PREDICTION OF OPERATIONAL IMPROVEMENTS IN WIND POWER PLANTS

by
ELİF SARAÇOĞLU

Submitted to the Sabancı Graduate Business School
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
September 2020

PREDICTION OF OPERATIONAL IMPROVEMENTS IN WIND POWER PLANTS

Approved by:

Prof. Abdullah Daşcı .....
(Thesis Supervisor)

Assoc. Prof. Raha Akhavan-Tabatabaei .....

Assist. Prof. Levent Erişkin .....

Date of Approval: September 1, 2020

ABSTRACT

PREDICTION OF OPERATIONAL IMPROVEMENTS IN WIND POWER PLANTS

ELİF SARAÇOĞLU

BUSINESS ANALYTICS M.Sc. THESIS, SEPTEMBER 2020

Thesis Supervisor: Prof. Abdullah Daşcı

Keywords: Wind Turbine Upgrade, Wind Farm Performance Evaluation, Power Curve, Machine Learning, Power versus power

The operational optimizations, referring to the upgrades on wind turbines, can be very expensive; on the other hand, it is very complicated to assess the level of improvement they provide. Because of the inability to make reliable estimates on improvement levels, the plant owners are often reluctant to invest in upgrades. Like the OEM power curves, the improvement percentages for the upgrades, represent merely a reference and might differ for better or worse in the actual environmental conditions of the plant. The evaluations can not be done with a simple comparison of the pre-upgrade and post-upgrade performance, due to the complexity of the variables affecting power production and high levels of uncertainty of the environmental variables. In this research, we aim to study a machine learning approach implemented on wind farm level to evaluate the impact of operational improvements. Our approach consists of modeling the power output of the farm using a group of turbines referred to as the control turbines. The control group will not be upgraded to form the baseline for the pre-upgrade conditions. This baseline is later used to make a reliable comparison with the conditions after improvements are implemented.

ÖZET

RÜZGAR SANTRALLERINDEKI OPERASYONEL İYILEŞTIRMELERIN TAHMINLENMESI

ELİF SARAÇOĞLU

İŞ ANALİTİĞİ YÜKSEK LİSANS TEZİ, EYLÜL 2020

Tez Danışmanı: Prof. Dr. Abdullah Daşcı

Anahtar Kelimeler: Rüzgar Türbini İyileştirmeleri, Rüzgar Santrali Performans Değerlendirmesi, Güç Eğrisi, Makine Öğrenmesi, Güçe Karşı Güç

Rüzgar türbinlerinin performans iyileştirmesi için uygulanabilen operasyonel optimizasyonlar çok pahalı olabilir; ancak, sağladıkları iyileştirme düzeyini değerlendirmek çok karmaşıktır. İyileştirme seviyeleri hakkında güvenilir tahminler yapılamaması nedeniyle, tesis sahipleri genellikle yükseltmelere yatırım yapma konusunda isteksizdir. OEM (Orijinal Ürün Üreticisi) güç eğrilerinin gerçek performansı yansıtmaması gibi, iyileştirmeler için öngörülen yüzdeler de yalnızca referans olarak kullanılabilir. Güç üretimini etkileyen, karmaşık ilişkilere sahip değişkenler ve çevresel faktörlerin yüksek düzeydeki belirsizliği nedeniyle, değerlendirmeler iyileştirme öncesi ve sonrası sahip olunan performans koşullarının basit bir karşılaştırması ile yapılamaz. Bu araştırmada, rüzgar türbinlerinin çalışma koşullarına ait iyileştirmelerin etkisini ele almak adına rüzgar çiftliğinin sağladığı veriler temelinde bir makine öğrenimi yaklaşımını uygulamayı hedefliyoruz. Bu yaklaşım, kontrol türbinleri olarak adlandırılan bir grup türbin üzerinden çiftliğin güç çıkışını modellemeyi benimsemektedir. Kontrol grubu, bu süreçte iyileştirme öncesini niteleyen koşullara temel oluşturması açısından herhangi bir iyileştirmeye tabi tutulmayacaktır. Bu temel, iyileştirmeler uygulandıktan sonra değişen koşullar ile tutarlı ve güvenilir bir değerlendirme yapmak için kullanılır.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Prof. Abdullah Daşcı, for his invaluable guidance and insights, never-ending patience and constant encouragement. I would also like to thank him for the understanding and help from the first day I knocked on his door. I feel very honored for the opportunity to work under the supervision of him. I would also like to thank Assist. Prof. Levent Erişkin, Assoc. Prof. Raha Akhavan-Tabatabaei for their detailed evaluation of my thesis.

Secondly, I would like to thank my friends and family that helped me on this path. I want to thank Tibet Scarlet for the help and encouragement he gave me in the last minutes. I want to thank Eylül Akdemir for letting me experience the 'Eureka' moment in my darkest time and Burcu Sarı for her on-the-point, realistically absurd comments that made everything more tolerable. I would also like to thank Arda Ağababaoğlu, Berna Ünver, and Hatice Çakır for the great times we had at Sabancı University, especially the nights in the balcony of dorm B8. I am also profoundly thankful and would like to express my love to Ayşegül Sınav, Esra Urkan, Mert Altun, Ozan Temmuz Binicier, and Deniz Baran for their genuine friendship, constant support even when I was a bit unreachable; thus, I would also like to thank them for proving that true friendship is not bound by time. I also want to state my deepest gratitude to my mother, Serpil Saraçoğlu, for keeping up with my constant nagging and my father, Arif Saraçoğlu, for all the times he drove me to Tuzla because I missed the shuttle. I thank them for all the support they provided me.

Last but in no means least, I would like to thank my love, Ekin Türe, for his endless support from the start till the end, during all the ups and downs. I want to thank him for all the care and love he gave me, as well as the one last push that made me reach the end. During this time, I have seen again and again how lucky I am to have him by my side.

Dedicated to
Bıdık and Prozac

TABLE OF CONTENTS

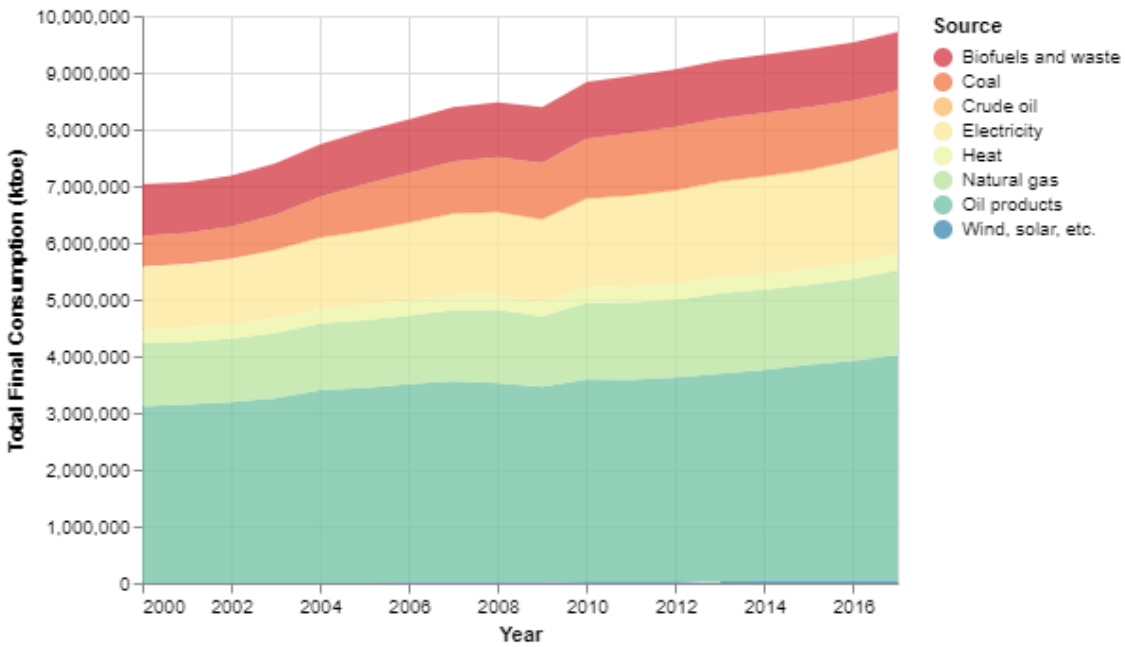
LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The total energy consumption of the world is increasing with population growth and the overall improvement in economic welfare. The world population grows by around 1% every year and with increasing life expectancy and improvements in healthcare, it is projected to reach 10.9 billion by 2100 (Roser, Ritchie & Ortiz-Ospina, 2019). The gross domestic product (GDP) of the world, on the other hand, is expected to double by 2040, increasing the demand for energy further (BP, 2019). The total final energy consumption through 2000 and 2017 can be seen in Figure 1.1; the leading sources of energy production in 2017 are oil products, electricity, and natural gas.

Figure 1.1 Total Final Consumption by Source, World 2000-2017

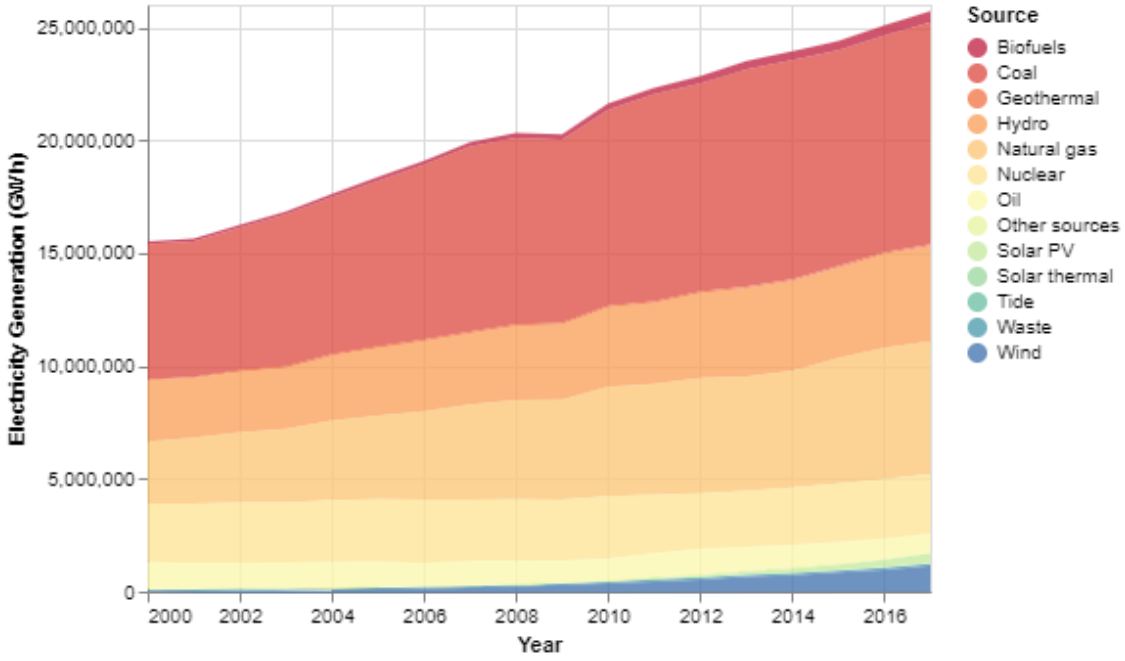


Source: IEA World Energy Balances 2019, https://www.iea.org/subscribe-to-data-services/world-energy-balances-and-statistics

Among the different energy sources, electricity is becoming more and more critical for two reasons: technological and environmental. The electricity demand is further increasing as technology is developing at a tremendous pace, getting cheaper and

more accessible, creating a shift to the digital in every aspect of our lives. The trend in technology is to make everything *smart*, from the thermostats in our homes to factories and even cities; however, even though *smart* systems promise more efficient and thus sustainable solutions, most of them depend on electricity, which is still produced using non-renewable resources. In 2017, the primary sources used in the production of electricity were coal and natural gases (see Figure 1.2), contributing to global warming in terms of CO_2 emissions.

Figure 1.2 Electricity Generation by Source, World 2000-2017



Source: IEA Electricity Information 2019, https://www.iea.org/subscribe-to-data-services/electricity-statistics

One of the main reasons for global warming is the burning of fossil fuels to supply energy (WWF, n.d.); from 2014 to 2019, CO_2 emissions due to energy production, increased every year by approximately 1.3% (IRENA, 2019). If the average global temperature rises by 1.5 °C, the effects of global warming might be irreversible, and to prevent this from happening, carbon emissions need to be limited by 45% until 2030 (IPCC, 2018). Policies worldwide shift towards renewable resources to tackle climate change and still meet the energy demand. The goal to reduce CO_2 pollution can only be reached by electrification of transportation, manufacturing, and industry: while at the same time, the resource for electricity is switched to renewables (IRENA, 2017).

In 2017, 19% of the world's energy supply was generated by renewable resources. Green energy accounted for 11% of the total final energy consumption (TFEC) in

2018, where 5.7% was made up of renewable electricity (IRENA, 2017). With the advancements in technology, the cost of wind and solar energy decreased to a level at which they could compete with fossil fuels and are forecasted to decrease further. As the production of renewable energy has become more cost-effective, many regions, including China, the US, EU, and India, started to switch to renewable energy (REN21, 2020). In the Table 1.1, the capacity of renewable resources used for power production for the years 2018 and 2019 for the World are given. In 2019, renewable energy capacity without including hydropower was 1437 GW, while wind power accounted for 45.3%, and solar PV accounted for 43.6%.

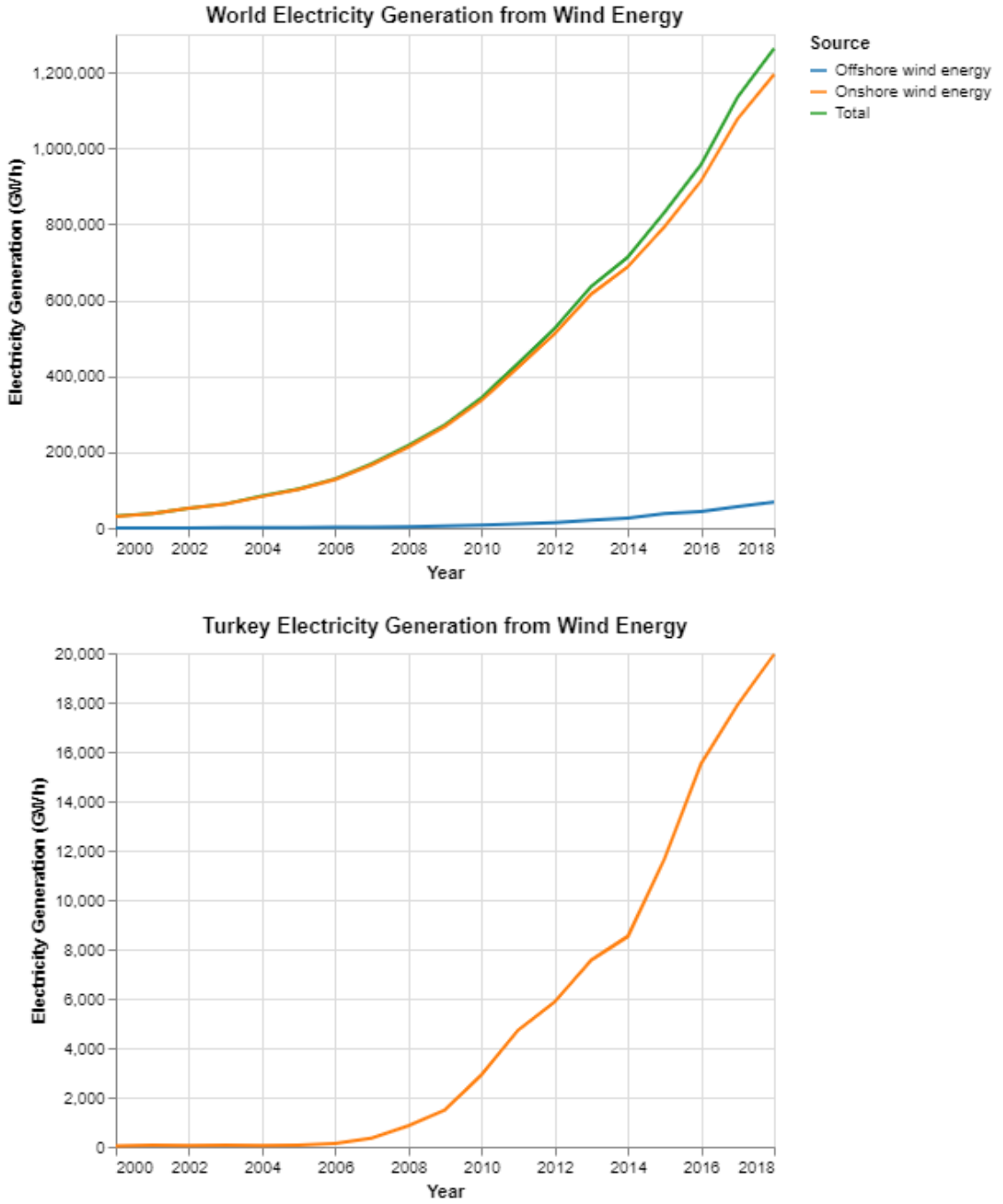
Table 1.1 Renewable Energy Indicators for Power Production

		2018	2019
Renewable Power Capacity (including hydropower)	(GW)	2387	2588
Renewable Power Capacity (not including hydropower)	(GW)	1252	1437
Wind power capacity	(GW)	591	651
Solar PV capacity	(GW)	512	627

Notes: Table data from REN21 (2020)

Solar power and wind power are forecasted to generate half of the world's energy capacity in 2040 (IEA, 2019). Wind turbines are accounted for 28.3% of the renewable electricity produced in OECD in 2019 and accomplished the second-fastest growth rate with an average growth of 20.7% since 1990. The growth rate since 2019 for electricity generation in the world and Turkey can be seen in Figure 1.3. The OECD region with the highest electricity production from wind turbines is OECD Europe, where the trend is towards the offshore wind plants. The United Kingdom had the lead in shares of offshore wind production in OECD with 45.4%, followed by Germany (33.1%), Denmark (7.9%) and the Netherlands (6.2%) (IEA, 2020). The highest total wind power capacity (including both onshore and offshore), however, belongs to China, followed by the United States, Germany, India, and Spain (REN21, 2020).

