TLS function is selected for this project. Transport Layer Security (TLC) is a cryptographic protocol that provides end-to-end security of data sent between applications over the Internet.

The gateway has bidirectional communication disabled due to the requirement of this project, specifically the inability to modify any building configurations and settings.

As a result of some on-site complexities, 4G routers were installed to connect to the internet.

2.3 Data Collection Completion

BMS (Delta Control System) and PV System (Solar Edge) are the main sources of the building data of Boyd School. Gateway accesses HVAC and metering system data through BMS while data ingestion from the PV system is challenging, a large proportion of the data of the PV system has to be manually inputted.

The modelling process is significantly impacted due to the limited availability of historical data in BMS. For example, Delta BMS can only save 6000 historical numeric values and 200 to 600 historical binary values.

Boyd School

Area:	29
Equipment:	63
Total Points:	725

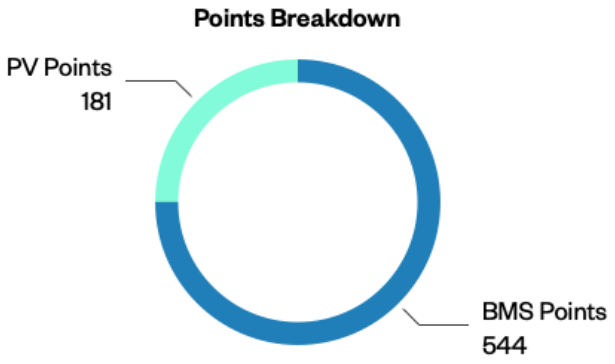


Figure 2.3.1: Data Collection Points of Boyd School

Fitzroy Garden Service Centre

Area:	9
Equipment:	35
Total Points:	611

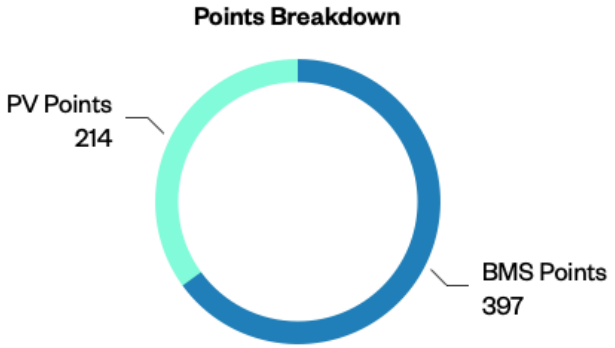


Figure 2.3.2: Data Collection Points of Fitzroy Garden Service Centre

East Melbourne Library

Area:	4
Equipment:	20
Total Points:	251

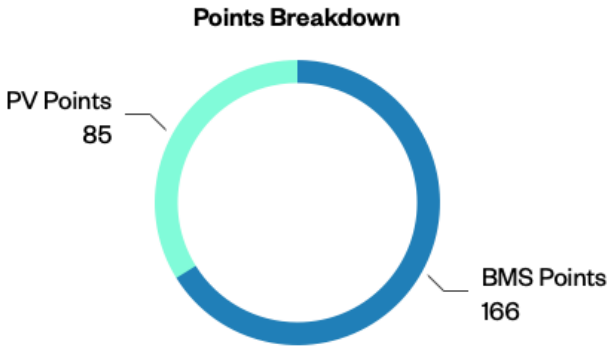


Figure 2.3.3: Data Collection Points of East Melbourne Library

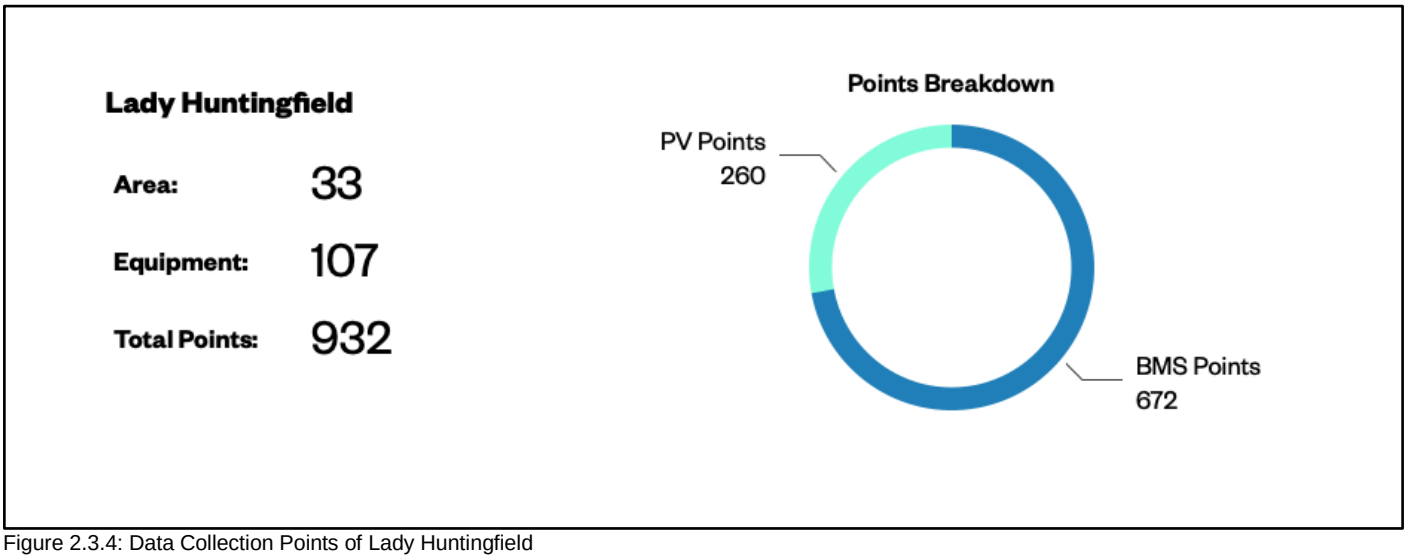


Figure 2.3.4: Data Collection Points of Lady Huntingfield

Note:
Refer to Appendix 1 for full data ingestion record.

3. APPLICATION 1 – PV ANALYTICS

3.1 Data Query

For the application of PV analytics, PrediQ has to query the DCH and VBIS tagging system on:

- the total number of solar panels;
- panel orientations;
- panel locations;
- whether panels are obstructions;
- panel installation angles;
- historical total power generation of all sites as well as the designated power generation;
- the amount of power generation of the respective solar panel type (e.g. single crystal or polycrystalline);
- local weather conditions;
- inverters' power and power generation, and their locations.

3.2 PV System Overview

Building 1: Boyd Community Hub
System Points: Total Energy Production, Total Power Production, Site kWh/kWp
Microinverter Points: 89 modules (Energy & Power Data Available)
Data Date Range: 2013-01-01 to 2020-01-12 and 2021-08-01 to date
API/ Data Source: SolarEdge

Building 2: Lady Huntingfield Childrens Centre
System Points: Total Energy Production & Consumption, Total Power Production & Consumption, Lifetime Energy, Energy Today, Device Reporting
Microinverter Points:122 modules (Energy & Power Data Available)
Data Date Range: 2020-06-05 to date
API / Data Source: Enphase Energy

Building 3: Fitzroy Gardens Visitor Centre
System Points: Total Energy Production, Total Power Production, Lifetime Energy, Energy Today, Device Reporting
Microinverter Points:104 modules (Energy & Power Data Available)

Data Date Range: 2016-06-24 to date
API/ Data Source: Enphase Energy

Building 4: East Melbourne Library
System Points: Total Energy Production, Total Power Production, Lifetime Energy, Energy Today, Device Reporting
Microinverter Points: 40 modules (Energy & Power Data Available)
Data Date Range: 2016-01-08 to date
API/ Data Source: Enphase Energy

Building 5: Library at The Dock
System Points: Total Energy Production, Total Power Production
Microinverter Points: -
Data Date Range: 2019-11-05 to date
API/ Data Source: Aurora Vision

3.3 Data Analytics

3.3.1 API/ Data Source & Quality Overview

Enphase Energy – Lady Huntingfield, Fitzroy Gardens, East Melbourne Library
Due to technical difficulties encountered with the Enphase Energy API, self-developed software is used to access the Enphase Energy Web Portal to obtain data for system and microinverter energy/ power production and device status. The collected data is then simultaneously uploaded to DCH via MQTT and to our database, where further data quality assessments, analysis, and forecasting can be done with ease.

SolarEdge – Boyd Community Hub
Currently, only data for the 4 string inverters can be accessed through the SolarEdge API as a module by module data is not made publicly available. It should also be noted that there is a data shortage period between the periods around 2020-01-12 to 2021-08-01.

Aurora Vision – Library at The Dock
Only total system power and energy production data is available. There exists no module/panel data from this provider, making analysis and forecasting to be nearly impossible and extremely inaccurate.

3.3.2 Microinverter Data and Groups

There are two groups of microinverters for East Melbourne Library, 26 in group J (Southward) and 14 in group I (Northward).

Figure 3.3.2.1: Microinverter Group - East Melbourne Library

Energy data is collected from 2016-01-08 to 2021-08-10 (2042 days). With 40 microinverters, 81680 entries are expected. Totalling 1066 missing entries, resulting in a rate of 1.3% missing data. However, 554 days (27.13%) contain at least one or more missing data points. It was observed that all missing values only originated from 4 inverters. Their IDs, missing values, and distributions are labelled below:

Missing Values

	year	month	whna
48	2020	1	2
49	2020	2	24
50	2020	3	31
51	2020	4	30
52	2020	5	31
53	2020	6	38
54	2020	7	62
55	2020	8	62
56	2020	9	62
57	2020	10	62
58	2020	11	60
59	2020	12	62
60	2021	1	62
61	2021	2	63
62	2021	3	62
63	2021	4	62
64	2021	5	64
65	2021	6	85
66	2021	7	107
67	2021	8	35

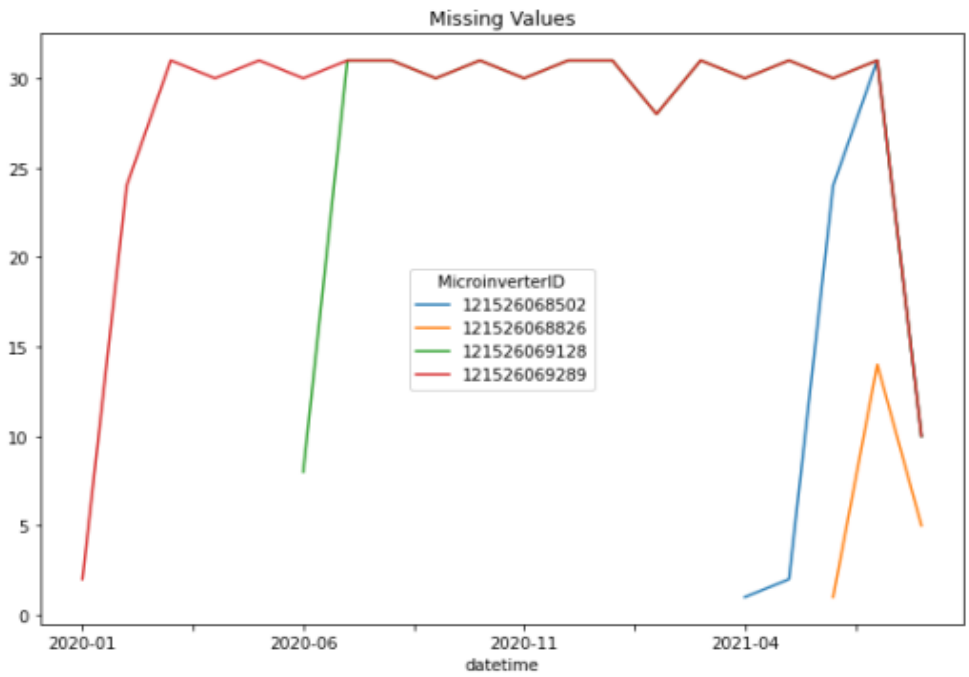


Figure 3.3.2.2: Missing values - East Melbourne Library

It observed that distributions of energy production between the two groups are very similar, this is much anticipated as the two groups are relatively close to each other and have the same orientation.

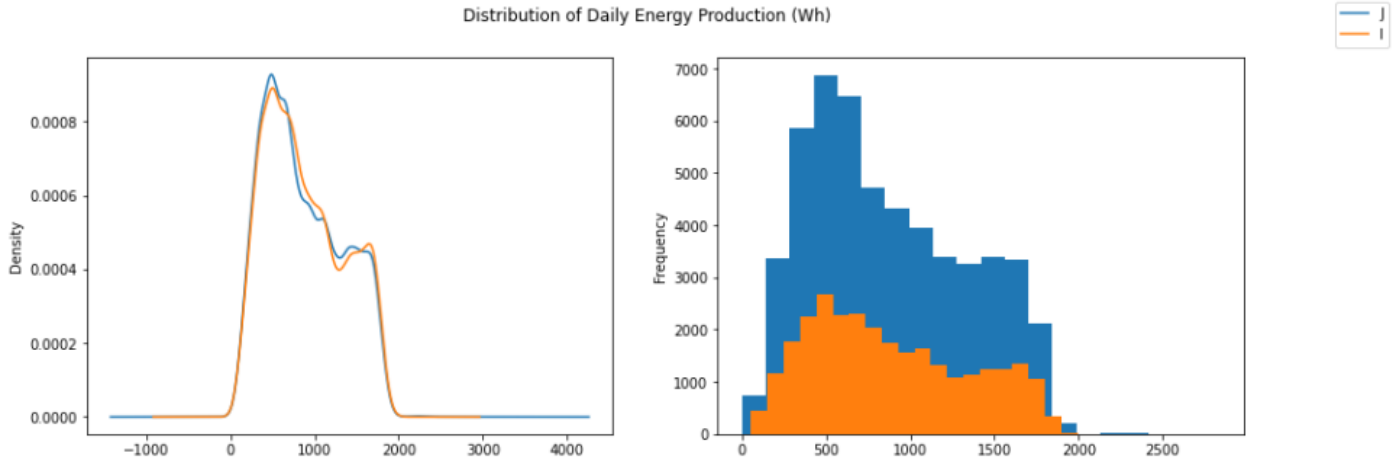
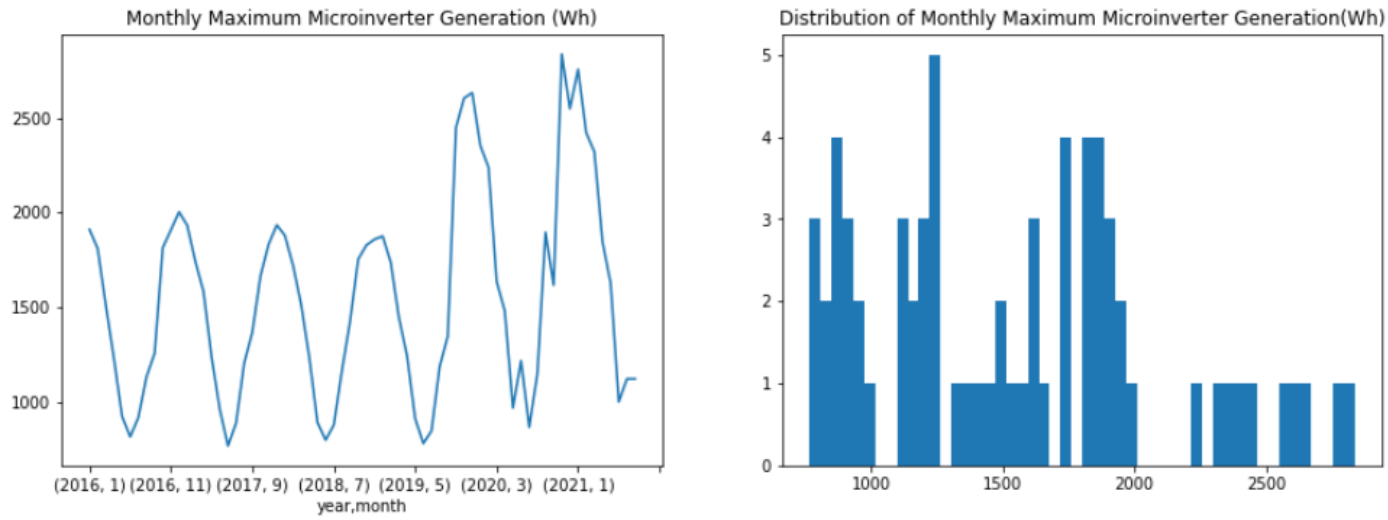


Figure 3.3.2.3: Distribution of Daily Energy Production - East Melbourne Library

The maximum microinverter energy production beyond 2019 Oct does not seem normal. Using an outlier cut-off of 2003, we have again identified the select few troubling microinverters



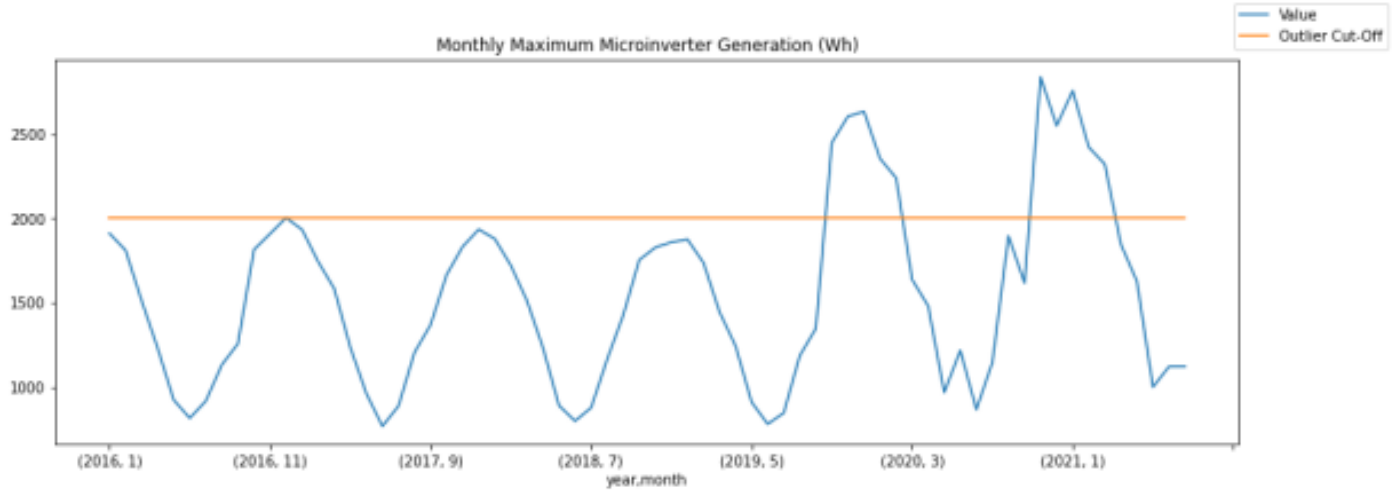


Figure 3.3.2.4: Monthly Maximum Generations - East Melbourne Library

Due to the nature of the microinverter ids being fixed points, it was very easy to use Euclidean distance and cluster the microinverters/ panels together into groups that are representative of their performance and generations. The reproducibility of this grouping method is also tested on Carlton Baths with outstanding results.

