

Modelling the Impact of Wind and EVs to the Power Grid in Orkney

by

Yifei Jing

An Honours report submitted for the Degree of:

BEng in Electrical and Electronic Engineering

Febrary 2020

Contents

Table of Contents	i
List of Figures	iv
List of Tables	vi
Synopsis	vii
Acknowledgements	viii
Statement of Authorship	ix
1 Introduction	1
1.1 Literature Review	1
1.1.1 Integration of wind power generation to the grid	1
1.1.2 The impacts of EV integration into power grid	2
1.2 Project objectives	4
2 Background	5
2.1 Challenges of building the power grid in Orkney	5
2.2 Establish a data set from available source	7
2.2.1 Data sources	7
2.2.2 Tools and implementations	9
2.2.3 Data bias and loss	11
2.3 Statistical analysis	12
2.3.1 <i>Covariance and Correlation</i>	12
2.3.2 Regression models	13
2.3.3 Binomial regression	14
2.4 Simulation frameworks	15
2.4.1 <i>Event-driven simulation</i>	15
2.4.2 <i>Analytic simulation</i>	18

2.5 Household loading profile	19
2.6 EV charging profile	20
3 Work	23
3.1 Analyzing Orkney over-all demand	23
3.1.1 Characterizing the attributes of demand	23
3.1.2 Constructing Weekday profile	24
3.1.3 Constructing Month profile	28
3.1.4 Analyzing the month profile of November	29
3.2 Analyzing Orkney Wind generation	29
3.2.1 Analyzing the power generation attributes	30
3.2.2 Analyzing the weather attributes	31
3.2.3 Analyzing the curtailment attributes	32
3.2.4 Linking wind power generation to wind speed and curtailment	32
3.3 Modeling the over-all EV effects	35
3.3.1 Characterizing EV charging events	35
3.3.2 Analyzing different penetration levels	37
3.4 Implementing the simulation model of the grid	39
3.4.1 Simulating Household load	40
3.4.2 Simulating EV charging load	41
3.4.3 Simulating wind power generation	42
3.4.4 Simulating grid storage	44
3.5 Determining the storage	44
3.5.1 Determining the scale of the wind farm	44
3.5.2 Determining the storage	46
3.6 Simulating the impacts of EV penetrations	48
3.6.1 Impacts of normal charging strategy	49
3.6.2 Impacts of delay charging strategy	50
3.6.3 Impacts of smart charging strategy	51
4 Conclusion	52
4.1 Impact of wind	52
4.2 Impact of EVs	53
4.3 Limitations and problems	55
Appendices	57
Bibliography	63

References	64
-------------------	-----------

List of Figures

2.1	The cumulative wind power generation capacity between 2000 and 2019	6
2.2	The EV number evolution from 2013 to 2019	6
2.3	The Orkney Live bar chart to represent the energy generation and demand of the island	8
2.4	ANM zones	9
2.5	An implementation of Half-adder	16
2.6	Structure of <i>Event-driven</i> simulation framework	17
2.7	Master controller of <i>Monte Carlo</i> simulation in <i>Trick</i>	19
2.8	Slave controller of <i>Monte Carlo</i> simulation in <i>Trick</i>	19
2.9	Average kWh/day over course of week for residential charging for an average EV	21
2.10	Weekly demand profile of an average EV for residential charging . .	22
2.11	Arrival and departure study of a day	22
3.1	The demand of mondays and their aggregated sum in one-day period, the differences of two outliers are specially selected	26
3.2	The plot of Aggregation sum and its mean of Monday	27
3.3	The aggregations of demand in 7 weekdays of November and the aggregation of the month	28
3.4	Seasonal effect of heating	29
3.5	The aggregated demand profile of November	30
3.6	The wind speed versus power generation on 12-02	33
3.7	The wind speed versus power generation on the whole data set . .	33
3.8	The Histogram of the samples generated by a Normal distribution of $\mu = 18, \sigma = 1$	36
3.9	The demand profile of 234 EVs under two charging mode	37
3.10	The Histogram of the samples generated by two normal distributions at 12:00 and 18:00	37

3.11	The demand profile of 234 EVs with two charging peak time	38
3.12	The Synthesized demand profile of 1.4% EV penetration	38
3.13	The demand profile under the 30% penetration of EVs	39
3.14	A one-day-interval simulation result of the household	41
3.15	Simulation results of three charging strategies	42
3.16	An illustration of the results of wind speed during a 5-day-interval ($5*24=140$ h)	44
3.17	The plot of the simulated total generation to demand under various scale factors. (Generated by <i>trick-dp</i>)	45
3.18	The scores and over-discharges in the range of storage from 15 MW h to 100 MW h	47
3.19	The performance of the storage under the normal charging strategy .	49
3.20	The performance of the storage under the delay charging strategy .	50
3.21	The performance of the storage under the smart charging strategy .	51
4.1	Daily average curve of all appliances	57
4.2	Daily average curve of refrigerator	57
4.3	Daily average curve of washing machine	58
4.4	Daily average curve of cooking	58
4.5	Daily average curve of computer	59
4.6	Daily average curve of audiovisual	59
4.7	Daily average curve of heating	60
4.8	Daily average curve of lighting	60
4.9	The wind speed with a typical increasing tendency	60
4.10	The wind speed with a typical decreasing tendency	61
4.11	The wind speed with a high level	61
4.12	The wind speed with a low level	61
4.13	The wind speed with a level near the global mean	62
4.14	An illustration of an extreme case on November 30th. The upper chart is the condition of curtailment, where no curtailment was hap- pened on this day. The bottom chart the subtraction of power gen- eration and demand.	62

List of Tables

2.1	Storage technologies and characteristics	7
2.2	Table of error types	15
2.3	Typical household electrical devices	20
2.4	Standard Electric Vehicle Charging Level	20
2.5	Typical EV Charging Time	21
2.6	Current condition of EVs in Orkney	22
3.1	Statistical properties of two demand attributes and their sum	24
3.2	The correlation matrix of Monday demand data	25
3.3	Mean and standard deviation of six Mondays	25
3.4	Correlation matrix of weekdays' aggregation on November Live de- mand	27
3.5	Statistical characteristics of Generation attributes	31
3.6	Wind speed statistical characteristics	31
3.7	Parameters of two linear regression model on wind speed to power generation on 12-02	34
3.8	Type errors of parameter k	35
3.9	The statistics of demand under three penetration levels	39
3.10	The implementation of the 5 coefficients of the p.f.	43
3.11	The average SoC of some high storage simulations	48
3.12	The average SoC of some high storage simulations under 40% pen- etration	49
4.1	The capacity of storage (MW h) at the highest score among simu- lations. The percent following the capacity is the ratio to the total daily consumption of one day without EV	54

Synopsis

In this project, the impacts of EVs to the power grid in Orkney, which has a large dependence on wind power, is analyzed from two perspectives: grid storage and energy consumption. The storage works as a compensation for wind power generation, as it is an intermittent power resource. For different EVs penetrations, a viable range of the storage is analyzed. The consumption has an increasing tendency resulted from electrification on the island, which affects the relationship of electricity generation and demand. Additionally, some developed charging strategies which aims to increase robustness of grid are implemented to estimate their effects on storage and consumption. The work is associated as: (1) Analysis collected data on total renewable power generation and demand to get relationships of loading profile, renewable generation and curtailments. (2) Design a small model of grid to simulate various EV penetrations, charging strategies, and curtailments.

In work part (1), a data set is established to collect power and weather data from available data sources. Then, evaluations and analysis are performed on this data set. In work part (2), a model of grid is designed with *C++* programming language base on a general simulation environment. To find optimized solution to the problems, *Monte Carlo* simulation is used.

Acknowledgements

- Thanks to my parents, Xing Li and Jianlin Jing, who supported me all the time.
- Thanks to my supervisor, Professor David Flynn, who gave me the chance to take part in this project.
- Thanks to my tutor, Doctor Wenshuo Tang, who gave me advice.

Statement of Authorship

I, Yifei Jing

State that this work submitted for assessment is my own and expressed in my own words. Any uses made within it of works of other authors in any form (eg. Ideas, figures, text, tables) are properly acknowledged at their point of use. A list of the references employed is included.

Date: April 15th 2020

Chapter 1

Introduction

1.1 Literature Review

1.1.1 Integration of wind power generation to the grid

The increasing energy consumption against the ever-decreasing amount of energy resources like the oil and gas have led to growing concerns from the customers, governments, and researchers [1] [2]. Seeking for available solutions, renewable energy resources including photovoltaics (PV), wind, tidal, biomass and hydrogens have attracted more and more attention [3]. Such renewable energy resources not only act as the substitution of the traditional power resources, but also have better performance on efficiency. The traditional power like the fossil fuel has only about 30% overall efficiency from sources to end users, while the renewable energy resources have an about 70% overall generation to grid connection efficiency [4]. Having been providing 24% power demand in 2017, the renewable power is assumed to increase to almost 30% of power demand in 2023 [5]. Among these renewable power resources, wind power is a fast-growing renewable energy resource with a large number of wind turbines already installed in wind farms [6]. By the beginning of February 2020, the UK consists of 10,429 wind turbines with a total installed capacity of over 22 gigawatts: 13,575 megawatts of onshore capacity and 8,483 megawatts of offshore capacity [7]. The integration of wind power is especially high in Scotland with 8,423 megawatts of installed wind power capacity by December 2018, which contributes over 50% of Scotland's electricity generation [8] [9]. For those places with a typical strong windy climate but lack of fossil and gas production, it is viable to connect wind generation to the grid for the sake of diminishing the budget on importing those non-renewable energy resources and on the other hand achieving a further target of de-carbonization.

However, the integration of wind power source into grid has led to challenges.

These challenges can be divided into two parts: storage and consumption.

The energy storage system (ESS) installed in an electrical grid can provide various ancillary services as it is summarized in [10]. For wind power generation, because of its intermittent nature, the storage should be capable of providing regulation and compensation for the grid [11] [12]. Generally, the storage stores energy by recharging when the embedded renewable generation (ERG) is high and the demand is relatively low, and it discharges when the demand is at peak value to avoid the cost of using expensive generators or imported energy. If sometimes, the ERG reaches a point that the grid cannot keep the voltage below the maximum to sustain its stability, the curtailment will happen, which causes the excess of output to be cut out by turning off some wind turbines. It is not possible to make profits at this moment.

On the other hand, when demand reaches the peak value of the day, but the ERG is at a low level, it is expected the ESS will be capable of providing enough energy, whose capacity and technologies attach an important role on the period and frequency of provision. For example, the super-capacitor is capable of charging and discharging frequently, but it does not generally have large capacitance, thus it can only support services on short duration, the batteries like Li-ion, Ni-MH have larger capacitance, but limited life cycles which prevent them from switching between charging and discharging states frequently. Because of the price issue, they are used for mid term duration and mostly on portable devices like mobile phones and EVs. For the grid, the storage used for frequency regulation and spinning reserve should be storage with smaller capacitance and longer cycle life, and for storage reservation, storage with large capacitance should be used. A comprehensive survey of battery types, characteristics, and manufactures is provided in [13].

1.1.2 The impacts of EV integration into power grid

Along with the increasing concerns of renewable energy sources, the electrification of transportation, such as Plug-in Hybrid Electric vehicles (PHEVs) and Plug-in Electric vehicles (PEV), has attracted more attention, because of economical and environmental reasons. The state-of-art architectures and energy management models are reviewed in [14]. An economical evaluation based in battery technologies, and the requirement of infrastructure of EVs is presented in [15]. If the EVs are connected to the grid, a bidirectional power transfer channel can be set up between the parked EV and the electrical network. The network can receive power from

EV at peak power, and the EV is recharged at a low load time. This concept is known as Vehicle-to-Grid (V2G), which is first proposed by Kempton and Ledendre in 1997 [16]. This concept implies that higher efficiency and stability can be achieved with V2G services. However, this concept is still limited by technical and economical issues [17], and the communication technologies are extremely important for V2G [18]. Thus, it is not possible to come into reality in the short term. The EVs then act as an additional load for the grid as a charging battery. If EVs are recharged at peak hours, additional generation would be required to meet this demand. The challenges on consumption derive mainly from the tendency of electrification as the EVs contribute a large amount of demand to the grid when being charged [19]. Generally, two charging modes are provided: 110V/15A as the normal mode and 240V/30A as the quick mode [20]. From the “Household Electricity Survey of UK” [21], the peak load value of one household is generally happens at night, at about 700W which is lower than half of the charging power. Thus the challenges for the grid is to satisfy new peak demand resulting from increasing penetration of EVs.

The instant power consumption of the charging outlets shown in [Table 2.4](#) implies that a lot of EVs charging at the same time will increase the peak demand and may cause the instability of the grid. Previously, many papers have focused on this scenario.

Three levels of penetration are considered in [19]: 35% in 2020, 51% in 2035, and 62% in 2050, along with two types of charging mode: normal charging, and quick charging according to Level 1 and Level 2 of [Table 2.4](#). The author uses a replica of power grid in two areas and tests impacts of three charging strategies. The model of EV charging is a presumption that 85% of the EVs are connected in the morning and 40% are connected in the peak time. The results show that the investment could increase up to 40% under off-peak charging strategy and the largest penetration level. However, the power source of the grid is not considered, as it is assumed that the main power source is the wind, an intermittent power resource. Nor the model of grid is too complicated to be recovered.

Another research also focuses on two charging modes without a replica of the grid [20]. Instead, the author builds a probability model to simulate the EV arrivals and departures. Three charging strategies are also considered. But the objective is on Time-of-use pricing.

A hybrid of these two researches is [22], where a replica of power grid and probability of arrival and departure model is used. Three penetration levels: 0%, 10%, and 20% are tested in this research. The objective is the demand profile of the grid. This

research should be a good start of this project. However, the probability model is not able to implemented from the formulas provided.

The impact of PHEV charging in residential distribution network is analyzed in [23]. Three levels of penetration are test on a replica of grid. A method based on dynamic programming is designed to minimize the cost function on the loss.

Various typical systems are tested with the penetration of EVs in [24]. The loss of life of the distribution transformers is calculated to estimate the upgrades needed for the system. A methodology of simulation is also provided, by defining a base case, the impacts of penetration of EVs can be estimated from energy source consumption and cost.

A large-scale PHEV charging infrastructure developed by FREEDM/ATEC is introduced in [25] to test for optimal power allocation. *Monte Carlo* method is used to simulate various scenarios at a municipal parking deck to analyze the performance of intelligent charging algorithm and find the optimal SoC of EV battery.

1.2 Project objectives

In this project, the models of power demand and generation in Orkney from an over-all landscape is first established based on available data collected from ANM [26]. Statistical features are extracted from these models and will be further discussed and compared with characteristics gathered from paper. Later, a small grid model is designed to simulate with different storage capacities. Problems on determining wind power generation and household demand models in this simulation model will also be mentioned. To evaluate the performance of the storage in the grid, a metric will be defined and improved.

In this project, the effects of EV charging will be analyzed first from an over-all perspective based on three levels of penetration. Then different penetrations of EVs will be analyzed in the simulation framework established as a smaller replica of Orkney to determine optimal storage for the grid with high dependence on wind power. After that, three charging strategies, which are summarized from [22], will be implemented in this simulation model. Their effects on the grid demand profile and storage will be discussed.

The results of this project should provide a good enough estimation of storage and demand in a power grid with a increasing penetration of EVs.

Chapter 2

Background

In this section, the information of the power grid in Orkney, as well as the concepts and techniques in data analysis using in this project are provided. Then, a brief description of the simulation framework used to build the model is presented.

2.1 Challenges of building the power grid in Orkney

Orkney is an island in the south of UK, where fossil fuels and natural gas are expensive owing to its geographic reasons, but the windy climate and tides imply large capacity of renewable energy generation. From this perspective, the community in Orkney is planning to eliminate the dependence on oil fuel and gas using completely clean electricity power resources like wind, hydrogen, and sunlight [27] [28].

To increase wind power generation capacity, an innovative Active Network Management (ANM) system was set up in 2009 to allow operators to connect wind farm generation to the grid [26]. Behold to ANM, the capacity of wind generation has been growing dramatically from 15.5 MW in 2009 to 47.7 MW in 2019 [29], which is depicted in [Figure 2.1](#).

To increase hydropower capacity, a project called *Surf'N'Turf* was launched to build hydrogen storage in Eday, which includes three tube trailers each capable of storing 250kg of hydrogen and one 500kg static hydrogen storage [30].

In the past few years, some transitions of electrification have been posed; like increasing the number of electrical vehicles (EVs) and ferries (EFs) to switch from fossil demand to electricity demand, improving the gas heating types of equipment to the electrical heating ones, and installing wind turbines with local storage for households to achieve energy trades among customers. Additionally, the number of EVs has been increasing drastically from a relatively small group to 234 in the first quarter of 2019 [29], as is indicated in [Figure 2.2](#). The rate of growth is 975%

2.1. CHALLENGES OF BUILDING THE POWER GRID IN ORKNEY

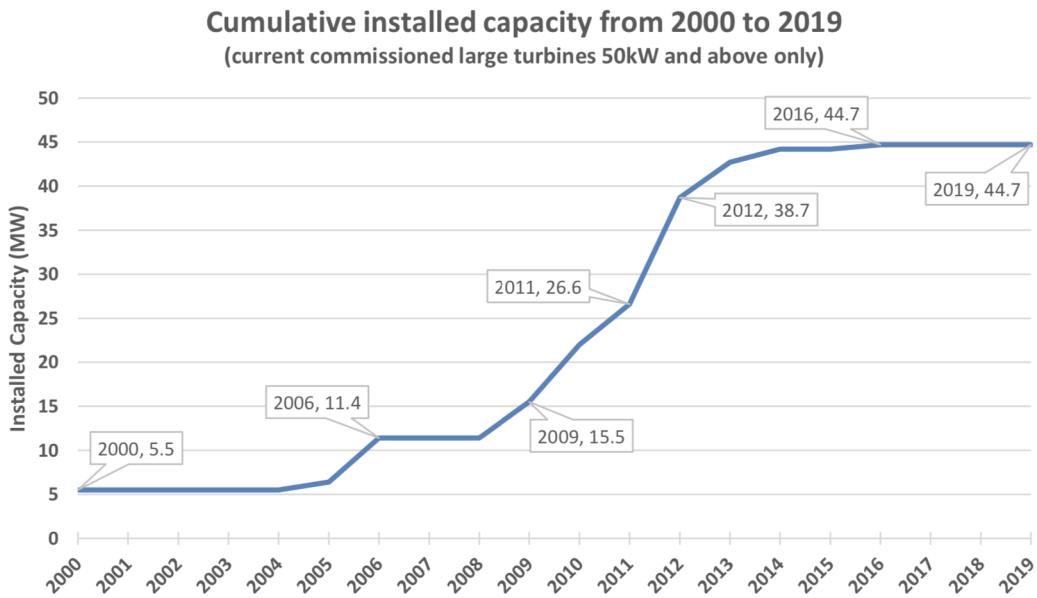


Figure 2.1: The cumulative wind power generation capacity between 2000 and 2019

in 5 years. The reason of this high grow speed is that on the one hand distributed wind turbines and local energy reserve storage can be installed in Orkney, which increase the accessibility of electric power with cheaper price, on the other hand, the community of Orkney provides charging stations which can charge EVs without fee (This service was turned to be paid charging scheme in 2019 April [29]).

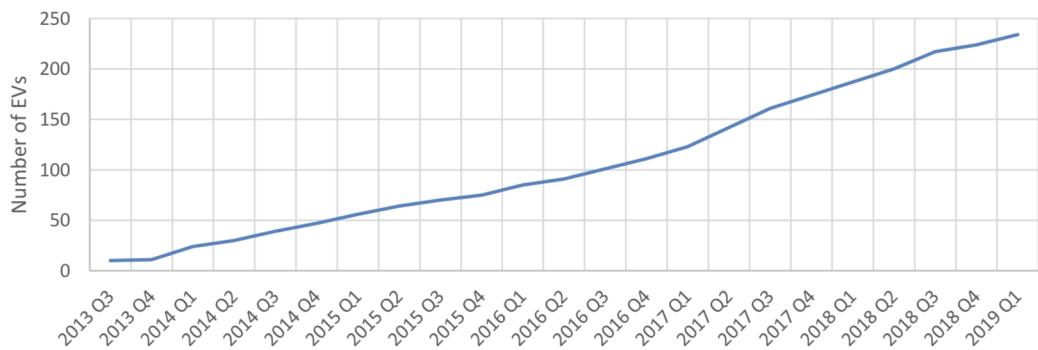


Figure 2.2: The EV number evolution from 2013 to 2019

To solve this problem, the over generated energy can be exported to the mainland of UK via a 20MW transmission line [31]. However, there is still 16% wind power generation having been curtailed in the last year [29], which implies that the current wind power capacity has been excessive.

Under the analysis of local power demand and generation, it is possible to estimate the amount of local storage in an appropriate range and determine its technique.

However, it is estimated that 99% of the total power generation in Orkney is from the wind [29], which is intermittent in nature. Its large variance leads to an unstable connection to the grid, which imposes large challenges to the design of the storage. The characteristics of typical storage technologies are summarized in [Table 2.1](#) [32] [33].

