Norwegian School of Economics
Bergen, Spring, 2020

NHH

Empirical Comparison of Load Forecasting Methods for Skagerak Energilab

A Perspective of the Operational and Economic Efficiency Gain as a Result of Increased Forecasting Accuracy in a Microgrid Environment

Emil Karsten Skarholt & Alexandr Vornicov

Supervisor: Lars Jonas Andersson

Master Thesis, Economics and Business Administration

Major: Business Analytics & Energy, Natural Resources and the Enviroment

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

First and foremost, we would like to acknowledge and thank professor Lars Jonas Andersson for helpful guidance throughout the research and writing of this master thesis. Additionally, we would like to thank professor Mette Bjørndal for clarifications and guidance on the power market perspective of the analysis. We would also like to express our gratitude to Signe Marie Oland at Skagerak Nett for excellent support and cooperation with the thesis by supplying data, technical and general perspectives of the research project. Lastly, we would like to express thankfulness to our friends and family for support, feedback, and motivation throughout the process of writing this master thesis.

Norwegian School of Economics

Bergen, June 2020

Abstract

This master thesis is analyzing short-term load forecasting. Power consumption in kW will be forecasted 24 hours ahead, for each day of a week and finally averaged to derive mean performance. The forecast will be conducted by selected methods and models and compared against a simple yet reasonable benchmark model. To evaluate the performance in detail, we select to compute MAPE values for each individual hour, day and average over one week. In addition, we construct a tailored evaluation metric to estimate the economic consequences of inaccurate load forecasts. This master thesis is intended to provide a theoretical and empirical link between contemporary forecasting techniques and actual economic benefits that can be derived from improved accuracy of load forecasts at Skagerak Energilab.

Obtained results show a tendency of increased forecasting accuracy when utilizing machine learning algorithms with Neural Network structures. However, no single method could outperform an ensemble average model. Compared to the benchmark model, our proposed Ensemble consisting of BATS, seasonal ARIMA, and a multivariate AR ANN increased forecasting accuracy by a notable degree. Also, improved performance was shown to result in a decreased direct economic cost.

Keywords – NHH, Master Thesis, Forecasting, Multi-Step Forecasting Machine Learning, Deep Learning, Time Series, Electricity Market, Economic Analysis, Microgrid

Table of contents

List of Figures

List of Tables

List of abbreviation

Abbreviation	Explanation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AR ANN	Autoregressive Artificial Neural Network
DER	Distributed Energy Resources
DNN	Deep Neural Networks
DSO	Distribution System Operator
ELFE	Economic Load Forecasting Error
kV	Kilovolt
kW	Kilowatt
kWh	Kilowatt Hour
LF	Load Forecasting
LFE	Load Forecasting Error
LSTM	Long-Short-Term Memory
LTLF	Long-Term Load Forecasting
ML	Machine Learning
MTLF	Medium-Term Load Forecasting
MW	Megawatt
MWh	Megawatt Hour
NN	Neural Networks
PEC	Percentage Economic Cost
PV	Photovoltaic
RE	Renewable Energy
RES	Renewable Energy Source
RNN	Recurrent Neural Networks
STLF	Short-Term Load Forecasting
TSO	Transmission-System Operator
VSTLF	Very Short-Term Load Forecasting

1.0 Introduction

The Introduction section intends to give an overall idea about the thesis. It familiarizes the reader about the topic and the main goal of the study, while further providing preliminary perception about the contribution of the thesis. Lastly, to prepare the reader on what will be presented further into the study, a brief outline of the thesis structure will be exhibited.

1.1 General Introduction

Electricity is a commodity that is readily available in most parts of the western world. It is a product that differs from other commonly tradable goods as it is not readily available to store in larger quantities and for the most part, must be consumed at the moment as it is produced. This distinct feature has direct ramifications on the electricity market, creating an essential need for real-time balance in supply and demand. As Renewable Energy Sources (RES) and Distributed Energy Resources (DER), known for intermittency, increase their presence in the generation mix, the electricity market experiences further complications.

To bring clarity and some degree of certainty in such a complex system, load forecasts serve as the primary tool for all power market participants. This instrument, within the field of forecasting, directly focuses on the objective to analyze, decompose, and estimate future electricity demand (load) in a selected customer area. In this thesis, we intend to analyze Skagerak Energilab, a newly developed microgrid in Skien, Norway. The main focus will be set on short-term load forecasts within the confines of the mentioned microgrid. Further, relating the obtained forecasts to the existing structure of the power market, analysis of the potential economic effect would be quantified and discussed.

Different models have been constructed in an attempt to fulfil the intended objective. Statistical, as well as more advanced AI-based methods, have been tried and tested when developing the multi-step forecasts, and revealed both strengths and weaknesses that were consequently analyzed. As a benchmark, it was opted to use seasonal naïve -simple yet reasonable model that sets a threshold for further improvement. Forecasting the load at such a small resolution has proven to be a challenging task, due to volatile nature and sudden changes in electricity consumption. The best approach for this task was to combine different individual forecasts.

1.2 Contribution of Thesis

This master thesis intends to evaluate and compare a wide variety of forecasting methods and models, in a newly developed and compact microgrid environment called Energilab, operated by Skagerak Energi. The fundamental research contribution is to determine a forecasting method that can systematically predict electricity load[1] with a sufficient level of accuracy. The core empirical intention of the thesis is to identify the underlying statistical problem in which Energilab is faced by its customer's load pattern. Further, we present an appropriate method to solve the problem by identifying to what extent the statistical errors can be lowered, with additional interest linked towards establishing the economic efficiency gain. While the empirical analysis is performed in a small customer setting, the structural methodology is scalable and can be applied to different sizes of microgrids.

1.3 Outline of Thesis

The initial chapter 2 of the thesis covers background material of the Nordic electricity market and structure for the economic analysis. Further, the material will provide details encompassing microgrids and Energilab specifically. In addition to background material for load forecasting, aspects of factors affecting load, the economic value of the forecasts and finally, a perspective from the literature will be provided.

Chapter 3 describes and analyze the data before the actual forecasting task. Data is preprocessed in terms of missing data imputation, one-hot encoding and further aligned correctly in relation to granularity. Descriptive analysis and identification of the load curve and economic data are provided to identify important nuances before the forecast task was conducted.

Chapter 4 introduces the methodology section with in-depth information about the methods proposed in the thesis. The methodology section initializes choices made about cross-validation while further establishing how we decide to approach the evaluation of the forecasting models

[1] Also known as electricity demand.

and methods. The methodology section then presents the different methods and models, ranging from simple, too time-consuming, complex deep learning algorithms.

Chapter 5 presents the results from each method proposed and summarizes the performance of each model. The reader will be provided with a detailed deconstruction of the results based on our established model evaluation criteria, as presented in the methodology section.

Chapter 6 provides a discussion of the forecasting results and a perspective of the operational and economic efficiency gain from the results of the thesis. Finally, chapter 7 concludes the thesis and gives an outlook for future work.

2.0 Background

In the following section, we will give background knowledge relevant to the task in hand. Initially, this section presents the Nordic energy market and its design, which is split in two parts, technical and economic structure. Analysis of microgrid development in the market and the newly launched project Skagerak Energilab will be exhibited. The section will then present load forecasting (LF), factors affecting it, and different categories of forecasting. Finally, the economic perspective of LF will be given and establish a base for the discussion.

2.1 Technical Structure of the Nordic Power Market

From a traditional point of view, electricity is generated in substantial quantities from multiples of large generators across the country, which is then distributed towards the end-users. The physical attributes of the product in which the end-user is consuming are, however, a special kind of product, in which it must be consumed at the same time it is generated to be in perfect balance (Olje og Energidepartementet, 2020a). In Norway, this balance is settled upon a frequency in the grid of 50 Hz, with a normal variation of +/- 0.1 Hz (Olje og Energidepartementet, 2020b). In the case of a period where generation is higher than consumption, the frequency will rise, contrary, if consumption is higher than the generated quantity, the frequency will drop below 50 Hz. If the frequency is not kept at the established level, it can cause major overheating and power failures. The result could be substantial socio-economic consequences caused by a failure in electric appliances, machinery and increased maintenance and operational cost for the participants in the electricity market.

At the top of the technical hierarchy, we have the transmission system operator (TSO). The TSO is in charge of the transmission lines that connect the power generators and the end-users across the whole of Norway. In Norway, this capacity is also connected with countries outside of Norway (Norsk Vassdrag og Energidirektorat, 2020). TSOs operates the most powerful transmission lines, mainly the 420, 320 and 132 kV voltage lines and has the responsibility to maintain the quality of frequency of 50 Hz in the grid (Olje og Energidepartementet, 2020b). The Norwegian TSO is Statnett.

Moving beyond the TSO, the next in line of the hierarchy is the Distribution- and regional system operators (DSO). The regional grid operates in many similar ways like the TSO grid, but it is geographically enclosed to one region, maintaining the connections from transmission grids with the distribution grid (Norsk Vassdrag og Energidirektorat, 2020). The distribution grid is the lowest level of the hierarchy and distributes the power out to the end-users, like households, businesses and cottages. Within the distribution grid, there is a technical difference between high- and low voltage lines. High voltage lines have a voltage above 1 kV, while the low voltage lines are the lines that serve energy in the form that the end-users consume, at a voltage typically between 230 V and 400 V (Norsk Vassdrag og Energidirektorat, 2020). Within the Skagerak Energi group, Skagerak Nett is the DSO for Vestfold and Telemark county, where Skagerak Energilab is located.

2.2 Organization and Economical Structure of the Nordic Power Market

The power market can be separated into two categories, wholesale and end-user market. The wholesale market is a collection of power producers, brokers, power suppliers and large industry customers (Olje og Energidepartementet, 2020a). Within the wholesale market, there are large quantities of power, where notably power suppliers trade with the interest to supply small to medium-sized households and smaller-scale businesses. The said supply is what is known as the end-user market. A simplified schematic of the power market can be seen in Figure 1.

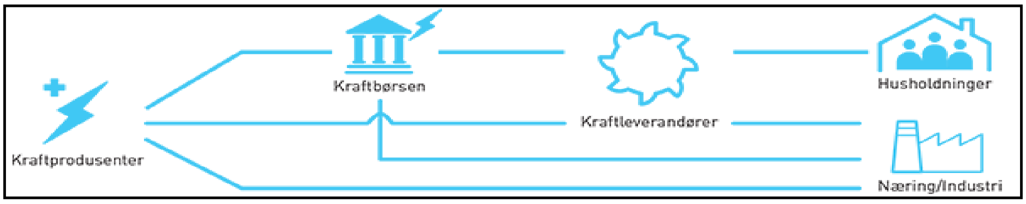

Figure 1 Schematic Representation of the Power Market. Source: (Olje og Energidepartementet, 2020a)

Within the wholesale market there are also three distinct marketplaces as displayed in Table 1:

Table 1 Marketplaces in the Nordic Power Market

Marketplace	Responsible Party	Settlement period
Day-ahead (Elspot)	Nord Pool	12:00 AM
Intraday	Nord Pool	One hour ahead of consumption
Balancing market(Real TIme)	TSO (Statkraft)	Intra Hour

The day-ahead market is the main marketplace for power trading in the Nordic region (Olje og Energidepartementet, 2020a). In this market, physical contracts for the supply of power, hour-by-hour the next day are traded with a corresponding spot price that is settled in the equilibrium of supply and demand of electricity on the Nord Pool power Exchange (Statnett, 2020a). Coupled with much of the day-ahead market in Europe, it becomes a large auction that closes at 12:00 AM each day. Following this closing of the auction, the prices for each hour, next day, will be calculated based upon all the purchase and sell orders received, and the transmission capacity available creating what is known as unit commitment (UC) (Olje og Energidepartementet, 2020a).

Uncertainty plays a significant role in the market, as well. There will always be uncertainty connected to the supply of power, but also, the consumption of power. The day-ahead market is based on forecasts for both production and consumption the following day, and unknown information that occurs after the day-ahead market closes, like updated weather forecasts, must be accounted for in the intraday market. In the intraday market, a participant can trade and correct their bids, given new information on production or consumption. Contracts are continuously traded from the closing of the day-ahead market up until one hour of operation (Olje og Energidepartementet, 2020a). Both the day-ahead and intraday market are traded through the Nord Pool power exchange and are known as the market for planned energy.

After the day-ahead and intraday market closes, the role in creating balance between production and consumption and persisting the 50 Hz frequency within the hour of operation, is handed over to the TSO, Statnett (Statnett, 2020b). If imbalances occur, Statnett utilizes three different levels of power balancing reserves to maintain instantaneous balance, with their own respective response time. As observed in Figure 2, Primary reserve (FCR) and secondary reserves (FRR-A) are both reserves that activate automatically in response to imbalances and the amount of time these imbalances occur. Tertiary reserves (FRR-M) are manually activated by the TSOs and have an activation time of 15 minutes (Olje og Energidepartementet, 2020a). These reserves are also linked to a marketplace called the regulation power market and are known as the market for unplanned energy. In the market for unplanned energy, Statnett can either up-regulate, that is supply more energy through the mentioned reserves with a price higher than the corresponding spot price to incentivize new generation. Down-regulation, the opposite case, draws energy from zones where the supply is too high, to alternative zones where supply is too low to maintain balance. Down-regulation is incentivized by pricing the down-regulated energy at a price lower than the corresponding spot price, so that generators with a higher marginal cost would profit from purchasing energy rather than produce it themselves. These prices are set in terms of the dominating market volume, thus, if up-regulation is dominating the regulation market in terms of volume, the up-regulating price will be higher than the spot price, subsequently maintaining the down-regulation price at spot price level. Conversely, if down-regulation market is dominating, the price will be set at a lower rate than the spot market, holding up-regulation price at the spot price level. However, if there are no dominating market, then all three prices are balanced at the spot price level.

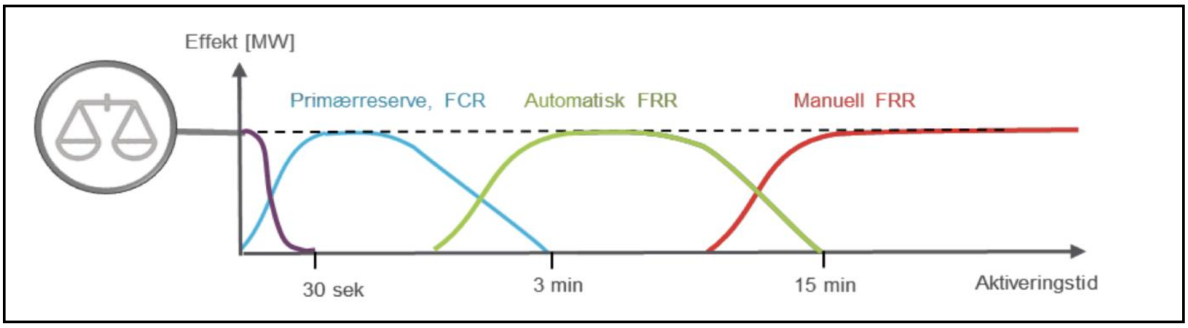

Figure 2 Regulating Market Reserves. Source: (Olje og Energidepartementet, 2020a)

Besides the physical power market, there also is the financial power trading market. This marketplace is mainly used for risk management and speculation (Olje og Energidepartementet, 2020a), and does not involve physical power trading. In this thesis, we will solely make use of the physical power market to perform an economic analysis of load forecasting results.

2.3 Microgrid Development

Traditional energy production and distribution systems were characterized by the top-down structure, where electricity flowed from large transmission-connected generation to a passive distribution network. Nowadays it is noticeable that the electricity flow is becoming more dynamic. Necessities to reduce electricity costs, improve resilience, curb CO2 emissions and provide reliable power supply are some of the driving factors that bring conceptual changes in conventional energy production and distribution systems. One of the most significant changes is the rapid increase in the number of distributed energy resources (DER). DER are small-scale power generation sources, mainly wind and solar or controllable loads that are located close to final users and are mainly distribution-network-connected. Large penetration of such generation technologies and the necessity to perform control and management of electrical systems at a much higher resolution facilitated the emergence of a concept called "microgrid".

According to the U.S. Department of Energy Microgrid Exchange Group, the following criteria defines a microgrid:

> *"a microgrid is a group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island-mode."* (Berkeley Lab, 2019)

For industrialized countries, microgrids are subject to coexistence within a mature "macro grid" that features gigawatt-scale generating units, thousands of miles of high voltage transmission lines, minimal energy storage[2], and carbon-based fossil fuels as a primary energy source[3]. According to (Hirsh, Parag, & Guerrero, 2018), factors driving microgrid development in these countries fall into three broad categories: Energy Security, Economic Benefits, and Clean Energy Integration. The degree to which particular category drives the advancement of microgrids changes from place to place. In the United States, energy security and ability to provide power reliability for "critical facilities" such as hospitals, water and waste treatments, in case of electricity outage or natural disaster motivates the development of microgrids. In the case of Europe, it is mostly a need for clean energy and environmental concerns.

For some developing countries, limited access to reliable electricity hinders human potential development and constraints economic growth. Connecting scarcely populated reclusive areas to the main centralized grid represents a challenge for many countries as the cost of building infrastructure is burdensome. Thus, a large portion of the population in developing countries lives with limited or no access to electricity at all. In this case, microgrid emerges as a feasible solution. Remote microgrids combining clean DER and storage, in some cases facilitated by innovative mobile payment platforms, can provide a lifeline to those people, allowing children

[2] As a result of large hydropower reserves, this is usually not the case for Norway.

[3] Again, not a Norwegian case.

to study at night, medical systems to provide reliable service and entrepreneurs to improve their livelihoods (Hirsh, Parag, & Guerrero, 2018). These innovations can allow developing nations to potentially leapfrog to a world of microgrids, in the same way, that mobile communications allowed them to connect to each other and the outside world without building up extensive landline networks (Hirsh, Parag, & Guerrero, 2018).

A wide range of possible applications bundled with the virtues of technological developments foreshadows the increasing popularity of microgrids in the future. Though providing some advantages as a possible reduction in electricity costs, integration of non-dispatchable energy sources and reduction of CO2 emissions adds value to the concept of microgrid, at the same time it poses some challenges. For instance, while operating in a grid-connected mode, DER in microgrids introduce some different operating conditions that require flexible resource capabilities in conventional power plants. As it was reported by California ISOs[4] net load, that is the difference between forecasted load and expected electricity production from non-dispatchable generation resources, can experience extreme oscillations during the day due to the intermittent nature of power production from DER. This phenomenon was given an industry moniker of "The Duck Chart", where the daily net load resembles a duck figure with a significant decrease in net load during the day due to a large degree of solar production from 8 AM until 8 PM[5]. Highly varying net load during the day amplifies the need for conventional plants with ramping flexibility and the ability to start and stop multiple times per day. Thus, if microgrids emerge as an entity that has the potential to manage net load, it is vital to establish good forecasting practices on an appropriate scale and assess economic benefits. Hence it is timely, appropriate, and reasonable to analyze load forecasting in a microgrid setting in order to ensure better operations and grid safety.

2.4 Skagerak Energilab

Microgrids are developing in the market today. Skagerak Energilab is one of them. Launched summer 2019, it represents one of the most recent developments within the Norwegian power market. The system is developed as a pilot and is supposed to aid Skagerak Energi in its focus on the future electricity market. At the same time, the project will provide the company with insightful knowledge of the technical, operational, and regulatory aspects behind systems on

[4] ISO is commonly known as North American regional transmission operators (RTO)
[5] An example of the duck chart can be exhibited in the Appendix

this scale. Energilab, as we will name the system for ease of reading, is also connected with the main grid, thus requiring an in-depth analysis of the market coordination to be able to operate most efficiently in the well-established traditional grid network.

Energilab, located on Skagerak Arena, is a football stadium in Skien, Norway. The system is a combination of 5000 m^2 PV-panels on the roof of the stadium and has, in combination, a 1 MWh capacity battery. These two energy sources are connected to two substations, in which further supplies the local low-voltage grid. Today the battery mainly revolves around shedding the peak loads in the area, which on a routine basis occurs in periods where the stadium is powering their floodlights during football matches, an operational period when the load increases tenfold (Skagerak Energi, 2019) — a schematic figure of the system is presented below.

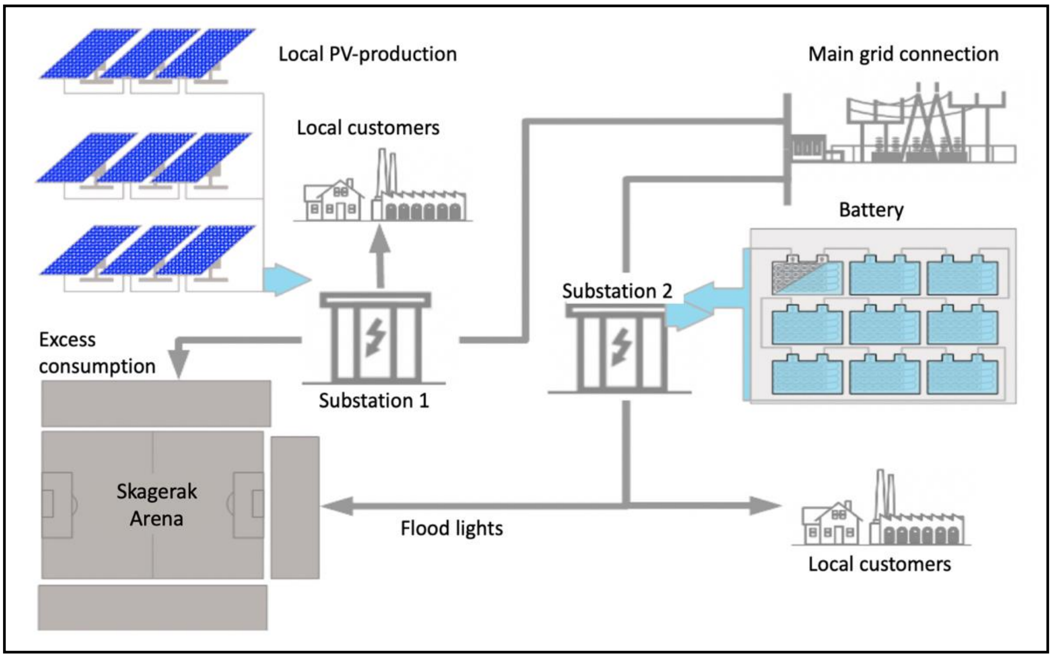

Figure 3 Techincal Structure of Skagerak Energilab. Source: (Skagerak Energi, 2019)

As represented by Figure 3, the PV-system connects to substation 1. This comes as a result that the load patterns connected to that substation coincide with the expected generation curve from the PV-system. Further, the power output from the floodlights is drawn from substation 2; thus, from an operational point of view, this was the most efficient solution[6]. Finally, both substations join together at a larger transformation station, which is then connected to the

[6] From a data collection point of view, electricity load data from customers connected to Energilab through substation 1 and 2, are aggregated together in the same data points.

distribution grid. Currently, Skagerak Nett is performing simulations at the project, to observe and identify if Energilab excels further than only shedding the floodlight peaks, specifically in terms of load shedding, island mode operation, load shifting or battery charge and discharge processes. These simulations usually run across one week.

