Unsupervised Machine Learning Anomaly Detection for Multivariate Time-Series Data in Wind Turbine Converters

Chuang Yu

Master's Thesis
Master of Engineering - Big Data Analytics
May 30, 2020

MASTER'S THESIS	
Arcada	
Degree Programme:	Master of Engineering - Big Data Analytics
Identification number:	
Author:	Chuang Yu
Title:	Unsupervised Machine Learning Anomaly Detection for Multivariate Time-Series Data in Wind Turbine Converters
Supervisor (Arcada):	Leonardo Espinosa Leal
Commissioned by:	ABB Drives Oy

Abstract:
Because wind power is one the main clean energy sources, the demand for wind generated energy has been rapidly increasing all over the world. As wind turbine converter is one of the key components in wind turbine, it is critical to ensure the reliability of its operation without human monitoring in addition to cost efficiency. This thesis studies and experiments two unsupervised machine learning models to detect anomaly turbine converters: Hidden Markov Model and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) model. The aim is to compare the results from the two selected models and cross validate the results with existing models and visualized graphs for data analytics. With Hidden Markov Model, three distance computation methods are applied: Discrete Frechet, Dynamic Time Warping (DTW) and Partial Curve Mapping (PCM). The experiments show that PCM is the fastest but also produces the worst result, while Discrete Frechet is the slowest, and produces the similar result as DTW. HDBSCAN is very intuitive to use and relatively fast to produce the clusters, and it works exceptional good on certain data Analysis Group. The experiment results show that both models do not provide satisfactory result compared to the existing models.

Keywords:	Anomaly detection, Hidden Markov Model, HDBSCAN, ABB wind turbine converter, unsupervised machine learning
Number of pages:	63
Language:	English
Date of acceptance:	

CONTENTS

Figures

Tables

FOREWORD

This thesis is written for ABB Drives Oy, Drive Products business unit, while working for Drive Products PC tools team.

I would like to take this opportunity to express my sincere thanks to many people that support me especially during the COVID-19 pandemic time. It would not be possible for me to go this far without their supports.

First of all, I would like to thank my advisor Dr. Teppo Pirttioja, for providing guidance, comments, ideas and time.

I want to thank my supervisor Prof. Leonardo Espinosa for giving exemplary advice and support during the writing process.

Last but not the least, I would also like to thank my family, thanks for their supports, patience and understanding. It is no fun to work during the nights, weekends and holidays, the COVID-19 pandemic makes the life even harder, I probably had already given up without their encourage.

1 INTRODUCTION

This chapter describes the necessary definitions used in the master thesis, the aims of the research, the methodology of the research and the structure of the master thesis.

1.1 Background

In effort of global CO_2 emission reduction, wind energy become one of the rapidly growing clean and sustainable energy sources due to the reliability and cost efficiency. In Europe along, Krohn et al. (2009) reported that the investment for wind energy will be up to 20 billion Euros by 2030, while the cost of wind energy will significantly decline. In Krohn's report, approximately 75% of the wind energy cost came from the wind turbine equipment and installation. This aligns with Du's (2016) study, which estimated the operations and maintenance related cost is around 25%-30% of the wind energy investment.

According to Saeed (2008, p.14) wind turbine is a device that converts the wind's kinetic energy into electrical energy. Figure 1 shows wind turbines in a wind park. A typical wind park or wind farm consists of several wind turbines, which are normally installed in an offshore or distant from population area. This makes the maintenance and supervision of the wind park challenging.

Figure 1. Wind turbines in a wind park (ABB commercia 2020)

In Saeed (2008, p.14) study, he introduced how the wind turbine works. First, the wind energy turns three blades around a rotor. The rotor is connected to the main shaft, which spins a generator to create electricity.

Due to the nature of wind power, a wind turbine converter that is capable of adjusting the generator frequency and voltage to the grid is required. Most importantly, the wind turbine converter plays an important role in helping create the perfect wind economy. ABB is one of the technical leaders to offer the state-of-the-art turbine converters to the market. In addition, ABB is a pioneer to provide the most sophisticated cloud services, which is available in the market, for turbine converter's maintenance and supervision.

In recent years, many researches in ABB have been conducted to aim improving the turbine converter maintenance cost efficiency and enhancing the reliability and maintainability of the turbine converters. This thesis is inspired by the latest researches and try to explore some new ideas.

1.2 Motivation and Aim of the Study

The thesis is motivated by the clean and sustainable energy development trend. It is easier and faster for the market to take the advantages of wind energy if the cost of wind energy can be reduced so that it is considerably cheaper than the overall cost of fossil fuel energy. One way to reduce the operation and maintenance cost of wind energy is to alert faults to the operators as earlier as possible.

In the effort of further reducing the operations and maintenance related cost, the thesis researches on possible machine learning models to automatically detect anomalies of the wind turbine converter, which is the key component inside a wind turbine.

The thesis experiments on two different unsupervised machine learning models based on the data collected from a wind park located in offshore area throughout the whole year 2018. Because ABB only provides the turbine converters as the one of the components of the wind turbine, the collected data is only limited to the turbine converters in the wind park. Thus, the research mainly focuses on the anomaly detection of the turbine converters. This is important to the wind park operation and maintenance because turbine

converters are the key component in the wind turbine. Chapter 1.4.3 introduces wind turbine and turbine converters in details.

The research questions of the thesis are:

> What are unsupervised machine learning models suitable for automatically detecting anomalies of wind turbine converters?
>
> What are the models' performance and accuracy?

The aim of the thesis is to answer the research questions and point out possible directions to future study or commercial applications. The conclusion is to present how the experimental models perform and if they are worth of continuing exploration.

1.3 Data and Methods

The data used in this research was collected from a wind park, which consists of 50 wind turbines. Each wind turbine equipped with an ABB wind turbine converter, which is the key component of the wind turbine. From each ABB wind turbine converter, the data was collected in one-minute interval throughout the whole year 2018. However, the amount of data does not distribute evenly for all wind turbine converters. Figure 2 shows the amount of data for each wind turbine converters. In order to achieve better accuracy, the wind turbine converters, which have significantly less data compared to the others, are excluded from the research, for example, wind turbine converter number 5 and 25. The missing data could be caused by different reasons, for instance, maintenance break, turbine converter's faults or network communication error.

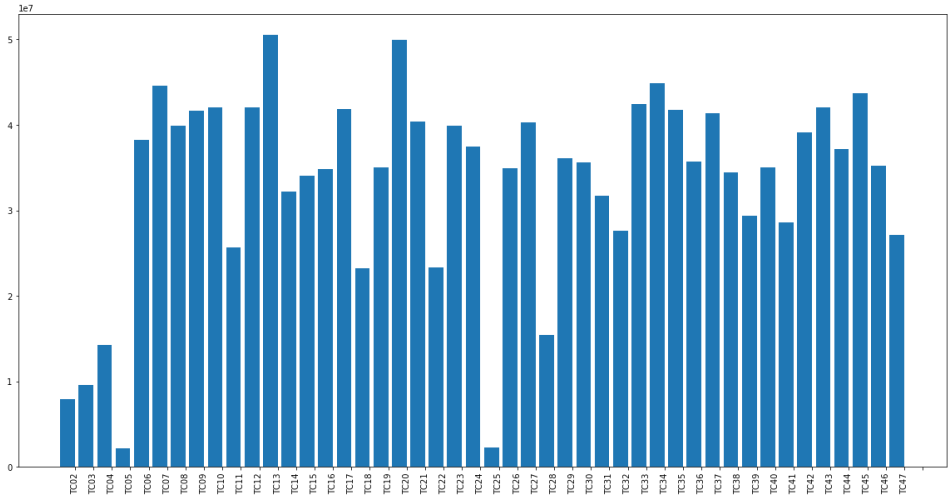


Figure 2. Counts of wind turbine converters collected data

The collected data includes measurements of speed, current, voltage, power, torque, temperature and pressure from different parts of the device. These measurements are identified as parameters, which are predefined by manufacture. The parameters are identically configured for all wind turbine converters. For all wind turbine converters, 84 parameters' values were simultaneously collected via ethernet connection. The parameters are categorized into predefined parameter groups. Therefore, each collected data point is identified as the combination of wind turbine converter ID, parameter group index and parameter index. Figure 3 shows example data from wind turbine converter #13.

	timestamp	value	par_group	par_index
0	2018-05-10 12:27:18	0.0	5	51
1	2018-05-10 12:26:21	19.0	5	51
2	2018-05-10 12:25:24	34.0	5	51
3	2018-05-10 12:24:20	0.0	5	51
4	2018-05-10 12:19:23	19.0	5	51
5	2018-05-10 12:17:21	0.0	5	51
6	2018-05-10 12:16:28	19.0	5	51

Figure 3.Example data for wind turbine converter #13

1.4 Definitions

This chapter introduces key definitions and terminologies used in the thesis.

1.4.1 Anomaly

There are many definitions of anomaly. The one defined by Cook et al (2019) is the most relevant one to this thesis: "*the measurable consequences of an unexpected change in state of a system which is outside of its local or global norm*".

Typically, anomaly defines partners that do not follow normal or expected behavior in a system. Often it is also called exception or outlier. However, an anomaly is not necessarily a fault in the system, it could be caused by system running in different mode, which is completely normal.

1.4.2 Anomaly Detection

Rana et al (2016) presents that anomaly can be detected with different techniques, for instance, statistics and machine learning. Anomaly detection works on different type of data, typically on time-series data, which can be either univariate or multivariate. This thesis focuses on anomaly detection with unsupervised machine learning techniques on multivariate time-series data.

Anomaly detection is widely applied in different areas, for instance, cyber intrusion detection, stock price manipulation detection, faults detection and so on.

1.4.3 Wind Turbine Convertor

A wind turbine turns kinetic energy from the wind into electrical energy into the power grid. Figure 4 depicts a typical structure of the wind turbine along with the grid feed-in.

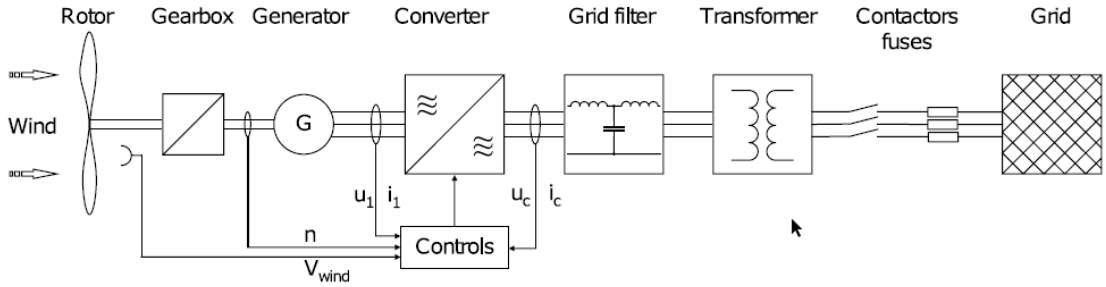


Figure 4. Structure and components of the drive train and the grid feed-in of a wind turbine with a fully-rated converter (Fuchs 2014, p.275).

The wind turbine is capable of operating in variant wind speed. The task of the converters is to convert the alternating current of the generator-side frequency to the grid-side stable frequency, therefore, the converter is also named as frequency converter. The frequency converter performs two functionalities. On generator side, the converter performs speed control and is operated with a variable frequency of the converter. On the power grid side, it allows the power to be fed into the grid and is operated with the stable grid frequency of 50 or 60 Hz (Fuchs 2014).

1.4.4 Machine Learning

David B (2012) defines machine learning as the body of research related to automated large-scale data analysis, which includes many of the traditional areas of statistics. However, machine learning mainly focuses on mathematical models and prediction.

There are two types of machine learning: supervised machine learning and unsupervised machine learning. Supervised machine learning analyses data which contains significant information, referred as labeled data, while the unsupervised machine learning analyses unlabeled data, which no difference between training and test data set. Typically, anomaly detection falls into unsupervised machine learning according to Shai (2014).

1.5 Structure of the Thesis

This thesis is structured as follow: Chapter 2 discusses the theoretical base of data analytics, its applications in the wind industry, and presents the concepts of machine learning and anomaly detection. Chapter 3 introduces the algorithms used for machine learning in this research, while Chapter 4 discusses and evaluates the results of the work. Finally, Chapter 5 concludes and sums up this thesis, and outlines the possible directions of future works.

2 RELATED WORK

Rana et al (2016) summarized a few machine learning techniques for time-series anomaly detection, including statistical, classification, clustering, knowledge based and so on. As introduced in Chapter 1.4.4, classification works on labeled data, thus it falls under the supervised machine learning. On the other hand, clustering is primarily unsupervised machine learning. In Rana's report, a few examples of clustering models are introduced, such as DBSCAN, ROCK, SNN and K-Means.

Cook et al (2019) present that Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) are proved to be effective for IoT data. Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) are also commonly used to detect anomaly by predicting the trend in the future. Cook et al (2019) also point out that Principal Component Analysis (PCA) can be used to reduce the complexity of

multivariate time-series data. When detecting anomaly on multivariate time-series data, Multiple Kernel Anomaly Detection (MKAD) is often used as one of the clustering models.

Both Rana et al (2016) and Cook et al (2019) summarize that the anomaly detection technique can highly depend on the data and system under the analysis. There is no silver bullet works for generic purpose, and it is common to apply ensemble models to detect the anomaly on time-series data.

This thesis works with the un-labeled data; therefore, classification technique is not applicable, clustering approach is used.

2.1 Definition

Anomaly detection is one of the data analysis areas, which is widely applied to scientific and financial field. Anomaly detection for time-series data is one of the hot topics in machine learning research. In Wu's (2016) research, the time-series data can be defined as

$$S = \{v_i(1), v_i(2), \cdots, v_i(t), \cdots, v_i(n)\} \quad (1)$$

In which t $(t = 1,2, \cdots, n)$ is defined as time, i $(i = 1,2, \cdots, m)$ is defined as variable. $v_i(t)$ is defined as the record of the variable i on time t . When $m = 1$, S is defined as univariate time-series. When $m > 1$, S is defined as multivariate time-series.

2.2 Anomaly Detection with Density Based Cluster

In previous study of the topic, Stikhin (2019) in his master thesis uses unsupervised machine learning model to detect anomaly in wind turbine converters. With his model, the result is very encouraging and positive. The topic of this thesis is inspired by Stikhin's research. The same data set is used in this research as in Stikhin's.

Stikhin implements an algorithm includes dimensionality reduction with principal component analysis (PCA), density-based clustering and distance-based nearest neighbor analysis.

One month of collected data from selected wind turbine converters is processed before feed into the model. The data from all wind turbine converters is taken from the same parameters, and grouped into four categories, which are closely correlated to certain function of the device. These categories are separately analyzed with the same model. This ensures the anomaly detection in the device, because the fault can happen to some parts of the device, while the other parts work normally. The four categories are described in Chapter 3.2 in details. Then the data is normalized and optimized with PCA model. Two components are selected in order to create a 2-dimensional grid. A number of unique values, i.e., the number of converters with similar values, is calculated for every cell of the grid. If this number does not pass the defined threshold for a normal cluster, the cell and all its values are marked as suspected, while the other points are marked as normal. K-d tree is used to calculate the mean distance between the points in the normal clusters. If a point is too far away from the normal ones (according to a threshold), it is marked as outlier. In the end, for each converter, the percentage of outliers are calculated from all points. If the percentage is higher than the defined value, the converter is considered as behaving anomaly.

The advantage of this model is that it is simple to implement and easy to understand, while the performance is considerably fast. However, the main drawback of the model is that it is not fully automated: a predefined threshold and a predefined percentage value are required in order to produce the result. In addition, the predefined threshold and percentage values may vary depending on the time, weather or even location of the wind park. One of the aims of this thesis is to experiment models which can be automated without any predefined value.

2.3 Time Series Data Anomaly Detection with LSTM

Standard Neural Network is proved to be effective to analyze unstructured data such as image, but also has the limitation of analyzing sequential data. To overcome the problem, Sherstinsky (2019) described that Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) algorithms have been used to analyze sequential data, such as time-series data.

Unlike the other neural networks, which all the inputs are independent from each other, in RNN, all the inputs are related to each other. LSTM is an improved version of RNN, in which the vanishing gradient problem of RNN is resolved. Therefore, with time-series data such as language, Yulius (2018) experimented that LSTM is working more effective.

LSTM is a generic algorithm to analyze time-series data, which has been widely used in different fields. In Nguyen's (2018) study, the LSTM model from anomaly detection of social network was used to detect earthquake. And Zhu (2019) demonstrated LSTM can be used to detect anomalies on prognostic and health data.

2.4 Clustering Analysis Based on Dirichlet Process Gaussian Mixture Models

In study made by Tan et al (2019), the authors proposed another way of clustering multivariable data: clustering analysis based on Dirichlet Process Gaussian Mixture Models (DP-GMM). The research concluded that the usage of DP has made the training procedure unsupervised without the need for knowing the number of clusters. The simulated calculation shows that the Non-stationary Discrete Convolution kernel has improved the ability of kernel PCA in accounting for the heterogeneity of multivariate data.

In a GMM, an m-dimensional random variable x follows a Gaussian mixture model with J components:

$$x \sim N(\mu_j, \sum_j) \text{ with probability } \pi_j$$

$$\text{s.t.} \qquad \pi_j > 0 \; \forall \, j, \sum_{J=1}^{J} \pi_j = 1 \tag{2}$$

where ~ denotes that the random variable on the left side follows the probability distribution on the right hand side. $\mu_j, \sum_j$ are the mean vector and the covariance matrix of the j-th Gaussian component, respectively. The mixture proportion π_j is the probability of drawing from the j-th component.

In DP-GMMs, the sample x is drawn from a GMM. The Gaussian components in this mixture have parameters, for example, $\mu_j, \sum_j$ and π_j. These parameters follow the distributions generated by the DP in a Bayesian way. Therefore, a Dirichlet mixture of Gaussian distributions can be written as:

$$\begin{aligned} G(\cdot) &\sim DP(\alpha, G_0) \\ \mu_j, \sum_j &\sim G(\cdot) \quad for\ j = 1, \dots, J \\ \pi &\sim Dir(\frac{\alpha}{J}, \dots, \frac{\alpha}{J}) \end{aligned} \tag{3}$$

where $\mu_j, \sum_j$ are mean and covariance of the j-th component, J is the number of components in this mixture, and DP and Dir respectively stand for the Dirichlet Process and the Dirichlet Distribution.