Type	Energy efficiency (%)	Energy density (Wh/kg)	Life cycle (cycles)	Services
Ni-Cd	60 - 90	40 - 60	500 - 2000	Wind power smoothing, spinning reserve
Li-Ion	70 - 85	100 - 200	500 - 2000	Wind power smoothing, spinning reserve, PV, portable devices
Ni-MH	50 - 80	60 - 80	< 3000	Wind power smoothing, spinning reserve, PV, portable devices
Pumped hydro	65 - 80	0.3	> 20 years	Daily load cycle, weather variations
Flywheel	95	5 - 30	> 20000	Wind power smoothing, spinning reserve
Super-capacitor	100	-	-	Voltage and frequency control, governor controlled generation

Table 2.1: Storage technologies and characteristics

2.2 Establish a data set from available source

A data set is constructed to acquire understanding of power generation and demand in Orkney. The data set began to scrape data from 19th October. This section describes the available data source used in this constructing the data set and the details on implementation.

2.2.1 Data sources

The data set contains data of three parts: generation and demand in Orkney, curtailment status in Orkney and the weather in Orkney.

2.2.1.1 Generation and demand data source

The ANM website [26] contains a bar chart to indicate the current power generation and demand of the whole island. Figure 2.3 shows this chart. There are 5 records: Live Demand, Winter Peak Demand, Orkney ANM, Non-ANM Renewable Generation, Total renewable capacity. Some explanations of them can be found in the left of the web page. However, the website itself does not expose APIs to get these data, thus it is necessary to analyze its http request and response and use a server to scrape the data periodically.

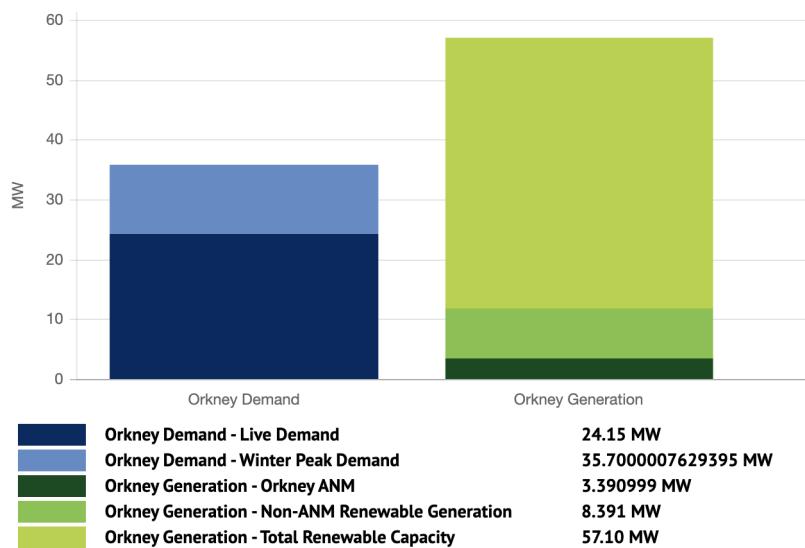


Figure 2.3: The Orkney Live bar chart to represent the energy generation and demand of the island

2.2.1.2 Curtailment data source

On the ANM status website [34], the island is divided into 9 zones as it is indicated in Figure 2.4. Then there is a table contains the current status of each zone. Unfortunately, there are neither public APIs to extract these data. The same method should be applied to scrape the data from this website.

2.2.1.3 Weather data source

The Open Weather Map website [35] provides public APIs to clients to get weather information of all over the world. Depending on the functions, some APIs which can request for historical data are paid services and the free ones can only acquire data at current time point. As the former two sources are both extracted from real time, it is not necessary to get historical data. The free APIs are used to get the

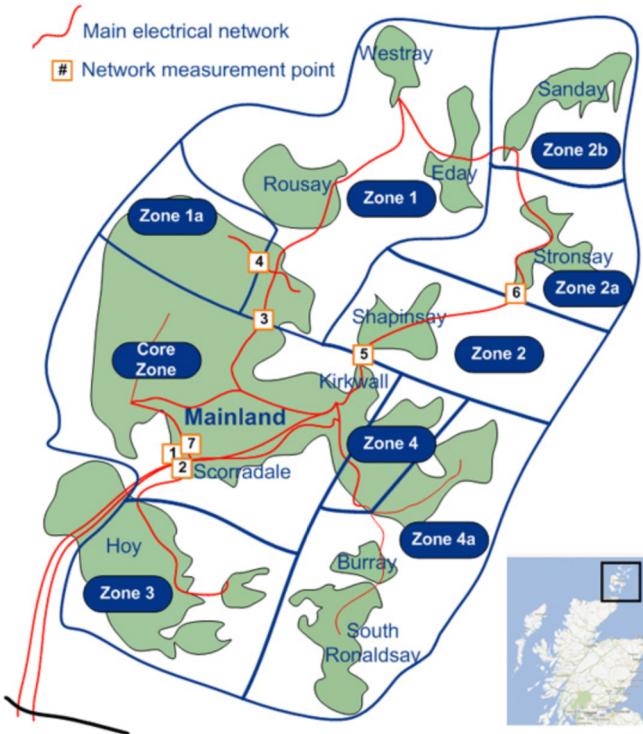


Figure 2.4: ANM zones

weather information in Orkney. Various kinds of weather information are available from this API, such as temperature, visibility, wind speed and degree, and clouds. As only the wind speed is directly related to wind power generation [20], it is selected along with the temperature and humidity to be recorded in the data set.

2.2.2 Tools and implementations

The data set is established on a cloud server with Cent OS 7 installed. As the data needs to be requested periodically all the time, a cloud server is the most efficient method. The scripts used in scrape the data are written in python 3 [36] and bash [37]. Libraries and tools based in these two languages are mentioned below.

2.2.2.1 Fetching Generation and demand data

By tracing the http response and request, the from of the generation and demand data are transformed in a JSON format file. Then, it is possible to use the Request python library [38] to imitate this http request and receive this file. After parsing the content, current power generation and demand data can be extracted and stored in the database built on the server. The store and program launch actions are controlled by crontab [39] to execute in a 10-minute period.

2.2.2.2 Fetching curtailment data

The form of curtailment data is not JSON as it is embedded in the html file fetched from the request. The web page contains some javascript programs which demonstrate information on the table. Because this file is a mixture of programming logics and data, it cannot be analyzed using a JSON parser as the generation and demand data did. The method is that, first save this html file as a text file to the disk, then use grep [40] to find the clauses which contain the status data. The status data is again saved as a file. Relaunch a python script to parse this simplified file and store the data of the table to the local database. The Status of the zone is represented using 1s and 0s, as it is assumed that the state of each zone can only be off and on. Actually, there is one more state called partly off which was not noticed when writing the script file, thus only two states are recorded.

2.2.2.3 Fetching weather data

As the APIs are provided, the implementation is that to write a python script using a library which wrap the API functions of Open Weather Map in python [41]. Use the verification code provided after registering a client of the website and find the region code of Orkney, then the weather condition can be fetched in JSON format. Apply the same procedure are the power generation and demand data to store the weather data in the database.

2.2.2.4 Synchronizing database to local

The database in the cloud server needs to be synchronized periodically to avoid possible loss due to abrupt. The method is to use github [42]. Create a git repository [43] and add the database file into it. Then, use crontab to execute a push command of git to synchronize the repository on the remote server of github. To access the repository on remote server, it is mandatory to enter login id and password of a github user. As it is using crontab to synchronize, the interaction of git push will cause problems. The solution is to use the pexpect library [44], which create a shell for each command needs interaction and interact with that command as the user defined. After solving this problem, the database can be updated without manually controlling extra steps. When it is needed for the user to get or update the data set, directly pull the repository from github, as it is always synchronized to the cloud server. The frequency of synchronization is set at 12:30 every day.

2.2.2.5 Data analyzing and visualizing

Various libraries are used. The Numpy and Scipy libraries [45] [46] are used to process statistical and vector-based computing. The Pandas library [47] is used to loading data set into memory and applying complex data set enquiries. The Scikit-Learn library [48] provides machine learning models and data preprocessing methods. The matplotlib library [49] provides plotting and rendering tools to produce plot. These libraries all provide APIs in python. Thus, python programming language [36] is used. Another reason is that the task of data analyzing is generally classified to fast development, as the cycle of editing code and fetching the results must be short. The programming language python [36], which is a dynamically typed and script-based language, is satisfying all demands. Further more, the browser based programmable Jupyter notebook under Jupyter Project [50] provides a flexible and user-friendly UI for programming using python. This makes visualizing data like plots and tables in an easier and more efficient way.

2.2.3 Data bias and loss

It is definite that the data accuracy is limited under various reasons. No data sources listed above has the guarantees or warranties that the data is accurate, complete and up-to-date. Since they are the only sources for Orkney, there are not other choices.

2.2.3.1 Sample loss

As it has been mentioned above, the sample rate is 10 minute. That is to say information between two neighboring sampling slots is out of observation. If some fierce disturbances happen at that interval, this information is lost because of low sampling rate. However, the ANM website does not provide the definition of the real-time data, nor data of each zones. Additionally, the data of the over-all island has been in aggregation pattern that cannot indicate spikes caused by individual household load or power generation. It is assumed that the information in each 10-minute interval is recoverable be observing its two neighboring samples.

2.2.3.2 Source stability

Under the assumed sample rate, there should be $6 * 24 = 144$ samples of each attribute each day. However, there are not always 144 samples in a one-day period. For example, the number of samples of the whole November should be $24 * 6 * 30 = 4320$, but the data set only contains 4252 samples. There are 68 10-minute samples

lost. Additionally, from iterating these samples day by day, there are only two days satisfying the assumption of 144 samples a day. Most of the days contain 141 or 142 samples. The reason is that the ANM website always runs into corruption: the website cannot respond to the requests and accessing through browser results in another page of maintenance. It seems that it will become blocked for a very short period each day from the samples. However, it is recorded that the website went to a longer corruption from 13th December to 19th December and 25th January to 29th January.

The methods of analyzing data of time series in this project are all designed under the presumption that there are no samples lost. The days without enough samples cannot be used sequentially. However, if only several days with enough samples are used, there will be only two days in November. There are just one or two samples missing for most of the days. Thus, it is possible to recover these lost samples from linear prediction of samples before and after the point. A method of fix these loss points is discussed in Data cleaning section. Under the condition that the sample of half or the whole day is lost, there is not possible ways to recover, thus, these days are excluded from analysis.

2.3 Statistical analysis

To analyze energy scenarios in Orkney, some statistical tools have been used in this project. This section summaries concepts and tools used in statistical analysis.

2.3.1 Covariance and Correlation

The *covariance* of two random variables X and Y is define as:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (2.1)$$

It gives a measurement of the degree of X and Y vary together. However, this numeric value is influenced by the magnitudes of X and Y . For example, $\text{Cov}(2X, Y) = 2\text{Cov}(X, Y)$. Thus, the definition of *Correlation* is introduced:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2.2)$$

It can be verified that $-1 \leq \rho(X, Y) \leq 1$ ¹. It is said that X and Y are *positively*

¹Cauchy-Schwarz Inequality

correlated is $\rho(X, Y) > 0$, that X and Y are *negatively correlated* if $\rho(X, Y) < 0$, and X and Y are *uncorrelated* if $\rho(X, Y) = 0$.

2.3.1.1 Correlation Matrix

In pragmatic analysis, the *Correlation Matrix* is a common method to measure the relationships among at least 2 attributes. Assume the number of attributes is N , the *Correlation Matrix* is calculation the *Correlation* between each pair of attributes. The *Correlation* between the i th and j th attributes, $\text{corr}(\text{attr}[i], \text{attr}[j])$, is located in the a_{ij} and a_{ji} entries. For example, Table 3.2 calculates the *Correlation Matrix* of the demand of Mondays to find their degree of variations. It can be figured out that the diagonal entries are always 1 as it is the *Correlation* is itself.

2.3.2 Regression models

The Regression method is used to quantize the relationship among attributes as the *Correlation matrix* merely provides the degree of relation. Parameters are derived from regression models in order to infer from known attributes to the attributes with strong relation. For example, the wind speed is related to the wind power generation and a regression model constructed to predict the wind power generation from the wind speed is shown in Section 3.2.4. The fundamental concepts are introduced in this section.

2.3.2.1 Linear Regression model

The linear regression model can be characterized with a set of parameters: Θ , which quantize the relation between the target y , and the attributes x :

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \quad (2.3)$$

In this formula, the target y can be seen as a weight sum of attributes x plus a bias θ_0 .

2.3.2.2 Gradient Descent

Gradient Descent uses the slop of the cost function to find the parameters of the model. By calculating derivation on the cost function ², the slops of each parameter tell the direction it grows fast. Thus, the negative of the slops can be chosen and added to the parameters to form a set of new parameters. Each projection of x to y can be used to compute a set of new parameters, then this operation can iteratively evaluate the whole set to reach the global minimum of the cost function.

2.3.2.3 Polynomial linear regression model

When the linear regression can not have better fitness with the power of 1 for the attributes, the polynomial linear regression is introduced, with can extend the power of each feature to n.

For example, for a vector $x = [a, b, c]$, if it is extended to degree $d = 2$, $x = [a, a^2, b, b^2, c, c^2, ab, ac, bc]$. The number of the attributes (n) will be extended to:

$$n_{poly} = \frac{(d + n)!}{d! n!} \quad (2.4)$$

When the degree of the polynomial linear regression is increased, the cost function is becoming smaller, as the new degree applies more limitation on the model to well fit in the training set.

There is also a trade-off between the degree and generalization of the model. As high degrees are easy to introduce over-fitting, but that does not mean low degree is good. When the cost function can not decrease to an acceptable range no matter how much data has been fed, it is time to increase the degree. Low degree always introduces under-fitting or high cost function.

2.3.3 Binomial regression

The Binomial regression model uses boolean value as the target value. It is used to implement classifier, for example, to judge whether a mail is a junk mail from the information of title, and to identify whether a curtailment has happened based on the wind generation power and demand in this project. Because the target attribute

²Normally, the *Mean Squared Error (MSE)* is used:

$$MSE(\theta) = \sum_{i=1}^m \frac{1}{m} (\theta^T x^{(i)} - y^{(i)})^2$$

which is also known as the *norm-2* metric emphasizing the *mean* of distribution of the error.

is not a numeric value, it is not appropriate to use metrics like *MSE* or *RMSE* to train and evaluate the model. Instead, the *Type I* and *Type II* errors are introduced.

2.3.3.1 Type I and Type II error

Type I error is also called false positive, which refers to true hypothesis that is rejected by the decision. *Type II* error is called false negative, which refers to false hypothesis not rejected by the decision. The concepts of these errors can be demonstrated in [Table 2.2 \[51\]](#).

	True	False
Don't reject	Correct inference (true positive) $possibility = 1 - \alpha$	Type II error (false positive) $possibility = \beta$
Reject	Type I error (true negative) $possibility = \alpha$	Correct inference (false negative) $possibility = 1 - \beta$

Table 2.2: Table of error types

It is necessary to notice that to eliminate one type error, the other type error must grow, but by selecting a low threshold, an optimal α can be achieved to increase the quality of judgement.

In the machine learning field, the *Confusion matrix* is implemented based on the error types. It has two rows and two columns which count the predictions under different error types or correct inferences.

2.4 Simulation frameworks

To simulate the power grid with high EV penetration and dependence on wind power generation, two kinds of simulation frameworks have been considered in this project: *event-driven simulation* and *analytic simulation*. Both of them are introduced in this section.

2.4.1 Event-driven simulation

This simulation framework is mostly used in circuit simulation software design. The basic idea is that the actions will trigger further events happen at a later time. Such events then also act as one action which can cause new events happen at a later time. This is a recursive procedure, which starts from an action and ends with all actions processed.

To elaborate, think of an example from Harold [\[52\]](#).

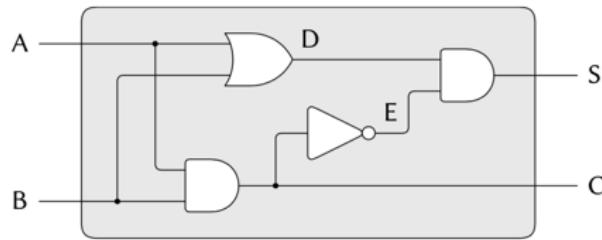


Figure 2.5: An implementation of Half-adder

Figure 2.5 shows a circuit of half adder linked using logical gates. Ports A and B are inputs representing two logical number to be added, S is the sum, and C is the carry. Its bool expression can be written as $S = (A|B) \& \overline{(A \& B)}$, $C = A \& B$. From both the expression, either a change of A or B will result in the changes on S and C . Assume $A = 0$ and $B = 0$, causing $S = 0$ and $C = 0$. At time $t = 0$, A is changed from 0 to 1. As A is linked to *OR-gate* D and *AND-gate* F, this will drive the output of D to 1 at a later time (an interval of *OR-gate delay*). After the delay, the output of D changes to 1, this change becomes an action to the *AND-gate* of S and triggers a change after a *AND-gate delay*. The wire S has a new value after the delay, and the simulation system stops as there is no new action triggered by S .

The key to this simulation scheme is a structure called *Schema*³, which stores the information of the current time, and the *agenda* which stores the actions that should be processed at a specified time. The simulation processing logic is to compare the current time and the specified time of the first *agenda item* in the *Schema*. If the two values are equal, process the actions in that *agenda item*. If they are not equal, update the current time by an increase. After processing the the actions in one *agenda item*, new actions may be added to be processed at a later time, and the current *agenda item* is deleted. After the deletion, the originally second *agenda item* becomes the first the *agenda item* and the procedure calls itself at the end to perform a tail iteration. The procedure then run to the next *agenda item* with the same logic till there are no *agenda items* in the *agenda table* which denotes the end of the simulation.

Apart from the trigger propagation strategy, a log system is a necessary part of simulation framework, as it is the interface to the users to check simulation details

³One implementation is a record of an integer value which represents the current time and a table of which the tuple is a record of an integer value which represents the specified time and a list of actions which are procedures on the wires. The wire can also be a record of a value and a list. The value is its current state: 0 or 1. The list is the actions added by each logical gates. So that when the value is changed by a start action or other triggered events, the actions are performed to propagate the effects and trigger further events to the gates ‘after it’.

and results. There are two available strategies: time-based log and event-based log. The differences can be understood from literal meaning that the time-based log records the state of each component on every time slot⁴ and the event-based log records the information of a state change of one component⁵. Generally, it is reluctantly to tell which is better, as the log can be converted or interpreted to the opposite form. It is the complexity of adding it to the specific simulation framework that causes the priority. For example, the lionized idea of the discussed digital circuit simulation system is based on *Event-driven* strategy. A event-based logger system can be implemented by adding a new action called *probe* which writes information of a state change to a log file whenever it is called. This new action is added to any wire to be logged and the details of log item can be changed by merely editing this action. However, if using the time-based log, the logger cannot be implemented as an action as the action is only called when it is triggered. It should be new structure which records each wire in the circuit. When the current time is updated in the *Schema*, the logger procedure is launched to read this structure and record the state of each wire. Obviously, there are cumbersome codes to write for this logger.

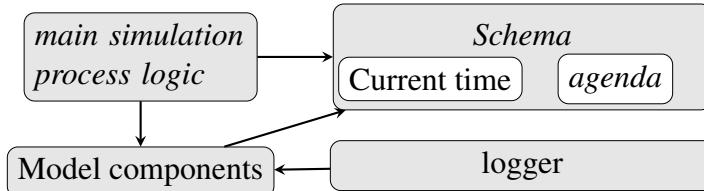


Figure 2.6: Structure of *Event-driven* simulation framework

Figure 2.6 shows the basic structure of a *Event-driven* simulation system. Fow a power grid simulation system, the model components could be wind power generator, grid, EV, house, appliance. The state should be the time, because the consumption is assumed to be a function of time, for example, the household demand profile in [21] is a function of time of the day, the season, the house type and house owner.

⁴Every time the current time in the *Schema* is updated, the logger will record the states of all components. Thus, the time-based logger structure is a timetable.

⁵It can be embedded with the *Event-driven* strategy as an action of the wire as a possible implementation. The structure is thus a diary.

2.4.2 Analytic simulation

Unlike the former strategy, this strategy is more generally applied, as its idea is easy to understand: analyze the model based on some variables⁶ for a specified amount of cycles. The core of this framework is the *scheduler*, which initializes the model, runs analytic functions for instructed number of cycles and terminates the model. It is like the *Schema* in previous framework, but all of the actions have been determined before the simulation begins, and no new actions will be added during simulation. The action in the *Analytic simulation* is called the *Scheduled job* which is provided by the model. Take the power grid as an example. Models of grid, wind generation, house, EV and other objects in one grid should be implemented first, so that they can be added to the scheduler to be initialized. These models should contain an analytic function to be called on each simulation cycle by the scheduler, a termination function to be called at the end of the simulation, to summarize the results, such as the average value, the score of special metrics.