Figure 1.3 Electricity Generation from Wind Energy, 2000-2017



Source: IRENA (2020), Renewable Capacity Statistics 2020; & IRENA (2020), Renewable Energy Statistics 2020, The International Renewable Energy Agency, Abu Dhabi.

The power production and efficiency of a wind turbine depend on a wide range of parameters. For the sake of better explaining the purpose of this research, the parameters are vaguely grouped into two: design parameters and environmental parameters. This grouping solely aims to better distinguish the uncertain effects as well as the parameters which can be improved with upgrading. As with design parameters, we aim to describe the mechanical components that determine the power

performance of a wind turbine, i.e., the gearbox design, nacelle blade material and design, and the controller. These parameters can be optimized or can be improved to increase the power performance of a turbine. The environmental parameters represent the actual working conditions: i.e., the wind speed, the geographical topology, the relative position of the turbines in the farm, temperature, and dampness. As it is not possible to control the environment wind farm operates in, the power production ultimately depends on the environmental conditions, whereas the efficiency is determined by design. The aim of the upgrades is mainly to utilize the operating conditions by improving the design in order to transfer more wind energy to power.

To further increase the power production of a wind farm, a variety of upgrades can be implemented. These upgrades can be categorized into four groups: Improvements in wind turbine controls (i.e., Control system updating, Wind farm control, Pitch control, Intermittent wind energy capture, Handle special conditions), Tuning and optimization (i.e., Site-specific tuning, Individual turbine tuning, Nacelle misalignment), Aerodynamic performance (i.e., Blade add-ons, Increase blade size, Blade cleaning/restoration), Retrofits and modernization (i.e., Overhaul and modernization, Retrofit control systems, Retrofit drivetrain components, Retrofit electrical systems, Grid compatibility, Restoring power performance). These upgrades can improve the overall performance or the performance in a specific wind speed range, increase the maximum level of power output or extend the range of operation by increasing the maximum wind speed that the wind turbine can operate (Carlberg, 2015).

The plant owners are often reluctant to invest in upgrades, as site-specific improvement rate is not easily evaluated. Moreover, upgrades are often expensive, and the inability to assess the cost/income ratio makes the investors reluctant. The computation of the improvement is very complicated because it is not statistically correct to compare the performance of pre-upgrade and post-upgrade conditions without detailed analysis. The distribution of the environmental parameters (i.e., wind speed) may differ between the compared periods. Furthermore, the differences in these parameters affect the uncertainty levels from sensor measurements and wake effects.

In this research, we aim to study a machine learning approach implemented on wind farm level to evaluate the impact of operational improvements. Our approach consists of modeling the power output of the farm using a group of turbines referred to as the control turbines. The control group will not be upgraded to form the baseline for the pre-upgrade conditions. The method we apply uses data from SCADA systems. SCADA data is collected from the turbine controllers and are easily accessed. The literature mainly focuses on proving the improvement levels of upgrades for

only one turbine, to help the investor in the decision of upgrading. This research addresses one step ahead, the evaluation of the improvement from upgrading for the whole farm, for the investor to track the turnover from upgrading. In this research, one of the key findings is the selection of control turbines; with the right selection a small number of turbines can reflect the farm performance accurately.

This thesis is organized as follows: Literature review can be found in Chapter 2. The methodology for data preprocessing and analysis is explained in detail in Chapter 3, and Chapter 4, respectively. The final results are discussed in Chapter 5.

2. Literature Review

In this chapter, the literature on the upgrade assessment of wind turbines is introduced. To better understand the concepts in the literature, a brief explanation of the power production of a wind turbine is given, in addition to a general introduction to the methods of evaluating power performance. The power production of a wind turbine is a nonlinear function depending highly on volatile wind speed. Because of the uncertainties in the parameters that affect power production, obtaining a reliable evaluation of the performance of a wind farm requires complicated analysis. The level of uncertainty can quickly be apprehended when the calculation of the theoretical power output of a wind turbine is investigated:

$$P = 0.5\rho C_p A V^3 \tag{2.1}$$

Power output P depends on the wind speed V, the air density ρ, the swept rotor area A, and the power coefficient C_p which is a function of blade pitch. Power output is mainly dependent on the wind speed where the relation is cubic; thus, wind speed is widely used in the literature to predict the performance. The power output is also dependent on the absolute atmospheric pressure p, the absolute temperature T, the specific gas coefficient R, and humidity, as ρ is defined as $\rho = \frac{p}{RT}$ and R is dependent on humidity. The effects of T, R, and p have often been neglected in the literature. However, the variance in air temperature may affect the power output by 20%, and the variance in pressure can affect the power output by 10% depending on the geography (Schlechtingen, Santos & Achiche, 2013a). Additional parameters that influence the power production are local orography, wake effects caused by other turbines, wind direction, vertical and horizontal sheer, atmospheric stability, drive train temperature, and turbulence intensity (Schlechtingen et al., 2013a).

The main approaches to evaluate the performance of the wind turbine are modeling of the power curve and calculation of annual energy production (AEP). The power curve is the plot of power production versus the wind speed; the properties of the

power curves are expounded in Chapter 3. An initial power curve is modeled in controlled conditions by the manufacturers; however, these plots cannot be relied upon in the uncontrolled environment of the wind farm. The reason why manufacturer power curves do not account for the real performance is that power production from wind not only depends on the wind speed but also the characteristics of the air, topography of the area, direction of the wind and many other parameters that are specific to the location and local environmental conditions of the wind farm as explained previously. The empirical power curves that are obtained from the operating conditions may reflect the actual performance; however, the measurement of the wind speed can be unreliable. The wind speed is generally obtained through nacelle anemometers, but the measurements vary according to the position of the anemometer. In some cases, to obtain more reliable power curves, an external met mast sensor is used to obtain an undisturbed wind speed (Evans, Zhang, Iyengar, Chen, Hilton, Gregg, Eldridge, Jonkhof, McCulloch & Shokoohi-Yekta, 2014). After the power curve of the wind turbine is modeled, AEP can be calculated through the integration of the power curve to a given wind distribution. In AEP calculations, assumptions on the downtime and turbine failure are also included to account for the losses.

Wind turbines are upgraded to improve the efficiency of power generation from the kinetic energy of wind. These upgrades may improve the power coefficient C_p and the swept rotor area in order to extract more power, or they may decrease the losses. However, installing upgrades are costly and might halt production (Lee, Ding, Xie & Genton, 2015). Furthermore, the assessment of the improvements is not an easy task. Due to the uncertainty in weather conditions, (i.e., the wind speed distribution, humidity, temperature) and the multivariate dependency of power production to these conditions, a comparison between pre-upgrade and post-upgrade conditions is not reliable for assessment. On the other hand, wind operators need the means to evaluate the impact of the upgrades in order to validate the costs of upgrading (Lee et al., 2015).

In order to evaluate the difference in performance, the power curve analysis provides a relatively simple process. However, even though power curves map the relation between wind speed and power output, as stated before, power output is dependent on a lot of different parameters. The general approach in the literature to account for the uncertainties of environmental parameters is to compare a pair of turbines working in similar conditions and are positioned close to each other. One of the turbines from this pair is upgraded, and the other turbine is not upgraded to form a baseline. The turbine that is not upgraded is referred to as the control turbine or the reference turbine throughout this study: the turbine that is expected to have

a change in performance due to upgrades is referred to as the test turbine. In the literature, using the control turbine, the performance of the test turbine is modeled using different methods. Modeling is generally conducted for the period before the upgrade, as it is later used to assess the pre-upgrade conditions of the test turbine. There are two main approaches to model the relation between the test turbine and the control turbine: multivariate prediction of the power output and power-to-power relation. After the modeling step, in both approaches, the evaluation of the improvement is done using the power curves or AEP.

In multivariate power output prediction, the power output of a turbine is modeled using several other parameters in addition to wind speed. The resultant model is inherently a multi-dimensional power curve (surface) (Lee et al., 2015). Several different methods are applied to find the best fit for modeling. In this approach, it is essential to have reliable measurements to incorporate reliable models. One of the procedures for this approach is the KERNEL Plus method, introduced by Lee et al. (2015). In this method, the power curve modeling is done using multivariate kernel regression. The environmental parameters used to model the power output are obtained through a mast. The variables used in the model are the wind speed, wind direction, air density (calculated using temperature and air pressure), turbulence intensity, and vertical wind shear (both calculated using wind speed measurements). This method does not require the use of control turbines, as the modeling can be done using the turbines' previous data. There are other promising methods in the literature to model the baseline performance. In the work of Evans et al. (2014), power from other turbines, temperature, pressure, wind speed measured from a mast, lidar sensors, and the wind speed from nacelle anemometer are used. In addition to these parameters, the power output predicted from neighboring turbines were also used. These measurements were calculated using the Bayesian Power Curve method, which is a robust nonlinear regression model that uses Bayesian methods. In the study by Evans et al. (2014), stepwise linear regression, Lasso regression, and M5P regression were used to estimate the power output. At the farm level, the best performing model was found to be Stepwise Regression. In the research from Evans et al. (2014), AEP is modeled using natural splines and robust regression. Using the results of AEP prediction, a comparison between pre and post-upgrade states was conducted.

The second approach to upgrade evaluation is the side-by-side testing method (Albers, 2014). In the side-by-side method, the power-to-power relation of a control and test turbine is modeled. The relation is modeled by sections or bins: a second parameter is used for binning. Usually, when modeling a power curve, the binning parameter is wind speed; however, wind speed varies depending on the wake ef-

fects. To eliminate the wake effect but represent the condition of the wind, nacelle direction was used as the binning parameter in (Albers, 2014). Furthermore, as the power production of the test turbine is estimated through a turbine that operates at roughly the same conditions, this method can eliminate the environmental uncertainties (Lee et al., 2015).

A case study to evaluate the impact of vortex generators was conducted, and the two approaches were compared by Hwangbo, Ding, Eisele, Weinzierl, Lang & Pechlivanoglou (2017). KERNEL Plus (multivariate power output prediction) and "side-by-side testing" (power-to-power relation) was used in this study. Power-to-power method estimates had a lower degree of uncertainty given a large dataset compared to multivariate prediction; there were also fewer assumptions made in this method. However, the power-to-power method requires a one-by-one pairing, whereas multivariate methods do not require a second turbine.

In this study, our aim is to propose a method on the identification of the effect of operational optimizations on wind farm level. To summarize, the operational optimizations, referring to the upgrades on wind turbines, can be very expensive; however, it is very complicated to assess the level of improvement they provide. Like the OEM power curves, the improvement percentages for the upgrades, represent merely a reference and might differ for better or worse in the actual environmental conditions of the plant. The task of identifying the level of improvement cannot be made by a simple comparison of the pre-upgrade and post-upgrade performance, due to the complexity of the variables affecting power production and high levels of uncertainty of the environmental variables. There is very little research that addresses this problem, and mostly they are not conducted on the wind farm level.

What we propose in this study differs from the literature in two main ways. Firstly, in our method, instead of assessing the improvement on the turbine level, we are addressing the problem at the farm level: the models built in this study predict the total power output of the farm instead of one turbine. Secondly, We combine the power to power relation method with multivariate prediction to build the prediction models. In the power-to-power method, a control turbine that performs similarly to the test turbine is used to create a baseline; in our method, we chose a group of turbines for the same purpose. The turbines with similar behavior are clustered into groups, and the control turbines are chosen from each group as a representative. The predictive models are built using the control group's power production.

3. Data Preprocessing and Descriptive Analytics

For this research, we were provided with the Supervisory Control and Data Acquisition (SCADA) data for 52 wind turbines in a power plant in Turkey. In this Chapter, the attributes of the data, data filtering, and preprocessing steps are explained in detail and the descriptive statistics are provided.

3.1 Data Collection

SCADA systems are automation control systems and are mainly used to provide a centralized control unit to monitor and control the working conditions of a plant. These systems are adjustable; therefore, they are used in many areas such as manufacturing, oil and gas, water, and most commonly in power and energy (Roy, 2015). They can be used for monitoring a single piece of equipment in a plant, the entire plant, or even multiple plants in a region. To be able to monitor and control the site, SCADA systems collect the measurements and status of the sensors from the equipment at regular intervals and store these measurements as a distributed database with an associated timestamp (Krambeck, 2015).

For wind turbines, using SCADA is a convenient choice as it can provide essential data such as wind parameters, energy conversion parameters, vibration parameters, and temperature parameters (Kusiak & Li, 2011). The information is collected at the controller and can have more than 150 features in a single timestamp. Some of the parameters are listed below To provide a better view on how much data SCADA systems provide: wind speed, wind direction, wind intensity, turbulence, power output, reactive power, power factors, blade pitch angle, generator torque, rotor speed, drive train acceleration, tower acceleration, bearing temperature, nacelle interior temperature, ambient temperature, spinner temperature, etc. (Schlechtingen, Santos & Achiche, 2013b). In addition to these parameters, fault information

and turbine status are also logged. The list above is just a small subset of the sensor information obtained with SCADA; furthermore, these parameters are collected with a frequency of 5 to 10 minutes, creating a very large, detailed time-series data (Gill, Stephen & Galloway, 2012).

Even though SCADA systems can provide a wide variety of parameters, the data we were provided only consisted of wind speed, nacelle direction, and the power output for each turbine individually from the power plant. The data points were collected in 10-minute intervals for two years from 2016 to 2017, adding up to 105264 timestamps. We were neither provided with the position of the wind turbines nor the failure and alarm logs. We were also not provided with the OEM specifications of the turbines in the plant.

Table 3.1 Parameters Provided in the Data

Parameter Name	Unit	# of Datapoints
Wind Speed	m/s	105264∗52
Power Output	KWh	105264∗52
Nacelle Direction	°	105264∗52

3.2 Data Preprocessing

For this research, to be able to evaluate even a small difference in power production, the prediction model needs to have high accuracy; thus, detailed data preprocessing was required. Even though SCADA data provides a large amount of value-timestamp points, its quality is fairly low. Some of the potential causes that affect SCADA data quality are sensor accuracy, EMI, information processing errors, storage faults, fault in communication systems, and alarms; these circumstances might cause the storage of false values or of null data points. In the industry, specialists filter the data manually because of the complexity of the potential errors in the data: the fluctuation due to errors is hard to detect, and alarm records need to be thoroughly analyzed. However, manual filtering is a very time-consuming task (Llombart, Pueyo, Fandos & Guerrero, 2006).

In addition to the requirement of thoroughly cleaned data, we also needed to have consistent data points for all turbines as the research was aiming to compare each turbine with the others to be able to choose the most representative ones. Having

consistent data points implies that the same timestamps should be used for all 52 turbines for wind speed, power output, and nacelle direction to preserve the correlation. It was also crucial that all the turbines were in working conditions and online for the analysis. In other words, even if one of the turbines had faulty measurements at a timestamp, all the data for that timestamp needed to be removed, causing a massive data loss (Carlberg, 2015).

The initial step for filtering is mainly the handling and removal of the alarm records Carlberg (2015); Llombart et al. (2006). In our case, we were not provided with these records or the status logs; thus, the failure timestamps could not be marked in such fashion. The steps we conducted to clean the data are as follows: visual inspection of the power curves, removal of negative power production data points, removal of null values, filtering of erroneous data points, filtering of the points below the rated wind speed. In the table below, the percent of the data removed in each step is shown.

Table 3.2 Steps of Data Filtering

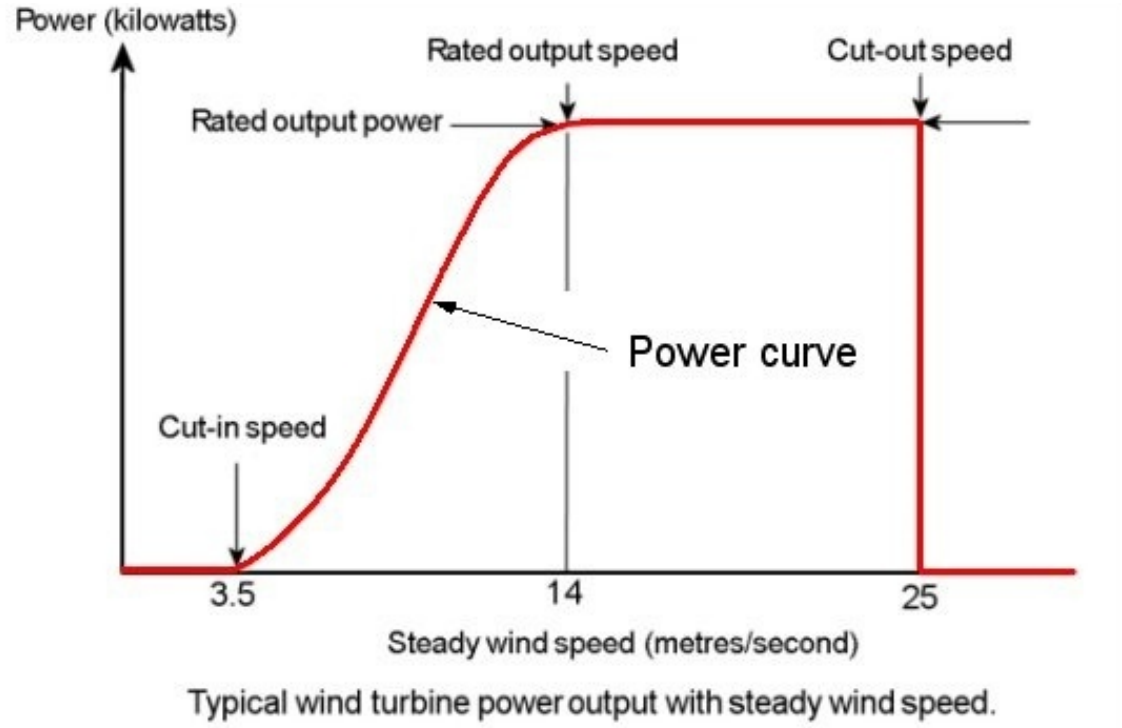
Step No	Explanation	Final # of Timestamps	% of Data Removed
1	Visual inspection of the power curves	105264	-
2	Removal of negative power production data points	12634	88 %
3	Removal of null values	11062	12.4 %
4	Filtering of erroneous data points	5901	46.6 %
5	Filtering of the points below the rated wind speed	5814	1.4 %

3.2.1 Visual Inspection of the Power Curves

Firstly, a visual inspection to evaluate the quality of the data was conducted by drawing the empirical power curves of the turbines to determine what steps to take for filtering. A power curve is an important indicator of the power performance of a wind turbine. A power curve maps the relation of the power produced with the wind speed, and the relation typically resembles a sigmoid function. Generally, power curves are supplied by the OEM; however, these power curves are obtained in ideal meteorological and topographical conditions, such as reduced turbulence and air-density corrections (Gill et al., 2012). Thus, they do not represent the actual working power performance of the turbine but are used as a reference. The reason for this difference between the OEM power curve and the actual power curve is mainly because of the topographical attributes of the location of the wind farm as

well as the environmental attributes such as wind speed distribution, air density, and wind direction. Also, mechanical condition, malfunctions and control issues of the turbine itself, in addition to the uncertainties in the measurements, change the shape of the power curve for each turbine (Kusiak & Verma, 2012; Shokrzadeh, Jafari Jozani & Bibeau, 2014). Obtaining the empirical power curve, at the actual working conditions, is a research topic by itself and is widely studied in the literature.