2.5 Clustering Multivariate Time Series Using Hidden Markov Models

Hidden Markov Model is based on Markov chain, which models the state of a system with a time-series random variable. Gagniuc (2017) states that the Markov property suggests that the distribution for this variable depends only on the distribution of a previous state. In other words, if Markov chain property applies to a system, the system's future states can be predicted by its current state.

A Hidden Markov Model is a Markov chain for which the state is only partially observable. In other words, some of the states of the system are not observable, such as hidden to the observer. Therefore, the model is called Hidden Markov Model.

Hidden Markov Model has been used to cluster multivariate time-series data in the past in many different fields, such as financial and health care. In Shima's (2014) research, it was proved that Hidden Markov Model is a good model to cluster not only continues multivariate time-series data, but also categorical time-series data. The data is a set of N health trajectories T_i corresponding to N distinct individuals, where each trajectory is a matrix with d columns. Each column is a time-series of length l_i that takes values in either categorical or continuous variables. The d time-series will be in general correlated, and the variables are referred as the "observables". Although the assumption is that the

same d measurements are taken for all individuals, the length of trajectories is not necessarily the same across individuals. Then, Shima (2014) defined a meaningful distance $D(T_i; T_j)$ between trajectory T_i and trajectory T_j, and apply any clustering method that takes as input a distance matrix. Shima (2014) proposed to take advantage of Discrete Frechet distance to compute the distance between trajectories for continuous variables.

3 RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the method and algorithm used in the thesis. The thesis researched two approaches to identify the anomaly turbine converters. The goal is to compare the results and evaluate the advantages and disadvantages of the approaches, finally analyze the differences between the results of two approaches.

The first approach is based on Hidden Markov Model. The open source library hmmlearn (2019) is used to implement the model. Then Discrete Frechet, Dynamic Time Warping (DTW) and Partial Curve Mapping (PCM) are used to calculate the distance between any two turbine converter datasets. The open source library similaritymeasures (2020) is used to calculate the distances. Then the distance matrix is used to cluster the turbines converters. Density-Based Spatial Clustering of Applications with Noise (dbscan) (2011) is selected to perform the clustering.

The second approach is based on Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which is unsupervised machine learning clustering model. The HDBSCAN is implemented in open source library called hdbscan (2020).

The details of the libraries used in the experiments are introduced in Chapter 3.4.

3.2 Data

Theoretically every turbine converter data contains values from 84 parameters continuously collected every minute for a year. But as shown in Figure 2, the data counts are quite different between turbines. There are many reasons for this, for example, some of the turbines may be in maintenance service for some period, some of the turbines may be out of service due to fault. Nevertheless, most of the turbine converters data is good enough for data analytic purpose, especially during April and May, therefore, in all models, the April data is used for training and May data is used for testing. The reason to focus on these months is that Stikhin (2019) experimented data in these two months. In order to compare the results with Stikhin's, the research uses the same data. In addition, Stikhin had researched the raw data for the whole year, and concluded that these two months data is the most reliable and meaningful from data analytics perspective.

The data can be summed up as follows:

- Generator-side values from the inverter unit (INU): speed, current, torque, DC voltage, output frequency, voltage, power
- Grid-side values from the insulated-gate bipolar transistor (IGBT) supply unit (ISU): line current, active and reactive power, converter current, negative sequence current
- Generator-side diagnostic values: temperatures for the control board, inverter, insulated-gate bipolar transistor (IGBT), main circuit interface (INT) board, incoming unit (ICU), LCL filter, inductors and capacitors
- Grid-side diagnostic values: temperatures for the converter, IGBTs, INT board, switching frequency, miniature circuit breaker (MCB) closing time counter, inlet and outlet cooling liquid temperatures and coolant pressures
- Ambient temperatures as recorded by the sensors for both generator- and grid-side modules
- Faults and warnings

Every converter contains four generator-side modules and six grid-side modules, and the temperature values are collected from two sensors for each of the modules.

The data is stored in a database in Microsoft Azure Databricks platform. The schema of the database is shown in Table 1.

Table 1. Schema of the turbine converter data in Databricks

Name	Data Type
Timestamp	timestamp
Value	double
Par_name	string
Turbine	string
Par_group	int
Par_index	int

According to Stikhin (2019), in order to maintain focus on the most common fault of the turbine converters, the parameters are categorized to four Analysis Groups by functionality of the turbine converter. The four Analysis Groups are shown in Table 2.

Table 2. Turbine converter data Analysis Groups

Analysis Group	Source converter side	Selected parameters
1	Generator	Inductor 1 temperature
	Generator	Inductor 2 temperature
	Grid	Active power P
	Grid	Reactive power Q
2	Grid	Inlet cooling liquid temperature
	Grid	Outlet cooling liquid temperature
	Grid	Active power P
	Grid	Reactive power Q
3	Grid	Ambient temperature measured by the sensors at the upper parts of modules 1–6

	Grid	Ambient temperature measured by the sensors at the lower parts of modules 1–6
	Grid	Active power P
	Grid	Reactive power Q
4	Grid	Phase voltage U1
	Grid	Phase voltage V1
	Grid	Phase voltage W1
	Grid	Main voltage 1 positive sequence
	Grid	Main voltage 1 negative sequence
	Grid	Active power P
	Grid	Reactive power Q

These Analysis Groups and converter parameters are selected by domain experts. The same model is applied to all four Analysis Groups.

3.3 Methods

3.3.1 Hidden Markov Model

Hidden Markov Model is a proved effective model to detect anomaly with time-series data. As shown in Table 2, for each data Analysis Group, Hidden Markov Model is used to analyze multivariate time-series data. This thesis applied a method proposed by Shima (2014). The goal is to detect the anomaly turbine converters. In order to achieve the goal, the turbine converters need to be clustered, and to cluster the turbine converters, the distance between the Hidden Markov Model needs to be calculated, because the distance between each turbine converter's original dataset is ill-defined.

The following steps describe the algorithm:

1. Query Analysis Group 1 (as shown in Table 2) raw data from certain time interval, the raw data is not sorted and structured as per parameter value per row. The data contains timestamp, parameter group, parameter index and value.
2. Transform the raw data to pandas dataframe, and combine the raw data to per turbine converter and sorted according to time stamp
3. Train Hidden Markov Model with transformed data as input, which is multivariate time-series data. One Hidden Markov Model per turbine converter
4. Test Hidden Markov Model with transformed data from another timeframe
5. Calculate the emission probability distribution from each Hidden Markov Model, saved the models and the emission probability distribution for each turbine converter
6. Calculate the distance matrix between every turbine converter's emission probability distribution. The distance is calculated with one of the three different methods: Discrete Frechet, Dynamic Time Warping (DTW), and Partial Curve Mapping (PCM).
7. Use DBSCAN cluster model with distance matrix, so that the model finds multiple clusters.
8. Output all turbine converter's clusters with the turbine converter's index.
9. Repeat Step 1 to Step 8 for rest of the Analysis Groups.

3.3.2 HDBSCAN clustering library

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is a clustering algorithm extended from DBSCAN by converting it into a hierarchical clustering algorithm, and then extract the hierarchical clusters into flat and stable clusters. According to Brendan Bailey (2017), it has the advantages over other clustering algorithm such as performance, intuition and suitable for data of varying density. It was used in other ABB researches, and proved to be an effective clustering method on time-series data. For comparison and validation purpose, the thesis also used the algorithm to cluster the same data used in Hidden Markov Model. The comparison is analyzed in Chapter 4.2.

The following steps describe the algorithm:

1. Query Analysis Group 1 (as shown in Table 2) raw data from the same time interval as the Hidden Markov Model algorithm, the raw data is not sorted and

structured as per parameter value per row. The data contains timestamp, parameter group, parameter index and value.

2. Transform the raw data to pandas dataframe, and combine the raw data to per turbine converter and sorted according to time stamp.
3. Normalize the data between -1 and 1
4. Transform the dataframe so that the row of the dataframe represents the turbine converters, and the column of the dataframe represents the normalized parameter value. The columns contain the concatenated sorted time-series data, e.g. column 0 to n contains the parameter values of Inductor 1 temperature, column n+1 to 2n contains the parameter values of Inductor 2 temperature and so on.
5. Use HDBSCAN cluster model with data processed in step 4, so that the model finds multiple clusters.
6. Output all turbine converters clusters with the turbine converter's index.
7. Repeat Step 1 to Step 6 for rest of the Analysis Groups.

The result dataframe of step 2 is the same as the dataframe of step 2 in Hidden Markov Model algorithm.

3.4 Research Design

The raw data is stored in Microsoft Azure data lake. The data analysis is done with Microsoft Azure Databricks (2020), which provides similar notebook features as Jupyter notebook. Microsoft Azure Databricks (2020) integrates python runtime libraries and Apache Spark. In addition, it provides services such as installing python libraries and access to file system. The data analysis of the thesis is written in python and executed in Microsoft Azure Databricks (2020).

In order to keep focus on the research topic, the thesis used ready-made open source libraries instead of implementing own libraries. The python libraries are used in this thesis are list in Table 3.

Table 3. Python libraries used for data analysis

Name	Purpose	Version
hmmlearn (2019)	Hidden Markov Model	0.2.3
hdbscan (2020)	HDBSCAN cluster model	0.8.26
matplotlib (2019)	plot graphs	3.0.3
numpy (2019)	mathematical computing	1.16.2
pandas (2019)	data analytics	0.24.2
pyspark (2019)	spark data query	2.4.4
scipy (2019)	required by hmmlearn	1.2.1
similaritymeasures (2020)	calculate distance between two Hidden Markov Models' emission probability distribution	0.4.2
scikit-learn (2019)	machine learning library	0.20.3

One of the obstacles of the thesis is the performance of the Discrete Frechet distance calculation, the problem is solved by using multi-threads and splitting the calculation into multiple hosts. In addition, DTW and PCM are also studied to compare the performance and accuracy.

The thesis also research on how the amount of data effects the performance and accuracy of different algorithms introduced in Chapter 3.3.1 and Chapter 3.3.2. The findings are described in Chapter 4.3.

In addition to compare the results of two algorithms designed in this thesis, the results are also compared to the work done by Stinkhin (2019). The results presented by Stikhin are considerably accurate, because it uses mathematical computing on 2-dimension space to count the outliers. The details of Stinkhin's work is described in Chapter 2.2.

4 RESULTS

This chapter describes the results of the research, analyze and compare the results from different experiments.

In order to validate the results, the data is visualized by using heat map with active power and reactive power as the axis. The graphs are shown in Appendix 1 for Analysis Group 1, Appendix 2 for Analysis Group 2, Appendix 3 for Analysis Group 3, and Appendix 4 for Analysis Group 4.

4.1 Result of Hidden Markov Model

This chapter describes the results of clusters with Hidden Markov Model, which algorithm is described in Chapter 3.3.1.

The results are compared between different ways of calculating distance matrix, the methods used in this thesis are: Dynamic Time Warping (DTW), Discrete Frechet and Partial Curve Mapping (PCM).

Discrete Frechet is suggested by Shima (2014). There are two reasons to use DTW and PCM as well. The first reason is to compare and cross validate the results from different distance matrix calculation. The second reason is that Discrete Frechet distance calculation is slow, the more data the slower the calculation, while DTW and PCM are much faster. The experiments are to compare the performance versus the accuracy in order to determine the best model from both accuracy and performance point of view.Table 4 shows the calculation time with random sample data:

Table 4. Performance of different distance calculations

	100 samples	1000 samples	10000 samples
PCM	0,007257s	0,025864s	0,117573s
Discrete Frechet	0,151545s	14,222789s	1419,680883s
DTW	0,044014s	1,333476s	140,136014s

The calculation is done with Intel i7 2.7 GHz processor on windows 10 64-bits OS without multiple threads. Python version is 3.7.

As the Table 4 shows, Discrete Frechet distance calculation is the slowest among the three methods. The calculation time increases exponentially when the input data amount increases. The average daily data is around 1440 samples per parameter per turbine converter. In Analysis Group 1, 4 parameters data are collected, which results 5760 samples per day per turbine converter. The theoretical one-day data Discrete Frechet distance calculation time between any two turbine converters is roughly 6 minutes. The theoretical one-month data Discrete Frechet distance calculation time between any two turbine converters is roughly 90 hours. However, different turbine converters normally contain different number of samples due to variant reasons, such as network connectivity, power break or turbine converter maintenance. On the other hand, the Hidden Markov Model library requires the number of the data points is the same among all multivariate time-series data within any specific turbine converter. Therefore, the model always takes the minimum number of data points among all multivariate time-series data. This un-intentionally reduces the distance matrix computing time.

The experiment is also done by taking random samples from the raw data so that the number of data is reduced, but the result of this approach is not satisfactory. The reason could be the fact that Hidden Markov Model strongly depends on the order and samples of the time-series data in order to make accurate clustering.

In order to compare and cross validate the result, this thesis chose the same date as in Stikhin's (2019) research. There are 38 turbine converters contain valid data in the selected date. Therefore, the total calculation time for the Discrete Frechet distance matrix is around 6 days with the same computing power as used to calculate the result in Table 4.

4.1.1 Analysis Group 1

Analysis Group 1 studies the correlation between two inductors' temperature and active and reactive power as shown in Table 2.

When using DTW to calculate the distance matrix, the cluster result is shown in Table 5.

Table 5. One-day Analysis Group 1 data clusters with Hidden Markov Model by using DTW

	Cluster 1	Cluster 2	Cluster 3
DTW	TC02, TC06, TC07, TC08, TC09, TC10, TC13, TC14, TC17, TC18, TC20, TC22, TC23, TC24, TC26, TC27, TC31, TC33, TC35, TC36, TC37, TC38, TC39, TC41, TC43, TC44, TC45, TC47	TC12, TC21, TC32, TC42, TC46	TC03, TC28, TC29, TC30,

When using Discrete Frechet to calculate the distance matrix, the cluster result is shown in Table 6.

Table 6. One-day Analysis Group 1 data clusters with Hidden Markov Model by using Discrete Frechet

	Cluster 1	Cluster 2
Discrete Frechet	TC02, TC03, TC06, TC07, TC08, TC10, TC12, TC13, TC14, TC17, TC18, TC21, TC22, TC23, TC24, TC27, TC28, TC29, TC30, TC31, TC32, TC33, TC35, TC36, TC37, TC38, TC41, TC43, TC45, TC46, TC47	TC09, TC20, TC26, TC39, TC42, TC44

When using PCM to calculate the distance matrix, the cluster result is shown in Table 7.

Table 7. One-day Analysis Group 1 data clusters with Hidden Markov Model by using PCM

	Cluster 1	Cluster 2	Cluster 3	Cluster4	Noise
PCM	TC02, TC10, TC14, TC18, TC21, TC22, TC24, TC27, TC28, TC29, TC30, TC32, TC33, TC39, TC43, TC44, TC45, TC46	TC03, TC08, TC12, TC23, TC35, TC42, TC47	TC06, TC26, TC31,	TC07, TC09, TC17, TC37, TC41,	TC13, TC20, TC36, TC38,

As shown in the tables, three results are completely different in both terms of number of clusters and turbine converters in clusters. The validation shows that the result with Discrete Frechet distance matrix is the most accurate among three calculations by referring to Appendix 1. For example, turbine converter #07, #10, #45 and #46 are clearly belong to the same cluster as shown in Figure 5.

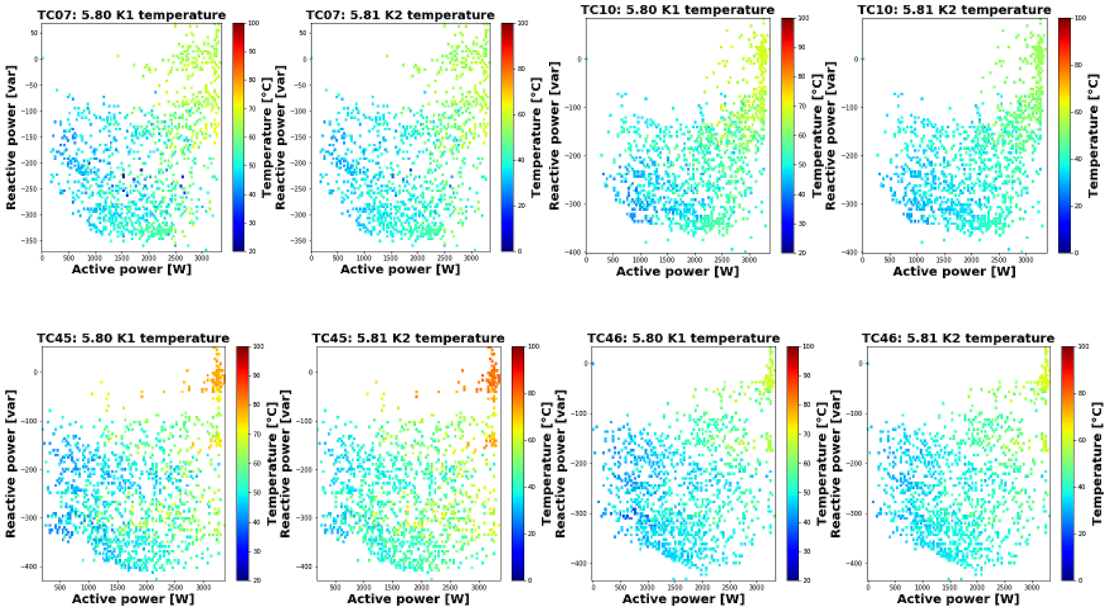


Figure 5. Example heat maps from same cluster.

However, the visual presentation of the data is not 100% reliable, turbine converter #20 and #44 belong to the same cluster in both Hidden Markov Model and Stikhin's (2019) model, but visually they belong to the different clusters as shown in Figure 6. Therefore, the heat maps must be used together with other references for cross validation.