In this project, a simulation environment based on *Analytic Simulation* called *trick* [53] is used for simplicity. The reason is that it provides the interface of implementing model in C/C++. Then the jobs can be scheduled using a script which includes the model and functions exposed to be called for initialization, analysis, and termination. Additionally, a versatile logger is provided, for example, it can automatically parse the source code of model and extract possible variables to be recorded. Because of this, the documentation also elaborates a special style of comment which can be understood by the interpreter to make sure the properties are rightly unified⁷, such as the unit, i/o, and type. This framework also provides *Monte Carlo* simulation method, which can repeatedly run copies of a simulation with different input values. By calculating values from previous *Monte Carlo* runs, optimization can be achieved. For the design of *Monte Carlo* simulation, Figure 2.7 depicts the master controller, and Figure 2.8 shows the slave controller.

The master controller for any given Monte Carlo simulation delegates run information to distributed slave instances. The master is responsible for spawning and managing the state of all slaves for a given simulation.

A Monte Carlo slave simulation is responsible for the execution of the runs delegated by the master controller. Should a simulation run fail, the slave will inform the master and continue running until explicitly killed or disconnected. Slaves consume

⁶Usually take the time as the variable, as most real-world models are based on time, or should be simulated by time.

⁷Because of this function, the data recording function can convert all the variables automatically to the same unit. This releases the need for users to convert from different units.

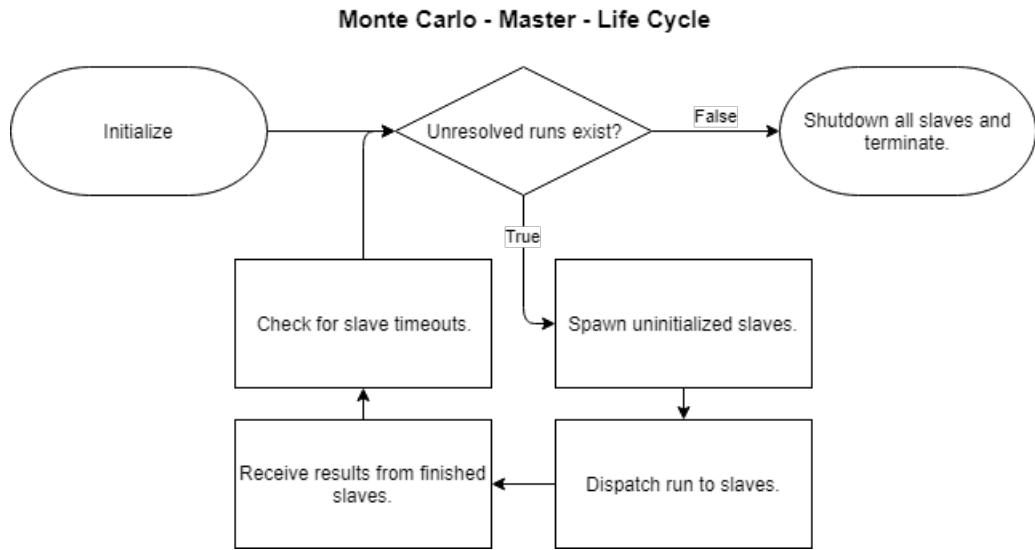


Figure 2.7: Master controller of *Monte Carlo* simulation in *Trick*

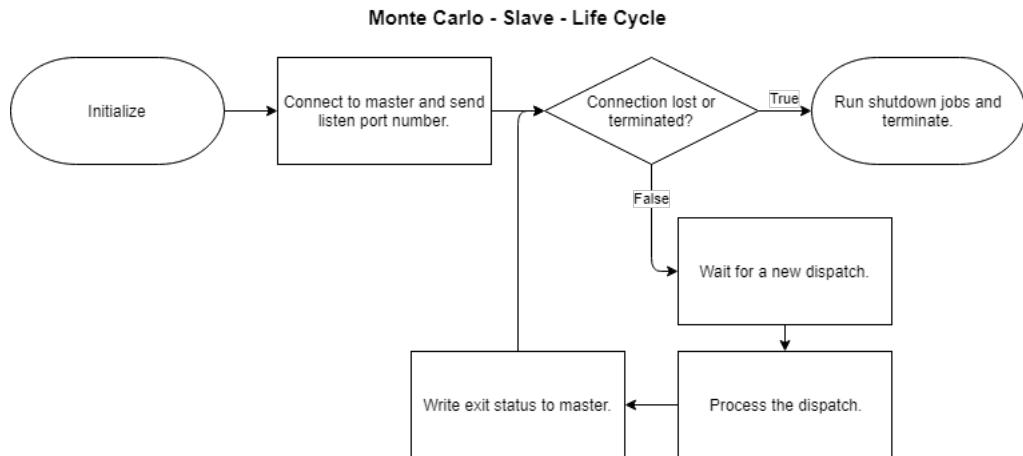


Figure 2.8: Slave controller of *Monte Carlo* simulation in *Trick*

only a single CPU and run only one job at a time. If you want to increase parallelism, you should create multiple slaves. Creating one slave per CPU is a reasonable approach.

2.5 Household loading profile

In Section 3.4.1, a daily household loading profile is needed to support simulation. The basic strategy is to first divide all electrical devices to three types based on their behaviors, and apply each appliance with the model of its type and general consumption data. A research has been delivered in [54], and the results can be

summarized in [Table 2.3](#).

	Even-smooth type	Stochastic behavior type	Fixed-program type
Device	Fridge freezer, Wine cabinet	TV-set, Total light, Electrical cooker, Microwave, Dish washer, Washing machine, Clothes dryer	Water heater, Air conditioner
Feature	Base load, Continuously working, Smooth load, Low power consumption	Randomly plug in, Randomly working period, Possible peak load, Medium power consumption	Randomly plug in, Fixed working period, High power consumption

Table 2.3: Typical household electrical devices

To model the Even-smooth type device, the consumption is generated periodically. To model the Stochastic behavior type and Fixed-program type, statistical algorithms are used to simulate the randomness. Typical consumption data is extracted from the UK household electricity survey [21].

2.6 EV charging profile

Three categories of typical voltage and power levels of EV charging have been derived from [55] and summarized in [Table 2.4](#). The Level 1 and Level 2 are mostly used in single-family residential and multifamily residential. Level 3 is for commercial situations. These three charging consumptions are defined by the Society of Automotive Engineers [56], and all U.S. EV makers must follow this standard.

	Voltage (VAC)	Current (Amps)	Power (kVA)	Phase
Level 1	120	12	1.44	Single
Level 2	208/240	32	6.7/7.7	Single
Level 3	208/480/600	100	16.8	Three

Table 2.4: Standard Electric Vehicle Charging Level

In addition, [Table 2.5](#) provides typical charging time for various sizes of EVs. Combine the information of two tables, the Level 1 charging is more suitable for charging overnight, and Level 2 could be the standard method of charging as it

EV Configuration	Battery Size (kWh)	120 V and 12 Amps	240 V and 32 Amps	480 V and 100 Amps
PHEV-10	4	3h 5m	35m	n/a
PHEV-20	8	6h 10m	1h 10m	n/a
PHEV-40	16	12h 20m	2h 20m	63m
BEV	24	18h 30m	3h 30m	1h 34m
PHEV Bus	50	n/a	5h 50m	3h 17m

Table 2.5: Typical EV Charging Time

support all five types. Level 3 is designed for public charging facilities for large types of EVs.

However, these profiles only estimate ideal conditions where the charging outlets are fully powered and the EVs have run out of battery capacity.

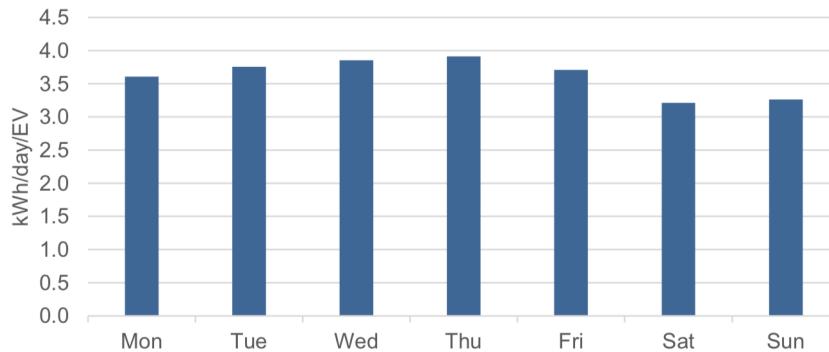


Figure 2.9: Average kWh/day over course of week for residential charging for an average EV

In the Electric Vehicle charging behavior study [57], the amount of demand of an average EV and the time of charging differs from different week days as it is demonstrated in Figure 2.9.

The amount is merely reaching 4.0 kWh for an EV, thus it should be acknowledged that if the EVs are recharging every day, they are not likely to cost as much as the time in the table. A daily trend can also be detected that there is a gradual increase between Monday to Thursday before a slight decrease on Friday and then a lower level on Saturday and Sunday. The distribution of charging demand among the period of a day is also studied in Figure 2.10. The profile indicates a demand peak on evening time with a maximum between 7-8pm on weekdays and 6-7pm on weekends. The demand profile spread broadly on weekends with a lower maximum. Additionally, this demand profile can be combined with the histogram of departure and arrival of a typical household in Figure 2.11. The departure behaviors can be projected to the low demand in the morning, and the arrivals in the evening

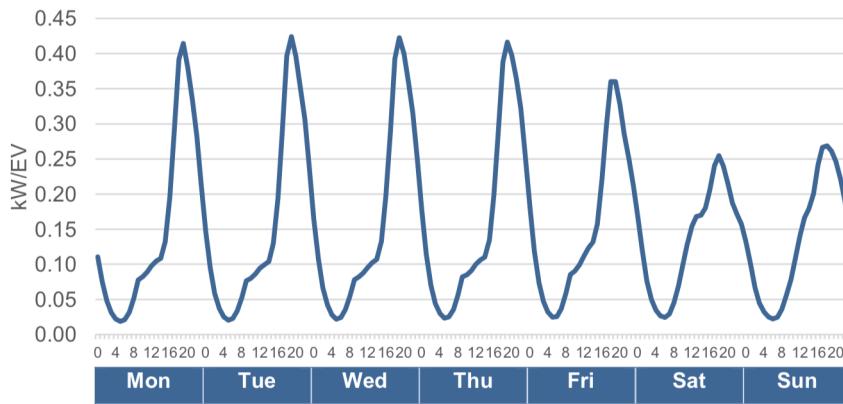


Figure 2.10: Weekly demand profile of an average EV for residential charging

indicates that people are recharging their EVs, thus this results in a peak demand in that morning.

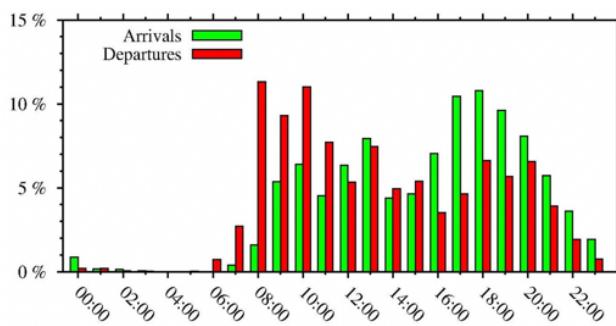


Figure 2.11: Arrival and departure study of a day

The statistics of EVs in Orkney is summarized in [Table 2.6](#), which is derived from [29]. The number of EVs in Orkney takes up to 1.4% of all vehicles on the island. This number is much lower than the predicted penetration of 35% in 2020 from [19]. It is expected that the number will grow dramatically in the following years, as the electrification of the island has brought large opportunity and merit to transform to EVs. Later, analysis on current EV penetration and higher penetration levels will be performed, and the impacts can be quite obvious under large penetration levels.

Average Miles Per Day	Average MWh Per Mile	Number of EVs Registered in Orkney	Estimated Annual Energy Consumed by Orkney EVs (MWh)
21.74	0.000299	234	555

Table 2.6: Current condition of EVs in Orkney

Chapter 3

Work

3.1 Analyzing Orkney over-all demand

Based on the data of over-all Orkney demand, a demand profile can be analyzed using techniques on aggregation and smoothing. This profile provides a bird view of daily demand pattern in Orkney. In this section, the details of analysis of demand data in Orkney is discussed.

3.1.1 Characterizing the attributes of demand

As it has been recorded in section 2.2.1.1, there are two demand attributes in the data set: Live demand, Winter peak demand. From literal sense, the Live demand could be the demand of current time and the Winter peak demand could be an estimation of current maximum demand based on historical data or learning models. There is no explanation on thw website of them.

Actually, the two attributes can be added to one constant, which is demonstrated in Table 3.1, where the mean, standard deviation, and various percentiles have been calculated. The sum attribute is calculated by adding the two demand attributes and it becomes a constant over all time.

The Winter peak demand attribute is taken out of consideration. One reason is that it can be calculated from the sum 35.7. The other reason is that the peak value is even lower than the Live demand. Curiously, the implication of why ANM chooses to use this attribute is impossible to understand; on the one hand, the live demand has been enough to represent the load of the grid on the island, on the other hand, the peak demand is not the peak demand in strict sense. The analysis on demand will thus only focus on the Live Demand.

	Live Demand	Winter Peak Demand	Sum
mean	19.04	16.66	35.7
std	2.92	2.92	0
min	10.57	8.05	35.7
25%	16.91	14.52	35.7
50%	18.9	16.8	35.7
75%	21.18	18.79	35.7
max	27.65	25.13	35.7

Table 3.1: Statistical properties of two demand attributes and their sum

3.1.2 Constructing Weekday profile

The first step of analyzing demand is to classify them into weekdays. The reason is that, intuitively, the demand is determined by individuals' behaviors, their behaviors have disciplines under the routine plans. For example, a white collar gets up at 7:00 in the morning, he opens the oven to cook food, then this will result in a consumption increase in the morning, he goes to company, turns on his computer and prepares coffee, this forms a consumption by computer and pot. At night, he watches TV, then takes a shower, this results in audiovisual and heating consumption. This description implies that specific vocation has generate demand under defined pattern. Generally, most of people work in the day and stay at home at night. They take meals and go to sleep at the same intervals so that these consumptions appear significantly in the over-all demand. The possible differences should be that the behaviors are different on various weekdays.

Take the demand of all mondays from the end of October and the whole November as the first group to be analyzed. If all these mondays are workdays or citizens' behaviors are enough disciplinary, large degree of correlation should be able to be acquired and the standard cross-correlation should tend to 1 to demonstrate large relation. That is to say, the demands of these days increase and decrease at the same time which appear to be similar patterns. If this is true, the assumption to build weekday profile is reasonable, and the aggregated demand trace can be calculated using the average value of these days. Then, profiles of other weekdays can be analyzed using the same method.

The extracted mondays are 10-22, 10-29, 11-05, 11-12, 11-19, 11-26. Using these six days, first calculate the their correlation matrix ¹. Table 3.2 records correlation matrix of the six days calculated using '*corcoef*' method of Numpy. From this matrix, two days are to be specially paid attention to: 10-22, 11-19. The 10-22

¹A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables.

3.1. ANALYZING ORKNEY OVER-ALL DEMAND

demand appears to be negatively correlated to other days except 11-19, and the 11-19 data is positively correlated with all other days under low degrees ². While other days can be characterized as largely similar from their high correlation values.

	10-22	10-29	11-05	11-12	11-19	11-26
10-22	1	-0.48	-0.59	-0.49	-0.35	-0.34
10-29	-0.48	1	0.77	0.88	0.30	0.63
11-05	-0.59	0.77	1	0.83	0.34	0.76
11-12	-0.49	0.88	0.83	1	0.22	0.61
11-19	-0.35	0.30	0.34	0.22	1	0.61
11-26	-0.34	0.63	0.76	0.61	0.61	1

Table 3.2: The correlation matrix of Monday demand data

The problem should be more obvious if the six days are plotted in one graph in one-day interval. Figure 3.1 depicts the data of these six days and also the average of their sum as the aggregated data. The four days characterized as largely similar above have most of the time point coinciding, and the two days defined as outliers have significant differences with the common pattern and are emphasized using red colored rectangle. The 10-22 data has a very low level in the afternoon, and inversely, the 11-19 data has a much higher consumption in the morning. Those are what effect the correlation scores. Though, these two days have shown different patterns of mondays, the aggregated sum is not largely influenced, as it is coinciding with other four days in the graph. Therefore, if most of the days of Monday adhere to the main pattern, then the outliers cannot interfere the aggregation.

There should be reasons to explain the abnormal pattern on 10-22 and 11-19, however, no valid news or information can be found through web on these two days in Orkney. There must be something happen, for example, a maintenance which results in a lot of households turning off the appliances or a festival which increase the demand. Because 10-22 has lower average demand and 11-19 has larger average demand, and both of the days have larger standard deviations. The details of means and standard deviations are shown in Table 3.3. It should be viable to infer the two special conditions.

	10-22	10-29	11-05	11-12	11-19	11-26	Aggregation
Mean	15.82	20.68	17.67	20.31	21.83	19.05	19.23
Standard deviation	2.31	2.18	2.18	2.09	2.86	2.43	1.5

Table 3.3: Mean and standard deviation of six Mondays

²From experience, the standard correlation value larger than 0.7 can be detected with large similarity

3.1. ANALYZING ORKNEY OVER-ALL DEMAND

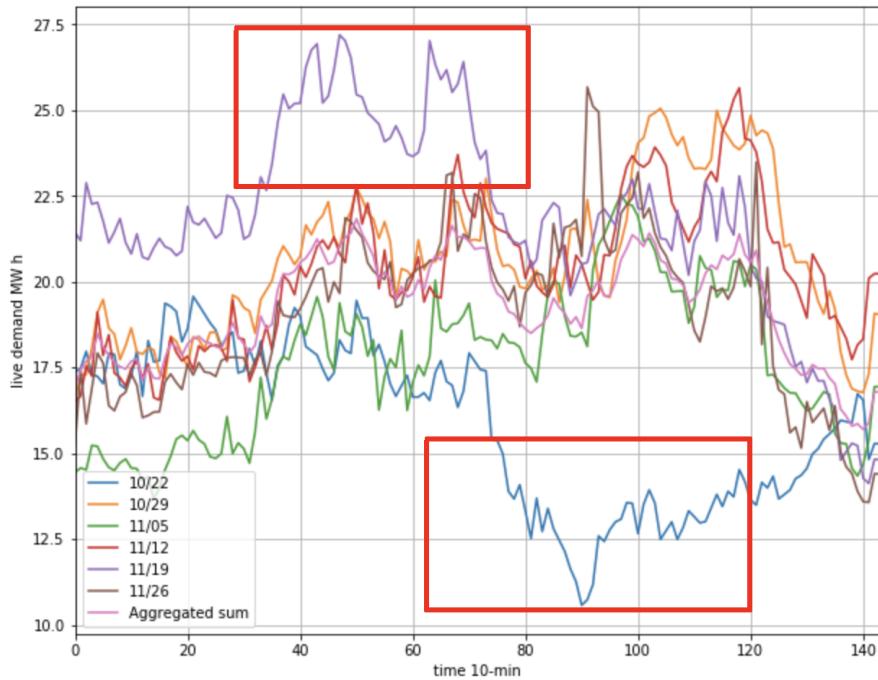


Figure 3.1: The demand of mondays and their aggregated sum in one-day period, the differences of two outliers are specially selected

From Table 3.1, the average demand and the standard deviation in a global scope are 19.04 MW h and 2.92, and the four normal days have their means near to the global value, while the two outliers have larger differences. The aggregation data also has a closer average, the deviation is lower adhere to *Variance*³. A more clean plot can be viewed in Figure 3.2, where only the aggregated data and the mean is plotted. From the plot, it can be identified that there are four peak demand periods at almost the same level on the Monday, and they demonstrate the people's behaviors in the morning, noon, afternoon and evening. If 7:00 am - 8:00 pm (42 - 120 10-minute) is defined as the diurnal interval, most of consumption in this period will be over the mean consumption. Then the compliment of this period, the nocturnal period, has energy consumption lower than the average in most of the time.