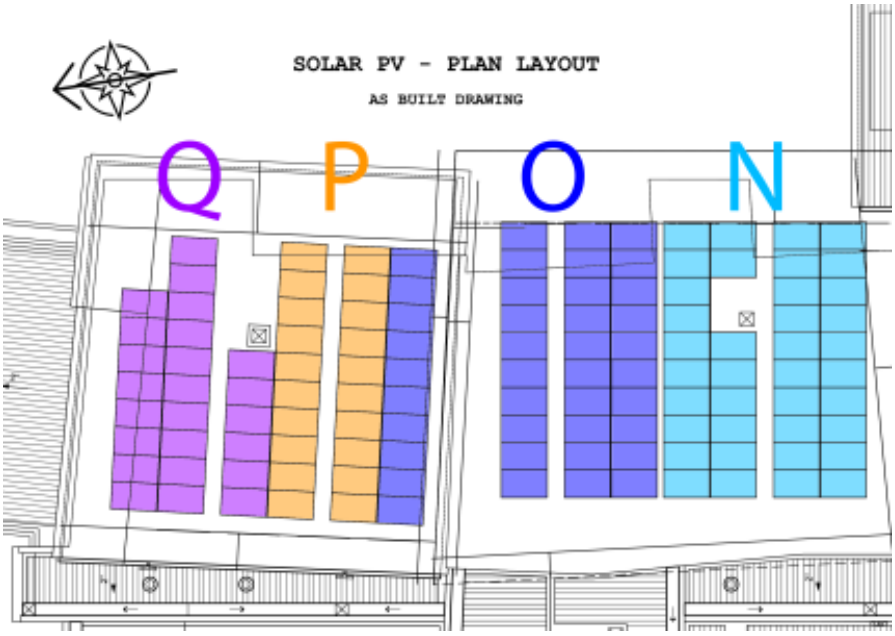


Figure 3.3.2.5: Microinverter/ Solar Panel Groupings for Fitzroy Gardens and Lady Huntingfield

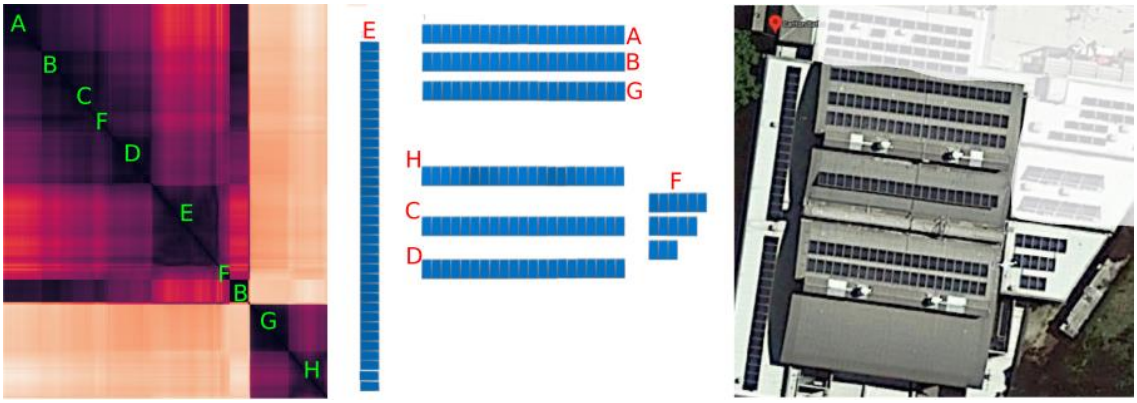


Figure 3.3.2.6: Carlton Baths Solar Panel Grouping/ Clustering Performance

3.3.3 Comparing Performance

After importing several years' electricity generation data, the data is cleaned and noisy data is removed. Various comparisons are carried out, including:

1. Comparing Actual Electricity Generation and Designated Electricity Generation

The graph below shows the actual electricity generation and the designated electricity generation of East Melbourne Library. It is observed that the actual electricity generations are all below the designated generations, which could be caused by weather and/or product discrepancies. The orange bars (undisc_specYield_kWh) represents the figures without the 83.3% discount denoted by Performance Ratio (PR) on the proposal page.

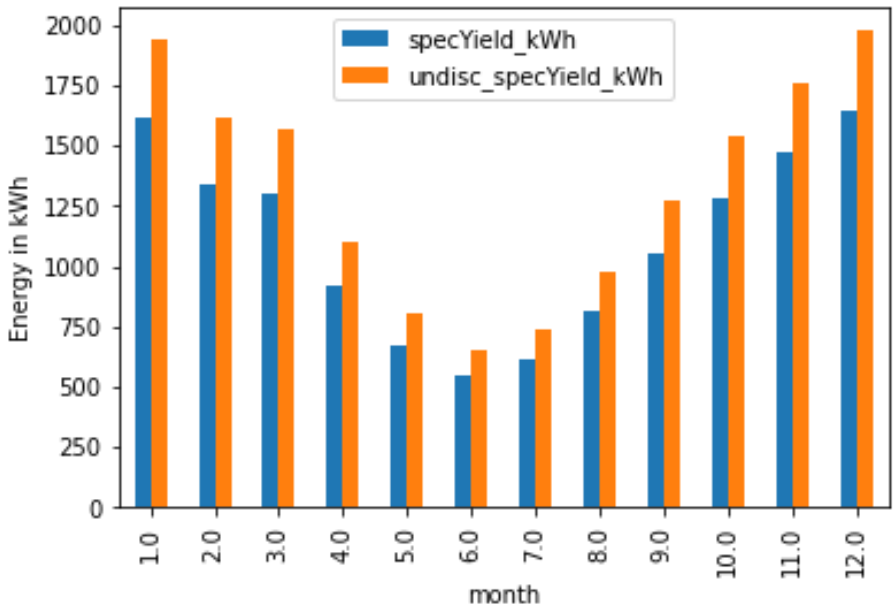


Figure 3.3.3.1. Comparing Actual and Designated Electricity Generation of East Melbourne Library

Combining the graph above, the analytic model concluded that the gaps of the PV systems' anticipated electricity generation claimed during the sales process and their actual electricity generation are as follows:

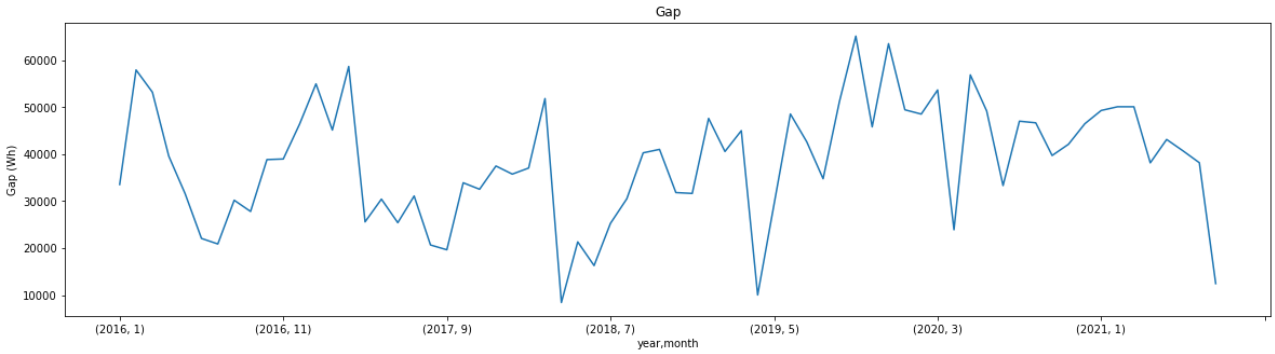


Figure 3.3.3.2. The gap between Actual and Designated Electricity Generation of East Melbourne Library

An approximately 500kWh gap per year is regarded as acceptable. Yet, the other buildings have a larger variation owing to the frequent malfunctioning of the system and shading problems. We are going to quantify these issues in the next stage, so as to provide respective solutions for the clients on solar system installation and maintenance.

2. Comparing Actual Attenuation and Claimed Attenuation by Warranty

The new 25-year linear performance warranty proposed on page 34 of the operation and maintenance manual for City of Melbourne – East Melbourne Library (C284416:3) will be referenced in the following discussion about system performance and this warranty.

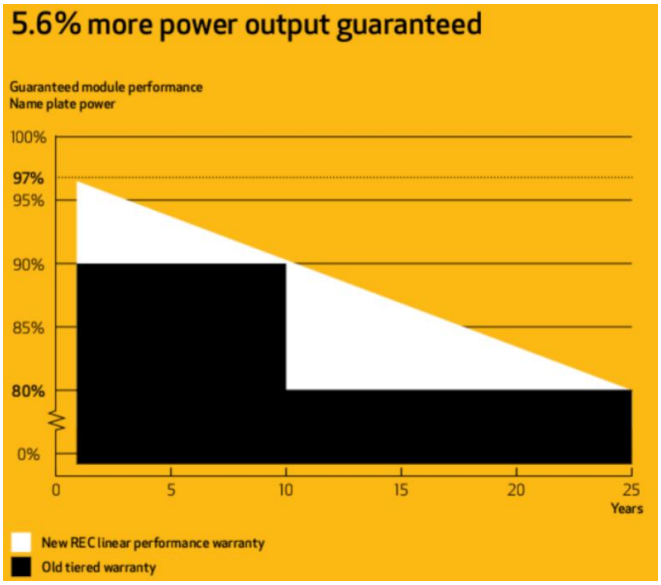


Figure 3.3.3.3: Warranty proposed performance

The reference for 100% guaranteed module performance and continued performance was calculated as follows:
1. Record maximum microinverter generation values of every month
2. Obtain yearly sums of month maximums from above
3. Sum of the first year was used as the 100% referencing point
4. Sums of the following years was then compared

Note:
For step (1), we are not concerned with the maximums over the two different panel groups because their distributions are nearly similar as shown above. This will be different for other systems where groups are not so similar. Though here it is justifiable to treat the monthly maximums over different groups the same as over the whole system. Additionally, as full-year data was not available at the time for 2021, the 100% reference period was Jan - July 2016 instead of the full of 2016, this is only for 2021. The results are showing as below:

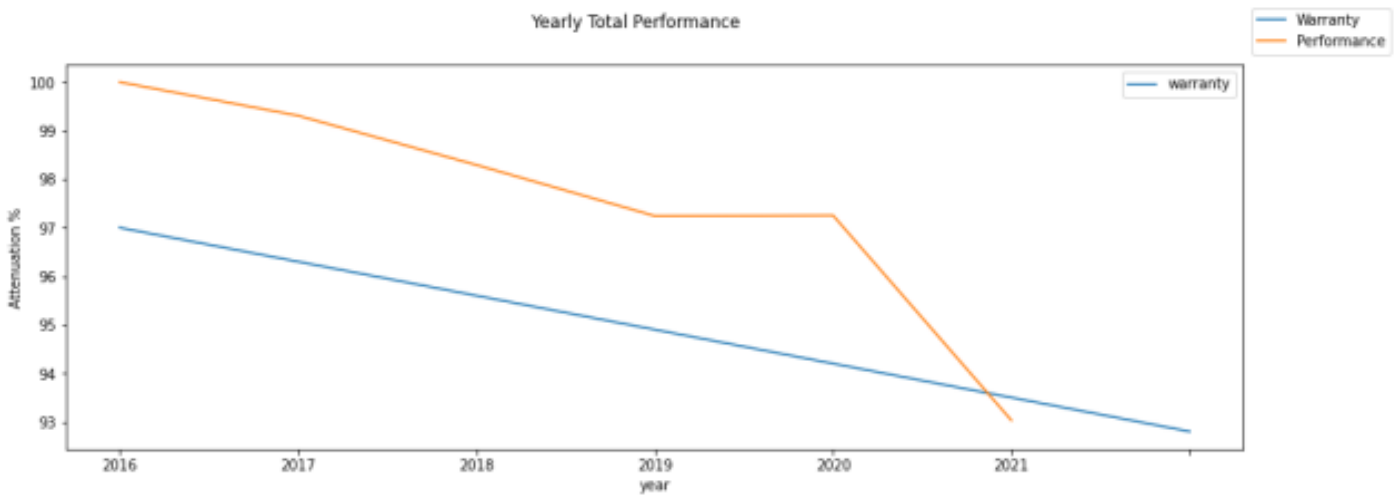


Figure 3.3.3.4: Yearly total performance vs Warranty

Though the point for 2021 is not representative of the whole year, it should still be noted that performance from Jan - July is subpar. Therefore, a monthly analysis is showing as below for more detail:

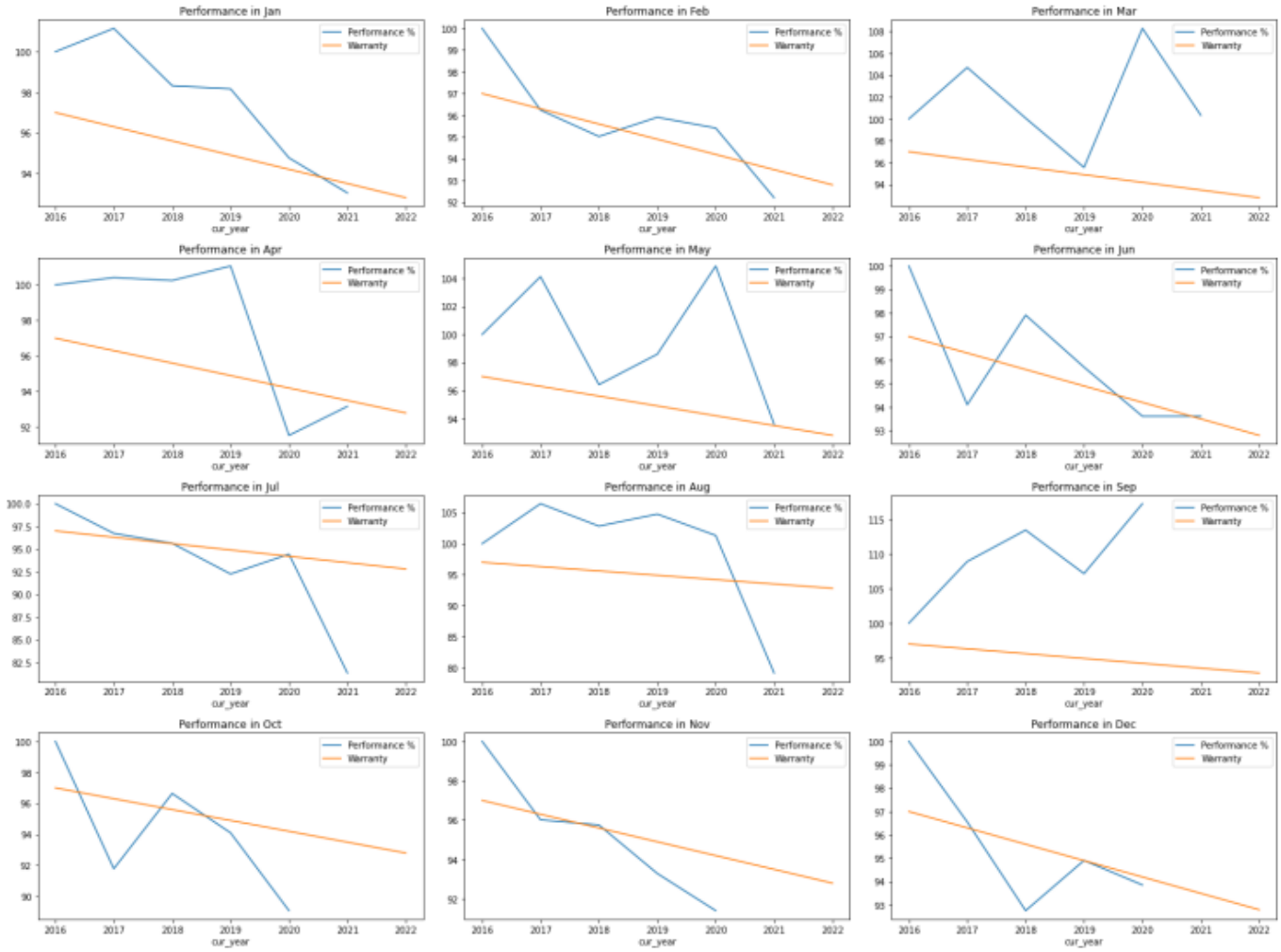


Figure 3.3.3.5: Monthly total performance vs Warranty

It observed that actual attenuation is more than that claimed by the warranty. The weather conditions and data quality accounts for the errors and deviations of the calculation model which will be referenced in later analysis.

3. Comparing PV Energy Generation in similar suburbs

Below is the table of PV performance comparisons between Boyd Community Hub and state performance of Victoria, obtained from pv-map.apvi.org.au.APVI (performance data are sourced at up to 5 minute intervals from more than 6000 PV systems in 57 regions across Australia), also labels 2015-06-16 (top right) as one of the bottom performing/ cloudiest days, and 2020-01-14 as one of the top-performing/sunniest days.

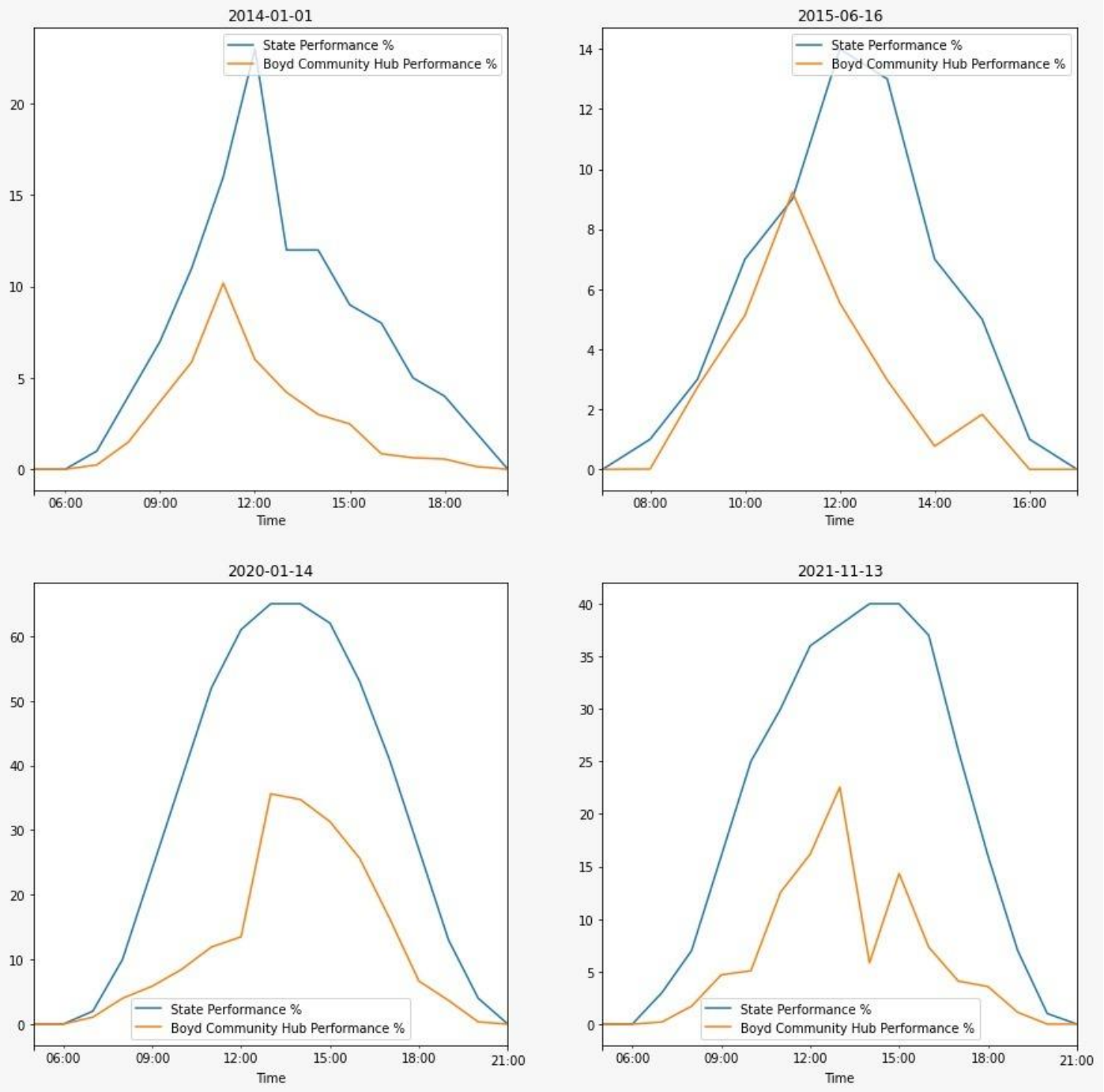


Figure 3.3.3.6: Comparing PV Energy Generation in Similar Suburbs

It observed that the whole PV site performance of Boyd Community Hub is lower compared with Victoria's performance, this could be possibly due to the orientation differences and the nature of the system design. However, more importantly, it is observable that in recent years, performance in the morning has decreased, since there is a tall building beside the East and frequent PV system malfunctioning.