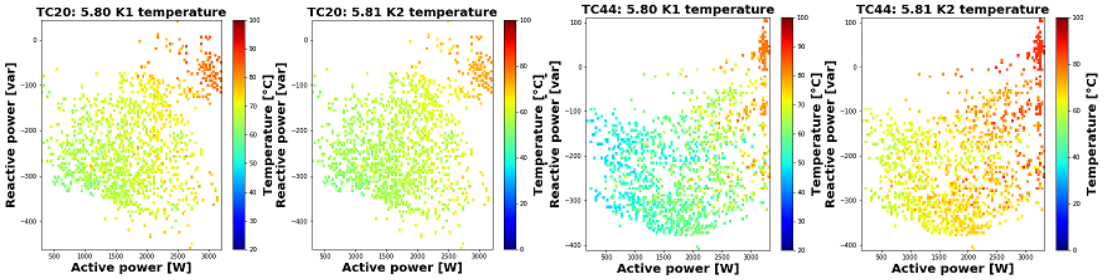


Figure 6. Example heat maps from same cluster, but visually different.

By comparing with Stikhin's (2019) result, Hidden Markov Model does not give satisfactory result. As in Stikhin's result, turbine converter #20 and #37 clearly belong to

the same cluster. However, these two turbine converters are in different clusters as shown in Table 6. The heat maps of turbine converter #20 and #37 are shown in Figure 7.

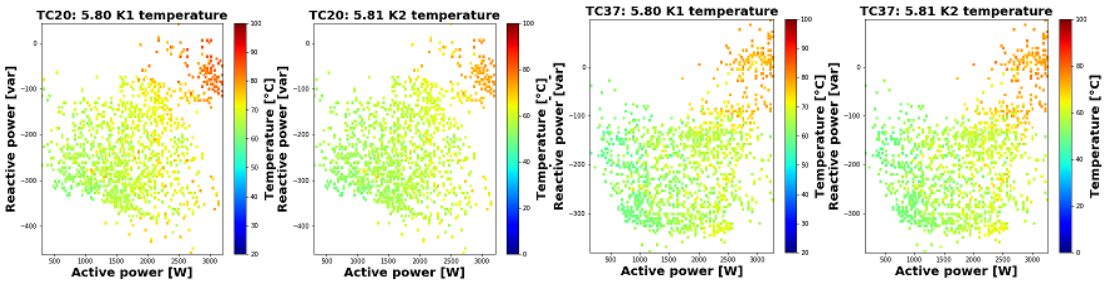


Figure 7.Heat map of turbine converter #20 and #37.

It is also worth to note that the distribution of the data points in the heat map does not matter, the color of the data points is more important on determining the cluster.

4.1.2 Analysis Group 2

Analysis Group 2 studies the correlation between inlet and outlet cooling liquid's temperature and active and reactive power as shown in Table 2.

When using DTW to calculate the distance matrix, the cluster result is shown in Table 8.

Table 8. One-day Analysis Group 2 data clusters with Hidden Markov Model by using DTW

	Cluster 1	Cluster 2
DTW	TC02, TC03, TC06, TC07, TC08, TC09, TC10, TC13, TC14, TC16, TC17, TC22, TC24, TC26, TC27, TC29, TC31, TC32, TC33, TC35, TC37, TC38, TC39, TC41, TC42, TC44, TC46, TC47	TC18, TC20, TC21, TC23, TC28, TC30, TC43, TC45

When using Discrete Frechet to calculate the distance matrix, the cluster result is shown in Table 9.

Table 9. One-day Analysis Group 2 data clusters with Hidden Markov Model by using Discrete Frechet

	Cluster 1	Cluster 2
Discrete Frechet	TC02, TC03, TC06, TC07, TC08, TC09, TC10, TC13, TC14, TC16, TC17, TC18, TC20, TC21, TC22, TC23, TC24, TC27, TC28, TC29, TC31, TC33, TC35, TC37, TC38, TC39, TC40, TC41, TC42, TC45, TC46, TC47	TC26, TC30, TC32, TC38. TC43, TC44

When using PCM to calculate the distance matrix, the cluster result is shown in Table 10.

Table 10. One-day Analysis Group 2 data clusters with Hidden Markov Model by using PCM

	Cluster 1	Cluster 2	Cluster 3	Noise
PCM	TC22, TC47	TC02, TC03, TC07, TC08, TC14, TC16, TC17, TC24, TC37, TC39, TC40, TC41, TC43, TC44, TC45, TC46	TC06, TC09, TC21, TC20, TC26, TC27, TC28, TC29, TC30, TC31, TC32, TC35,	TC10, TC13, TC18, TC20, TC23, TC33, TC38, TC42

As shown in the tables, the three results are significantly different in both terms of number of clusters and turbine converters in clusters. The validation shows that the none of the results is satisfactory by referring to Appendix 2 and results from Stikhin's (2019) research.

4.1.3 Analysis Group 3

Analysis Group 3 studies the correlation between 6 delta temperature of upper and lower part of the module and active and reactive power as shown in Table 2.

When using DTW to calculate the distance matrix, the cluster result is shown in Table 11.

Table 11. One-day Analysis Group 3 data clusters with Hidden Markov Model by using DTW

	Cluster 1	Cluster 2	Cluster 3
DTW	TC02, TC03, TC08, TC09, TC13, TC14, TC16, TC18, TC17, TC21, TC22, TC23, TC24, TC26, TC27, TC28, TC30, TC31, TC32, TC33, TC35, TC37, TC38, TC40, TC41, TC43, TC44, TC47	TC06, TC07, TC10, TC20	TC12, TC17, TC29, TC36, TC42, TC45, TC46

When using Discrete Frechet to calculate the distance matrix, the cluster result is shown in Table 12.

Table 12. One-day Analysis Group 3 data clusters with Hidden Markov Model by using Discrete Frechet

	Cluster 1	Cluster 2	Cluster 3	Noise
Discrete Frechet	TC02, TC14, TC29	TC06, TC07, TC10, TC20	TC03, TC08, TC09, TC12, TC13, TC16, TC17, TC18, TC21, TC22, TC23, TC24, TC26, TC28, TC30, TC31, TC32, TC33, TC36, TC37, TC38, TC40, TC41, TC42, TC43, TC44, TC45, TC46, TC47	TC27

When using PCM to calculate the distance matrix, the cluster result is shown in Table 13.

Table 13. One-day Analysis Group 3 data clusters with Hidden Markov Model by using PCM

	Cluster 1	Cluster 2	Noise
PCM	TC02, TC08, TC09, TC14, TC16, TC18, TC21, TC23, TC24, TC26, TC27, TC28, TC30, TC31, TC32, TC37, TC38, TC40, TC41, TC43, TC44, TC45, TC47	TC03, TC06, TC12 TC22, TC29,	TC07, TC10, TC13, TC17, TC20,

		TC33, TC36	TC42, TC46

As shown in the tables, DTW and Discrete Frechet give more closer clusters than previous two Analysis Groups: one of the three cluster overlaps completely, and one of the clusters overlaps significantly. While PCM gives quite different result from both DTW and Discrete Frechet.

However, when validate the results by referring to Appendix 3 and results from Stikhin's (2019) research, none of the methods provides satisfactory result.

4.1.4 Analysis Group 4

Analysis Group 4 studies the correlation between 5 voltages and active and reactive power as shown in Table 2.

When using DTW to calculate the distance matrix, the cluster result is shown in Table 14.

Table 14. One-day Analysis Group 4 data clusters with Hidden Markov Model by using DTW

	Cluster 1	Cluster 2
DTW	TC02, TC03, TC06, TC10, TC12, TC13, TC16, TC17, TC18, TC20, TC21, TC22, TC23, TC24, TC26, TC27, TC28, TC29, TC30, TC32, TC33, TC35, TC36, TC37,	TC08, TC09, TC31
	TC38, TC39, TC40, TC41, TC42, TC43, TC44, TC45, TC47	

When using Discrete Frechet to calculate the distance matrix, the cluster result is shown in Table 15.

Table 15. One-day Analysis Group 4 data clusters with Hidden Markov Model by using Discrete Frechet

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Discrete Frechet	TC08, TC17, TC27, TC40	TC41, TC42, TC43, TC44, TC45, TC47	TC02, TC03, TC06, TC09, TC10, TC12, TC13, TC16, TC18, TC20, TC21, TC22, TC23, TC24, TC26, TC28, TC29	TC30, TC31, TC32, TC33, TC35, TC36, TC37, TC38, TC39

When using PCM to calculate the distance matrix, the cluster result is shown in Table 16.

Table 16. One-day Analysis Group 4 data clusters with Hidden Markov Model by using PCM

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Noise
PCM	TC03, TC16, TC18, TC27, TC38	TC12 TC23, TC42, TC47	TC39, TC40, TC43, TC44	TC02, TC08, TC18, TC22, TC33, TC35, TC36, TC41, TC45	TC24, TC26, TC28, TC29, TC32, TC37	TC06, TC30	TC10, TC13, TC17, TC20, TC21, TC31

As shown in the tables, three results are completely different in both terms of number of clusters and turbine converters in clusters. When validate the results by referring to Appendix 3 and results from Stikhin's (2019) research, none of the methods provides satisfactory result.

4.2 Result of HDBSCAN clustering model

This chapter describes the results of HDBSCAN clustering model, which algorithm is described in Chapter 3.3.2.

The results are compared with Hidden Markov Model results.

4.2.1 Analysis Group 1

The cluster result for one-day Group 1 data is shown in Table 17.

Table 17. One-day Analysis Group 1 data clusters with HDBSCAN cluster model

Cluster 1	Cluster 2	Noise
TC08, TC09, TC12, TC13, TC14, TC18, TC20, TC21, TC22, TC23, TC24, TC26, TC27, TC28, TC29, TC30, TC31, TC32, TC33, TC35, TC37, TC38, TC40, TC41, TC42, TC43, TC44, TC47	TC06, TC07, TC10	TC02, TC03, TC17, TC45, TC46

As shown in the table, HDBSCAN does not provide similarity with any of the results from Hidden Markov Model. While Stikhin's (2019) research suggests that most of the turbine converters in the noise group should have been part of the cluster 2. Therefore, HDBSCAN does not provide satisfactory result for Analysis Group 1.

4.2.2 Analysis Group 2

The cluster result for one-day Analysis Group 2 data is shown in Table 18.

Table 18. One-day Analysis Group 2 data clusters with HDBSCAN cluster model

Cluster 1	Cluster 2
TC02, TC06, TC07, TC10, TC45, TC46	TC03, TC08, TC09, TC13, TC14, TC16, TC17, TC18, TC20, TC21, TC22, TC23, TC24, TC26, TC27,

	TC28, TC29, TC30, TC31, TC32, TC33, TC35, TC37, TC38, TC39, TC40, TC41, TC42, TC43, TC44, TC47

As shown in the table, HDBSCAN does not provide similarity with any of the results from Hidden Markov Model. However, it gives a perfect match result with Stikhin's (2019) research. Therefore, HDBSCAN provides a satisfactory result for Analysis Group 2.

4.2.3 Analysis Group 3

The cluster result for one-day Analysis Group 3 data is shown in Table 19.

Table 19. One-day Analysis Group 3 data clusters with HDBSCAN cluster model

Cluster 1	Noise
TC02, TC03, TC06, TC07, TC08, TC09, TC10, TC12, TC13, TC14, TC16, TC17, TC20, TC21, TC22, TC23, TC24, TC26, TC27, TC28, TC29, TC30, TC31, TC32, TC33, TC37, TC40, TC41, TC42, TC43, TC44, TC45, TC46, TC47	TC18, TC36, TC38

As shown in the table, HDBSCAN does not provide similarity with any of the results from Hidden Markov Model. Most importantly, the HDBSCAN model does not provide a clear cluster among all the data, therefore, HDBSCAN model does not provide a satisfactory result for Analysis Group 3.

Analysis Group 3 and 4 are more difficult to validate visually because there are more data points in these Analysis Groups. As shown in Appendix 3 and Appendix 4, Analysis Group 3 contains 6 data points, and Analysis Group 4 contains 5 data points, while Analysis Group 1 and 2 contains only 2 data points.

4.2.4 Analysis Group 4

The cluster result for one-day Analysis Group 4 data is shown in Table 20.

Table 20. One-day Analysis Group 4 data clusters with HDBSCAN cluster model

Cluster 1	Noise
TC02, TC03, TC08, TC09, TC10, TC12, TC13, TC16, TC17, TC18, TC20, TC21, TC22, TC23, TC24, TC26, TC27, TC28, TC29, TC30, TC31, TC32, TC33, TC35, TC36, TC37, TC39, TC40, TC41, TC42, TC43, TC44, TC45, TC47	TC06, TC17, TC38, TC42

As shown in the table, HDBSCAN does not provide similarity with any of the results from Hidden Markov Model. Most importantly, the HDBSCAN model does not provide a clear cluster among all the data, therefore, HDBSCAN model does not provide a satisfactory result for Analysis Group 4.

4.3 Effects of amount of data

The comparison between Hidden Markov Model and HDBSCAN cluster model is also done for one-month data. To save the calculation time, in this research, the distance matrix is calculated by DTW only and for Analysis Group 2 only.

The result of Hidden Markov Model is shown in Table 21.

Table 21. One-month Analysis Group 2 data clusters with Hidden Markov Model by DTW

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Noise
TC02, TC07, TC31, TC35, TC45, TC46	TC03, TC06, TC08, TC11, TC13, TC14, TC16, TC17, TC20, TC23, TC28, TC29, TC30, TC32, TC33, TC43, TC44, TC47	TC21, TC24, TC27,	TC12, TC39, TC40	TC09, TC10, TC15, TC18, TC26, TC37,TC41, TC42	TC22

The result of HDBSCAN model is shown in Table 22.

Table 22. One-month Analysis Group 2 data clusters with HDBSCAN

Cluster 1	Cluster 2	Noise
TC02, TC06, TC07, TC10, TC45, TC46	TC03, TC08, TC09, TC13, TC14, TC16, TC17, TC18, TC20, TC21, TC23, TC24, TC26, TC27, TC28, TC29, TC30, TC32, TC33, TC35, TC37, TC39, TC40, TC41, TC42, TC43, TC44, TC47	TC22, TC31

As can be seen that the amount of data does not affect to the clustering result between Hidden Markov Model and HDBSCAN cluster model. By cross validating the result provided by Stikhin (2019), the HDBSCAN gives a better result than Hidden Markov Model with one-month data.

Compared with one-day results in Chapter 4.1.2 and Chapter 4.2.2, HDBSCAN model gives a quite similar clusters in one-month and one-day data. While Hidden Markov Model produces different clusters in both terms of number of clusters and turbine converters in each cluster. By cross validating the result in Stikhin's (2019) research, HDBSCAN model's result is more accurate. By examine the data, the selected day data aligns that in the selected month, meaning that anomaly group of turbine converters keep as anomaly through all month. Therefore, theoretically the clusters of one-day data should be the same as the clusters in on month data.

5 CONCLUSION

This chapter concludes the research and points out possible directions for future research.

5.1 Summary

In this thesis, two unsupervised machine learning models are used to detect the anomaly in turbine converters: Hidden Markov Model and HDBSCAN model. Hidden Markov Model is a bit more complex than HDBSCAN. Both models' input is the same data in order to compare and validate the results. In addition, the data visualization method and Stikhin's (2019) results are used to cross validate the results.

The data used as input of the models is categorized into 4 Analysis Groups, which are introduced in chapter 3.2. Each the Analysis Group represents a critical module of the turbine converter. By examining the parameter values belonging to the model one can provide reliable evaluation on the anomaly turbine converters.

In Hidden Markov Model, three distance computation methods are used to compare and validate the results: Discrete Frechet, Dynamic Time Warping, and Partial Curve Mapping. One-month data is used to train the model, and one-day data is used to test the model. The model is applied to all 4 Analysis Groups with 3 different distance computation methods used for each of the Analysis Group. The results are described in chapter 4.1. One of the challenges of the research is the computation time of Discrete Frechet distance matrix. Experiments have been conducted with the effort of reducing the number of data, but the results are not satisfactory. In the end, the problem is solved with multi-threads and distribute the computation to different hosts.

In HDBSCAN model, the same dataset is used to test as in Hidden Markov Model. However, HDBSCAN model has an additional step with the input data in order to achieve better results: input data is normalized between -1 and 1. The model is applied to all 4 Analysis Groups. The results are described in chapter 4.2.

One-month data is tested as well to evaluate the effects of amount of data feed to the models. Because of the distance computation time, one-month data is only applied to Analysis Group 2 with DTW as the distance computation method. But the result does not show any improvement compared to one-day data.

5.2 Discussion

Conceptually, multivariate time-series data analysis is a challenging mathematical task for a few reasons. In practice, time-series data are finite observations within an infinite time domain, in other words, the observations are always incomplete. Due to the nature of the world, small part of the observations are always noises, which have negative impact on the analysis result. With multivariate time-series data, the analysis is even more challenging, because the correlation between the univariate time-series data are normally unknown. Despite the challenges, the development in the recent year's research, especially in the machine learning field, also poses new opportunities. The research in this thesis is inspired by the latest scientific efforts and the experiments are aiming to explore new ideas on the topic.

In summary, both Hidden Markov Model and HDBSCAN model have problem of providing reliable clustering result. In addition, Hidden Markov Model requires significant computing power and time depending on the distance calculation method. Among all three distance calculation methods used in the thesis, Discrete Frechet takes the longest time, and PCM takes the shortest. However, PCM provides relatively the worst result, while DTW and Discrete Frechet give the similar results.

One assumption was made in Hidden Markov Model, the number of states is always set to 2, which is may not be the optimal value in certain situation. The reason to make such assumption is to assume that there is always a group of turbine converters which behave a bit differently from the rest of the turbine converters. On the other hand, the anomaly also depends on the defined threshold value. For instance, if the threshold represents the mathematical distance between turbine converters, when the threshold is large enough, all turbine converters are in the same cluster, and when the threshold is small enough, every turbine converter is in a separate cluster. According to Abou-Moustafa (2004) Hidden Markov Model is quite sensitive to its structural parameters, for example, the number of states and topology of the model. In practice, having a hard-coded number of states is not ideal for all situations.

One of the reasons to experiment with HDBSCAN model is that this model provides a way to make a cluster prediction to the new input data, which is not usual in most of the

available cluster models. The experiment however shows that the prediction does not provide reliable enough result.