The next step is to apply this analysis method to each weekday. From the assumption and the experience of the Monday case, there will be some outliers in each weekday though, most weekdays appear a similar pattern and they are indicating the main tendency of demand. The date range is selected to be the whole November, as the data set was established at 19th October, and there are long term data loss in December and January 2.2.3.2. The data is first classified by seven weekdays, then

³For the sample average of i.i.d.: $S_n = (X_1 + \dots + X_n)/n$, the deviation $\sigma_n^2 = \sigma^2/n$

3.1. ANALYZING ORKNEY OVER-ALL DEMAND

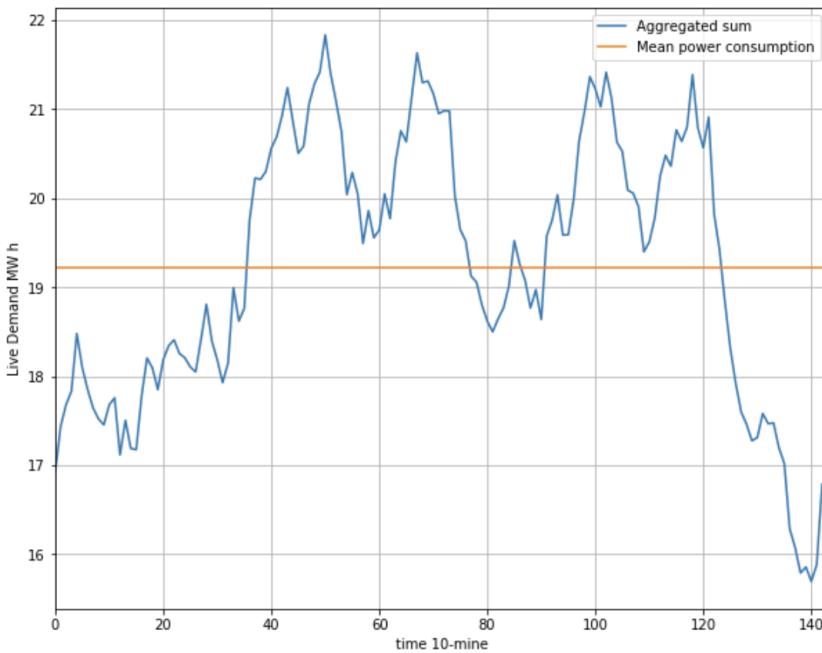


Figure 3.2: The plot of Aggregation sum and its mean of Monday

apply aggregation to each weekday group. Finally, the seven aggregation weekdays are plotted and analyzed to find their similar patterns.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Sunday	1	0.96	0.93	0.93	0.98	0.94	0.84
Monday	0.96	1	0.97	0.94	0.94	0.90	0.80
Tuesday	0.93	0.97	1	0.95	0.93	0.90	0.79
Wednesday	0.93	0.94	0.95	1	0.93	0.90	0.81
Thursday	0.98	0.94	0.93	0.93	1	0.92	0.81
Friday	0.94	0.90	0.90	0.90	0.92	1	0.95
Saturday	0.84	0.80	0.79	0.81	0.81	0.95	1

Table 3.4: Correlation matrix of weekdays' aggregation on November Live demand

Table 3.4 shows the correlation matrix of the seven aggregations. As most of them are over 0.9, the profile of each weekday appears to be similar in large degree and it implies that they can be aggregated for the second time to form a new aggregation of the month. What can be verified is that the demand pattern of a day in Orkney is always following a similar pattern, the value may at a degree though, the peak value and the lowest value can always occur at the same time. This result should be implicitly true before the experiment, as this is the overall demand data, which records the total electricity demand of the island to represents the behavior of most people in common.

Figure 3.3 shows a plot of the seven aggregation data to visualize their similarity. A

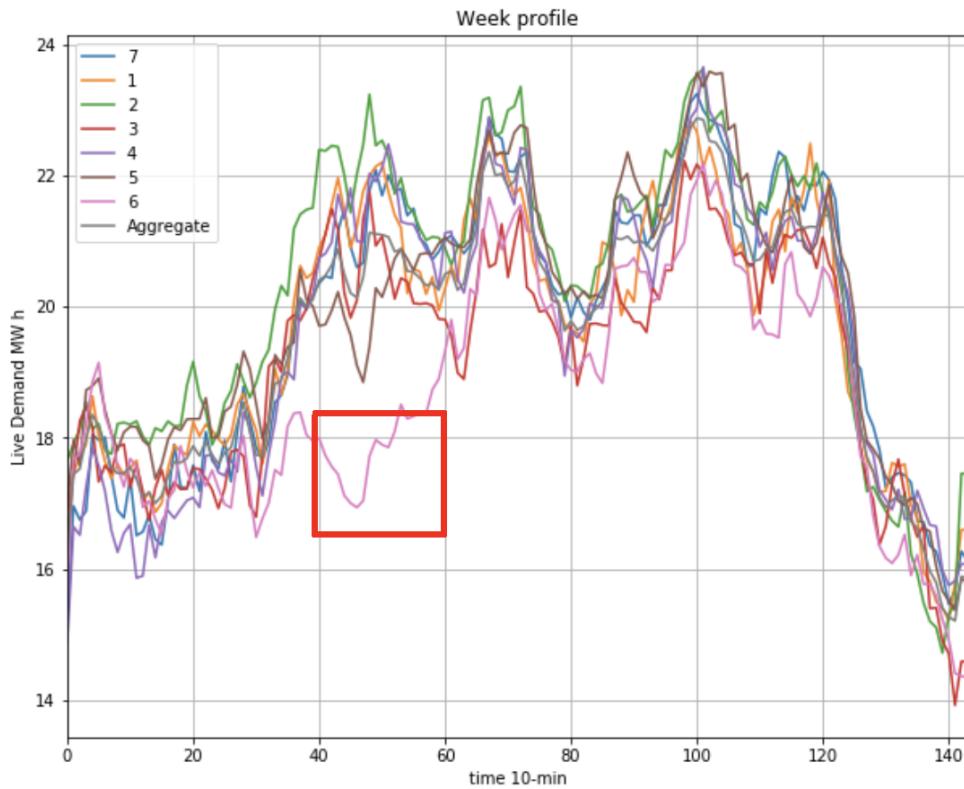


Figure 3.3: The aggregations of demand in 7 weekdays of November and the aggregation of the month

discrepancy of them is at the red colored rectangle, where the demand of Saturday in the morning is lower. it can be verified from the correlation matrix that the scores for Saturday (at the bottom row) is lower about 0.1 than those of other weekdays on average. Except this disagreement, other aggregations appear to be largely similar and seem to follow the same path of growing and decaying during a day period. The aggregation of these seven weekday profiles is thus following the same pattern of the day.

3.1.3 Constructing Month profile

As it is mentioned in the last section, a profile of the month can be made using the aggregation of each weekday which is actually the aggregation of each day in the month. The reason is that the results of the aggregations of each weekday appear highly correlated, so that a new aggregation can be calculated based on this result. Apart from the behaviors in a day definition, the behaviors in a season definition can differ in various ways. The most general case is the heat is the heating consumption which is approximately zero in the summer but has a large load profile in the winter.

[Figure 3.4](#) shows the seasonal effect on heating appliances which is collected from [21]. The factor can reach about 3.0 in January and drop to 0 in summer months, thus causing big variances for month profiles in different seasons.

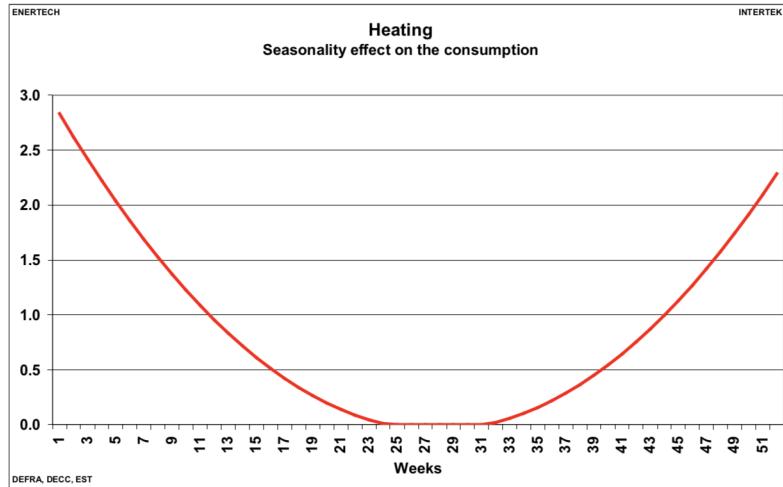


Figure 3.4: Seasonal effect of heating

3.1.4 Analyzing the month profile of November

Limited by the coverage of data set that it is collected in late October and huge corruptions happened in December and January (in [Section 2.2.3.2](#)), the current available month profile analysis can only be applied to November. Additionally, the overview of this month has been enough for later analysis on the effects of EVs. [Figure 3.5](#) shows the final aggregation demand profile of November with three constant patterns: 21.9, 20.2, 17.7.

3.2 Analyzing Orkney Wind generation

The wind power generation analysis is fundamental to this project of building a generation-to-demand model. This analysis will provide statistical wind power characteristics and real-time cases for later simulation. In this section, the ANM record of wind power generation is first discussed, then a linear regression model will be created to link wind speed with wind power generation with curtailment introduced.

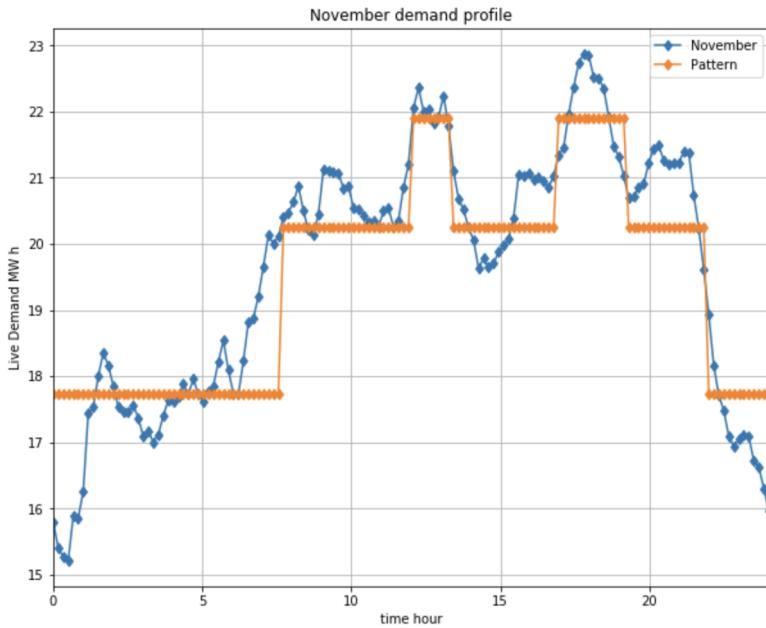


Figure 3.5: The aggregated demand profile of November

3.2.1 Analyzing the power generation attributes

As it has been mentioned in [Section 2.2.1.1](#), there are three attributes relating to the power generation: Orkney ANM, Non-ANM Renewable Generation, Total renewable capacity. The Orkney ANM is indicating the power generation from ANM system, and the Non-ANM Renewable Generation is the generation which is not controlled by ANM. As it is recorded in [29], there are three kinds of connections of generation to grid. The Total renewable capacity should be the estimated full load generation at current time from its literal meaning. Adding the first two attributes as the Total generation to compare to the capacity by statistical characteristics. [Table 3.5](#) shows their statistical characteristics, where Total Generation has the same standard deviation as Total Renewable Capacity, implying that they are vibrating at the same amplitude. This condition is similar to [Section 3.1.1](#), where the Peak Winter Demand and Live Demand can be added up to a constant at each point. Thus, the sum and subtraction of Total Generation and Total Renewable Capacity are performed as the Sum and Subtraction attributes in the table. As previously inferred, their sum appears to be a constant again. It is preposterous that the ANM choose to use these two attributes which can always be calculated from others, nor are their names misleading the readers.

Originally, another explanation is assumed that the Total Renewable Capacity attribute is the estimation of the available capacity of storage to the grid. There are two reasons to refute this position. One is that the ANM network has been commissioning

	Orkney ANM	Non-ANM Renewable Generation	Total Generation	Total Renewable Capacity	Sum	Sub
mean	10.09	12.43	22.52	34.58	12.07	57.1
std	6.92	6.99	13.54	13.54	27.08	0
min	-2.09	0.26	-1.53	15.08	-26.95	57.1
25%	2.89	5.40	8.32	22.25	-12.61	57.1
50%	11.23	14.73	27.25	29.85	2.61	57.1
75%	16.54	18.47	34.85	48.78	40.46	57.1
max	21.26	25.96	42.02	58.63	60.17	57.1

Table 3.5: Statistical characteristics of Generation attributes

no battery storage since 2015 and the only available storage is the hydrogen fuel installed by *Surf'n'Turf* [29]. The other reason is that the storage should be represented by energy not power and even if it is represented using power, the mean, which is 15.07 from Table 3.5, is not reasonable. Because the wind is strong in Orkney and there is 61% of the time the generation is higher than the demand. This implies that the battery should be at approximately full storage for most of the time. Having proved that Total Renewable Capacity is of nonsense, the two renewable generation attributes are applied to an additive operation on element wise to form a Total Generation attribute, as they are both renewable generations installed by different sponsors.

3.2.2 Analyzing the weather attributes

As it is discussed in Section 2.2.1.3 that wind speed, temperature and humidity attributes are recorded in the data set, a first attempt is on the wind speed for its strong relation to wind power generation. Table 3.6 shows the wind speed characteristics. Though the average speed is 6.01 m/s, the max value can reach 19.5 m/s, which must be identified as extreme weather. The 50 percentiles reach only 5.7 m/s, indicating that about half of the days have smaller wind speed than the average. However, the 75 percentiles construct a strong wind group with larger deviation than other intervals, implying the various conditions of wind, which result in challenges for the wind generation.

Mean	Std	Min	25%	50%	75%	Max
6.01	3.42	0.5	3.1	5.7	8.2	19.5

Table 3.6: Wind speed statistical characteristics

Another special scenario of wind speed is its distribution. From the data of Orkney,

the mean of is more near to the min than the max. If it is subtracted by two standard deviation, it will be a minus value and it is impossible. While the adding of two standard deviation is possible and it results in at least 75 percentiles, which is adhere to *Chebyshev's Inequality*⁴ to be smaller than 0.25. The distribution is *Weibull distribution* [59], which is a modified version of the general *Rayleigh distribution*.

3.2.3 Analyzing the curtailment attributes

The curtailment data is collected as binary type in [Section 2.2.2.2](#). There are three kinds of issues in a zone: ANM Operation, SHEPD Equipment, Generator Site Issue. For the sake of simplicity, the curtailment of one zone is the result of any issues in the zone.

3.2.4 Linking wind power generation to wind speed and curtailment

The difficulty of this problem is that the influence of the other attribute cannot be removed while analyzing one attribute. For example, the wind power increases along with the growth of wind speed, it then reaches a maximum limited by the largest rotation speed of the wind turbines. Thus, the model of generation should first increase linearly with wind speed, then appears horizontally at higher wind speed. With the addition of curtailment, the model reaches the maximum level earlier and even the generation with lower wind speed is affected, as the curtailments happen whenever the generation exceeds the demand to some degree. It is thus possible to make a regression model which is allowed to under-estimate the generation.

3.2.4.1 Choosing and Analyzing the sample data

It is deeply acknowledged that the data matters more than algorithms in machine learning, which is discussed in [60]. However, the data size of this model is just from one day which is specifically selected to demonstrate the main feature of wind speed to generation. As it is depicted in [Figure 3.6](#) that 12-02 is selected as the sample source. The reason is that an apparent linear relation of wind speed to generation and the upper limitation can be detected in the figure. It can be found that the the generation grows linearly reciprocal to the wind speed before the wind

⁴In probability theory, Chebyshev's inequality also called the Bienaymé–Chebyshev inequality) guarantees that, for a wide class of probability distributions, no more than a certain fraction of values can be more than a certain distance from the mean.

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad [58]$$

speed reaches 8 m/s. Then it stops at about 38 WM h. The red line in the figure depicts the expected generation characteristics.

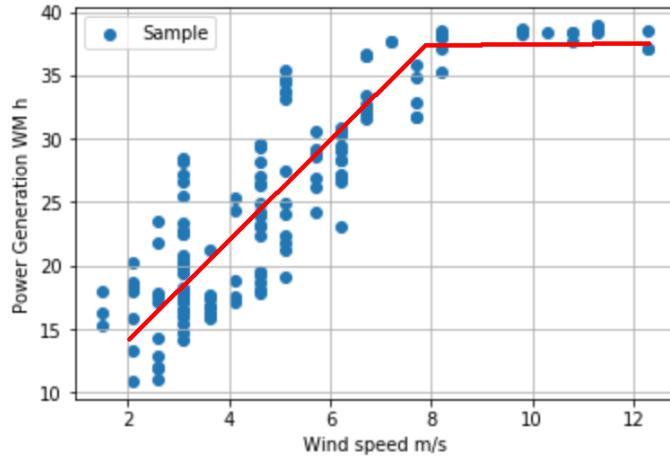


Figure 3.6: The wind speed versus power generation on 12-02

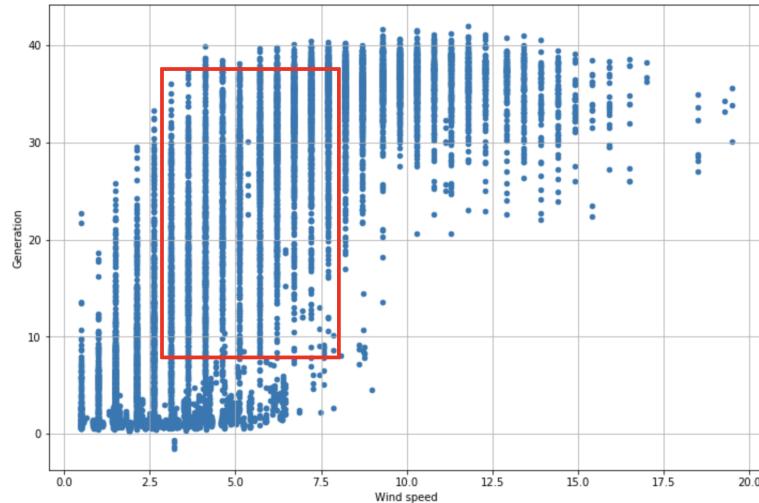


Figure 3.7: The wind speed versus power generation on the whole data set

If the whole data set is selected, as Figure 3.7 indicated, various curtailment conditions will definitely affect the model. As the red colored rectangle emphasized, the value generation of 2.5 m/s to 7.5 m/s has approximately the same range of value which cannot be appropriately understood by the algorithms.

3.2.4.2 Modeling the relation of wind speed and power generation

Two linear regression methods are tested on this relation: linear regression, and polynomial linear regression. As the red line shows, the linear regression of the first

order should be separated into two regions. The case for the second order linear regression is easy to apply on the whole wind speed interval.

The resultant parameters of two regression models along with their Mean squared error are shown in [Table 3.7](#). These two regression strategies both result in an acceptable **MSE**, thus it is appropriate to use both of them to predict wind power generation from the speed.

model	intercept	a_1	a_2	MSE
linear	7.5	3.75	0	14.4
polynomial	5.1	5.22	-0.19	15.05

Table 3.7: Parameters of two linear regression model on wind speed to power generation on 12-02

It should be noted that the reason why the conditions happen in the rectangle region of [Figure 3.7](#) is that the curtailment occurs whenever the demand cannot fully follow the generation capacity to sustain the stability of the grid.

3.2.4.3 Modeling the relation of curtailment and power generation

Under the condition that the demand is enough large that no curtailment occurs, the wind speed-to-generation model derived in the previous section is approximating the power generation well. However, according to statistics of last few years, the time of curtailment takes up to 31% of all time periods. Another model modeling the generation under curtailment should be added to include all conditions in the wind power generation model.

The assumption is that, when the curtailment happens, some wind turbines are cut out, causing the power generation being limited to a level slightly above the demand. This assumption is expressed as:

$$\text{Generation} \leq \text{LiveDemand} + k \quad (3.1)$$

The crux is to find the optimal value of k . Using the wind speed-to-generation model to predict the power generation of the full capacity, then add k to the current demand. If the generalization is higher than this value, this time slot will be marked as curtailment happen. Comparing the predictions of curtailment with the collected ANM status data using the *Confusion matrix* in [Section 2.2](#). To acquire the value k which fits in a general scenario, the curtailment data of nine regions is transformed into a single boolean value which is true if at least two regions have reported curtailment. The linear speed-to-generation model is the polynomial model, and

the values of k range from 1 to 20.

k	1	2	3	4	5	6	7	8	9	10
Type I	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.11	0.13	0.18
Type II	0.53	0.50	0.46	0.41	0.37	0.34	0.31	0.25	0.20	0.17
k	11	12	13	14	15	16	17	18	19	20
Type I	0.22	0.28	0.35	0.42	0.52	0.60	0.68	0.74	0.79	0.86
Type II	0.14	0.11	0.09	0.07	0.05	0.05	0.04	0.02	0.02	0.01

Table 3.8: Type errors of parameter k

Table 3.8 shows the scores of this binomial regression model. The performance is relatively high ⁵ when $k = 8, 9, 10$. Taking the median of them and the model is constructed as:

$$Generation \leq LiveDemand + 9 \quad (3.2)$$

3.3 Modeling the over-all EV effects

To analyze the demand of EVs, it is necessary to first construct the model of demand, then it can be synthesized with the Live demand model derived in the previous section.

As it is elaborated in [Section 2.6](#) that the demand of EVs to the grid is determined by the charging station type and the capacity of batteries. Besides the power approximation, the demand in time-series should be characterized using the charging events model, which is statistically fitting to the histogram of arrivals and departures in [Figure 2.11](#).

In this section, the charging events model is first designed. It is fundamentally necessary to the model of demand of EVs and the implementation of simulating EV charging. Then, the effects of different penetration levels of EVs will be analyzed.

3.3.1 Characterizing EV charging events

To characterize the charging events of EVs, several presumptions are first promoted: (1) All of the EVs are charged only one time each day, and they are charged equivalent amount of energy. (2) Only one type of EV is considered to unify the capacity of batteries. (3) Bidirectional power transportation technologies: V2G, V2H, and V2V [61], are not considered to decrease the degree of complexity.

⁵By taking the average of the two type error for each k or the weighted sum of them, as the target is to predict the occurrence of curtailment, not the normal mode.