Pointing out the primary intention of this thesis, Skagerak Energi is currently supplying over 200.000 customers with electricity (Skagerak Energi, 2020), in their operating region. From the perspective of Energilab, the number of consumers connected to this microgrid is a small percentage of these total number of customers. Also, the clientele is not homogeneous, as it consists of 42 residential buildings, nine businesses and two cooperative residential buildings. We will examine LF and the potential economic value of these forecasts on the specified microgrid but emphasize that the methods proposed can be scaled up.

2.5 Categories of Load Forecasting

Within LF, there are currently two distinct categories. These two distinctions are power systems planning and power systems operation (Shahidehour, Yamin, & Li, 2002), which in broader terms, relies upon the forecasting horizon that is of interest.

Power system planning is, within the field of LF, the category where the main objective is to support and aid in the long-term decision-making concerning grid investments and significant trends in the market. The forecasting horizon typically spans from 1 to 10 years ahead and represents the aggregated projection of the marketplace to come. These forecasts have little to no value for the day-to-day operational perspective of the power producers.

Alternatively, power system operation is the opposite perspective within these two categories of LF, where the main objective is to maximize the accuracy[7] in the short-term horizon. These forecasts span from seconds beforehand, up to 168 hours, or seven days in the future (Shahidehour, Yamin, & Li, 2002). These forecasts are not suitable for grid planning and decision-making for long term development of the power system, but rather an operational tool for day-to-day management practices.

[7] By accuracy we mean the degree of closeness of the forecasted value compared to the actual realized value

Besides the points leading towards grid investment discussed above, evidence also prove that accurate LF leads to increased efficiency for the distribution companies, by helping the planning process of operation to supply all connected customer with reliable electricity every hour of the day. LF aids the decision process of maintenance within the power system, by understanding the demand at the area of interest which would lead to as low as possible impact on customers. Secondly, LF also minimizes the level of risk within the company, by proceeding to induce more well-informed decisions. Thirdly, from the perspective of a grid-connected microgrid, LF aids a more operational efficient decision-making process in terms of energy supply towards and from the microgrid. In this thesis, we will focus solely on the power system operation category when discussing LF.

2.6 Factors Affecting Load Patterns

The initial step, when designing an efficient and accurate LF model, is to build a good understanding of the underlying factor and characteristics of the system. Electricity load is built upon a set of different factors that influence the behavior of the customers. Availability and high-quality data regarding these factors are essential to have access to, to minimize forecasting errors. The said factors can broadly be characterized into economical, time, weather and random disturbances (Shahidehour, Yamin, & Li, 2002). Naturally, the availability of historical load data would also be critical.

Economic conditions within a specified area could have an impact on the load pattern. Possible factors like these are demographic conditions, income and type of customers, industrial activities and population. These factors typically move in a more long-term perspective and would thus be more relevant for LTLF. Electricity prices also represent a complicated relationship in terms of LF[8].

Time is an important variable to consider in all scenarios of LF. Seasons, weekdays, and holidays affect the load pattern in multiple ways. Seasonality affects the load pattern by the number of daylight hours within a day. Weekdays affect the load patterns of industries and

[8] As identified by (Holstad & Pettersen, 2011), Norwegian electricity consumers short-term price elasticity is close to zero. Thus, we conclude that prices will not play as a significant feature in our short-term load forecasts and will not consider it further as a feature in our forecasting models.

commercial firms by lowering their activity during the weekend, while residential load patterns differ within the weekend by having a different set of routines during these days. Holidays or significant events also affect the load pattern greatly by lowering or increasing the overall load below or above typical values.

As temperatures rise and fall, it impacts the level of energy needed to power air-conditioning in the summer heat and heating during the cold winter months. There are also factors like humidity, precipitation, wind and solar radiation within the day that affects the load.

Random disturbances will always be an essential risk when considering LF. These random events much depend on the size of the customer and their activities. Large industrial firms can suddenly experience unexpected load changes due to shutdowns or operational difficulties. In the UK, observations have been made that popular TV-shows, and their viewers demand tea, at the same time, has caused such a sudden change in the load, that the system nearly collapsed from the unanticipated electricity demand.

In addition to the factors explained above, it has also been widely accepted in the literature that lagged values of electricity load are able to predict the future well. Inclusion of these values in the model thus could become an important feature to consider.

2.7 Economic Value of Forecasts

The successful and economically efficient operation of electricity markets is a complicated task. Distinct nature of electricity as a commodity requires a constant balance between generation and consumption without the possibility of storing energy (in substantial amounts) for later use. Thus, load forecasts are an important prerequisite for unit commitment (UC), security analysis, planning of power development, and many other vital decisions in the power market. In an ideal setting, the day-ahead load forecasts would exactly match the real-time load, and optimal dispatch of energy would be achievable. However, as discussed earlier, energy consumption is highly stochastic, and thus load forecasting will never be an exact science. The consequence of this reality is that load forecasting errors (LFE) have direct economic implications and result in increased operational costs.

Generally speaking, besides an actual possibility of power blackout[9] due to mismatch of consumption and generation, even small errors in estimated load, either positive or negative, would cause suboptimal UC. On one hand, in the event of a positive LFE, where models predict a higher load than the actual real-time load, it would result in spoilage and unnecessary commitment of units that had to be down regulated and sold elsewhere. On the other hand, if the models were to result in negative LFE, the predictions would be too low, and the corresponding supply would be inadequate, resulting in up-regulation and increased costs to supply more energy than expected. Consequences of such inaccuracies are that the costs of over or under-contracting in a day-ahead market and then selling or buying power on the balancing market can lead to the financial distress of the utility company. Moreover, as it was noticed by Sangrody and Zhou (2016), errors in both directions resulting in economic losses, however, the prices tagged to different error directions are often different. Thus, minimization of electricity volumes traded on the balancing market is an important economic objective for many power exchange participants.

As a result, LF has gradually become the central and integral process in the planning and operation of electric utilities. Existing research usually explores advanced forecasting techniques for reducing statistical errors, although these models have been shown effective in improving accuracy, it is seldom that these studies demonstrate if, and how much such improvements in load forecast accuracy might bring economic value to power system operations, or any other market participants. Being grid-connected, the Energilab microgrid provides a potential case study for this purpose, utilizing load forecast within the framework of the actual physical power market would allow seeing the connection between accuracy and economic value[10]. Nevertheless, the cumulative socio-economic benefit of improved forecasts is difficult to estimate as load predictions are integrated into almost all of the steps in the power exchange process, therefore it is important to determine the exact setting and case of forecast usage.

[9] In the event that the grid is overloaded or heavily undersupplied by power, blackouts could occur, resulting in periods where power is not being supplied and customers are left without energy.
[10] Comprehensive description of proposed approach would be provided in chapter 4.2.

2.8 Literature Review

LF has attracted researchers since the 1960s (Lin & Santra , 2019) but development in computing power in the last decade has revived the LF field by utilizing advanced machine learning techniques and computational expensive artificial intelligence methods. The literature review will primarily surround the said topic, and thus build an overview of the different proposed techniques and considerations to provide a solid background material. The literature review will initially overview the available STLF methods, in addition, different perspectives on feature selection will be presented. A review of the literature surrounding the economic potential of LF will be presented at the end.

Since the 1960s, STLF initially was solely driven by standard statistical methods like linear and non-linear regression (Papalexopoulos & Hesterberg, 1990), time series analysis, least squares approximation and curve-fitting techniques (Hagan & Behr, 1987). As described in the book by Shahidehour, Yamin and Li (2002), the statistical model, Autoregressive Integrated Moving Average (ARIMA) has been proven as a practical method with overall good accuracy and efficiency. ARIMA models have also proven itself in Juberias, Yunta, Monero and Mendivil (1999) article, that applied the method on real-time LF in a Spanish transmission system. Overall, the said statistical approach reappears in the literature as a well-established method. As computational power and the advancements in semiconductors increased, methods commonly known as artificial intelligence (AI) and machine learning (ML) techniques began to flourish the LF field. Since the 1990s, AI and ML techniques developed quickly and a wide variety of methods has been applied in LF as well as other economic prediction tasks (He, 2017). Neural networks (NN) have been gaining increased popularity the last couple of decades. Application of Artificial Neural Networks (ANN) was tested in a Portuguese dataset containing 93 households (Rodrigues , Cardeira, & Calado, 2014), while (Rahman, Smith, & Srikumar, 2017) researched the application of Deep Neural Networks (DNN) for LF in a commercial and residential building. The article published by Energies, written by Bouktif, Fiaz, Ouni and Serhani (2018) explores the comparison of several machine learning techniques, specifically Linear regression, Ridge regression, K-Nearest Neighbor, Random Forest, Gradient Boosting, Long Short-Term Memory Recurrent Neural Network (LSTM) and Extra Trees. Their results yielded a lower forecasting error compared to all the said models, by applying a univariate LSTM model with parameter optimization, on their given dataset. While most of the literature on STLF is focused on the application within day-ahead markets and hour-based predictions, little focus has been given to high-resolution forecasts based on 15-10 min intervals. In the article by Kobylinski, Wierzbowski and Piotrowski (2020), the authors performed net-load forecasting, with a timestep resolution of 15-min, within a residential microgrid environment (comprising of 93 single-family households) by utilizing ANN method. Their result concluded with electricity load forecasts of single households with MAPE from 10.0 % up to 23.5 %, three days ahead.

Given current technological developments, widespread use of smart meters and increasing role of load management at distribution level, forecasting on different aggregation levels deserves separate attention. As noted by (Ahlert, 2010) methods for this purpose are still in the development stage. For low, residential level loads, the pattern is often dominated by residents' behaviors, and most social behaviors are highly stochastic which results in poor predictability. Due to these facts, when applied to disaggregated electricity consumption, most forecasting methods have relatively large errors, no matter how advanced the methods are and how delicately their hyperparameters are tuned (Peng, et al., 2019). Analyzing small unit load forecasting, in his book - "Economics of Distributed Storage Systems" by Klaus-Henning Ahlert (2010), summarized results of several studies. According to Ahlert the accuracy of forecasts measured by MAPE[11] varies from 1.6% to 11.5% depending on time horizon of predictions from hour ahead to week ahead (Ahlert, 2010). As it was underscored by Marinescu, Harris, Dusparic, Clarke and Cahill (2013) the level of accuracy tends to drop significantly as the level of aggregation decreases, from 1.97% MAPE at the national level and 5.15% at university campus level to 13.8% at the village level.

Besides academic papers and scientific articles, power companies and grid operators have challenged academics and professionals to find the best sets of tools to be utilized in LF settings. A widely known competition named Global Energy Forecasting Competition (GEFCom) has been conducted three times in 2012 (GEFCom2012), 2014 (GEFCom2014) and 2017 (GEFcom2017). Participants were required to backcast and forecast 20 US utility zones and system levels in 2012 (Hong, Pinson, & Fan, Global Energy Forecasting Competition 2012, 2014) while the 2014 competition focused on rolling forecasts of the quantiles of hourly loads for one US utility company (Hong, et al., Probabilistic Energy Forecasting: Global Energy

[11] By a sample of methods of Autoregression, Neural Networks, Support Vector Machines, Kalman Filtering, Fuzzy Regression, Fuzzy Inference System, Genetic Algorithm and Discrete Wavelet Transformation.

Forecasting Competition 2014 and Beyond, 2016). GEFCom2017 focused on real-time hierarchical probabilistic load forecasting for 10 ISO New-England zones (Ziel, 2018).

Electricity consumed and the load pattern is highly dependent on consumers' everyday life routines. In essence anything related to the production and pattern in people's life will influence and impact the load in the microgrid environment. As presented in the paper written by (Diamantoulakis, Kapinas, & Karagiannidis, 2015), modern microgrids can incorporate not only traditional factors but also what they refer to as *Smart Grid Factors*. These factors can be used to reveal patterns that were not revealed before in the traditional grid and are highly representative for the area in which one wishes to forecast since they are collected at the place in real-time. Fahad and Arbab (2014) extended the traditional weather variables of air temperature and wind speed to create a new index called Wind Chill Index, measuring the effective felt temperature and the impact on electricity load.

Even though load forecast is a cornerstone and vital prerequisite for the UC and many other economically essential processes in the power exchange market, disproportionately more literature was focused around technical specifications of forecasting models rather than economic considerations. Due to the increased popularity of advanced statistical methods among the business community, and the rise of data-driven management practices, economic values of improved prediction accuracy eventually started to attract more attention. For instance, in early research by Ranaweera, Karady and Farmer (1997), the authors discussed the practical needs on load forecast accuracy. The economic impact of the inaccurate LF as a function of power system parameters, in a simulated setting, was evaluated in the research. Authors assessed if extra costs of providing more accurate load forecasts could be justified by the economic benefit that they could bring into system operations. It was concluded that LFE within 5% would probably be adequate in practice, while economic value by further reducing forecast errors could be negligible (Ranaweera, Karady, & Farmer, 1997). Moving further Hobbs, Jitprapaikulsarn and Maratukulam (1999) estimated savings in generation costs by improved load forecasts. Obtained results revealed that when mean absolute percentage error (MAPE) is in the range of 3% to 5%, reduction of 1% in MAPE will decrease variable generation costs by approximately 0.1 %-0.3%. Translated to approximate actual numbers, the estimate is that a 1% reduction in forecasting error for a 10,000 MW utility can save up to $1.6 million annually (Hobbs, Jitprapaikulsarn, & Maratukulam, 1999). Sangrody and Zhou (2016) furthered the research field. They noticed that economic loss corresponding to not meeting actual demand (due to negative forecasting errors) is often different from that corresponding to resource wasting (due to positive forecasting errors). Taking the difference into account, a new model evaluation metric and objective function with different economic coefficients for positive and negative errors were proposed. Wang and Wu (2017) tried to close the bridge in the literature between load forecasts and UC. Their paper contributed by proposing two effective strategies to establish the coordination between load forecasting and day-ahead UC (DAUC) tasks, in order to derive improved load forecasts with higher economic values. Considering the asymmetric economic impact of errors and assigning weights of individual forecasting models based on their economic impacts on UC solutions, strategies proposed were tested on a large-scale power system (Wang & Wu, 2017).

Despite doubtless assertion that load forecast is a crucial ingredient for efficient operations in the power market, it remains quite challenging to estimate the exact economic value of improved load predictions. The literature review reveals that this task is highly dependent on a particular setting of the research. The horizon of the forecasts, utility size, and its nature - be it residential, commercial, industrial, or mixed type, are factors that shape the problem at hand and make it difficult to compare results among different studies, as every setting requires a unique approach.

3.0 Data

The first step of any empirical data analysis is to examine the underlying data. The analysis aims to reveal anomalies, significant patterns and relationships between the dependent variable and the available features. In this section, we will guide the reader through the data source and their origin. Further, data preprocessing will be performed to adjust it to the desired state so that it is ready for the final forecasting task. Data description will then be performed to obtain an impression of the data, to detect said significant patterns. Lastly, we present feature selection and software used.

3.1 Data Collection

The historical data collected to be used in developing the forecasting models in this thesis is provided in Table 2**.**

Table 2 Description of Data Sources

Description	**Source**	**File Format**	**Dimension**	**Type**	**Resolution**
Load data	Skagerak Nett	xlsx	18089x2	Continuous	15 min
Regulating market Prices	Nord Pool	xlsx	4524x4	Continuous	Hourly
Day-ahead spot prices	Nord Pool	xls	4524x3	Continuous	Hourly
Weekdays	Own	One-hot encode	-	Dummy variable	Daily
National holidays	norskkalender.no	One-hot encode	-	Dummy variable	Daily
Weather data	https://rp5.ru	xls	8943x13	Mixed	30 min
Football matches on Skagerak Arena	Fotball.no	One-hot encode	-	Dummy Variable	Daily

The majority of data came from Skagerak Nett and Nord Pool consisting of load data from 42 residential houses, 9 business and 2 cooperative residential buildings[12]. The table describes

[12] Also commonly known as “Borettslag” in Norwegian.

each dataset used, the source of the data, file format, dimension, type of data and resolution. In our quest to write the thesis, we have been given access to load data from 25. October 2019[13] until 30. April 2020**.** Particular attention must be sighted at the different resolution of the data extracted from Nord Pool and Skagerak Nett. Nord Pool prices are posted on an hourly basis, while Skagerak Nett collects data on a 15 min resolution.

Concerning the purpose of building models that are relevant in a practical setting, we have opted to use data that would be readily available for all market participant when forecasting day-ahead electricity load over one week. In respect to the economic perspective, Nord Pool data is collected for the relevant period this thesis is analyzing. Energilab is located within area code NO2 (Kristiansand)[14].

3.2 Data Preprocessing

In most cases, the historical data from Energilab, market data from Nord Pool and other sources, as explained in the subsection above, appears to be formatted for human readability. This raises the need for a selection of pre-processing steps necessary to extract relevant data, prepare and store it in such structure that would be universally accepted by our different sets of methods and models. The final goal of mentioned pre-processing steps is to get a tidy and easily understandable dataset for computers. Within the dataset, each row represents a consumption period, while each column represents the input variables and one response variable.

As described in Table 2, datasets were obtained from different sources. Each data file, for the sake of efficiency, will be merged into one large time-series data file. However, as a result of the data resolution, each data file will create concerns surrounding time-series consecutive nature. The available data files must be either aggregated or disaggregated to match the time series date and time of delivery accurately.

[13] Data collected before this period were not readily available. However, since Energilab is still in its early age, we found the data period as sufficient.
[14] Every area code within the Nord Pool marketplace is available at nordpoolgroup.com.

The subsequent subsections will describe more in detail the different pre-processing steps. These individual steps are performed and compiled into one data file that includes all the relevant features. The resulting data file is stored, and used in the final forecasting task in R studio and Python.

3.2.1 Missing Data and Outliers

The expectation, when utilizing different sets of forecasting methods is that most methods, except the simplest ones, do not work when missing values are present in the historical time series. Missing values creates obstacles when creating an efficient STLF model, and thus, needs to be imputed. Imputation is a technique where estimated values replace missing observations before the forecasting task. An important side note to remember when doing this is not to introduce bias and a non-zero average error.

We observed that 3.63% of the total load data was missing, and the majority of these values originate at the end of the time series. Discussions made with Skagerak Nett clarified that the reason behind these missing values was because of maintenance and shutdowns in favour to set up the system to run simulations in the future.

In this thesis, we opted to use the *Hmisc*[15] package to impute the observed missing values fully. The said package utilizes time-varying regression to predict the missing load value by using time and weekday values observed in the past. The imputation process was then iterated ten times to retrieve the mean predicted value, which we then used to fill the missing observations. The resulting data file, after imputation, could then be merged with the other data files to create a large time-series data file without any missing values. With respect to the forecasting task at hand, the thesis did not find it suitable to select a forecasting horizon within a period of a high degree of missing data. As such, the period after 13. march, that is the period after Covid-19 restrictions hit Norway in 2020, will be discarded because of a large collection of missing data during this period[16]. We took this decision since forecasting within this period with a high number of missing data would practically mean the models would to a large extent forecast values imputed by the *Hmisc* function, and thus predict, already predicted values.

[15] Multiple methods were tested out but failed to not introduce bias and non-zero average errors given our multi-seasonal time series.
[16] Mainly as a consequence of maintenance and system checks at Energilab after discussion with a company representative at Skagerak Nett.

Outliers within the time-series pose a question if one wish to either keep them or omit and replace with values that resemble the mean value at that given time of the day. In this thesis, we chose not to omit nor to replace, in order to train our models to perform forecast with a realistic appeal. The underlying reasoning behind this is based on the market structure and pricing model in the electricity market, that would yield a hypothesis that peak periods and the ability to model the said period is highly relevant in terms of economic efficiency.

3.2.2 Public Holidays & Special Events

As described in section 2.6, public holidays should be considered in LF. Currently, there are 12 national public holidays in Norway each year. Table 3 lists them chronologically from 1. January to 26. December. An important notice must be made regarding the actual date these holidays occur. Easter is based on the moon movement, and not the specific date itself. The result is that only five days reoccur on the same date, while the remaining 7 days occur on different dates.

Table 3 Norwegian National Holidays

Holiday name	**Date**
New Year's Day	January 1st
Maundy Thursday	Three days until Easter Sunday
Good Friday	Two days until Easter Sunday
Easter Sunday	First Sunday after the first new moon on or after 21. march
Easter Monday	The first day after Easter Sunday
Labor Day	May 1st
Norwegian Constitution Day	May 17th
Ascension Thursday	40th day after Easter Sunday
Whit Sunday	50th day after Easter Sunday
Whit Monday	51th day after Easter Sunday
First Day of Christmas	December 25th
Second Day of Christmas	December 26th

In addition to the public holidays, as discussed in the background material, Skagerak Arena is the main football Arena for Odd Ballklubb. Odd Ballklubb is currently playing in the

Eliteserien[17] in Norway, thus resulting in the regular arrangement of football matches at the Arena. The implication of this, will be handled by dummy variables on the day Odd Ballklubb and other clubs are playing football at the Arena[18].

3.2.3 Handling Different sets of Data-Resolution

As a consequence of the market structure within the electricity market and a quest to produce a thesis with a practical appeal, data resolution from each database will be aggregated into 1-hour intervals. Each 15- and 30-minute data points will be averaged to form a 1-hour observation. A consequence of aggregating the observations is a closer resemblance towards the electricity market and thus strengthen the economic analysis. However, a lower degree of resolution would be interesting in terms of the operational perspective, but in regard to different sets of resolution from each datafile, we did not find it viable in this thesis.

3.3 Data Description

The data description section intends to provide the reader with a solid understanding of the features in the data of the study. Supported with visual data representation, it would lay a foundation for consequent data analysis in this master thesis. Useful insights about the nature of the data are necessary for the appropriate model formulation and result evaluation. As it was mentioned previously, the final dataset consists of three subsets: Physical Load Time Series, Nord Pool market data (Spot and Regulating market prices) and records of weather observations.

3.3.1 Load data

Amongst various characteristics of time series, stationarity (or nonstationary) is the most basic data characteristic. Series are considered to be following stationary process when statistical properties such as mean, and variance do not depend upon time. Formally it is defined as follows:

[17] Eliteserien is the top football league in Norway.
[18] As the datafile only consisted of 3 football games in october and november, the dummies were not included in the final forecasting proceedure. However, we emphesise that the dummies should be considerd in a future analysis.

$$E(Y_t) = \mu \ , for\ all\ t\ ,$$
$$Var(Y_t) \ = \ \sigma^2\ for\ all\ t,$$
$$cov(Y_t, Y_k) \ = \ cov(Y_{t+s}, Y_{k+s})\ , for\ all\ t, k, s$$

Simply stated, stationary data portrays a horizontally looking series (constant μ), without trend, with constant variance over time and with no seasonality. For many statistical tools, particularly ARIMA, it is important to ensure that series under study are following the stationary process. Thus, it is important to analyze underlying processes of data formation in great detail. One should carefully evaluate the trend, cyclical, seasonal and irregular components of the time series. Trend is a pattern within a series where there is a long-term development in the mean of the dependent variable over time, it can be characterized by having a linear or nonlinear behavior. Seasonality occurs when the time-series data fluctuates over a given fixed time period. These time periods could span from daily, weekly to monthly and yearly intervals. Cycles are similar to seasonality, but with a major difference in that the time period is said to be unknown. Irregular or random variations in a time series are caused by unpredictable influences, which are not regular and also do not repeat in a particular pattern. Time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series over different periods (Hyndman & Athanasopoulos, 2018)

Graphical representation and visual inspection of the time series are often of great help when trying to spot different components of data formation.