Figure 3.1 Example Power Curve



Source: https://www.quora.com/What-is-a-power-curve-and-how-do-we-draw-one

A power curve has three essential distinctive wind speeds: cut-in, rated speed, and cut-out speed. Under the cut-in speed, the turbine does not generate power, rated speed is the speed at which rated power is produced, and cut-out speed is the highest speed where the turbine can work without incurring damage (Shokrzadeh et al., 2014). As we were not equipped with the OEM power curve, we made assumptions on these features in order to use for filtering in the later steps by visually assessing the power curves.

3.2.2 Removal of Negative Power Production Data Points

Cut-in speed is the minimum wind speed at which the wind turbine can effectively produce a power output: below this speed, either negative power or no power at all is produced. Our initial aim was to filter out the data points below the cut-in speed; however, from the power curves drawn, we were not able to distinctly distinguish the cut-in speed. Nevertheless, throughout the data, there were negative power output measurements, meaning that this was either an erroneous measurement or the wind speed was below cut-in speed. Both cases were unacceptable; thus, all the timestamps that included a negative power output value for any turbine were dropped. With this step, 92630 timestamps were removed, and 12% of the data remained. The distribution of the number of timestamps containing a negative value for each turbine is given in the figure below.

Figure 3.2 Distribution of Negative Power Production

3.2.3 Removal of Null Values

The null value ratio was very high in the data we were provided. Even though we explained the possible reasons for null values at the beginning of this Chapter, we did not have the data to analyze the reasons for the null values and could not differentiate between the sensor errors and the downtime of the turbines. As the research required to use the data where each turbine was online and working, and that we did not have the means to characterize the causes behind the missing values, specifically downtime, we did not use imputation. Instead, we decided to remove the timestamps containing missing values. In the Figure 3.3, the missing rates for the turbines sorted by each parameter are shown.

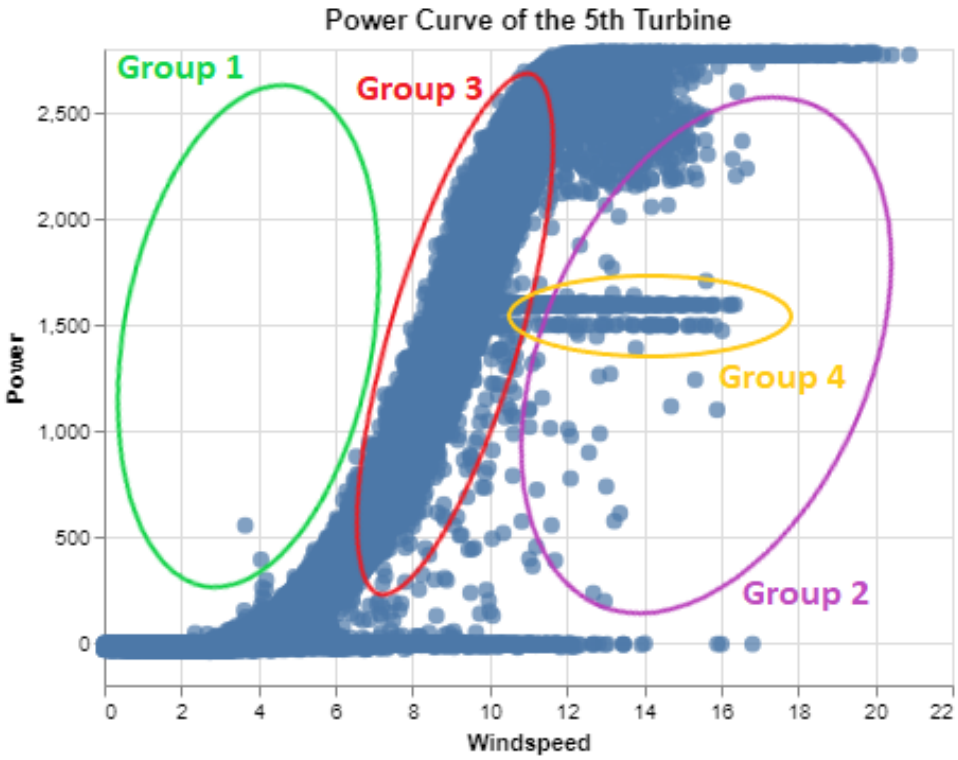
Figure 3.3 Distribution of Null Values for each Parameter

The missing values for power output and the wind speed were snychronious; however, most of the nacelle direction data was erroneous and missing. When the missing values from the nacelle direction were also taken into consideration, a massive portion (around 53.5%) of the data was lost. Therefore, we decided not to use nacelle direction parameter and instead, work with only wind speed and power output as these two parameters were the main indicators of performance. The percentage of data lost from removing the null values, for the whole data set and the filtered dataset for negative values is shown in the table below. At this step, all the timestamps containing a null value for any of the turbines were filtered out, removing 12.4% of the data. After this step, we were left with 11062 timestamps.

3.2.4 Filtering of Erroneous Data Points

In the visual inspection step, a lot of erroneous data points were observed. These measurements can be grouped into four: points above the power curve points below the power curve, points scattered relatively close to the power curve, and curtailment. The cases where the data points are located above the power curve are mainly caused by wake effects when the wind speed is reduced due to an obstacle before reaching the turbine. These points are labeled as group 1 in the figure below. The points which are located below the power curve, shown as group 2 in the figure below, are due to the averaging when the data is logged. If the turbine does not fully work in the 10-minute interval, the average of the power output will decrease, carrying down the data point logged. The points scattered relatively close to the power curve compared to group 1 and group 2, shown as group 3 in the figure below, cannot be directly labeled as errors, as the reasons are generally not clear for these points (Llombart et al., 2006). Curtailment, shown as group 4 in the figure below, is generally caused when there is no more capacity to receive more energy as transmission systems are working at the highest possible rate or when the demand is low, and thus the production of the turbine is lowered by discarding some of the wind energy (Qiggle, 2017).

Figure 3.4 Type of Errors Indicated on a Power Curve



Manual filtering was not considered in this research, as we did not have the expertise to perform this task, and even if we did, it would be very time consuming to conduct it for 52 turbines. Instead, we used two different methods based on the research by Llombart et al. (2006) and IEC 61400-12 standard.

The first method used is implemented as follows: Firstly, the data is partitioned by grouping the wind speed into 0.5 m/s bins. Afterward, the mean value (μ) of power output and the standard deviation (σ)are computed for each bin. The third and last step is to go back to the initial data and filter the power output by $\mu \pm 3\sigma$ for each bin. This method performed well; however, for some turbines, curtailment was not filtered.

The second method is the same as the first one except for the second step. For this method, instead of the mean and standard deviation, the median and the median absolute deviation (MAD) are calculated for each bin. In the third step, filtering is done the same way; however, this time, median and MAD are used. For each bin, values outside the scope of $median \pm 3MAD$ are removed. The second method performed better at dealing with curtailment; therefore, it was decided to use this method. In this step, 46.6% of the data was removed, and 5901 timestamps were maintained.

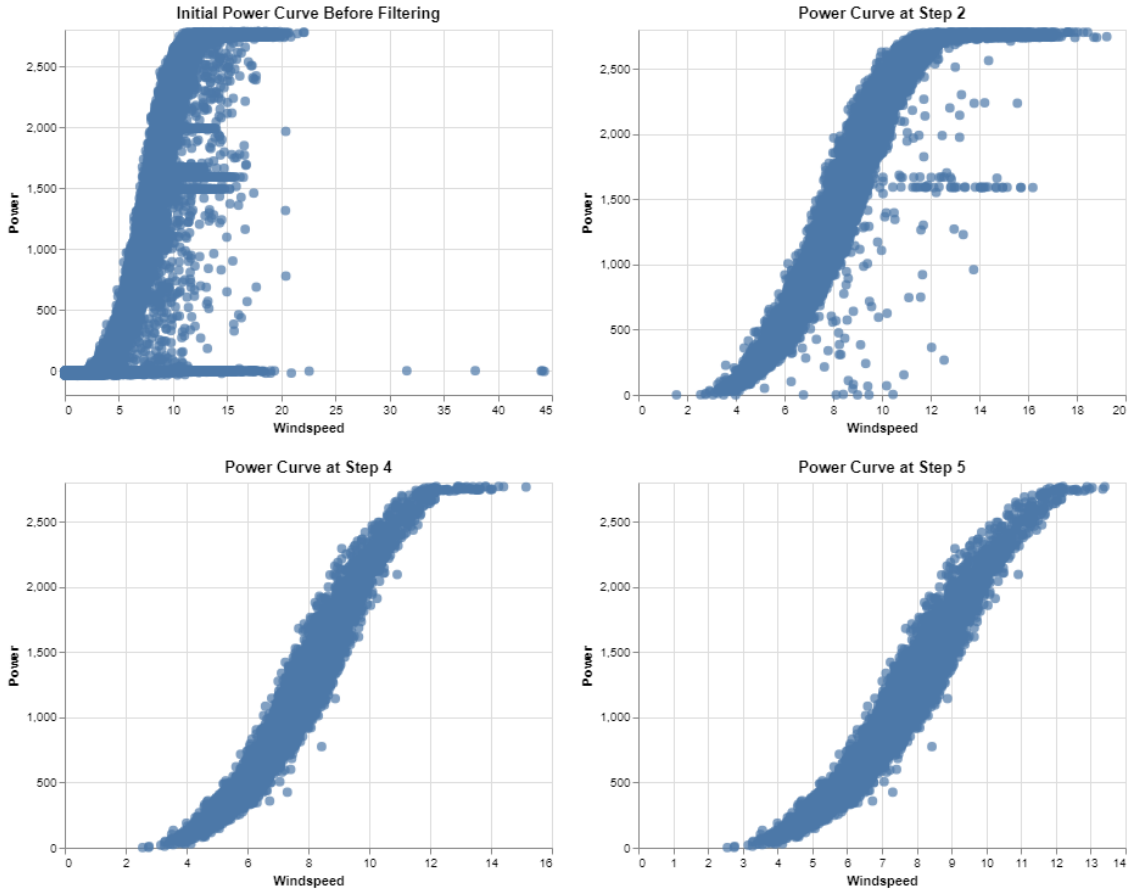
3.2.5 Filtering of the Points Below the Rated Wind Speed

After the fourth step, the power curves were drawn again for inspection. It was observed that filtering with the median method caused a more substantial loss from the data points at rated power or close to the rated power. At rated wind speed, the power output is regulated to stay constant at rated power; at this point, the maximum capacity of the generator is reached, and by adjusting the blades, constant power output is obtained until the wind speed reaches the cut-off value. Rated power is also referred to as the maximum power output that the turbine can reach.

From the power curves, it was assumed that the rated wind speed was 14 m/s, and above this limit, there were only a handful of data points left. In order to prevent these few datapoints from acting as outliers and disrupting the analytical models, all the measurements where the wind speed was above 14m/s were removed. The portion of the data removed was 1.4%, and there were 5814 timestamps left. With this step, data filtering was completed.

In the figure below, the resultant power curves are visualized for each step applied in data preprocessing.

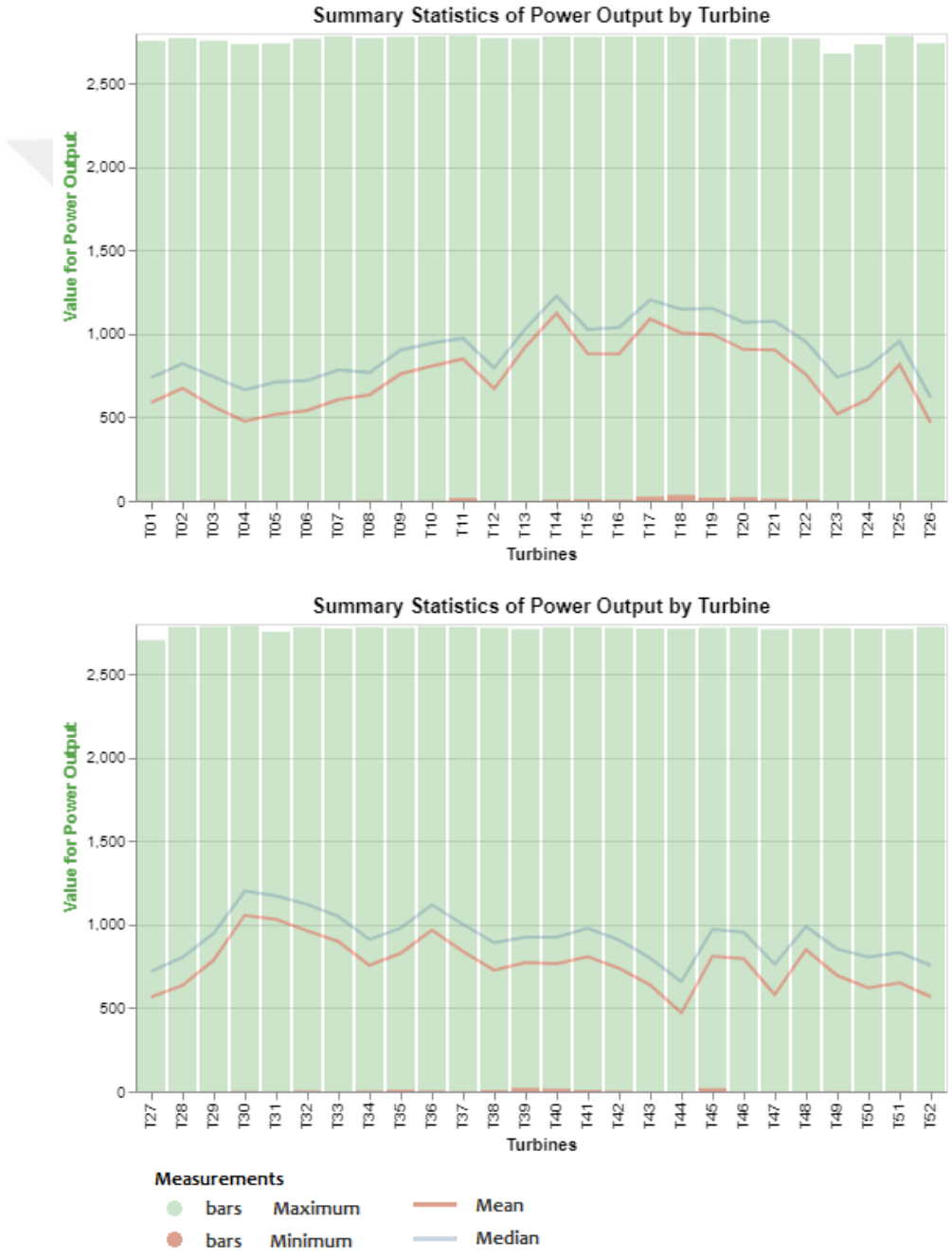
Figure 3.5 Power Output vs. Wind Speed Distribution in each Filtering Step



3.3 Descriptive Statics

In this section summary statistics of the data is provided. As filtering steps eliminated a high percentage of the data, time relation could not be maintained and was not regarded. The maximum, minimum and average power output for each turbine is in the Figure 3.6.

Figure 3.6 Summary Statistics of Power Output By Turbine



The summary statistics of power output for all of the data can be found in the table below.

Table 3.3 Descriptive Statistics of Power Output

Total count of data points	302328
Mean	916.14
Standard Deviation	658.32
Min	0
25%	385.31
50%	746.44
75%	1330.42
Max	2789.02

The maximum, minimum and average wind speed for each turbine is visualized in the Figure 3.7.

The summary statistics of wind speed for all of the data can be found in the table below.

Table 3.4 Descriptive Statistics of Wind Speed

Total count of data points	302328
Mean	7.08
Standard Deviation	1.749
Min	0
25%	5.78
50%	6.95
75%	8.27
Max	14

Figure 3.7 Summary Statistics of Wind Speed By Turbine

4. Analysis and Results

The method proposed in this research focuses on the prediction of the total power output of pre-upgrade conditions using several turbines as a control group, while the rest of the turbines are used for assessing the post-upgrade conditions. By this method, we aim to evaluate the increase in performance by comparing these two groups, while minimizing the effect of uncertainty of environmental conditions. The selection of the control group is the most critical task, as these turbines should represent the behavior of the farm itself.

The method proposed in this study is based upon by the case study conducted by Marcus Carlberg, which is inspired by 'Side-by-Side Testing to Verify Improvement of Power Curves' by Axel Albers (Carlberg, 2015). In 'Side-by-Side Testing to Verify Improvement of Power Curves,' Axel Albers (2014) presents an approach to identify the improvement on the power curve of a turbine by comparing it to an identical turbine positioned as neighbors. One of the turbines is the test turbine, where a change in the power curve is expected, while the other is the reference turbine used to form a baseline relation. The method uses only SCADA data to models the power-to-power relation between two turbines by using wind direction as the reference parameter. The power-to-power relation is modeled during the training period, where no change in behavior is expected. In the testing period, the power curve of the turbine is recreated from the relation with the reference turbine to represent the behavior before the change. This power curve is later compared with the empirical power curve of the test turbine in the testing period, in order to identify any changes (Albers, 2014). In the case study by Carlberg, this method was used to acquire the level of improvement provided by vortex generators (Carlberg, 2015).