From [Figure 2.11](#), it seems that there are two peak arrival time each day: 12:00 and 18:00, and the latter is larger. The first distribution is extracted from [\[20\]](#), which uses a normal distribution with $\mu = 18 : 00$, $\sigma = 1$, implying that the peak time is at 18:00, and most EVs arrive in the range $15 : 00 - 21 : 00$ ⁶. Using the random number generation module of *Numpy*, it is possible to produce a set of charging start points with the number of EVs provided. [Figure 3.8](#) shows a histogram of samples in this distribution.

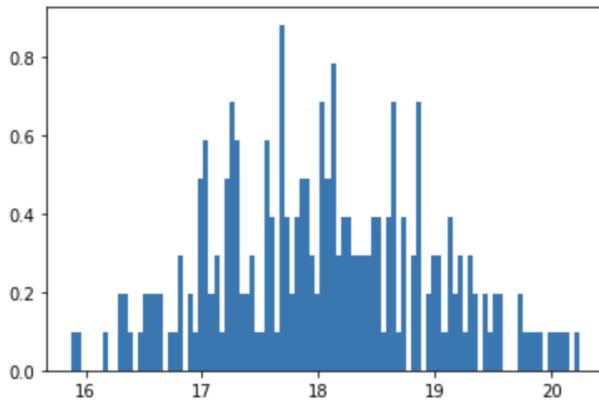


Figure 3.8: The Histogram of the samples generated by a Normal distribution of $\mu = 18, \sigma = 1$

There are 234 EVs in Orkney ([Table 2.6](#)), and the total daily demand can be calculated: $W_{perday} = 21.74 * 0.299 = 6.5kWh$. Two charging modes are considered in this scenario: 110V/15A * 5.5h as the Normal charging, and 240V/30V * 1.5h as the quick charging. Under such condition, the demand caused by the charging events can be visualized in [Figure 3.9](#). The differences on instant charging power and length of charging periods lead to significant demand profiles of the two modes. The peak value of Quick charging is over 2 times higher than that of the normal charging mode and the peak value reaches about 2 hours earlier than the Normal charging mode⁷.

Despite the demand profile derived from one peak time, a more pragmatic approximation can be constructed by two normal distributions centering at 12:00 and 18:00, respectively. The idea behind the choice of these two center point is to achieve a replica of the Arrivals vs Departure Histogram in [Figure 2.11](#).

The histogram of the charging events and the demand profile of this new approximation are demonstrated in [Figure 3.10](#) and [Figure 3.11](#).

⁶The interval is 3 standard deviations of the mean, which covers 99.7% of the data [\[62\]](#). Actually, σ can be slightly larger as there are still arrivals after 22:00.

⁷It also implies a possible solution to delay the demand peak time by transform Quick charging to normal charging.

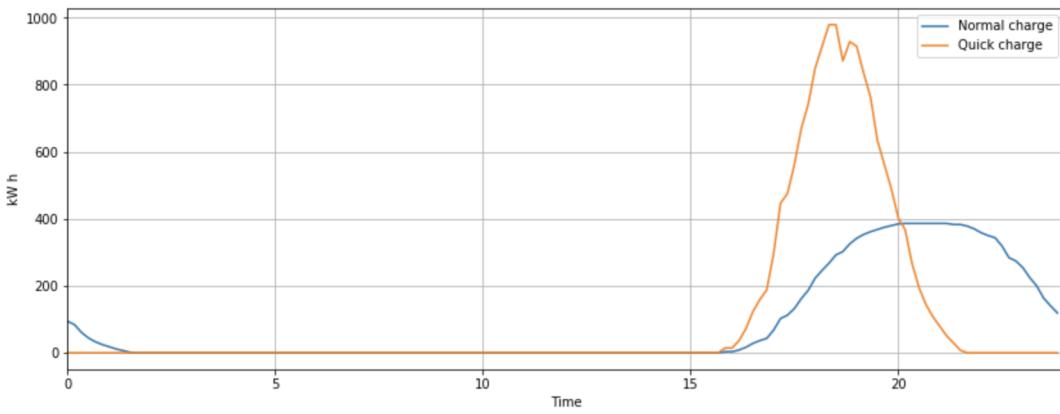


Figure 3.9: The demand profile of 234 EVs under two charging mode

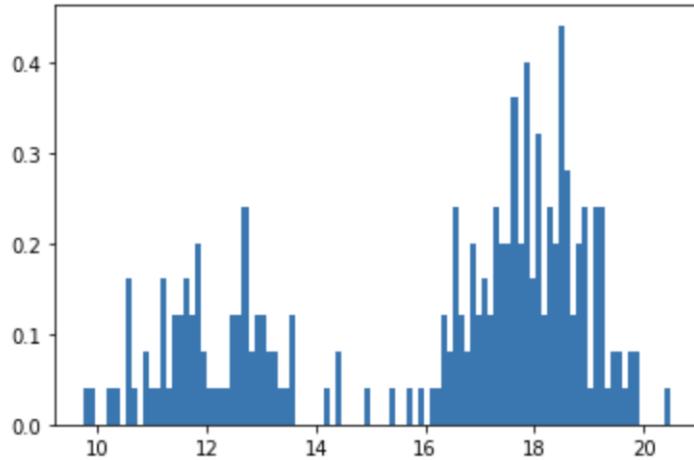


Figure 3.10: The Histogram of the samples generated by two normal distributions at 12:00 and 18:00

3.3.2 Analyzing different penetration levels

The current penetration level 1.4% has not attached a threat to the grid. If all EVs are recharged in one day, the total consumption is $6.5 * 234 = 1521\text{ kWh}$, while the island total demand on a daily average is 469 MWh . This little consumption can be visualized in Figure 3.12, where merely a slight increase at night can be detected. Comparing to the prediction of EV penetration in [33] that the percentage is 35% in 2020, this penetration level is over-estimated. Three penetration levels: 10%, 20%, and 30% are tested on current demand profile. Table 3.9 shows the results after adding the charging events at centering at 18:00. The scenario without EVs is also estimated by subtracting the 1.4% EV consumption from the demand profile, then new penetrations are tested. During the process of increasing the penetration from 10% to 30%, the total energy demand, which can be estimated from the mean values, is increasing over 10%, the max instant power consumption increases about

3.3. MODELING THE OVER-ALL EV EFFECTS

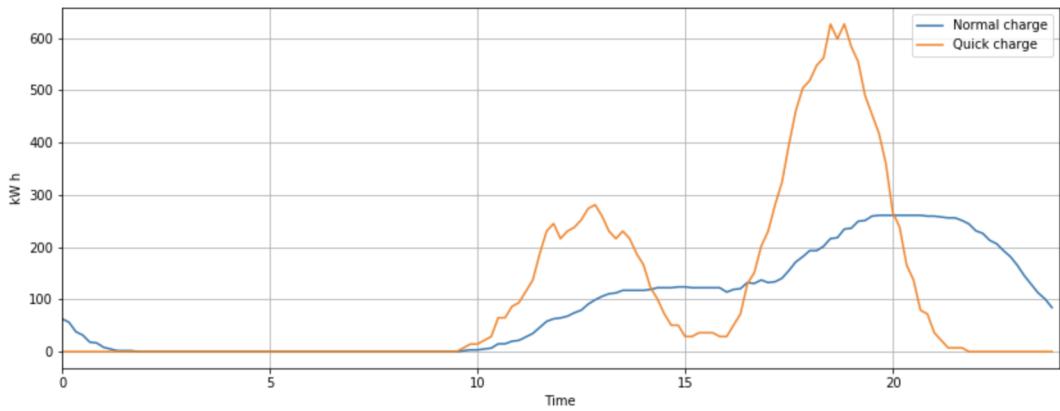


Figure 3.11: The demand profile of 234 EVs with two charging peak time

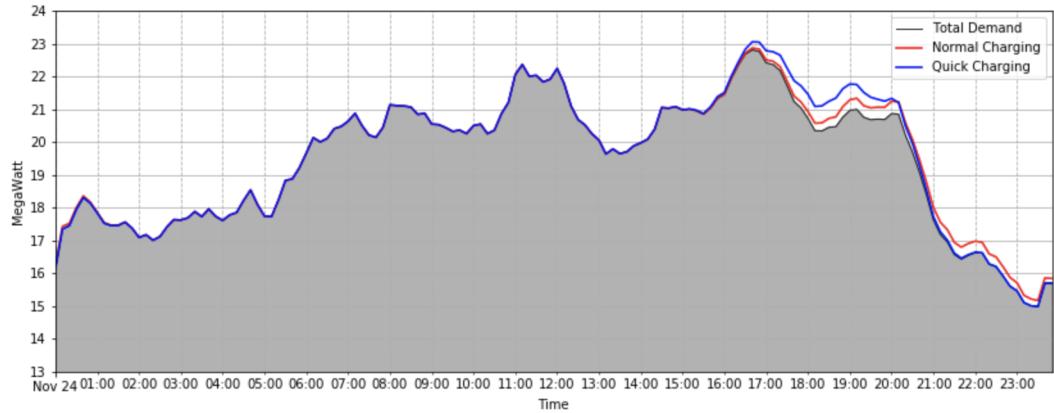


Figure 3.12: The Synthesized demand profile of 1.4% EV penetration

30% for Normal charging and 80% for Quick charging. The time is delayed to after 8 o'clock in the evening for Normal charging, and 6 o'clock in the evening for Quick charging.

A more obvious impression is available by plotting the demand profile of the 30% penetration in [Figure 3.13](#), where a single charging interval is too significant to accept as a rational estimation. The reason may be that the 30% is impossible for current demand level, and it is also an estimation of maximum with all EVs using Quick charging mode, which implies that the real consumption should lie between the Quick mode and the Normal mode. Additionally, if the distribution of two peak times is considered, the peak value will be alleviated and the peak value will be distributed to two mountains on the graph.

	Mean	Max	Peak Time
No EV	19.6	22.88	15:40
10% (Normal)	20.11	24.14	18:10
10% (Quick)	20.23	28.91	18:20
20% (Normal)	20.74	26.44	20:10
20% (Quick)	20.98	34.7	18:20
30% (Normal)	21.37	29.19	20:10
30% (Quick)	21.73	41.36	20:10

Table 3.9: The statistics of demand under three penetration levels

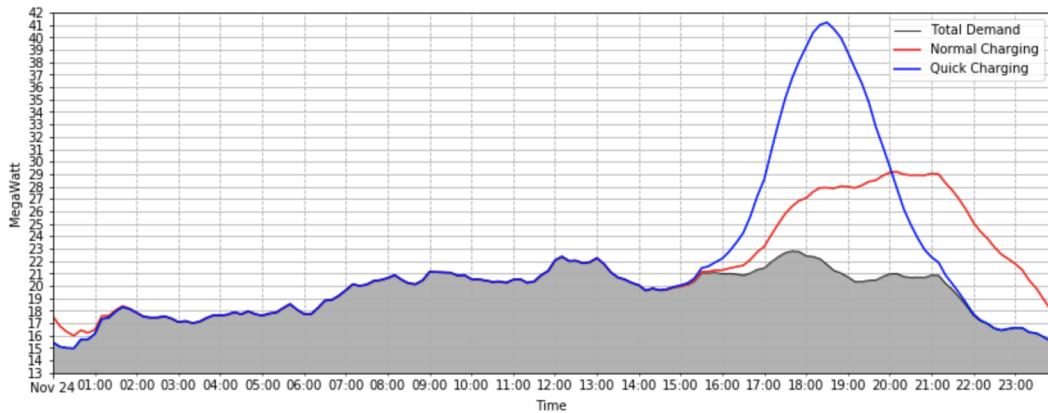


Figure 3.13: The demand profile under the 30% penetration of EVs

3.4 Implementing the simulation model of the grid

In this section, the details of the implementation of the simulation model are recorded. This simulation model is designed under the *NASA Trick* simulation environment, which has been introduced in [Section 2.4.2](#). The whole grid model is separated into several independent models: household loading model, EV charging model, wind model, and storage model. The deliberate design of these parts makes it possible to simulate each single part without the couple of other elements, and it is better for unit test. Later these models are integrated into the whole grid model.

Each model has four common methods: default, init, analytic, and terminate. The ‘default’ method is called before ‘init’ method to assign parameters to default value, the ‘init’ method is called to initialize the state of some variables and manage memory blocks. These two methods are both called once at the start up of the simulation. The ‘analytic’ method is called on every time slot to process the simulation, which means the number of times of calling this methods is equivalent to the number of time slots. The ‘terminate’ method is called at the end of the simulation to summarize the result and clear applied memories.

To exemplify, the household model identifies all appliances in the house, reads in the data of profiles of appliances, and creates variables for them in ‘default’ method. Then, the ‘init’ method is called to synchronize all appliances to the state of the same time and prepare other miscellaneous parameters for later analysis. The simulation control logic create the *schema*, which stores the procedures on each time slots, and starts the timer. The ‘analytic’ method updates the state of each appliance on each calling. If some data record processes are register, for example, the total energy consumption is recorded on every 10 minutes, and the consumption of light is recorded on every minute. When the *schema* runs to the end, the ‘terminate’ method is called, the mean and sum of the total consumption can be recorded in this method, and scores based on objective functions can be calculated, then the applied memory blocks are cleared to avoid memory leak.

3.4.1 Simulating Household load

The house contains various appliances which have specific characteristics, which have been summarized in [Table 2.3](#). There are 8 types of appliances considered in this model: refrigerator, cleaning, heating, cooking, lighting, audiovisual, computer, unknown. Among them, the cleaning, lighting, audiovisual, computer, and unknown are characterized using stochastic model, and others are based on periodic constants from their consumption profile. Because these appliances have different demand profiles, they are independently designed small models. The data of the consumption of appliances is based on [\[21\]](#), and the details are included in [Figure 4.1](#). The consumption profiles of specific appliance are also provided in the [Appendices](#).

3.4.1.1 Simulating periodical appliances

The simulation of appliances which consume energy with predefined period is implemented by storing the data of a period in the model. As the time updates, the consumption of the appliance is adhere to the data at that time ⁸.

3.4.1.2 Simulating stochastic appliances

It is known in [Section 2.5](#) that the stochastic appliances can be classified into two group: randomly plug in with randomly working period, randomly plug in with fixed working period. For the former type, the time of plug in should be random

⁸Thus, a simple implementation is to store the time-series data in to an array, each time the consumption is updated by reading the corresponding data by calculating the index

and the working period should also be random, the later one is only random on the time of plug.

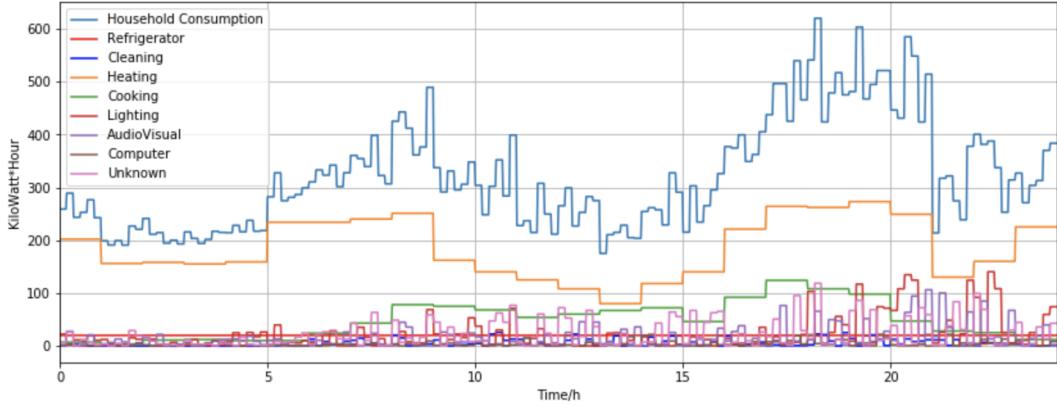


Figure 3.14: A one-day-interval simulation result of the household

Figure 3.14 shows the simulation result. Comparing to the source of data in Figure 4.1, the simulation successfully recover the scenario of household consumption.

3.4.2 Simulating EV charging load

To simulate the charging process of electrical vehicles, two models are designed: EV, and charging station. The model of EV has attributes indicating the SoC, the time of arrival and departure, which use the statistical model derived in Section 3.3.1. Each time the EV arrives at home, it is connected to the charging station, which controls the charging rate⁹.

Three charging strategies are implemented in this charging station. The first is to charge the EV whenever it is connected to the charging station. The charging power is the normal mode: 110V/15A. The reason is that the charging station generally ignores the current state of the grid to complete the charging and the normal mode causes less instant power consumption to the grid. The second one is to delay the charging time to the late night, generally after 24:00, as it is the time interval that the demand of other appliances is at the lowest level. It is also using the normal mode charging power as this is a period of time that most EVs are charging together and the night gets enough time for them to be fully charged. The third strategy is based on a bi-directional communication channel, which makes it possible to let the charging station determine the charging power dynamically. In this strategy,

⁹As the strategies on charging should be tested later, it is necessary to abstract a layer between the grid and EV, the charging station, which communicates to the network and determines whether to charge the EV at the current time

the charging station predicts the expected departure time of the EV and determine whether to charge the car based on the current load of grid. Charging power is the quick mode as the smart strategy has the feedback from the grid.

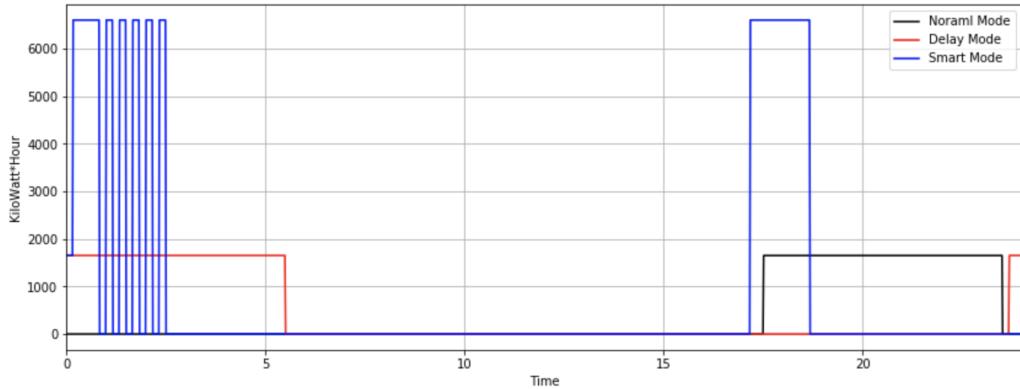


Figure 3.15: Simulation results of three charging strategies

An illustration of the three strategies is depicted in Figure 3.15, where the real-time consumption of three strategies are separated obviously: the normal strategy has a continuous distribution starting from about 18:00, the delay strategy has a continuous distribution located at around 24:00, and the smart strategy results in a non-deterministic consumption as it is based on current state of the grid.

3.4.3 Simulating wind power generation

As it has been discussed that the wind power generation is directly related to the wind speed, and two linear regression models have been established in Section 3.2.4 to predict the power generation based on wind speed. Therefore, the concentration is attached on building a model for the wind speed. The statistics of wind speed has been analyzed in Section 3.2.2.

It is the easiest idea to use a normal distribution, which is feeded by the analyzed mean and variance. The generated wind speed must be the same to the assumed statistics. However, the data should be located on the time series, which means that each generated random sample should be related to the previous ones. Assume all the wind speed of the whole day is $a_{day} = a_1 a_2 a_3 \dots a_n$, where each a_x is a random variable following the same normal distribution. From the *Weak Law of Large Numbers*¹⁰, it is these i.i.d rvs make a_n the same distribution as the sampled data. If the samples are related to the previous outputs, this generation should

¹⁰In probability theory, the Law of large numbers states that the sample average of the sequences of i.i.d rvs: $A_Y^n = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$ has the mean: $\mu_A = \mu_Y$, and the variance: $\sigma_A^2 = \frac{\sigma_Y^2}{n}$.

be a *Markov Chain*, which models dependent rvs best. For *Markov Chain*, the states and the transition matrix should be predefined to reflect the distribution of wind speeds. Therefore, the problem actually changes to the determination of the transition matrix. If the whole project is to find a better model of simulation wind speed, this should be a good question. However, there is neither enough content, nor time to discuss this problem in this project.

The implementation of a normal distribution cannot be used because the generated data should be continuous. The Markov model is neither able to be implemented in a short term. One compromised solution in this project is to select the wind data in real world. To let the data fully reflects all the scenarios of wind speed, various days should be chosen carefully. Finally, there are 5 days selected in this project. [Figure 4.9](#) depicts a typical increasing tendency, it has an average wind speed of 5 m/s. [Figure 4.10](#) is a typical decreasing tendency, it has an average wind speed of 7.5 m/s. [Figure 4.11](#) is a day with high wind speed all day, the average wind speed is 7.6 m/s. [Figure 4.12](#) is a day with low wind speed all the time, the average wind speed is 2.3 m/s. The last one is in [Figure 4.13](#), which has an average of 5.7 m/s and is close to the mean from the global view. It is assumed that these five scenarios can reflect all weather conditions, and thus support reasonable analysis.