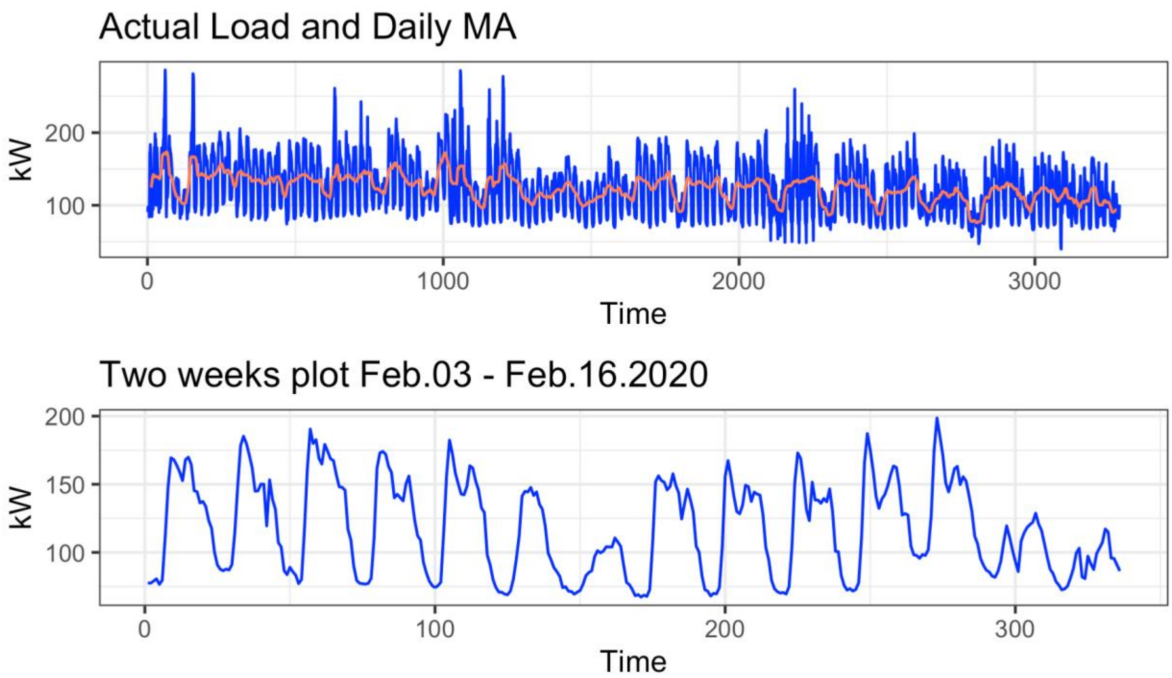

Figure 4 Visual Inspection of Seasonal Patterns of Time Series Data

In many instances economic or energy-related time series data have been proved to consist of several coexisting or cointegrated seasonal characteristics at the same time (Tang, Wang, & Wang, 2013). On the above Figure 4, while it is difficult to judge if the trend is present, the load curve clearly depicts both a daily and weekly seasonality. On a daily basis, electricity consumption rises from morning till noon gradually decreasing afterwards until it plateaus during the night. The weekly cycle is characterized by the high load during weekdays and significantly lower during the weekends. Apart from repetitive cycles, it is evident that the electricity load at this resolution level (microgrid) is highly volatile. It is said to experience a weak statistical pattern due to a lack of aggregation. Thus, when compared to big-volume consumption entities, small unit load forecasting happens to be a much more challenging task.

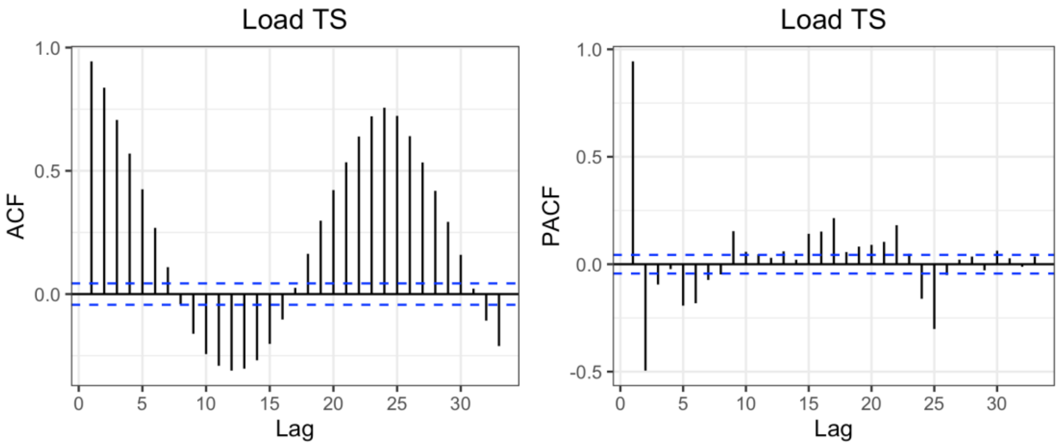

Figure 5 Autocorrelation and Partial Autocorrelation Plots

The autocorrelation and partial autocorrelation plots in Figure 5 help to graphically visualize the relationship of a variable with itself in a previous time period. It is an important characteristic to take into account when working with time series as contemporary variables tend to be influenced by their past values, and this feature affects model formulation. Figure 5 clearly exhibits significant correlation of contemporary observations with its lags, sine wave pattern of ACF confirms previously observed seasonal patterns (daily on this figure). The same is evident from PACF graph spikes at early lags (1,2) and later on (24,25) suggest the presence of an autoregressive process. Thus, one can conclude that the series under study are not stationary.

Differencing is a popular method for achieving stationarity, it can help to stabilize the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality (Hyndman & Athanasopoulos, 2018). Seasonal differencing can be described as follows:

$$y_t{}' = y_t - y_{t-m}\ , where\ m\ is\ the\ seasonal\ period$$

If seasonally differenced data appear to be white noise, then an appropriate model for the original data is:

$$y_t \ = y_{t-m} + e_t$$

There are several objective tests to check if differencing resulted in obtaining stationary data

Augmented Dickey-Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) are among the most popular ones, however, they are not identical, and it is usually advised to perform both rather than using them interchangeably.
The ADF tests for the presence of “Unit root”. Formally it tests if the coefficient θ in the following equation is equal to 0:

$$\Delta y_t = a + \theta y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + e_t$$

Controlling for autocorrelation by including a specified number of $\gamma_i \Delta y_{t-i}$ terms, ADF tests if the autoregressive process is present in the differenced time series. The null hypothesis assumes the presence of unit root. (H_0: $\theta = 1$).

Sometimes it is convenient to have stationarity as the null hypothesis and reject it only in the presence of significant statistical evidence. In the KPSS test, the null hypothesis of stationarity corresponds to the hypothesis that the variance of the random walk equals zero (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). The test breaks series in three components a deterministic trend (δt) a random walk (r_t) and stationary error (e_t) resulting in the following regression:

$$y_t = \mu + \delta t + r_t + e_t,$$

$$where\ r_t = r_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2)\ and\ is\ uncorrelated\ with\ e_t$$

$$H_0(y_t\ level\ stationary): \sigma^2 = 0\ and\ \delta = 0$$

Table 4 below reports the results obtained while performing ADF and KPSS tests on the weekly differenced Load data obtained from Skagerak Energilab for the period from 29.10.2019 to 13.03.2020.

Table 4 ADF and KPSS Test Result

	Statistic	**p-value**	**H1**	**Result**
ADF	-10.65	0.01	Stationarity	Stationary
KPSS	0.11	0.1	Nonstationarity	Stationary

Results suggest that seasonal differencing produces stationary time series of load data. Continuing exploration of the dependent variable Table 5 summarizes some of the measures of central tendency and dispersion statistics for the electricity load observed in studies.

Table 5 Summary Statistics of Weekday and Weekend Load Data

	N	Min.	Median	Mean	Max.	Sd.Dev.	Variance
Overal	3287	39.75	119.63	123.64	286.29	38.32	1468.29
Weekday	1068	69.19	140.16	138.41	286.29	41.94	1759.26
Weekend	432	69.63	114.24	116.54	261.16	28.88	834.14

Reinforcing previous remarks, reported mean statistics for grouped observations portrays the structural difference between weekday and weekend load profiles. Additionally, standard deviation, when analyzed in line with corresponding mean values, reveals significantly different magnitudes of load variability during weekday and weekend.

Figure 6 allows investigating the consumption levels at a smaller time resolution.

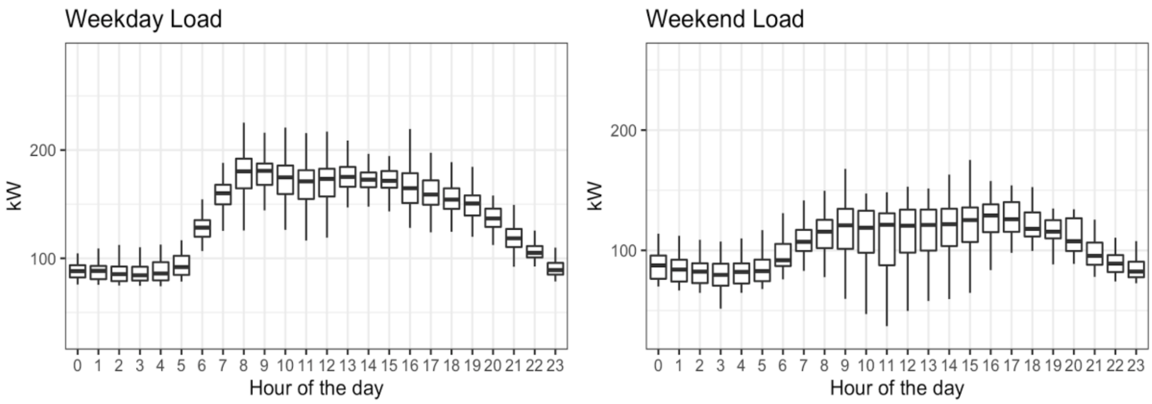

Figure 6 Distribution of Consumption Levels at Each Hour of the Day, Split by Weekday and Weekend

Visualizing dispersion of hourly load reveals some interesting insights. The variability of consumption is much higher during weekends, particularly during daily hours from 7 AM until 5 PM[19]. It can be attributed to the fact that time-dependent social practices shape the energy demand, and a more stochastic social behaviours usually characterizes weekends. Thus, load forecasting is not merely a practice of blind model formulation, it also involves the analysis of

[19] Empirical evidence can be observed in Appendix A3.

underlying social context. Not surprisingly techniques used to predict electricity usage in industrial districts, for instance, would not work equivalently well in other settings. However, overall seasonality tends to be persistent. As a final step in data representation of the electrical consumption Figure 7 depicts the average hourly load profile for each day of the week.

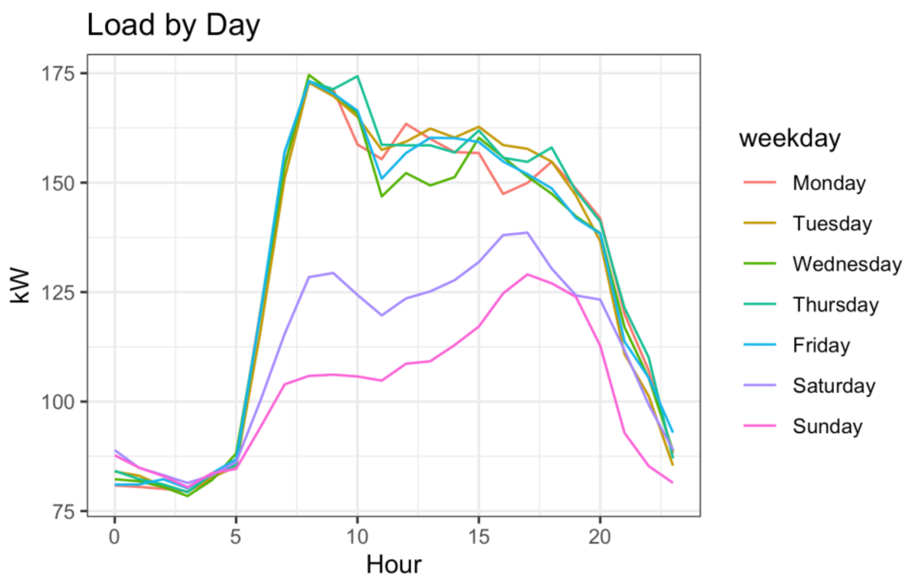

Figure 7 Average Load by Hours and Day

Five of seven days are comparatively similar to each other; however, Saturday and Sunday have unique load patterns. This fact can suggest that representing weekday and weekend as only separate attributes, might be not sufficient for an effective model; instead, one might be willing to use a broader set of dummy variables.

At this point, it is worth noticing that electricity load data, apart from displaying previously mentioned patterns also follows a yearly seasonality, that is usually discernible when observations are collected over a longer time horizon. Additionally, if collected over several years, one may observe a general trend in consumption development.

3.3.2 Economic Data

The data used to measure the economic effect of the forecasting model's accuracy is obtained from the Nord Pool database[20] and represents actual Spot and Regulating prices for the same period as the load data from Skagerak Energilab. After the liberalization of power markets, the price of electricity, in line with load data, became the most important mechanism to ensure operational stability and efficiency. Due to the unique nature of power as a commodity and due to many factors, that affect the power market, electricity price is characterized by high volatility, multiple seasonality, and the presence of spikes. Descriptive statistics of electricity prices are summarized in Table 6. While analyzing presented data, it is essential to remember that the Elspot price meanwhile represents the minimum reference price for up-regulating power and the maximum price for down-regulating power bids (Bang, Fock, & Togeby, 2011). In this regard, it is not surprising that the mean values of up and down-regulation are respectively higher and lower than the Spot price. However, minimal and maximum values for regulatory prices might attract interest.

Table 6 Summary Statistics Nord Pool Data

	Min.	Median	Mean	Max.	St.Dev	Variance
Spot	66.24	292.47	279.45	857.02	131.07	17179.69
Down	10.47	262.10	261.45	857.02	131.40	17266.42
Up	71.09	300.48	293.55	2737.42	152.02	23110.38

Reported values of minimal down-regulating price (10.47 NOK/MW) and maximum value for up-regulation (2737.42 NOK/MW) seem to fall rather far from the mean. As it was mentioned previously, electricity prices exhibit a phenomenon which is called spikes or in more general terms: outliers (Duffner, 2012). It is one of the peculiar characteristics of the power market as the presence of price spikes is not necessarily related to corresponding variations in electricity demand (Zedda & Masala, 2019). In some rare events, due to limited transmission capacity, breakdowns or excess production from renewable sources (e.g. wind) exceptionally high or low prices might occur.

[20] The said datasets is available at: https://www.nordpoolgroup.com/historical-market-data/

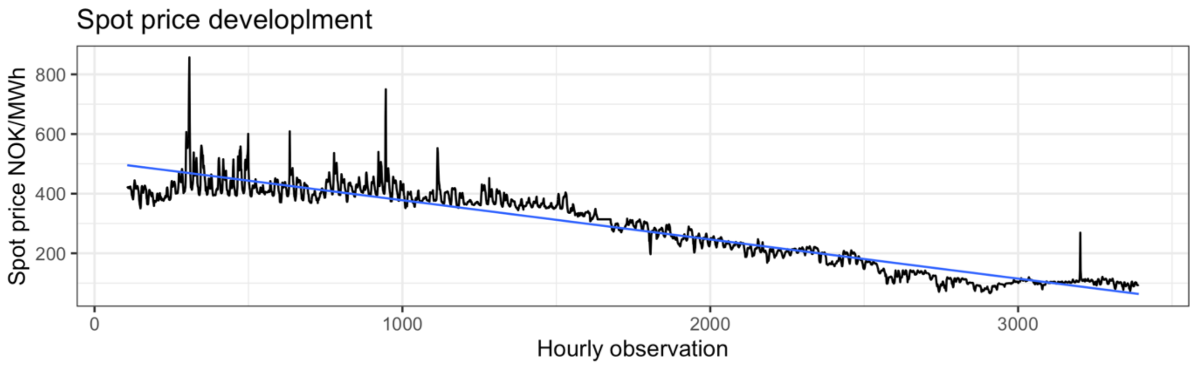

Figure 8 Spot Price Development

Figure 8 exhibits the dynamics of the spot price for the period of interest. The fitted linear model emphasizes the overall trend for the price reduction from October to March. Starting at levels close to 500 NOK/MWh prices gradually fell reaching 100 NOK/MWh and lower. Sharp spikes on the graph represent the type of an event described above. Hypothetically in case of ideal load forecasts, utility companies would contract power purchases concerning spot prices, however, in practice, companies do need to adjust actual deviations from the expected load by participating in balancing markets. This endeavour results in incurring an opportunity cost as regulating prices tend to deviate from the spot price.

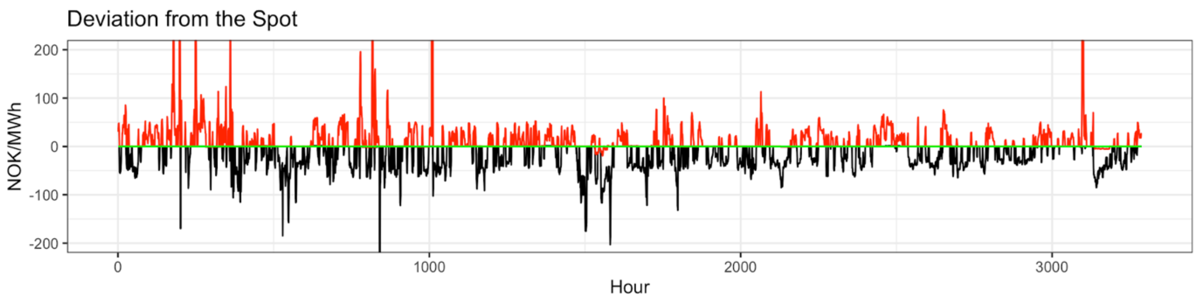

Figure 9 Regulating Market Price Deviation from Spot Price

Figure 9 displays the empirical variation of regulating prices relative to the day ahead spot market price. In every particular hour, the dominating direction of regulation determines the price deviation from the spot, e.g. if most of the market participants are willing to sell the overcommitted amount of power then the down-regulation price would be significantly lower than the spot, and vice versa. For the period of study, most of the variations are occurring in the range of (+/-) 200 NOK/MWh, either way, correction of erroneous commitment can be regarded as opportunity cost calculated by the absolute difference in regulation market prices and spot market prices.

$$OC_t^{down} = P_t^{spot} - P_t^{Down}$$

$$OC_t^{up} = P_t^{Up} - P_t^{spot}$$

Where OC_t^{down} and OC_t^{up} represents the opportunity cost of down and up-regulation in period *t*. Spot market prices in period *t* represented by P_t^{spot}while subsequently P_t^{Down} and P_t^{Up} represents the regulating market prices for both down and up-regulation in period *t*. Summarized statistical data for opportunity cost is represented in the table below.

Table 7 Summary Statistics Opportunity Cost

	Min.	Median	Mean	Max.	St.Dev	Variance
Opp.cost Down	-1.48	30.19	29.30	329.90	27.96	781.91
Opp.cost Up	-19.21	15.91	26.22	2184.84	83.56	6981.63

Table 7 has several interesting points. First, the median and mean values of opportunity cost for downward regulation are higher, at the same time standard deviation and variance of up-regulation are larger. Second, maximum values that are usually associated with the price spikes, represent unexpected events in the market and have tremendous economic bearing for market participants. Third, negative values hint towards some market inefficiencies or price settlement errors.

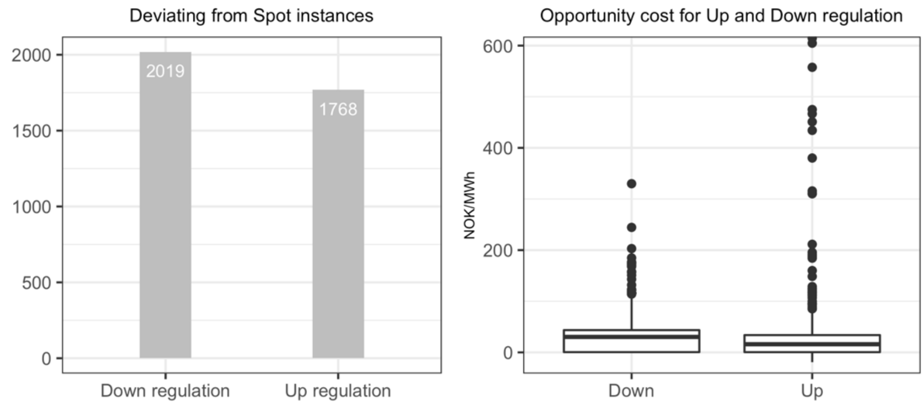

Figure 10 Instances of Deviances and Distribution of Opportunity Costs

Finally, Figure 10 displays the number of instances when regulating price deviated from the Spot, for the period starting from 29.10.2019 to 13.03.2020, there were 2019 instances when the overcommitted load was prevalent in the market, and 1768 instances when contracted power was inadequate. On average, down-regulation in this period had slightly higher opportunity cost, however extreme values are more often associated with up-regulation.

3.4 Variable Selection

Variable selection and identification of variable importance is in broad terms a usual concern when building statistical models and machine learning algorithms. The consequences of high dimensionality[21] are that inclusion of features that are not truly associated with the dependent variable will lead to a reduction of test errors. Despite this, our task in load forecasting is to identify the forecasting power of different methods within a microgrid environment, and our available features are small in number, henceforth we will thus not consider variable selection as a focus area in this thesis.

3.5 Software used

All of the above analysis, data preprocessing, visualization, and the subsequent forecasting experiments, model evaluation and result visualizations were used through the open-source statistical programming software R and Python[22].

The main package[23] used for the forecasting experiments was the *'forecast'* package. It was developed by Professor Rob J. Hyndman and includes a wide variety of methods that could be used in forecasting tasks. Also, the forecasting experiments were conducted by utilizing the TensorFlow library in Python. TensorFlow is developed by the Google Brain team and is an integrated tool that allows designing, build, and train deep neural networks (DNN).

---

[21] Often commonly known as the curse of dimensionality.
[22] The full ensemble of codes and data used is available upon request.
[23] By packages, we refer to downloaded statistical packages in R that can contain machine learning algorithms, mathematical processing and data handling tools.

4.0 Methodology

The methodology section will build the framework for our LF task at hand. First, we present a perspective on relevant cross-validation methods, model evaluation and choices made for performance assessment. Further, the forecasting procedure will be presented to enlighten the basic workflow of the analysis in this thesis. Then, the benchmark model will be explained, while lastly, proposed statistical techniques and machine learning methods will be defined in detail.