In our case, we did not have the data needed for Albers' method. Nacelle direction was a key feature to minimize the uncertainty from wake effects in the method he proposed; however, nacelle direction data we were provided was mostly erroneous (Carlberg, 2015). Furthermore, we needed to find a solution on the wind farm level instead of the turbine level. Therefore, we decided to do some-to-rest testing instead of side-by-side. In this some-to-rest model, some turbines in the wind farm

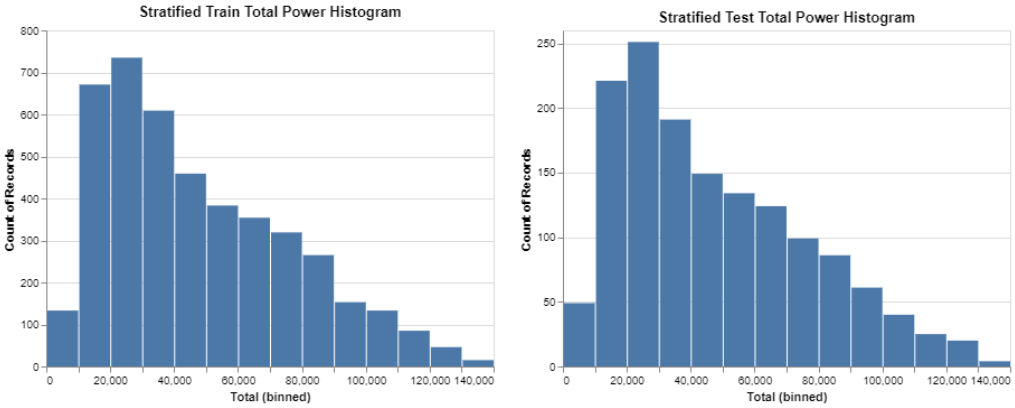
are chosen as the reference/control turbines, while the rest of the turbines in the wind farm are the test turbines that undergo an optimization. The idea behind this method is to model the behavior of the test turbines, using the control turbines, instead of a one-to-one comparison. Parallel to the side-by-side method, the behavior of each control turbines needs to represent a group of turbines similar to itself; in other words, each control turbine would be the representative of a group. In order to identify turbines with similar performance, the pairwise correlations were calculated using power output. The correlations between the turbines were very high, with a minimum of 0.45192 and a maximum of 0.980692. To decide on the groups that have similar behavior, we decided to use clustering. Clustering was implemented in several different ways, for a different number of clusters. The Elbow method was used to identify the ideal number of clusters. Clustering by visual inspection was also considered. The details for the implementation of the clusters is explained in Section 4.1. We decided upon using three different results that had 5, 6, and 7 clusters.

A successful control turbine selection required the turbine to be highly correlated with the cluster it is representing while having a lower correlation with the other control turbines; in order to achieve a higher score from the prediction model we needed to maintain as much information as possible while maintaining a small control group. The need for a small control group is purely financial, as the control group would not have the upgrade implemented; thus, the power output for the farm would decrease the more turbines are used for the control group. Instead of using an optimization algorithm to find the combination of control turbines that meet the requirement explained above, we decided to find the most successful combination-model pair. We aimed to model the performance of the farm using every combination of control turbines and obtain the control group that provides the highest accuracy.

To model the performance of the farm, we needed to decide upon an indicator that would be later be used as the dependent variable for the predictive models. In the literature, to evaluate differences in the performance of turbines, mostly power curves and annual energy production (AEP) was used. However, comparing the power curves for individual turbines was not efficient for the method we proposed, and the calculation of AEP required several assumptions increasing the uncertainty. Using wind speed required much additional information, such as the Nacelle Transfer Function, air density, and temperature, as wind speed measurements are not as trustworthy as power output because of its nature. (Carlberg, 2015). It was decided to use only the power output parameter for the analysis, and farm performance at each timestamp was represented with the total power output for the test turbines.

The algorithms used for clustering are k-means, k-medians, and hierarchical clustering, whereas linear regression, lasso regression, ridge regression, k-nearest neighbor (KNN) regression, and gradient boosting machine (GBM) regression was used for predictive modeling. Before starting the clustering step, the data was split into train and test sets by 75-25%. The train set contained 4360 timestamps, whereas the test set contained 1454 timestamps. Two methods were used to split the data: stratified sampling and random sampling. Stratification parameter for stratified sampling was total power output for all 52 turbines. There was no significant difference between the random sampling and the stratified sampling, except the power output distributions obtained from stratified sampling had less differences between the train and the test sets as expected. The distribution of the total power output for stratified sampling is shown in Figure A.1. The research proceeds with using the sets produced with stratified sampling: nevertheless, both methods were used in the clustering process.

Figure 4.1 The Total Power Output Distribution for Stratified Train and Test Sets



4.1 Clustering

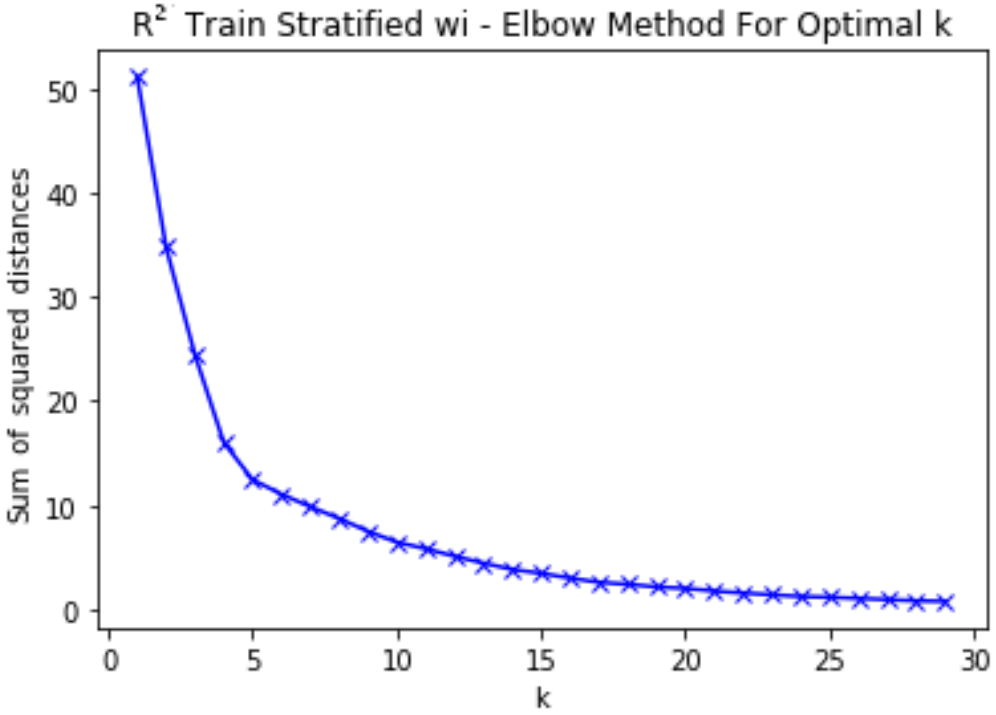
The purpose of clustering in this research is to identify and group the turbines that better explain the performance of each other. To better understand the relation between the turbines, we found the coefficient of determination for every pair and created a baseline matrix. Using the power output of each turbine in the farm as the independent variable, the production of the rest of the turbines were predicted using linear regression models with intercept and without intercept. During the modeling, only the train set that was obtained via stratified sampling (STRS) was

used. For each pair, the R^2 values were recorded, and two matrixes of R^2 coefficients were formed. One was formed using the results of regression with intercept, and the other was formed using the results of regression without intercept. The same process was repeated for the train set that was formed by random sampling (RTRS). From these matrices, a pattern could be clearly seen. At this point, we built the first clusters by visual inspection.

We constructed several different pipelines to perform clustering. First of all, we decided to use both RTRS and STRS. We also decided to use R^2 matrices we obtained in the previous step, to perform clustering using the pairwise R^2 coefficients between each turbine in addition to RTRS and STRS. As clustering algorithms, we decided to use k-means, k-medians, and hierarchical clustering.

To decide upon the number of clusters, the elbow method was used. The elbow methods was implemented for all the pipelines. The results of clustering with R^2 matrix obtained from regression with intercept is given in the Figure 4.2 .

Figure 4.2 The Elbow Curve for R^2 Matrix Obtained with STRS



The resulting clusters of all the pipelines can be found in Appendix A. We decided to use the results of the three pipelines: k-means clustering using R^2 matrix with intercept for 5, 6 and 7 clusters. From the clusters obtained, combinations were composed to be used as the dependant variables for the predictive models.

4.2 Predictive Models

The combination of the control turbines indicates the independent variables for the predictive model. The power output of these turbines would be the input features for the models to be built. The dependent variable is the total power output of the rest of the turbines. Using the power output of the control turbines, the total power output of the rest of the farm is predicted in the models. Due to the time constraint of this project, the initial aim of inspecting every possible combination-model pair could not be accomplished: instead, a subset of 3000 combinations was chosen (1000 for each clustering method). The models, linear regression, lasso regression, ridge regression, KNN regression, and GBM regression, were built and hyper tuned for each individual combination. As with each combination, the input dataset of the model changed; it was critical to do the hyperparameter tuning for each combination to assure the low error rates. For the error, calculations root mean squared error (RMSE) was used. The details and the best results of each model are explained in detail below. Full results can be found in Appendix B.

4.2.1 Linear Regression

Linear regression was the first model built in this research, and as it had fairly low RMSE for both the train and test data sets it was taken as a baseline. Linear regression was built using initially 10000 combinations and created a baseline for the selection of the subset of 1000 combinations. The best five results for linear regression are shown in the table below. Full results can be found in Appendix B.

Table 4.1 Best Results from Linear Regression with 5 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T11, T19, T33, T45	0.9837	0.9837	3348.64	3351.21
2	T06, T12, T20, T31, T45	0.9833	0.9834	3404.05	3385.20
3	T06, T13, T16, T32, T45	0.9821	0.9834	3516.15	3385.74

Table 4.2 Best Results from Linear Regression with 6 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T05 ,T13 ,T21 ,T29 ,T42, T49	0.9868	0.9878	2959.61	2846.65
2	T06 ,T08 ,T18 ,T31 ,T39, T49	0.9871	0.9876	2932.50	2860.80
3	T02 ,T09 ,T19 ,T31 ,T42, T47	0.9870	0.9875	2926.64	2871.88

Table 4.3 Best Results from Linear Regression with 7 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T06, T09, T17, T25, T31, T41, T49	0.9906	0.9904	2436.21	2454.15
2	T06, T12, T19, T35, T33, T41, T51	0.9903	0.9901	2476.40	2499.91
3	T07, T10, T14, T22, T31, T48, T42	0.9901	0.9910	2498.08	2380.24

4.2.2 Lasso Regression

Linear regression with L1 regularization was built using the subset of 1000 combinations for each of the clustering methods. In order to find the right α for each combination, grid search with three-fold cross-validation for the values 1, 10, 100, 1000, 10000, 100000, 1000000 was implemented. The best model parameters for each combination were chosen according to the average RMSE of the folds. Using the whole training set, and the obtained 'best parameter' the model was rebuilt for each combination. The best results are given below. Full results can be found in Appendix B.

Table 4.4 Best Results from LASSO Regression with 5 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T11, T19, T33 , T45	0.9837	0.9837	3346.34	3351.22
2	T06, T12, T20, T31 , T45	0.9833	0.9835	3401.71	3385.24
3	T06, T13, T16, T32 , T45	0.9821	0.9834	3513.73	3385.76

Table 4.5 Best Results from LASSO Regression with 6 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.9868	0.9878	2957.23	2846.67
2	T06, T08, T18, T31, T39, T49	0.9871	0.9877	930.15	2860.81
3	T02, T09, T19, T31, T42, T47	0.9870	0.9875	2924.29	2871.83

Table 4.6 Best Results from LASSO Regression with 7 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	0.9901	0.9910	2495.79	2380.24
2	T06, T09, T17, T25, T31, T41, T49	0.9906	0.9905	2433.97	2454.16
3	T06, T12, T19, T35, T33, T41, T51	0.9903	0.9901	2474.13	2499.92

4.2.3 Ridge Regression

Linear regression with L2 regularization was built in a similar fashion to lasso regression. Grid search with three-fold cross-validation to find the best α between 1, 10, 100, 1000, 10000, 100000, 1000000 was implemented for each of the 3000 combinations. To decide upon the best model, RMSE was used; for each combination, the models were built again with the best parameters. The best results are given below. Full results can be found in Appendix B.

Table 4.7 Best Results from Ridge Regression with 5 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T11, T19, T33, T45	0.9837	0.9837	3346.34	3351.28
2	T06, T12, T20, T31, T45	0.9833	0.9835	3401.71	3385.28
3	T06, T13, T16, T32, T45	0.9821	0.9834	3513.94	3385.87

Table 4.8 Best Results from Ridge Regression with 6 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.9868	0.9878	2957.23	2846.67
2	T06, T08, T18, T31, T39, T49	0.9871	0.9877	2930.28	2862.38
3	T02, T09, T19, T31, T42, T47	0.9870	0.9875	2924.41	2872.30

Table 4.9 Best Results from Ridge Regression with 7 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	0.9901	0.9910	2495.89	2381.01
2	T06, T09, T17, T25, T31, T41, T49	0.9906	0.9905	2433.98	2454.25
3	T06, T12, T19, T35, T33, T41, T51	0.9903	0.9901	2474.13	2499.92

4.2.4 KNN Regression

For the hyperparameter tuning of KNN regression, a grid search with 4-fold cross-validation was used to find the best combination of parameters. The parameters used for the hyperparameter tuning is as follow:

- Number of neighbours: 4, 6, 8, 10, 15, 20, 25, 30
- Weights: Uniform and distance
- Euclidean and Manhattan distance

The best model parameters found for each combination were used to retrain the models using the whole dataset. The results for all the 3000 combinations can be found in Appendix B, and the best results can be found below.

Table 4.10 Best Results from KNN Regression with 5 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T34, T10, T20, T31, T46	1	0.9841	0	3303.12
2	T06, T10, T16, T31, T46	1	0.9839	0	3328.70
3	T06, T13, T16, T32, T45	1	0.9837	0	3351.55

Table 4.11 Best Results from KNN Regression with 6 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T06, T08, T18, T31, T39, T49	1	0.9876	0	2869.27
2	T34, T10, T20, T28, T48, T43	1	0.9874	0	2894.01
3	T02, T09, T19, T31, T42, T47	1	0.9873	0	2895.47

Table 4.12 Best Results from KNN Regression with 7 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	1	0.9912	0	2348.75
2	T06, T09, T17, T25, T31, T41, T49	1	0.9902	0	2474.05
3	T07, T10, T15, T20, T31, T41, T49	1	0.9898	0	2530.42

4.2.5 GBM Regression

GBM Regression was built using 500 estimators, max depth of 4, and learning rate of 0.01. Hyperparameter tuning was not implemented in GBM due to the computation load. The best results obtained from GBM Regression are as follows.

Table 4.13 Best Results from GBM Regression with 5 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T06, T10, T16, T31, T46	0.9867	0.9821	3031.32	3516.88
2	T05, T08, T15, T31, T46	0.9860	0.9821	3118.46	3526.36
3	T06, T10, T21, T31, T41	0.9863	0.9817	3062.44	3551.16

Table 4.14 Best Results from GBM Regression with 6 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.9898	0.9863	2600.90	3014.54
2	T02, T09, T19, T31, T42, T47	0.9904	0.9861	2517.95	3032.27
3	T05, T11, T16, T29, T40, T49	0.9900	0.9858	2581.23	3074.81

Table 4.15 Best Results from GBM Regression with 7 Clusters

Rank	Combination	R^2	R^2 Test	RMSE Train	RMSE Test
1	T06, T09, T17, T25, T31, T41, T49	0.9929	0.9889	2117.03	2644.48
2	T07, T10, T14, T22, T31, T48, T42	0.9922	0.9888	2208.42	2657.08
3	T01, T08, T19, T35, T31, T40, T47	0.9917	0.9885	2293.12	2707.49

4.3 Results

Overall the best 20 models for 5 Clusters are listed in Table 4.16. With less turbines KNN regression performed better than the other modeling algorithms; 10 of the 20 best results were obtained using KNN regression. The best combinations was T34, T10, T20, T31, T46. The turbines that occurred the most in the 20 combinations are T06 and T31 (13 times), followed by T05 and T45 (9 times).

Table 4.16 Best 20 Combination-Model Pairs for 5 Clusters

Rank	Combination	Model	R^2	RMSE Train	RMSE Test
1	T34, T10, T20, T31, T46	KNNRegression	1	0	3303.12
2	T06, T10, T16, T31, T46	KNNRegression	1	0	3328.70
3	T07, T11, T19, T33, T45	LinearRegressionWithIntercept	0.9837	3348.64	3351.21
4	T07, T11, T19, T33, T45	RidgeRegression	0.9837	3346.34	3351.28
5	T06, T13, T16, T32, T45	KNNRegression	1	0	3351.55
6	T06, T12, T20, T31, T45	KNNRegression	1	0	3354.39
7	T07, T11, T19, T33, T45	KNNRegression	1	0	3355.28
8	T06, T10, T21, T31, T41	KNNRegression	1	0	3355.95
9	T05, T08, T15, T31, T46	KNNRegression	1	0	3364.85
10	T05, T10, T18, T31, T42	KNNRegression	1	0	3365.18
11	T06, T10, T20, T31, T46	KNNRegression	1	0	3384.75
12	T06, T12, T20, T31, T45	LinearRegressionWithIntercept	0.9833	3404.05	3385.20
13	T06, T12, T20, T31, T45	RidgeRegression	0.9833	3401.71	3385.28
14	T06, T13, T16, T32, T45	LinearRegressionWithIntercept	0.9821	3516.15	3385.74
15	T06, T13, T16, T32, T45	RidgeRegression	0.9821	3513.94	3385.87
16	T34, T08, T19, T32, T41	KNNRegression	1	0	3399.31
17	T06, T10, T16, T31, T46	RidgeRegression	0.9819	3534.77	3403.14
18	T06, T10, T16, T31, T46	LinearRegressionWithIntercept	0.9819	3537.20	3403.19
19	T06, T10, T20, T31, T46	LinearRegressionWithIntercept	0.9811	3608.01	3425.67
20	T06, T10, T20, T31, T46	RidgeRegression	0.9811	3605.53	3425.72

Overall the best 20 models for 6 Clusters are listed in Table 4.17. KNN performed worse compared to Linear, LASSO and ridge regression. The combination, T05, T13, T21, T29, T42, T49, performed best. The most commonly used turbine was 49. 13 combinations, followed by 31 and 42 with 9 combinations.