The actual implementation of the weather condition is based on a nonuniform random distribution. The reason is that the first four selected cases have the means not near the global the mean, and the 5th weather condition should be the normal one. The problem can be expressed as: $5.0a_1 + 7.5a_2 + 7.6a_3 + 2.3a_4 + 5.7a_5 = 6.01$, where a_1, a_2, \dots, a_5 are the coefficients of the probability function of random variable X as the weather condition. A solution used in the simulation system is indicated in [Table 3.10](#).

a_1	a_2	a_3	a_4	a_5
0.12	0.23	0.08	0.12	0.45

Table 3.10: The implementation of the 5 coefficients of the p.f.

It can be found that the 5th weather condition has the most significant weight. The reason is the it is the most common weather. The weight of the extreme weather, the 3rd and the 4th conditions, have a sum of 0.2. The synthesized mean of this implementation is 5.8. Now the wind speed simulation satisfies both the requirements on the statistical characteristics and the real world nature of wind.

[Figure 3.16](#) shows one result of the simulation system with a 5-day-interval. Though 5 days are simulated, three weather conditions are happened, as the 5th weather condition has a larger weight. The average wind speed of these five days is 5.9 m/s,

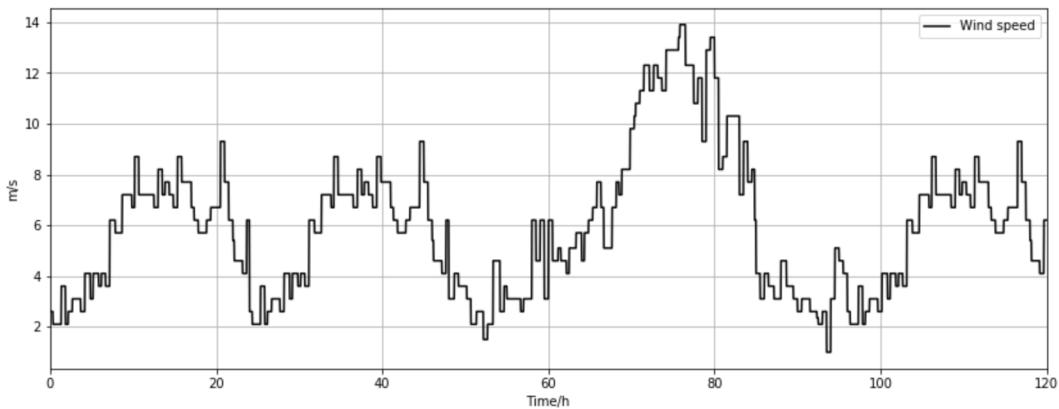


Figure 3.16: An illustration of the results of wind speed during a 5-day-interval ($5 \times 24 = 140$ h)

and the standard deviation is 2.76 compared to that of data set at 3.42. This result is accepted.

3.4.4 Simulating grid storage

The storage is simply a battery, which is recharged when the demand is lower than the generation, and is discharged when the generation cannot satisfy current demand. When SoC is empty and the demand is higher than the generation, the imported energy is used. When SoC is full and the generation is higher than the demand, the curtailment is set to cut down some wind turbines and export energy.

3.5 Determining the storage

The implementation of the simulation has been elaborated, while two important parameters have not been determined: the scale factor of the wind farm and the storage of the grid.

3.5.1 Determining the scale of the wind farm

Because it is a small grid with only 10 to 20 households, the scale of the wind farm should be at an appropriate range to reflect the condition in Orkney. If the scale is too low, the generation will fail to satisfy the demand most of the time. If the scale is too high, there is no need to use storage to compensate the shortcut. A practical understanding of the generation to the demand is extracted from the collected data set that by measuring the $Generation to Demand (G2D) = Generation Power -$

Demand Power, 39% of the time, the generation cannot satisfy the demand. From the perspective of the overall generation to overall demand, the average generation is $22.52MWh$ from [Table 3.5](#), and the average of demand is $19.04MWh$ from [Table 3.1](#). Thus, the simulated total generation to demand value should be positive and the line in the figure should have many downward parts to indicate the 39% unsatisfied time.

To determine the scale factor, a metric based on assumed performance is established. The overall generation to demand on the plot is used. The line of the appropriate scale should often above $y = 0$ to satisfy the feature of average, and the line should have many downward parts to satisfy the statistics of generation to demand. *Monte Carlo simulation* is used to find the optimized scale factor. The time interval is 4 days, as there are four typical weathers in this simulation, and the scale factor is tested from 50 to 1100.

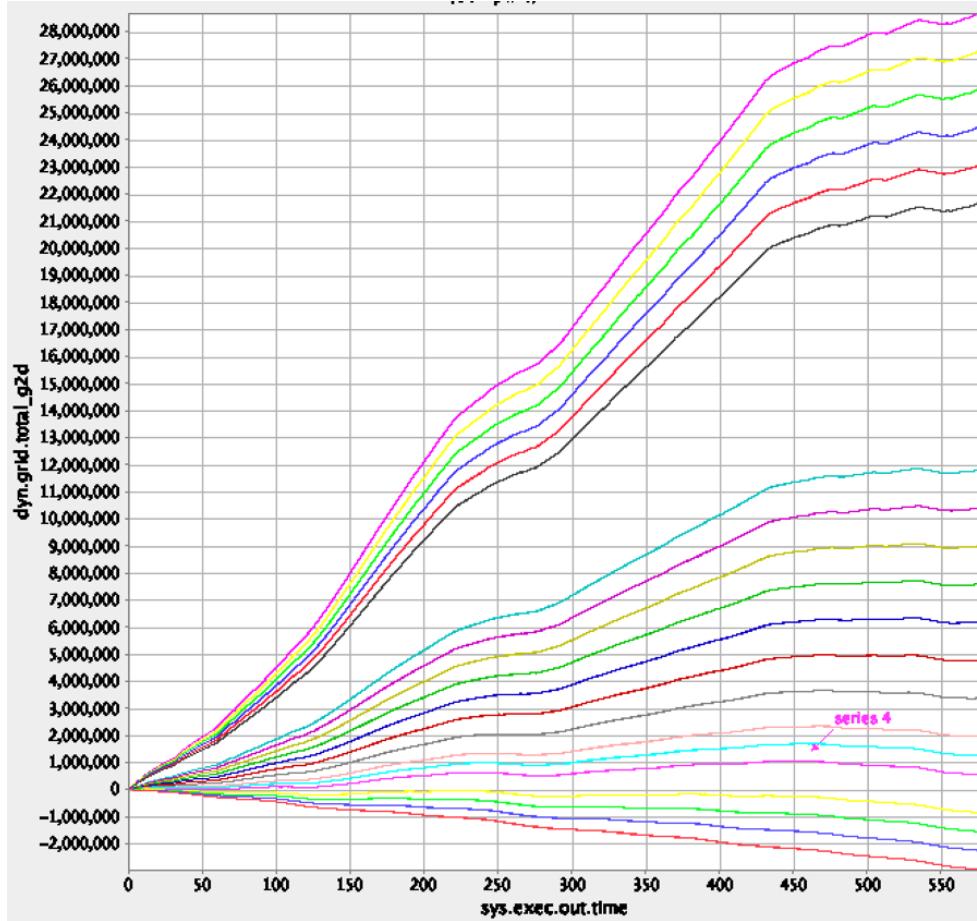


Figure 3.17: The plot of the simulated total generation to demand under various scale factors. (Generated by *trick-dp*)

[Figure 3.17](#) shows the results of testing different scale factors. Apparently, the

fifth and sixth lines are the scale factors that mostly satisfy the requirements of the model. The value is 300 on the *series 4* arrow, and this value is selected as the scale of the wind farm. On the other hand, this value should estimate the lower bound of the wind farm, as there are more wind turbines added later.

3.5.2 Determining the storage

The determination of storage is based on all previous work, as only until the demand and generation have been simulated, will it be possible to estimate the storage of the grid. A proper storage can not always be empty, as it should provide energy to compensate the unstable feature of wind generation, nor this storage be always full, as the curtailment will happen frequently, causing more waste on potential generation.

Therefore, a metric is defined to estimate the quality of the current storage: the portion of time the SoC of the grid storage is in the range 15% to 80%.

$$\text{Quality Of Storage} = \frac{\text{Total time of } (15\% \leq \text{SoC} \leq 80\%)}{\text{Total time}} \quad (3.3)$$

It is the percentage of health condition, and it is quantified as a score between 0 and 1. From the definition provided, the closer it is approaching to 1, the better the storage is satisfying the requirements.

However, this single numeric value also has some limitations. Assume the storage is so large that any power generation or demand in the grid cannot attach any apparent change to the SoC. If the initialized SoC is at 90%, the score will be nearly 0. A similar result can be acquired from a initial SoC at 10%. The storage of the battery definitely satisfies the requirement, but the metric fails to detect it ¹¹.

On the other hand, the task is to find a storage that appropriately satisfies the need. That is to say, too large storage introduces waste as most of the storage will not be efficiently used, too small storage is not adequate for providing energy in peak hours.

To have more information on how the current storage of battery behaves, the average SoC and the portion of time that the SoC is at 0 are also estimated.

Then, the initialize SoC of the grid storage should be determined. Two kinds of initialization are considered: the initial SoC is at a constant value for all value of capacities, and the initialize SoC is at a constant percent of the capacity. The reason

¹¹Although the two assumed initial SoCs result in the battery always in a condition not beneficial to the battery life, the large storage makes it possible to accommodate the need to store enough energy to compensate for the less windy days.

why these two initializations differ is that if the constant SoC is chosen, the constant can be too small for larger capacities so that it will possibly be out of the health range at the beginning, and it will take longer time to reach the health range with the scale of the wind farm not changed. The initial state definitely affects the score, and this is what should be avoided. A compromising method is to set initial SoC to 0 for all tests, then the constant and percent initial states are both satisfied.

Additionally, a longer simulation interval is also necessary to minimize the effects of initial SoC. Theoretically, as the simulation introduces randomness, the longer the simulation period is, the more the results adhere to the predefined statistical characteristics ¹².

The length of simulation is set to 30 days. The number of houses is 20. And there are no EVs in the grid. It is necessary to first find an appropriate storage for the grid without penetration of EVs, then the impact can be quantized by discrepancies of scores.

A new problem is that range of the test data should be determined first. Figure 4.14 in the Appendices shows an extreme case, where there is no higher wind speed all day long that the generation cannot support the demand. If the battery is to assume this case as the worst case, it should be able to support the demand for the whole day without outer generation. From simulation, this value for 20 houses is 152728 kW h. In practical cases, there must be some generation from the wind farm, thus, the storage should be smaller than this value. The simulation tests the range of storage from 15,000 kW h to 100,000 kW h.

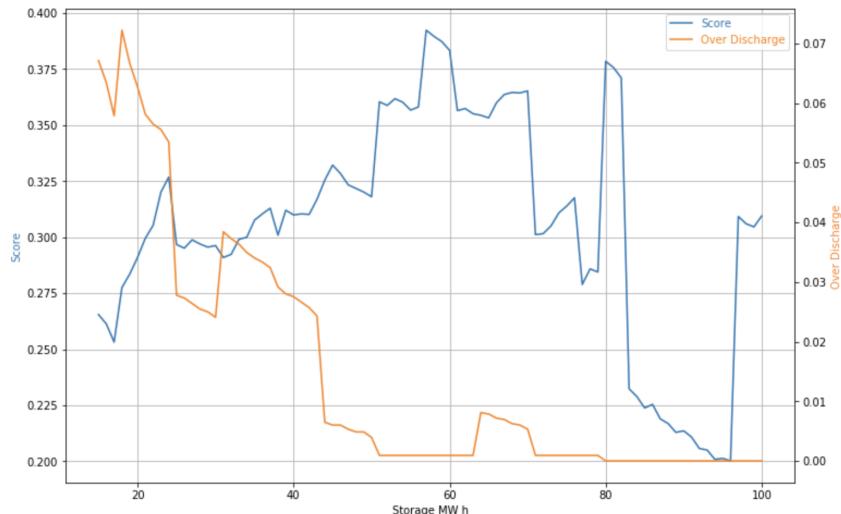


Figure 3.18: The scores and over-discharges in the range of storage from 15 MW h to 100 MW h

¹²This can be elaborated by *Weak Law of Large Numbers* [63].

Figure 3.18 illustrates the results. It can be found the percent of over-discharge has a decreasing tendency with the increasing storage. It can be explained intuitively that larger storage is unlikely to be fully discharged. However, the score is not increasing with the storage grows. The highest score is at when the storage is 57 MW h, and the storage larger than 80 MW h perform badly. This can be explained by the average SoC¹³, as the rate of over-discharge is decreasing, the grid storage can only be fully charged which cannot satisfy the range of health.

Capacity (MW h)	81	85	90	100
Average SoC (MW h)	65.7 (81%)	73.9 (87%)	78.7 (87%)	83.6 (83%)
Score	0.38	0.22	0.21	0.31

Table 3.11: The average SoC of some high storage simulations

Table 3.11 records the average SoC of some of the high storage simulation along with their ratio to the capacity. An average at over 80% definitely introduces most of the time the battery being over charged.

From the perspective of battery health, the storage does not have a good performance. On the other hand, it reflects the fact that the wind power generation has the potential to support the demand and provide more energy to be stored and exported. Though the high capacities have low score, but they are never over-discharged, this can make them much safer if the penetration of EV is increased.

3.6 Simulating the impacts of EV penetrations

The previous work makes it possible to quantize the impacts of EVs and the efficiency of charging strategies to fully use the wind power generation. The results in previous section indicate that the generation is often so large that the grid storage is fully charged. It is reasonable that the penetration of EVs will somehow make the score higher than before, as it increases the demand, however, introduces more chances to let the battery fully discharged.

There are three charging strategies, and penetration from 0 to 100%. It is exhausting to test every penetration level on each strategy, thus, 10%, 20%, and 40% are selected, and the impacts are compared. To be more accurate, the length of simulations increased to 60 days.

As the strategies have been implemented in previous sections. This section only shows the results. The results are plotted on the parasite figure by each penetration

¹³Remember that three metrics are recorded to fully evaluate the current storage: Quality factor, average SoC, percent of over-discharge.

and each charging mode. Thus, there are 9 figures in all.

3.6.1 Impacts of normal charging strategy

The normal strategy is to charge the EV whenever it is connected to the grid. This section shows the performance of normal charging strategy.

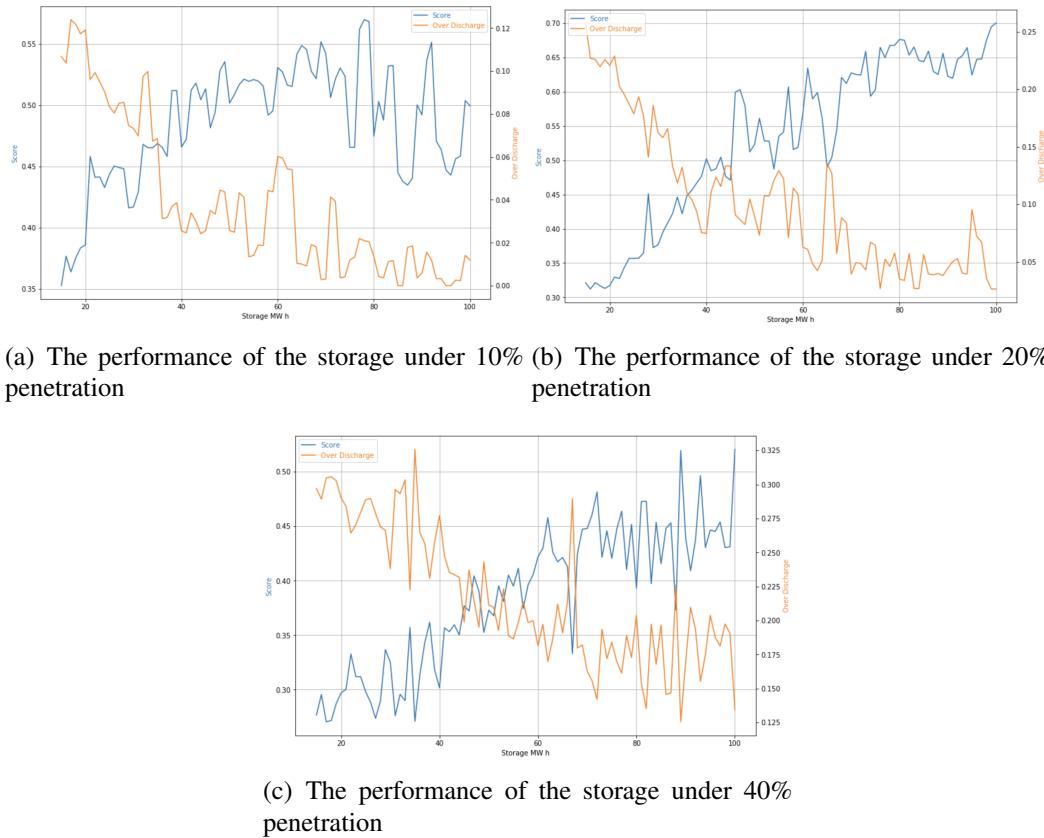


Figure 3.19: The performance of the storage under the normal charging strategy

Capacity (MW h)	81	85	90	100
Average SoC (MW h)	25.9 (31%)	21.4 (25%)	22.2 (25%)	25.6 (25%)
Score	0.47	0.42	0.44	0.52

Table 3.12: The average SoC of some high storage simulations under 40% penetration

3.6.2 Impacts of delay charging strategy

This section shows the results of delay charging strategy based on the same simulation schema in testing normal charging strategy.

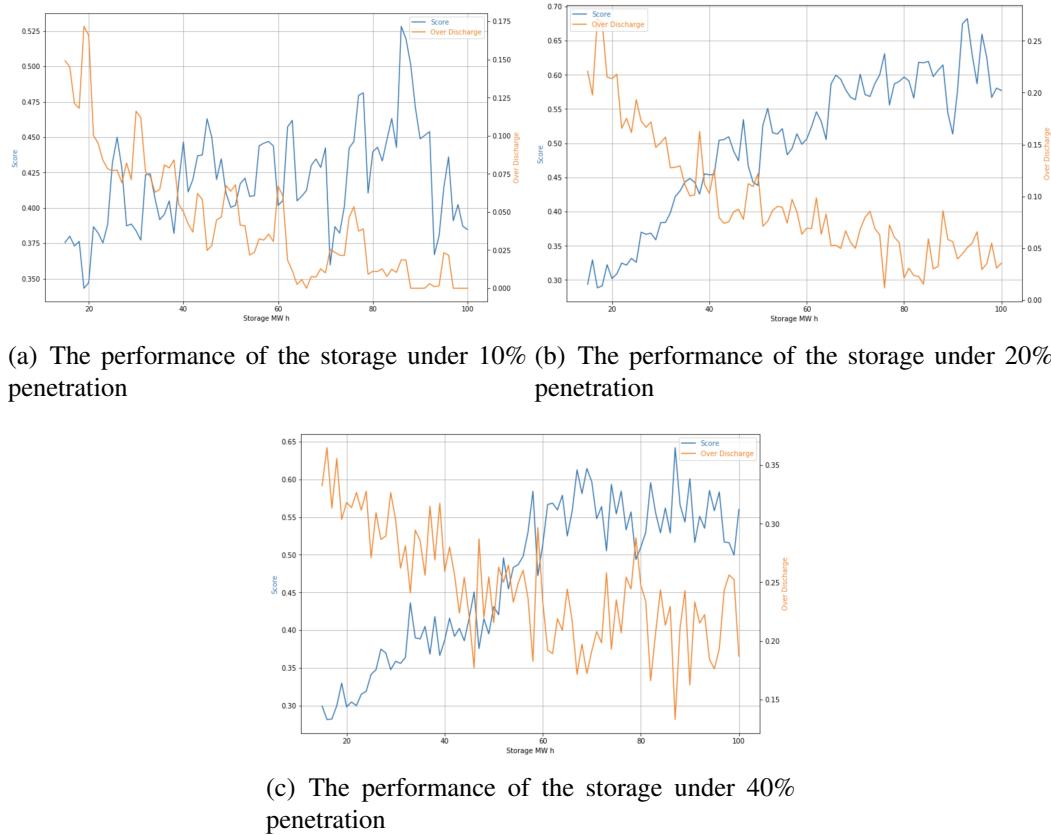


Figure 3.20: The performance of the storage under the delay charging strategy

3.6.3 Impacts of smart charging strategy

This section shows the results of smart charging strategy based on the same simulation schema in testing normal charging strategy.