4.1 Cross-Validation

Estimating how robust the overall performance of a model, is an essential task in the forecasting procedure. It is usually done by exposing the model to previously unseen data that was not utilized in the process of building the models in the first place. The primary reasoning behind this is to avoid overfitting and assess the performance on an entirely new data set (test data). Overfitting usually refers to the case when the model has not only been capturing the overall signal of the training data, but also most of the noise that it was exposed to, yielding a model that performs excellent when predicting using the training data, but fails to generalize on unseen datasets.

To overcome the pitfall of overfitting and the drawbacks discussed above, cross-validation has been established as an appropriate method to assess the models' performance. In its purest form, it involves splitting the dataset into training and test samples, creating two individual datasets that are either only used to train the models or to test the performance of the proposed models.

K-fold cross-validation is a more sophisticated technique, where the data is split into roughly equal k folds. Models are then trained on the k-1 folds and consequently evaluated on the remaining hold-out data. The process is repeated k times, then obtained results are averaged out to get the overall performance. While k-fold cross-validation yields the advantage of using all the data, it has the drawbacks of slower computational time, since models must be fitted for each iteration.

In cases when one does not handle time-series data, the standard k-fold cross-validation methodology discussed above is recommended (Hastie, Tibshirani, & Friedman, 2017). However, when dealing with time-series, like in our case, specific problems might arise. First, by splitting the data into k folds without respecting the timeline sequence, creates a bias, as the model might learn the future movements, creating the potential of getting artificially increased forecasting accuracy, by knowing the actual future values. Secondly, it does not represent the real-world application, predicting the forthcoming observations, by utilizing future data to train the models, that still does not exist at that point of time. Hyndman & Athanasopoulos (2018) suggest using time series specific cross-validation in cases when one is dealing with sequential data where the timeline is essential.

Time series cross-validation resembles the standard K-fold cross-validation but differentiates itself by only using past data to train the models. This method is also known as an expanding window approach (Hyndman & Athanasopoulos, 2018). A consequence of expanding window approach is that the size of training data increases for each iterative forecast into the future. The disadvantage, however, is that comparing models over time is difficult, as for each period the models have more data to train on. The main benefit, on the other hand, is that while respecting the timeline, one always forecasts and consequently measures the validation error only on future values. To overcome the flaws of the expanding size of the training data, rolling window forecasts is another variety of cross-validation technique. It keeps the training size constant by dropping an equal amount of the oldest observation for each iteration.

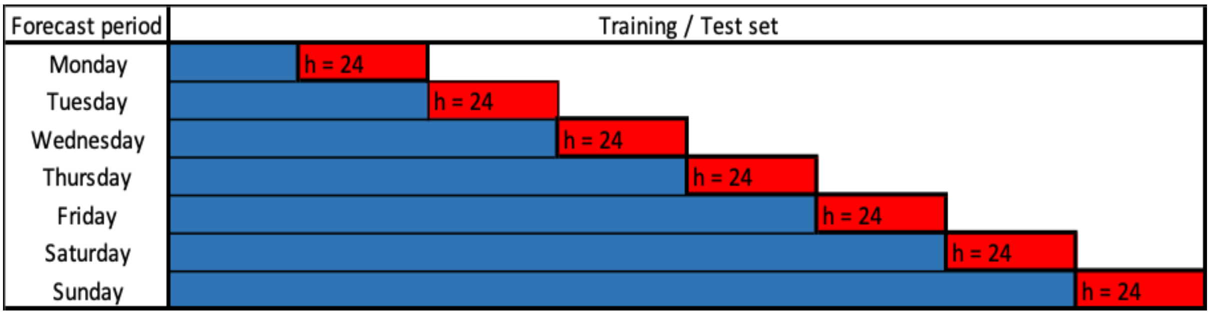

Figure 11 Expanding Window

Since the collected data from Skagerak Energi only spans from 29.10.2019 up until 13.04.2020, we found that expanding the window would be the most sensible approach given the already low number of observations.

For our analysis, the last full week of data observations was set aside as a test sample. The multistep forecasts will be produced for one day ahead or 24 hours in the future as the data has an hourly resolution. Figure 11 schematically represents the proposed procedure.
Iteratively forecasting seven individual days, the overall performance would be averaged over a whole week, this way allowing to assess the mean performance of each model.
Thus, applying expanding window cross-validation technique, the training data will consist of observation from the end of October 2019 until the last observation of the day before the forecast (expanding every iteration by 24 observations). Starting from Monday, 2. March 2020 until Sunday 8. March 2020 will serve as the test data[24].

4.2 Model Evaluation

Estimating the operational and economic efficiency gains that can be derived from an increased forecasting accuracy is a highlighted intention of this thesis. Thus, it is important to line up the objective with the appropriate measurement criteria. Assessment of forecasting accuracy is a fundamental evaluation criterion that will allow us consequently to estimate economic gains and discuss operational nuances regarding obtained forecasts.

While there are many different metrics available to perform this task, selecting a particular one is not as straightforward as it may seem at first glance. Each metric has its strengths and weaknesses. Moreover, the scale at which the models are evaluated has to be analyzed (percentage or absolute value). In this subsection, we will present the accuracy metric that was selected, and propose an economic loss function that would be used to evaluate incurred financial cost due to imperfect forecasts.

Most available forecasting methods tune model parameters in order to minimize a predetermined loss function. Commonly used evaluation metrics in cases of supervised

[24] The reason for discarding observations from later period are discussed in section 3.2.1.

learning are Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root-Mean Squared Error (RMSE) or Mean Absolute Percentage Error (MAPE).

In this thesis, MAPE has been chosen as the first of two loss functions that we wish to optimize. There are three main reasons for covering this decision. First, the models that we develop, and their corresponding results, have in our intention the potential to be scaled up to a larger customer dimension than the Energilab project. Thus, displaying the results in absolute percentages will create a comparable outcome. Secondly, percentage errors can easily be interpreted as high or low in a day-to-day decision-making process. Lastly, MAPE has shown to be an established loss function across the literature, thus enabling us to compare our results to available and relevant literature.

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \frac{y_t - \hat{y}_t}{y_t},$$

Besides these established model evaluation metrics described above, there is also the possibility to utilize problem-specific tailored loss functions, to penalize different model characteristics. As it was discussed, many decisions in the electricity market are taken in advance, usually based on some forecasts of expected load, but as we know LFE can be positive or negative, resulting in additional costs due to participation in the regulatory market (opportunity cost). At the same time as was noted in the background section, the purpose of the day-ahead market is to minimize total operational cost, this can be achieved by contracting the power at the spot price that reflects the market balance at each hour. The actual load multiplied by the spot price thus enables us to derive the ideal day-ahead operational cost, whereas the imbalance between the actual load and the day ahead forecasts has to be settled in the regulatory market.

The described market structures enable us to construct Economic Load Forecasting Errors (ELFE). The particular loss function in mind is built upon a set of relevant opportunity costs, that is computed based on market data from Nord Pool.

$$ELFE = COST_{ideal} + \sum_{t=1}^{t} (y_t - \hat{y}_t^u)\, OC_t^{up} + \sum_{t=1}^{t} (\hat{y}_t^o - y_t) OC_t^{down},$$

Where OC_t^{down} and OC_t^{up} both represent the opportunity cost associated with participation in either the down-or-up regulation market in each individual time period that resulted due to imperfect forecasts. These opportunity costs are set in NOK per kW deviance from the spot price[25]. $\hat{y}_t^u$ and $\hat{y}_t^o$ represents the under and over point forecast in period *t* while y_t represents the actual load in period *t*. These two sums are then added to the ideal cost. Thus, ELFE represents the realized cost after participation in the regulation market. Furthermore, ELFE enables us to quantify the percentage of economic costs (PEC) that was incurred due to inaccurate forecasts.

$$PEC = \frac{ELFE - COST_{ideal}}{COST_{ideal}} * 100$$

PEC reflects what percentage of the ideal cost was incurred due to imperfect forecasts. It indicates how load forecasting errors will affect the cost. In terms of load forecasting models, a higher PEC indicates a model that would lead to a higher economic cost and loss. Henceforth, this evaluation metric will be assessed to estimate the economic effect of forecasting accuracy of different sets of proposed models in this thesis.

In addition to our two evaluation metrics, we will also conduct an in-depth analysis into the distribution of the LFE. General description of error occurrences, relevant statistics, and actual error incidence throughout each hour of the day[26], will be further discussed. This discussion will contribute to operational insights that might be helpful for many management decisions.

4.3 Forecasting Procedure

The important foundation of forecasting future load is to gather historical sequential data over a certain specified timeframe. The data granularity, commonly known as data resolution, could span from seconds to years. In our case, electricity load is a time series, sequentially collected data points. Mathematically, time series can be represented as the following,

$$\{y_t\}_{t=1}^{T}$$

[25] Refresher on these deviances can be found in the data description of this thesis.
[26] See heat map in appendix A4

where t represents the point in time, while y_t represents the electricity consumption in a particular point in time. Plotting these collections of data points over time creates what is referred in the industry as a load curve. In LF, we wish to predict the load beyond the observed data points in a predetermined timeframe, often referred to as the forecasting horizon. Referring back to 2.3, in this thesis, the forecasting horizon will be short-term and have a market based and operational perspective, henceforth with forecasts 24 hours ahead, over seven individual days. This interacts well with the current operational simulations performed by Skagerak Nett in the microgrid. Thus, the thesis will create valuable insight regarding consumption forecasts, for their upcoming simulations[27] and a market orientation view.

Different forecasting methods, as will be explained in the subsequent sections, create individual forecasting models. These models try to describe best the relationship between the dependent variable, electricity load at a given time, and the selected independent variables for specified methods. To find that particular model and method that best describes the relationship of the underlying nature of electricity load, model parameters will be optimized based on a selection of loss functions.

Initially, this thesis will create a benchmark model that serves as a reference point, in terms of a basic model performance. Further, more advanced methods will be applied to the same dataset that the benchmark model was trained and evaluated on. These results can then be compared to each other to evaluate which model that produces the most accurate predictions with respect to the selected evaluation metrics discussed above. Referring back to the data section, we acknowledge that 15- and 30-min resolution could potentially be too high-resolution to get acceptable results at a small microgrid with highly stochastic nature. With respect to the market structure, spot prices and regulating prices, that are priced at an hourly level, the load forecast and economic analysis will be based on the results from an hourly data resolution. This approach will yield the most sensible results from a market-based perspective and would favour a more practical appeal. Finally, expanding window forecasts will be performed for each individual day of the forecast horizon, observing the consequences of being able to utilize a

[27] As Energilab is intended to create knowledge and insight into microgrids, the engineers at Skagerak Nett are currently testing out different scenarios where the microgrid can perform stabilization, load shedding, island mode etc. These simulations are run across a whole week in which it reveals the capability of Energilab as an operational tool to increase efficiency in the grid.

larger amount of data, while also being able to spot model performance in different periods of volatility.

4.4 Benchmark Model

Evaluation of computationally time-consuming and complex methods discussed in the subsequent chapter requires a foundation of a benchmark model to compare performance towards. Without a benchmark foundation, results from complex methods do not inherently tell if their performance is worth the additional computational cost. In this thesis, the seasonal naïve method will be established as a benchmark model. The motivation behind this method is to represent a simplistic but logical approach to our load modelling, which requires a low degree of pre-analyzing and feature optimization. Also, this method has proven to be a much-suited benchmark across the literature for decades as they generally perform adequately over a selected time interval and does capture seasonality. Thus, this thesis will provide empirical evidence if the state-of-the-art methods are more valuable in terms of lower LFE performance, with respect to their increased computational complexity, in addition to the financial impact in relation to the electricity regulating market. A brief introduction on seasonal naïve will be presented in section 4.4.1**.**

4.4.1 Seasonal Naïve

The seasonal naïve method presented is a simple approach to capture the complex multi-seasonal pattern of the load curve at Energilab. The introduced naïve approach will only be used to aid the reader towards the sophistication level of the more advanced proposed methods of this thesis, and naturally creating a performance to be surpassed at an acceptable level. The basic naïve model assume that what has happened in the past, will repeat itself in the future. An implication of this yields a model that forecasts a straight line after the last observation in the training data, thus, the mathematical formulation is expressed as the following,

$$\hat{y}_{T+h|T} = y_T$$

An extension of the basic naïve approach is the seasonal naïve method. Seasonal naïve enables a simple form of seasonality into the model by forecasting h periods to be equal to the last observation from the same historical season, essentially replicating the last observed day, week or year indefinitely. In our case with multiple seasonality present, seasonal naïve could either replicate the last observed day or week. By replicating last day, the model would quickly misrepresent the actual pattern of the weekly load curve. Hence, we proposed to model with weekly periodicity, to more closely represent the main seasonal pattern. Mathematically, seasonal naïve forecast for $T+h$ is formulated as the following,

$$\hat{y}_{T+h|T} = y_T + h - m(k+1)$$

Where m represents the seasonal period and $k = (h\text{-}1)/m$, which represents the number of complete periods in the forecast horizon prior to $T+h$. Translating this into practice, our benchmark model will forecast next Monday until Sunday by simply replicating the equal values of the similar day of the last week.

4.5 Proposed methods

Moving beyond the established benchmark models, the thesis will now present a selection of proposed methods and models, in order to capture the most accurate forecasting model and best performance based on evaluation metrics for the Energilab microgrid. In spite of the availability of multiple methods to choose from, based on our findings regarding the complexity of seasonality and volatility, the methods proposed seems to be able to best model the load curve present at the microgrid.

4.5.1 Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average, often referred to as the ARIMA model is a common sight in forecasting tasks across the literature. ARIMA is a univariate time series model based upon three main components, Autoregressive (AR), (I) Integrated and (MA) Moving Average. Thus, the ARIMA model accounts for past observed values and can take into account all of these three components or at a minimum only one component. These kinds of models were first introduced by Box & Jenkins (1970) and further revised in Box, Jenkins and Reinsel (1994).

The AR component of the ARIMA model is a linear regression that forecasts future points of the dependent variable y. The difference however, from the standard linear regression is that the feature variables are all past observations of the time series dependent variable. Hence, the AR component uses a combination of the time series lagged values, random errors and a constant, as model parameters to linearly predict the future values of the dependent variable. Mathematically explained as the following:

$$AR[p]: y_t = c + \sum_{i=1}^{p} \varphi_i \, y_{t-i} + \varepsilon_t$$

Contrary to the AR component, the MA component of the ARIMA model uses a regression like model of the q past errors instead of past observations to forecast future points (Hyndman et al., 2018). Furthermore, the constant of the linear combination explains the mean (μ) of the time series. Mathematically explained as the following:

$$MA[q]: y_t = \mu + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t$$

Combining both the AR and MA component yields what is known as the ARMA model. However, ARMA models are only suitable to stationary time series. The ARIMA model was thus invented to overcome this issue, by including the [I] component by applying a differencing parameter to differentiate a finite set of data points, resulting in a stationary time series. The arbitrary characteristics of an ARIMA model are consequently explained by the following parameters ARIMA*(p,d,q)* and in cases of seasonality *(P,D,Q)m* will be specified. Summarizing this, *p* and *P* represents the order of the AR component, *d* and *D* describes the degree of differentiation (non-seasonal and seasonal respectably), *q* and *Q* describes the order of the MA components, while *m* represents the frequency of the seasonality in the time series. In cases of multiple seasonality, additional seasonal components could be added.

Within the '*forecast*' package, Hyndman includes the *auto.arima* function that automatically selects the optimal model parameter discussed above by utilizing the AIC criteria. If one wishes to find and select these parameters manually, one usually identifies the orders by comparing the ACF and pACF of the sample and the properties of the model.

4.5.2 Exponential Smoothing Method

Exponential smoothing (ES) is a well-established time series forecasting tool. The origin of these techniques was developed in the 1950s and has ever since been a popular choice in the forecasting field. The backbone of these models is the technique of predicting the values as weighted averages of past observation, by weighing recent observation more heavily, and the less recent observations by a geometric decreasing ratio. ES methods is considered as peers to the ARIMA model framework discussed above. Despite ES methods, being around since 1950, it was not until the influential article by Hyndman, Koehler, Snyder and Grose (2002), that ES began to flourish with the inclusion of more sophisticated methods.

In the following subsections of ES methods, we will guide the reader through our selected ES model of choice, specifically the double-seasonal Holt-Winters.

4.5.2.1 Double-Seasonal Holt-Winters model

The Double-Seasonal Holt-Winters (DSHW) method, has origins from Holt's linear trend method (Holt, 1957) and the additional extension Single-Seasonal Holt-Winters (SSHW) method, who, as the name suggests, were invented by Charles C. Holt (1957) and his student Peter Winters (1960). The Holt-Winters method is able to capture seasonality, in addition to the underlying level and trend, by estimating smoothing equations. Essentially, for this single-seasonal method, there are two versions, the additive and the multiplicative. The additive is more appropriate in which the time series seasonal variation does not change in size in relation to the level of the series, while the multiplicative approach is used when the seasonal variation is proportional to the level of the series (Hyndman & Athanasopoulos, 2018).

As mentioned before, our data inhibits both intraday and weekly seasonality, yielding need to implement methods that do handle multiple seasonality. The DSHW method was an adaptation of SSHW, invented by Taylor (2003), who purposefully used this method for STLF in the electricity market. This adaptation leads to the inclusion of a second seasonal component, making it able to more efficiently capture the actual load curve, henceforth creating more accurate forecasts.

Taylor's DSHW method uses a combination of additive trend and multiplicative seasonality, where the two seasonal components are multiplied together (Taylor, 2003). Again, there exist two variations of the DSHW, an additive and a multiplicative variation. The variation of choice depends on the same seasonal variation of the time series as with SSHW (Hyndman & Athanasopoulos, 2018). By analyzing the load curve from Energilab, it is safe to conclude that it roughly inhibits a constant seasonal variation, and the additive variation should be the best choice for our purpose.

Formally the additive DSHW method can be mathematically formulated as the following:

Level component	$S_t = \alpha(Y_t - S_{t.-s1} - W_{t-s2}) + (1-\alpha)(S_{t-1} + T_{t-1})$
Trend component	$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$
Seasonality 1	$D_t = \delta(Y_t - S_t - W_{t-s2}) + (1-\delta)D_{t-s1}$
Seasonality 2	$W_t = \omega(Y_t - S_t - D_{t-s1}) + (1-\omega)W_{t-s2}$
Forecast function	$\hat{Y}_{t+h\|t} = S_t + hT_t + D_{t-s1+h} + W_{t-s2+h},$

where the level component is estimated by smoothing the difference of the observed value, the intra-day and intra-weekly seasonality and one seasonal cycle ago. The trend component is additive and is estimated by the weighted first differences of the level component and the previous trend T_{t-1}. The seasonality 1 component D_t, that is the daily (*h=24)* seasonality, is estimated by the weighted average of the difference between the observed value at time *t* and the level at time *t* minus the intra-week index one week ago, with the intra-day index one day ago. The seasonality 2 component W_t , that is the weekly (*h*=168) seasonality, is estimated by the weighted average of the difference between the observed value at time *t* and the level at time *t* minus the intra-day index one day ago, and the intra-week index one week ago. Every component explained above concatenate together to form the final forecasting function for Energilab. The smoothing parameters $\{\alpha,\beta,\delta,\omega\}$ are established by optimization.

4.5.3 Exponential Smoothing State Space Models

Exponential smoothing state space (ETS) models, as presented in Hyndman , Koehler, Ord and Snyder (2008), is compared to the classic ES models from the 50s a newly developed approach based on the innovation state space approach. These univariate models consist of three main components, namely error, trend and season. These components are then represented as a simple exponential smoothing equation, and further combined either additive or multiplicative. The primary strength of these methods are their ability to model seasonality resembling a sinus-pattern, but struggles in cases when complexity rises and non-seasonality is present (Hyndman R. J., Koehler, Ord, & Snyder, 2008). Further in this thesis, we will investigate two ETS models, mainly BATS and TBATS models, both of which are available in the ‘*forecast*’ package.

4.5.3.1 BATS model

Time series forecasting using BATS models incorporates a framework by using Box-Cox transformation, Fourier representation and utilization of ARMA error correction. The model was constructed by the intention to handle intricate seasonality patterns (De Livera, Hyndman, & Snyder, 2010). As observed in the Data description, it revealed such complex seasonality pattern at Energilab.

Decomposing the BATS acronym, gives us Box-Cox transformation, ARMA errors, Trend and Seasonal components. The system of equations to represent the BATS model can be written as the following (De Livera, Hyndman, & Snyder, 2010):

Box-Cox Transformation	$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}; & \omega \neq 0 \\ \log y_t & \omega = 0 \end{cases}$
Level component	$\vartheta_t = \vartheta_{t-1} + \varphi b_{t-1} + \alpha d_t$
Trend component	$b_t = (1-\varphi)b + \varphi b_{t-1} + \beta d_t$
Seasonal component	$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t$
Error component	$d_t = \sum_{i=1}^{p} \phi_i d_{t-1} + \sum_{i=1}^{q} \theta_i \epsilon_{t-1} + \epsilon_t$
Forecast function	$y_t^{(\omega)} = \vartheta_{t-1} + \varphi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t,$

The BATS model is established by a set of parameters, namely BATS($\omega, \varphi, p, q, m_1, m_2, \dots, m_T$), where ω represents the Box-Cox parameter, φ the dampening parameter. p and q represent the ARMA error and $m_1, m_2, \dots, m_T$ the seasonal periods (1,2,…,T). Referring back to Holt-Winters method, a BATS(1,1,0,0,m_1, m_2) signals the double seasonal Holt-Winters method, as we described in the section above. We will fit the multi-seasonal time series of load from Energilab within the *BATS* function and use parallel processing to speed the computational time up. Additionally, this function will choose the best BATS model based on AIC criteria and utilize both ARMA errors and Box-Cox transformation.

4.5.3.2 TBATS model

The next innovation state-space model, the TBATS model, is an extension to the BATS model explained above. TBATS models incorporate an additional component, namely a trigonometric representation of the seasonal component that has been obtained by applying Fourier transformation. The system of equation of TBATS models can be formulated as in (De Livera, Hyndman, & Snyder, 2010):

Seasonality	$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$
Level of seasonality	$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} cos\lambda_j^{(i)} + s_{j,t-1}^{*(i)} sin\lambda_j^{(i)} + \gamma_1^{(i)} d_t$
Growth of seasonality	$s_{j,t}^{*(i)} = -s_{j,t-1} sin\lambda_j^{(i)} + s_{j,t-1}^{*(i)} cos\lambda_j^{(i)} + \gamma_2^{(i)} d_t,$

Where both gamma parameters $\gamma_1^{(i)}$and $\gamma_2^{(i)}$are smoothing parameters, $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$, where m_i represents the seasonal period of the *i*-th component (De Livera, Hyndman, & Snyder, 2010). The resulted fitted TBATS model is set up by the following parameters TBATS($\omega, \varphi, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \dots, \{m_T, k_T\}$). The model naturally resembles the BATS model with the inclusion of $k_1, k_2, \dots, k_T$seasonal fourier terms (1,2,...,T), used within each seasonality (De Livera, Hyndman, & Snyder, 2010). Again, we will fit the multi-seasonal time series object to the *TBATS* function and use parallel processing, Box-Cox transformation and ARMA errors. Model parameters will be chosen by the AIC criteria.