Table 4.17 Best 20 Combination-Model Pairs for 6 Clusters

Rank	Combination	Model	R^2	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	LinearRegressionWithIntercept	0.9868	2959.61	2846.65
2	T05, T13, T21, T29, T42, T49	RidgeRegression	0.9868	2957.23	2846.67
3	T05, T13, T21, T29, T42, T49	LassoRegression	0.9868	2957.23	2846.67
4	T06, T08, T18, T31, T39, T49	LinearRegressionWithIntercept	0.9871	2932.50	2860.80
5	T06, T08, T18, T31, T39, T49	LassoRegression	0.9871	2930.15	2860.81
6	T06, T08, T18, T31, T39, T49	RidgeRegression	0.9871	2930.28	2862.38
7	T06, T08, T18, T31, T39, T49	KNNRegression	1	0	2869.27
8	T02, T09, T19, T31, T42, T47	LassoRegression	0.9870	2924.29	2871.83
9	T02, T09, T19, T31, T42, T47	LinearRegressionWithIntercept	0.9870	2926.64	2871.88
10	T02, T09, T19, T31, T42, T47	RidgeRegression	0.9870	2924.41	2872.30
11	T34, T10, T20, T28, T48, T43	KNNRegression	1	0	2894.01
12	T02, T09, T19, T31, T42, T47	KNNRegression	1	0	2895.47
13	T05, T13, T21, T29, T42, T49	KNNRegression	1	0	2939.67
14	T34, T10, T21, T26, T41, T49	KNNRegression	1	0	2941.45
15	T05, T12, T18, T31, T42, T47	KNNRegression	1	0	2951.16
16	T07, T09, T18, T30, T41, T49	KNNRegression	1	0	2957.35
17	T05, T11, T16, T29, T40, T49	LinearRegressionWithIntercept	0.9866	2983.72	2976.54
18	T05, T11, T16, T29, T40, T49	LassoRegression	0.9866	2981.32	2976.60
19	T05, T11, T16, T29, T40, T49	RidgeRegression	0.9866	2981.56	2976.62
20	T34, T10, T20, T28, T48, T43	LinearRegressionWithIntercept	0.9859	3067.01	2978.69

Overall the best 20 models for 7 Clusters are listed in table 4.18. The best combination is T07, T10, T14, T22, T31, T48, and T42; it has performed the best compared to every combination-model pair. For each model, except for GBM, this combina-

tion produced the lowest error scores. In 16 of the best possible combinations T31 was used.

Table 4.18 Best 20 Combination-Model Pairs for 7 Clusters

Rank	Combination	Model	R^2	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	KNNRegression	1	0	2348.75
2	T07, T10, T14, T22, T31, T48, T42	LassoRegression	0.9901	2495.79	2380.24
3	T07, T10, T14, T22, T31, T48, T42	LinearRegressionWithIntercept	0.9901	2498.08	2380.24
4	T07, T10, T14, T22, T31, T48, T42	RidgeRegression	0.9901	2495.89	2381.01
5	T06, T09, T17, T25, T31, T41, T49	LinearRegressionWithIntercept	0.9906	2436.21	2454.15
6	T06, T09, T17, T25, T31, T41, T49	LassoRegression	0.9906	2433.97	2454.16
7	T06, T09, T17, T25, T31, T41, T49	RidgeRegression	0.9906	2433.98	2454.25
8	T06, T09, T17, T25, T31, T41, T49	KNNRegression	1	0	2474.05
9	T06, T12, T19, T35, T33, T41, T51	LinearRegressionWithIntercept	0.9903	2476.40	2499.91
10	T06, T12, T19, T35, T33, T41, T51	LassoRegression	0.9903	2474.13	2499.92
11	T06, T12, T19, T35, T33, T41, T51	RidgeRegression	0.9903	2474.13	2499.92
12	T07, T10, T15, T20, T31, T41, T49	KNNRegression	1	0	2530.42
13	T04, T10, T19, T36, T33, T39, T49	KNNRegression	1	0	2533.44
14	T04, T09, T16, T20, T31, T48, T43	LinearRegressionWithIntercept	0.9899	2527.80	2535.04
15	T04, T09, T16, T20, T31, T48, T43	RidgeRegression	0.9899	2525.48	2535.08
16	T04, T09, T16, T20, T31, T48, T43	LassoRegression	0.9899	2525.48	2535.08
17	T03, T08, T19, T35, T31, T46, T43	KNNRegression	1	0	2537.78
18	T03, T08, T19, T35, T31, T46, T43	LassoRegression	0.9885	2696.64	2546.26
19	T03, T08, T19, T35, T31, T46, T43	LinearRegressionWithIntercept	0.9885	2699.12	2546.27
20	T03, T08, T19, T35, T31, T46, T43	RidgeRegression	0.9885	2696.90	2547.48

5. Conclusion

In this study, using an analytical approach, the total power production of a wind farm with 52 turbines was predicted using a subgroup of turbines from the farm. The approach investigated the best possible turbine combination and model pair. In order to decide upon the control group combinations, first, clustering methods were implemented to identify similar turbines. This similarity was not computed by the raw data; instead, it was computed using the pair-wise coefficient of determinations. After a set of combinations were formed, linear regression, lasso regression, ridge regression, KNN regression, and GBM regression algorithms were used to build the prediction model. The data used for predictions consists only of the power output obtained through SCADA systems. As expected, the prediction power of the models increased as more turbines were included. KNN regression was the best performing model in cases where fewer turbines were used. All of the four modeling algorithms, except GBM, performed well; the reason for the underperformance of GBM regression might be due to the lack of proper hyper tuning. The modeling power increased with as the number of control turbines increased, however the choice of how many control turbines to use is mainly financial.

In future directions of this study, a case study can be conducted to verify the accuracy of the proposed methods. The data from before and after the upgrade is required for this purpose. In future work, the data processing steps can be optimized for different sections of the power curve. To further extend the analysis, all of the combinations can be investigated, and different feature selection methods could be compared with the clustering method implemented in this thesis.

BIBLIOGRAPHY

Albers, A. (2014). Side-by-side testing to verify improvements of power curves. 6th Nordic Wind Power Conference.

BP (2019). Bp energy outlook 2019 edition. Technical report, BP.

Carlberg, M. (2015). Quantify change in wind turbine power performance using only scada data. Master's thesis, KTH School of Industrial Engineering and Management.

Evans, S. C., Zhang, Z., Iyengar, S., Chen, J., Hilton, J., Gregg, P., Eldridge, D., Jonkhof, M., McCulloch, C., & Shokoohi-Yekta, M. (2014). Towards wind farm performance optimization through empirical models. In *2014 IEEE Aerospace Conference*, (pp. 1–12).

Gill, S., Stephen, B., & Galloway, S. (2012). Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy*, *3*, 94–101.

Hwangbo, H., Ding, Y., Eisele, O., Weinzierl, G., Lang, U., & Pechlivanoglou, G. (2017). Quantifying the effect of vortex generator installation on wind power production: An academia-industry case study. *Renewable Energy*, *113*.

IEA (2019). World energy outlook 2019. Technical report, IEA, Paris.

IEA (2020). Renewables information: Overview (2020 edition). Technical report, International Energy Agency.

IPCC (2018). Summary for policymakers. in: Global warming of 1.5c. Technical report, International Energy Agency.

IRENA (2017). *CLIMATE POLICY DRIVES SHIFT TO RENEWABLE ENERGY*. Abu Dhabi.

IRENA (2019). *Global energy transformation: A roadmap to 2050* (2019 ed.). Abu Dhabi: International Renewable Energy Agency.

Krambeck, D. (2015). An introduction to scada systems - technical articles.

Kusiak, A. & Li, W. (2011). The prediction and diagnosis of wind turbine faults. *Renewable Energy*, *36*(1), 16 – 23.

Kusiak, A. & Verma, A. (2012). Monitoring wind farms with performance curves. *IEEE Transactions on Sustainable Energy*, *4*.

Lee, G., Ding, Y., Xie, L., & Genton, M. G. (2015). A kernel plus method for quantifying wind turbine performance upgrades. *Wind Energy*, *18*(7), 1207–1219.

Llombart, A., Pueyo, C., Fandos, J., & Guerrero, J. (2006). Robust data filtering in wind power systems. *European wind energy conference EWEC, Athens*, *2*, 149–54.

Qiggle (2017). Automatically detect power curtailment in wind turbines.

REN21 (2020). Renewables 2020 global status report. Technical report, Paris: REN21 Secretariat. ISBN 978-3-948393-00-7.

Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2019). World population growth.

Roy, S. (2015). Performance assessment of scada based wind turbine: Condition monitoring approaches. *International Journal of Electrical Power System and Technology*, *1*, 1–9.

Schlechtingen, M., Santos, I. F., & Achiche, S. (2013a). Using data-mining ap-

proaches for wind turbine power curve monitoring: A comparative study. *IEEE Transactions on Sustainable Energy*, *4*(3), 671–679.

Schlechtingen, M., Santos, I. F., & Achiche, S. (2013b). Wind turbine condition monitoring based on scada data using normal behavior models. part 1: System description. *Applied Soft Computing*, *13*(1), 259 – 270.

Shokrzadeh, S., Jafari Jozani, M., & Bibeau, E. (2014). Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy*, *5*, 1262–1269.

WWF (n.d.). Causes of global warming: Wwf-australia.

APPENDIX A

Table A.1 Clusters Used in Prediction Models

turbines	r2_kmean_5_st_wi	r2_kmean_6_st_wi	r2_kmean_7_st_wi
T01	4	1	2
T02	4	1	2
T03	4	1	2
T04	4	1	2
T05	4	1	2
T06	4	1	2
T07	4	1	2
T08	2	0	4
T09	2	0	4
T10	2	0	4
T11	2	0	4
T12	2	0	4
T13	2	0	4
T14	0	2	5
T15	0	2	5
T16	0	2	5
T17	0	2	5
T18	0	2	5
T19	0	2	5
T20	0	2	0
T21	0	2	0
T22	0	2	0
T23	3	4	3
T24	3	4	3
T25	0	2	0
T26	3	4	3
T27	3	4	3
T28	3	4	3
T29	3	4	3
T30	3	4	3
T31	3	4	3
T32	3	4	3
T33	3	4	3
T34	4	1	0
T35	0	2	0
T36	0	2	0
T37	3	4	3
T38	1	5	6
T39	1	5	6
T40	1	5	6
T41	1	5	6
T42	1	5	1
T43	1	3	1
T44	1	3	1
T45	1	5	6
T46	1	5	6
T47	1	3	1
T48	1	5	6
T49	1	3	1
T50	1	3	1
T51	1	3	1
T52	1	3	1

Figure A.1 R^2 Matrix for STRS Obtained Using Linear Regression with Intercept

APPENDIX B

Table B.1 Best 40 Results from Linear Regression with 5 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T11, T19, T33, T45	0.983775426743524	3348.64689154179	3351.21884903854
2	T06, T12, T20, T31, T45	0.983333630524103	3404.05426774627	3385.2022947781
3	T06, T13, T16, T32, T45	0.982194958619643	3516.15549272218	3385.74301834057
4	T06, T10, T16, T31, T46	0.981957724419656	3537.20681433545	3403.19931348721
5	T06, T10, T20, T31, T46	0.981197774198585	3608.01234485213	3425.67360550792
6	T05, T10, T18, T31, T42	0.982859285959416	3443.22574238466	3451.4629373588
7	T03, T09, T21, T33, T46	0.981767010989142	3554.86941655944	3469.00464435887
8	T06, T12, T18, T29, T46	0.981238698007404	3622.03494866195	3477.41474921032
9	T05, T08, T15, T31, T46	0.980121273803261	3719.87951088908	3488.91173450515
10	T06, T10, T21, T31, T41	0.982511188275425	3475.05559137855	3496.69982351351
11	T34, T10, T20, T31, T46	0.980006706412567	3715.96734901569	3510.58628551931
12	T05, T13, T21, T31, T41	0.982031537891458	3524.98390460693	3539.77828483413
13	T34, T08, T19, T32, T41	0.981029228525627	3617.11369111184	3543.12660260619
14	T34, T11, T17, T28, T42	0.980743552489598	3660.61875177418	3575.645182138
15	T02, T09, T19, T31, T42	0.980674350846566	3655.62779986661	3595.12672941552
16	T04, T11, T36, T31, T47	0.980369024387148	3693.90177764895	3600.79475955395
17	T05, T13, T21, T31, T44	0.980399129264833	3697.24809534838	3606.75909558932
18	T05, T10, T17, T33, T46	0.979526414758379	3766.06867384435	3606.99488719784
19	T02, T10, T22, T37, T46	0.980414097710529	3693.76247590812	3617.85718750502
20	T04, T08, T21, T29, T48	0.980555973562898	3684.84323466573	3618.82013411611
21	T05, T10, T18, T32, T45	0.982435144123006	3485.33837481105	3619.30198258685
22	T07, T12, T19, T29, T47	0.982068754197587	3540.9552151186	3628.7878343496
23	T06, T13, T19, T32, T44	0.978511734317071	3869.67686514252	3632.15501060931
24	T05, T13, T19, T30, T45	0.980080247492172	3708.09033377703	3632.24831334408
25	T04, T13, T21, T29, T45	0.981924735662094	3551.45731734393	3638.768175716
26	T02, T09, T18, T28, T50	0.978896178830792	3842.22630448091	3646.94203253619
27	T06, T13, T18, T32, T42	0.978727280819105	3840.42131263167	3650.2296563131
28	T03, T13, T22, T31, T48	0.979476278898008	3778.56711964842	3650.88201635799
29	T05, T08, T18, T32, T45	0.981573225668666	3574.80646644834	3655.87267580789
30	T05, T12, T19, T33, T48	0.980656191287148	3665.71226574605	3660.99498306503
31	T07, T13, T21, T29, T48	0.9804623843676	3688.43079271168	3662.88779203261
32	T05, T10, T20, T29, T46	0.979438417539798	3780.96746159141	3666.99263944722
33	T04, T11, T19, T33, T45	0.981142763172891	3614.3879671941	3667.12780683626
34	T06, T13, T17, T33, T47	0.979999541990458	3733.31306990081	3676.8309053877
35	T34, T12, T19, T33, T47	0.98095392694681	3639.69127013081	3678.15996411195
36	T06, T12, T35, T31, T45	0.980102341674647	3724.93759679425	3681.34845285617
37	T03, T08, T20, T31, T48	0.979654948951635	3760.36039754939	3685.27780128756
38	T34, T13, T19, T31, T47	0.978474369654269	3863.31742586949	3688.02172710718
39	T06, T11, T21, T33, T48	0.980173136268194	3705.76950946717	3692.17255341424
40	T03, T11, T36, T30, T47	0.978460709923681	3863.16502757461	3696.1408025248

Table B.2 Best 40 Results from Linear Regression with 6 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.986877950471716	2959.61135450363	2846.65199744711
2	T06, T08, T18, T31, T39, T49	0.987106267872599	2932.50702443898	2860.80405065031
3	T02, T09, T19, T31, T42, T47	0.987097663295829	2926.64149160848	2871.88938256176
4	T05, T11, T16, T29, T40, T49	0.986678991787743	2983.72121756493	2976.5464083196
5	T34, T10, T20, T28, T48, T43	0.985929225987141	3067.01388458895	2978.69862528979
6	T06, T09, T20, T33, T40, T49	0.985731234666261	3079.99444964954	2979.48728217678
7	T06, T12, T19, T30, T48, T43	0.987532571765993	2878.76252060273	2980.12114075548
8	T05, T10, T15, T31, T42, T47	0.985245252510769	3134.65453357803	2994.012798909
9	T34, T10, T21, T26, T41, T49	0.985879863027015	3076.71620028984	2995.34876550328
10	T07, T13, T21, T29, T38, T51	0.986051146111292	3055.02848883716	2996.12717100389
11	T02, T10, T22, T33, T40, T49	0.986270667671238	3026.44194432072	3001.41761708339
12	T05, T12, T18, T31, T42, T47	0.985851353418848	3072.1553766146	3009.96562113354
13	T03, T10, T21, T30, T42, T51	0.986339265978501	3006.87458211843	3015.12814884035
14	T05, T13, T20, T31, T41, T50	0.986371690474058	3005.35838711393	3019.38056147601
15	T34, T13, T20, T33, T42, T47	0.985785828665302	3073.99484769139	3019.89157268415
16	T05, T13, T21, T29, T38, T50	0.985279079864284	3137.32312793432	3025.71830330776
17	T07, T11, T16, T31, T42, T52	0.984891540420207	3167.39661172289	3032.1240992518
18	T04, T08, T18, T29, T38, T49	0.985538733079014	3116.81764540466	3039.95001015819
19	T07, T11, T18, T31, T40, T50	0.985553677411894	3095.82985103568	3044.39404907227
20	T01, T08, T20, T33, T40, T49	0.985791114873671	3084.93488225688	3071.86778192845
21	T04, T09, T21, T32, T48, T43	0.986290298811684	3017.97688720966	3072.78694788339
22	T04, T10, T22, T32, T41, T49	0.986091116436124	3038.51006881935	3074.35367722159
23	T02, T08, T19, T29, T42, T47	0.984513972900552	3220.07051985226	3074.61870804178
24	T07, T09, T17, T31, T38, T47	0.986253843203321	3023.66802017964	3076.30251772465
25	T01, T10, T18, T28, T41, T50	0.985035544007852	3167.08188937831	3077.9295902077
26	T06, T13, T22, T37, T42, T50	0.986030562740065	3053.75479427986	3084.41002107785
27	T01, T09, T18, T28, T41, T49	0.985343203239098	3135.04922726796	3086.89261116856
28	T34, T10, T21, T27, T40, T51	0.985423271478192	3128.09727581159	3094.23305316631
29	T05, T10, T21, T31, T39, T50	0.984722821130157	3183.05823269625	3103.86273843423
30	T03, T09, T20, T27, T42, T49	0.9856149132852	3110.31588255984	3107.25720409697
31	T06, T09, T17, T32, T42, T51	0.985791875549746	3068.6602603881	3108.59708345721
32	T34, T08, T20, T33, T41, T50	0.984303439807803	3225.40145112986	3111.01723597312
33	T03, T08, T20, T33, T40, T49	0.984970897697507	3168.21257162761	3119.76615262108
34	T06, T13, T22, T32, T48, T44	0.984936080512836	3174.45433835471	3125.87974770643
35	T05, T09, T19, T33, T40, T51	0.98466599059317	3189.48096464245	3126.29871923398
36	T06, T08, T14, T31, T41, T52	0.984998533033797	3161.21907845004	3128.17416482567
37	T06, T09, T18, T30, T39, T49	0.984820810944167	3171.54332557181	3129.93782290297
38	T06, T12, T22, T31, T48, T43	0.985593917583242	3103.66570173811	3130.62484341168
39	T07, T09, T14, T33, T42, T51	0.983879961828218	3272.84033260241	3134.52541769009
40	T01, T08, T18, T31, T42, T49	0.98592026713919	3067.68463804472	3134.8517958771