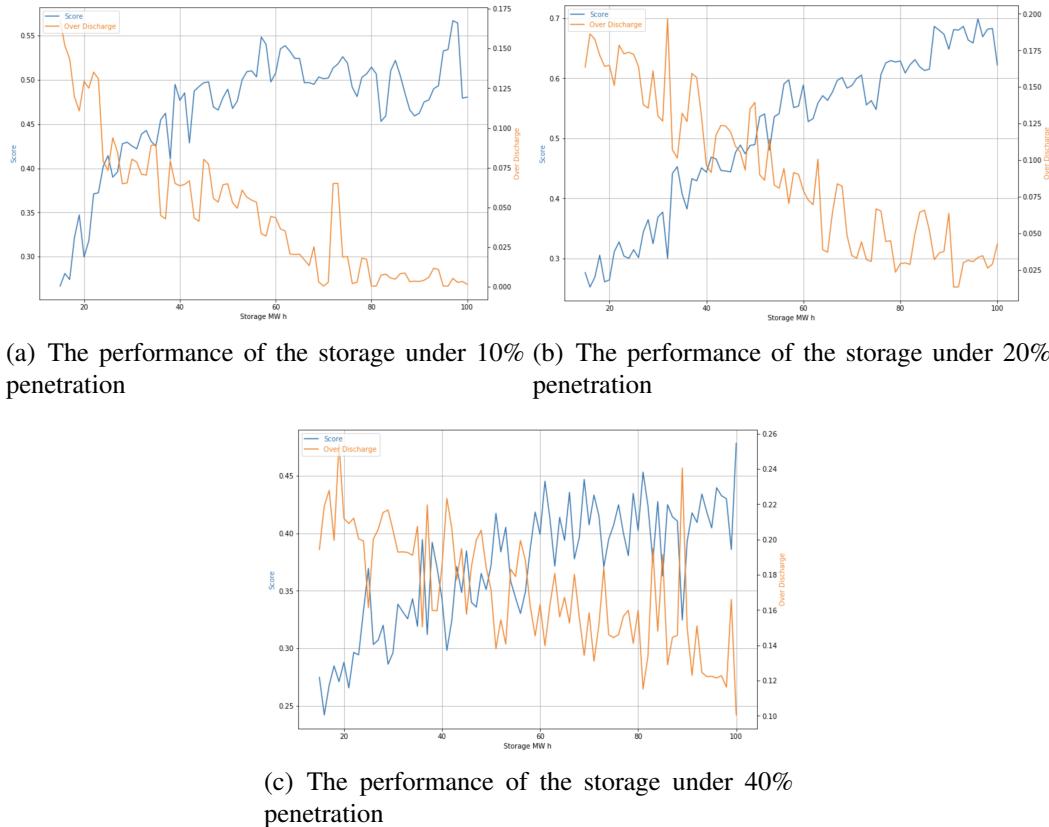


Figure 3.21: The performance of the storage under the smart charging strategy

Chapter 4

Conclusion

It is the non-determinism of both the wind and charging events of EVs that causes the instability of the grid. The extent of this non-determinism has been analyzed both by statistical methods and illustrating methods.

4.1 Impact of wind

It has been explained in Introduction that the impacts of using wind as the resource of power generation introduces the need of storage to store energy when generation is higher than demand, and the compensate the shortage when generation is lower than the demand. In Orkney, wind power generation can satisfy the demand 69% of the time. The wind resource still has large potentials, as the curtailments, which indicate the wind speed is too strong, are frequently happening that it is reported that region A has over 20% of the total time in a year under curtailment [29].

The simulation results in [Section 3.5.2](#) also shows that the wind power can fully satisfy the daily demand if EV is not introduced. A capacity of 57 MW h, which is about $\frac{1}{3}$ of the total daily power demand, has the best performance in the grid. The grid storage is largely likely to be fully charged, though, the extra power can be exported or cut off.

When the EVs are added to the grid, over generated power can be absorbed and the score of the storage increases. The increasing trend meets the turning point before 40% penetration level. This is the point that the wind farm generation can fully support. However, one limitation is that the wind power generation model is extracted from the Orkney data set, which has already introduced curtailment. If the curtailment level is alleviated to a higher value, the grid will be able to accommodate more EVs.

4.2 Impact of EVs

[Figure 3.19\(a\)](#) shows the score of normal under 10% penetration. It can be found that most scores increases to 0.5, and the maximum score moves to higher capacity. The higher capacities behave better with lower over-discharge scores, while the performance is still not significant enough. For the score of delay strategy under this penetration in [Figure 3.20\(a\)](#), the performance does not change too much. However, it is difficult to understand why the score on the map oscillates drastically. The pre-assumed results should be an increasing arc. A possible explanation is that the generation from the wind is still too large for the penetration of 10%. In the end, the delay strategy does not result in progress on performance and the best score is also over 80 MW h. The smart strategy in [Figure 3.21\(a\)](#) shows an apparent tendency of performance progress, where the score grows significantly from 20 MW h to 40 MW h. It also shows that for storage larger than 50 MW h, the performance converges to score 0.5. The advantage of this strategy is obvious, as the score is much more stable, and the rate of over-discharge is at an extremely low level.

[Figure 3.19\(b\)](#) illustrates the results of 20% penetration. As more EVs drive the demand to a higher value, lower capacities tend to show the lack of storage. Though most of the capacities in the middle region still have a score over 0.5, the over-discharge ratio is increasing. The higher capacities behave much better than before, and 100 MW h reaches 0.7 with a small over-discharge ratio. The result of delay strategy in [Figure 3.20\(b\)](#) shows the same tendency: the score increases with the capacity of the storage, and the rate of over-discharge decreases. Though the score over 80 MW h is not stable enough, the apparent increasing tendency makes it the better choice of storage under such penetration level. The result appears much more significant for the case of smart strategy in [Figure 3.19\(b\)](#) that the score of the storage over 80 MWh has a stable range over 0.6.

[Figure 3.19\(c\)](#) plots the scores under 40% penetration. With a high penetration, the score drops and the over-discharge rate increases significantly. At this penetration level, only several capacities over 80 MW h can reach the score of 0.5. The result of this penetration level is implying that the generation cannot fully satisfy the demand anymore, as the average SoC in [Table 3.12](#) indicates that all capacities have a low average SoC near the low boundary of battery health. The reason why the averages are all at 20s is that the power generation cannot fully charging the battery at this penetration level. The case for delay strategy in [Figure 3.20\(c\)](#) also shows some decrease in performance, while the storage over 60 MW h has an average score around 0.55 with the rate of over-discharge grows higher. This result indicates that

the delay strategy has the ability to manage the performance when the penetration is higher, but the rate of over-discharge which introduces instability to the grid is relatively higher. The smart strategy under 40% penetration in [Figure 3.21\(c\)](#) has a lower average score than that of the delay strategy. It is only at about 0.4. However, the rate of over-discharge is lower than other two strategies.

To sum up, appropriate penetration of EVs can help to increase the performance of the grid storage and fully uses the power generation from the wind. The delay strategy shows progress on performance when the penetration is at a higher value over 20%, but it is not stable enough. It can be explained from theory that this strategy is simply delaying the charging to a later time different from the peak hours. The smart strategy performs well under the condition that the total wind power generation is large enough to satisfy the demand. It is stable because of its dynamic nature, the strategy has always made the best effort to decrease the danger to the grid.

	Normal	Score	Delay	Score	Smart	Score
10%	78 (0.51)	0.58	85 (0.56)	0.53	97 (0.64)	0.58
20%	100 (0.66)	0.7	95 (0.625)	0.68	97 (0.64)	0.7
40%	100 (0.66)	0.53	86 (0.56)	0.64	100 (0.66)	0.48

Table 4.1: The capacity of storage (MW h) at the highest score among simulations. The percent following the capacity is the ratio to the total daily consumption of one day without EV

[Table 4.1](#) shows the best score on each mode under each penetration level. Except the delay mode, other modes meet an increase on the storage capacity to accommodate more EVs. The main tendency is that the grid storage needs more than $\frac{2}{3}$ of the daily consumption as the penetration increases.

The test results of capacities larger than 100 MW h are not plotted. The reason is that there is no progress on the performance. It has been mentioned while analyzing the performance of 40% penetration and in [Table 3.12](#). It is the wind power generation that first reaches saturation point, that no matter how the capacity increases the performance can not further grow. Therefore, another conclusion is that the scale of the wind farm or the curtailment limitations should increase with the growth of storage capacity.

4.3 Limitations and problems

Originally, the objective of this project is not to model and simulate the impacts of EVs and the charging strategies. It was assumed that enough data can support the model of the energy use of a house and its appliances. Then, methodologies in machine learning and deep learning can be used to make predictions. However, searching for a data set with so many details is difficult and the existing researches on this topic have been enough many. Based on the understanding that the electrification in Orkney must introduce problems on EVs and clean power resources. The topic of modeling the impacts is selected.

For the limitations, the lack of data source has been mentioned in [Section 2.2.3.2](#), and the only available data set has some attributes not able to be used. The overall analysis of the EV impact is not useful as it is simply adding charging events to possible time points.

The simulation should have used more specific data for the appliances to characterize the randomness and instant power consumption. The lack of knowledge on power grid makes it impossible to analyze some specific attributes like loss, voltage transformation. This is also the reason why the metric is fully dependent on the current SoC, the only property that can be used.

For the wind power modeling and simulation, the *Markov Model* is a good solution. However, constrained by both time and knowledge, it is not used. It is believed that this probabilistic model can produce data with higher quality in simulation.

Additionally, the simulation framework did not take power grid specialized characters into consideration. For example, the loss of voltage conversion and the overall voltage level. If the accuracy or the degree of performance need to be improved, these details should be considered. Further more, the type of the batteries in the grid storage is not specialized. [Table 2.1](#) has summarized prevailing battery technologies. Various types of batteries should be included to compose a multilayered grid storage to support short and long term services.

However, to support these considerations, data is the most important reason. The readers can acquire an impression from this project that most of the data used in the final simulation framework is not from what was collected from the official website. Instead, only the statistics were used, and the data of the grid was generated from probabilistic methods. The statistical features can be guaranteed, while shape of the data can vary drastically from that of the reality. Therefore, most of the time, the probabilistic methods can only be used to approximate the real scenario, but nothing can be better than using carefully collected real-world data. This principle is more

effective in machine learning. In a famous paper published by *Microsoft* researchers Michele Banko and Eric Brill, very different Machine Learning algorithms were tested on a very large data set [64]. Even a fairly simple algorithm performed almost identically well on a complex problem of natural language disambiguation once they were given enough data.