4.5.4 Artificial Neural Networks

In recent years, computational power has risen, and alternative methods have gained popularity in the field of time series forecasting. Artificial Neural Networks (ANN) is inspired by neuroscience and attempts to mimic interaction between neurons with the help of mathematical representation. These methods are able to recognize, regularize and develop knowledge of the output variable y_t , by modeling a non-linear function of input variables x_{t-j} ; $j = 1,2, \dots, J$.

The general structure of an ANN with absence of hidden layers resembles a standard linear regression, with a set of input variables, predicting an output value. Immediately, as we include one or more hidden layers, as depicted in Figure 12, our model becomes non-linear and is often referred to as the multilayer feed-forward network (Hyndman & Athanasopoulos, 2018).

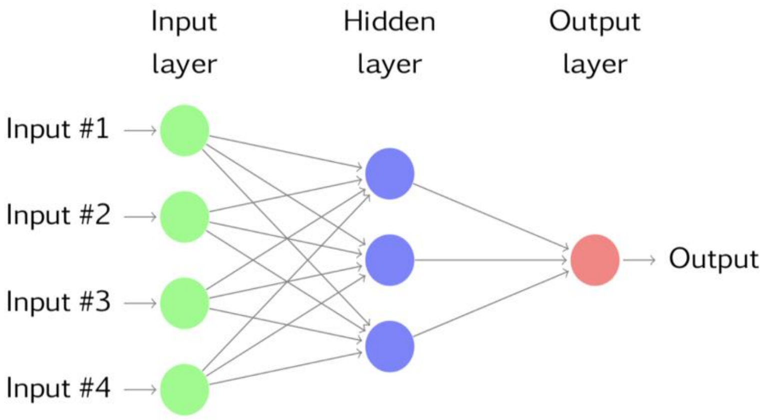

Figure 12 Neural Network with One Hidden Layer. Source: (Hyndman & Athanasopoulos, 2018)

Within each hidden layer, there consists multiple neuron nodes, that feed input values from each neuron into the input layer, and further transmits information to the following layer with each of their corresponding weights. Mathematically, the function for the output layer, based on the structure from Figure 12 and notations from Zhang (2007), can be formulated as the following function:

$$y_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \sigma \left(\beta_{0i} + \sum_{k=0}^{n} \beta_{ki} x_{t-j} \right) + \epsilon_t$$

where y_t denotes the output value in time t, based on the inputs from x_{t-j}. α_i and represents the weights between the neurons while $\epsilon_t$is the random error not captured by the model. The inputs fed into the hidden layer also modified by using activation functions, most commonly the sigmoid activation function $\sigma(x) = \frac{1}{1+e^{-x}}$. This tends to reduce the effect of extreme input values, thus making the network somewhat robust to outliers (Hyndman & Athanasopoulos, 2018). ANN is capable of being both univariate or multivariate models, that is including independent variables as predictors and not only using past observations of the dependent variables as predictors.

4.5.4.1 Autoregressive Artificial Neural Network

An extension of the multilayer feed-forward network is the specialized approach to include time-lagged variables within the input space. The R package ‘*forecast*’ has the inbuilt function ‘*nnetar*’ that fits a neural network with time-lagged values of the time series as input variables, making it able to build forecasting models not only as a classic ANN but also an integration of AR components discussed in the ARIMA section, thus making it an Autoregressive Artificial Neural Network (AR ANN) model.

ANN has proven themselves as viable methods within the field of STLF. However, a crucial aspect to take into consideration is the design of the NN, which includes the number of hidden layers, size of the hidden layers, number of inputs, number of time-lags to include and more. Henceforth, skipping the part of optimal model design has the potential to lead the NN into poor forecasting performance. In opposition to models as exponential smoothing, ANN may also integrate feature values as discussed in section 2.6, thus creating a multivariate model. The function also incorporates the possibility of growing an ensemble of models. In our attempt, we take the ensemble average of 20 neural networks for each day-ahead forecast to produce our final point forecast for each hour of the day.

4.5.4.2 Long Short-Term Memory

Neural networks as discussed in the subsection above are recognized by their directed computation. A major shortcoming of the traditional feed-forward neural network is that it is not able, by their mathematical formulation, to have nodes that are reachable from themselves. Specifically, information fed one node forward into the network is not able to reach back to itself again, for each iteration of new input variables. However, Recurrent neural network (RNN) does include the feature of including cycles[28] inside the neural network, giving it an internal state of memory and allowing information to be saved. RNNs introduction of memory, and their capabilities of persisting information, makes them well suited for data that has a sequential nature, such as speech-and handwriting recognition, translation and time-series forecasting (Goodfellow, Bengio, & Courville, 2016)

[28] Also described as *self-loops* (Goodfellow, Bengio, & Courville, 2016)

An extension of the standard RNN method, the long short-term memory network, commonly known as LSTMs, were intended by Hochreiter and Schmidhuber (1997). Hochreiter and Schmidhuber contributed by arming RNNs with the capability of learning long-term dependencies of the underlying data. Fundamentally, the standard RNNs, should be able to learn long-term dependencies, but as explored by Bengio, Simard and Frasconi (1994**)**, experimental results found evidence that even though not impossible, RNN and the gradient descent of an error criterion may be inadequate to train models in tasks involving long-term dependencies (Bengio, Simard, & Frasconi, 1994) because of increased inefficiencies when the temporal span expands.

LSTMs avoids this obstacle by including a selection of *gates* and a *cell state*. The cell state is feeding information forward in the network and picks up new information from three essential gates. These gates are composed of sigmoid neural net layer and a pointwise multiplication operation (Olah, 2015). The sigmoid layer optionally let new information through by a sigmoid function, with values between 0 and 1, where 0 represents not letting any new information through, while 1 represents that all new information should be let through to the cell state. In LSTM, there are three types of these gates to control and persist the cell state.

LSTM is composed of three gates; The *forget gate layer*, *Input gate layer* and lastly the *Output gate layer.* RNNs model its input time-series using recurrence:

Hidden state $h_t = f(h_{t-1}, x_t),$

where x_t is the input at time *t* and $h_{t-1}$is the hidden state which is a vector representation of the historical inputs up until time *t*. As mentioned, LSTM includes a selection of gates in the recurrence function f. Formally these gates and cell states of LSTM can be represented by the following system of equation:

Forget gate layer	$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
Input gate layer	$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
Candidate cell state	$\overline{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
Cell state	$C_t = f_t * C_{t-1} + i_t + \overline{C_t}$
Output gate layer	$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

Hidden state $h_t = o_t * tanh(C_t)$.

The *forget gate layer* decides what information should be discarded or not from the cell state by looking at h_{t-1} *and* x_t. The *input gate layer* then decides what information that should be updated and stored in the cell state, by subsequently creating a vector of new *candidate cell states*. The combination of the forget gate, input gate, candidate cell state and lagged cell state will update the current *cell state* at time t. Finally, determination of the output will be based upon the cell state that has been put through a *tanh*[29] and the multiplication of the *output gate layer*.

Henceforth, as we feed a time-series sequence into an LSTM, it will substantially create a vector of hidden state outputs $\{h_1, h_2, \dots, h_n\}$.

4.5.5 Ensemble Average Model

It has been well-known since at least 1969, when Bates and Granger wrote their famous paper on "The Combination of Forecasts", that combining forecasts often leads to better forecast accuracy (Hyndman & Athanasopoulos, 2018). An ensemble average model is simply a collection of forecasting models put together to produce an average point forecast.

In our setting, it is a collection of the three most accurate models based on the MAPE evaluation criteria, evaluated on the same test data as all other models. Each point forecast from these models is averaged together to produce one final forecast for each hour. The main advantage of this method is that it utilizes the advantages and disadvantages of each of the best performing models, to purposefully produce a higher performing model.

[29] By using *tanh* or formerly known as Hyperbolic tangent activation function, the cell state values will be between -1 and 1. The *Tanh* function is s-shaped.

5.0 Results and Analysis

The following section presents the forecasting results, beginning with the benchmark model, which will set the basis for comparison of the proposed methods. As was discussed in the methodology section, the initial assessment would handle the overall accuracy of the models. Subsequently, benchmark and proposed methods will be funnelled through our tailored economic load forecasting error function, to assess the economic impact of each individual model. The results from the economic analysis will reveal each model's performance in terms of their level and magnitude of error over time, which will translate into direct opportunity costs within Energilab, as a result of grid-connected market participation.

5.1 Benchmark Result

Every result in this thesis should be compared to and seen in reference to the benchmark model. Given the nature of data under the study (evident seasonality), the seasonal naïve method represents a commonsense approach, where one can reasonably assume that the load pattern repeats itself with a certain periodicity. Thus, results from this model would represent a logical baseline that would have to be beaten in order to demonstrate the usefulness of the proposed models.

5.1.1 Performance Day-Ahead Forecast

The benchmark model was capable of forecasting load with a varying degree of accuracy across the seven individual days of the week starting from Monday 00:00 AM, 2. March, until Sunday 11:59 PM, 8. March. Results depict a not so surprising resemblance to the volatile nature of the load curve, with LFE ranging from 2.30 % (forecasts for Tuesday) up until 25.63% (forecasts for Saturday) MAPE on the test data (Table 9 summarizes the forecasting results of each individual day of the forecasting horizon.**)**. Further, into the analysis, we conduct a more in-depth look at the errors, and analyze if the observed patterns iterate across the proposed methods.

In Figure 13**,** point forecasts from seasonal naïve are depicted in comparison to the actual load values at Energilab in the given forecast horizon. It is evident that performance is accurate from Monday until Wednesday but fails to anticipate a lower load level on Thursday afternoon.

Further, low forecasting accuracy on Saturday can easily be disentangled by an unpredictably low consumption level during the mid-day, this fact led to the highest MAPE over all individual days from the test sample. Referring back to the data description, observation during this particular Saturday deviates from the average values one would expect on this hour and day.

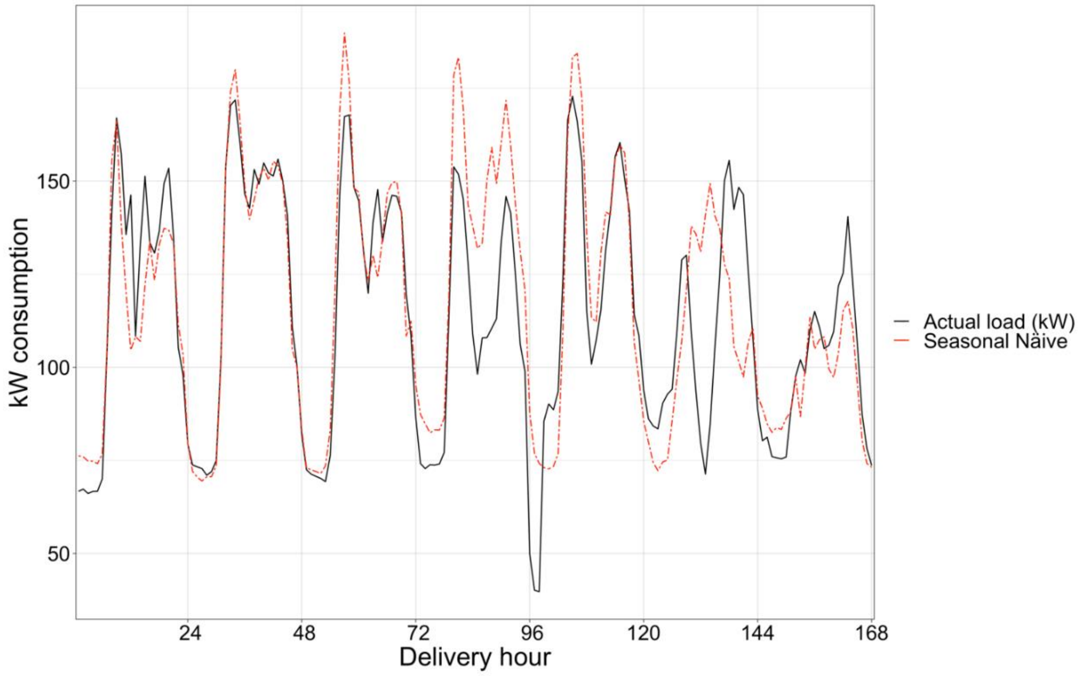

Figure 13 Point Forecast from Seasonal Naïve

Figure 14 represents the fitted values of seasonal naïve (fit in training data). It is clear to observe the consequence of using seasonal naïve as a forecasting approach, as it iteratively repeats the last week's observation, which consequently could lead to larger LFE if the last observed week was predominantly cluttered with high degrees of volatility. However, the benchmark seems to fit the data relatively well.

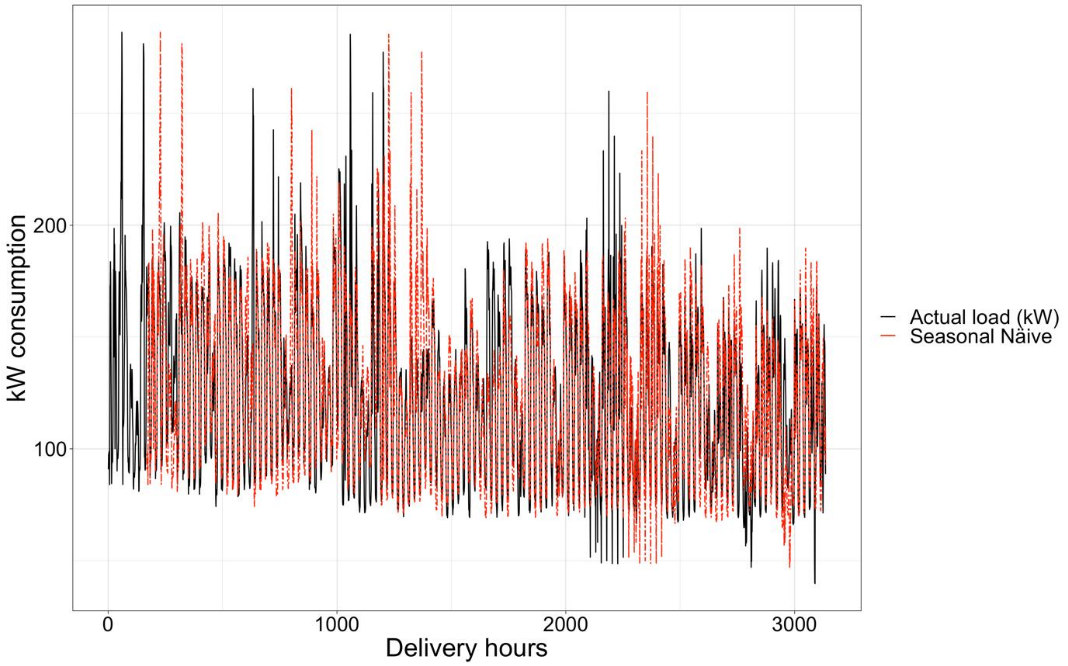

Figure 14 Fitted Values Seasonal Naïve (Training data)

5.1.2 Performance Across Delivery Hours

For operational purposes, it is essential to analyze the hourly performance of obtained forecasts (ex-post analysis). Plotting errors for each hour of the day ahead forecast, for the test data, can depict patterns where the model might be performing exceptionally good or bad. As displayed in Figure 15**,** observations visualize the difficulty of forecasting the mid-day hours. These observations validate what was noted in the data description, where load values had the highest degree of dispersion from 7 AM until 5 PM. Furthermore, the figure also disentangles the cause behind Friday's low performance, unexpected load activity in the morning hours resulted in significant errors. Another spike (large MAPE in this case) of comparable magnitude can be observed during mid-day on Saturday.

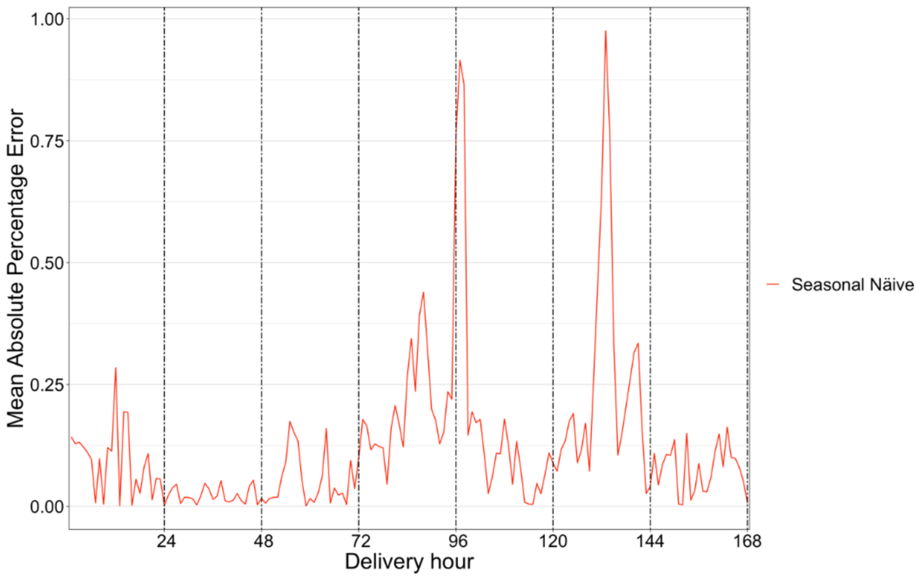

Figure 15 Benchmark MAPE for each Delivery Hour

Generally, seasonal naïve performed as expected, given the cyclical historical load curve. Early mornings and late-night forecasts seemingly tend to be the most precise hours, with increasing errors observed during mid-day forecasts. Unexpected volatile load values on Friday and Saturday were however not present in the previous week, thus leading to a high degree of errors on these days.

5.1.3 Distribution of Forecasting Errors

In addition to evaluating seasonal naïve as a forecasting method by using MAPE as a metric of choice, including other performance criteria might be of interest concerning the economic analysis and value of forecasts. The electricity regulating market moves in waves of dominating directions and could endure grid companies with financial distress as a result of low forecasting accuracy. With seasonal naïve, one would expect that what has happened will happen again. As depicted in Table 8**,** this assumption leads to a dominating presence of over predictions with 90 instances, compared to 78 instances of under predictions. This makes sense given the downward trend of kW load at Energilab in the observed historical data.

Table 8 Distribution of Benchmark Forecast

	# Under pred.	# Over. pred	Max. Diviance	Avg. Diviance	**MAPE**
S-naive	78	90	69.58	12.81	**12.63**

Furthermore, the maximum deviance of kW is recorded at 69.58 kW with average deviance of 12.81 kW for each hour of the forecasting horizon (in absolute terms). What this means in terms of economic performance will be presented in the following subsection, but regardless, given the overall average load recorded at Energilab at 123.54 kW, the maximum deviance from seasonal naïve is high.

5.2 Results from Proposed Methods

Following the estimation of the benchmark model by assessing MAPE, the direction of LFE, and kW deviance, the proposed methods will be presented in the following subsection. Each proposed method and model will be presented to give a foundation for comparison, and the final economic analysis as a consequence of forecasting accuracy would be displayed.

5.2.1 Performance Day-Ahead Forecasts

Most of the proposed methods and models did, on average, outperform the seasonal naïve benchmark. A general point of interest is to what extent each technique is capable to successfully model periods of higher volatility as these periods would potentially lead to a higher economic loss. Table 9 summarizes the forecasting results of each individual day of the forecasting horizon.

Table 9 MAPE for All Models within Each Day of the Forecasting Horizon

	Mon.	Tue.	Wed.	Thurs.	Fri.	Sat.	Sun.	**Average**
Benchmark Model								
S-naive	8.96	2.30	5.53	22.52	15.82	25.63	7.66	**12.63**
Univariate Models								
Arima	7.71	2.39	5.47	21.03	13.10	24.48	7.73	**11.70**
DSHW	8.66	6.89	7.34	18.04	19.99	14.50	11.67	**12.44**
BATS	9.76	5.12	4.86	15.86	18.36	13.73	10.88	**11.22**
TBATS	9.61	5.55	8.22	16.18	22.63	18.40	10.04	**12.95**
AR-ANN.uni	11.13	3.39	8.58	15.99	16.46	21.19	11.66	**12.63**
Multivariate Models								
AR-ANN.multi	8.49	3.69	6.23	16.28	15.44	18.00	8.95	**11.01**
LSTM	5.66	4.45	5.81	13.37	19.49	20.72	8.81	**11.19**
Combined								
Ensemble Average	7.05	3.01	4.50	17.63	14.83	17.76	7.62	**10.34**

Seasonal ARIMA has proven itself, as a reliable model as observed in the literature review. Through trial and error, our empirical results show that an ARIMA(5,0,0)(0,1,0)[168] model performed the best on a 24-hour horizon. Inclusion of moving averages was thus, not a vital part of the performance of the model; however, a seasonal differencing as the correction for stationarity was vital. Performing several other fits based on theoretical reasoning (ACF and PACF graph analysis of seasonally differenced data and residuals check) the model that produced comparable results had the following specification ARIMA(5,1,1)(0,1,1)[168] (additional non-seasonal differencing was included as a precaution against evidence of some autocorrelation processes left in differenced data). However, following the parsimony principle in forecasting procedure, the simplest model was selected as one potentially having higher generalization power. These empirical findings contradict some theoretical interpretations, but the fundamental requirement for stationarity that was tested using unit root tests had a significant impact on the model formulation. As a consequence, obtained results show relatively better forecast performance compared to seasonal naïve across the whole forecast horizon. Despite this, performance did not increase significantly during the volatile periods on Thursday and Saturday.

Double-Seasonal Holt-Winters method, first suggested by Taylor (2003), gave a more stable result compared to seasonal Naïve and particularly when comparing most volatile periods (Thursday and Saturday). However, it performed less well on Tuesday and Wednesday leading towards only a slightly lower average MAPE than the benchmark model.

Results from the proposed Exponential State Space Models gave an interesting observation. On the one hand, BATS outperformed all of the univariate models, modelling on a satisfactory level, compared to the benchmark. The final BATS model is formulated as BATS(0.596, {4,3}, 0.996, {24,168}), which captured both daily and weekly seasonality, in addition to the volatile periods. On the other hand, TBATS resulted in the lowest average MAPE of all models in this thesis, with generally high error margin within each weekday and weekend. The final optimal TBATS model is formulated as TBATS(0.665, {3,3}, 0.871, {<24,11>, <168,6>}).

For the final univariate model, an AR ANN(24,7,16)[24], that is an autoregressive neural network with 24 lagged variables($y_{t-1}, y_{t-2}, \ldots, y_{t-24}$), 7 of the last observed value of the 24-

hour seasonal period, and 16 neurons in the hidden layer, gave surprisingly low results with an average MAPE exactly similar to seasonal naïve. This result led to curiosity to what extent a multivariate version of the same model would increase or decrease the performance. After trial and error, including dummy variables for both Saturday and Sunday resulted in a model[30] that increased LFE performance on average and most of the days of the forecasting horizon. However, including any weather data generally decreased performance significantly.