Table B.3 Best 40 Results from Linear Regression with 7 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	0.990119932390893	2498.08865481008	2380.24504817154
2	T06, T09, T17, T25, T31, T41, T49	0.99062614351861	2436.214449153	2454.15670730868
3	T06, T12, T19, T35, T33, T41, T51	0.990340167757141	2476.40981102807	2499.91028014921
4	T04, T09, T16, T20, T31, T48, T43	0.989918527979294	2527.80717941289	2535.04170524279
5	T03, T08, T19, T35, T31, T46, T43	0.988545609782939	2699.12105559202	2546.27862990747
6	T01, T08, T19, T35, T31, T40, T47	0.989384780808785	2603.39393220328	2561.53394003068
7	T04, T10, T19, T36, T33, T39, T49	0.98813522568956	2740.45736562221	2586.32244530041
8	T04, T11, T17, T35, T33, T46, T42	0.988541171859719	2694.3204823246	2588.25691947262
9	T05, T12, T18, T36, T33, T40, T47	0.988175518619338	2743.16637105804	2600.89207072429
10	T02, T09, T18, T36, T28, T40, T50	0.988941701624657	2655.79984112006	2603.53735486864
11	T07, T10, T15, T20, T31, T41, T49	0.988649322006064	2677.82984884493	2607.72372048961
12	T01, T09, T14, T21, T27, T41, T50	0.988198610609731	2751.16812358528	2610.71720250659
13	T05, T12, T18, T36, T29, T45, T43	0.988367248962694	2724.80784464362	2614.60412033097
14	T05, T13, T18, T21, T29, T48, T42	0.988441449255114	2707.73667160524	2630.08990362807
15	T07, T10, T17, T25, T33, T46, T42	0.987365707838817	2828.67852538192	2634.4186982365
16	T05, T13, T16, T21, T28, T48, T43	0.9884221738223	2720.9371920582	2650.24663413163
17	T06, T13, T18, T21, T29, T48, T42	0.988487695367282	2703.42725227444	2660.31404861799
18	T06, T12, T18, T36, T32, T45, T42	0.987469406016663	2815.54547477035	2665.54172851688
19	T06, T12, T16, T22, T31, T41, T49	0.988462485249387	2706.98452226834	2687.15980001568
20	T04, T12, T19, T34, T32, T39, T47	0.988525651055536	2707.82845985201	2688.30933483489
21	T06, T08, T19, T35, T31, T45, T51	0.986770668649028	2897.24204253167	2691.01184299452
22	T06, T11, T19, T35, T31, T46, T51	0.986371237673341	2936.42500119854	2691.8134130901
23	T07, T13, T19, T34, T31, T39, T50	0.988534385998506	2698.48103550285	2697.72690818272
24	T05, T12, T14, T22, T32, T46, T42	0.987775055761998	2784.6844876791	2701.06619776292
25	T06, T09, T19, T25, T31, T38, T52	0.987045696037944	2868.54988007944	2703.50004350573
26	T05, T08, T18, T25, T31, T46, T51	0.987370944760025	2835.31293468148	2704.25185375175
27	T06, T13, T18, T21, T32, T45, T43	0.987545683963115	2808.5841463451	2704.33808122506
28	T06, T13, T19, T22, T29, T38, T49	0.987289686828877	2847.18215490679	2711.25285505635
29	T07, T13, T16, T22, T30, T41, T50	0.98818460426173	2729.01457634602	2711.34584188559
30	T06, T12, T15, T35, T31, T46, T42	0.987399226462357	2833.4037612408	2711.70191248289
31	T06, T11, T17, T35, T33, T48, T43	0.988165041586083	2741.49710351279	2720.53711213297
32	T07, T10, T19, T35, T31, T38, T50	0.988430212932666	2705.49216759483	2722.28245733095
33	T02, T08, T17, T20, T31, T40, T50	0.987736394352353	2789.60618316226	2724.3977856189
34	T04, T12, T18, T36, T30, T41, T47	0.986685677855133	2903.23870570782	2725.33529066011
35	T05, T11, T19, T35, T33, T39, T44	0.98776221640588	2790.92070052365	2726.8137074002
36	T05, T13, T18, T22, T28, T41, T49	0.987966836253754	2770.50850339692	2727.21661269033
37	T05, T11, T17, T21, T33, T48, T42	0.987591010094148	2795.88408734244	2727.2532386837
38	T04, T11, T18, T36, T28, T41, T44	0.988038695604447	2765.93304388464	2734.41143080473
39	T01, T08, T14, T22, T28, T40, T52	0.988088313775369	2771.46297013987	2743.86349014619
40	T05, T09, T14, T21, T28, T38, T49	0.987579085882928	2816.41818543199	2774.17991623204

Table B.4 Best 40 Results from Lasso Regression with 5 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T11, T19, T33, T45	0.98377542477552	3346.34218641403	3351.2224333196
2	T06, T12, T20, T31, T45	0.983333628604741	3401.71141807698	3385.24003393227
3	T06, T13, T16, T32, T45	0.982194956700901	3513.73547607751	3385.76587554726
4	T06, T10, T16, T31, T46	0.981957722490566	3534.77230749248	3403.22383782041
5	T06, T10, T20, T31, T46	0.981197772258391	3605.52909889133	3425.72769341502
6	T05, T10, T18, T31, T42	0.982859283963259	3440.85593506568	3451.45800510984
7	T03, T09, T21, T33, T46	0.981767009071687	3552.42275018941	3469.02661492862
8	T06, T12, T18, T29, T46	0.981238696099793	3619.542048786	3477.46010993502
9	T05, T08, T15, T31, T46	0.980121271952208	3717.31925265458	3488.95167048451
10	T06, T10, T21, T31, T41	0.982511186299323	3472.6638710836	3496.67050276106
11	T34, T10, T20, T31, T46	0.980006704470115	3713.40979087759	3510.62672992807
12	T05, T13, T21, T31, T41	0.982031535949167	3522.55781233288	3539.75642387591
13	T34, T08, T19, T32, T41	0.981029226654511	3614.62417283214	3543.12800775509
14	T34, T11, T17, T28, T42	0.98074355056541	3658.09929302271	3575.66704829154
15	T02, T09, T19, T31, T42	0.980674348917925	3653.11177595672	3595.07727128892
16	T04, T11, T36, T31, T47	0.980369022479205	3691.35940646442	3600.79760995965
17	T05, T13, T21, T31, T44	0.980399127282801	3694.70342828	3606.75197868427
18	T05, T10, T17, T33, T46	0.979526412770681	3763.47663275133	3607.0693417917
19	T02, T10, T22, T37, T46	0.980414095827036	3691.22019870844	3617.85493206203
20	T04, T08, T21, T29, T48	0.980555971700286	3682.30709555774	3618.86220210601
21	T05, T10, T18, T32, T45	0.982435142186584	3482.93957256514	3619.32117532514
22	T07, T12, T19, T29, T47	0.982068752268486	3538.51812958549	3628.78212381733
23	T06, T13, T19, T32, T44	0.978511732303746	3867.01350793415	3632.17094363671
24	T05, T13, T19, T30, T45	0.980080245504624	3705.53820194623	3632.26685626441
25	T04, T13, T21, T29, T45	0.981924733755714	3549.01299991876	3638.75783306309
26	T02, T09, T18, T28, T50	0.978896176965408	3839.58183030929	3646.95677450659
27	T06, T13, T18, T32, T42	0.978727278839444	3837.77808973631	3650.22714283547
28	T03, T13, T22, T31, T48	0.979476276995363	3775.96646808108	3650.89768136448
29	T05, T08, T18, T32, T45	0.981573223827511	3572.34606887813	3655.90750981494
30	T05, T12, T19, T33, T48	0.980656189329577	3663.18930366519	3661.03539177493
31	T07, T13, T21, T29, T48	0.980462382464538	3685.89218739211	3662.89539190579
32	T05, T10, T20, T29, T46	0.97943841561301	3778.36515985093	3667.07307474514
33	T04, T11, T19, T33, T45	0.981142761199787	3611.90033576369	3667.13129704267
34	T06, T13, T17, T33, T47	0.979999540014799	3730.74357639889	3676.85738772104
35	T34, T12, T19, T33, T47	0.980953924998207	3637.18621927364	3678.15078346421
36	T06, T12, T35, T31, T45	0.980102339782017	3722.3738609868	3681.36415491976
37	T03, T08, T20, T31, T48	0.979654947083387	3757.77227536578	3685.33366024186
38	T34, T13, T19, T31, T47	0.978474174759227	3860.67574210875	3687.89991537909
39	T06, T11, T21, T33, T48	0.98017313430827	3703.21897325356	3692.20561572092
40	T03, T11, T36, T30, T47	0.978460708010181	3860.5061428519	3696.12861767341

Table B.5 Best 40 Results from Lasso Regression with 6 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.986877948431629	2957.23479521185	2846.67011416763
2	T06, T08, T18, T31, T39, T49	0.98710626591265	2930.15222477271	2860.81102604661
3	T02, T09, T19, T31, T42, T47	0.987097661285592	2924.29140750882	2871.83428083943
4	T05, T11, T16, T29, T40, T49	0.986678989784366	2981.32529055399	2976.60847213569
5	T34, T10, T20, T28, T48, T43	0.985929223882368	3064.55107236045	2978.77253421615
6	T06, T09, T20, T33, T40, T49	0.985731232648485	3077.52120145127	2979.55538239334
7	T06, T12, T19, T30, T48, T43	0.987532569673099	2876.45090052119	2980.14996560186
8	T05, T10, T15, T31, T42, T47	0.985245250454947	3132.13738985159	2994.0134787156
9	T34, T10, T21, T26, T41, T49	0.985879861070473	3074.24558016133	2995.35315095215
10	T07, T13, T21, T29, T38, T51	0.986051144089631	3052.57529378313	2996.10042152839
11	T02, T10, T22, T33, T40, T49	0.986270665583504	3024.01171510968	3001.41618505141
12	T05, T12, T18, T31, T42, T47	0.985851351370847	3069.68842834572	3009.99707699208
13	T03, T10, T21, T30, T42, T51	0.986339263845803	3004.46007156714	3015.11109069928
14	T05, T13, T20, T31, T41, T50	0.98637168842296	3002.94508562996	3019.41354490169
15	T34, T13, T20, T33, T42, T47	0.985785826581314	3071.52642518604	3019.8852831406
16	T05, T13, T21, T29, T38, T50	0.985279077837885	3134.80383868193	3025.72371142309
17	T07, T11, T16, T31, T42, T52	0.984891538384618	3164.85316865811	3032.11852679513
18	T04, T08, T18, T29, T38, T49	0.985538731121674	3114.31481858997	3039.97846127833
19	T07, T11, T18, T31, T40, T50	0.985553675392969	3093.34388437978	3044.39843453651
20	T01, T08, T20, T33, T40, T49	0.985791112891224	3082.45766395728	3071.93183712401
21	T04, T09, T21, T32, T48, T43	0.986290296763455	3015.55345141277	3072.80481264981
22	T04, T10, T22, T32, T41, T49	0.986091114385403	3036.07014189931	3074.36341603834
23	T02, T08, T19, T29, T42, T47	0.984513970948108	3217.48476532507	3074.61426539077
24	T07, T09, T17, T31, T38, T47	0.986253841221379	3021.24000651926	3076.27525695827
25	T01, T10, T18, T28, T41, T50	0.98503554199219	3164.53869898243	3077.93573985418
26	T06, T13, T22, T37, T42, T50	0.986030560722172	3051.3026212683	3084.40101482963
27	T01, T09, T18, T28, T41, T49	0.985343201280131	3132.53175770869	3086.90614766242
28	T34, T10, T21, T27, T40, T51	0.985423269450492	3125.58539724111	3094.24917449625
29	T05, T10, T21, T31, T39, T50	0.98472281907396	3180.50221301217	3103.86734046605
30	T03, T09, T20, T27, T42, T49	0.985614911299608	3107.81828087291	3107.30628368487
31	T06, T09, T17, T32, T42, T51	0.985791873548935	3066.19611268508	3108.61097506277
32	T34, T08, T20, T33, T41, T50	0.984303437817188	3222.81141699582	3111.07045446025
33	T03, T08, T20, T33, T40, T49	0.98497089571784	3165.66846857754	3119.83928667702
34	T06, T13, T22, T32, T48, T44	0.98493607845598	3171.9052307538	3125.91290299596
35	T05, T09, T19, T33, T40, T51	0.984665988559355	3186.91978433558	3126.35421240821
36	T06, T08, T14, T31, T41, T52	0.984998531038535	3158.68059326253	3128.19049260392
37	T06, T09, T18, T30, T39, T49	0.984820808952392	3168.99654709662	3129.93140879718
38	T06, T12, T22, T31, T48, T43	0.985593915516261	3101.17344863317	3130.68236339152
39	T07, T09, T14, T33, T42, T51	0.983879959799516	3270.21220292154	3134.54212779208
40	T01, T08, T18, T31, T42, T49	0.985920265171383	3065.22127214623	3134.85477312371

Table B.6 Best 40 Results from Lasso Regression with 7 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	0.990119930196205	2495.796055294	2380.24063066419
2	T06, T09, T17, T25, T31, T41, T49	0.990626141442234	2433.97863339901	2454.16799755984
3	T06, T12, T19, T35, T33, T41, T51	0.990340165591514	2474.13710962908	2499.92101868284
4	T04, T09, T16, T20, T31, T48, T43	0.989918525834343	2525.48729414142	2535.08776764661
5	T03, T08, T19, T35, T31, T46, T43	0.98854560766337	2696.64391030067	2546.26403829521
6	T01, T08, T19, T35, T31, T40, T47	0.989384778728698	2601.00465562206	2561.4963614164
7	T04, T10, T19, T36, T33, T39, T49	0.988135223443041	2737.94228940541	2586.32119770846
8	T04, T11, T17, T35, T33, T46, T42	0.98854095607924	2691.8728386388	2588.37964290728
9	T05, T12, T18, T36, T33, T40, T47	0.988175516461344	2740.64879925534	2600.95272366687
10	T02, T09, T18, T36, T28, T40, T50	0.988941699561394	2653.36245634377	2603.55083178576
11	T07, T10, T15, T20, T31, T41, T49	0.988649319827557	2675.37225297871	2607.72102817107
12	T01, T09, T14, T21, T27, T41, T50	0.988198608532992	2748.64319911242	2610.72227599967
13	T05, T12, T18, T36, T29, T45, T43	0.988367246768298	2722.30712993879	2614.68597316715
14	T05, T13, T18, T21, T29, T48, T42	0.988441447091696	2705.25162212207	2630.13090659896
15	T07, T10, T17, T25, T33, T46, T42	0.987365705600896	2826.08246624717	2634.44655290142
16	T05, T13, T16, T21, T28, T48, T43	0.988422171643652	2718.44002904402	2650.30219866966
17	T06, T13, T18, T21, T29, T48, T42	0.988487693224175	2700.94615643211	2660.32953421877
18	T06, T12, T18, T36, T32, T45, T42	0.987469403858923	2812.96146173775	2665.5772097706
19	T06, T12, T16, T22, T31, T41, T49	0.98846248313091	2704.50015827115	2687.16583813814
20	T04, T12, T19, T34, T32, T39, T47	0.98852544067134	2705.36787430514	2688.34869470338
21	T06, T08, T19, T35, T31, T45, T51	0.986770666578075	2894.5830284766	2691.01407311386
22	T06, T11, T19, T35, T31, T46, T51	0.986371235545543	2933.73002542248	2691.8173616083
23	T07, T13, T19, T34, T31, T39, T50	0.98853438384434	2696.00448141969	2697.7119387589
24	T05, T12, T14, T22, T32, T46, T42	0.987775053635173	2782.12880029467	2701.10398592417
25	T06, T09, T19, T25, T31, T38, T52	0.987045693897088	2865.91721144622	2703.49112406415
26	T05, T08, T18, T25, T31, T46, T51	0.987370942717455	2832.71076492048	2704.31044337782
27	T06, T13, T18, T21, T32, T45, T43	0.987545681720319	2806.00653324736	2704.35834074589
28	T06, T13, T19, T22, T29, T38, T49	0.987289684668223	2844.56910365527	2711.27047617394
29	T07, T13, T16, T22, T30, T41, T50	0.9881846021045	2726.5099926153	2711.30438504607
30	T06, T12, T15, T35, T31, T46, T42	0.987399224349077	2830.80335212153	2711.6936271154
31	T06, T11, T17, T35, T33, T48, T43	0.988165039393402	2738.98106749168	2720.57359528486
32	T07, T10, T19, T35, T31, T38, T50	0.988429991846466	2703.03475073853	2721.80598378603
33	T02, T08, T17, T20, T31, T40, T50	0.987736392265802	2787.0459734749	2724.42316169652
34	T04, T12, T18, T36, T30, T41, T47	0.986685675675967	2900.574198408	2725.33963790757
35	T05, T11, T19, T35, T33, T39, T44	0.98776221422585	2788.35929556498	2726.80109447088
36	T05, T13, T18, T22, T28, T41, T49	0.987966834139463	2767.96582863789	2727.24056448325
37	T05, T11, T17, T21, T33, T48, T42	0.987591007906666	2793.31812457502	2727.3135027904
38	T04, T11, T18, T36, T28, T41, T44	0.98803869351146	2763.39456731808	2734.39918225514
39	T01, T08, T14, T22, T28, T40, T52	0.98808831170431	2768.91941686102	2743.88373781974
40	T05, T09, T14, T21, T28, T38, T49	0.987579083821549	2813.83336267937	2774.18879301923