Compared to Hidden Markov Model, HDBSCAN model provides better results especially for Analysis Group 2. The biggest difference compared to other Analysis Groups is that there is a clear gap between two clusters according to Stikhin's (2019) research. While the other Analysis Groups, the gaps between clusters are small. In some cases, there is no obvious anomaly in the selected month or day.

The experiment also shows that amount of data has no significant effects on the results for HDBSCAN model. On the other hand, Hidden Markov Model gives worse result when more data is used, but the time of calculation increases exponentially. From the raw data, some turbine converters behave differently in different days, which means they randomly change cluster within the month. Therefore, it is more challenging to cluster for longer period than shorter ones.

In practice, the models can be applied on daily base to notify the service engineers about the anomaly turbine converters. The models can be triggered automatically during the night, and the raw input data which was collected from last 24 hours is input to the models. There can be multiple models running simultaneously in order to compare and validate the results, so that the system provides the most reliable results to the users. The results can be classified according to predefined confidence levels. For instance, the intersection of the results from different models can be the highest certainty because the confidence of the anomaly is the highest. The certainty decreases as the confidence gets lower. Then the service engineer can examine the anomaly turbine converters more closely to determine the real faulty turbine converters.

The anomaly turbine converter provided by the models is not necessarily to be faulty, it can be normal behavior due to external condition changes. Or it could be totally new operational mode due to the change of configuration. The model only provides anomaly from mathematical perspective. It is still necessary to have a domain expert to examine the data and make the decision whether the turbine converter is in faulty state or not.

5.3 Future Work

In addition to turbine parameter values, the event logs can also be collected from ABB turbine converters. The event logs normally contain structured warnings and alerts happening in turbine converters. This is one of the critical sources for domain experts to investigate the faults in turbine converters. By combining the event logs with the parameter values, the Hidden Markov Model could give better results than using parameter values alone.

It is a known fact that the turbine converter operates differently in different seasons and in different weather conditions. Therefore, it is good to experiment with data from different day of the same month and the same day from different month, then compare the results to show that the external weather impacts on how the turbine converter works.

It is also good idea to compare the turbine converters in different wind parks. This of course depends on the data availability. The purpose of the comparison is to examine the portability and flexibility of the models. The aim is to fit the same model to all wind parks without significant changes.

The experiment results show that HDBSCAN model works much better with Analysis Group 2 than the other Analysis Groups. This suggests the possibility to use different machine learning models in different Analysis Groups. This may be even preferred approach, because the 4 turbine converter modules are rather working independently. In other word, one module's fault does not affect on the other models.

In addition, it is good to create an automatic workflow to use the models in real wind park. The anomaly detection models are the key component of the workflow, but other components of the workflow are necessary in order to provide a commercial service to the end users in practice, for instance, the data storage where the collected raw data are stored, a data platform where the models are executed, and a graphical user interface to configure and administrate the system.

Last but not the least, research on how the models can be used from the history anomaly detection results. The thesis studies the models with the input data independent from each other, meaning the result of yesterday's anomaly detection cannot be used in today's anomaly detection. While in practice, the anomaly can happen gradually from day to day

operation. By examining the past results, it can bring meaningful insight about the result in hand now. The user can also influence the result by manually conform the abnormal behaving turbine converters, which make the semi-supervised machine learning model, but should bring a better result.

REFERENCES

A. Cook, G. Mısırlı and Z. Fan. (2019). Anomaly Detection for IoT Time-Series Data: A Survey. *IEEE Internet of Things Journal*.

A. I. Rana, G. Estrada, M. Solé and V. Muntés. (2016). Anomaly Detection Guidelines for Data Streams in Big Data. *2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI), Dubai, 2016, pp. 94-98*.

Alex Sherstinsky. (2019). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena 404 (2020): 132306. Crossref. Web.*

Apache Spark. (2019). Apache Spark is a unified analytics engine for large-scale data processing. http://spark.apache.org/

Asif Saeed. (2008). Online Condition Monitoring System for Wind Turbine. *Blekinge Institute of Technology*. https://www.diva-portal.org/smash/get/diva2:830541/FULLTEXT01.pdf

Brendan Bailey. (2017). Lightning Talk: Clustering with HDBScan. https://towardsdatascience.com/lightning-talk-clustering-with-hdbscan-d47b83d1b03a

David Barber. (2012). Bayesian Reasoning and Machine Learning. *Cambridge University Press, ISBN-10: 0521518148*

Scikit-learn. (2011). Scikit-learn: Machine Learning in Python, Pedregosa et al. JMLR 12, pp. 2825-2830.

Fuchs, Friedrich W. (2014). Power Electronics and Generator Systems for Wind Turbines. *Chap. 8 in Understanding Wind Power Technology: Theory, Deployment and Optimisation, 1st ed., edited by Alois P. Schaffarczyk, translated by Gunther Roth, 273–339. Chichester: Wiley. isbn: 978-1-118-64751-6. Translation of "Leistungselektronik-Generatorsysteme für Windenergieanlagen" [in German] In Einführung in die Windenergietechnik, 1st ed. Munich: Hanser, 2012. ISBN: 978-3-446-43032-7*.

Gagniuc, Paul A. (2017). Markov Chains: From Theory to Implementation and Experimentation. *USA, NJ: John Wiley & Sons. pp. 1–256. ISBN 978-1-119-38755-8*.

Guangxuan Zhu, Hongbo Zhao, Haoqiang Liu, Hua Sun. (2019). A Novel LSTM-GAN Algorithm for Time Series Anomaly Detection. *2019 Prognostics & System Health Management Conference - Qingdao (PHM-2019 Qingdao)*

Hdbscan Library. (2020). Open source python library for hdbscan cluster model. https://hdbscan.readthedocs.io/en/latest/

Hmmlearn Library. (2019) Open source python library for unsupervised Hidden Markov Model. https://hmmlearn.readthedocs.io/en/stable/

Hu-Sheng Wu. (2016). A Survey of Research on Anomaly Detection for Time Series. *978-1-5090-6126-6/16 2016 IEEE, Page 426*

K. T. Abou-Moustafa, M. Cheriet, and C. Y. Suen. (2004). Classification of time-series data using a generative/discriminative hybrid. *in Proc. 9th Int. Workshop Frontiers Handwriting Recognit., Oct. 2004, pp. 51–56.*

K. T. Abou-Moustafa. (2003). A Generative-Discriminative Framework for Time-Series Data Classification. https://spectrum.library.concordia.ca/2392/

Matplotlib Library. (2019). Data visualization open source library for Python. https://matplotlib.org/

Mian Du, Shichong Ma, and Qing He. (2016). A SCADA Data based Anomaly Detection Method for Wind Turbines. *China International Conference on Electricity Distribution (CICED 2016)*

Microsoft Azure Databricks. (2020). Big data analytics and AI with optimized Apache Spark. https://azure.microsoft.com/en-us/services/databricks/

Numpy Library. (2019). Python open source library for mathematical computing. https://numpy.org/

Oleg Stikhin. (2019). Unsupervised Machine Learning for Anomaly Detection in Wind Turbine Converters. *Aalto University*

Pandas Library. (2019). Python open source library for data analytics. https://pandas.pydata.org/

Ruomu Tan, Tian Cong, Nina F. Thornhill, James R. Ottewill, Jerzy Baranowski. (2019). Statistical Monitoring of Processes with Multiple Operating Modes. *IFAC PaperOnLine 52-1 (2019) 635-642*

SciPy Library. (2019). Python open source library for scientific computing. https://www.scipy.org/

Shai Shalev-Shwartz, Shai Ben-David. (2014). Understanding Machine Learning: From Theory to Algorithms. *Cambridge University Press, ISBN 978-1-107-05713-5*

Shima Ghassempour, Federico Girosi, Anthony Maeder. (2014). Clustering Multivariate Time Series Using Hidden Markov Models. *Int. J. Environ.*

Res. Public Health 2014, 11, 2741-2763; doi:10.3390/ijerph110302741, ISSN 1660-4601

Similaritymeasures Library (2020). Open source python library for calculating similarity between two curves. (2020). https://jekel.me/similarity_measures/index.html

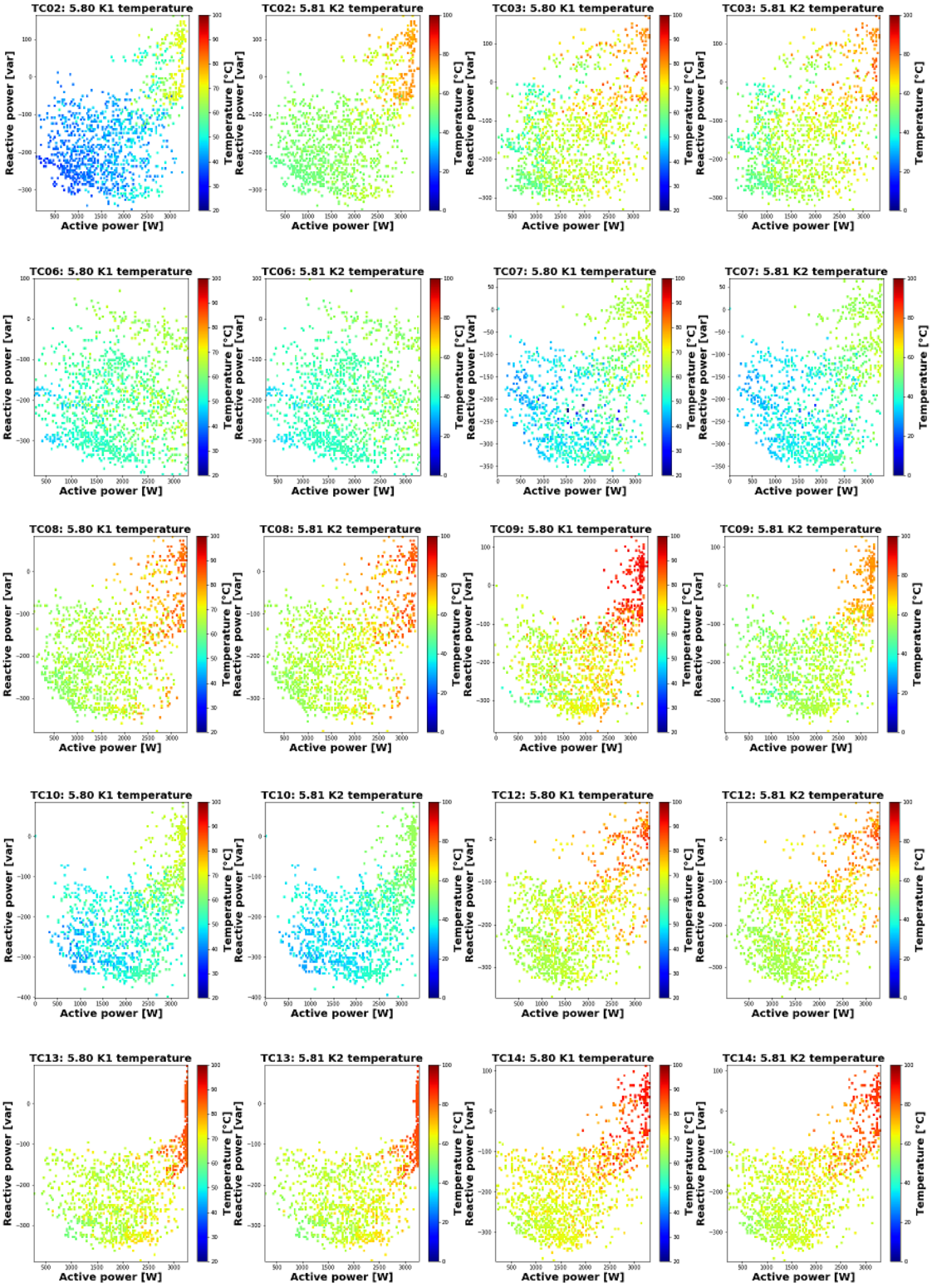
Søren Krohn, Poul-Erik Morthorst, Shimon Awerbuch. (2009). The Economics of Wind Energy. http://www.ewea.org/fileadmin/files/library/publications/reports/Economics_of_Wind_Energy.pdf

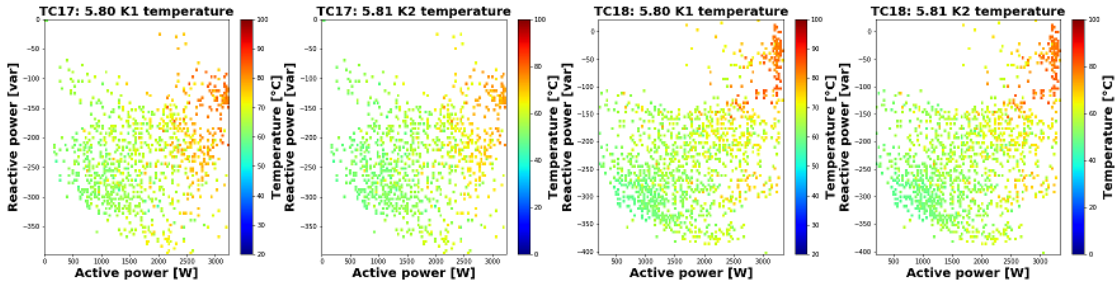
Van Quan Nguyen, Linh Van Ma, Jin-yong Kim, Kwangki Kim, Jinsul Kim. (2018). Applications of Anomaly Detection using Deep Learning on Time Series Data. *2018 IEEE 16th Int. Conf. on Dependable, Autonomic & Secure Comp., 16th Int. Conf. on Pervasive Intelligence & Comp., 4th Int. Conf. on Big Data Intelligence & Comp., and 3rd Cyber Sci. & Tech. Cong*.

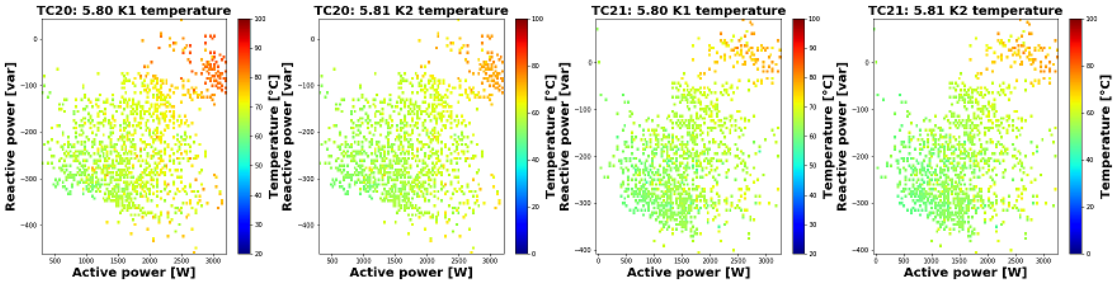
Yulius Denny Prabowo, at el. (2018). Lstm And Simple Rnn Comparison In The Problem Of Sequence To Sequence On Conversation Data Using Bahasa Indonesia. *2018 IEEE Indonesian Association for Pattern Recognition International Conference (INAPR)*.

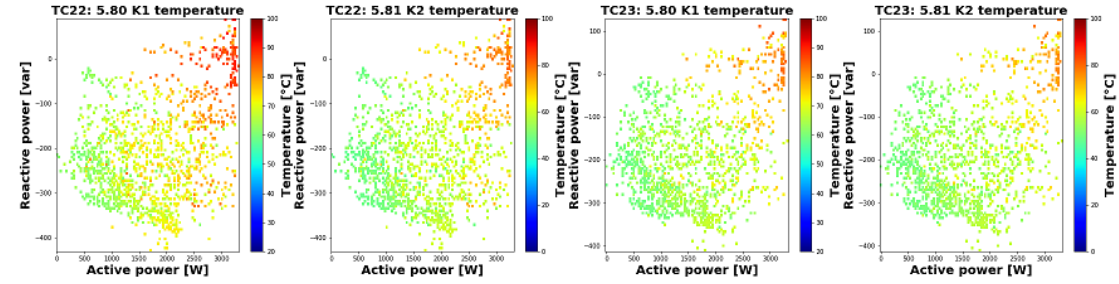
Yu Qin, YuanSheng Lou (2019). Hydrological Time Series Anomaly Pattern Detection based on Isolation Forest. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2019)*.