With a multivariate LSTM, the performance gave an interesting result compared to the multivariate AR ANN. Performance increased in the weekdays, but less so in the weekend, giving a final average MAPE slightly higher than the multivariate AR ANN. Despite being in the same category, these two methods are different in terms of modelling time, feature preparation, and implementation. Recurrent Neural Networks that are utilized in TensorFlow accept 3-dimensional input with the following structure: number of samples, number of timesteps to look back, and the number of features. This feature precluded us from utilizing the expanding window cross-validation technique, instead, following the instructions from the official TensorFlow tutorial for time series data as well as Chollet, 2017 in the book "Deep Learning with Python" the rolling window technique was utilized. Otherwise, the changing structure of input would have required a slightly different model formulation on each iteration. Thus, the LSTM model was specified by setting four weeks of past load data as the lookback window. Features as the hour of the day, temperature, and dummies for each day of the week were utilized. Furthermore, the LSTM model works best when inputs are normalized to a particular scale, it was done by subtracting the mean from each series and dividing it by its standard deviation. The selection of layer structure is a somewhat arbitrary choice, at this point in time this process is still not guided by strict, theoretically backed up rules, thus through trial and error, the selected model was set to have 160 neurons in the hidden layer and 24 neurons in the output layer. It is worth noticing that LSTM performance was not stable, the average obtained results (for seven days) on test data ranged from 10.88 up to 12.2 MAPE. Thus, considering all the factors listed above, while the LSTM model deserves the attention it was decided not to make a direct comparison of LSTM to other presented models, but instead use it as a showcase that might be interesting to explore further.

[30] The model formulation is identical to the univariate AR ANN.

Finally, our proposed Ensemble Average model gave a clear advantage of combining the average forecast of the three best performing models (LSTM excluded). Despite a sum computational cost higher than any individual model, the performance increase is noticeable compared to both the best performing univariate and multivariate models. In Figure 16**,** each model and their point forecast are plotted against the actual load value at Energilab during the forecasting horizon[31].

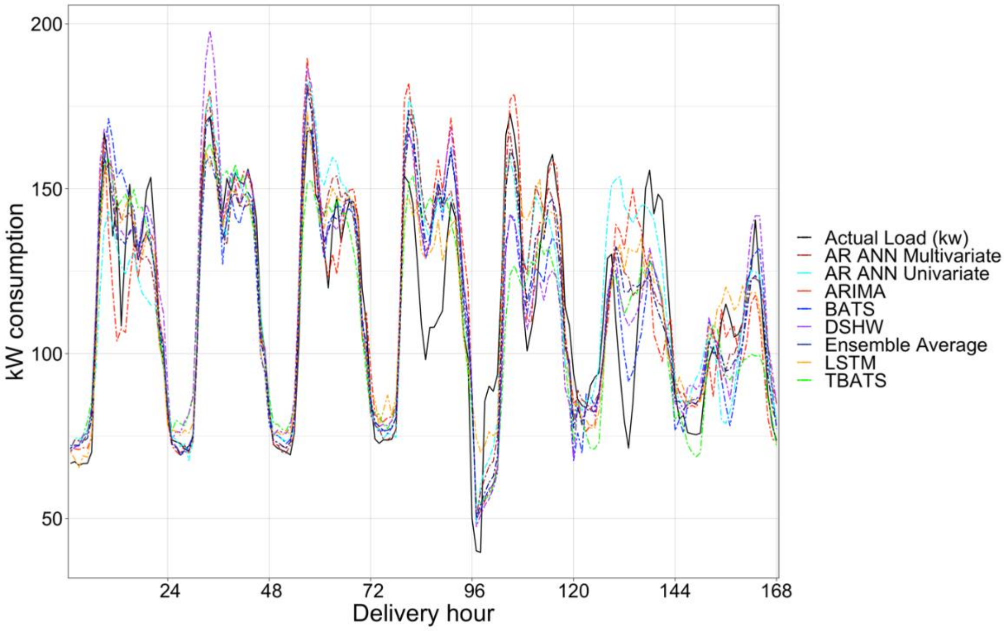

Figure 16 Point Forecast for Proposed Models

5.2.2 Performance Across Delivery Hours

Again, interest is turned towards the performance across the delivery hour of the forecasting horizon. Each of the proposed methods on a general basis does experience roughly the same patterns of errors each hour of the days that are forecasted. Errors are usually most prevalent in the hours between 7 AM and 5 PM, but each model does have its characteristics that are worth exploring. Most models experience a spike in LFE during the Thursday and Saturday with a shorter spike during early Friday morning. However, as depicted in Figure 17, BATS is again showing promising results by being the only model that relatively successfully predicted the drop-in load mid-day Saturday.

[31] A separate plot for each model can be exhibited in the appendix.

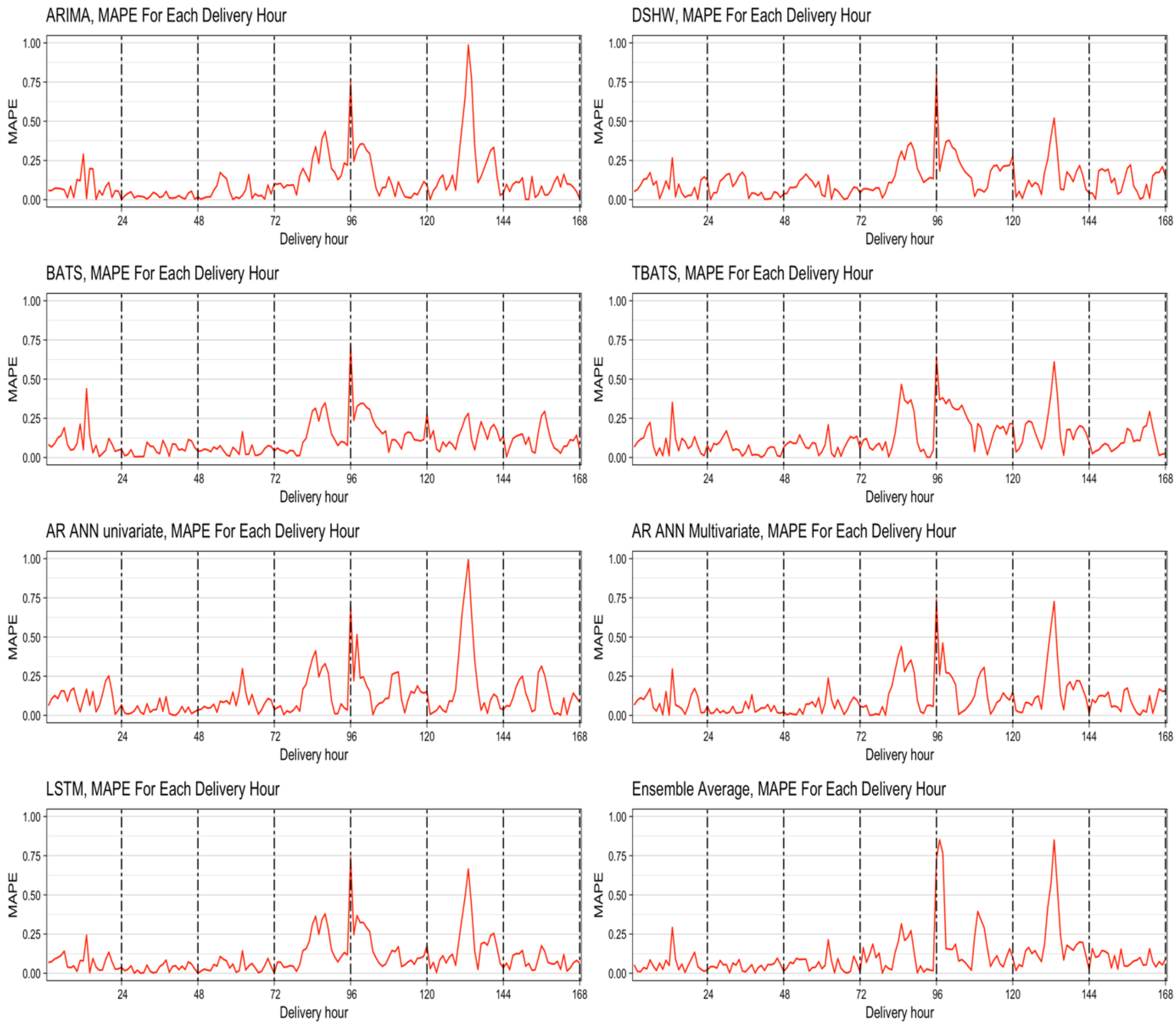

Figure 17 Proposed Models MAPE within each delivery hours

The proposed Ensemble Average model resulted in relatively accurate predictions during the first four days of the week, followed by LSTM and the multivariate AR ANN. Furthermore, the ensemble average model depicts the advantage of combining forecasting models to increase performance by the shorter duration of MAPE spikes. From the perspective of different methods, the univariate BATS model is highlighting the highest performance across delivery hours, while from a multivariate approach, LSTM and AR ANN would be perceived as on par.

5.2.3 Distribution of Forecasting Errors

The distribution of LFE from the proposed methods gave varying results. In Table 10, each result has been presented and compared towards the benchmark. As noted in the result section for seasonal naïve, there is observed a clear disproportionate representation of over predictions. Double-Seasonal Holt-Winters, Univariate AR ANN and the Multivariate AR ANN were models that surpassed the unequal distribution of seasonal naïve, however, except for the univariate AR ANN, these models did observe a relatively sizable drop in maximum deviance in kW.

Table 10 Distribution of Forecasting Errors - Proposed Models

	# Under pred.	# Over. pred	Max. Diviance	Avg. Diviance	**MAPE**
Benchmark Model					
S-naive	78	90	69.58	12.81	**12.63**
Univariate Models					
Arima	80	88	70.35	12.38	**11.70**
DSHW	60	108	40.20	13.33	**12.44**
BATS	83	85	47.47	12.20	**11.22**
TBATS	86	82	55.46	13.81	**12.95**
AR-ANN.uni	71	97	70.85	13.31	**12.63**
Multivariate Models					
AR-ANN.multi	77	91	51.82	11.59	**11.01**
LSTM	86	82	60.58	11.08	**11.19**
Combined					
Ensemble Average	84	84	47.42	10.87	**10.34**

The best performing model in terms of maximum deviance was DSHW, followed by Ensemble Average and BATS. In terms of average deviance however, the pattern is not the same. Ensemble Average resulted in a drop of roughly 2 kW on average, while LSTM and Multivariate AR ANN followed close by. Seasonal naïve results by average deviance were lower than most of the univariate models, except for BATS and DSHW. Ensemble Average has proven to show the most balanced results.

5.3 Economic Value of Results

As a final step in the result section, the economic performance of both benchmark and proposed models, based on the tailored economic loss function will be presented. Results will be displayed in PEC and compared to seasonal naïve.

5.3.1 Economic impact day-ahead forecast

Seasonal naïve, as a benchmark model, displayed an acceptable performance by being able to capture the multi-seasonal pattern at Energilab, but as explained in the subsection above, it also gave deviances and skewed distribution of the direction of errors. To what extent these results translate into economic performance is displayed in Table 11, where seasonal naïve had an increase in electricity regulation cost of 1.15% above the ideal cost within the week of operation.

Table 11 Economic Load Forecasting Error

	PEC	PEC difference from benchmark	MAPE	MAPE difference from Benchmark
Benchmark Model				
S.naive	1.1479 %	0	12.63 %	0
Univariate Models				
Arima	1.0652 %	0.0827	11.7 %	0.93
DSHW	1.125 %	0.0229	12.44 %	0.19
BATS	1.0223 %	0.1256	11.22 %	1.41
TBATS	1.2982 %	-0.1503	12.95 %	-0.32
AR.ANN.uni	1.1302 %	0.0177	12.63 %	0
Multivariate Models				
AR.ANN.multi	1.0899 %	0.058	11.01 %	1.62
LSTM	1.0867 %	0.0612	11.19 %	1.44
Combined				
Ensemble.Average	0.9204 %	0.2275	10.34 %	2.29

Ensemble average proved itself to be the most economically sound solution with an overall increase of 0.92% in regulation cost. Using the table above this result can be interpreted as follows: a 2.29% increase in forecasting accuracy (compared to benchmark) leads to a reduction of the additional incurred cost of regulation by 0.227 %. Following Ensemble average, the best performing models in terms of PEC were BATS, proving itself as not only a reliable model based on MAPE but also in terms of regulation cost. An interesting observation worth noticing is that seasonal ARIMA gave surprisingly better economic results than the multivariate models despite displaying a higher MAPE, maximum and average deviance. This observation reinforces the fact that the economic effect of over and under predictions might be different. Additionally, the presence of price spikes described previously might have a significant impact on overall incurred cost. Relatively, the said models performed substantially better compared to the last three proposed models. DSHW and Univariate AR ANN gave only a negligible PEC improvement. On the other hand, the most dejected results came from TBATS with the highest PEC of all models.

6.0 Discussion

Although machine learning and deep learning have gained massive popularity in the last decade and presented seemingly boundless opportunities, there still exist hurdles and obstacles to successfully implementing the said tools. Electricity markets are volatile and will most likely stay that way for a very long time. With the addition of collectively more microgrids within the traditional grid, analyzing and understanding this volatility could become even more complicated. However, organizations that wish to implement such tools for their day-to-day and market-based operation will have fewer obstacles if their current decision-making process is data-driven and well established. Also, it is not without saying that handling the ever so increasing mass of data must be an overall crucial foundation for companies in the power sector. In this section, we provide an in-depth discussion surrounding the forecasting results, economic considerations related to the accuracy of forecasts, as well as benefits and challenges of utilizing load forecasts in a microgrid environment.

6.1 Discussion of Forecast Results

The volatile nature of load at such a small resolution as presented at Skagerak Energilab is posing significant difficulties for the forecasting objectives, and it is a challenging task. The seemingly repetitive pattern throughout the week is highly susceptible to sudden unexpected changes in the overall load level, thus, producing hurdles for the predictability. All in all, substantiated by logical reasoning, the benchmark model used in studies, set a relatively high standard for consequent improvement potential. However, as it was displayed in the result section, most statistical and artificial intelligence-based methods were capable of producing moderately enhanced forecasts for the intended objective.

The most accurate and stable performance was obtained when using ARIMA, BATS, and multivariate AR ANN (performance of LSTM was somewhat volatile; thus, it is not included). While being different, all mentioned models successfully captured the slight variations in the load, however, certain shocks such as surprisingly low consumption during the midday on Thursday, a sudden drop in early hours on Friday, and unusual collapse on Saturday were poorly anticipated. The use of Ensemble Average, while also not capturing perfectly the unexpected shocks, nevertheless proved to be a reasonable choice, as the model produced a well-balanced (equal number of over and under predictions) and most accurate forecasts.

In short term load forecasting the historical data conveys the primary source of signal for future values, analyzing the past load it is evident that shocks observed during the “test week” were not present previously, nor could it be attributed to unusual weather (as no such was observed). As it was noted by Taylor (2003) univariate methods are considered to be sufficient for the short lead times involved because the weather variables tend to change in a smooth fashion, which will be captured in the demand series itself (Taylor, 2003). Thus, we can assume that unusual load patterns during “test week” might have been caused by some external (not present in the available data), most probably, socio-economic factors. Scheduled maintenance or equipment breakdowns of certain commercial customers could have caused described shocks. At a small resolution of electricity consumption, such an unexpected activity resonates with a significant change in the usual pattern of the load.

Additionally, analyzing the results, one can spot that not all models managed to outperform the benchmark. From both a statistical and economic perspective, the TBATS model generally performed worse than seasonal naïve on every evaluation metric. DSHW on the other hand, showed stable performance but without any significant improvement over the benchmark.

For the sake of forecasting accuracy, the data used in the study consisted of aggregated load values for all customers of all types, no additional information was available for disclosure. These customers represent a wide variety of consumption patterns. In general, businesses usually depict a relatively stable consumption within predictable patterns in terms of time and day, while residential customers might have a higher degree of volatility. Disaggregating the dataset into different types of customers or even individual consumers could detect interesting patterns not present within the current dataset. This might lead to deeper insights into the underlying causes of load volatility.

Henceforth, systematically collecting data for each type of customer (while respecting customers' privacy), as well as keeping a comprehensive record of socio-economic events that have a direct effect on load can potentially give some useful tools for following forecasting tasks within a microgrid environment.

6.2 Discussion of Economic Result

To bridge the forecasting results with particular economic rationale specifically tailored ELFE function was proposed. For scalability reasons and ease of interpretation, results were transformed to a percentage scale (PEC function). Obtained figures in Table 11 are displaying a link between forecast improvement and economic benefit associated with it. Ensemble average as a model with lower forecasting error and balanced distribution of under and over forecast was also proven to show the best performance when assessed by an economic loss function. As a general rule lowering the MAPE tends to produce lower PEC. However, this conclusion is not as straightforward as it appears. The previously discussed fact that in an economic context, consequences of under and over forecast can bear different costs manifested itself in the undertaken study. As was emphasized in the result section, though, the ARIMA model displayed a higher MAPE, it also produced better economic results than the multivariate AR ANN and LSTM.

Setting the minimization of economic performance as the primary goal of forecasting procedure might spur the idea of producing asymmetric loss function, which would severely punish the direction of forecasts that have higher costs. However, one should be cautious in these attempts. First of all, the presence of unpredictable price spikes can undermine the reasoning of such an initiative. Additionally, the benefit of well-balanced forecasting performance has many things to offer from a managerial perspective that can also be translated into the direct economic profit. For instance, an analysis of average and maximum deviance of a symmetric forecasting error might hint towards the deployment of an appropriate energy storage technology within a microgrid.

Further research of load forecasting effect on the economic operation is complicated due to the fact that most operational data is classified due to commercial reasons, and to our knowledge, it is not available for public access. Proposed economic assessment criteria are most suitable for utility companies that directly participate in the power market. However, the price of electricity for all users is linked to the results of day-ahead market operations. Thus, it allowed us to establish a link and evaluate economic loss due to imperfect forecasts.

6.3 Consequences of Missing Data

As initially explained, Energilab has been operational within just under one year. Thus, collected data is limited to the end of October, and as of writing this thesis, it is still collected. The consequence of this is that only a restricted number of seasons were represented. Norway, as an operational area with a high degree of hydropower typically is characterized by having flood and drought periods creating a multitude of incentives, throughout the different seasons, in terms of regulating market prices. Delving into further data accumulation over different periods would be beneficial for new potential studies and Skagerak Energilab in general. Doing so could reveal patterns that potentially would motivate new sets of loss functions and model specifications.

Furthermore, as displayed in the data section, a significant degree of missing data was observed at the end of March 2020 and data collected throughout April 2020. As explained, imputing these missing observations as a means of collecting more data creates obstacles for a robust empirical analysis of both the load curve and forecasting procedure. As imputed observations would already represent predicted values. For the successful development of the research project at Energilab and a higher degree of comprehension of the customer load pattern, the goal should be set to increase the quality of the collected data. This practice would boost potential accuracy of future forecasts practices, it will also decrease economic cost due to better analysis and ensure operational security to incentivize developments of RES within a microgrid further. However, as these missing observations arise from maintenance and simulation runs to further understand the capability of the project, Skagerak should settle on goals to minimize the number of unnecessary downtimes as it would favour a balance between simulated analysis and empirical analysis.

6.4 Robustness and Implementation

As our results indicate, our proposed methods are able to predict day-ahead load levels at Energilab with an acceptable level of accuracy. As deducted in the data description, load at Energilab has the presence of a highly cyclical pattern which to some extent, could be predicted to a high degree of accuracy with a relatively simple method like seasonal naïve. However, as displayed in the result section, volatile load levels at Energilab present challenges that were not modelled correctly by our simple benchmark, further stressing the need for more advanced methods. However, as with everything in life, nothing comes for free, to that extent that the proposed methods do include a higher level of computational time. Henceforth, as data observations grow over time, the real-life implementation of the proposed methods would meet some distinct barriers from a daily operational perspective if the models were to be re-trained and updated iteratively every day. This is especially true for the ensemble average model that are a combination of multiple time-consumption methods, further highlighting the need for automation if Skagerak were to implement the proposed tools in their daily operation successfully.

Also, there is a substantial benefit with these methods in which it lowers the barricade of long-term, in-depth domain knowledge of load forecasting, that could become lost if knowledgeable human resources no longer become available. Chiefly, deep learning methods like LSTM and AR ANN represents models that learn and decide future values by a mathematical representation of neurons, and together, if these methods integrated and developed with knowledgeable employees, could create even more accurate models over time. In addition to withstanding the potential danger of losing domain knowledge.

Lastly, Energilab with its small number of connected customers would favourably have an increased level of stable load curve if Energilab and Skagerak were to influence and incentivize the customers to exercise a less volatile consumption pattern. Integration of programs and cost-cutting behaviours over time could be such an incentive. These programs, if properly managed, could potentially yield a consumption pattern that was less volatile and henceforth increasing forecasting accuracy, creating a larger socio-economic profit with a decreased regulation cost for Skagerak as a company, and reduced electricity cost for the customers.

6.5 Challenges of a Microgrid

Referring back to Hirsh, Parag and Guerrero (2018), motivational categories behind the development of microgrids in the modern electricity market, as energy security and clean energy integration, are not perceived as a driving force behind the development of Energilab, but rather as a research project for future economic benefit in cases where energy must be transmitted to remote areas like islands and far away from urban areas. As maintenance and investments cost increase, systems like Energilab can draw energy companies into investing in

RES and DER as they are in theory able to self-sustain areas without the need for additional costly main grid connection. Being able to predict future load values confidently is thus recognized as a primary objective for the successful development of microgrids. As our result portrays, volatility does play an innermost role in the performance of all our proposed methods. Part of the observed volatility could be linked to the low number of customers connected. By not being able to benefit from a larger sample size mean, it introduces additional issues in terms of low confidence in battery discharge and charging optimization process in addition to the overall size of the RES and riskiness of total development costs. Again, disaggregation of customer type would collect additional information that could be valuable for development of similar systems in the future.

Furthermore, the results of this thesis are on par with what was observed by (Marinescu, Harris, Dusparic, Clarke, & Cahill, 2013)**,** (Kong, et al., 2017) and (Ahlert, 2010), in which all three found evidence that load forecasting and the subsequent results correlate highly with the area size and number of customers, despite the level of advanced method on wish to choose from. Microgrids thus inherit a fundamental issue related to energy security and raise debatable questions if the benefits of clean energy with potentially lower investment cost do outweigh the costs of lower energy security and customer satisfaction. These potential drawbacks, as presented in the result section, gave an overall idea about the direct cost incurred in terms of regulation market participation, leading towards higher energy costs for both parties. However, it could also in the long-term lead to increased indirect cost in terms of lower brand recognition and customer satisfaction.

6.6 Preferred Method of Choice at Skagerak Energilab

According to our results, the preferred method of choice based on both evaluation metrics is the ensemble average model, combining seasonal ARIMA, BATS, and Multivariate AR ANN it yields the most accurate load predictions at Energilab. The preferred method, despite its high total computational time, outperformed every single model and method in this thesis and most significantly decreased the economic impact of LFE in comparison to seasonal naïve. That being said, if computational time would become a matter of concern in the daily operational perspective from Skagerak's point of view, our proposed univariate BATS model resulted in the second-best economic performance after ensemble average and was the most accurate univariate model of all the proposed univariate models.