Table B.7 Best 40 Results from Ridge Regression with 5 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T11, T19, T33, T45	0.983775398685463	3346.34487696895	3351.28983113458
2	T06, T12, T20, T31, T45	0.983333612812505	3401.71302972404	3385.28180857076
3	T06, T13, T16, T32, T45	0.982192854254189	3513.94292358008	3385.87477674049
4	T06, T10, T16, T31, T46	0.981957686254712	3534.77585708392	3403.14842967377
5	T06, T10, T20, T31, T46	0.981197757120664	3605.53055030189	3425.72349676798
6	T05, T10, T18, T31, T42	0.982859285816781	3440.85574902617	3451.46965783592
7	T03, T09, T21, T33, T46	0.981766995593532	3552.42406319653	3469.13484886124
8	T06, T12, T18, T29, T46	0.981238663078114	3619.54523415375	3477.58237726868
9	T05, T08, T15, T31, T46	0.980119576374114	3717.47778570258	3489.9803137675
10	T06, T10, T21, T31, T41	0.982511172326633	3472.66525832589	3496.59366450885
11	T34, T10, T20, T31, T46	0.980006680757133	3713.41199301565	3510.65927458844
12	T05, T13, T21, T31, T41	0.982031537742265	3522.55763657243	3539.77801167115
13	T34, T08, T19, T32, T41	0.981029208357319	3614.62591597287	3543.11810298332
14	T34, T11, T17, T28, T42	0.980743504206332	3658.10369637888	3575.81230983185
15	T02, T09, T19, T31, T42	0.980674350624798	3653.11161463242	3595.13153304163
16	T04, T11, T36, T31, T47	0.980369012816271	3691.36031496115	3600.76148627498
17	T05, T13, T21, T31, T44	0.980398044899505	3694.80543981096	3606.36975541329
18	T05, T10, T17, T33, T46	0.979526386202862	3763.47907461293	3607.10266479462
19	T02, T10, T22, T37, T46	0.980414069372375	3691.22269157133	3617.85888280267
20	T04, T08, T21, T29, T48	0.980555973278925	3682.30694607646	3618.843209883
21	T05, T10, T18, T32, T45	0.982435124841377	3482.9412922571	3619.40757541022
22	T07, T12, T19, T29, T47	0.982068728190573	3538.52050532929	3628.99065802381
23	T05, T13, T19, T30, T45	0.980078354860013	3705.71404973601	3631.44291074965
24	T06, T13, T19, T32, T44	0.978510487198568	3867.12554043635	3631.8596958103
25	T04, T13, T21, T29, T45	0.981924709352698	3549.01539563981	3638.7979398068
26	T02, T09, T18, T28, T50	0.978896158520678	3839.58350820513	3646.99327532295
27	T06, T13, T18, T32, T42	0.978727267680411	3837.77909632792	3650.28122337865
28	T03, T13, T22, T31, T48	0.979476251312942	3775.96883061374	3650.88467938387
29	T05, T08, T18, T32, T45	0.981573201543151	3572.34822897987	3655.91406742363
30	T05, T12, T19, T33, T48	0.980656191062178	3663.18913961149	3661.03886479882
31	T07, T13, T21, T29, T48	0.980462356855939	3685.89460300152	3662.99068148344
32	T05, T10, T20, T29, T46	0.979438390203879	3778.36749442169	3667.02164484779
33	T04, T11, T19, T33, T45	0.981142745227842	3611.90186539004	3667.20249577329
34	T06, T13, T17, T33, T47	0.979999541830005	3730.74340710109	3676.83427095734
35	T34, T12, T19, T33, T47	0.980953926724548	3637.1860544359	3678.18883738978
36	T06, T12, T35, T31, T45	0.980102310496419	3722.37660030145	3681.29086635621
37	T03, T08, T20, T31, T48	0.979654934415009	3757.77344530309	3685.3979129503
38	T34, T13, T19, T31, T47	0.978474335470192	3860.66133025512	3687.99647048942
39	T06, T11, T21, T33, T48	0.980173113048632	3703.22095866762	3692.32012282019
40	T03, T11, T36, T30, T47	0.978459366086316	3860.62639807296	3695.64787630792

Table B.8 Best 40 Results from Ridge Regression with 6 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.986877929056018	2957.23697849129	2846.6700075866
2	T06, T08, T18, T31, T39, T49	0.987105098733192	2930.28484485272	2862.38182375294
3	T02, T09, T19, T31, T42, T47	0.987096566508429	2924.41546949522	2872.30664615156
4	T05, T11, T16, T29, T40, T49	0.986676817011421	2981.56842071848	2976.62468795645
5	T34, T10, T20, T28, T48, T43	0.985929197141884	3064.55398433693	2978.93388381953
6	T06, T12, T19, T30, T48, T43	0.987532555100509	2876.45258159446	2980.29940740418
7	T06, T09, T20, T33, T40, T49	0.985730277012439	3077.62425663744	2981.22282836313
8	T05, T10, T15, T31, T42, T47	0.985245239719331	3132.13852932935	2994.0182217878
9	T34, T10, T21, T26, T41, T49	0.98587984254466	3074.24759688651	2995.61089742024
10	T07, T13, T21, T29, T38, T51	0.986051126700662	3052.57719648826	2996.18589883163
11	T02, T10, T22, T33, T40, T49	0.986270654230664	3024.01296539288	3001.47802080667
12	T05, T12, T18, T31, T42, T47	0.985850260939543	3069.80671596281	3009.87939264138
13	T03, T10, T21, T30, T42, T51	0.986339254342109	3004.46111665959	3015.09869017459
14	T05, T13, T20, T31, T41, T50	0.986371678836717	3002.94614177534	3019.35782600608
15	T34, T13, T20, T33, T42, T47	0.985785801688714	3071.52911469339	3019.78036654921
16	T05, T13, T21, T29, T38, T50	0.985279056309512	3134.80613090253	3025.82234183792
17	T07, T11, T16, T31, T42, T52	0.984891525674817	3164.85449985408	3032.02352037117
18	T04, T08, T18, T29, T38, T49	0.985538713812457	3114.31668240758	3040.1899019403
19	T07, T11, T18, T31, T40, T50	0.98555205669673	3093.51718257132	3045.37492868536
20	T01, T08, T20, T33, T40, T49	0.985791099015857	3082.45916900907	3072.14768071409
21	T04, T09, T21, T32, T48, T43	0.986289401375627	3015.65192348478	3073.782233155
22	T04, T10, T22, T32, T41, T49	0.986091116315266	3036.06993127139	3074.35706577447
23	T02, T08, T19, T29, T42, T47	0.984513944482418	3217.48751467152	3074.72624774911
24	T07, T09, T17, T31, T38, T47	0.986253802003878	3021.24431628314	3076.46046409056
25	T01, T10, T18, T28, T41, T50	0.985035527322762	3164.54025005635	3077.9892473905
26	T06, T13, T22, T37, T42, T50	0.986030537983741	3051.30510461078	3084.37724899399
27	T01, T09, T18, T28, T41, T49	0.985343187214413	3132.53326080972	3086.92175836035
28	T34, T10, T21, T27, T40, T51	0.985423247358433	3125.58776576238	3094.48563413825
29	T05, T10, T21, T31, T39, T50	0.984721632172045	3180.6257590665	3103.70542390802
30	T03, T09, T20, T27, T42, T49	0.985614890088854	3107.82057210462	3107.29999956247
31	T06, T09, T17, T32, T42, T51	0.985791860845667	3066.19748340427	3108.70469591848
32	T34, T08, T20, T33, T41, T50	0.984303414994122	3222.813760006	3111.17538978335
33	T03, T08, T20, T33, T40, T49	0.984969524059876	3165.81292546642	3122.13532740138
34	T06, T13, T22, T32, T48, T44	0.98493482766839	3172.03691284383	3125.76307956817
35	T05, T09, T19, T33, T40, T51	0.984665975542736	3186.92113697943	3126.4026344067
36	T06, T08, T14, T31, T41, T52	0.9849985202809	3158.68172581589	3128.07921004895
37	T06, T09, T18, T30, T39, T49	0.984820801398111	3168.99733565937	3129.93544256749
38	T06, T12, T22, T31, T48, T43	0.985592709879509	3101.30321361076	3130.94883235432
39	T07, T09, T14, T33, T42, T51	0.983879961746944	3270.21200538788	3134.53250765147
40	T01, T08, T18, T31, T42, T49	0.985920252512706	3065.22265007124	3134.99635230345

Table B.9 Best 40 Results from Ridge Regression with 7 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	0.990119158420822	2495.89353215305	2381.01115014999
2	T06, T09, T17, T25, T31, T41, T49	0.990626123434577	2433.98097129601	2454.25437341733
3	T06, T12, T19, T35, T33, T41, T51	0.990340150855217	2474.1389968046	2499.92719490227
4	T04, T09, T16, T20, T31, T48, T43	0.989918515130061	2525.48863489388	2535.08508705004
5	T03, T08, T19, T35, T31, T46, T43	0.988543358579071	2696.90864201285	2547.48191560652
6	T01, T08, T19, T35, T31, T40, T47	0.989384762747575	2601.00661351628	2561.83084668735
7	T04, T10, T19, T36, T33, T39, T49	0.988135211046787	2737.9437196986	2586.41922371377
8	T04, T11, T17, T35, T33, T46, T42	0.988541149866988	2691.85007695876	2588.18694146145
9	T05, T12, T18, T36, T33, T40, T47	0.988174643998633	2740.7499059874	2602.78763439398
10	T02, T09, T18, T36, T28, T40, T50	0.9889416855934	2653.3641321049	2603.55921753951
11	T07, T10, T15, T20, T31, T41, T49	0.988649321801215	2675.37202038161	2607.70364413381
12	T01, T09, T14, T21, T27, T41, T50	0.988196717015714	2748.86346541775	2610.95050175607
13	T05, T12, T18, T36, T29, T45, T43	0.98836721483685	2722.31086624906	2614.39619926978
14	T05, T13, T18, T21, T29, T48, T42	0.988441449006298	2705.2513980681	2630.10409932575
15	T07, T10, T17, T25, T33, T46, T42	0.987365686062766	2826.08465142442	2634.56282964096
16	T05, T13, T16, T21, T28, T48, T43	0.98842215660026	2718.44179511556	2650.39018566107
17	T06, T13, T18, T21, T29, T48, T42	0.98848766063007	2700.94997994289	2660.31995750328
18	T06, T12, T18, T36, T32, T45, T42	0.98746785759208	2813.13501512719	2666.01221581682
19	T06, T12, T16, T22, T31, T41, T49	0.988462484835832	2704.49995844649	2687.13923645912
20	T04, T12, T19, T34, T32, T39, T47	0.98852563729421	2705.34469522438	2688.52191079719
21	T06, T08, T19, T35, T31, T45, T51	0.986770610458486	2894.58916796343	2691.09909859532
22	T06, T11, T19, T35, T31, T46, T51	0.986371194955435	2933.7343941357	2691.96904844716
23	T07, T13, T19, T34, T31, T39, T50	0.98853434525217	2696.00901866271	2697.60838954964
24	T05, T12, T14, T22, T32, T46, T42	0.987775044265843	2782.12986642118	2701.09942240744
25	T06, T09, T19, T25, T31, T38, T52	0.987044202444334	2866.08218582082	2702.57966846589
26	T05, T08, T18, T25, T31, T46, T51	0.987370916565756	2832.71369784564	2704.26408419714
27	T06, T13, T18, T21, T32, T45, T43	0.987545638660979	2806.0113839618	2704.33097666096
28	T06, T13, T19, T22, T29, T38, T49	0.987288649141147	2844.68497679671	2710.99534821991
29	T07, T13, T16, T22, T30, T41, T50	0.988184583299969	2726.51216227233	2711.27195229405
30	T06, T12, T15, T35, T31, T46, T42	0.987399209375371	2830.80503406575	2711.70117849665
31	T06, T11, T17, T35, T33, T48, T43	0.988164993666187	2738.98635884251	2720.68762257967
32	T07, T10, T19, T35, T31, T38, T50	0.988430174068888	2703.01346486426	2722.27941992062
33	T02, T08, T17, T20, T31, T40, T50	0.987736375355711	2787.04789498062	2724.35297286889
34	T04, T12, T18, T36, T30, T41, T47	0.986684589986468	2900.69245670919	2724.81065941858
35	T05, T11, T19, T35, T33, T39, T44	0.987760607495482	2788.54233498885	2727.07257793649
36	T05, T11, T17, T21, T33, T48, T42	0.987590997025066	2793.31934932249	2727.35154765869
37	T05, T13, T18, T22, T28, T41, T49	0.987965706014976	2768.09557574199	2728.00232884873
38	T04, T11, T18, T36, T28, T41, T44	0.988037302791042	2763.5552102023	2735.3742587444
39	T01, T08, T14, T22, T28, T40, T52	0.988086832947974	2769.09128292905	2744.52204495848
40	T05, T09, T14, T21, T28, T38, T49	0.987579066073767	2813.83537296923	2774.44383245099

Table B.10 Best 40 Results from KNN Regression with 5 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T34, T10, T20, T31, T46	1	0	3303.12879327609
2	T06, T10, T16, T31, T46	1	0	3328.70150993996
3	T06, T13, T16, T32, T45	1	0	3351.55465520954
4	T06, T12, T20, T31, T45	1	0	3354.3997229793
5	T07, T11, T19, T33, T45	1	0	3355.28693989007
6	T06, T10, T21, T31, T41	1	0	3355.95794928739
7	T05, T08, T15, T31, T46	1	0	3364.85247707672
8	T05, T10, T18, T31, T42	1	0	3365.18822590556
9	T06, T10, T20, T31, T46	1	0	3384.75323619769
10	T34, T08, T19, T32, T41	1	0	3399.31182642509
11	T05, T13, T21, T31, T41	1	0	3465.08930572952
12	T06, T12, T35, T31, T45	1	0	3483.56679047313
13	T34, T11, T17, T28, T42	1	0	3485.09883250582
14	T05, T13, T19, T30, T45	1	0	3493.65657709331
15	T34, T13, T19, T31, T47	1	0	3498.81320744141
16	T06, T12, T18, T29, T46	1	0	3504.12944176726
17	T04, T11, T36, T31, T47	1	0	3507.67688783042
18	T07, T12, T19, T29, T47	1	0	3515.28488846011
19	T06, T10, T16, T31, T46	0.987349872240692	2959.80582495996	3525.50058781132
20	T05, T10, T18, T32, T45	1	0	3532.45242770467
21	T06, T13, T18, T32, T42	1	0	3537.97514501076
22	T03, T09, T21, T33, T46	1	0	3541.07007846982
23	T34, T10, T20, T31, T46	0.987108378690504	2981.84081638576	3549.65284719683
24	T02, T09, T19, T31, T42	1	0	3553.09546861862
25	T05, T08, T18, T32, T45	1	0	3570.09106340123
26	T05, T08, T15, T31, T46	0.986945677653605	3012.40063868619	3571.96587635563
27	T34, T11, T17, T29, T42	1	0	3574.44455435537
28	T04, T13, T21, T29, T45	1	0	3585.02028122523
29	T05, T11, T25, T31, T46	1	0	3588.31881020069
30	T05, T13, T21, T31, T44	1	0	3591.87778746832
31	T03, T09, T21, T31, T42	1	0	3592.78506863719
32	T06, T10, T21, T31, T41	0.987627058427296	2920.91590864233	3597.32797202886
33	T04, T08, T21, T29, T48	1	0	3605.12553357198
34	T02, T09, T18, T28, T50	1	0	3606.105192413
35	T04, T13, T16, T31, T45	1	0	3606.50188118977
36	T05, T09, T18, T31, T51	1	0	3608.16650716166
37	T06, T09, T18, T31, T50	1	0	3611.75799525198
38	T04, T11, T19, T33, T45	1	0	3612.31659769891
39	T03, T08, T20, T31, T48	1	0	3614.97692432606
40	T03, T08, T19, T31, T45	1	0	3617.86353483604

Table B.11 Best 40 Results from KNN Regression with 6 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T06, T08, T18, T31, T39, T49	1	0	2869.27295612594
2	T34, T10, T20, T28, T48, T43	1	0	2894.01754297311
3	T02, T09, T19, T31, T42, T47	1	0	2895.4763337184
4	T05, T13, T21, T29, T42, T49	1	0	2939.67987094099
5	T34, T10, T21, T26, T41, T49	1	0	2941.45177746536
6	T05, T12, T18, T31, T42, T47	1	0	2951.16150147875
7	T07, T09, T18, T30, T41, T49	1	0	2957.35326490475
8	T03, T10, T21, T30, T42, T51	1	0	2990.09640092898
9	T05, T10, T15, T31, T42, T47	1	0	3000.2564672184
10	T34, T13, T20, T33, T42, T47	1	0	3008.89950023216
11	T05, T11, T16, T29, T40, T49	1	0	3008.91334531373
12	T05, T13, T20, T31, T41, T50	1	0	3013.12455799092
13	T05, T10, T21, T31, T39, T50	1	0	3018.67582575793
14	T06, T09, T18, T30, T39, T49	1	0	3025.54144479169
15	T06, T13, T19, T23, T46, T43	1	0	3030.89901586719
16	T07, T11, T16, T31, T42, T52	1	0	3037.82301021709
17	T01, T09, T18, T28, T41, T49	1	0	3038.372904301
18	T07, T13, T21, T29, T38, T51	1	0	3040.27199391728
19	T06, T08, T18, T31, T39, T49	0.990706087093889	2487.71215161727	3047.11436600308
20	T06, T08, T14, T31, T41, T52	1	0	3047.79355849091
21	T01, T08, T18, T31, T42, T49	1	0	3054.97093620159
22	T07, T11, T18, T31, T40, T50	1	0	3057.50261513794
23	T04, T10, T22, T32, T41, T49	1	0	3063.29709678901
24	T34, T10, T21, T27, T40, T51	1	0	3070.13039127212
25	T07, T12, T18, T31, T39, T49	1	0	3072.9862297377
26	T06, T12, T19, T30, T48, T43	1	0	3077.51901523686
27	T06, T08, T19, T32, T41, T52	1	0	3080.34054680194
28	T05, T13, T21, T29, T38, T50	1	0	3086.09929202465
29	T06, T09, T17, T32, T42, T51	1	0	3090.1626319769
30	T02, T10, T22, T33, T40, T49	1	0	3095.17460091902
31	T06, T09, T20, T33, T40, T49	1	0	3095.36618328639
32	T07, T09, T22, T29, T40, T49	1	0	3096.98264567734
33	T05, T13, T21, T29, T42, T49	0.990053452337239	2574.66732522776	3096.99676450322
34	T02, T09, T19, T31, T42, T47	0.991121562995378	2425.79830083988	3097.27298020023
35	T05, T11, T19, T32, T45, T49	1	0	3098.37589633947
36	T03, T10, T21, T30, T40, T50	1	0	3100.79166440837
37	T07, T09, T21, T28, T40, T51	1	0	3100.98137040376
38	T05, T11, T21, T31, T40, T50	1	0	3101.44554356247
39	T02, T08, T19, T29, T42, T47	1	0	3109.02303188063
40	T05, T13, T18, T32, T48, T44	1	0	3109.63034931403