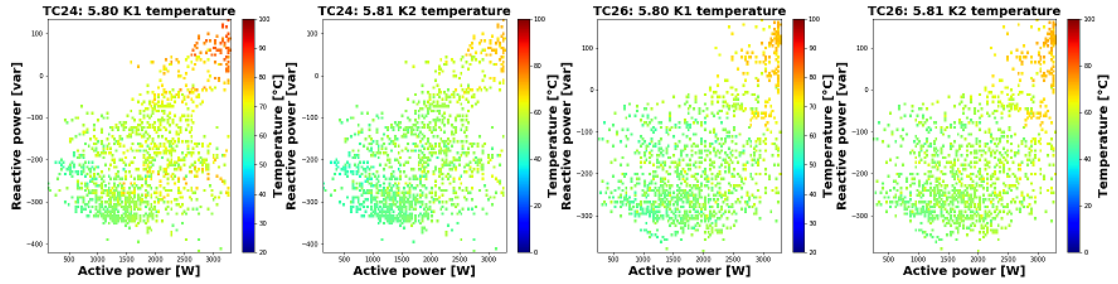
APPENDIX 1. HEAT MAP OF ANALYSIS GROUP 1 WITH ACTIVE POWER P AND REACTIVE POWER Q AS THE AXES

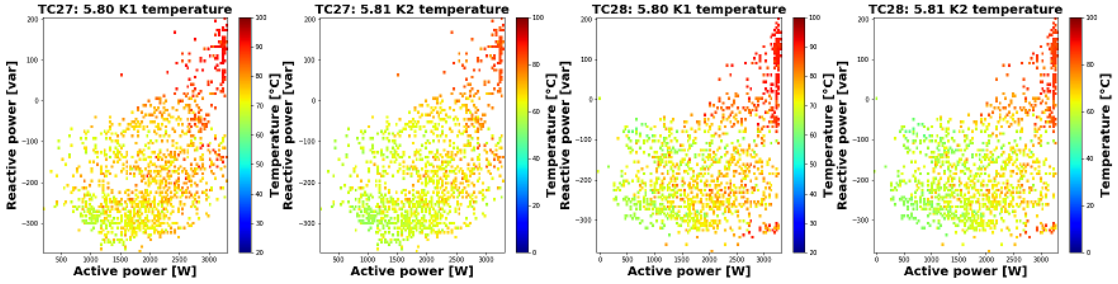


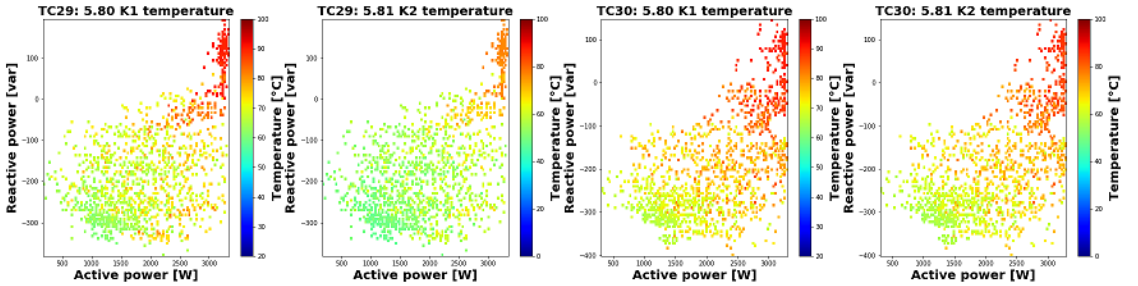

TC17: 5.80 K1 temperature
TC17: 5.81 K2 temperature
TC18: 5.80 K1 temperature
TC18: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

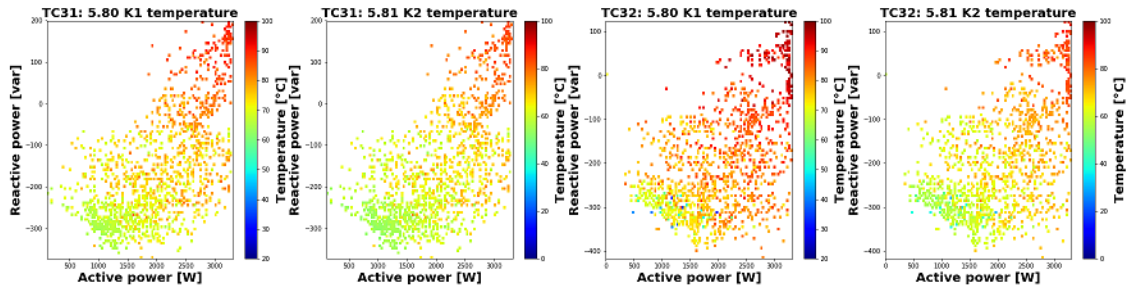

TC20: 5.80 K1 temperature
TC20: 5.81 K2 temperature
TC21: 5.80 K1 temperature
TC21: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

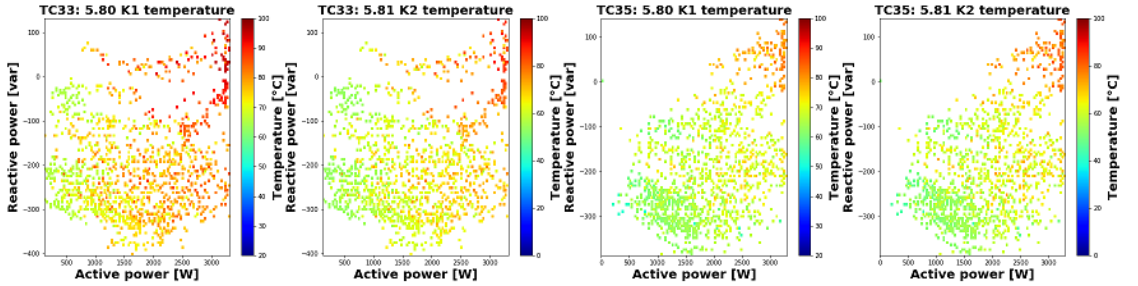

TC22: 5.80 K1 temperature
TC22: 5.81 K2 temperature
TC23: 5.80 K1 temperature
TC23: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

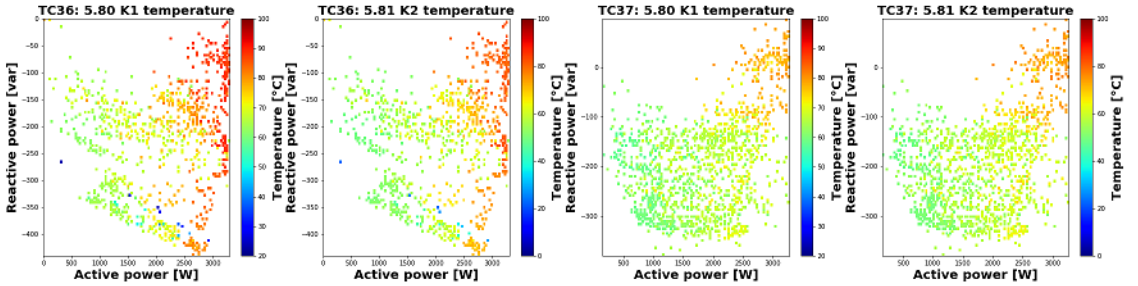

TC24: 5.80 K1 temperature
TC24: 5.81 K2 temperature
TC26: 5.80 K1 temperature
TC26: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

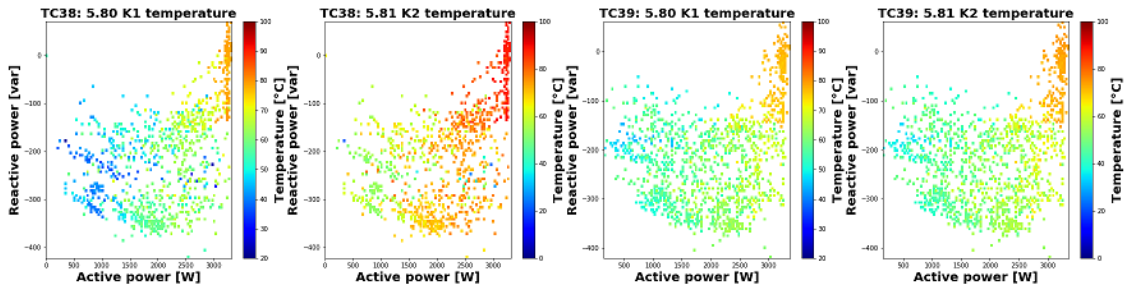

TC27: 5.80 K1 temperature
TC27: 5.81 K2 temperature
TC28: 5.80 K1 temperature
TC28: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

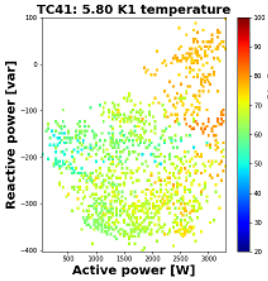

TC29: 5.80 K1 temperature
TC29: 5.81 K2 temperature
TC30: 5.80 K1 temperature
TC30: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

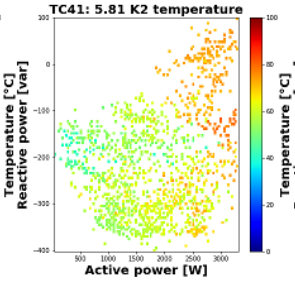

TC31: 5.80 K1 temperature
TC31: 5.81 K2 temperature
TC32: 5.80 K1 temperature
TC32: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

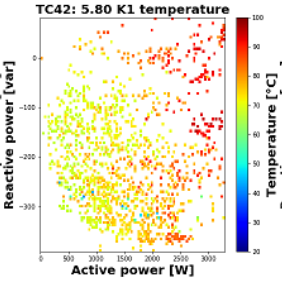

TC33: 5.80 K1 temperature
TC33: 5.81 K2 temperature
TC35: 5.80 K1 temperature
TC35: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

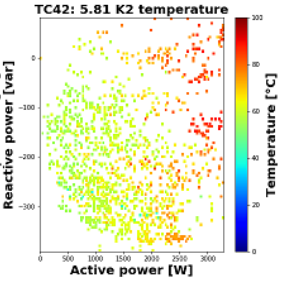

TC36: 5.80 K1 temperature
TC36: 5.81 K2 temperature
TC37: 5.80 K1 temperature
TC37: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

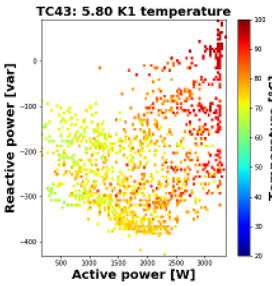

TC38: 5.80 K1 temperature
TC38: 5.81 K2 temperature
TC39: 5.80 K1 temperature
TC39: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

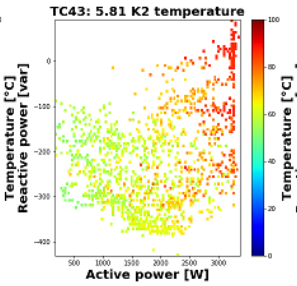

TC41: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

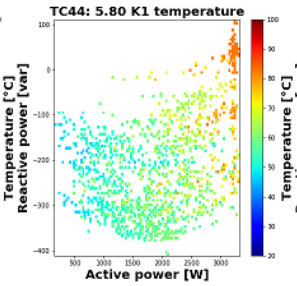

TC41: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

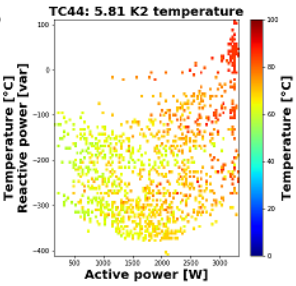

TC42: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

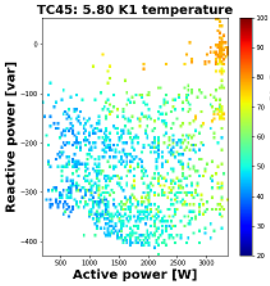

TC42: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

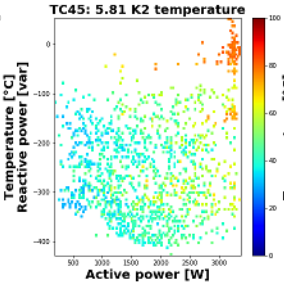

TC43: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

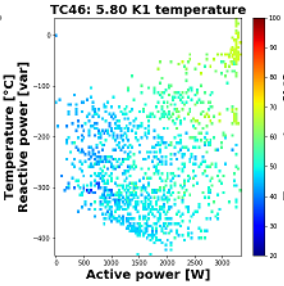

TC43: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

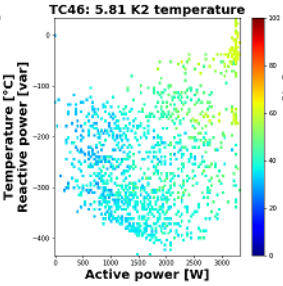

TC44: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

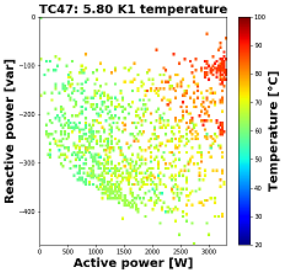

TC44: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

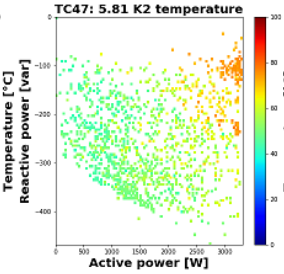

TC45: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]


TC45: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

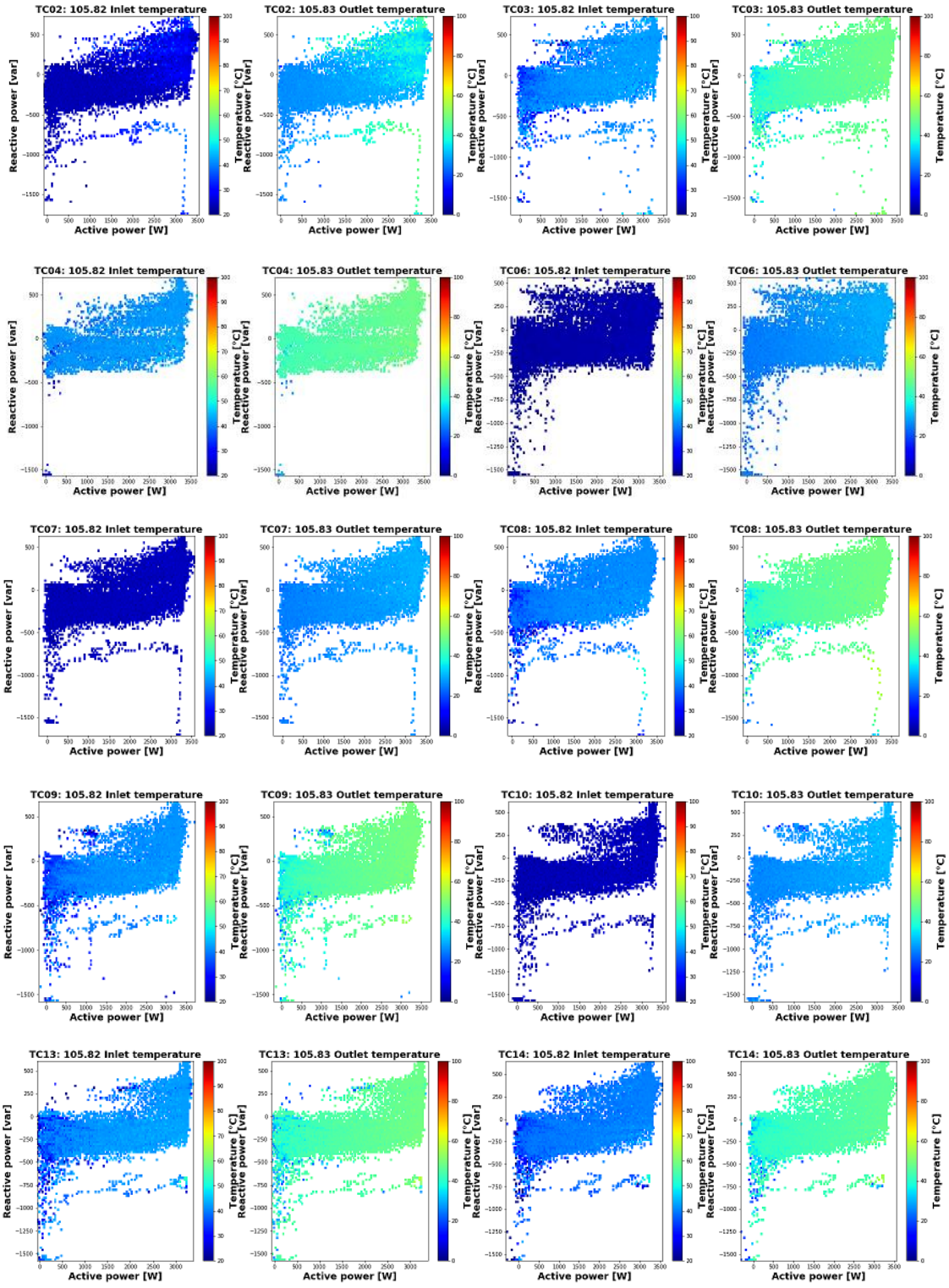

TC46: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

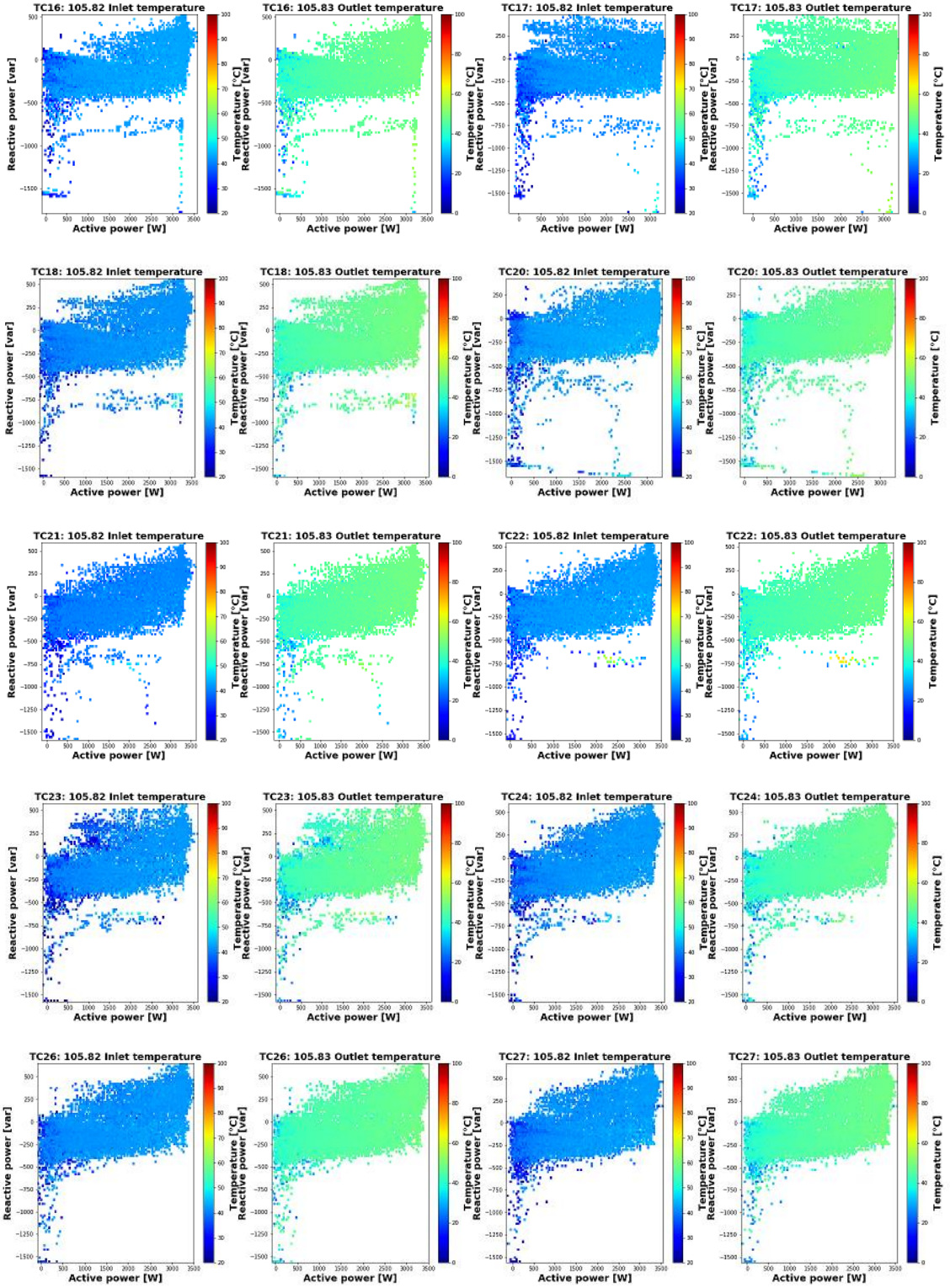

TC46: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

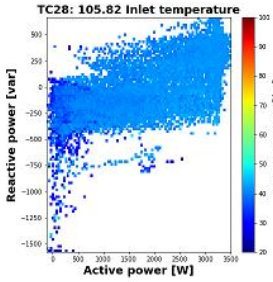

TC47: 5.80 K1 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

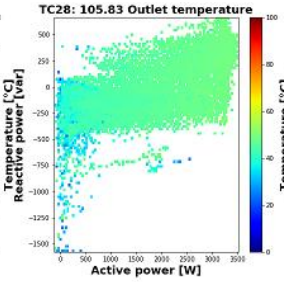

TC47: 5.81 K2 temperature
Reactive power [var]
Active power [W]
Temperature [°C]