Nevertheless, if additional control variables would be regarded as a useful complement to short term load forecasting, both multivariate AR ANN and LSTM models can be successfully implemented. Based upon the data preprocessing, AR ANN makes more explicit use of AR components and is more user friendly to operate with, while LSTM does need a separate approach in terms of preprocessing and model optimization, in addition to significantly increased computational time.

6.7 Further Research

The master thesis represents the first and initial study into the load pattern at Energilab, thus representing a benchmark for further research, for a given microgrid or any other similar project. As the results depict, the level of forecasting accuracy in a microgrid setting still is not on the same level compared to load forecasting on a larger regional grid. Thus, different projects which in turn could lower the volatility, are recommended to develop more precise forecasts further. We find this important if further microgrid projects should become a reality in the Nordic power market.

Also, as more data grow over time, further research should be projected towards forecasting tools tailored explicitly to microgrids. A result of this could become microgrid specific forecasting models, that excel in both MAPE and ELFE. Integration of socio-economic factors in the modelling process, as well as customer grouping techniques, could be a promising field of research. Additional insights about economic data might reveal information that would be possible to tailor into a specific loss function for achieving lower economic cost.

7.0 Conclusion

Load forecasting has been a fundamental necessity for the power industry since its inception. The multitude of methods that have been proposed to accomplish this task varies in complexity and ingenuity so that even an experienced researcher can be bewildered when trying to approach this task. However, it is essential to understand that an optimal universal technique, that suits every setting, simply does not exist. The context of the problem, the nature of the data, socio-economic interactions, and many other factors all play a notable role in the selection of an appropriate load forecasting technique.

With a varying degree of success, models presented in this thesis were able to capture the dynamics of load in the microgrid. Once again, it was proven that combining forecasts can lead to better accuracy. As an attempt to translate improved forecast performance to the actual economic gain, specific evaluation criteria based on the physical power market was established. Obtained results can be formulated as follows: 2.29%-point increase in forecasting accuracy can be associated with a 0.23% decrease in variable consumption cost (seasonal naïve - Ensemble Average), due to lesser activity in the regulatory market. The numbers are notably close to a study undertaken by Hobbs, Jitprapaikulsarn and Maratukulam (1999), where the authors analyzed the value of improved load forecasts for the unit commitment and derived that: "reduction of 1% in mean absolute percentage error (MAPE) decreases variable generation costs by approximately 0.1 %-0.3% when MAPE is in the range of 3%-5% " (Hobbs, Jitprapaikulsarn, & Maratukulam, 1999). Though the undertaken approach was different, and their focus was put on generation rather than consumption, the evaluation criterion was based on the day-ahead market structure. Thus, those activities representing two sides of the same process can potentially yield comparable results.

Though it was also shown that economic consequences of under and over forecasts could have different costs, it can be still considered optimal to derive well balanced and as accurate predictions as possible. For this reason, it is important to set good management practices to control for sudden and unexpected changes in load values. A comprehensive and thorough record of past data, as well as potential integration of demand response, would theoretically lower volatility and improve forecast performance.

8.0 References

Ahlert, K.-H. (2010). *Economics of Distributed Storage Systems : an economic analysis of arbitrage-maximizing storage systems at the end consumer level.* Germany: KIT Scientific Publishing.

Bang, C., Fock, F., & Togeby, M. (2011). *The existing Nordic regulating power market.* Denmark: Ea Energy Analyses.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks, Vol. 5, No. 2*, pp. 157-166.

Berkeley Lab. (2019). *Microgrids at Berkeley Lab.* Retrieved from https://building-microgrid.lbl.gov/microgrid-definitions

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). *Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches.* Energies.

Box, G. P., & Jenkins, G. M. (1970). *Time-series analysis, forecasting and control.* San Francsico, California: Holden-Day.

Box, G. P., Jenkins, G. M., & Reinsel, G. (1994). *Time-series analysis, forecasting and control (3rd ed.).* Upper Saddle River, New Jersey: Prentice-Hall.

Chollet, F. (2017). *Deep Learning with Python.* Manning Publications.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2010). *Forecasting Time Series with Complex Seasonal Patterns usng Exponential Smoothing.* Australia: Monash University, Department of Econometrics and Business Statistics.

Diamantoulakis, P. D., Kapinas, V. M., & Karagiannidis, G. K. (2015). *Big Data Analytics for Dynamic Energy Management in Smart Grids.* Greece: Big Data Research.

Duffner, S. (2012). *Forecasting German day-ahead electricity prices using multivariate time series models.* Bergen: NHH Master Thesis.

Fahad, M. U., & Arbab, N. (2014). Factor Affecting Short Term Load Forecasting. *Journal of Clean Energy Technologies, Vol. 2, No.4*, pp. 305-309.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* London, England: The MIT press.

Guo, Q., Feng, Y., Sun, X., & Zhang, L. (2017). *Power Demand Forecasting and Application Based on SVR.* Information Technology and Quantitative Management.

Hagan, M. T., & Behr, S. M. (1987). The Time Series Approach to Short Term Load Forecating. *IEEE Transaction on Power Systems, Vol. 2,* pp. 785-791.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning, 12th printing.* Stanford, California: Springer.

He, W. (2017). Load Forecasting via Deep Neural Networks. *Information Technology and Quantitative Management (ITQM 2017),* pp. 308-314.

Hipel, K. W., & McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems.* Amsterdam: Elsevier.

Hirsh, A., Parag, Y., & Guerrero, J. (2018). Microgrids: A Review of Technologies, Key Drivers, and Outstanding Issues. *Renewable and Sustainable Energy Reviews,* pp. 402-411.

Hobbs, B. F., Jitprapaikulsarn, S., & Maratukulam, D. J. (1999, November). Analysis of the Value for Unit Commitment of Improved Load Forecasts. *IEEE Transactionson Powcr Systems,Vol. 14,No. 4*, pp. 1342-1348.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation 9(8)*, pp. 1735-1780.

Holstad, M., & Pettersen, F. L. (2011). *Hvordan reagerer strømforbruket i alminnelig forsyning på endringer i spotpris.* SSB.

Hong, T., Pinson, P., & Fan, S. (2014). Global Energy Forecasting Competition 2012. *International Journal of Forecasting Vol. 7*, pp. 357-363.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and Beyond. *International Journal of Forecasting*, pp. 896-913.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice, 2nd edition.* Melbourne, Australia: OText.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothning: The State Space Approach.* Springer Science and Business Media.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). *A state space framework for automatic forecasting using exponential smoothing methods.* Australia: Monash University, Department of Econometrics and Business Statistics.

Juberias, G., Yunta, R., Monero, J. G., & Mendivil, C. (1999). A New ARIMA Model For Hourly Load Forecasting. *Transmission and Distribution Confrence, IEEE, Vol. 1*, pp. 314-319.

Kobylinski, P., Wierzbowski, M., & Piotrowski, K. (2020). *High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources*. Poland: Elsevier.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network. *IEEE Transaction on Smart Grid*.

Kwiatkowski, D., Phillips, P. B., Schmidt, P., & Shin, Y. (1992). Testeing the Null Hypothesis of Stationarity against the Alternative Unit Root. *Journal of Econometrics 54*, pp. 159-178.

Kwiatkowski, D., Phillips, P. B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics 54,* pp. 159-178.

Lin, J. L., & Santra , A. S. (2019, May 28). *Applications of Computational Intelligence to Power Systems*. Retrieved from Google Books: https://books.google.no/books?id=sAa9DwAAQBAJ&pg=PA2&lpg=PA2&dq=load+forecasting+has+research+1960&source=bl&ots=Qt-KmUJjCR&sig=ACfU3U2N3r2z1RDJATTcSaRB0E2UgPX9zg&hl=no&sa=X&ved=2ahUKEwizm4WQjtHnAhXswcQBHVM0CTQQ6AEwAXoECAsQAQ#v=onepage&q=1960&f=false

Marinescu, A., Harris, C., Dusparic, I., Clarke, S., & Cahill, V. (2013). Residential Electrical Demand Forecasting in Very Small Scale: An Evaluation of Forecasting Methods. *Software Engineering Challenges for the Smart Grid (SE4SG), 2nd International Workshop*, pp. 25-32.

Norsk Vassdrag og Energidirektorat. (2020, Februar 12). *NVE*. Retrieved from NVE.no: - https://www.nve.no/reguleringsmyndigheten/nettjenester/nettleie/utformingen-av-overforingsnettet-i-norge/

Olah, C. (2015, August 15). *Understanding LSTM Networks*. Retrieved from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Olje og Energidepartementet. (2020a, februar 12). *Energifakta*. Retrieved from Energifakta.no: http://energifaktanorge.no/norsk-energiforsyning/kraftmarkedet/

Olje og Energidepartementet. (2020b, Februar 12). *Energifakta*. Retrieved from Energifakta.no: http://energifaktanorge.no/norsk-energiforsyning/forsyningssikkerhet/

Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A Regression-Based Approach to Short-Term System Load Forecasting. *IEEE Transactions on Power Systems, Vol. 5*, pp. 1535-1547.

Peng, Y., Wang, Y., Lu, X., Li, H., Shi, D., Wang, Z., & Li, J. (2019). *Short-term Load Forecasting at Different Aggregation Levels with Predictability Analysis.*

Rahman, A., Smith, A. D., & Srikumar, V. (2017, February). *Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks.* Utah, USA: University of Utah.

Ranaweera, D. K., Karady, G. G., & Farmer, R. G. (1997, August). Economic Impact Analysis of Load Forecasting. *IEEE Transactions on Power Systems, Vol. 12, No. 3,* pp. 1388-1392.

Rodrigues , F., Cardeira, C., & Calado, J. F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal. *The Mediterranean Green Energy Forum 2014, MGEF-14*, pp. 220-229.

Sangrody, H., & Zhou, N. (2016). *An initial study on load forecasting considering economic factors.* University of New York: IEEE.

Shahidehour, M., Yamin, H., & Li, Z. (2002). *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management.* New York, USA: John Wiley & Sons, Inc.

Skagerak Energi. (2019). *Skagerak Energilab.* Retrieved from Skagerakenergilab.no: https://www.skagerakenergilab.no/forskning-og-utvikling/category2191.html

Skagerak Energi. (2020, February 11). *Skagerak Energi*. Retrieved from Om oss: http://www.skagerakenergi.no/om-oss/

Statnett. (2018, August). *Statnett.no*. Retrieved from Statnett: - https://www.statnett.no/contentassets/d27d9d5efd7a4371abe2b17c97ef4a64/27-august-2018-roller-i-balansemarkedene-og-aggregering.pdf

Statnett. (2020a, Februar 12). *Statnett*. Retrieved from Statnett.no: https://www.statnett.no/for-aktorer-i-kraftbransjen/tall-og-data-fra-kraftsystemet

Statnett. (2020b, Februar 12). *Statnett*. Retrieved from Statnett.no: - https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/

Tang, L., Wang, C., & Wang, S. (2013). Energy Time Series Data Analysis based on a Novel Integrated Data Characteristic Testing Approach. *Procedia Computer Science, vol 17*, pp. 759-769.

Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting using Double Seasonal Exponential Smoothing. *Journal of the Operational Research Society, 54,* pp. 799-805.

Wang, Y., & Wu, L. (2017, July). *Improving Economic Values of Day-Ahead Load Forecasts to Real-Time Power System Operations.* USA: IET The Insititution of Engineering and Technology.

Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices.* England: John Wiley & Sons Ltd.

Zedda, S., & Masala, G. (2019). Price spikes in the electricity markets: how and why. *3rd HAEE Annual Conference: Energy Transition: European and Global Perspectives.* Italy: University of Cagliari, Department of Economics and Business Science.

Zhang, P. G. (2007). A Neural Network Ensemble Method with Jittered Training Data for Time Series Forecasting. *Elsevier, Information Science*, pp. 5329-5346.

Ziel, F. (2018). *Quantile Regression for the Qualifying Match of GEFCom2017 Probabilistic Load Forecasting.* International Journal of Forecasting.

Appendix

A1 Missing Load Data Plot

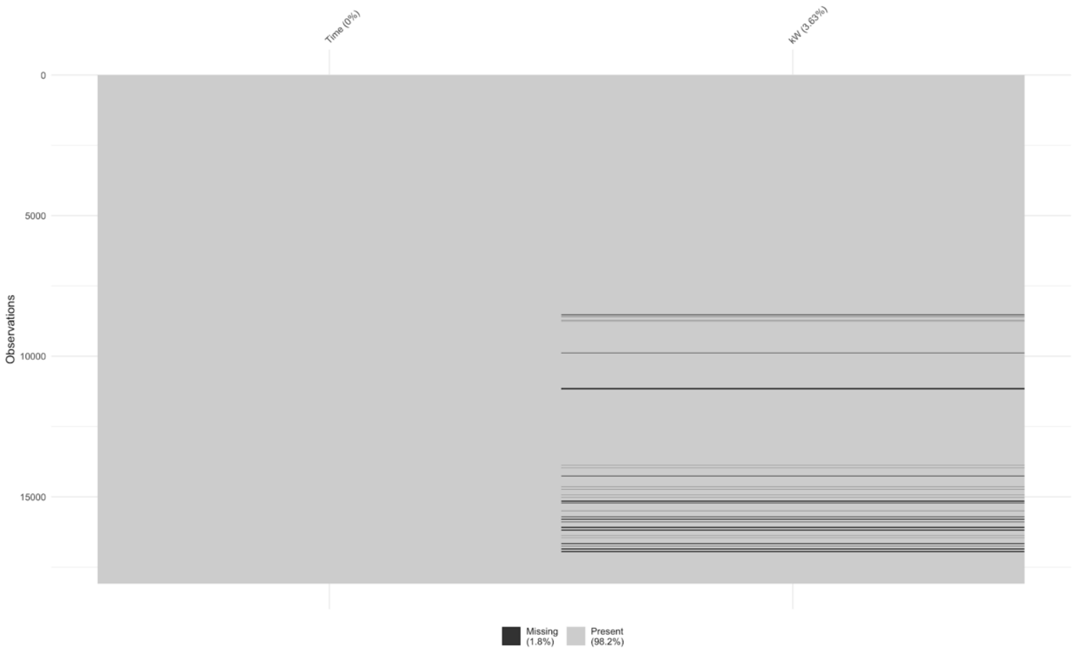

Figure A 1 Missing Load Data Plot

A2 The Duck Chart

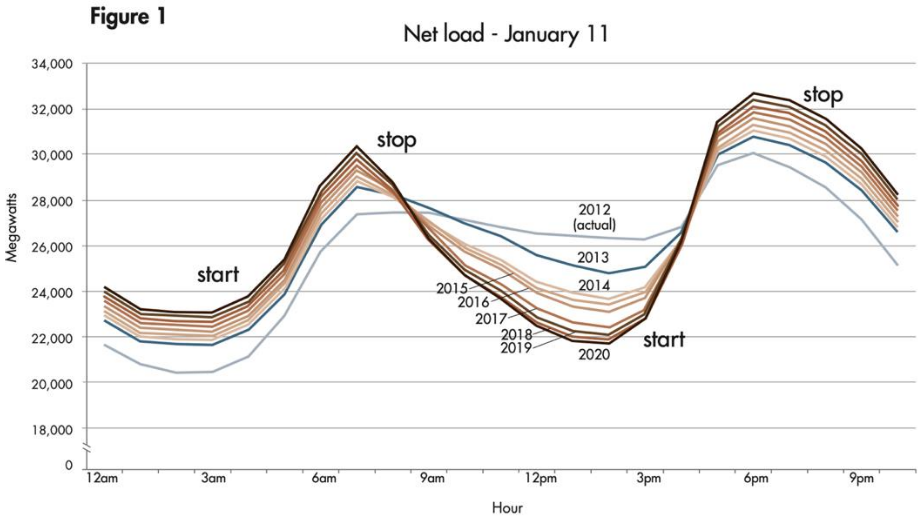

Figure A 2 The Duck Chart. Source: CASIO[32]

[32] The Figure can be found on: **https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf**

A3 Summary Statistics of Load Data by Weekday and Weekend

Table A 1 Summary Statistics of Load Data by Weekday and Weekend

Weekday						Weekend					
Hour	Min	Median	Mean	Max	St.Dev.	Hour	Min	Median	Mean	Max	St.Dev.
0	40.2	81.4	82.5	122.5	10.5	0	70.0	86.9	88.3	113.9	12.0
1	39.7	80.5	81.8	121.0	10.7	1	66.7	84.2	84.9	112.2	11.6
2	66.1	79.0	80.9	112.4	9.6	2	64.7	83.5	83.1	108.8	10.8
3	48.6	78.3	79.4	110.3	11.9	3	51.5	81.8	80.9	107.4	12.5
4	66.7	79.9	82.8	112.6	10.7	4	64.8	82.5	83.4	109.9	11.3
5	70.1	83.7	86.4	116.7	11.5	5	67.8	84.9	85.4	116.9	11.9
6	94.1	118.0	119.4	154.5	13.1	6	75.9	92.4	97.2	141.3	14.1
7	112.2	154.1	154.6	188.2	14.6	7	82.9	107.4	109.8	157.3	15.5
8	115.4	174.1	173.3	225.3	20.6	8	77.7	118.3	117.1	165.5	19.9
9	90.8	174.1	170.6	225.0	23.5	9	59.8	120.8	117.8	167.8	21.4
10	96.9	170.9	166.1	285.4	30.6	10	47.0	118.8	115.0	147.3	23.6
11	70.7	156.7	153.8	277.8	33.6	11	57.9	121.3	112.2	148.4	25.5
12	64.7	158.7	158.0	260.7	38.8	12	49.7	120.5	116.1	153.0	24.9
13	80.4	165.1	158.1	223.9	29.6	13	57.9	121.2	117.2	151.4	24.2
14	77.7	162.9	157.1	214.2	26.2	14	59.6	121.8	120.3	189.5	25.5
15	69.6	163.2	160.2	230.9	23.6	15	72.2	125.2	124.5	205.6	25.8
16	81.9	155.2	154.5	233.5	23.3	16	83.6	130.9	131.4	226.3	26.7
17	96.1	152.7	153.2	231.9	19.9	17	98.0	128.2	133.8	261.2	29.1
18	109.9	151.8	152.7	281.2	22.4	18	99.5	121.1	128.7	246.6	27.4
19	95.1	143.4	145.6	279.7	24.5	19	87.2	116.7	124.1	249.0	29.9
20	106.5	134.5	139.3	286.3	26.1	20	88.7	109.4	118.0	208.9	26.2
21	90.8	112.2	116.6	267.2	21.7	21	78.0	96.0	102.3	180.9	20.5
22	88.1	103.3	105.8	210.2	14.2	22	74.2	89.6	92.3	125.2	12.8
23	49.9	86.9	88.4	127.5	9.2	23	72.7	83.2	85.2	115.9	9.8