3.3.4 PV Forecasting Model

The PV forecasting model that we tried to build has not delivered the wanted result owing to the incompleteness of weather data. The framework of the prediction model is as follows:

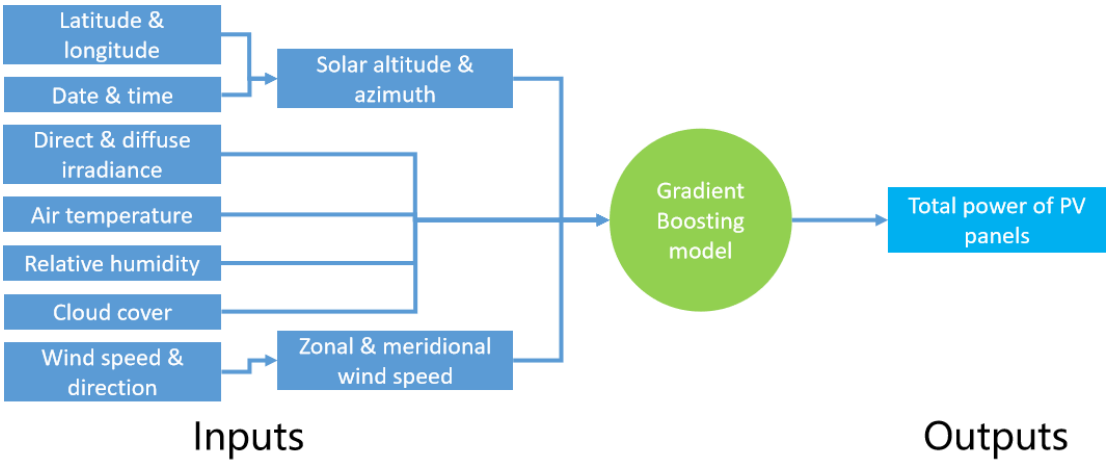


Figure 3.3.4.1. The framework of the prediction model

The following is a more accurate prediction model that we have developed with Lady Huntingfield's historical data:

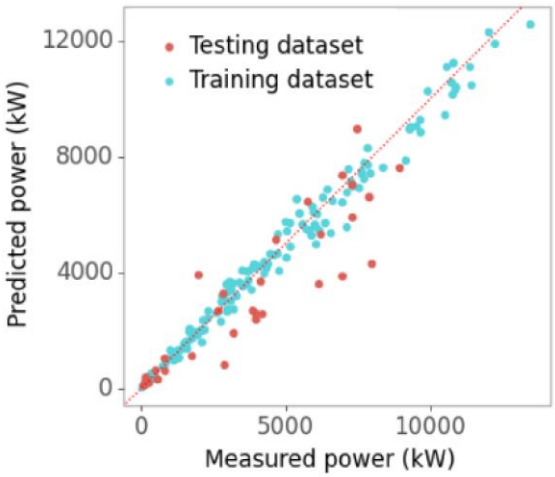


Figure 3.3.4.2: Prediction error in the training and testing dataset

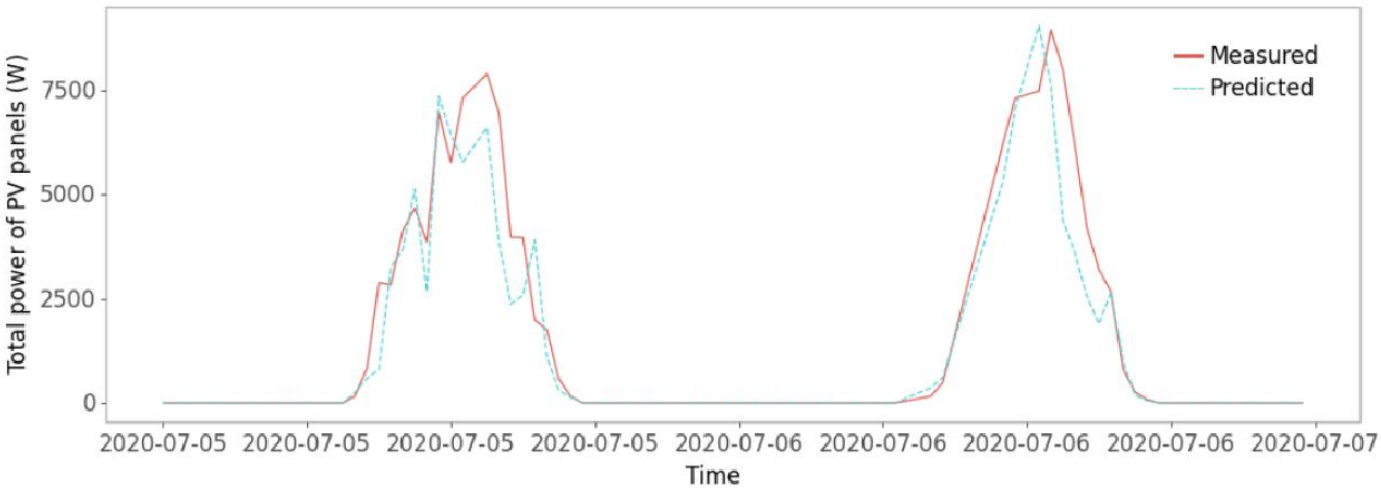


Figure 3.3.4.3: Measured and predicted power in the testing dataset

RMSE of PV power in testing dataset = 1532 W
MAPE of PV power in testing dataset = 16.7%
RMSE of daily PV production in testing dataset = 11.2 kWh
MAPE of daily PV production in testing dataset = 12.9%

The remained errors may be caused by:

a) Lack of training data;
b) Only 2 months' data are used for the model training;
c) The bias of input data;
d) The irradiance data were not collected on the location of the case study building, leading to a possible bias of irradiance value.

3.3.5 Reproducibility of PV Analytics & Visualisations

Now using all the data obtained from the building PV systems from their relevant APIs, which is now readily accessible from the database, combined with the clustering/ groups defined in table 3.3.3.1. We could conduct numerous analytics such as data quality, site energy generation, and assessments such as identifying low performing microinverters and the loss they cause. Some example visuals are indicated below to show the reproducibility of this entire process. (Below charts are internal testing for processing data which are not applications intend to deliver)

Figure 3.3.4.4: Data Quality & Generations Visual

Figure 3.3.4.5 Low-Performance Identification & Loss Assessment

4. APPLICATION 2 – ENERGY ANALYTICS

4.1 Data Query

For the application of energy analytics, PrediQ has to query the DCH and VBIS tagging system on:

- sub-metering architecture and system
- sub-metering historical energy data
- historical gas meter data
- historical water meter data
- local weather data
- HVAC main settings
- HVAC zonings
- room temperature
- room humidity

4.2 Energy Profile – Clustering

4.2.1 Introduction

While dealing with a massive dataset, clustering models understand the client's specific user characteristics and determine possible aspects for optimisation efficiently. Contributing to the development of demand-side response baseline, enhances the accuracy of forecasting models and assesses the feasibility of optimisation solutions.

4.2.2 Dataset

Building: Library at The Dock	
Input	Sum of meter readings
Resolution / Interval	1 hour (between meter readings)
Date Range	22/08/2020 to 21/08/2021
Pre-processing	Extreme values (> 1000 KWH) & days with incomplete readings were removed, resulting in 6912 records
Important Notes	Supplementary humidity and temperature data for the same time period were collected with the intent to guide clustering algorithms. However, the inclusion of such features results in noise and unstable clustering results for this dataset. More experiments can be performed under more data.

Building: Lady Huntingfield Children's Centre	
Input	LH_ElectricMeter_EMG_9(MSB)_14/07/2020-18/08/2021
Resolution / Interval	15 minutes (between meter readings)
Date Range	14/07/2020 to 21/08/2021
Pre-processing	Extreme values (> 1000 KWH) removed
Important Notes	Data quality is heavily influenced by COVID19, resulting in unstable results

4.2.3 Feature

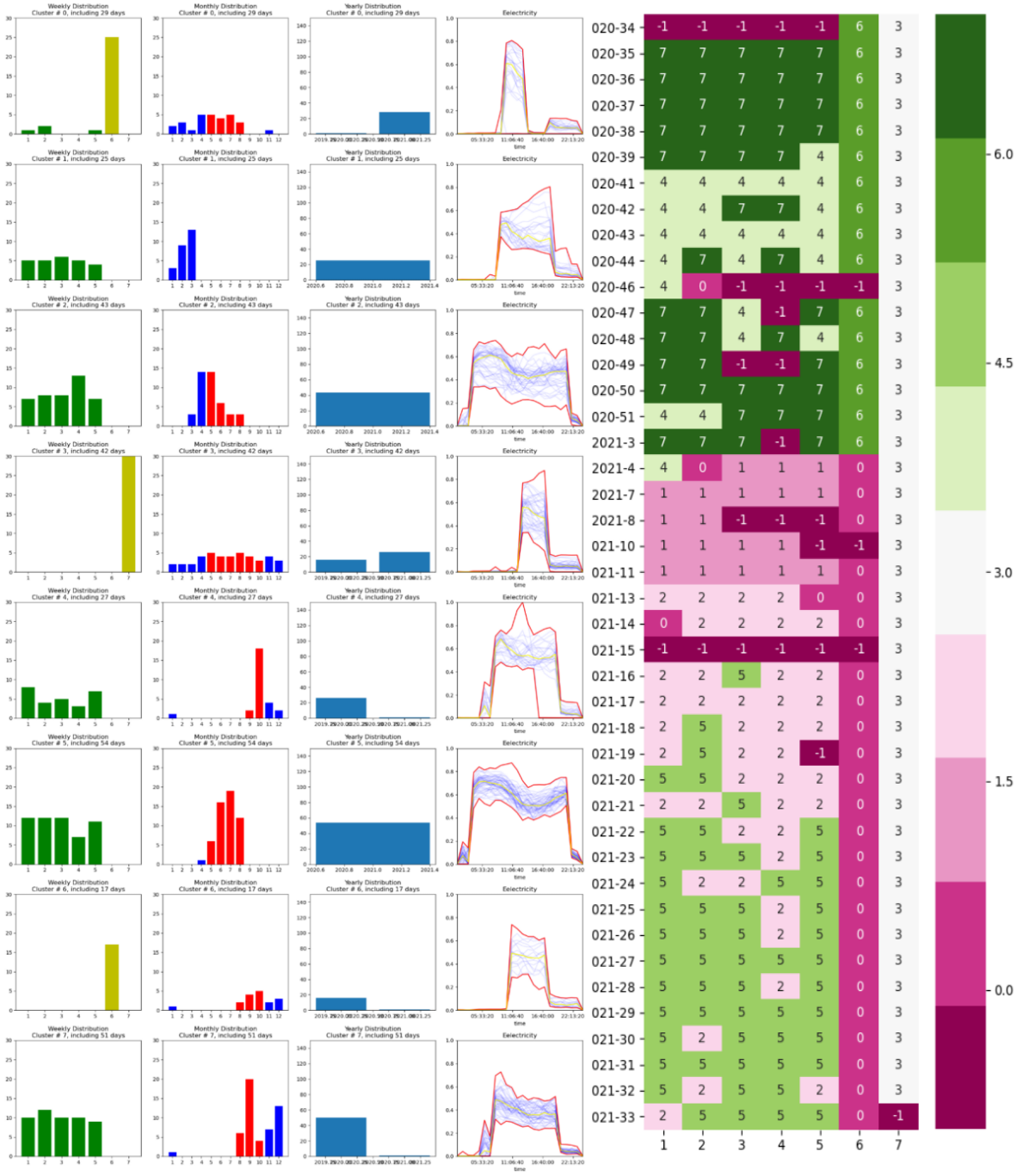
Values of meter readings for one day are concatenated, creating a feature vector representation of that day. Temperature and humidity features are also tested, but unfortunately, results are poorly clustered.

4.2.4 Experiments & Methodologies

A set of different algorithms are tested, including Agglomerative Clustering, Birch clustering, KMeans clustering (with DTW or Euclidean distance function) and KShape clustering. A range of different numbers of clusters (from 8 to 13) is also tested.

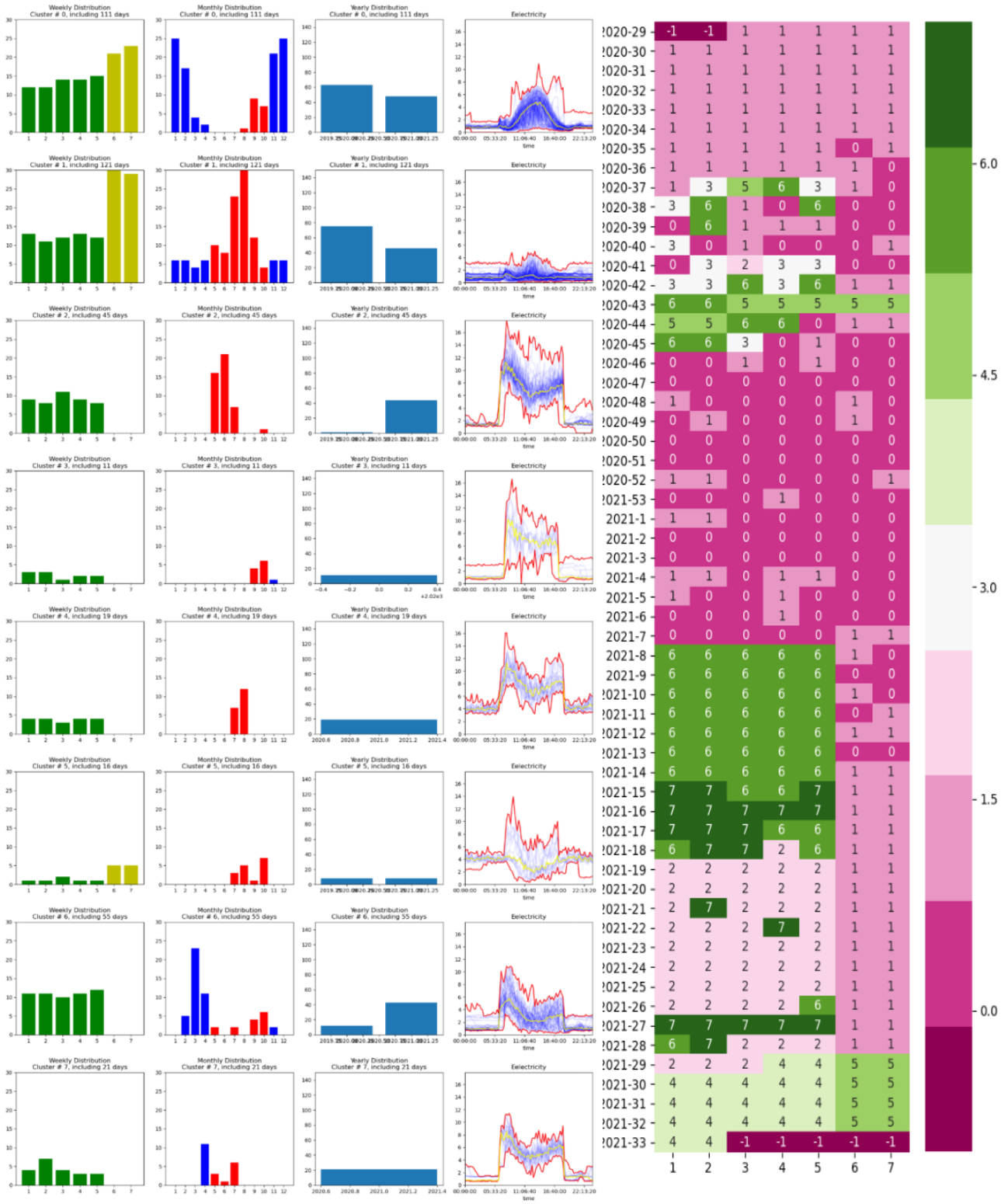
4.2.5 Optimal Results

Agglomerative Clustering with 8 clusters - Library at The Dock



Column one on the left diagram shows the frequency this row (cluster) appears on each day of week. Column two shows the frequency it appears in each month. Column three shows the frequency it appears in each year. The fourth column is the 24-hour curve of the historical data. The diagram on the right outlines the clusters on a calendar.

Lady Huntingfield:



4.2.6 Observations

Algorithm

Agglomerative clustering and KMeans (Euclidean) algorithm gives optimal results.

Number of Clusters

Currently in both datasets, using 8 clusters yields the best result. Increasing the number of clusters results in some empty clusters.

Quality of Results

Library at The Dock

Weekdays and weekends are assigned to different clusters. The candidate clustering model was also able to successfully distinguish seasonality of varying electricity consumption with the complete absence of explicit date information.

Lady Huntingfield Childrens Centre

Weekdays and weekends are not separated by clusters before the eighth week of 2021, a likely cause is due to the lockdown issues from COVID19. After the eighth week of 2021, weekdays and weekends are then properly separated by the unsupervised clustering algorithm.

Notes & Future Improvements

Time zone information and alignment must be double-checked before applying clustering algorithms. Offsets and misalignment in time zones result in unstable and uninterpretable results.

Increasing the number of clusters or designing more fine-tuned features/clustering methods may improve the results, but from our current experiments, we observed that using a larger number of clusters results in small clusters that only contain a few days. Further experiments are also limited by the amount of data. In order to improve the results and test other algorithms, more data is required. More data is not only helpful for clustering but also allows us to design better forecasting models by using it to conduct model pre-training and better evaluate them.

We can also try PCA based clustering methods or Fourier transformed based features for clustering. Another insufficiency is that we need to assign meanings (i.e., weekend, weekdays, etc) to clusters manually. Automated strategies can be developed, if more data is provided.

4.2.7 Reproducibility

The clustering methodology and model was developed by using the dataset of Lady Huntingfield Childrens Centre and then tested on Library at The Dock. We observed a good transferability of the current algorithms. We will be easily able to reproduce the results on other buildings when BRICK schemas and sufficient amounts of data are provided.

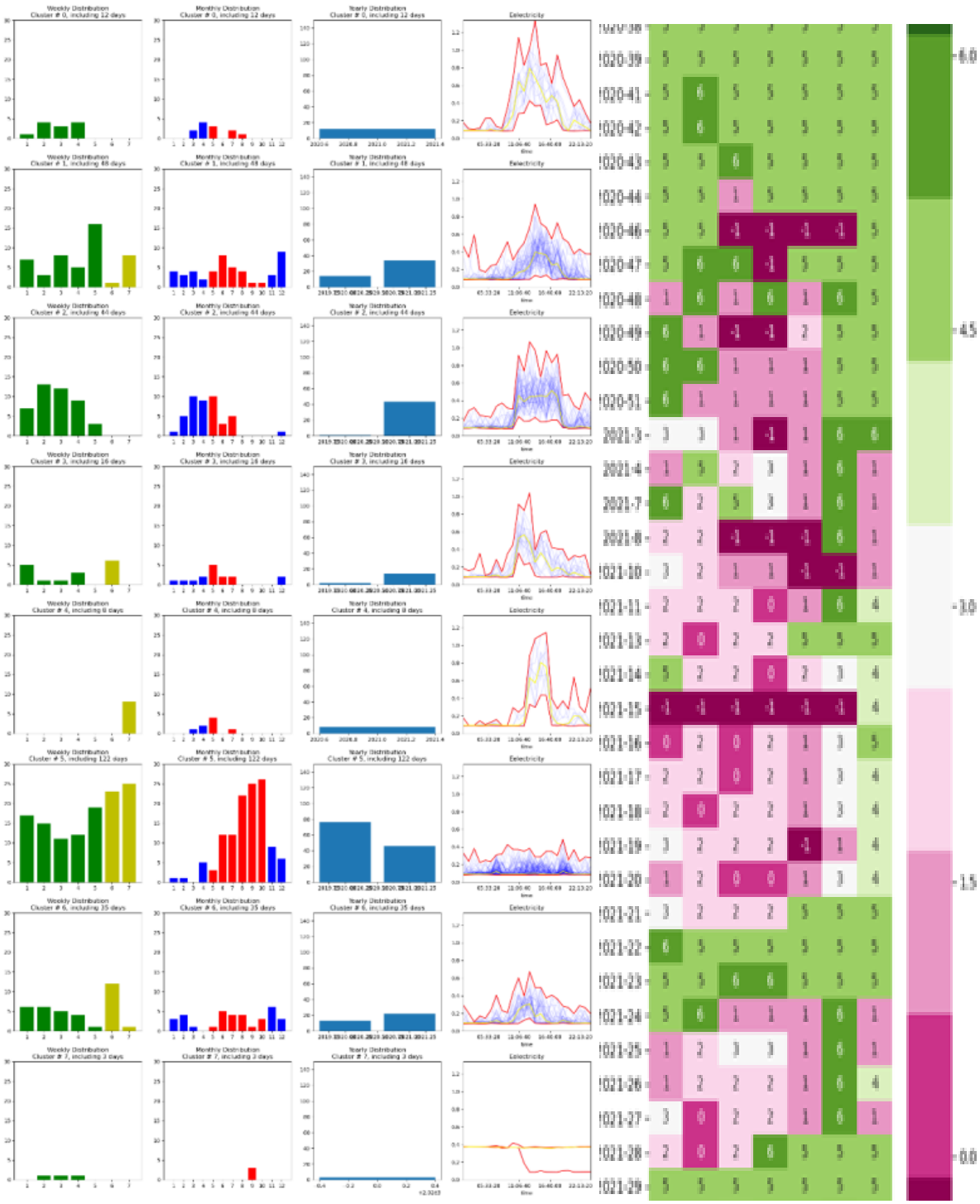
4.2.8 Library at The Dock Subsystems Clustering

Subsystem	Points
Lift	DB1.1 Light, DB1.2 Light, DB2.1 Light, DB2.2 Light, DB3.1 Light, DB3.2Light
Light	DB1.1 Light DB1.2 Light, DB2.1 Light, DB2.2 Light, DB3.1 Light, DB3.2Light
Power	DB1.1 Power, DB1.2 Power, DB2.1 Power, DB2.2 Power, DB3.1 Power, DB3.2 Power
Power & Light	sum of power and light
HVAC	L1 fan coil units + hw pumps, L2 fan coil units, L3 fan coil units + hw pump, L4 outside air AC units, L4 VRV condensers
MCC	MCC Incoming

Diving deeper, clustering algorithms mentioned above are also tested on Subsystems of the Library at The Dock listed above. Using the same methodology, clustering results are generated for these subsystems from 17/08/2020 to 17/08/2021. When there are multiple points within a subsystem, their total sum is used as the clustering vector.