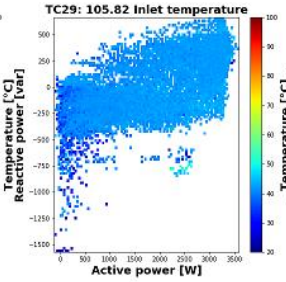
APPENDIX 2. HEAT MAP OF ANALYSIS GROUP 2, WITH ACTIVE POWER P AND REACTIVE POWER Q AS THE AXES

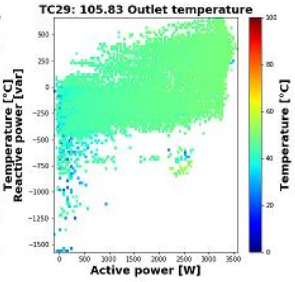


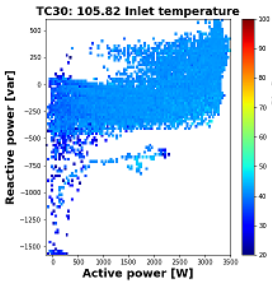

TC16: 105.82 Inlet temperature
TC16: 105.83 Outlet temperature
TC17: 105.82 Inlet temperature
TC17: 105.83 Outlet temperature
TC18: 105.82 Inlet temperature
TC18: 105.83 Outlet temperature
TC20: 105.82 Inlet temperature
TC20: 105.83 Outlet temperature
TC21: 105.82 Inlet temperature
TC21: 105.83 Outlet temperature
TC22: 105.82 Inlet temperature
TC22: 105.83 Outlet temperature
TC23: 105.82 Inlet temperature
TC23: 105.83 Outlet temperature
TC24: 105.82 Inlet temperature
TC24: 105.83 Outlet temperature
TC26: 105.82 Inlet temperature
TC26: 105.83 Outlet temperature
TC27: 105.82 Inlet temperature
TC27: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

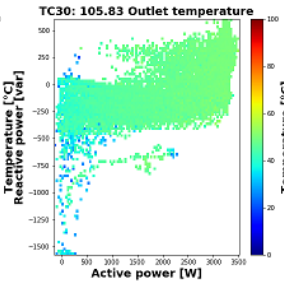

TC28: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

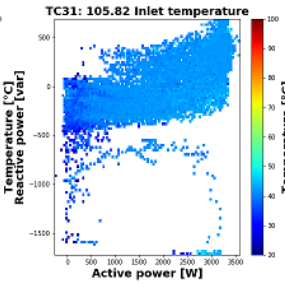

TC28: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

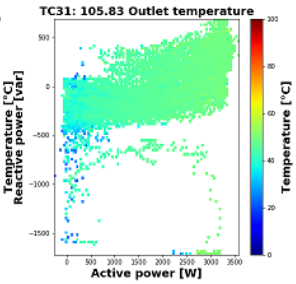

TC29: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

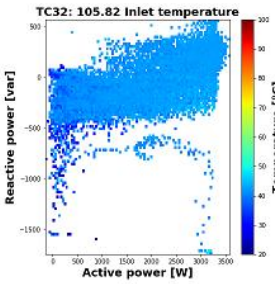

TC29: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

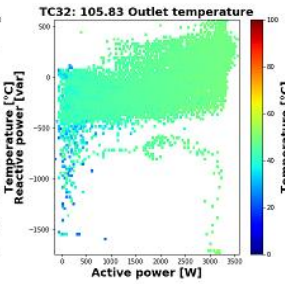

TC30: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

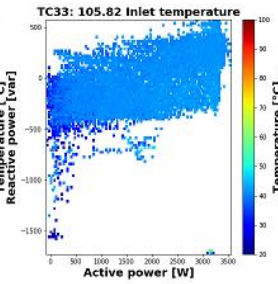

TC30: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

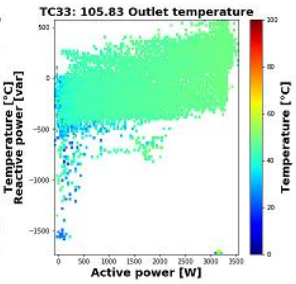

TC31: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

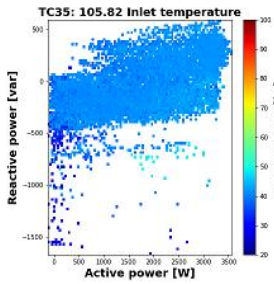

TC31: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

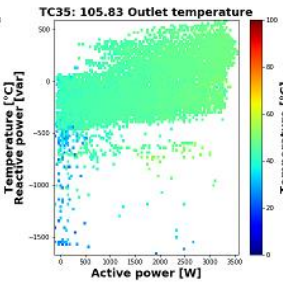

TC32: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

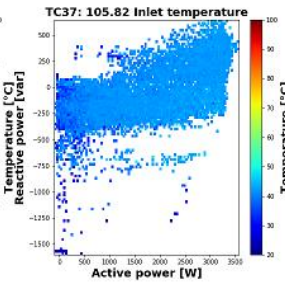

TC32: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

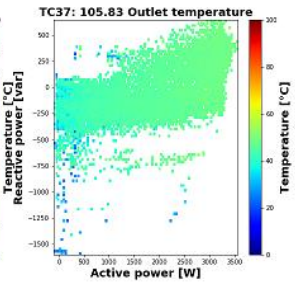

TC33: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

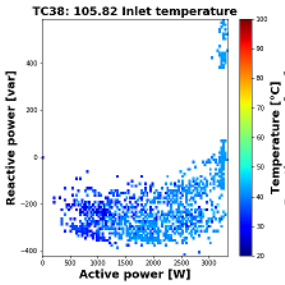

TC33: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

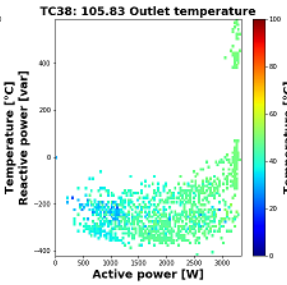

TC35: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

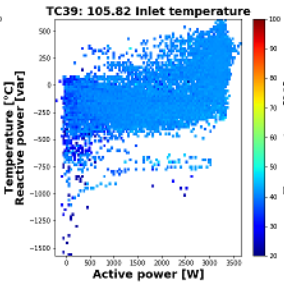

TC35: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

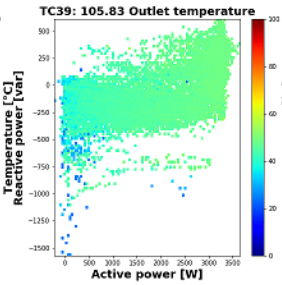

TC37: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

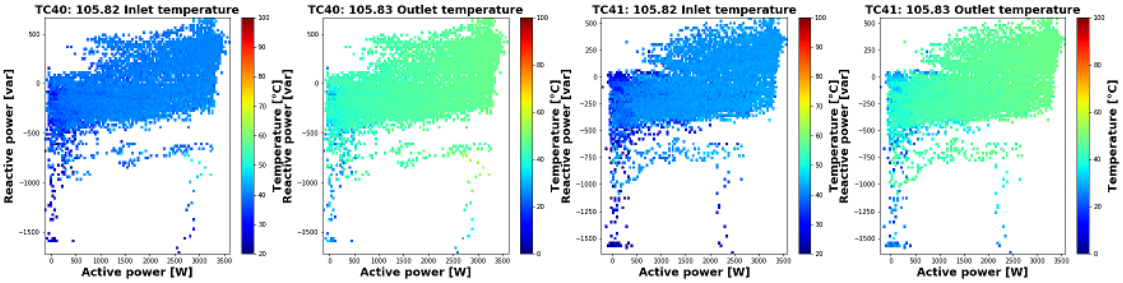

TC37: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

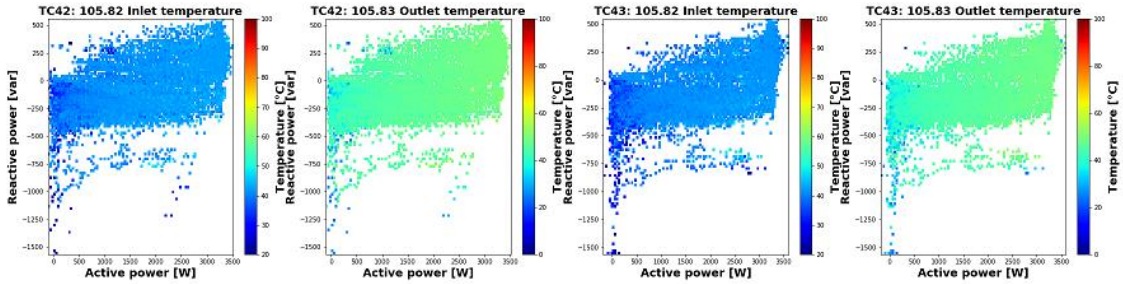

TC38: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

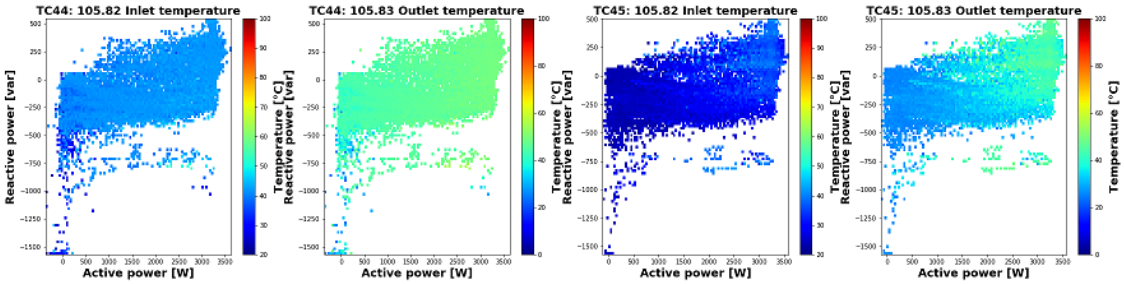

TC38: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

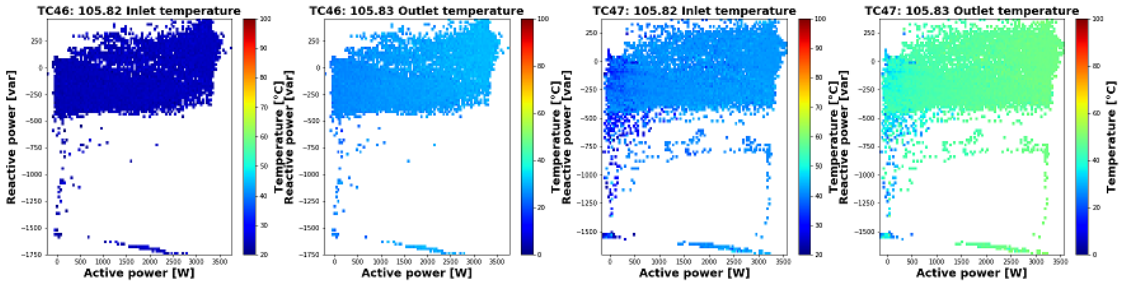

TC39: 105.82 Inlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]


TC39: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

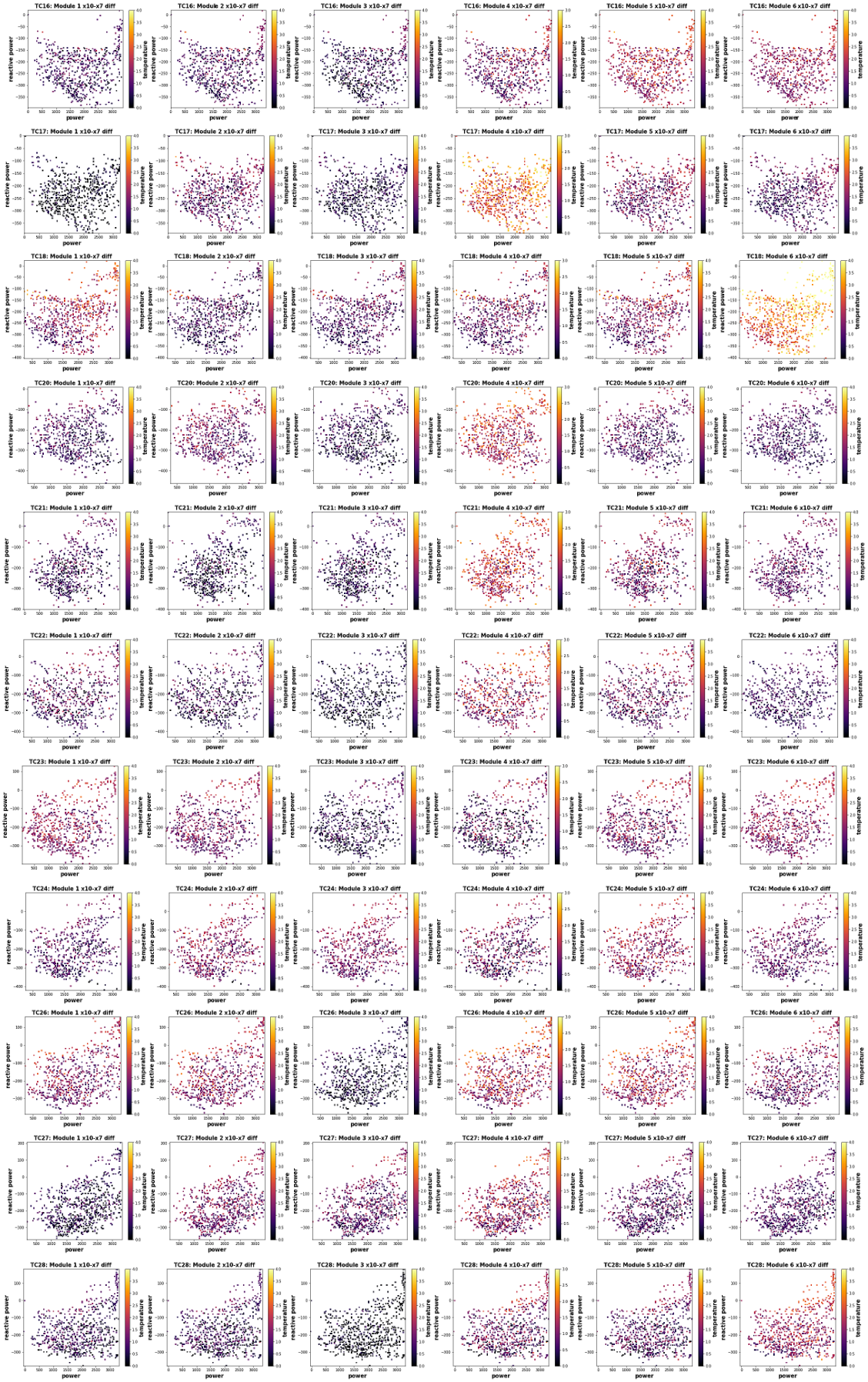

TC40: 105.82 Inlet temperature
TC40: 105.83 Outlet temperature
TC41: 105.82 Inlet temperature
TC41: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