Appendices

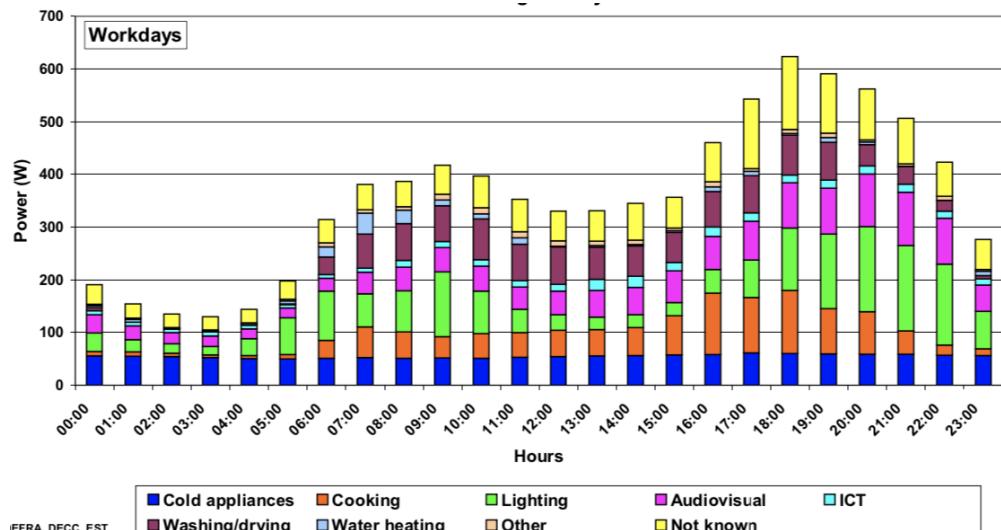


Figure 4.1: Daily average curve of all appliances

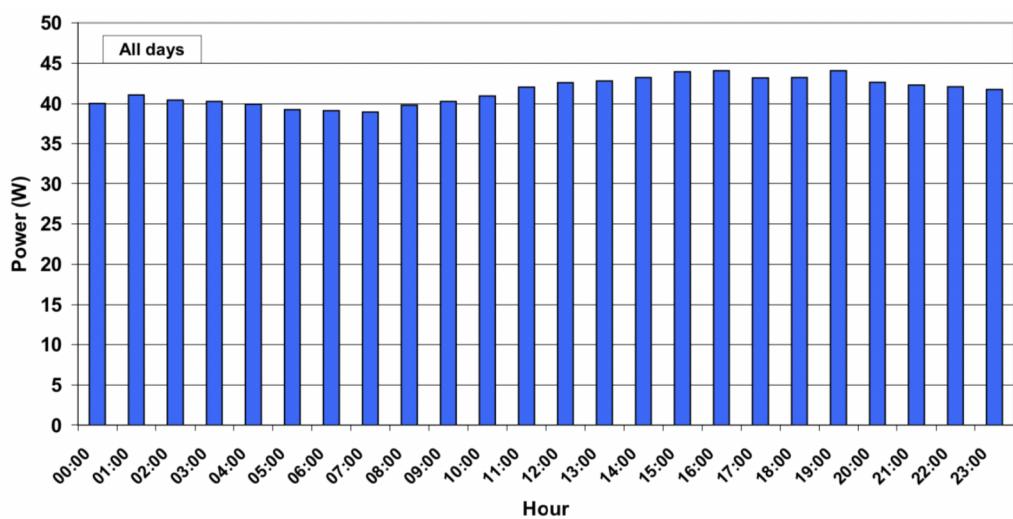


Figure 4.2: Daily average curve of refrigerator

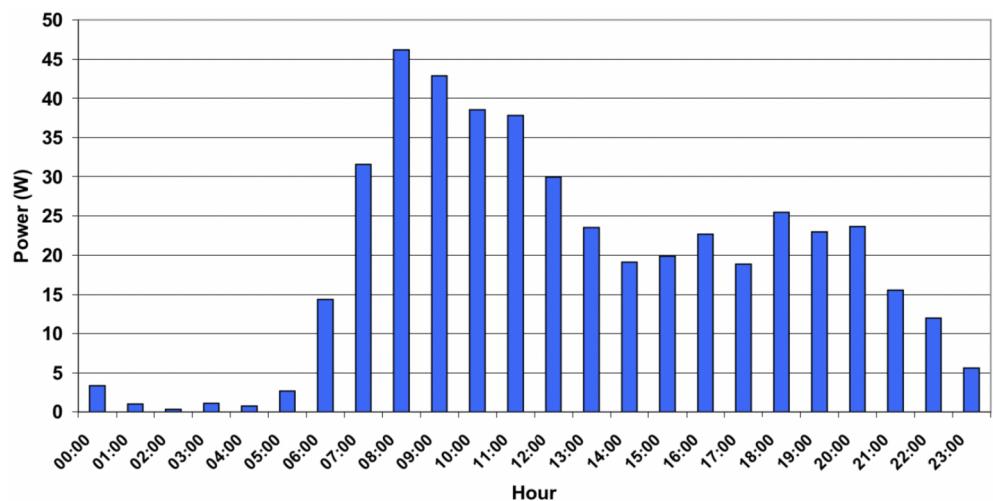


Figure 4.3: Daily average curve of washing machine

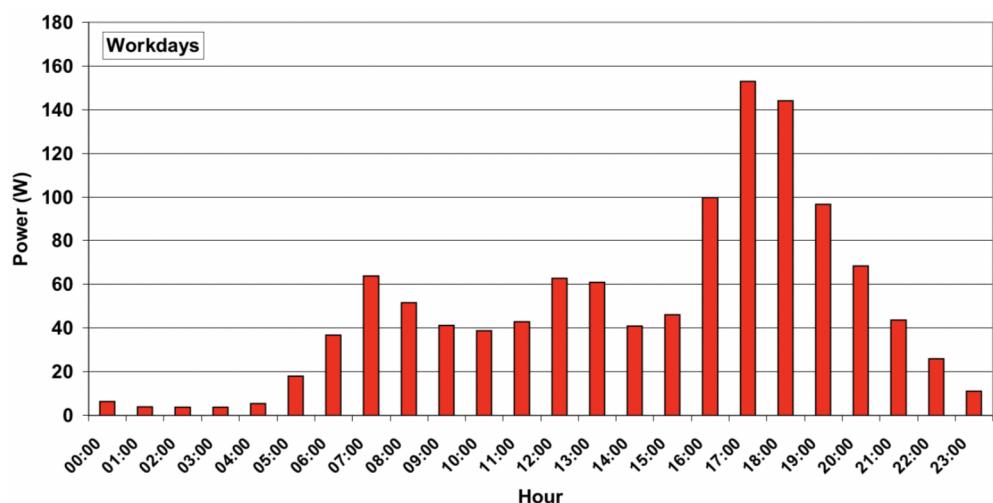


Figure 4.4: Daily average curve of cooking

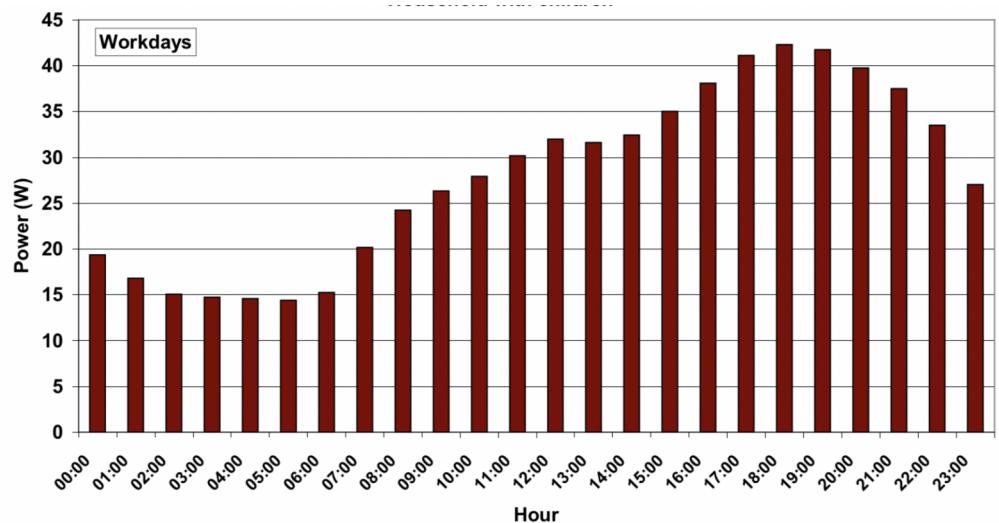


Figure 4.5: Daily average curve of computer

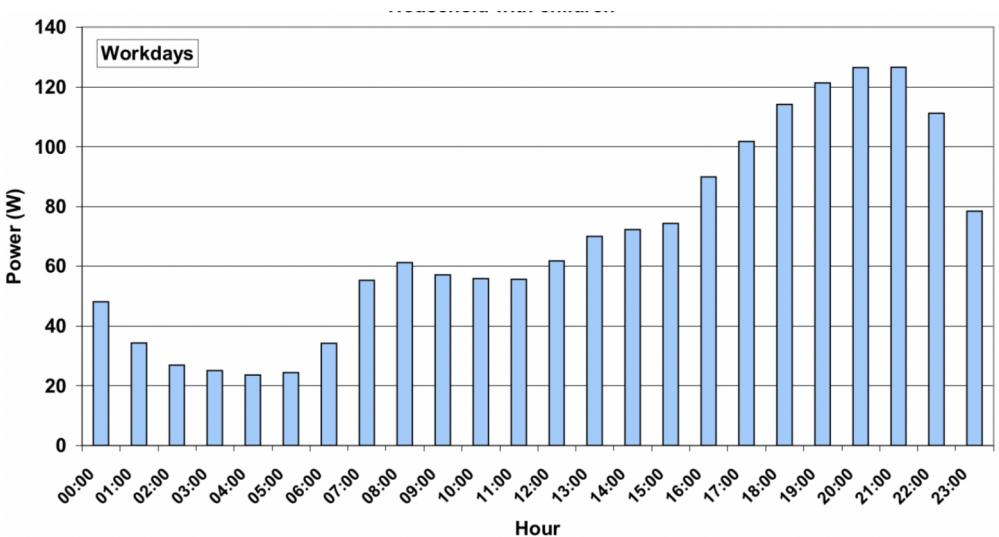


Figure 4.6: Daily average curve of audiovisual

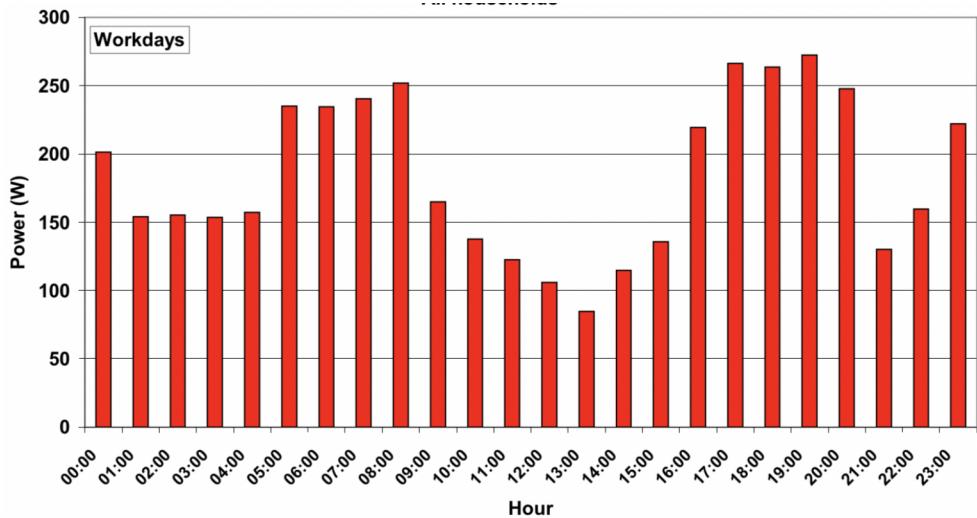


Figure 4.7: Daily average curve of heating

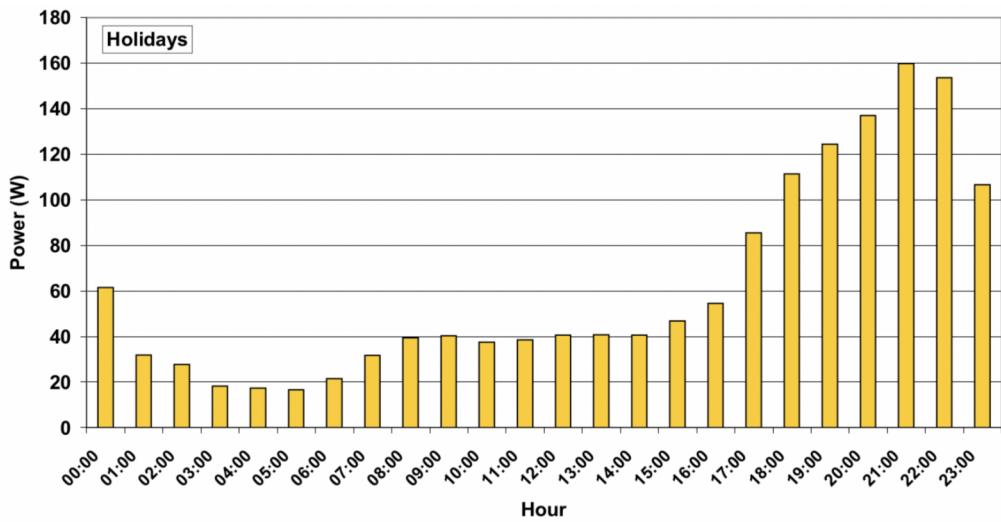


Figure 4.8: Daily average curve of lighting

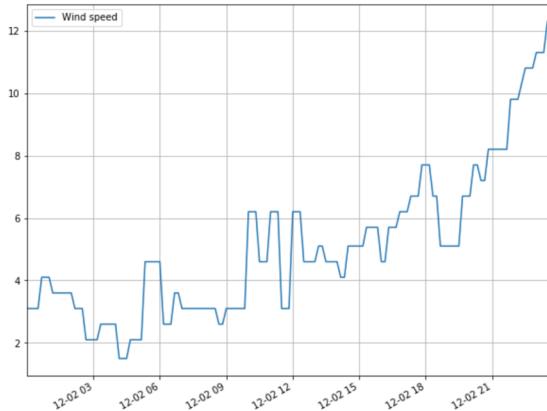


Figure 4.9: The wind speed with a typical increasing tendency

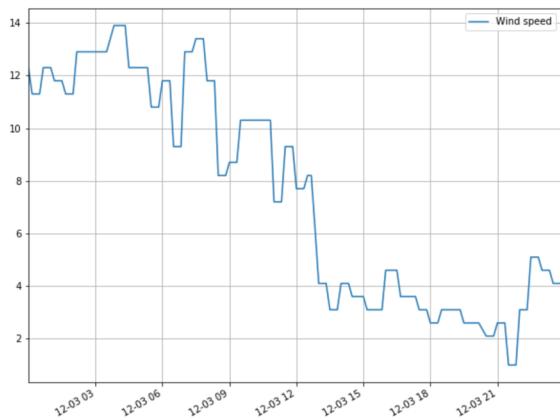


Figure 4.10: The wind speed with a typical decreasing tendency

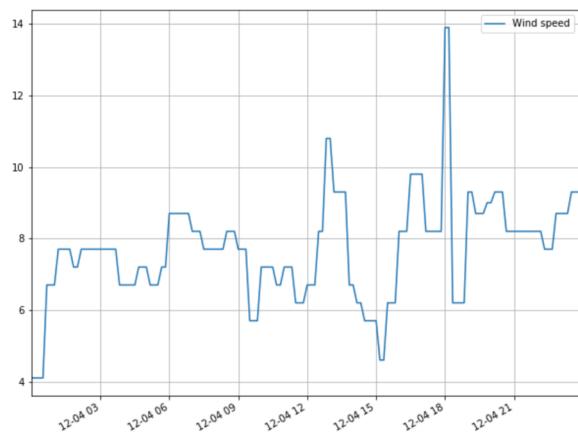


Figure 4.11: The wind speed with a high level

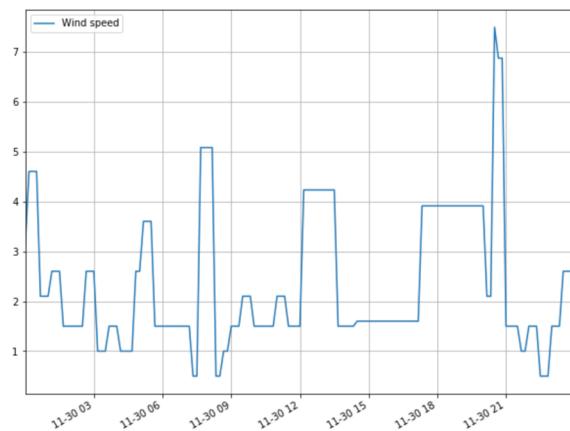


Figure 4.12: The wind speed with a low level

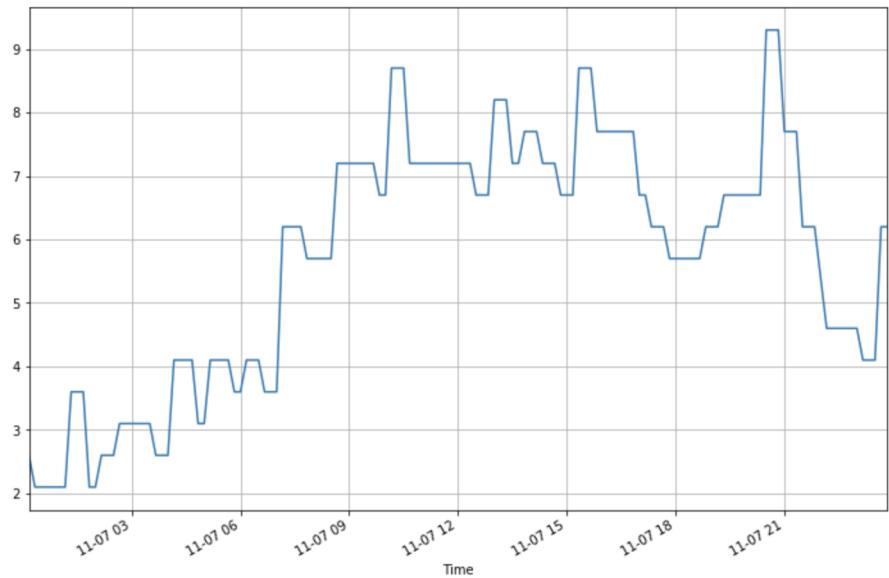


Figure 4.13: The wind speed with a level near the global mean

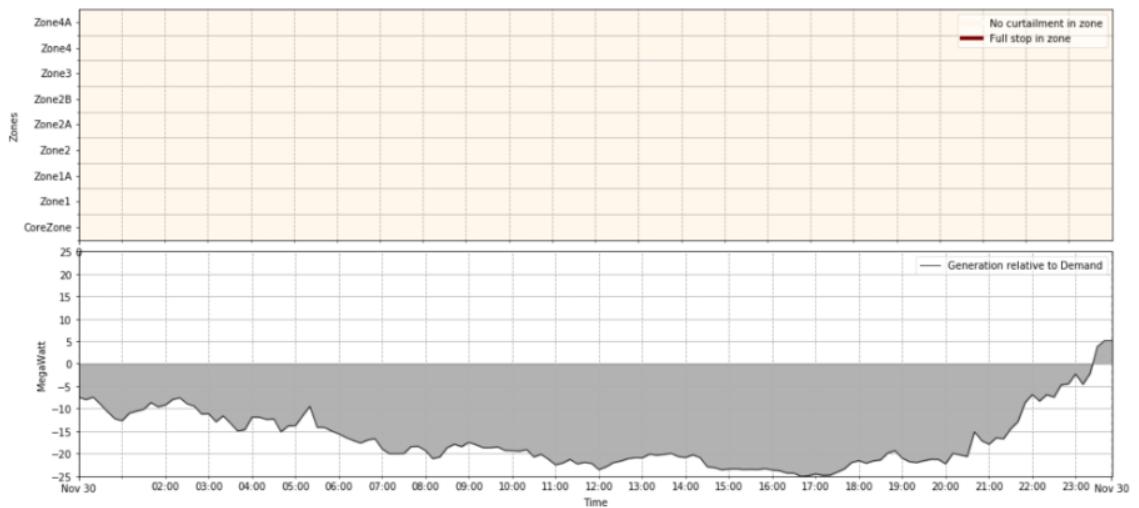


Figure 4.14: An illustration of an extreme case on November 30th. The upper chart is the condition of curtailment, where no curtailment was happened on this day. The bottom chart the subtraction of power generation and demand.

Bibliography

- [1] Abelson, Harold and Sussman, Gerald Jay. and Sussman, Julie. *Structure and interpretation of computer programs*. MIT Press, London, 1996.
- [2] Géron, Aurélien and Demarest, Rebecca. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly, Sebastopol, 2019.
- [3] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron. *Deep learning*. The MIT Press, Cambridge, MA, 2017.
- [4] DeGroot, Morris H. and Schervish, Mark J. *Probability and statistics*. Pearson Education, Inc., Hoboken, NJ, 2019.
- [5] Lippman, Stanley B. and Lajoie, Josee and Moo, Barbara E. *C++ Primer*. Addison-Wesley, Inc., Boston, 2013.
- [6] Strang, Gilbert. *Introduction to linear algebra*. Wellesley - Cambridge Press, Wellesley MA, 2016.

References

- [1] S. Ahuja, *Food, energy, and water: the chemistry connection*. Elsevier, 2015.
- [2] G. T. Miller and S. Spoolman, *Environmental science: problems, concepts, and solutions*. Brooks Cole, 2008.
- [3] J. Runyon, “6 key trends in sustainable and renewable energy,” Sep 2019. [Online]. Available: <https://www.renewableenergyworld.com/2019/02/15/6-key-trends-in-sustainable-and-renewable-energy/#gref>
- [4] X. Yu, C. Cecati, T. Dillon, and M. G. Simões, “The new frontier of smart grids,” *IEEE Industrial Electronics Magazine*, vol. 5, no. 3, pp. 49–63, Sep. 2011.
- [5] IEA, “Renewables 2018.” [Online]. Available: <https://www.iea.org/reports/renewables-2018>
- [6] E. . I. S. Department for Business, “Energy trends: Uk renewables,” Feb 2020. [Online]. Available: <https://www.gov.uk/government/statistics/energy-trends-section-6-renewables>
- [7] R. Association, “Wind energy statistics - renewableuk.” [Online]. Available: <https://www.renewableuk.com/page/UKWEDhome>
- [8] S. A. House, “Target for renewable energy now 80 per cent,” Sep 2010. [Online]. Available: <https://www2.gov.scot/News/Releases/2010/09/23134359>
- [9] “Europe’s largest onshore wind farm is switched on in scotland,” Sep 2019. [Online]. Available: <https://www.renewableenergyworld.com/2009/05/20/europe-s-largest-onshore-wind-farm-is-switched-on-in-scotland/>
- [10] M. T. Lawder, B. Suthar, P. W. Northrop, S. De, C. M. Hoff, O. Leitermann, M. L. Crow, S. Santhanagopalan, and V. R. Subramanian, “Battery energy

- storage system (BESS) and battery management system (BMS) for grid-scale applications,” *Proceedings of the IEEE*, vol. 102, no. 6, pp. 1014–1030, 2014.
- [11] P. P. Varaiya, F. F. Wu, and J. W. Bialek, “Smart operation of smart grid: Risk-limiting dispatch,” *Proceedings of the IEEE*, vol. 99, no. 1, pp. 40–57, Jan 2011.
 - [12] M. Kezunovic, J. McCalley, and T. Overbye, “Smart grids and beyond: Achieving the full potential of electricity systems,” *Proceedings of the IEEE*, vol. 100, no. SPL CONTENT, pp. 1329–1341, 5 2012.
 - [13] J. Benson and M. Labs, “Energy storage survey.” [Online]. Available: <https://www.energycentral.com/c/iu/energy-storage-survey>
 - [14] C. C. Chan, A. Bouscayrol, and K. Chen, “Electric, hybrid, and fuel-cell vehicles: Architectures and modeling,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 589–598, Feb 2010.
 - [15] G. Offer, D. Howey, M. Contestabile, R. Clague, and N. Brandon, “Comparative analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system,” *Energy Policy*, vol. 38, no. 1, p. 24–29, 2010.
 - [16] W. Kempton and S. E. Letendre, “Electric vehicles as a new power source for electric utilities,” *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, p. 157–175, 1997.
 - [17] W. Su and M. Y. Chow, “Performance evaluation of an EDA-based large-scale plug-in hybrid electric vehicle charging algorithm,” *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 308–315, 2012.
 - [18] N. DeForest, “Impact of widespread electric vehicle adoption on the electrical utility business - threats and opportunities.” [Online]. Available: https://ikhlaqsidhu.files.wordpress.com/2010/03/nr_utilities_final_8-31-09.pdf
 - [19] L. Pieltain Fernández, T. Gómez San Román, R. Cossent, C. Mateo Domingo, and P. Frías, “Assessment of the impact of plug-in electric vehicles on distribution networks,” *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 206–213, 2011.
 - [20] S. Shao, T. Zhang, M. Pipattanasomporn, and S. Rahman, “Impact of TOU rates on distribution load shapes in a smart grid with PHEV penetration,” *2010*

IEEE PES Transmission and Distribution Conference and Exposition: Smart Solutions for a Changing World, pp. 1–6, 2010.

- [21] I. Testing and Certification, “Household electricity survey final report issue 4,” vol. 1, p. 193, 2019.
- [22] K. Qian, C. Zhou, M. Allan, and Y. Yuan, “Modeling of load demand due to EV battery charging in distribution systems,” *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 802–810, 2011.
- [23] S. M. Mousavi Agah and A. Abbasi, “The impact of charging plug-in hybrid electric vehicles on residential distribution transformers,” *2012 2nd Iranian Conference on Smart Grids, ICSG 2012*, vol. 25, no. 1, pp. 371–380, 2012.
- [24] C. Roe, E. Farantatos, J. Meisel, A. P. Meliopoulos, and T. Overbye, “Power system level impacts of phevs,” in *2009 42nd Hawaii International Conference on System Sciences*, Jan 2009, pp. 1–10.
- [25] W. Su and M. Chow, “Evaluation on intelligent energy management system for phevs/pevs using monte carlo method,” in *2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, July 2011, pp. 1675–1680.
- [26] S. Distribution, “Anm generation,” 2020. [Online]. Available: <https://www.ssen.co.uk/ANMGeneration/>
- [27] O. I. Council, “Reflex orkney project.” [Online]. Available: <https://www.orkney.gov.uk/OIC-News/285m-ReFLEX-Orkney-project-to-create-a-smart-energy-island.htm>
- [28] “Project set to turn orkney into a ’smart energy island’,” Apr 2019. [Online]. Available: <https://www.orkney.com/news/reflex-orkney>
- [29] C. E. S. Aquatera, “Reflex orkney energy audit v1,” vol. 1, p. 37, 2019.
- [30] S. Turf, “Orkney hydrogen.” [Online]. Available: <http://www.surfnturf.org.uk/>
- [31] L. Watts, *Energy at the End of the World: An Orkney Islands Saga*. MIT Press, 2018.
- [32] J. P. Barton and D. G. Infield, “Intermittent Renewable Energy,” vol. 19, no. 2, pp. 441–448, 2004.

- [33] S. Vazquez, S. M. Lukic, E. Galvan, L. G. Franquelo, and J. M. Carrasco, “Energy storage systems for transport and grid applications,” *IEEE Transactions on Industrial Electronics*, vol. 57, no. 12, pp. 3881–3895, 2010.
- [34] S. Distribution, “Anm,” 2020. [Online]. Available: <https://www.ssen.co.uk/ANM/>
- [35] OpenWeather, “Open weather map,” 2020. [Online]. Available: <https://openweathermap.org/>
- [36] Python.org, “Python 3.0 release,” 2020. [Online]. Available: <https://www.python.org/download/releases/3.0/>
- [37] Gnu.org, “Bash - gnu project - free software foundation,” 2020. [Online]. Available: <https://www.gnu.org/software/bash/>
- [38] Requests.readthedocs.io, “Requests: Http for humans™ — requests 2.22.0 documentation,” 2020. [Online]. Available: <https://requests.readthedocs.io/en/master/>
- [39] Man7.org, “crontab(5) - linux manual page,” 2020. [Online]. Available: <http://man7.org/linux/man-pages/man5/crontab.5.html>
- [40] ——, “grep(1) - linux manual page,” 2020. [Online]. Available: <http://man7.org/linux/man-pages/man1/grep.1.html>
- [41] csparpa, “csparpa/pyowm,” 2020. [Online]. Available: <https://github.com/csparpa/pyowm>
- [42] GitHub, “Build software better, together,” 2020. [Online]. Available: <https://github.com/>
- [43] Git-scm.com, “Git,” 2020. [Online]. Available: <https://git-scm.com/>
- [44] Pexpect.readthedocs.io, “Pexpect version 4.8 — pexpect 4.8 documentation,” 2020. [Online]. Available: <https://pexpect.readthedocs.io/en/stable/>
- [45] Numpy.org, “Numpy — numpy,” 2020. [Online]. Available: <https://numpy.org/>
- [46] Scipy.org, “Scipy library — scipy.org,” 2020. [Online]. Available: <https://www.scipy.org/scipylib/index.html>

- [47] Pandas.pydata.org, “Python data analysis library — pandas: Python data analysis library,” 2020. [Online]. Available: <https://pandas.pydata.org/>
- [48] Scikit-learn.org, “scikit-learn: machine learning in python — scikit-learn 0.22.1 documentation,” 2020. [Online]. Available: <https://scikit-learn.org/stable/>
- [49] Matplotlib.org, “Matplotlib: Python plotting — matplotlib 3.1.2 documentation,” 2020. [Online]. Available: <https://matplotlib.org/>
- [50] Jupyter.org, “Project jupyter,” 2020. [Online]. Available: <https://jupyter.org/>
- [51] En.wikipedia.org, “Type i and type ii errors.” [Online]. Available: https://en.wikipedia.org/wiki/Type_I_and_type_II_errors
- [52] G. J. S. Harold Abelson and J. Sussman, *Structure and interpretation of computer programs*, 2nd ed. Cambridge, Mass: MIT Press, 2010.
- [53] NASA, “nasa/trick.” [Online]. Available: <https://github.com/nasa/trick>
- [54] G. Haines, A. Mcgordon, P. Jennings, and N. Butcher, “The simulation of vehicle-to-home systems—using electric vehicle battery storage to smooth domestic electricity demand,” *Proc. Ecologic Vehicles/Renewable Energies - EVRE*, 01 2009.
- [55] W. Su, H. Eichi, W. Zeng, and M. Y. Chow, “A survey on the electrification of transportation in a smart grid environment,” *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 1–10, 2012.
- [56] Sae.org, “J1772a: Sae electric vehicle and plug in hybrid electric vehicle conductive charge coupler - sae international.” [Online]. Available: https://www.sae.org/standards/content/j1772_201710/
- [57] E. E. Limited, “Electric vehicle charging behaviour study final report for national grid eso,” vol. 1, p. 55, 2019.
- [58] En.wikipedia.org, “Chebyshev’s inequality.” [Online]. Available: https://en.wikipedia.org/wiki/Chebyshev%27s_inequality
- [59] G. Bowden, P. Barker, V. Shestopal, and J. Twidell, “The weibull distribution function and wind power statistics,” *Wind Engineering*, vol. 7, no. 2, pp. 85–98, 1983. [Online]. Available: <http://www.jstor.org/stable/43749036>

- [60] A. Halevy, P. Norvig, and F. Pereira, “The unreasonable effectiveness of data,” *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 8–12, March 2009.
- [61] C. Liu, K. T. Chau, D. Wu, and S. Gao, “Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies,” *Proceedings of the IEEE*, vol. 101, no. 11, pp. 2409–2427, 2013.
- [62] En.wikipedia.org, “Normal distribution.” [Online]. Available: https://en.wikipedia.org/wiki/Normal_distribution
- [63] K. Yao and J. Gao, “Law of large numbers for uncertain random variables,” *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 3, pp. 615–621, June 2016.
- [64] M. Banko and E. Brill, “Scaling to very very large corpora for natural language disambiguation,” in *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, ser. ACL ’01. USA: Association for Computational Linguistics, 2001, p. 26–33. [Online]. Available: <https://doi.org/10.3115/1073012.1073017>