Lift

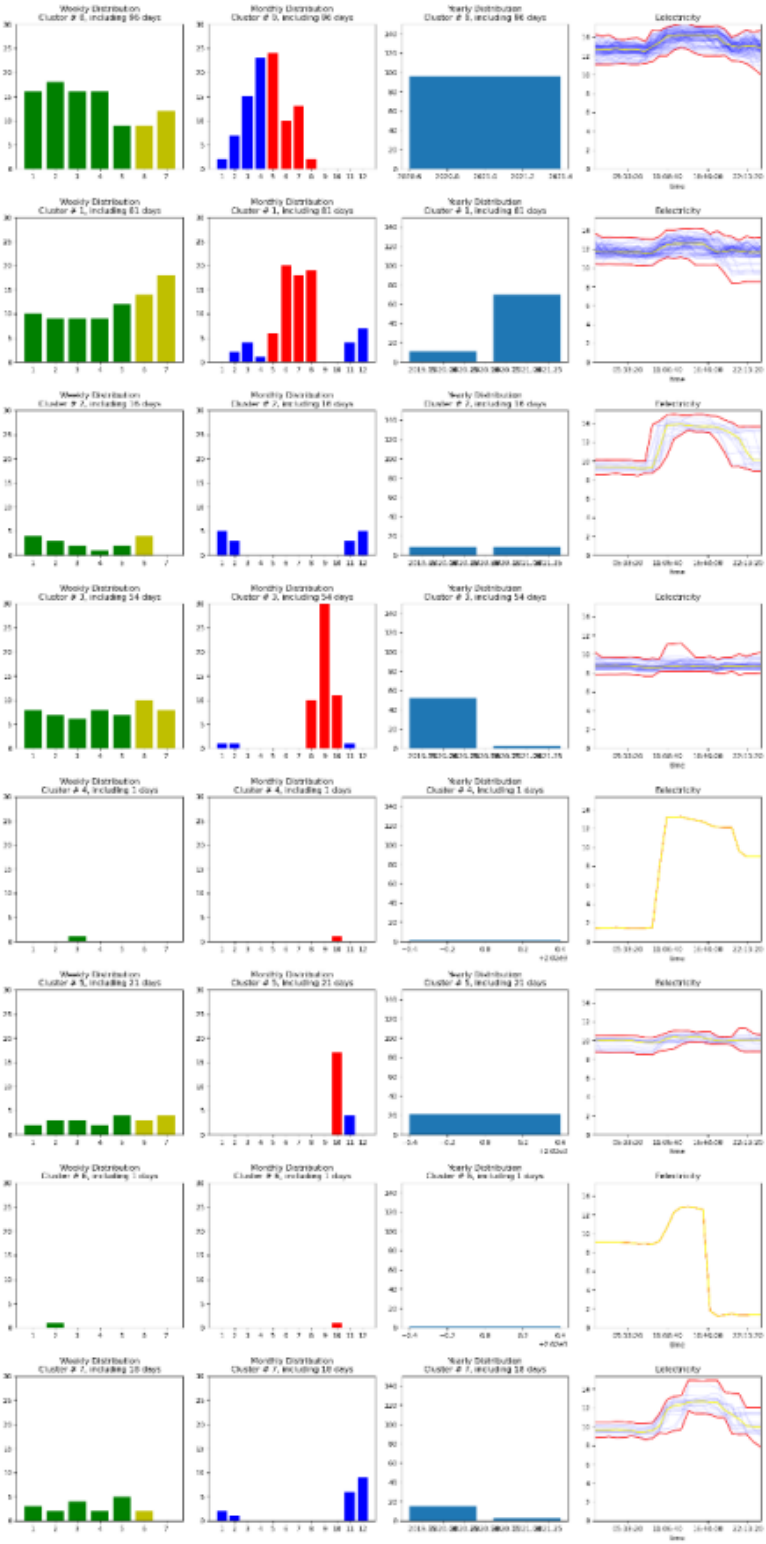
Agglomerative clustering gave the best results with k = 8. Clusters for the weekend strongly overlapped with cluster 5. Many clusters only contained a few days.

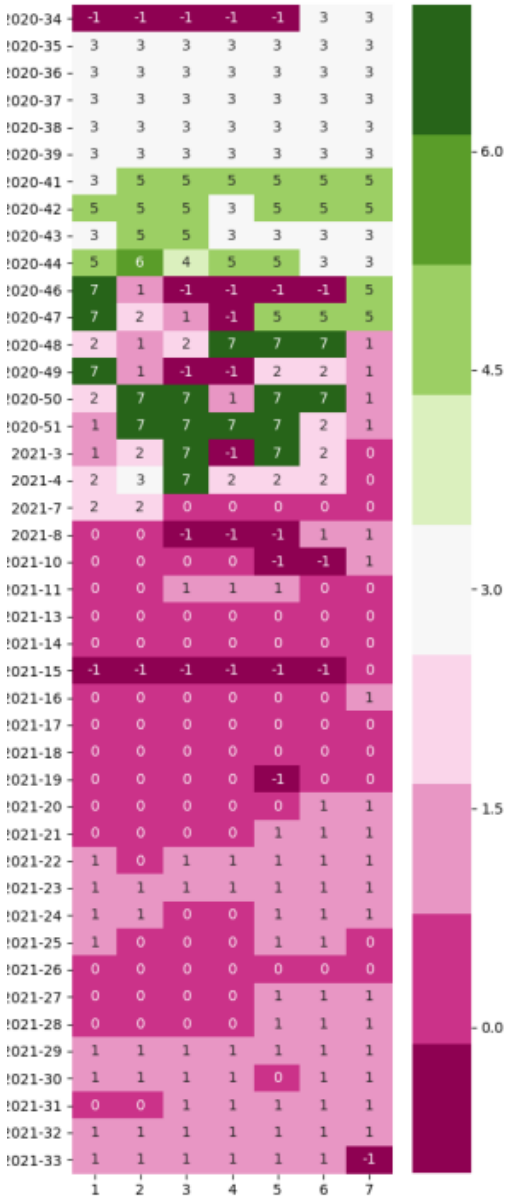


Light

All models gave similar results, except for kShape. No weekdays/weekends relationships observed between clusters. Strong relation between months and seasonality observed in clusters.

Agglomerative clustering results

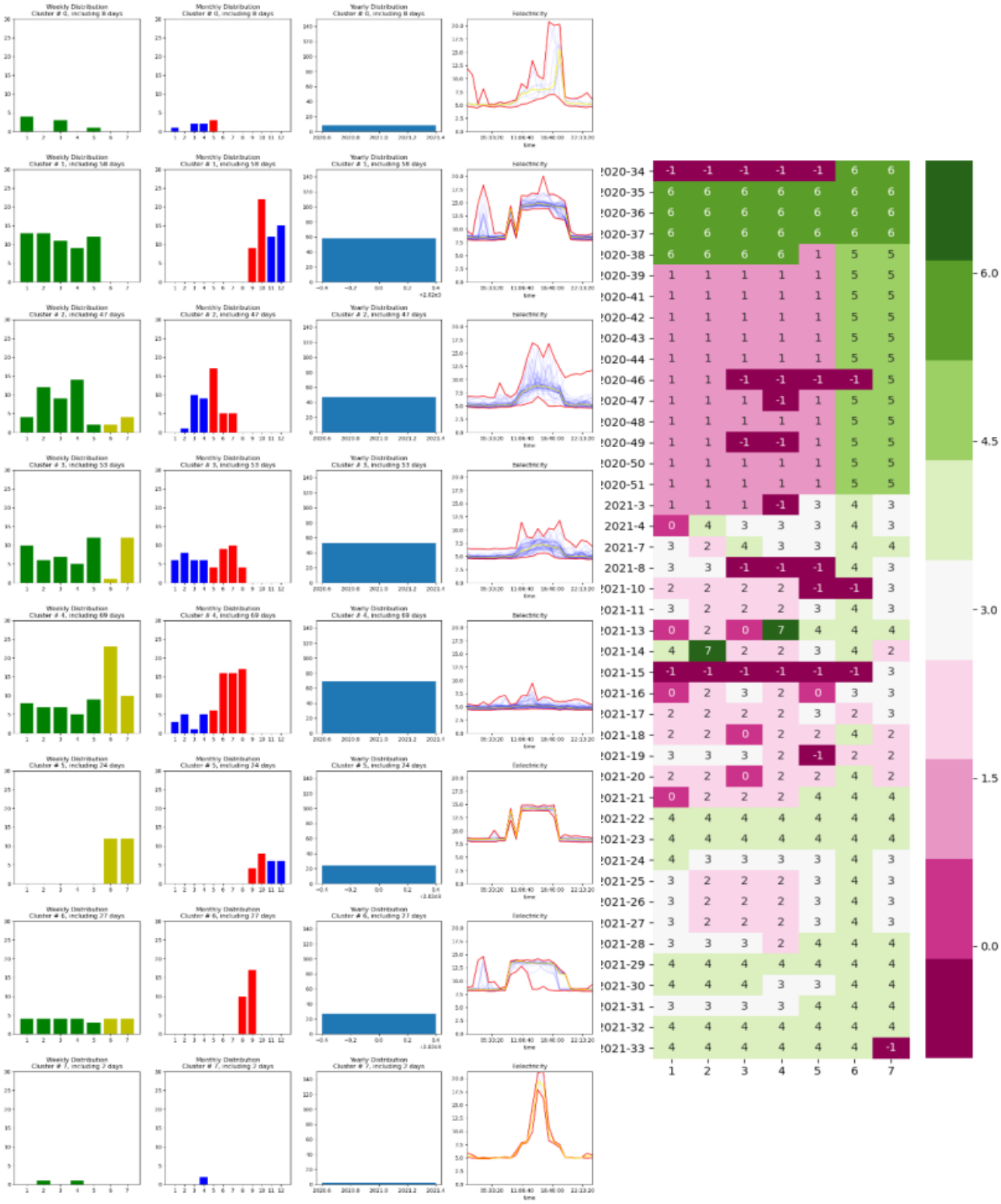




Power

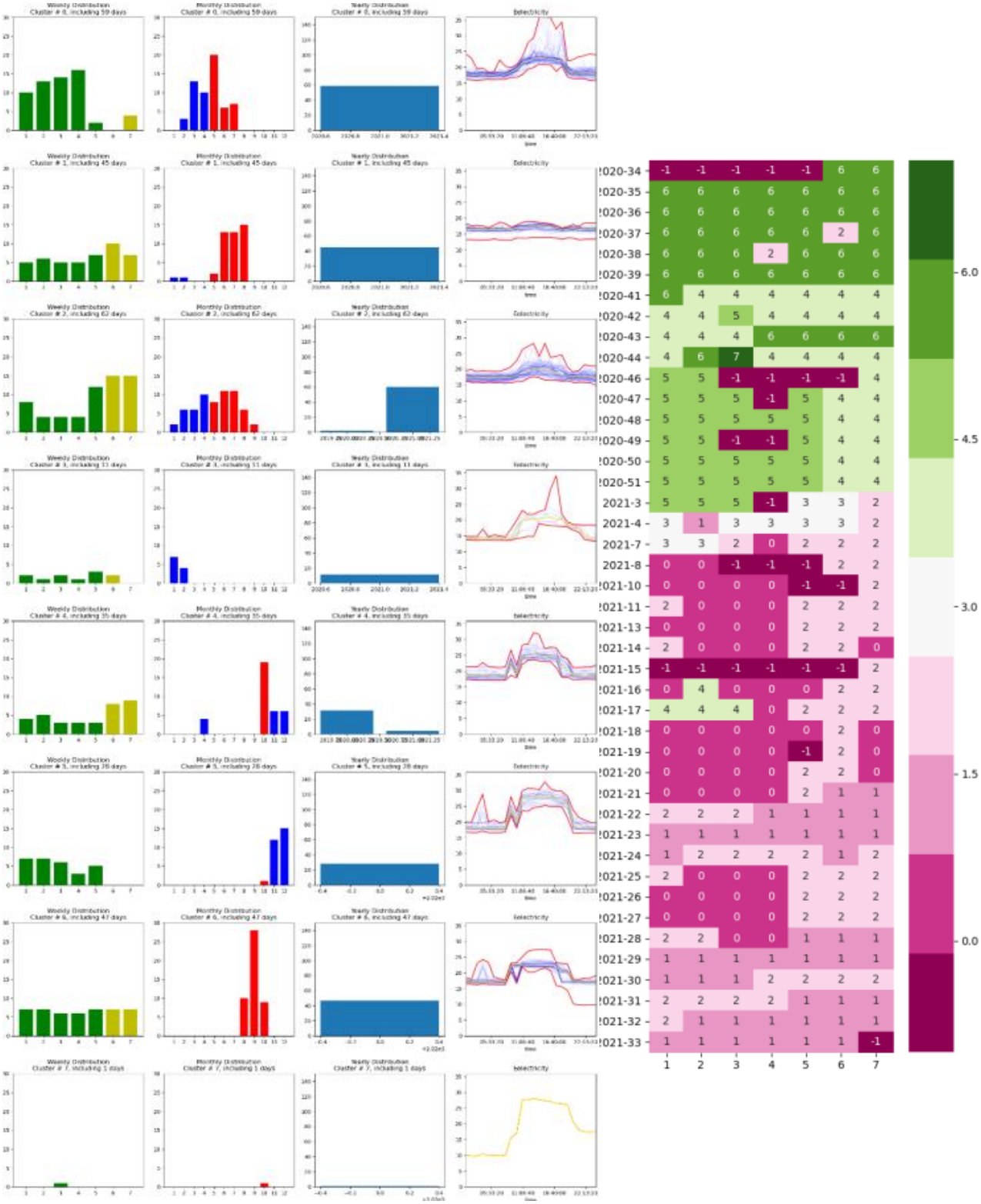
Agglomerative clustering and birch clustering gave good results. Observed relations between summer/ other seasons and clusters. In summer, the relations between weekdays/weekends and clusters are also observed.

Agglomerative clustering results k=8



Power & Light

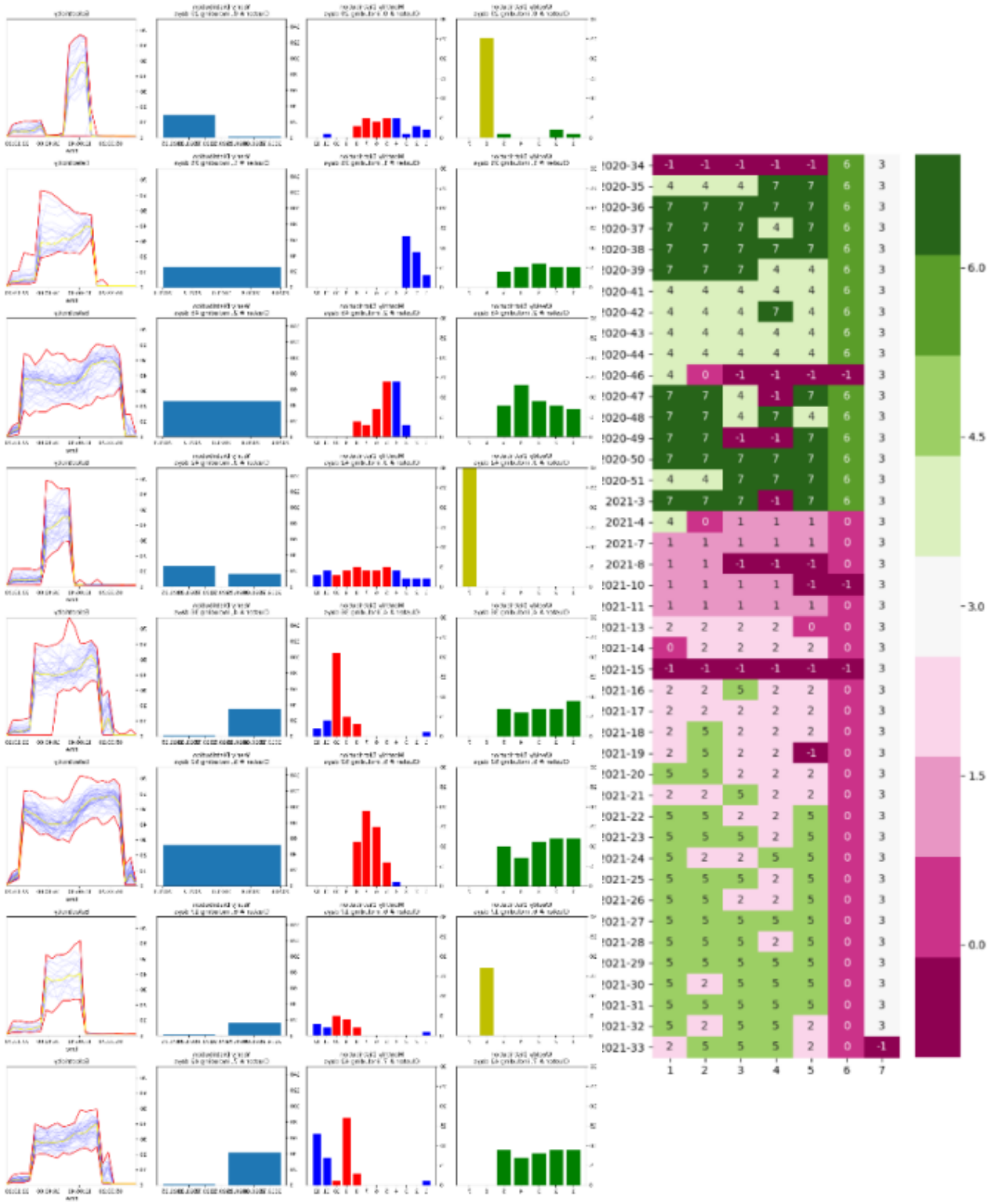
Agglomerative clustering gave best results. Patterns are well-clustered. Relation between months and clusters are observed.



HVAC

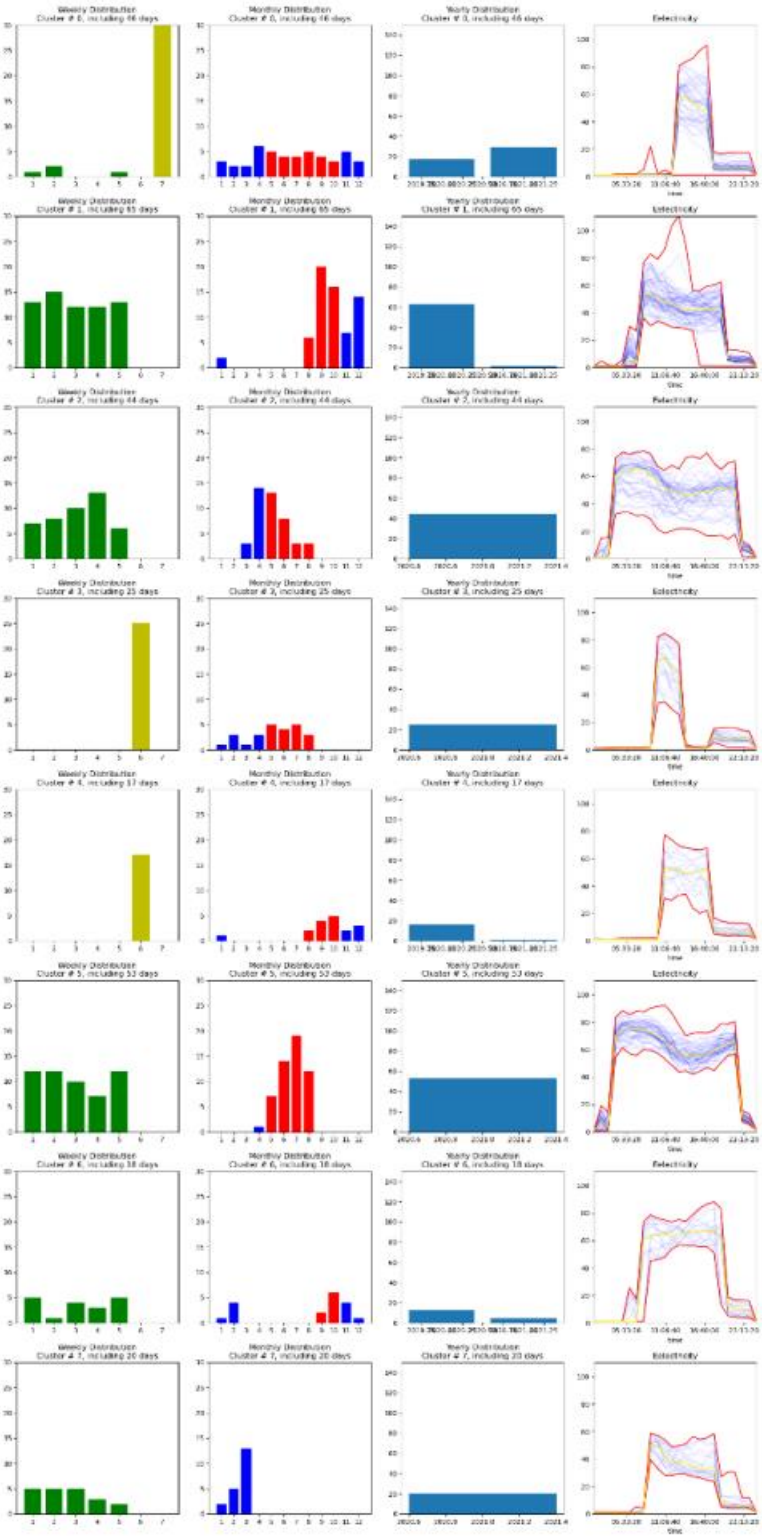
Agglomerative clustering and Birch clustering methods gave good results. We got best results when clusters=8. Patterns of clusters were similar to the sum of electricity usage.