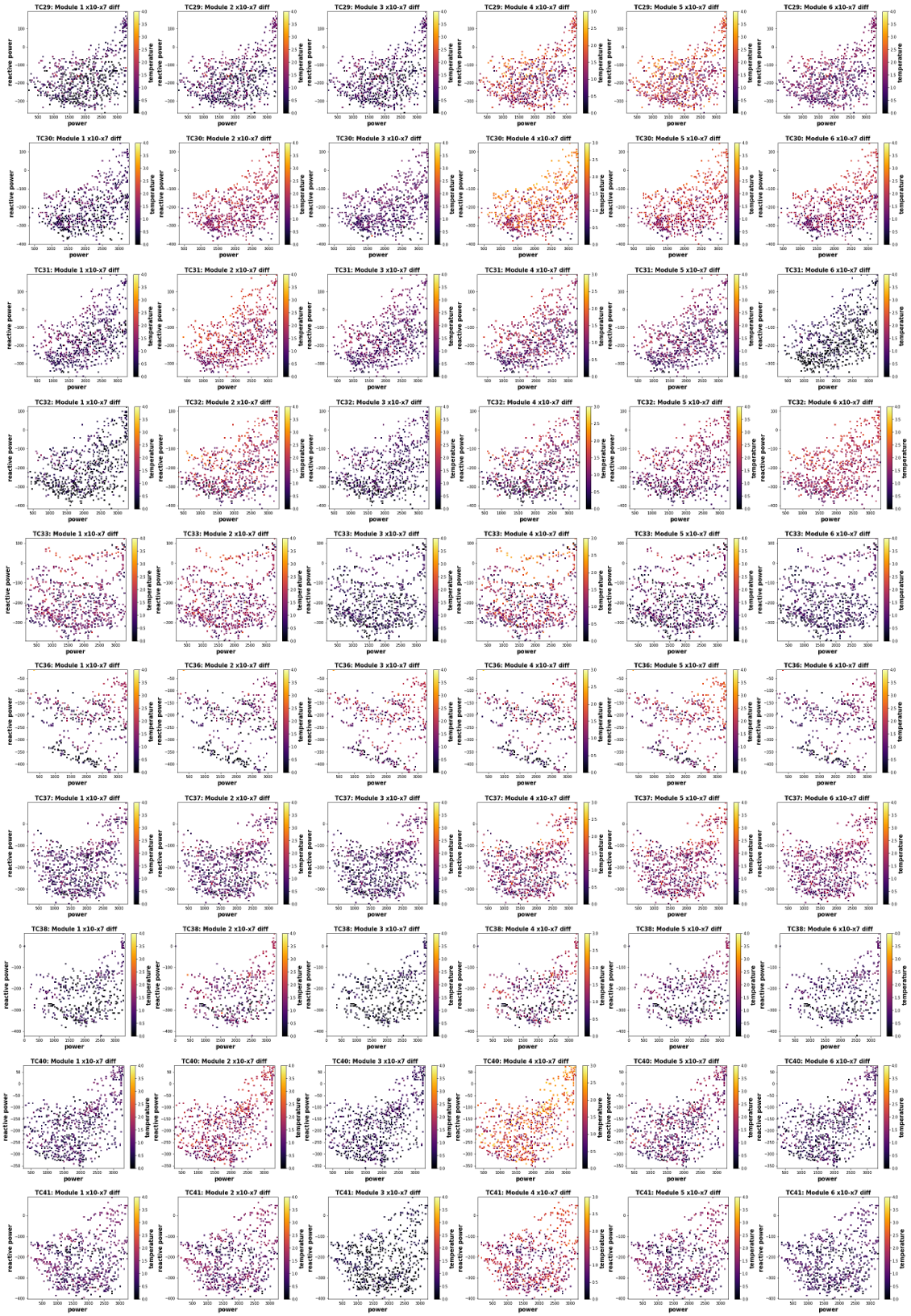

TC42: 105.82 Inlet temperature
TC42: 105.83 Outlet temperature
TC43: 105.82 Inlet temperature
TC43: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

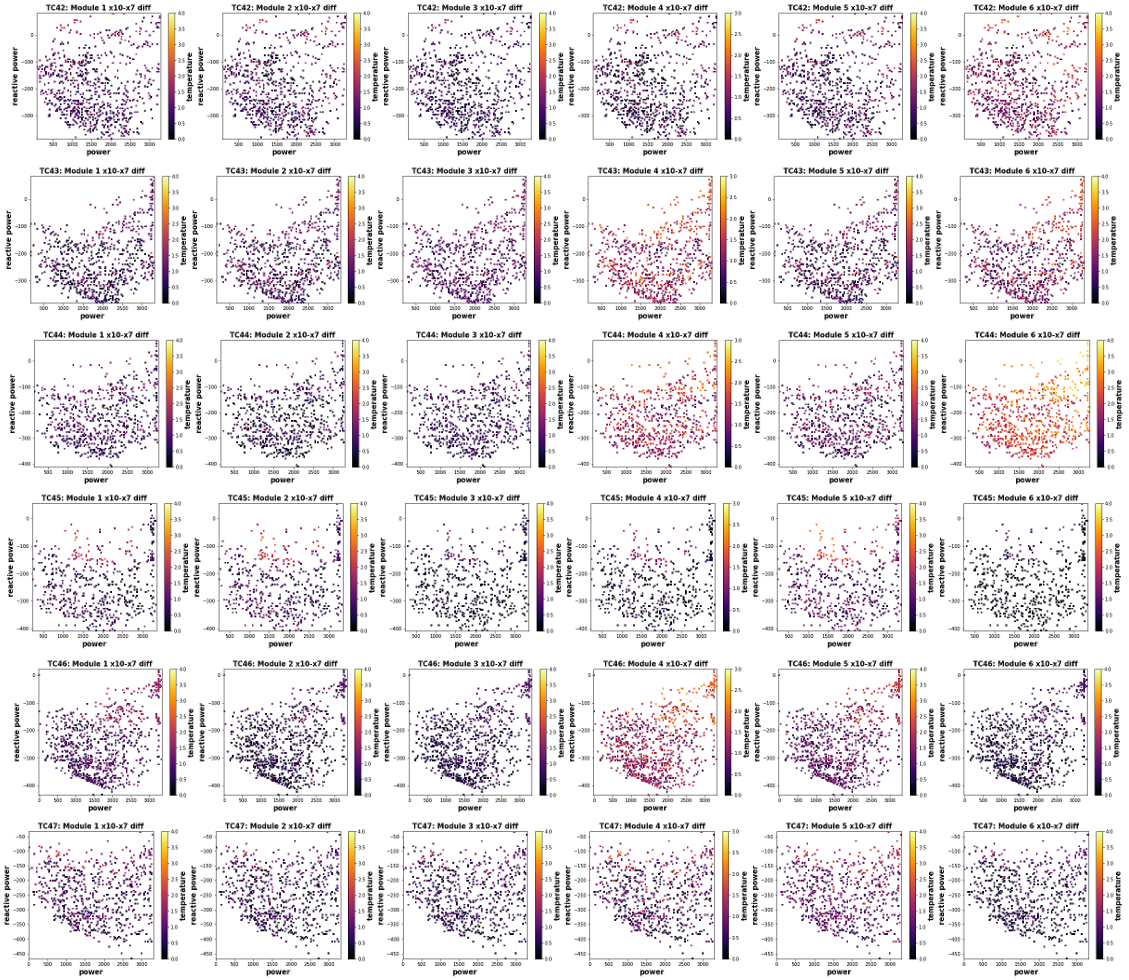

TC44: 105.82 Inlet temperature
TC44: 105.83 Outlet temperature
TC45: 105.82 Inlet temperature
TC45: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]


TC46: 105.82 Inlet temperature
TC46: 105.83 Outlet temperature
TC47: 105.82 Inlet temperature
TC47: 105.83 Outlet temperature
Reactive power [var]
Active power [W]
Temperature [°C]

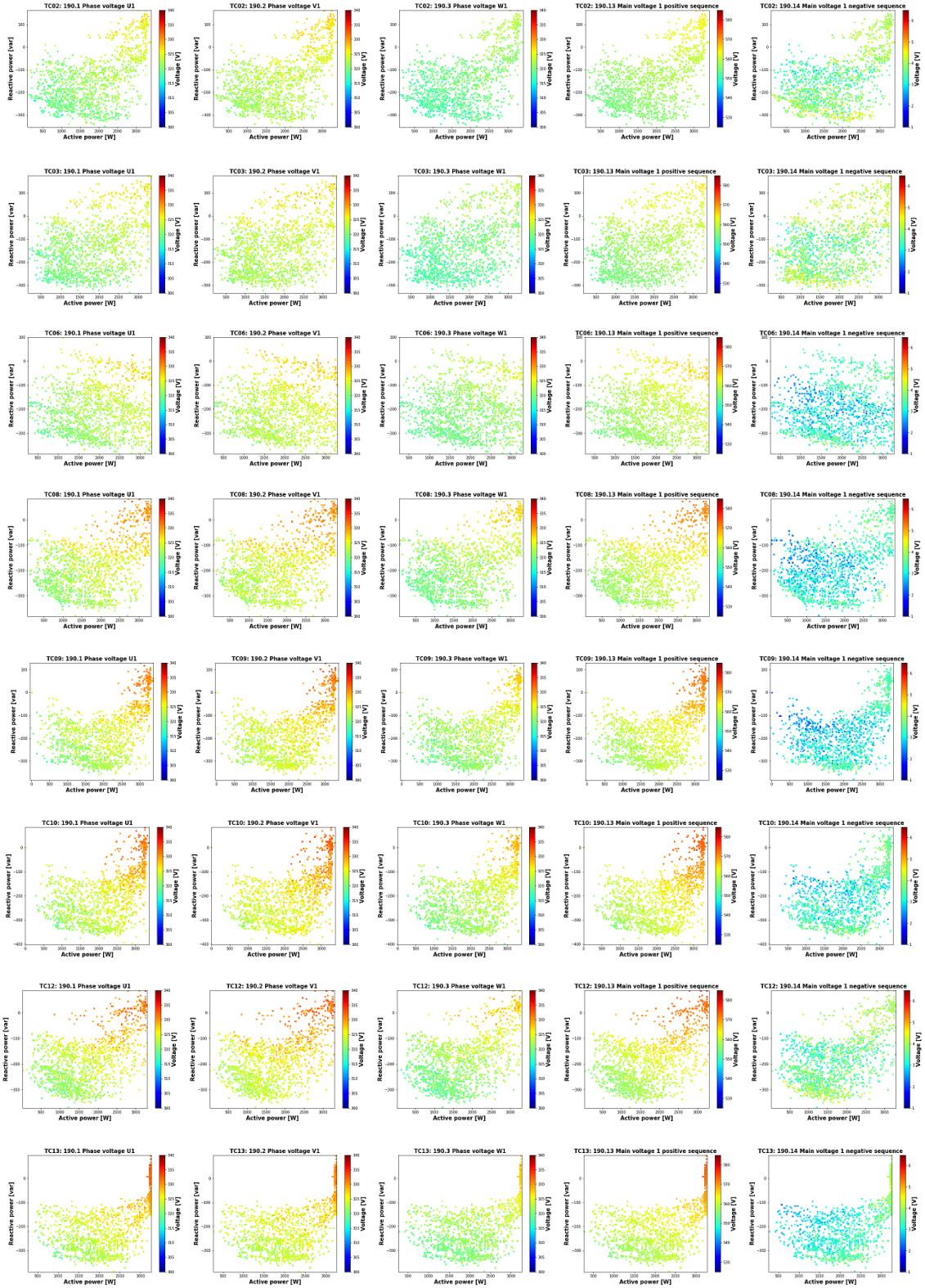
APPENDIX 3. HEAT MAP OF ANALYSIS GROUP 3, WITH ACTIVE POWER P AND REACTIVE POWER Q AS THE AXES

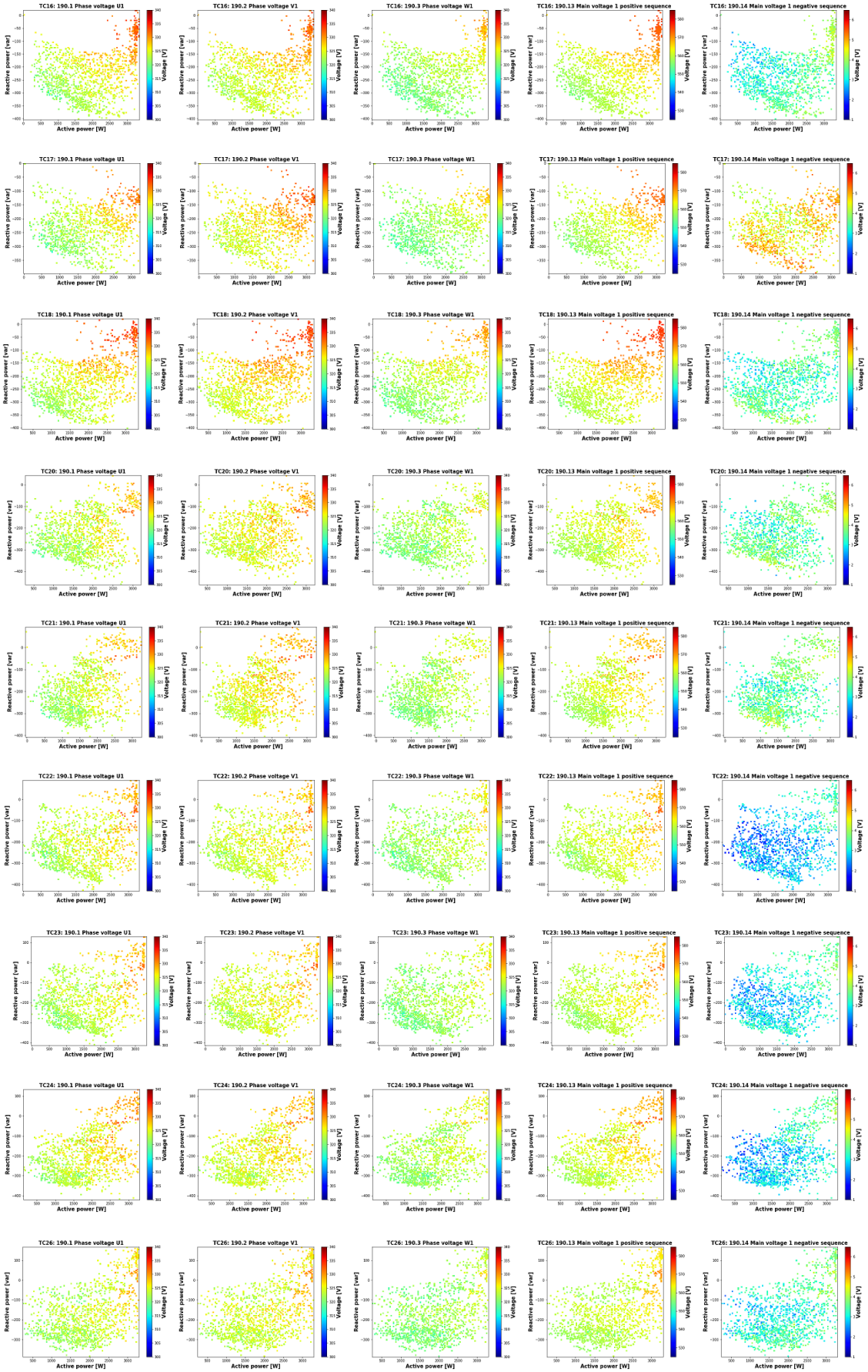

TC16: Module 1 x10-x7 diff
TC16: Module 2 x10-x7 diff
TC16: Module 3 x10-x7 diff
TC16: Module 4 x10-x7 diff
TC16: Module 5 x10-x7 diff
TC16: Module 6 x10-x7 diff
TC17: Module 1 x10-x7 diff
TC17: Module 2 x10-x7 diff
TC17: Module 3 x10-x7 diff
TC17: Module 4 x10-x7 diff
TC17: Module 5 x10-x7 diff
TC17: Module 6 x10-x7 diff
TC18: Module 1 x10-x7 diff
TC18: Module 2 x10-x7 diff
TC18: Module 3 x10-x7 diff
TC18: Module 4 x10-x7 diff
TC18: Module 5 x10-x7 diff
TC18: Module 6 x10-x7 diff
TC20: Module 1 x10-x7 diff
TC20: Module 2 x10-x7 diff
TC20: Module 3 x10-x7 diff
TC20: Module 4 x10-x7 diff
TC20: Module 5 x10-x7 diff
TC20: Module 6 x10-x7 diff
TC21: Module 1 x10-x7 diff
TC21: Module 2 x10-x7 diff
TC21: Module 3 x10-x7 diff
TC21: Module 4 x10-x7 diff
TC21: Module 5 x10-x7 diff
TC21: Module 6 x10-x7 diff
TC22: Module 1 x10-x7 diff
TC22: Module 2 x10-x7 diff
TC22: Module 3 x10-x7 diff
TC22: Module 4 x10-x7 diff
TC22: Module 5 x10-x7 diff
TC22: Module 6 x10-x7 diff
TC23: Module 1 x10-x7 diff
TC23: Module 2 x10-x7 diff
TC23: Module 3 x10-x7 diff
TC23: Module 4 x10-x7 diff
TC23: Module 5 x10-x7 diff
TC23: Module 6 x10-x7 diff
TC24: Module 1 x10-x7 diff
TC24: Module 2 x10-x7 diff
TC24: Module 3 x10-x7 diff
TC24: Module 4 x10-x7 diff
TC24: Module 5 x10-x7 diff
TC24: Module 6 x10-x7 diff
TC26: Module 1 x10-x7 diff
TC26: Module 2 x10-x7 diff
TC26: Module 3 x10-x7 diff
TC26: Module 4 x10-x7 diff
TC26: Module 5 x10-x7 diff
TC26: Module 6 x10-x7 diff
TC27: Module 1 x10-x7 diff
TC27: Module 2 x10-x7 diff
TC27: Module 3 x10-x7 diff
TC27: Module 4 x10-x7 diff
TC27: Module 5 x10-x7 diff
TC27: Module 6 x10-x7 diff
TC28: Module 1 x10-x7 diff
TC28: Module 2 x10-x7 diff
TC28: Module 3 x10-x7 diff
TC28: Module 4 x10-x7 diff
TC28: Module 5 x10-x7 diff
TC28: Module 6 x10-x7 diff
power
reactive power
temperature