Table B.12 Best 40 Results from KNN Regression with 7 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T07, T10, T14, T22, T31, T48, T42	1	0	2348.75376763343
2	T06, T09, T17, T25, T31, T41, T49	1	0	2474.05602613768
3	T07, T10, T15, T20, T31, T41, T49	1	0	2530.42844796111
4	T04, T10, T19, T36, T33, T39, T49	1	0	2533.44561728438
5	T03, T08, T19, T35, T31, T46, T43	1	0	2537.78843844101
6	T01, T08, T19, T35, T31, T40, T47	1	0	2548.13560849083
7	T06, T12, T19, T35, T33, T41, T51	1	0	2555.85862319113
8	T07, T10, T14, T22, T31, T48, T42	0.992776585386324	2134.02920797732	2561.40921431901
9	T06, T09, T17, T25, T31, T41, T49	0.993072792304169	2092.36124780184	2587.87010886749
10	T04, T09, T16, T20, T31, T48, T43	1	0	2608.83832235663
11	T05, T12, T18, T36, T29, T45, T43	1	0	2612.42575626989
12	T04, T11, T17, T35, T33, T46, T42	1	0	2612.98325894955
13	T02, T09, T18, T36, T28, T40, T50	1	0	2626.30175338294
14	T06, T12, T18, T36, T32, T45, T42	1	0	2633.5776990543
15	T04, T11, T18, T36, T28, T41, T44	1	0	2648.04512284816
16	T05, T12, T14, T22, T32, T46, T42	1	0	2648.10748000022
17	T06, T12, T15, T35, T31, T46, T42	1	0	2649.7307809853
18	T05, T12, T18, T36, T33, T40, T47	1	0	2655.17996921512
19	T07, T10, T19, T35, T31, T38, T50	1	0	2667.49657139121
20	T05, T13, T18, T21, T29, T48, T42	1	0	2668.67248864267
21	T05, T08, T18, T25, T31, T46, T51	1	0	2670.98213334606
22	T05, T13, T16, T21, T28, T48, T43	1	0	2673.52949556638
23	T06, T13, T18, T21, T29, T48, T42	1	0	2677.67439471817
24	T01, T09, T14, T21, T27, T41, T50	1	0	2680.544502142
25	T06, T11, T19, T35, T31, T46, T51	1	0	2684.70128940597
26	T04, T09, T17, T25, T32, T41, T49	1	0	2690.87361094256
27	T04, T12, T18, T36, T30, T41, T47	1	0	2694.33296797182
28	T07, T13, T19, T34, T31, T39, T50	1	0	2697.08651183084
29	T06, T08, T19, T35, T31, T45, T51	1	0	2703.14655379855
30	T07, T10, T15, T20, T31, T41, T49	0.992129344722311	2227.81293345544	2704.27673477748
31	T07, T13, T16, T22, T30, T41, T50	1	0	2705.76079352685
32	T02, T08, T17, T20, T31, T40, T50	1	0	2706.65286307656
33	T01, T08, T19, T35, T31, T40, T47	0.992222554616207	2226.358453182	2706.83730925865
34	T06, T12, T16, T22, T31, T41, T49	1	0	2708.36480263452
35	T07, T08, T16, T20, T29, T40, T47	1	0	2716.24876662692
36	T07, T10, T17, T25, T33, T46, T42	1	0	2719.85389351757
37	T05, T13, T18, T22, T28, T41, T49	1	0	2731.33904104448
38	T04, T10, T19, T36, T32, T45, T49	1	0	2732.48479958773
39	T04, T08, T19, T21, T32, T48, T42	1	0	2743.71475095755
40	T04, T13, T18, T36, T29, T48, T42	1	0	2746.38431804226

Table B.13 Best 40 Results from GBM Regression with 5 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T06, T10, T16, T31, T46	0.98673115837257	3031.32338576623	3516.88912278057
2	T05, T08, T15, T31, T46	0.986010188318224	3118.46948721366	3526.36713569918
3	T06, T10, T21, T31, T41	0.986398940478134	3062.44955461806	3551.16036697651
4	T05, T13, T21, T31, T41	0.986457865187827	3058.06152276692	3551.79582532805
5	T07, T11, T19, T33, T45	0.986992293271254	2996.29140983261	3559.078453061
6	T05, T10, T18, T31, T42	0.987404665140467	2949.55751094357	3579.34922594382
7	T34, T10, T20, T31, T46	0.985632014154965	3147.95583807518	3593.31287519322
8	T03, T09, T21, T33, T46	0.985651732061509	3151.34117380054	3619.01143993501
9	T06, T12, T20, T31, T45	0.986895236563537	3016.41806328475	3621.99302117605
10	T06, T10, T20, T31, T46	0.985640163035161	3150.93253930341	3630.11631005663
11	T06, T13, T16, T32, T45	0.986271206548779	3085.41603313729	3636.9054890866
12	T06, T12, T18, T29, T46	0.985069430703432	3228.94242233019	3647.85967887425
13	T07, T13, T21, T29, T48	0.984703694859719	3261.37017566648	3665.26509891869
14	T34, T08, T19, T32, T41	0.985818338048454	3125.24590936192	3666.56330119873
15	T05, T10, T18, T32, T45	0.986728581732964	3027.4870129671	3674.62530196057
16	T02, T09, T19, T31, T42	0.986230298402563	3083.60085912158	3684.65402503123
17	T05, T13, T19, T30, T45	0.985234305479387	3190.33467250443	3694.48755448457
18	T07, T12, T19, T29, T47	0.98661765084341	3056.90785662607	3696.72793527436
19	T34, T11, T17, T28, T42	0.986030218422225	3115.74703656517	3708.11277216678
20	T34, T13, T19, T31, T47	0.984383201323421	3288.36053614165	3708.9953654567
21	T34, T13, T21, T32, T42	0.984133808861137	3308.80146760812	3709.92106186509
22	T05, T08, T18, T32, T45	0.986145526160133	3097.5894927254	3737.57479262526
23	T05, T13, T21, T31, T44	0.984758046791447	3258.08335656878	3741.21770876629
24	T04, T11, T36, T31, T47	0.985411463063289	3182.15433221936	3741.91617250291
25	T06, T13, T18, T32, T42	0.984866763954829	3236.93693712012	3743.13568435644
26	T34, T11, T17, T29, T42	0.985995673083696	3114.64012251917	3747.3313532412
27	T02, T10, T22, T37, T46	0.985457706490162	3180.64056813979	3754.98145791332
28	T03, T09, T21, T31, T42	0.985174584404203	3200.58701013837	3757.64744958605
29	T05, T10, T20, T29, T46	0.984642267875978	3265.42101838331	3761.77053952947
30	T05, T10, T17, T33, T46	0.984825040175044	3240.08219244341	3772.43974190932
31	T07, T13, T22, T28, T48	0.984485363503241	3292.82080457332	3782.72294288441
32	T06, T12, T35, T31, T45	0.985139735389374	3216.86070782716	3783.00674545118
33	T04, T08, T21, T29, T48	0.984709255607494	3265.43615372982	3783.01508808574
34	T02, T10, T21, T33, T49	0.984638320303646	3265.13186946425	3783.04633066263
35	T05, T11, T25, T31, T46	0.984617582975467	3267.48819307617	3784.42470028273
36	T05, T12, T19, T33, T48	0.984637421708314	3264.52801995603	3791.91399013497
37	T07, T11, T18, T32, T46	0.983943080569694	3329.75621954828	3795.50591196807
38	T05, T13, T21, T37, T49	0.98280485855779	3458.78316799411	3796.56844033927
39	T03, T08, T14, T31, T47	0.985395267893685	3191.99878875465	3798.07278982994
40	T03, T08, T20, T31, T48	0.984394679248766	3291.07209225733	3804.19841608264

Table B.14 Best 40 Results from GBM Regression with 6 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T05, T13, T21, T29, T42, T49	0.989849672473116	2600.90785036468	3014.54676227493
2	T02, T09, T19, T31, T42, T47	0.990434144812022	2517.9571188981	3032.27592333231
3	T05, T11, T16, T29, T40, T49	0.990014402470218	2581.23658808156	3074.81379991802
4	T06, T08, T18, T31, T39, T49	0.990173456530398	2558.00399380849	3084.06093114932
5	T05, T10, T15, T31, T42, T47	0.989628381634716	2626.02071113019	3088.715631226
6	T07, T13, T21, T29, T38, T51	0.9888236431527	2732.41852314367	3115.37810572997
7	T34, T10, T21, T26, T41, T49	0.989367568788249	2667.69001498394	3118.41957781105
8	T07, T09, T17, T31, T38, T47	0.990347275967351	2531.74227729086	3128.73925677803
9	T34, T10, T21, T27, T40, T51	0.988989552391995	2716.46299717983	3136.33233479138
10	T06, T09, T17, T32, T42, T51	0.989578188776932	2626.05079947028	3142.72491775424
11	T06, T08, T17, T30, T41, T50	0.989584436944273	2623.38262866438	3146.4880696814
12	T34, T13, T20, T33, T42, T47	0.989658324637516	2619.92525601776	3164.87291055877
13	T06, T13, T22, T37, T42, T50	0.989519667074417	2642.91584392268	3187.83821481509
14	T34, T10, T20, T28, T48, T43	0.989306550254571	2671.57120217401	3188.89344729545
15	T05, T10, T21, T31, T39, T50	0.988759976402812	2728.08377466413	3189.0054108993
16	T01, T09, T19, T31, T38, T49	0.988744911250084	2736.81958783102	3189.83990447094
17	T06, T08, T14, T31, T41, T52	0.988591004426322	2754.62524992529	3190.06110800498
18	T05, T13, T21, T29, T38, T50	0.988521901151018	2768.07492997642	3194.87840258206
19	T05, T12, T18, T31, T42, T47	0.989173490137246	2685.22730267264	3196.14936699498
20	T07, T09, T21, T28, T40, T51	0.988769102252647	2738.17883872935	3207.10910046466
21	T06, T09, T20, T33, T40, T49	0.988731197206179	2734.93314416121	3209.24037614346
22	T07, T11, T16, T31, T42, T52	0.988827112356321	2721.61167298106	3213.59088205995
23	T06, T09, T18, T30, T39, T49	0.988895596621147	2710.47288616298	3214.45754322209
24	T03, T10, T21, T30, T42, T51	0.989348582697464	2652.97455710974	3223.72021214986
25	T34, T08, T20, T33, T41, T50	0.988011546618429	2816.52882630946	3231.15110517808
26	T02, T10, T22, T33, T40, T49	0.989074248364523	2697.64246492535	3231.27734863765
27	T04, T09, T20, T32, T41, T50	0.98892901144904	2704.59097981241	3241.95275757519
28	T06, T08, T19, T32, T41, T52	0.988202108550215	2795.27383414335	3244.74301260916
29	T06, T12, T19, T30, T48, T43	0.989996603447921	2576.57170359913	3252.2265398023
30	T05, T13, T20, T31, T41, T50	0.989250870625258	2666.93854584861	3260.8321597568
31	T01, T08, T20, T33, T40, T49	0.988951948995682	2718.06553579114	3266.75521428141
32	T04, T10, T22, T32, T41, T49	0.98912640843029	2684.43035499082	3273.35899977484
33	T01, T08, T18, T31, T42, T49	0.98992451876263	2592.97000133398	3273.57961023844
34	T07, T09, T18, T30, T41, T49	0.989291018923642	2655.79201705846	3275.89230863129
35	T07, T11, T18, T31, T40, T50	0.98885519223129	2716.97604416235	3278.69876848237
36	T04, T10, T18, T30, T40, T47	0.988404607262695	2773.6521183471	3280.90969738695
37	T03, T09, T20, T27, T42, T49	0.988859856704342	2734.92066959472	3282.0559943912
38	T06, T09, T25, T28, T41, T51	0.987958952730769	2839.82532650497	3282.16115385944
39	T34, T12, T20, T28, T40, T51	0.988266728637474	2804.52657176744	3285.09111604082
40	T04, T09, T21, T32, T48, T43	0.989449333455137	2645.40976598235	3293.21396328397

Table B.15 Best 40 Results from GBM Regression with 7 Clusters

Rank	Combination	R^2	RMSE Train	RMSE Test
1	T06, T09, T17, T25, T31, T41, T49	0.99290844627166	2117.03609254199	2644.48765501544
2	T07, T10, T14, T22, T31, T48, T42	0.99226417467034	2208.42378056974	2657.08192342875
3	T01, T08, T19, T35, T31, T40, T47	0.991749073640466	2293.12621593181	2707.49465009333
4	T04, T10, T19, T36, T33, T39, T49	0.990820979907756	2408.20064373519	2735.92219634606
5	T01, T09, T14, T21, T27, T41, T50	0.990679822659798	2442.66465042248	2789.6351491361
6	T04, T11, T17, T35, T33, T46, T42	0.991254005147574	2351.71645040918	2791.62461384353
7	T03, T08, T19, T35, T31, T46, T43	0.991380423669189	2339.27098815509	2800.82333237692
8	T06, T12, T19, T35, T33, T41, T51	0.991730344481904	2289.19417385024	2811.1826692468
9	T05, T13, T18, T21, T29, T48, T42	0.991030304754517	2383.11377680016	2820.37919501502
10	T07, T10, T15, T20, T31, T41, T49	0.991112892274094	2367.30108670237	2828.11087643331
11	T06, T13, T18, T21, T29, T48, T42	0.990905151407061	2400.67011608237	2837.3679484788
12	T04, T10, T19, T36, T32, T45, T49	0.989973229290557	2513.03977066464	2837.4215130537
13	T05, T12, T18, T36, T33, T40, T47	0.990636898247053	2438.77743162667	2854.59400823563
14	T06, T12, T15, T35, T31, T46, T42	0.990660164534313	2437.14110370992	2867.45752912882
15	T06, T12, T18, T36, T32, T45, T42	0.990416318349745	2460.05096923697	2872.67773360413
16	T06, T12, T16, T22, T31, T41, T49	0.991004828997282	2388.0067422752	2876.23737025514
17	T07, T10, T19, T35, T31, T38, T50	0.991150690536774	2363.95653986279	2876.48613381775
18	T05, T09, T14, T21, T28, T38, T49	0.990360757123331	2478.80875115675	2883.44087149106
19	T04, T09, T17, T25, T32, T41, T49	0.991469777765182	2322.62671505629	2887.41158809803
20	T04, T12, T18, T36, T30, T41, T47	0.989918131748993	2524.03238269345	2888.10171773337
21	T05, T11, T17, T21, T33, T48, T42	0.990509950497913	2442.79141087044	2890.27060079707
22	T06, T11, T19, T35, T31, T46, T51	0.989845195149376	2532.37311238497	2897.83376334967
23	T02, T09, T18, T36, T28, T40, T50	0.991537000708476	2321.21057040677	2899.06499010686
24	T05, T13, T16, T21, T28, T48, T43	0.991038485721441	2391.64603216144	2899.74547409467
25	T05, T12, T18, T36, T29, T45, T43	0.9910802357495	2383.81501512914	2905.8645272084
26	T06, T11, T17, T35, T33, T48, T43	0.991003694531084	2388.01867453431	2905.89332672747
27	T05, T08, T15, T34, T31, T41, T49	0.990936908794724	2400.95644101243	2910.0715329956
28	T04, T12, T19, T34, T32, T39, T47	0.991139313299614	2377.34390541689	2915.0377261902
29	T06, T09, T19, T25, T31, T38, T52	0.989922934014958	2527.68885951565	2916.46598009339
30	T01, T10, T15, T22, T29, T38, T47	0.989847384310272	2552.13848131636	2920.07501813004
31	T02, T08, T17, T20, T31, T40, T50	0.990948396713882	2394.40583968446	2920.65895063827
32	T06, T08, T19, T35, T31, T45, T51	0.990224506808446	2488.20620685746	2921.67622136433
33	T04, T09, T16, T20, T31, T48, T43	0.991860265726735	2269.28102006518	2930.37978751562
34	T05, T13, T18, T22, T28, T41, T49	0.99090353478765	2406.61768823198	2933.96065840086
35	T07, T13, T16, T22, T30, T41, T50	0.990462873404914	2449.58036199714	2937.06509605668
36	T07, T09, T17, T34, T31, T38, T47	0.991268209997141	2354.88021742125	2943.41975897353
37	T05, T11, T19, T35, T33, T39, T44	0.990139208862332	2502.9558614425	2947.23749762949
38	T07, T10, T17, T25, T33, T46, T42	0.99015983901307	2494.08032397802	2949.48408865808
39	T06, T09, T19, T21, T33, T46, T43	0.989633676829861	2556.28119216547	2950.33018269739
40	T04, T08, T19, T21, T32, T48, T42	0.990439778684052	2453.99003660402	2953.3597401603