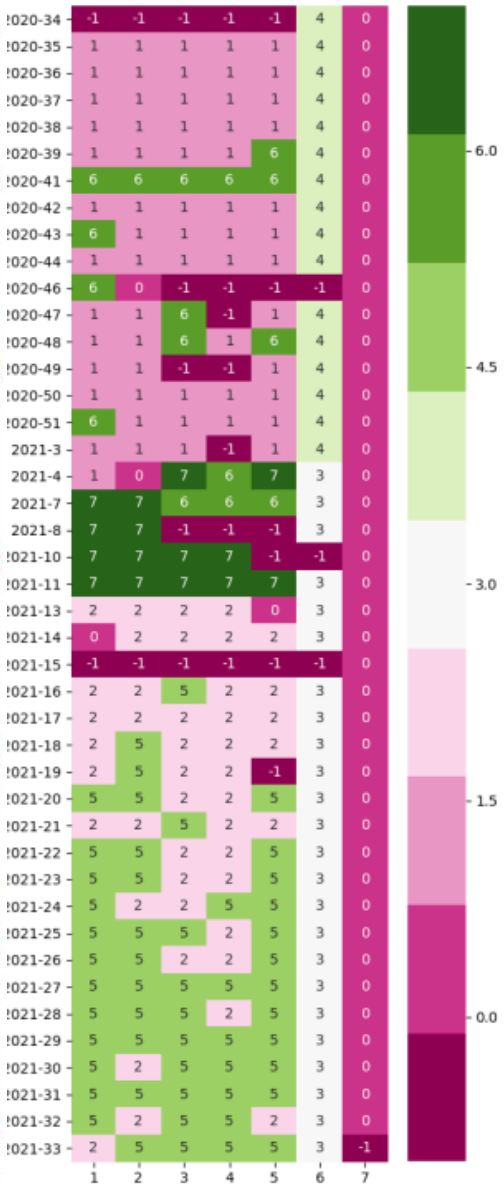
Agglomerative clustering results:



MCC

Agglomerative Clustering gave best results. Strong relationships between weekdays, seasons and clusters were observed.





4.3 Energy Consumption Forecasting Model

4.3.1 Introduction

Multiple forecasting algorithms have experimented on the dataset of Library at The Dock. Algorithms with better training results will be selected to be used in actual operation. Optimisation solutions are usually developed based on forecasting models, other than envisioning what could happen in the future, they can also be used to predict the effectiveness of energy savings and prevent risks.

4.3.2 Dataset

Library at The Dock	
Input	Sum of meter readings Supplementary humidity and temperature data
Resolution / Interval	1 hour (between meter readings)
Date Range	2020-08-22 to 2021-08-21
Pre-processing	Extreme values (> 1000 KWH) & days with incomplete readings were removed, resulting in 6912 records
Validation Method	5-Fold Cross-Validation (Randomly Split)
Testing Period	2021-08-01 to 2021-08-07

4.3.3 Feature engineering

Weather Features

Outdoor temperature, humidity, previous week's temperature, previous day's temperature, temperature/humidity, whether the temperature is within the comfortable range (14~22), the sum of squared temperature and humidity.

Time Features

Year, month, day, hour, weekday, day of the year, date distance, range of hour, Fourier series of time, in the frequency of day, week and year.

Other Features

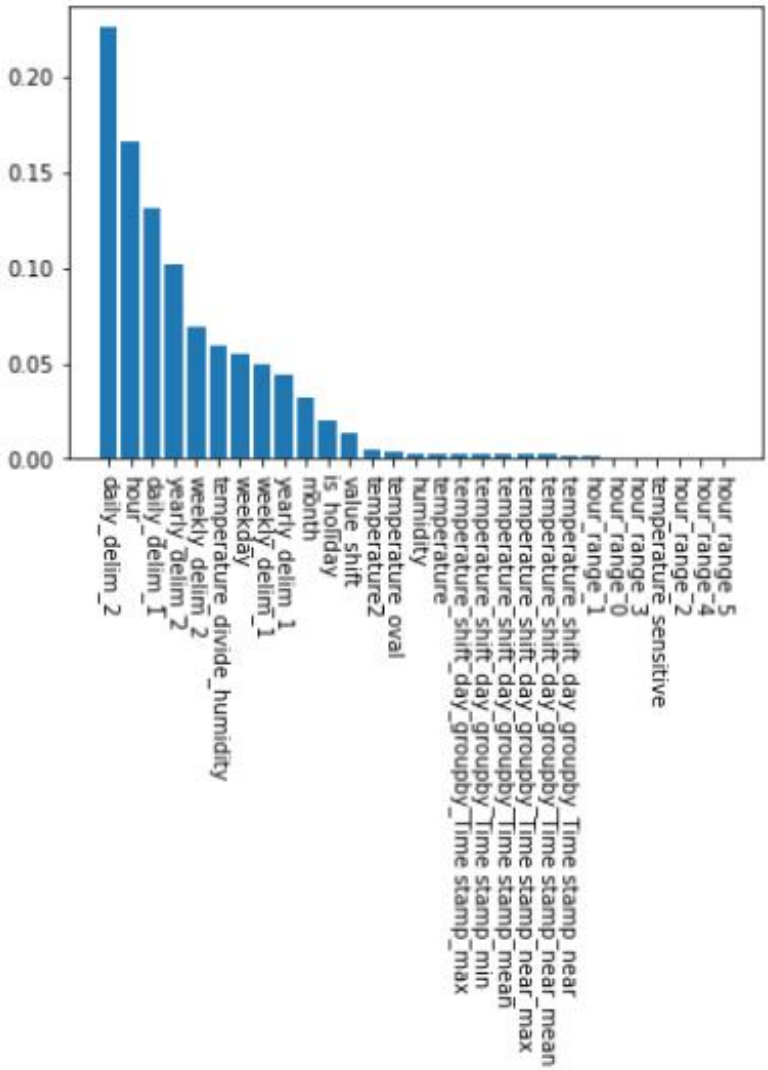
Meter value of last week and holiday details.

4.3.4 Experiment

We tested SVR, Random Forest and XGboost algorithms on the dataset. 5-fold validation was applied, feature importance was also investigated.

Mean absolute percentage error (MAPE) was used as the evaluation metric.

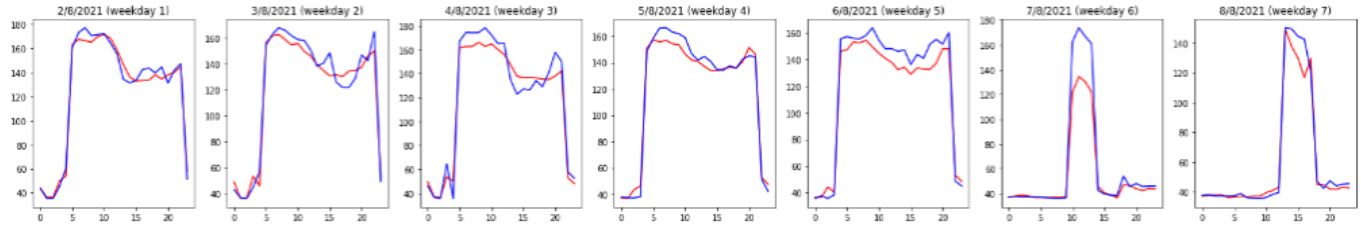
a. Random forest and SVR



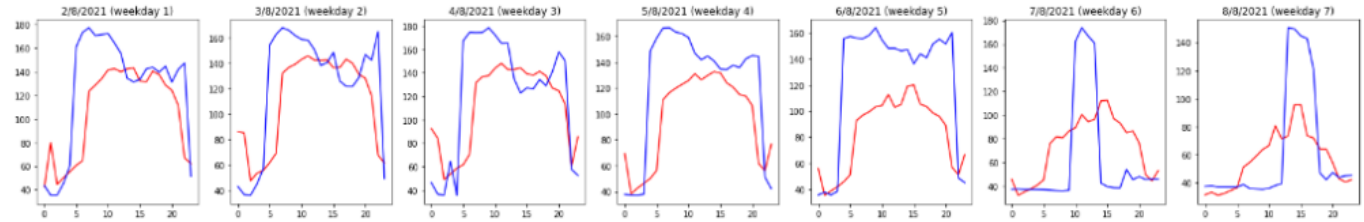
Importance of features

Model	Feature selection	MAPE
Random forest	/	5.48%
Random forest	threshold=0.001	5.41%
SVR	/	28.30%
SVR	threshold=0.001	28.60%

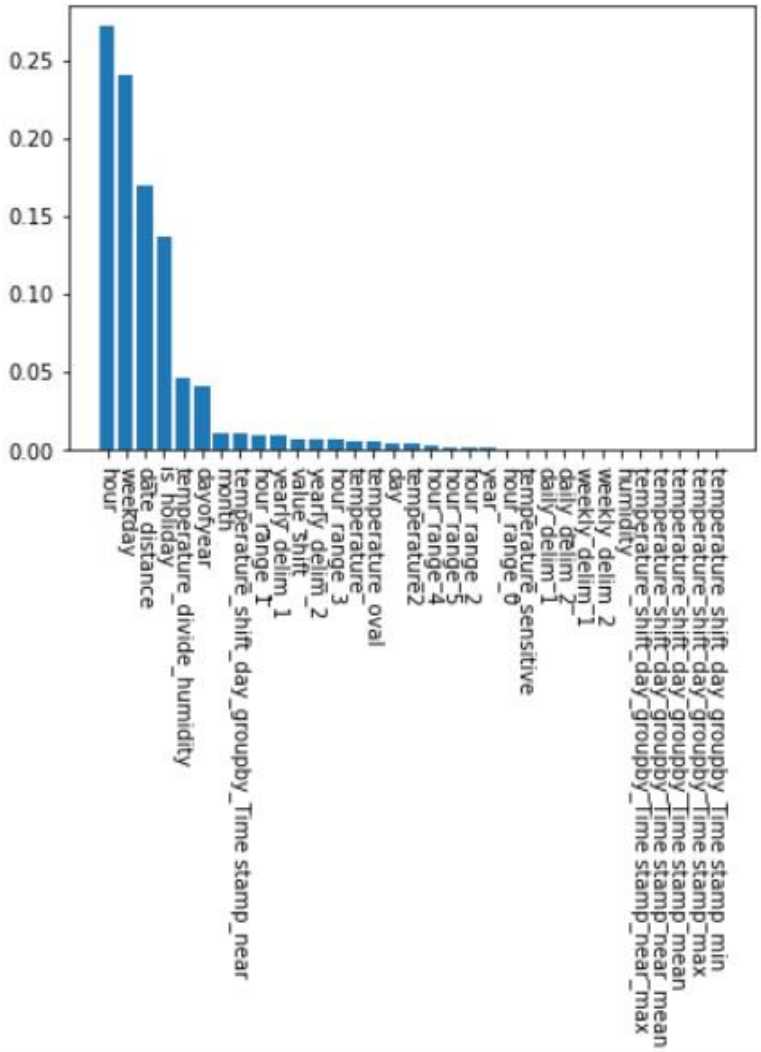
Random forest forecasting prediction (red line: prediction; blue line: ground truth)



SVR forecasting prediction (red line: prediction; blue line: ground truth)



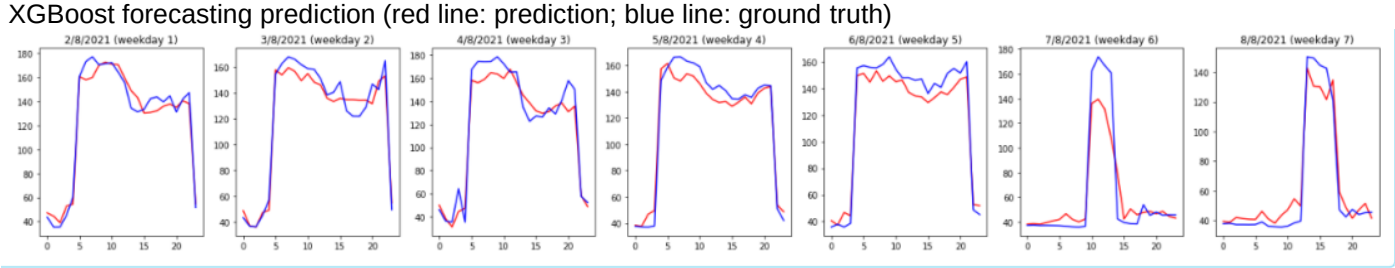
b. XGBoost



Importance of features

Model	MAPE
XGBoost (raw data)	8.26%
XGBoost (normalized data)	8.90%
XGBoost(raw, selected feature, threshold=0.001)	8.20%
XGBoost(raw, selected feature, threshold=0.002)	8.38%

XGBoost forecasting prediction (red line: prediction; blue line: ground truth)



4.3.5 Observations

Random forest gives the best result. We notice here that in most literature, which have more data for training, SVR gives better results (around 2% of MAPE). This indicates the importance of the size of the dataset. Besides, having data from the previous year allows us to design other features, and get a more robust validation result.

The random forest gets advantages from the number of features while the SVR may be overfitted to the training split. But from the experiment of training SVR using highly discriminative features, it is also observed that the performance is still not very satisfactory.

4.3.6 Notes and future improvements

It is important to have enough training data for each month. It observed that the random forest and XGBoost models obtained poor results when it was validated a month outside of the training set.

Potential Improvements

a) Designing other features (e.g. using previous years' values to conduct forecasting)
b) Parameter search for models
c) Apply other feature selection strategies
d) Use the clustering result to aid forecasting (e.g. firstly divide the dataset into subsets, then use the data in the cluster to train the model)

It is worth noting that most of the improvements require extra data for training. It would be very hard to find common patterns and features using only one year of data.

4.3.7 Reproducibility

We are able to extend the method and reproduce the results by using BRICK schemas as guidance.

Procedures

1) Given the BRICK schema, we are able to query the streams from the swagger API and sum the consumptions of electricity in different granularities. The sum of values will be stored in our database, for training purposes.
2) Then, we can use the data to train the model. We allow users to run the model training algorithms on the given buildings and get a trained forecasting model. The model will be stored locally.
3) The model can then be used for prediction purposes. The predicted results, the actual results, and other statistical values can be visualised by the front-end. These results will also be stored in the database so that the system can handle a number of buildings.
4) As we store the sum of results, we are able to implement and test different algorithms to improve the forecasting results once more data is available.

4.4 Thermal Model BIM Models

Full building thermal models are built and they simulate the operation of heating, cooling, ventilation, lighting and other loads in buildings. They guide decision-making regarding power capacity expansion by illustrating different strategies to meet future demands and environmental targets. Moreover, thermal models are employed to aid other modelling processes for energy forecasting and analytics when there is insufficient data.

We utilized EnergyPlus and TRNSYS to build the basic simulation model, the output of the model is used to aid energy analytics and forecasting.

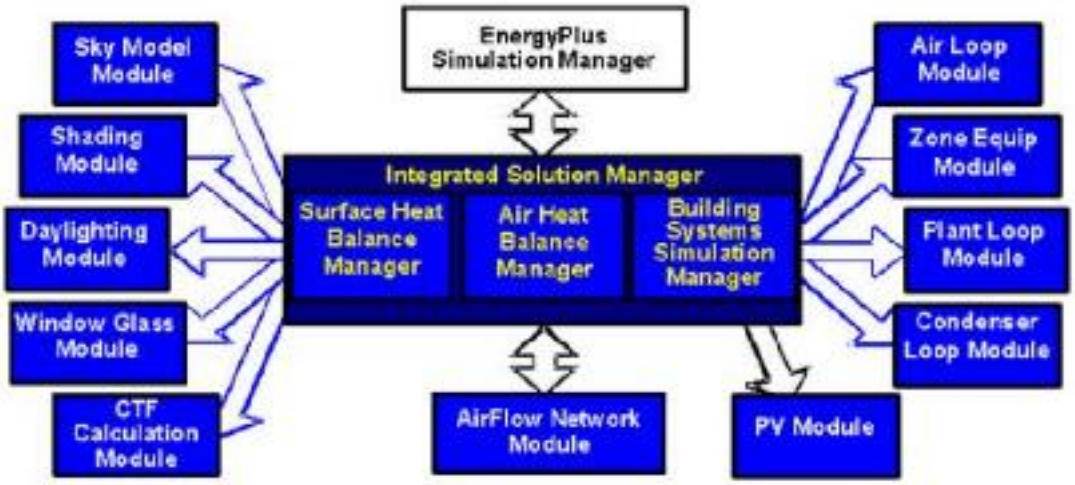


Figure 4.4.1: detailing of thermal models

Take Lady Huntingfield as an example, its building layout plan is transformed to BIM.

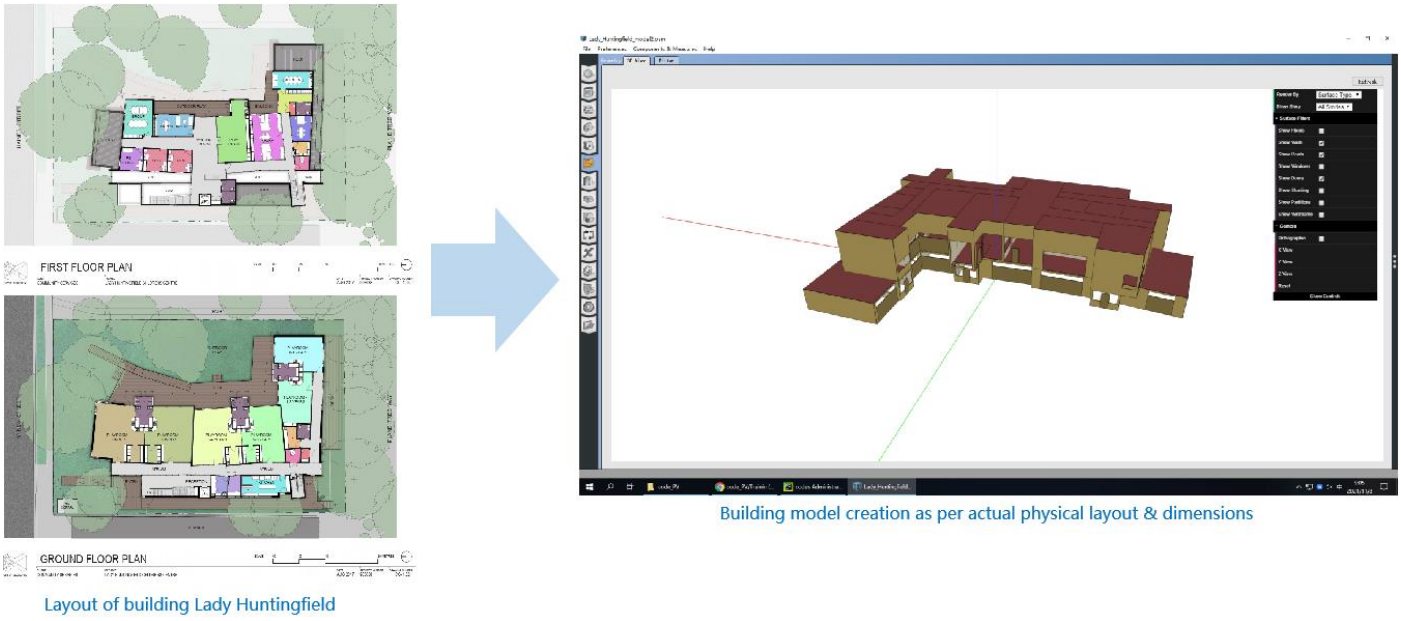


Figure 4.4.2: BIM Creation – Lady Huntingfield

We then set up the building thermal parameters. For instance, the images below are the enclosed structures settings, internal gains settings and weather data settings for Lady Huntingfield.