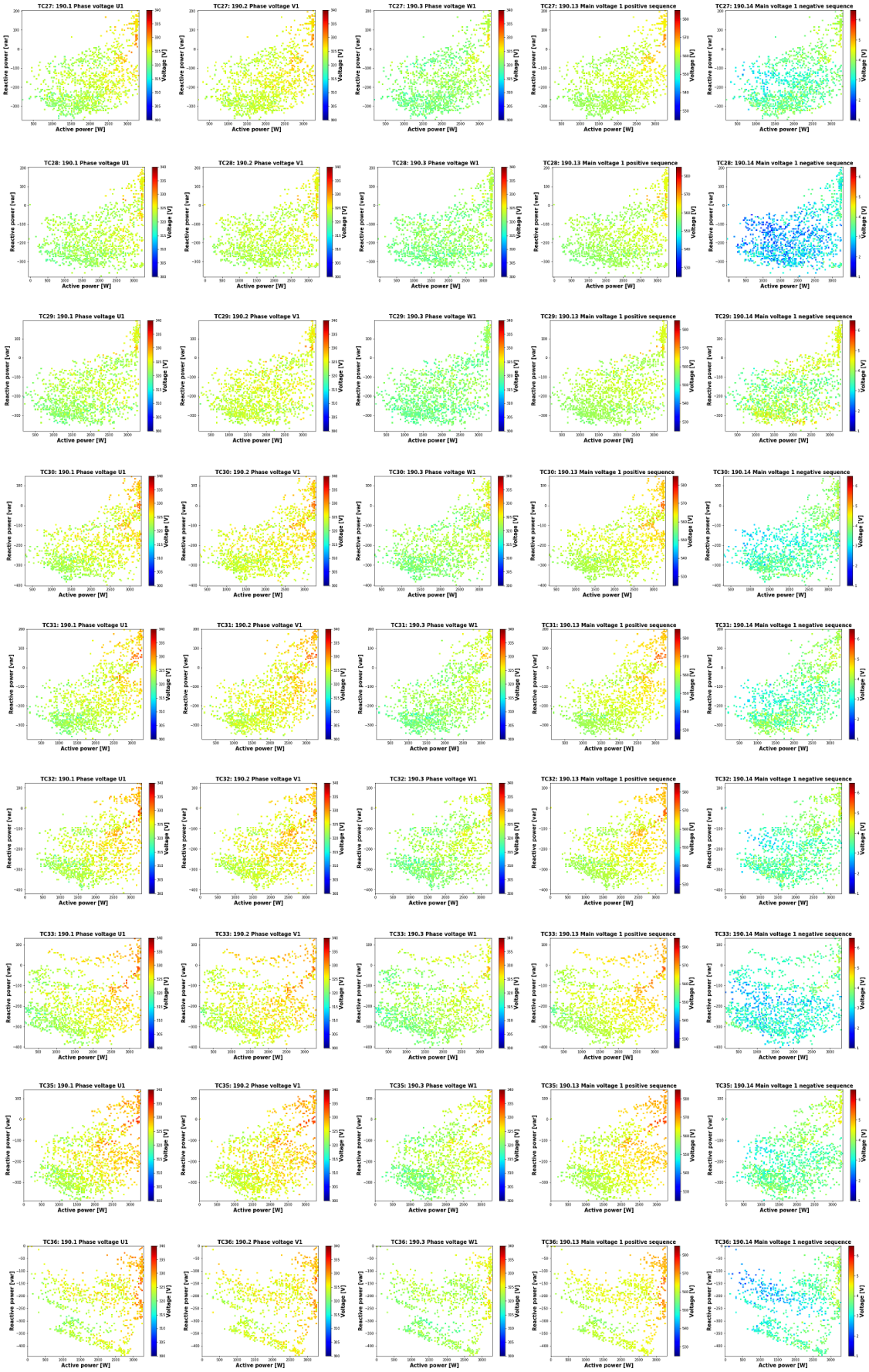

TC29: Module 1 x10-x7 diff
TC29: Module 2 x10-x7 diff
TC29: Module 3 x10-x7 diff
TC29: Module 4 x10-x7 diff
TC29: Module 5 x10-x7 diff
TC29: Module 6 x10-x7 diff
TC30: Module 1 x10-x7 diff
TC30: Module 2 x10-x7 diff
TC30: Module 3 x10-x7 diff
TC30: Module 4 x10-x7 diff
TC30: Module 5 x10-x7 diff
TC30: Module 6 x10-x7 diff
TC31: Module 1 x10-x7 diff
TC31: Module 2 x10-x7 diff
TC31: Module 3 x10-x7 diff
TC31: Module 4 x10-x7 diff
TC31: Module 5 x10-x7 diff
TC31: Module 6 x10-x7 diff
TC32: Module 1 x10-x7 diff
TC32: Module 2 x10-x7 diff
TC32: Module 3 x10-x7 diff
TC32: Module 4 x10-x7 diff
TC32: Module 5 x10-x7 diff
TC32: Module 6 x10-x7 diff
TC33: Module 1 x10-x7 diff
TC33: Module 2 x10-x7 diff
TC33: Module 3 x10-x7 diff
TC33: Module 4 x10-x7 diff
TC33: Module 5 x10-x7 diff
TC33: Module 6 x10-x7 diff
TC36: Module 1 x10-x7 diff
TC36: Module 2 x10-x7 diff
TC36: Module 3 x10-x7 diff
TC36: Module 4 x10-x7 diff
TC36: Module 5 x10-x7 diff
TC36: Module 6 x10-x7 diff
TC37: Module 1 x10-x7 diff
TC37: Module 2 x10-x7 diff
TC37: Module 3 x10-x7 diff
TC37: Module 4 x10-x7 diff
TC37: Module 5 x10-x7 diff
TC37: Module 6 x10-x7 diff
TC38: Module 1 x10-x7 diff
TC38: Module 2 x10-x7 diff
TC38: Module 3 x10-x7 diff
TC38: Module 4 x10-x7 diff
TC38: Module 5 x10-x7 diff
TC38: Module 6 x10-x7 diff
TC40: Module 1 x10-x7 diff
TC40: Module 2 x10-x7 diff
TC40: Module 3 x10-x7 diff
TC40: Module 4 x10-x7 diff
TC40: Module 5 x10-x7 diff
TC40: Module 6 x10-x7 diff
TC41: Module 1 x10-x7 diff
TC41: Module 2 x10-x7 diff
TC41: Module 3 x10-x7 diff
TC41: Module 4 x10-x7 diff
TC41: Module 5 x10-x7 diff
TC41: Module 6 x10-x7 diff
reactive power
power
temperature

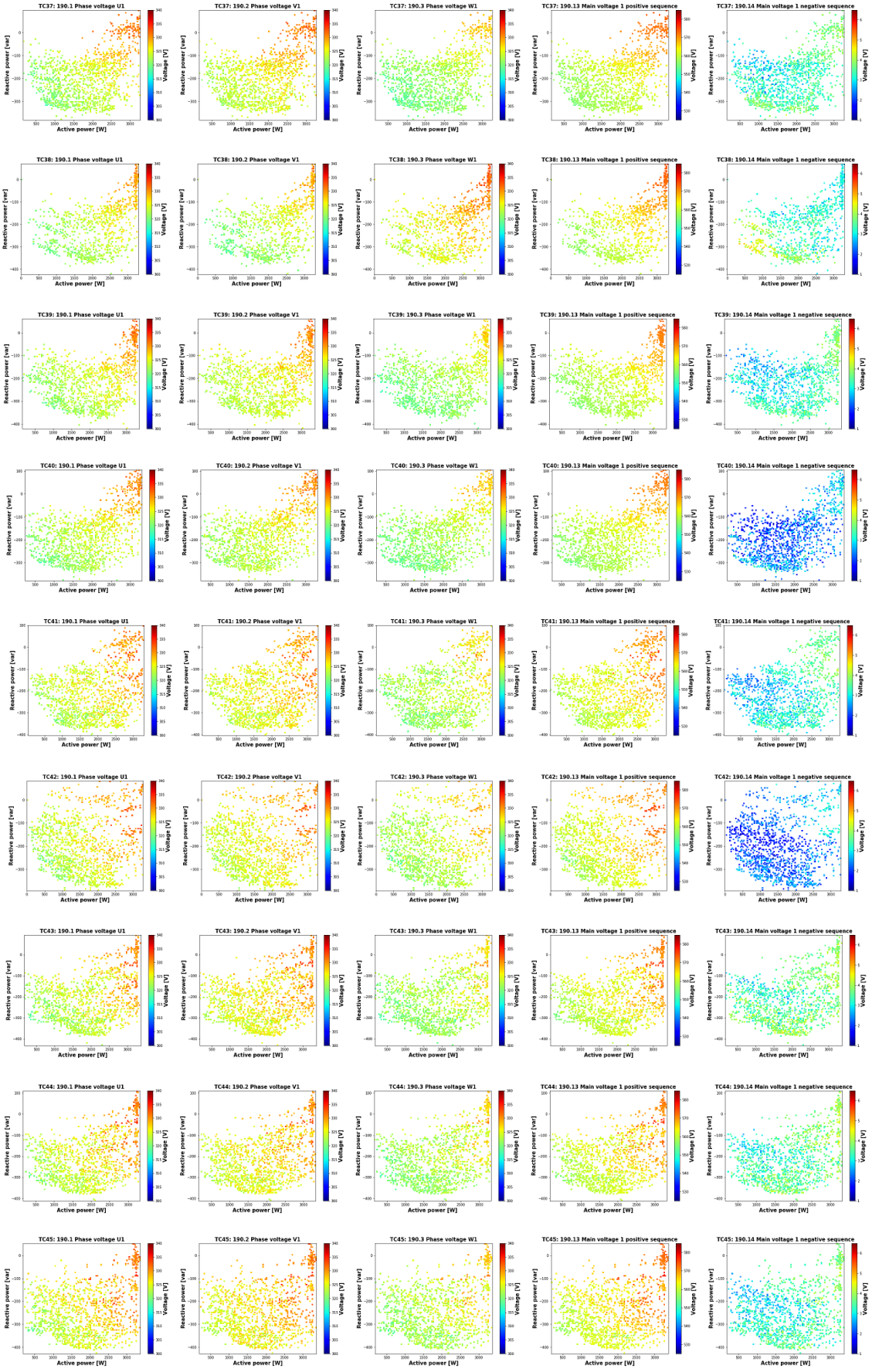

TC42: Module 1 x10-x7 diff
TC42: Module 2 x10-x7 diff
TC42: Module 3 x10-x7 diff
TC42: Module 4 x10-x7 diff
TC42: Module 5 x10-x7 diff
TC42: Module 6 x10-x7 diff
TC43: Module 1 x10-x7 diff
TC43: Module 2 x10-x7 diff
TC43: Module 3 x10-x7 diff
TC43: Module 4 x10-x7 diff
TC43: Module 5 x10-x7 diff
TC43: Module 6 x10-x7 diff
TC44: Module 1 x10-x7 diff
TC44: Module 2 x10-x7 diff
TC44: Module 3 x10-x7 diff
TC44: Module 4 x10-x7 diff
TC44: Module 5 x10-x7 diff
TC44: Module 6 x10-x7 diff
TC45: Module 1 x10-x7 diff
TC45: Module 2 x10-x7 diff
TC45: Module 3 x10-x7 diff
TC45: Module 4 x10-x7 diff
TC45: Module 5 x10-x7 diff
TC45: Module 6 x10-x7 diff
TC46: Module 1 x10-x7 diff
TC46: Module 2 x10-x7 diff
TC46: Module 3 x10-x7 diff
TC46: Module 4 x10-x7 diff
TC46: Module 5 x10-x7 diff
TC46: Module 6 x10-x7 diff
TC47: Module 1 x10-x7 diff
TC47: Module 2 x10-x7 diff
TC47: Module 3 x10-x7 diff
TC47: Module 4 x10-x7 diff
TC47: Module 5 x10-x7 diff
TC47: Module 6 x10-x7 diff
power
reactive power
temperature

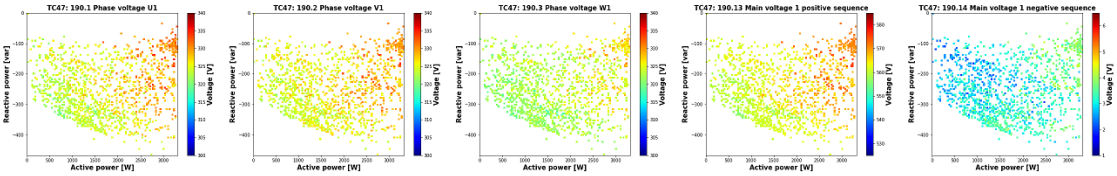
APPENDIX 4. HEAT MAP OF ANALYSIS GROUP 4, WITH ACTIVE POWER P AND REACTIVE POWER Q AS THE AXES




TC16: 190.1 Phase voltage U1
TC16: 190.2 Phase voltage V1
TC16: 190.3 Phase voltage W1
TC16: 190.13 Main voltage 1 positive sequence
TC16: 190.14 Main voltage 1 negative sequence
TC17: 190.1 Phase voltage U1
TC17: 190.2 Phase voltage V1
TC17: 190.3 Phase voltage W1
TC17: 190.13 Main voltage 1 positive sequence
TC17: 190.14 Main voltage 1 negative sequence
TC18: 190.1 Phase voltage U1
TC18: 190.2 Phase voltage V1
TC18: 190.3 Phase voltage W1
TC18: 190.13 Main voltage 1 positive sequence
TC18: 190.14 Main voltage 1 negative sequence
TC20: 190.1 Phase voltage U1
TC20: 190.2 Phase voltage V1
TC20: 190.3 Phase voltage W1
TC20: 190.13 Main voltage 1 positive sequence
TC20: 190.14 Main voltage 1 negative sequence
TC21: 190.1 Phase voltage U1
TC21: 190.2 Phase voltage V1
TC21: 190.3 Phase voltage W1
TC21: 190.13 Main voltage 1 positive sequence
TC21: 190.14 Main voltage 1 negative sequence
TC22: 190.1 Phase voltage U1
TC22: 190.2 Phase voltage V1
TC22: 190.3 Phase voltage W1
TC22: 190.13 Main voltage 1 positive sequence
TC22: 190.14 Main voltage 1 negative sequence
TC23: 190.1 Phase voltage U1
TC23: 190.2 Phase voltage V1
TC23: 190.3 Phase voltage W1
TC23: 190.13 Main voltage 1 positive sequence
TC23: 190.14 Main voltage 1 negative sequence
TC24: 190.1 Phase voltage U1
TC24: 190.2 Phase voltage V1
TC24: 190.3 Phase voltage W1
TC24: 190.13 Main voltage 1 positive sequence
TC24: 190.14 Main voltage 1 negative sequence
TC26: 190.1 Phase voltage U1
TC26: 190.2 Phase voltage V1
TC26: 190.3 Phase voltage W1
TC26: 190.13 Main voltage 1 positive sequence
TC26: 190.14 Main voltage 1 negative sequence
Active power [W]
Reactive power [var]
Voltage [V]


TC27: 190.1 Phase voltage U1
TC27: 190.2 Phase voltage V1
TC27: 190.3 Phase voltage W1
TC27: 190.13 Main voltage 1 positive sequence
TC27: 190.14 Main voltage 1 negative sequence
TC28: 190.1 Phase voltage U1
TC28: 190.2 Phase voltage V1
TC28: 190.3 Phase voltage W1
TC28: 190.13 Main voltage 1 positive sequence
TC28: 190.14 Main voltage 1 negative sequence
TC29: 190.1 Phase voltage U1
TC29: 190.2 Phase voltage V1
TC29: 190.3 Phase voltage W1
TC29: 190.13 Main voltage 1 positive sequence
TC29: 190.14 Main voltage 1 negative sequence
TC30: 190.1 Phase voltage U1
TC30: 190.2 Phase voltage V1
TC30: 190.3 Phase voltage W1
TC30: 190.13 Main voltage 1 positive sequence
TC30: 190.14 Main voltage 1 negative sequence
TC31: 190.1 Phase voltage U1
TC31: 190.2 Phase voltage V1
TC31: 190.3 Phase voltage W1
TC31: 190.13 Main voltage 1 positive sequence
TC31: 190.14 Main voltage 1 negative sequence
TC32: 190.1 Phase voltage U1
TC32: 190.2 Phase voltage V1
TC32: 190.3 Phase voltage W1
TC32: 190.13 Main voltage 1 positive sequence
TC32: 190.14 Main voltage 1 negative sequence
TC33: 190.1 Phase voltage U1
TC33: 190.2 Phase voltage V1
TC33: 190.3 Phase voltage W1
TC33: 190.13 Main voltage 1 positive sequence
TC33: 190.14 Main voltage 1 negative sequence
TC35: 190.1 Phase voltage U1
TC35: 190.2 Phase voltage V1
TC35: 190.3 Phase voltage W1
TC35: 190.13 Main voltage 1 positive sequence
TC35: 190.14 Main voltage 1 negative sequence
TC36: 190.1 Phase voltage U1
TC36: 190.2 Phase voltage V1
TC36: 190.3 Phase voltage W1
TC36: 190.13 Main voltage 1 positive sequence
TC36: 190.14 Main voltage 1 negative sequence
Active power [W]
Reactive power [var]
Voltage [V]


TC37: 190.1 Phase voltage U1
TC37: 190.2 Phase voltage V1
TC37: 190.3 Phase voltage W1
TC37: 190.13 Main voltage 1 positive sequence
TC37: 190.14 Main voltage 1 negative sequence
TC38: 190.1 Phase voltage U1
TC38: 190.2 Phase voltage V1
TC38: 190.3 Phase voltage W1
TC38: 190.13 Main voltage 1 positive sequence
TC38: 190.14 Main voltage 1 negative sequence
TC39: 190.1 Phase voltage U1
TC39: 190.2 Phase voltage V1
TC39: 190.3 Phase voltage W1
TC39: 190.13 Main voltage 1 positive sequence
TC39: 190.14 Main voltage 1 negative sequence
TC40: 190.1 Phase voltage U1
TC40: 190.2 Phase voltage V1
TC40: 190.3 Phase voltage W1
TC40: 190.13 Main voltage 1 positive sequence
TC40: 190.14 Main voltage 1 negative sequence
TC41: 190.1 Phase voltage U1
TC41: 190.2 Phase voltage V1
TC41: 190.3 Phase voltage W1
TC41: 190.13 Main voltage 1 positive sequence
TC41: 190.14 Main voltage 1 negative sequence
TC42: 190.1 Phase voltage U1
TC42: 190.2 Phase voltage V1
TC42: 190.3 Phase voltage W1
TC42: 190.13 Main voltage 1 positive sequence
TC42: 190.14 Main voltage 1 negative sequence
TC43: 190.1 Phase voltage U1
TC43: 190.2 Phase voltage V1
TC43: 190.3 Phase voltage W1
TC43: 190.13 Main voltage 1 positive sequence
TC43: 190.14 Main voltage 1 negative sequence
TC44: 190.1 Phase voltage U1
TC44: 190.2 Phase voltage V1
TC44: 190.3 Phase voltage W1
TC44: 190.13 Main voltage 1 positive sequence
TC44: 190.14 Main voltage 1 negative sequence
TC45: 190.1 Phase voltage U1
TC45: 190.2 Phase voltage V1
TC45: 190.3 Phase voltage W1
TC45: 190.13 Main voltage 1 positive sequence
TC45: 190.14 Main voltage 1 negative sequence
Active power [W]
Reactive power [var]
Voltage [V]


TC47: 190.1 Phase voltage U1
TC47: 190.2 Phase voltage V1
TC47: 190.3 Phase voltage W1
TC47: 190.13 Main voltage 1 positive sequence
TC47: 190.14 Main voltage 1 negative sequence
Reactive power [var]
Active power [W]
Voltage [V]