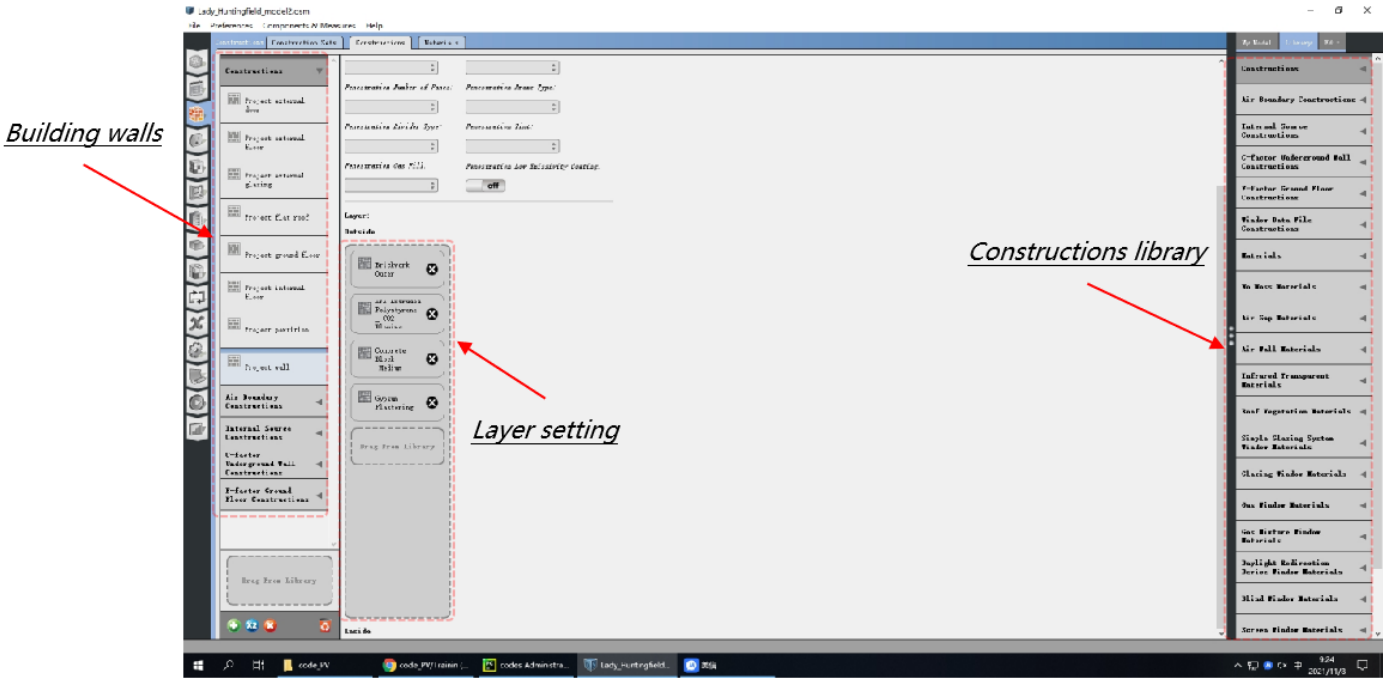


Figure 4.4.3: Enclosed structures setting

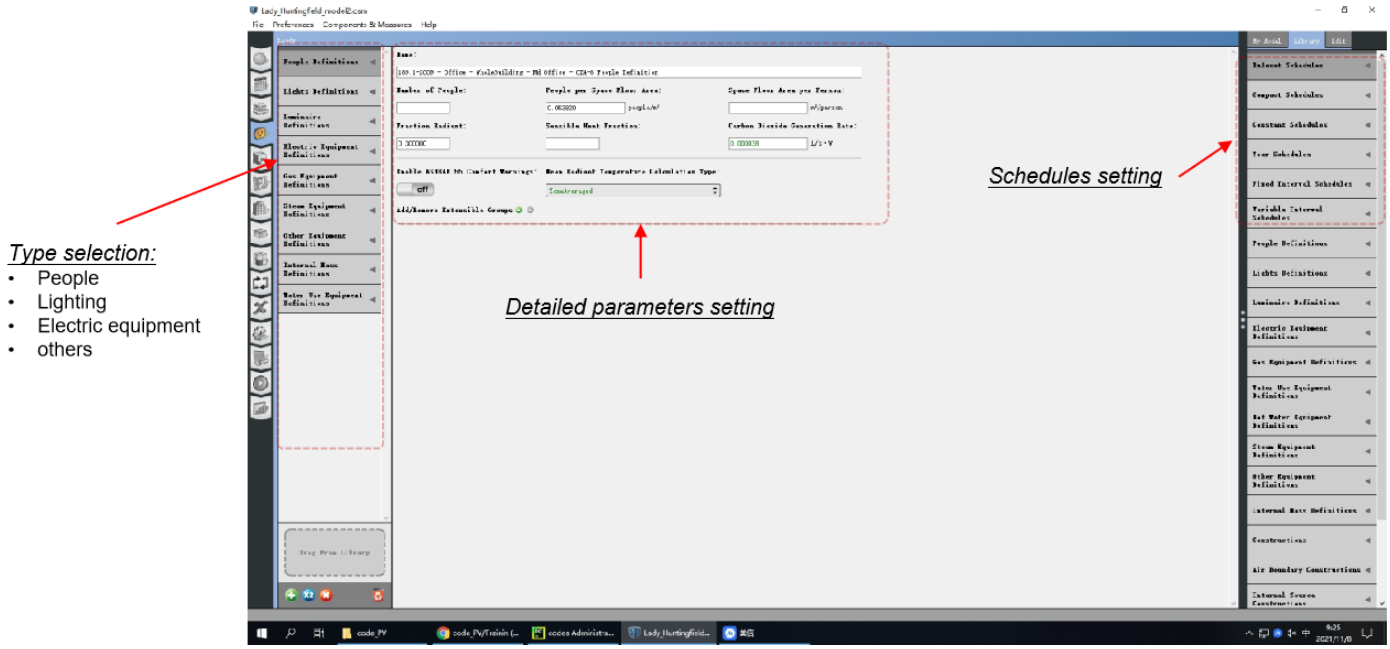


Figure 4.4.4: Internal gains setting

Weather data Melbourne input

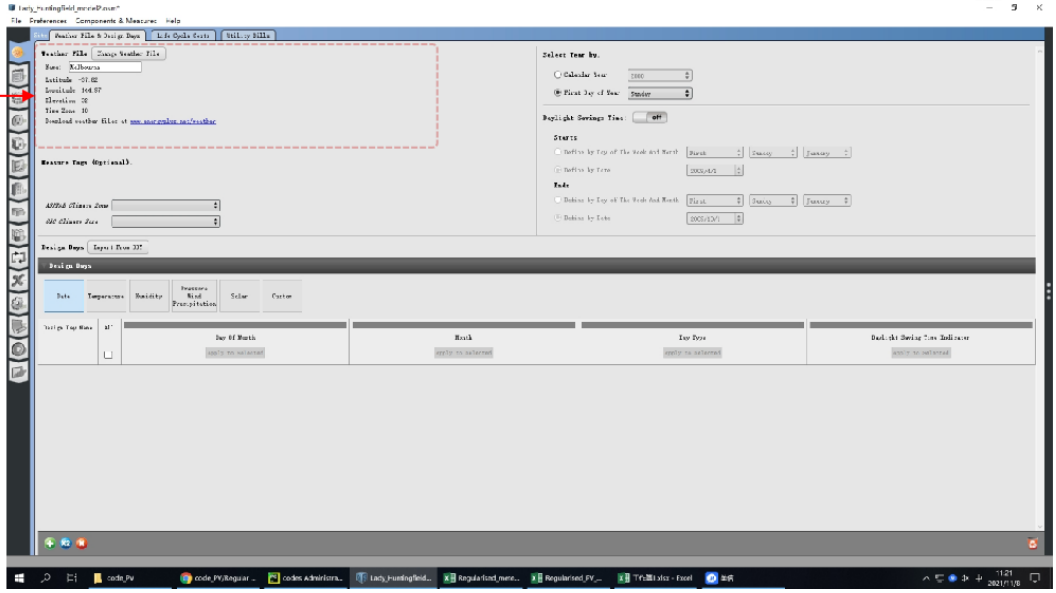
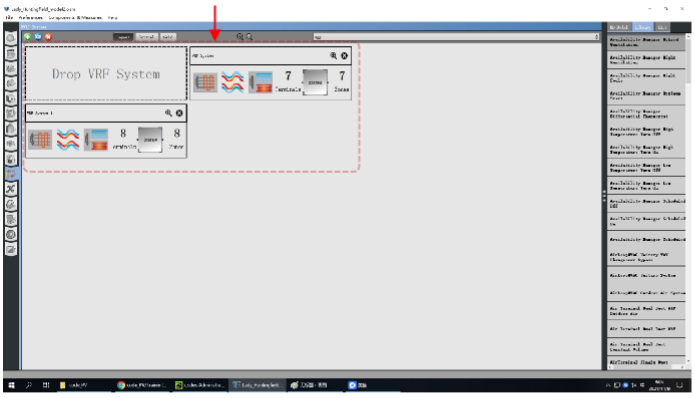


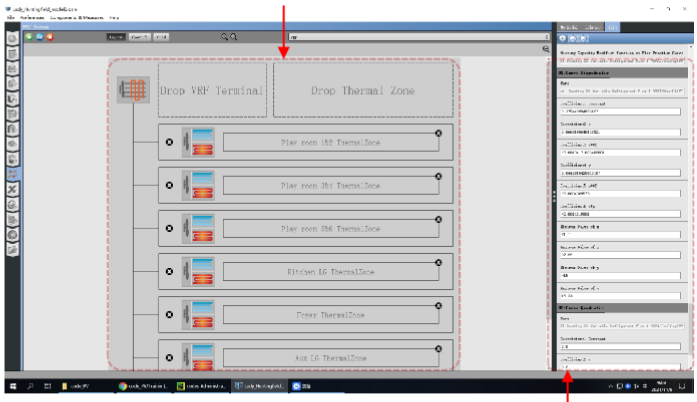
Figure 4.4.5 Weather data setting

After that, we add in HVAC information through VBIS Tags to have the thermal model set up as follows:

VRF system schematic



Terminal – zone topology



AC Plant Reference	Room Name	AC Plant Reference	Room Name
FCU-G-01	Playroom 01	FCU-1-01	Group
FCU-G-02	Playroom 02	FCU-1-02	Parent Room
FCU-G-03	Playroom 03	FCU-1-03	Waiting & Lobby
FCU-G-04	Playroom 04	FCU-1-04	L1 Play Room
FCU-G-05	Playroom 05	FCU-1-05	Staff Work Stations
FCU-G-06	Playroom 06	FCU-1-06	Meeting Room
FCU-G-07	Kitchen	FCU-1-07	Staff Kitchen
FCU-G-08	Corridor 1	AC-1-01	FS Consulting

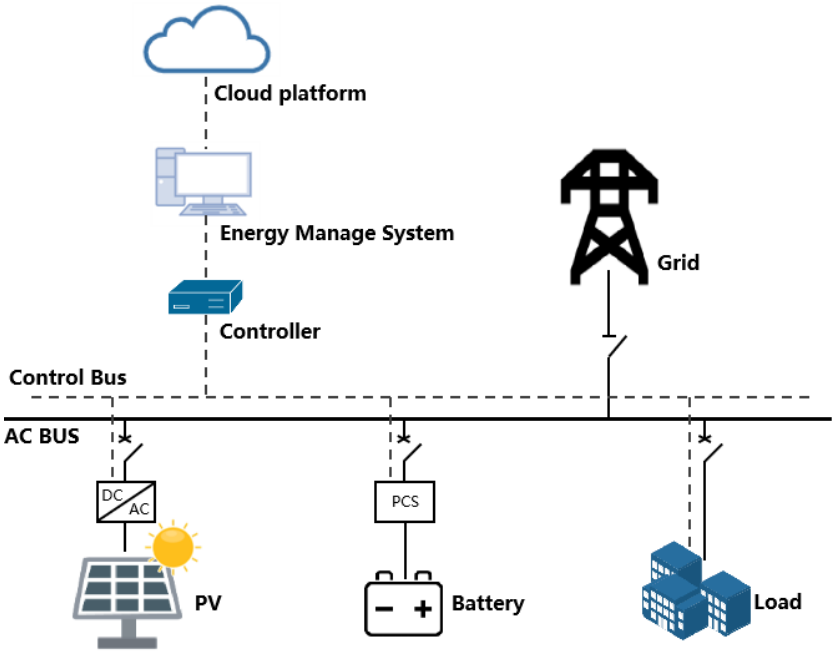
AC Plant Reference	Room Name	AC Plant Reference	Room Name
AC-G-01	Staff Co-Ord	AC-1-02	MCH2
CU-G-O2	GF Cond. Unit	AC-1-03	MCH1
		CU-G-O1	1st Fl Cond. Unit
AC-G-02	Comms Room		
CU-R-02	Cond Unit on Roof	AC-1-04	Comms Room
		CU-R-04	Cond Unit on Roof

Performance polynomial curves of VRF unit

Lastly, the thermal model was exported to other energy analytical models and forecasting models to produce energy-saving and cost-saving control strategies. Control algorithms could be developed with Matlab/Python etc, then be integrated with EnergyPlus via BCVTB for co-simulation.

4.5 Energy Optimisation Models

Energy optimisation models were still going through the training and optimisation phase. Referring to the topology below, the energy management system developed based on various algorithms has a range of data e.g. weather, rate, battery condition, and buildings' power consumption. Controlling the allocation of each distributed energy resource can help us to achieve goals of lowering energy costs and increasing the PV utilisation rate.

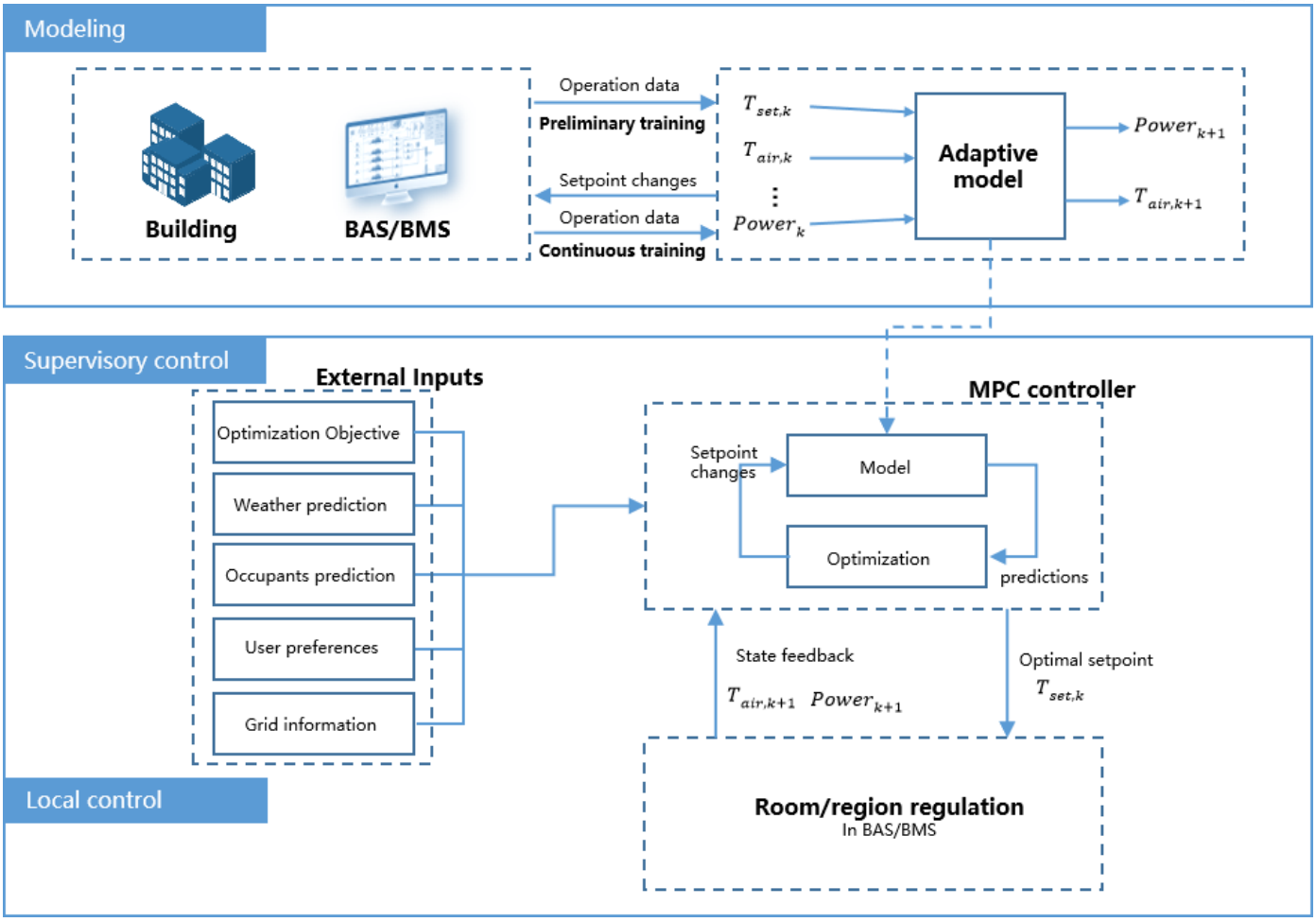


The model attempts to use 3 kinds of energy storage to lower the building's energy costs and increase the PV utilisation rate.

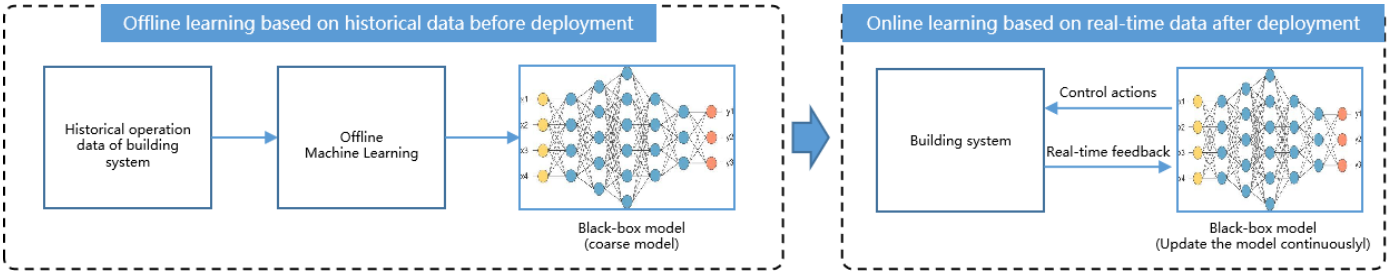
a) **Active Energy Storage**: Batteries store energy when the rate is low and supplies energy when the rate is high

b) **Passive Energy Storage:** Makes use of the buildings' thermal inertia to store thermal energy. For example, when the rate is low, the room temperature is set to be lower and the energy is stored and released by setting the temperature higher when the rate is higher. This reduces the HVAC energy costs.

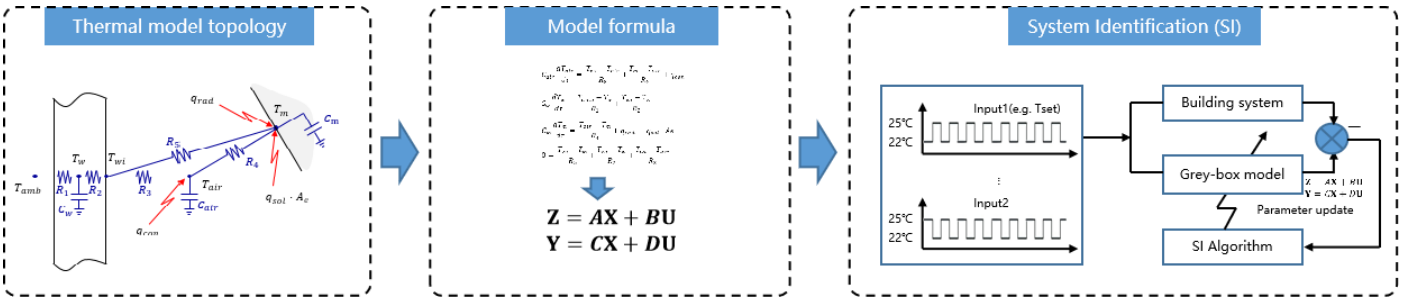
c) **Active + Passive Energy Storage**: the modelling and optimisation structure of passive energy storage is showing as below:



We would usually use the following modelling method when there is sufficient building data and when permission is granted by the building management:



However, as this project do not meet all the criteria required to use the above method, we have added the following methodologies in modelling:



Despite our current method of using physical simulation modelling (EnergyPlus, TRNSYS etc.) as a replacement for using actual building data, the connection between the simulation model and the actual building is still crucial and we will need to have access for more real-time data.

4.6 Carbon Neutral Models

Terms of 'Carbon neutrality' and 'net-zero emissions' are rising in popularity. Whether we are a business or an individual, this is an important step towards managing our carbon emissions efficiently and identifying potential for reductions and savings.

From this project, when we were running the PV analytics and energy efficiency analytics, we have also created emission models to estimate the carbon footprint for both building level and organisation level. Users will be able to find out how much they need to offset to compensate for their greenhouse gas emission, combined with our other analytics results, the system app indicates that another way to reduce emissions and to pursue carbon neutrality is to offset emissions made in one sector by reducing them somewhere else, this can be done through investment in renewable energy and energy efficiency.

We use emission factors that set by Environment Protection Authority (EPA), for the organisation level, calculation models for items of scope three were built including waste, air travel, transport, printing and publication services, courier services, office paper and reticulated water, couple examples of calculation methods are showing as below:

Purchased electricity	
Calculation Method	$Y = Q \times \frac{EF}{1000}$ Where: Y = scope 2 emissions (t.CO2-e) Q = quantity of purchased electricity (kWh) EF = scope 2 emission factor for purchased electricity (kg.CO2-e/kWh)
Assumptions	Where data was not available for the final months of 2018-19, estimations were made using average daily consumption, calculated using data provided for previous months. Data was not available for a number of air monitoring stations. Estimations were made based on known consumption of similar air monitoring stations or electricity use in prior years. Data was not available for all thirteen small footprint sites (including sites with sensors and cameras only) which are powered by 100% solar energy. As these sites are powered by 100% solar energy, there are no emissions associated with the electricity used by these sites. However, estimations for electricity use for these sites were made using the average power consumption and run-time for the year. In the instance that EPA is not the sole tenant of the building and data was provided for the entire building or one floor of a building, GHG emissions were apportioned using EPA's Net Lettable Area (NLA) as a portion of the total NLA for the building or Floor. Emissions associated with electricity powered by renewable energy are zero.
Factor	EF: 1.17 (kg.CO2-e/kWh)

Natural Gas	
Calculation Method	$X = Q \times \frac{EF}{1000}$ Where: X = scope 1 emissions (t.CO2-e) Q = quantity of natural gas purchased (GJ) EF = scope 1 emission factor for natural gas distributed in a pipeline which includes the effect of an oxidation factor (kg.CO2-e/GJ)
Assumptions	Where data was not available for the final months of 2018-19, estimations were made using average daily consumption, calculated using data provided for previous months. As EPA is not the sole tenant of these buildings, in the instance data was provided for the entire building or one floor of a building, GHG emissions were apportioned using EPA's Net Lettable Area (NLA) as a portion of the total NLA for the building or floor.
Factors	EF: 51.53 (kg.CO2-e/GJ)

In this project, main available inputs are read from building historical and live electricity metering data. As the graph shows below, by the basic calculation methods and live sub-metering data, our back-end carbon emission analytic models can be refined to list the sources of carbon emissions, we calculated emissions in the way of by its facilities such as from HVAC, power or lighting, thus we could find out those highest facilities that producing emissions and how we could possibly increase energy efficiency to offset emissions. (Below chart is internal testing for processing data which is not the final applications intend to deliver)

5. APPLICATION 3 – BUILDING PERFORMANCE ANALYTICS

5.1 Data Query

For the application of building performance analytics, PrediQ has to query the DCH and VBIS tagging system on:

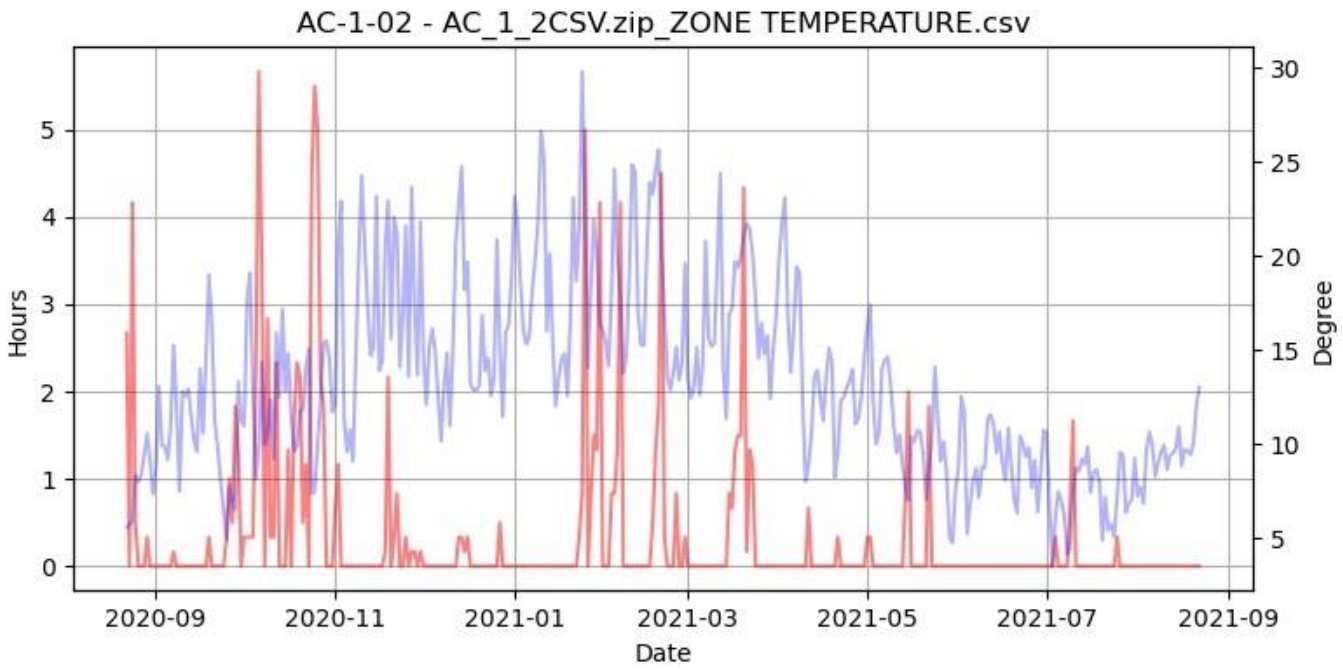
- local weather data
- building occupied period
- HVAC main settings
- HVAC zonings
- room temperatures
- room humidity
- other available IoT sensors

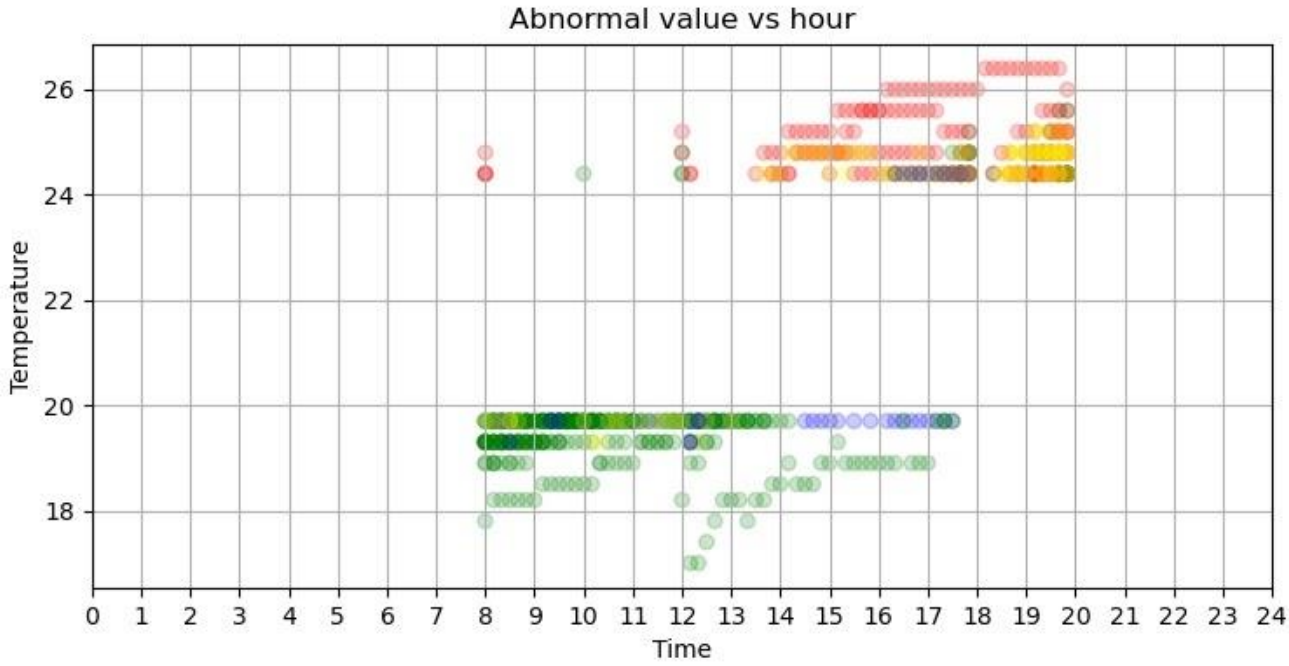
5.2 Building Performance Models

Building performance is analysed through generating the occupancy experience analytic model and analysing the relationship between the sensors inside the building and their corresponding control systems. Due to limited access to the data related to indoor comfort, BRICK models are used to locate the temperature sensors, humidity sensors and CO2 sensors, and the areas with poor indoor comfort are identified. The problematic devices and the controlling errors are then picked out.

Modelling Example: by recognising the abnormalities in room temperature, effectively identify the malfunctioning air conditioner and its temporal distribution.

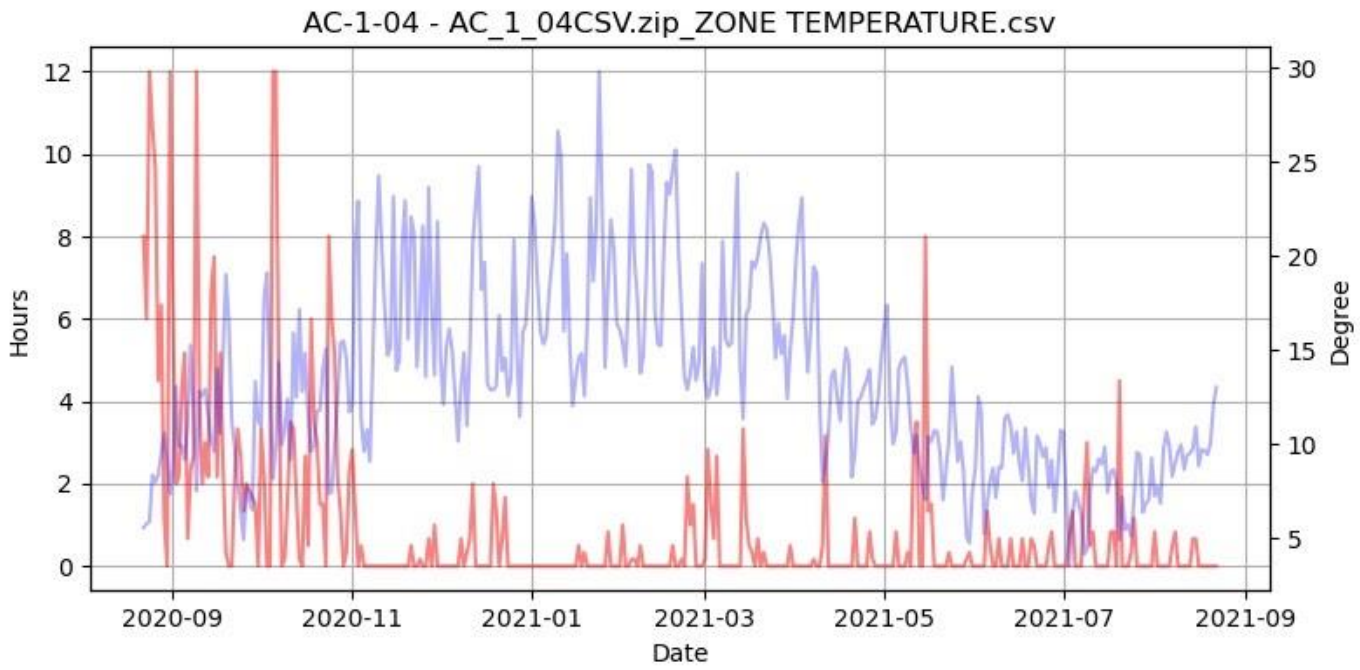
Making use of the BRICK model, the model uses the time series data of room temperature to expeditiously work out the full year building operating condition. The graph below shows one of the more seriously malfunctioning devices:

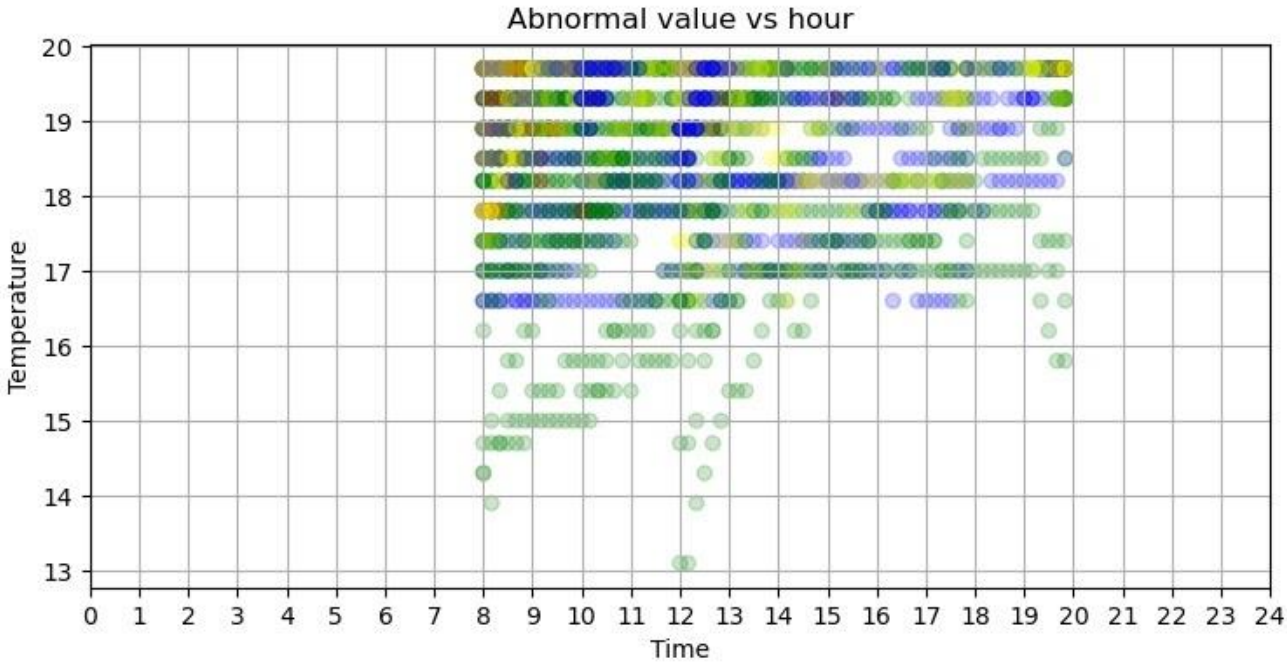




When referring to the first graph, the blue line represents the average outdoor temperature while the red line is the accumulated hours of malfunctioning. The second graph is the scatter plot that reveals the abnormal value and the time where abnormalities are recognised. X-axis is the time the abnormalities are recognised (00:00-24:00), y-axis is the temperature readings from the sensors when abnormalities are recognised. Blue colour means it happens in Winter, while red means Summer, yellow means Autumn and green means Spring.

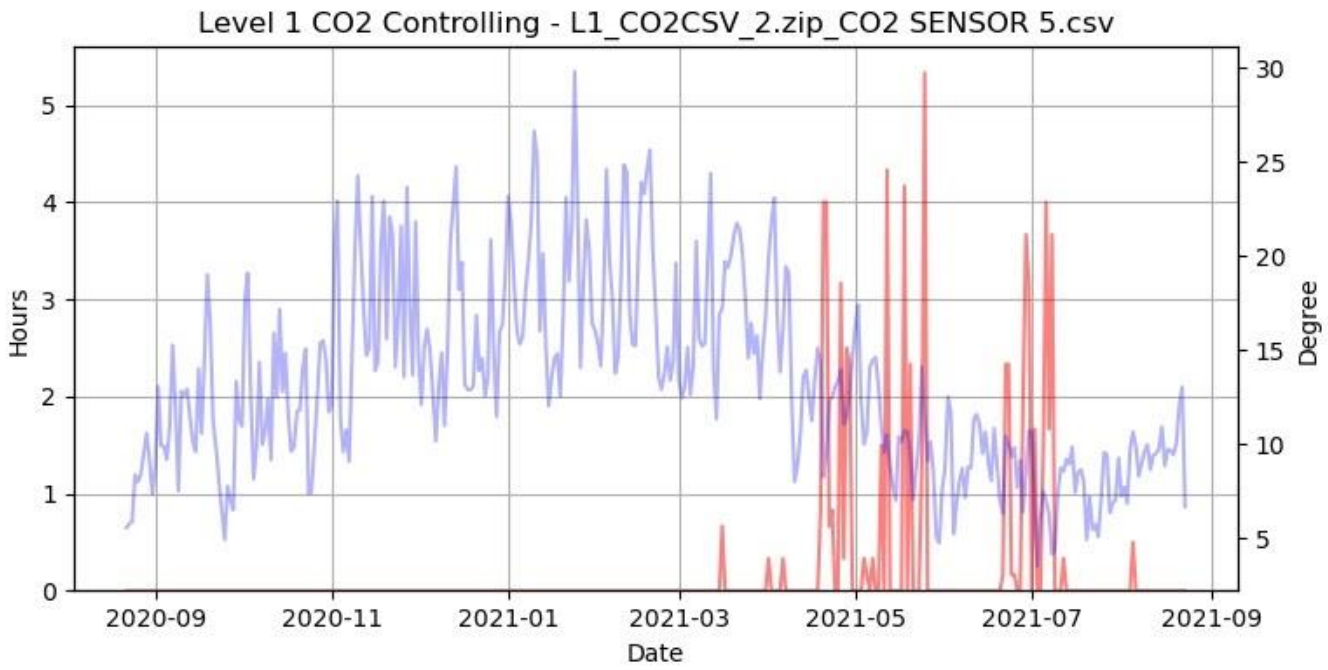
It is observed that temperature control occurs across the whole year for AC-1-02, and abnormalities are found during both heating and cooling. In the APP for clients, we identify the problems from raw data, then locate the malfunctioning device through data correlation while the identified device will provide information on the area where problems occur and possible solutions are raised.

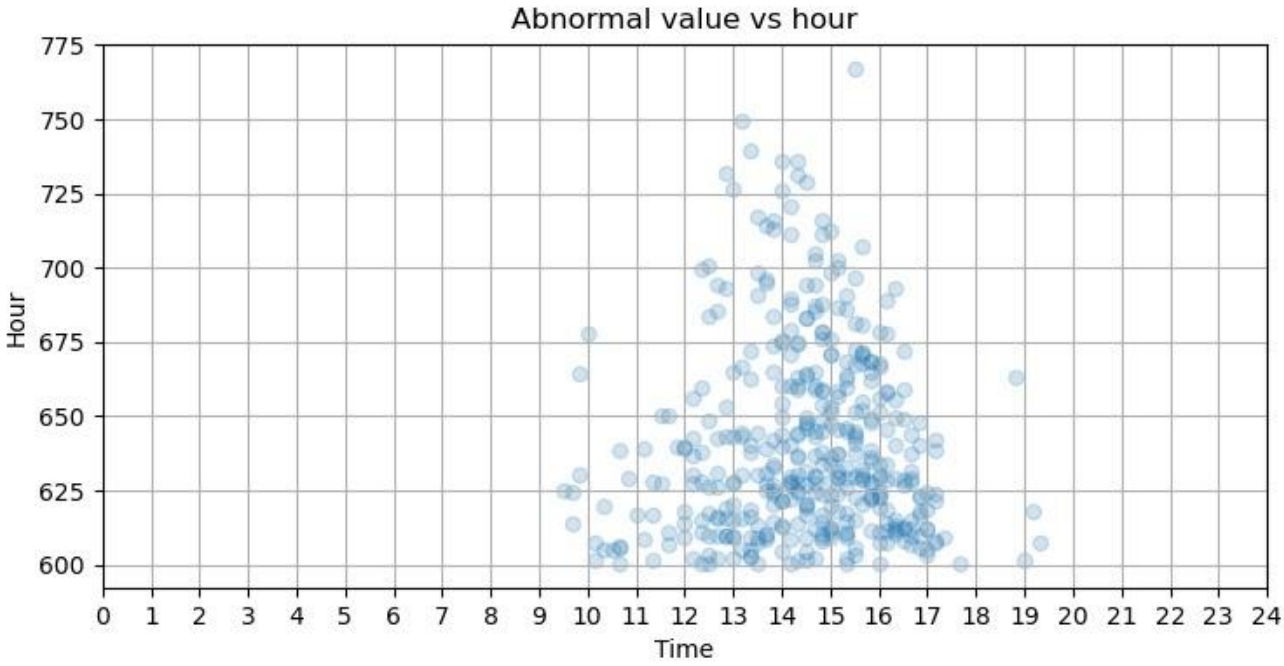




Similarly, temperature control problems mainly occur in Winter for AC-104, it had even reached some extremely low temperatures. Data analysis can help identifying the problem efficiently and provide the corresponding feasible solutions when dealing with a massive dataset.

Below are graphs about CO2 issues, where they usually occur when it is crowded but the control system's ventilation is poor.





When referring to the first graph, the blue line represents the average outdoor temperature while the red line is the accumulated hours of malfunctioning. The second graph is the scatter plot that reveals the abnormal value and the time where abnormalities are recognised. X-axis is the time the abnormalities are recognised (00:00-24:00), y-axis is the CO2 readings from the sensors when abnormalities are recognised. It is observed that Level 1 had frequent problems in May and July.

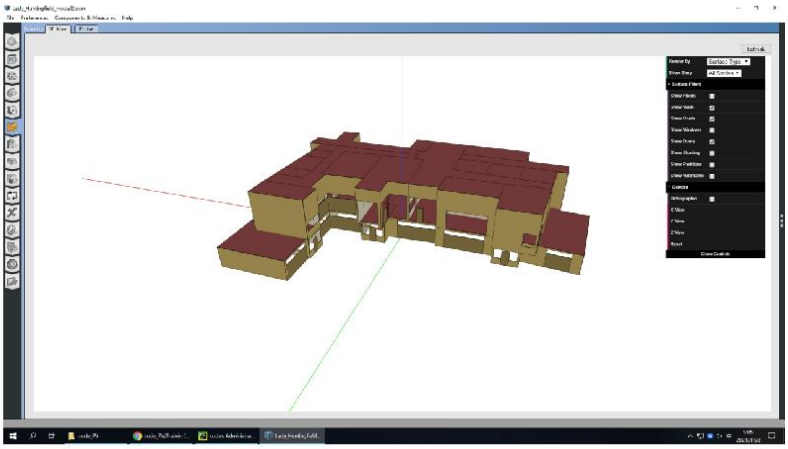
Performance analytics were done on 17 areas at Library at The Dock. The problems it has for several years of time have been uncovered in a few seconds, and the analytics can be duplicated to other buildings in a timely manner.

Note:
Refer to Appendix 2 for full analytics record of building performance for Library at the Dock.

5.3 BIM

BIM aids the development of thermal models and visualisation. We have built the preliminary thermal models for Lady Huntingfield and the visualisation model for Boyd School by using BIM and VBIS tags.

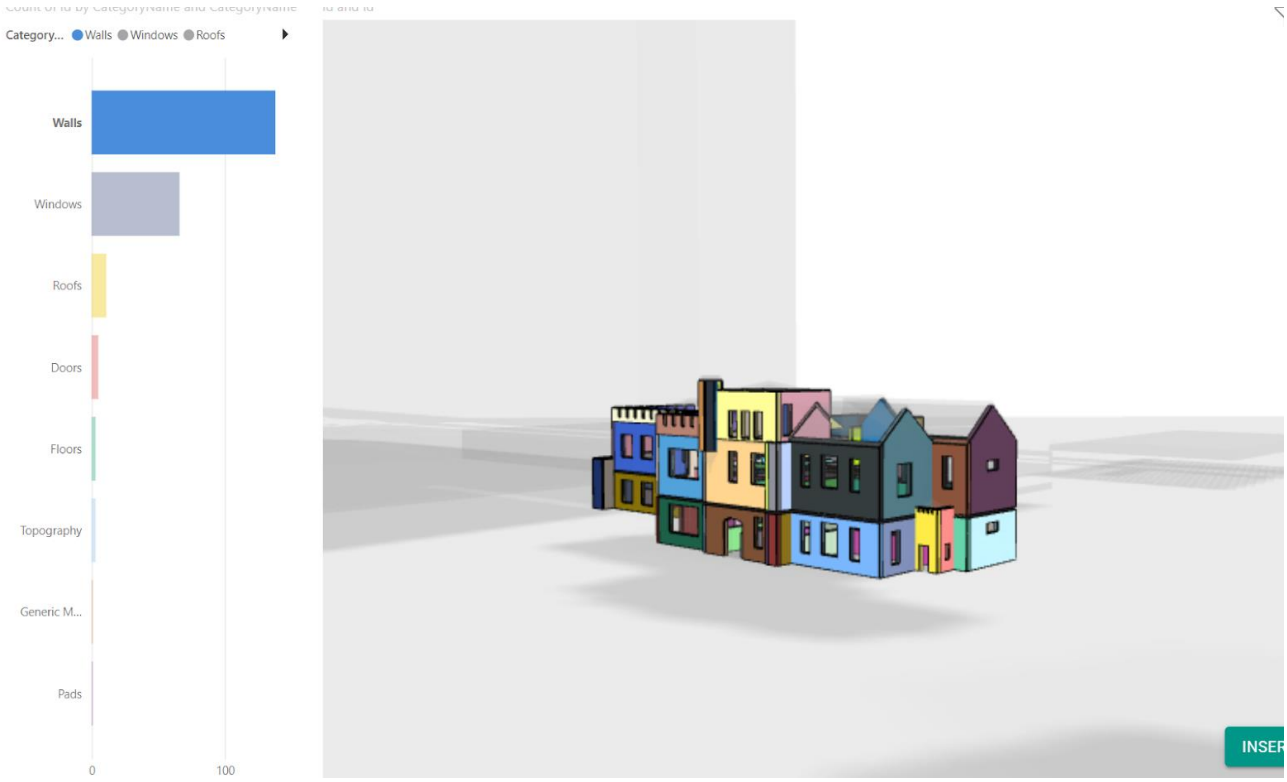
Layout of building Lady Huntingfield



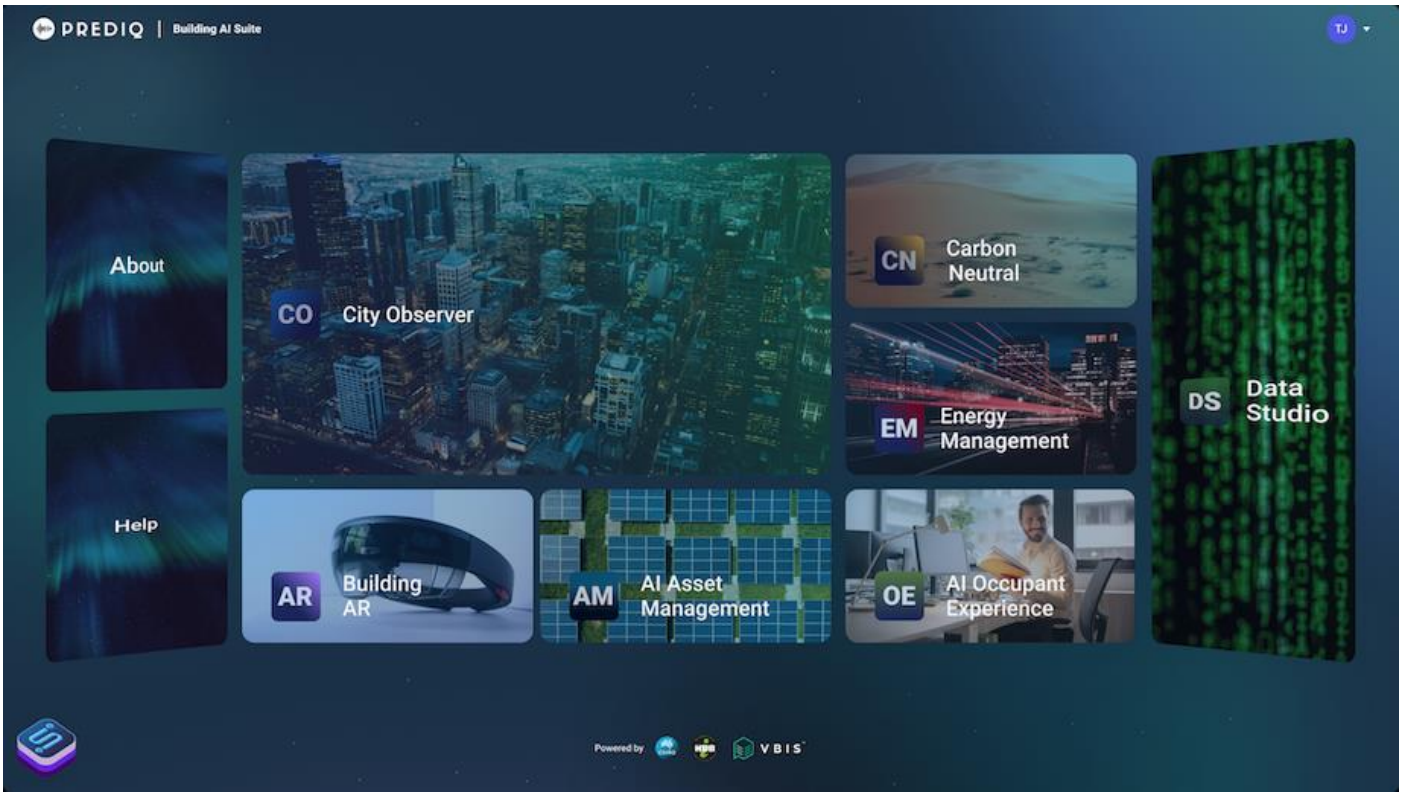
Building model creation as per actual physical layout & dimensions

Under circumstances where we do not have sufficient data, thermal models support the training of different energy models.

On the other hand, the 3D visualisation model of Boyd Community Hub has created, it might be easier for users to locate the area with issues and look into the connection between the issue area and its analytic results. Discussion on value with users will be needed to decide if this 3D model is necessary to be included in the final app user interface.



6. APPLICATIONS USER INTERFACE OVERVIEW

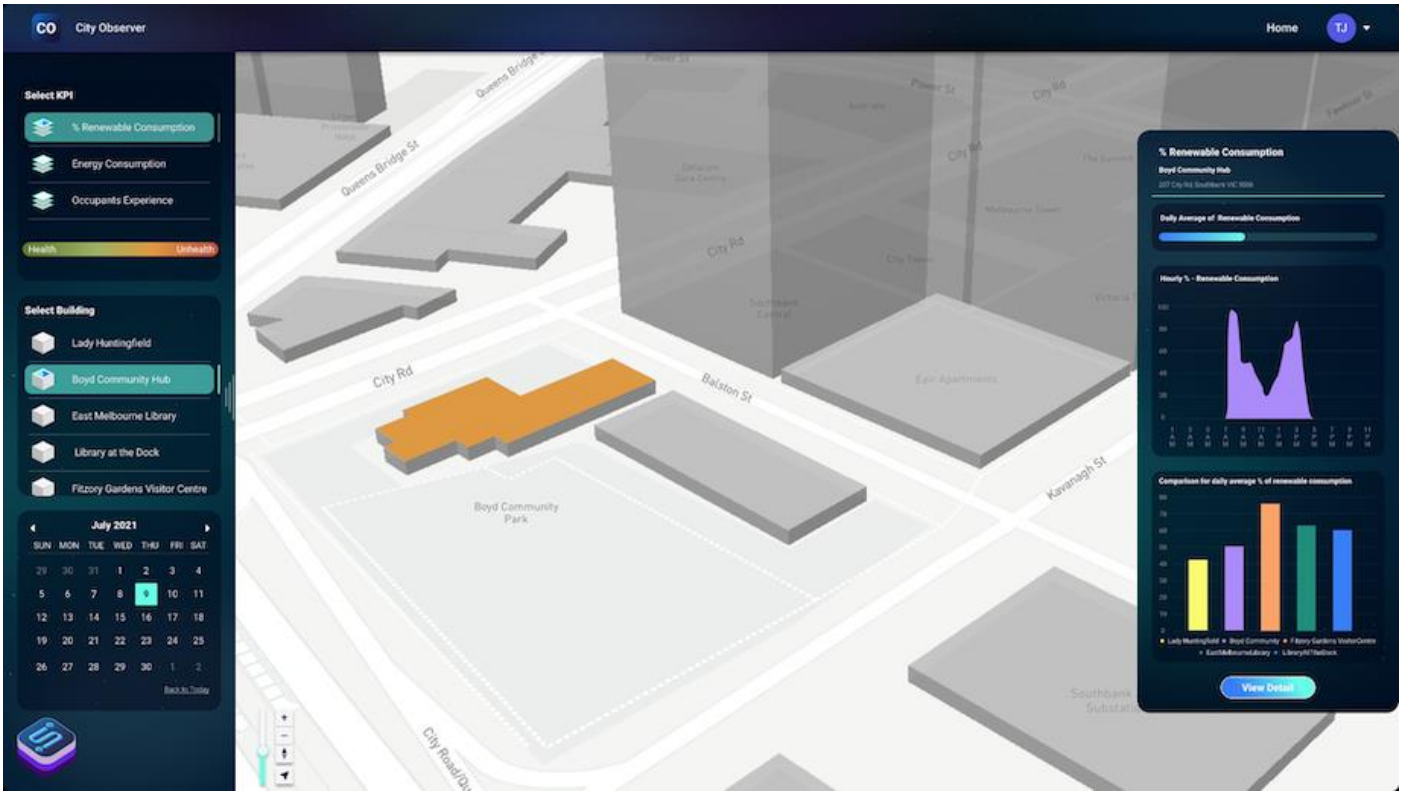


In this project, we have developed four (4) app modules at this stage, which including:

1) City Observer

This is for the project end user (City of Melbourne) to have an efficient multi-buildings management. Key functionalities including:

a) Visualisation on whole picture of multi-buildings management with selected KPI;
b) Display and visualization on key building performance;
c) Comparison of renewable generation and energy usage for buildings in the similar geographical and weather conditions.



2) Energy Management

This module Indicating both energy consumption and renewable generation and usage. Also calculating energy save from transformative solutions and enable users to see full saving of their actions. Key functionalities including:

a) Monitoring of specific building system parameters and by capturing changes in the external environment and analytic rules to generate response to prevent excessive energy consumption and operational inefficiency;
b) Formulate and optimise economic model that lively calculate the profit of ESS usage and EV charging in the energy market to provide feedbacks and recommendations on the way of using energy to maximize the return of investment of using renewable energy system;
c) Generate alerts for performance and maintenance needed for critical equipment such as PV systems to ensure energy efficiency and the most efficient forms of renewable energy.

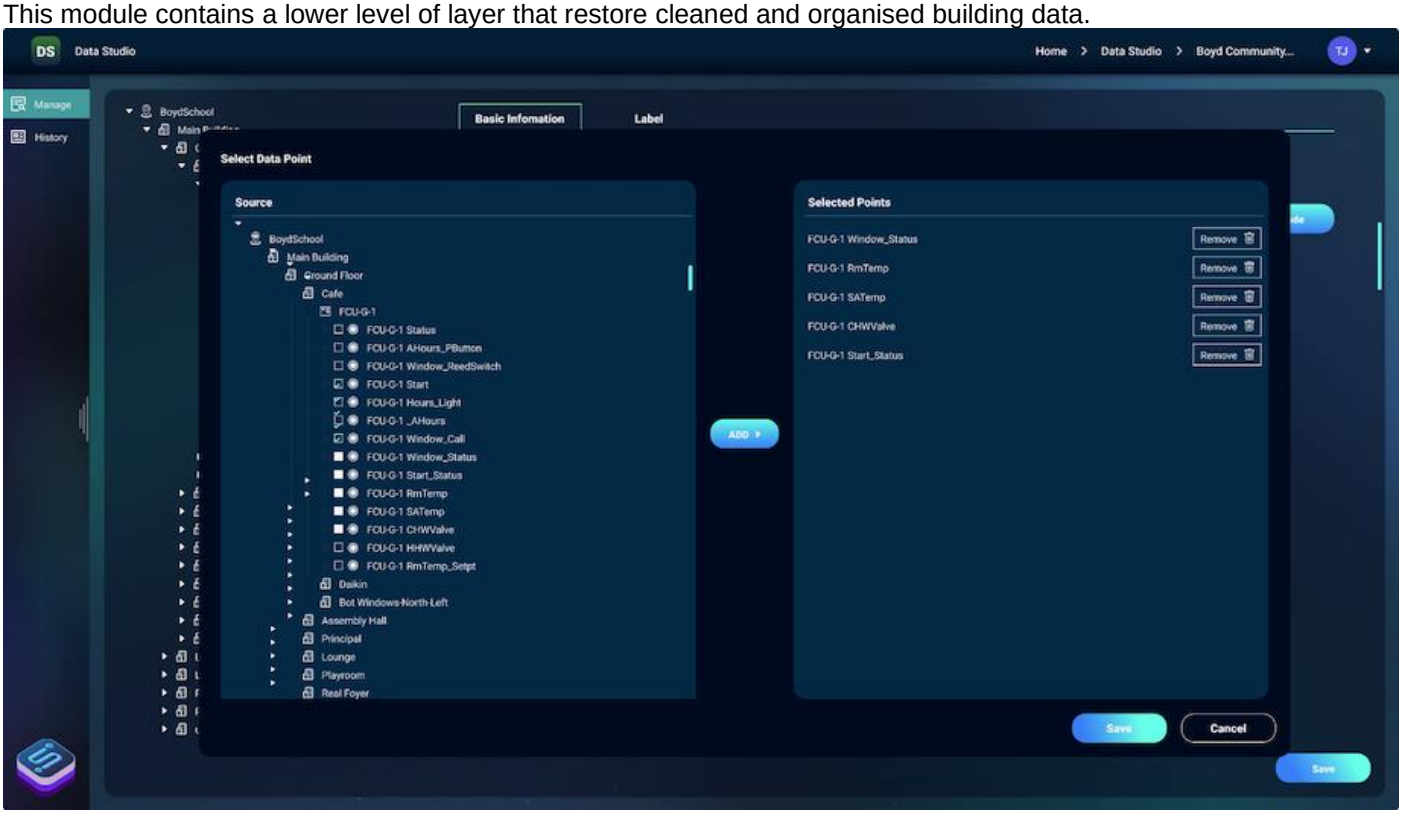
3) Carbon Neutral

Here calculates CO2 production and CO2 reduction from transformative solutions and enable users to see full saving of their action. Key functionalities including:

a) Visualise analytic highlights and monitor building performance on critical mechanical devices that causing CO2 emission;
b) Provides recommendations on the efficient and economic ways to reduce emission, and tailor carbon offset solutions to balance the unavoidable greenhouse gas emissions.

4) Data Studio

This module contains a lower level of layer that restore cleaned and organised building data.



Note:
Received end user feedback from City of Melbourne in November 2021. This will be taken on-board to ensure supplied interface addresses their requirements. It is anticipated that the user interface iterations will occur during the remainder of this project with functionality added and removed based on user feedback and scope limitations.

7. CONCLUSION

7.1 Use of BRICK Model and VBIS Tags

BRICK model and VBIS tags play an important role in the efficient location of data points and modelling when performing building analytics. When there are buildings lacking engineering information, the BRICK model and VBIS tags minimise the data inputting workloads in the early stage.

7.2. Analytical Summary

PV analytics can effectively identify the PV system running issues and provide optimisation solutions to provide a basis for the investment return calculation in the later stages. Meanwhile, the output of the PV forecast can aid energy optimisation and achieve the maximisation of solar energy usage.

Energy profile modelling can reveal the building use characteristics, which is used for energy-saving analysis and strategies and developing the demand side response baseline.

Data quality accounts for the unstable nature of the energy forecasting model, thermal model and other methods will be employed later on so as to provide more stability to the analytic model.

From the initial analytical results generated based on the 5 buildings' data, it is believed that not less than 5 optimisation solutions can be developed, including HVAC optimisation that considers good indoor comfort, energy storage optimisation, feasibility of VPP based HVAC, PV monitoring and optimisation, and self-controlled system strategies. These could help in accomplishing carbon neutral and energy optimisation, it is believed that this project will contribute to energy saving and cost saving notably. In addition, the visualisation interface assists the clients to locate areas and devices with issues through numerous algorithms and get professional advice that promotes better decision making in the future.

The indoor comfort model identifies the problems related to indoor comfort for the clients and can be easily duplicated to other buildings for assessment. The data knowledge graph and data correlation on the visualisation interface help clients recognize the solutions to the issues and the corresponding providers.

7.3 Uncertainties

There are numerous uncertainties when running the project. The historical data health of the buildings is concerning due to its low completeness and anomalies, there is a lack of historical data and some of the data coverage is minimal. Additionally, due to COVID-19, buildings are not operated in full capacity, it is anticipated that in the near future, there will still be fewer activities happening in the buildings and mandatory use of natural ventilation. These are going to alter the modelling process.

At this stage, we are not able to change the control system settings which has a significant impact on developing and validating of the optimisation models. Another uncertainty that affects the cost and returns analysis is the vague client's requirements on data analytics.