A4 Error Heat Map (MAPE) for Benchmark and Proposed Models

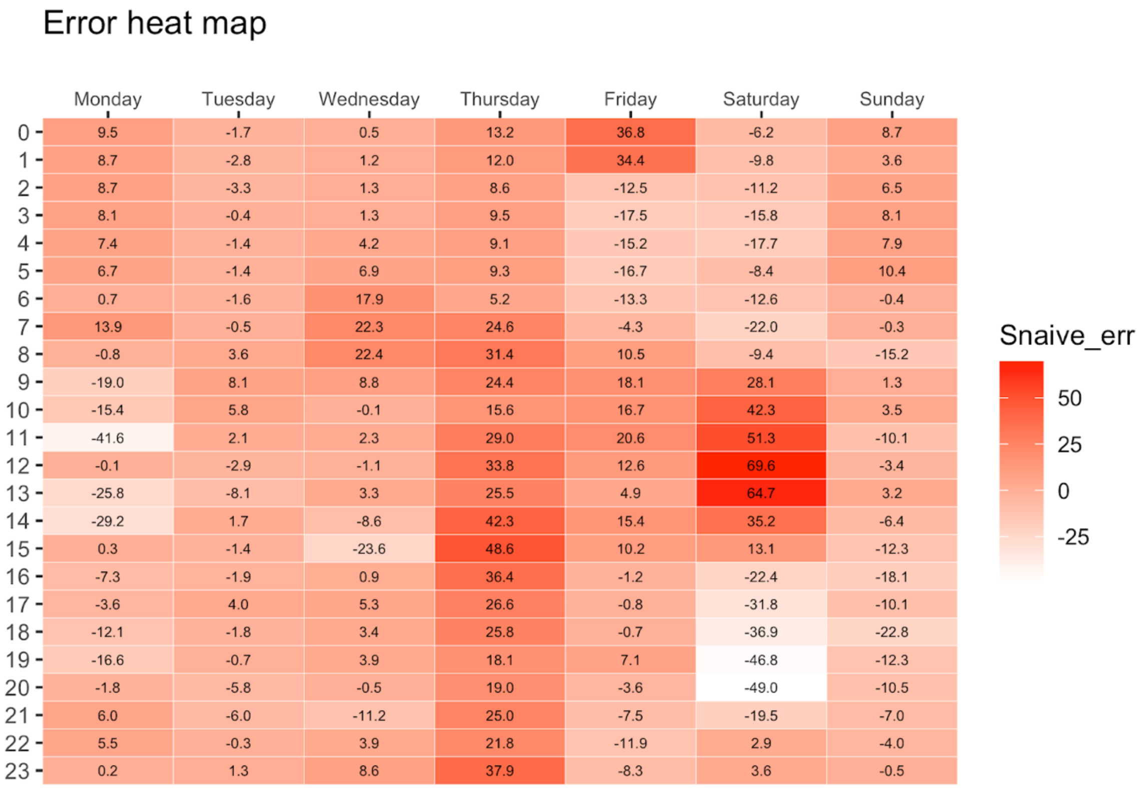

Figure A 3 Error Heat Map Seasonal Naïve

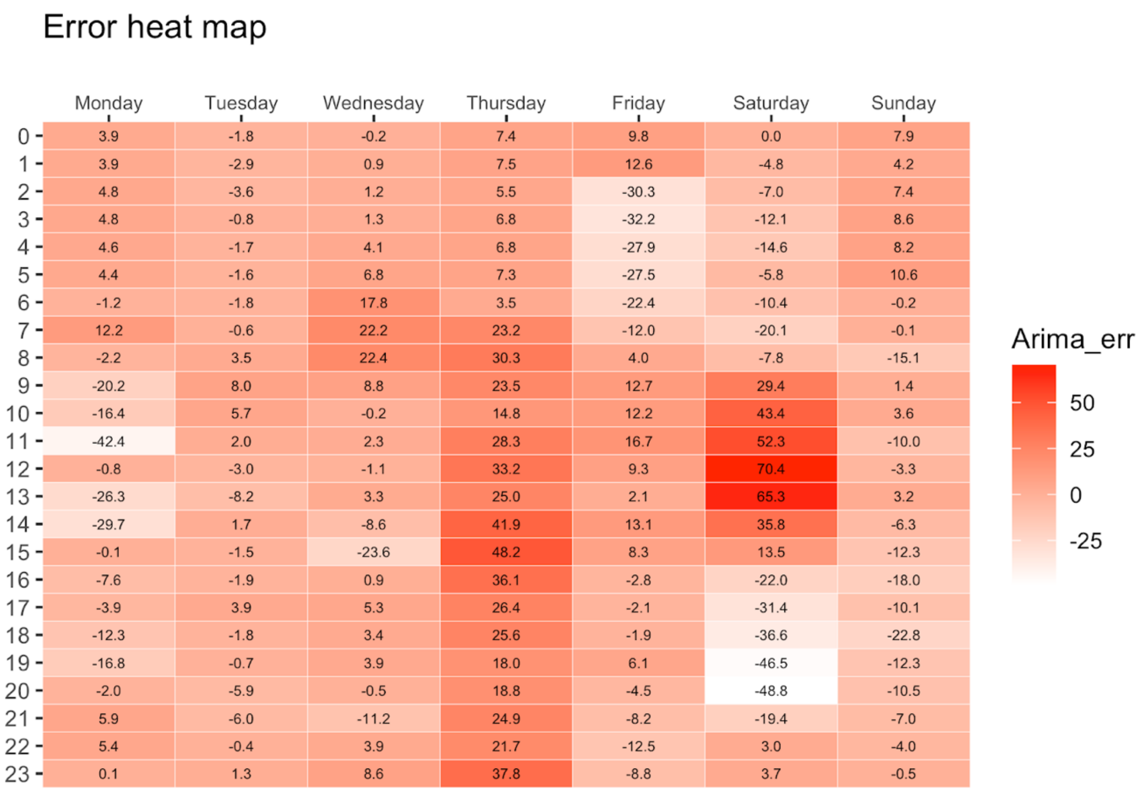

Figure A 4 Error Heat Map Seasonal ARIMA

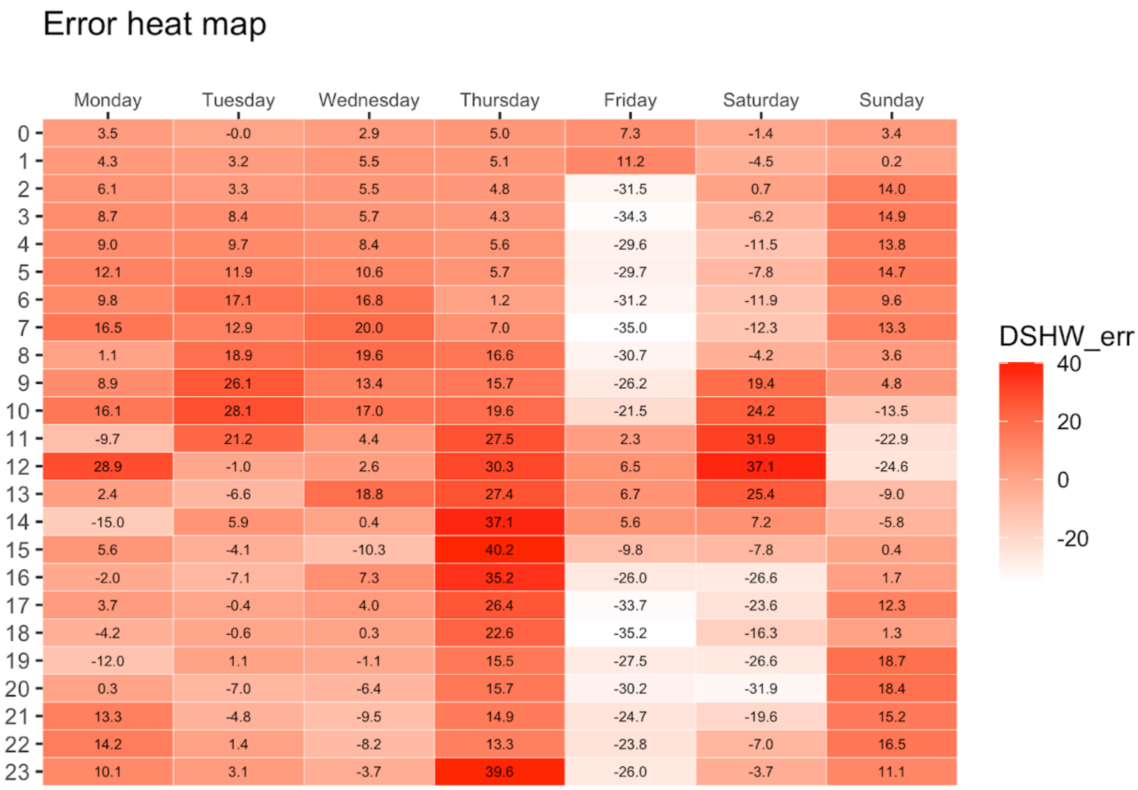

Figure A 5 Error Heat Map DSHW

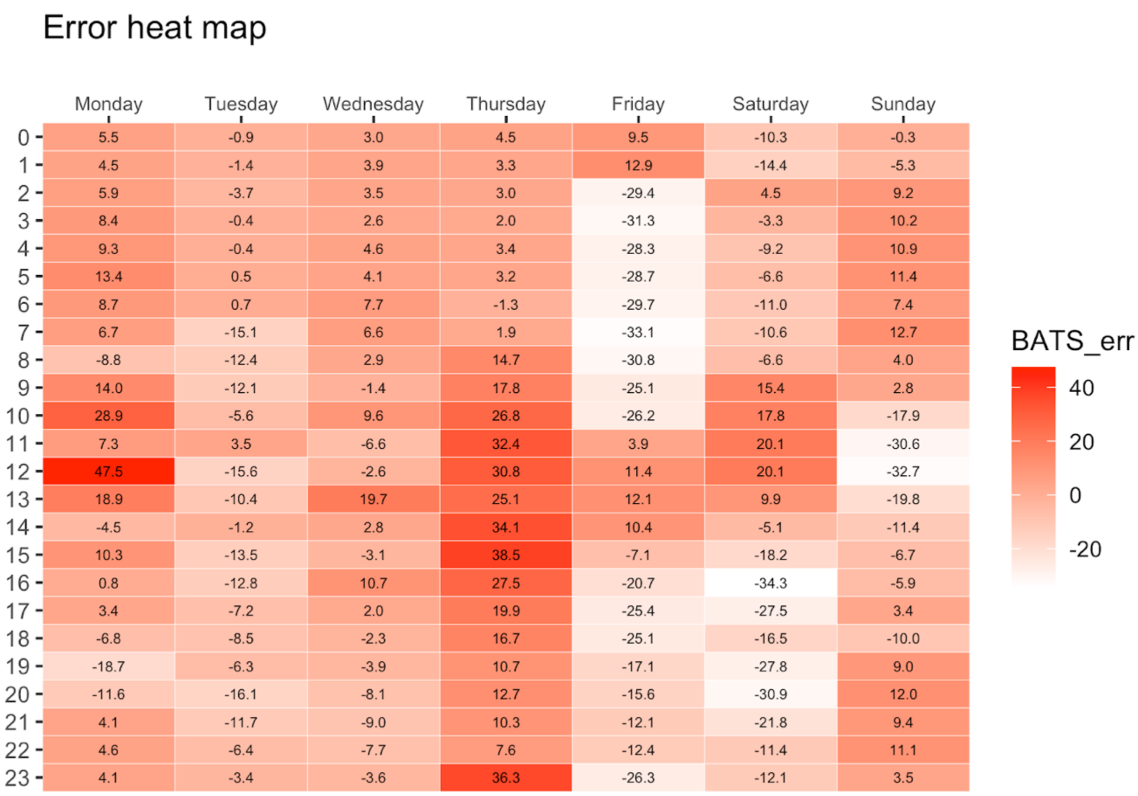

Figure A 6 Error Heat Map BATS

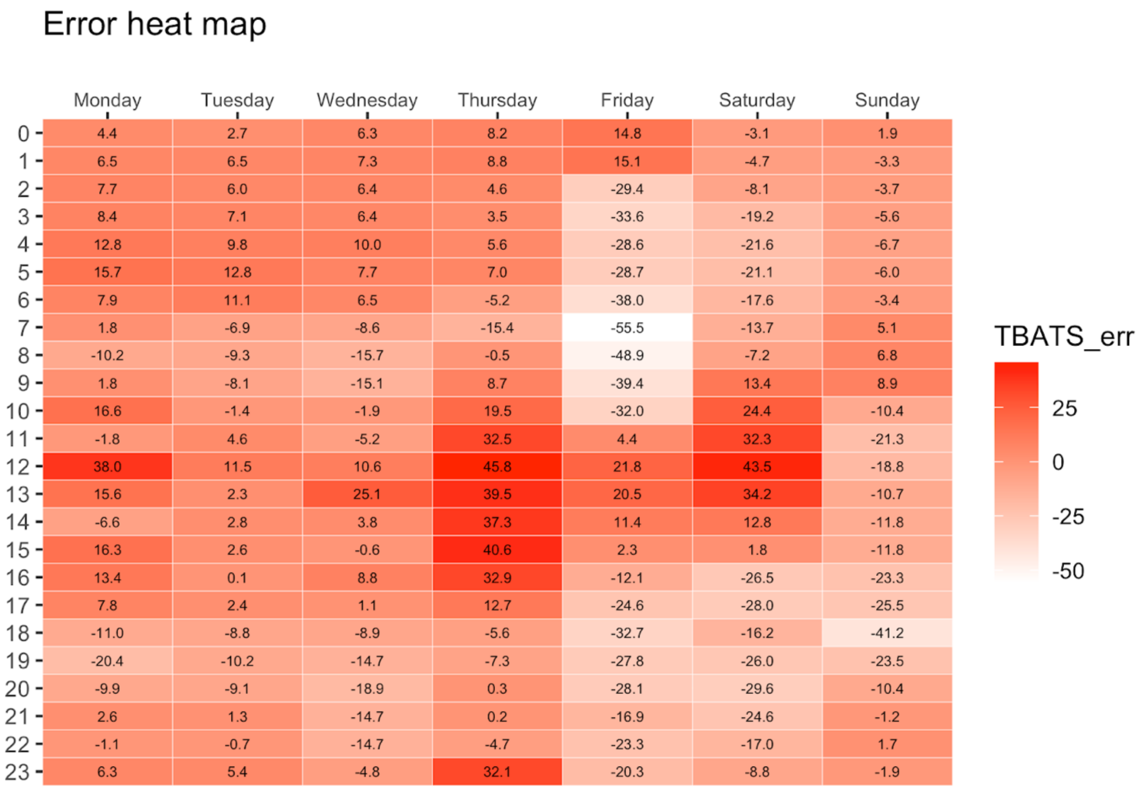

Figure A 7 Error Heat Map TBATS

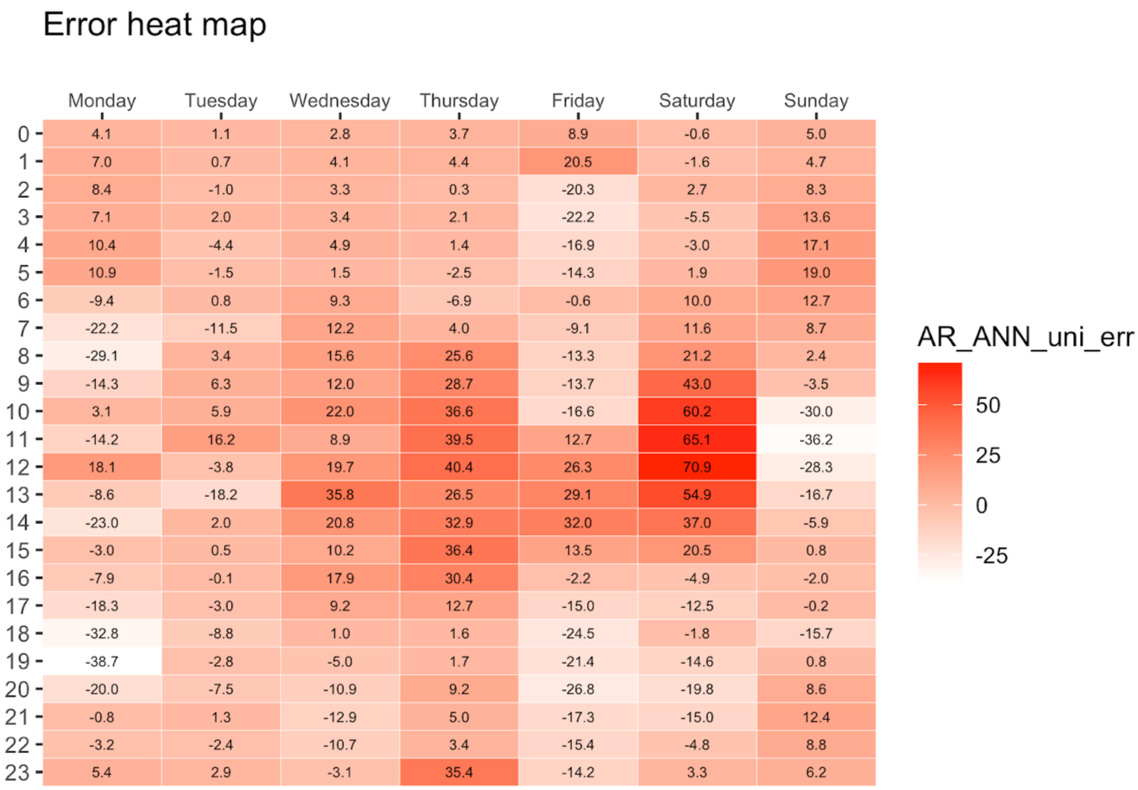

Figure A 8 Error Heat Map Univariate AR ANN

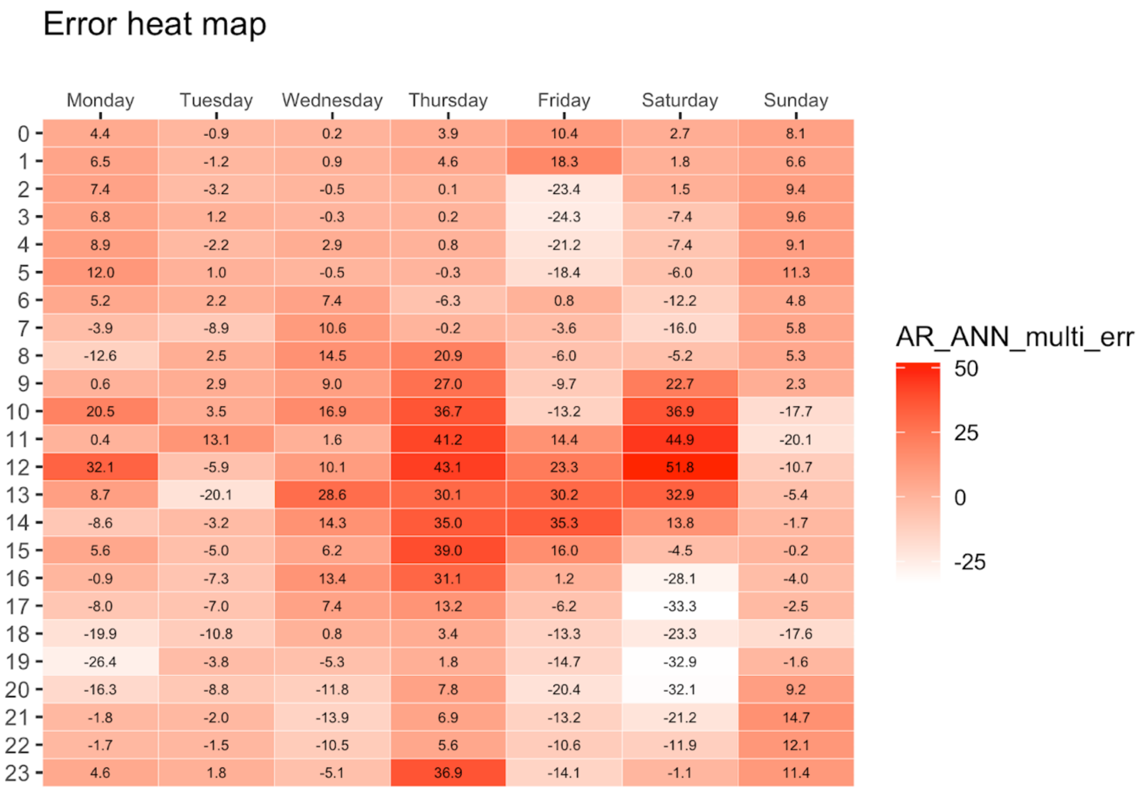

Figure A 9 Error Heat Map Multivariate AR ANN

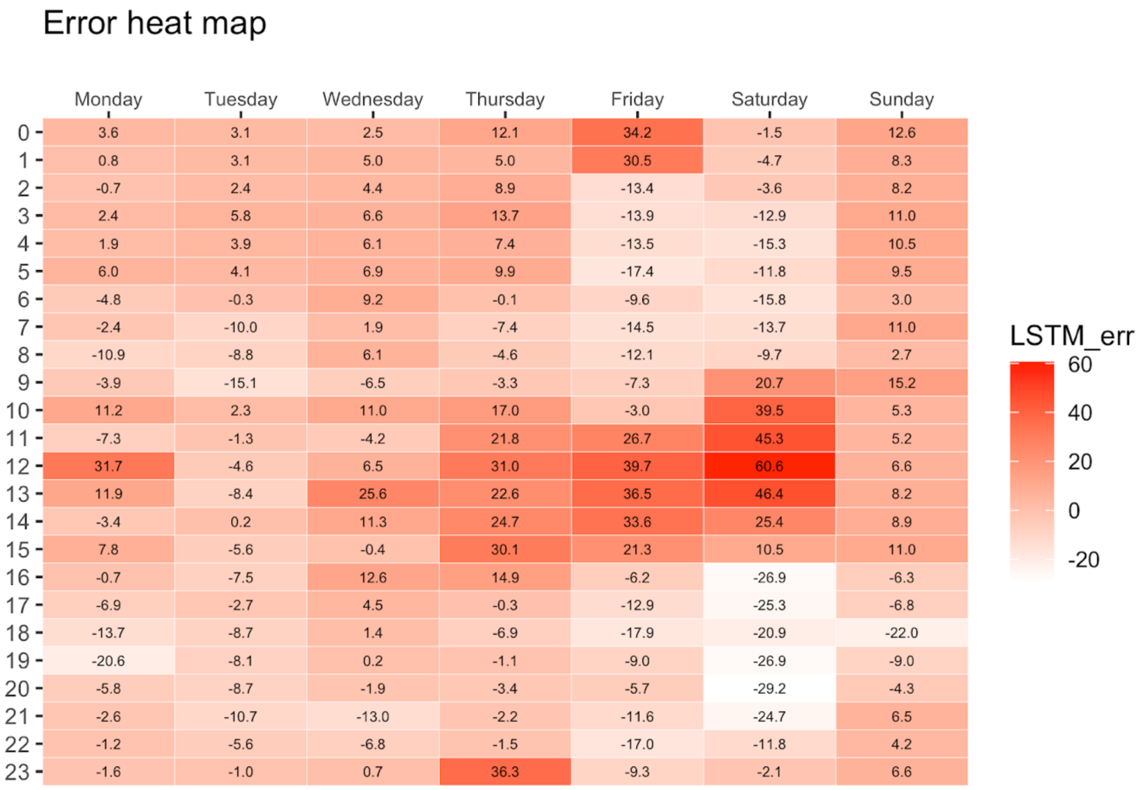

Figure A 10 Error Heat Map LSTM

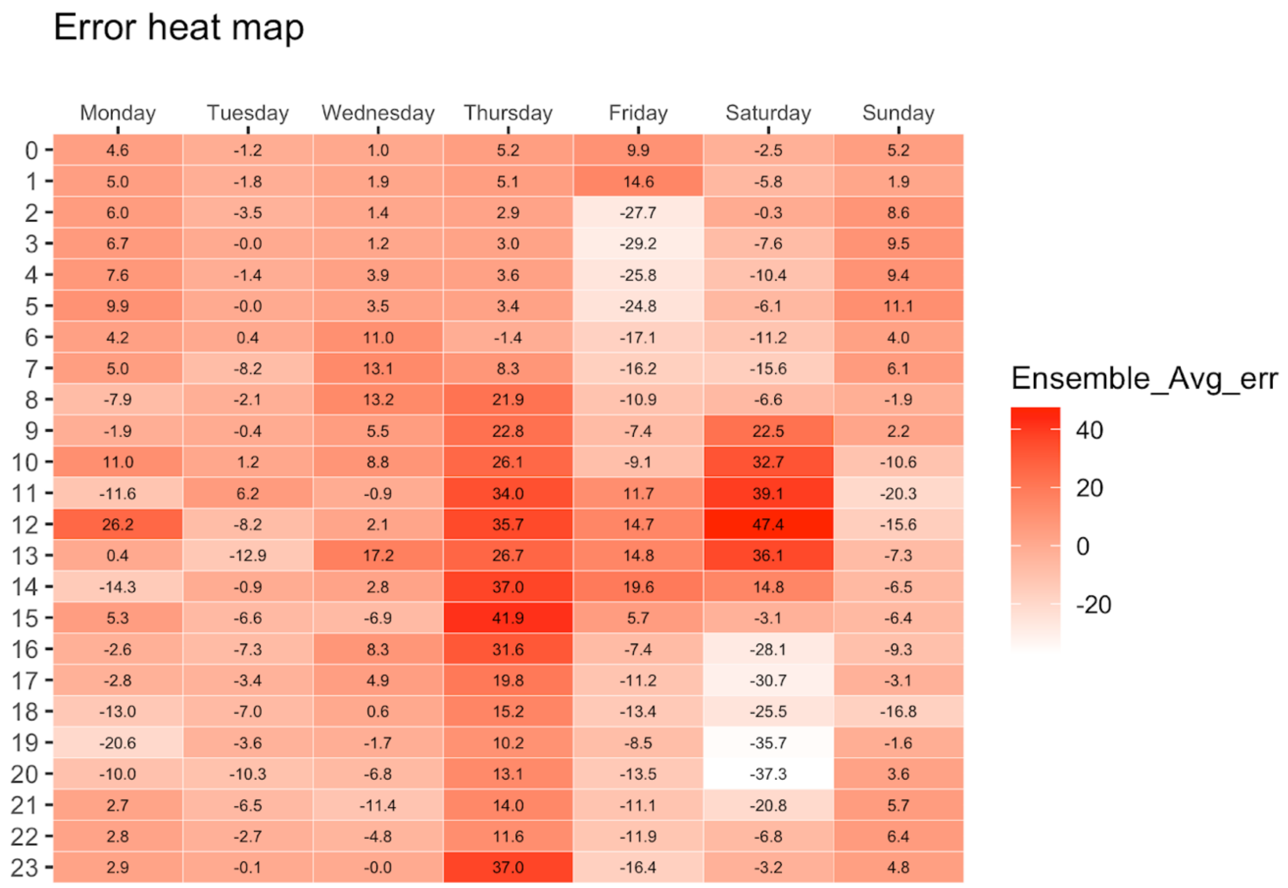

Figure A 11 Error Heat Map Ensemble Average

A5 Point Forecast for Benchmark and Proposed models

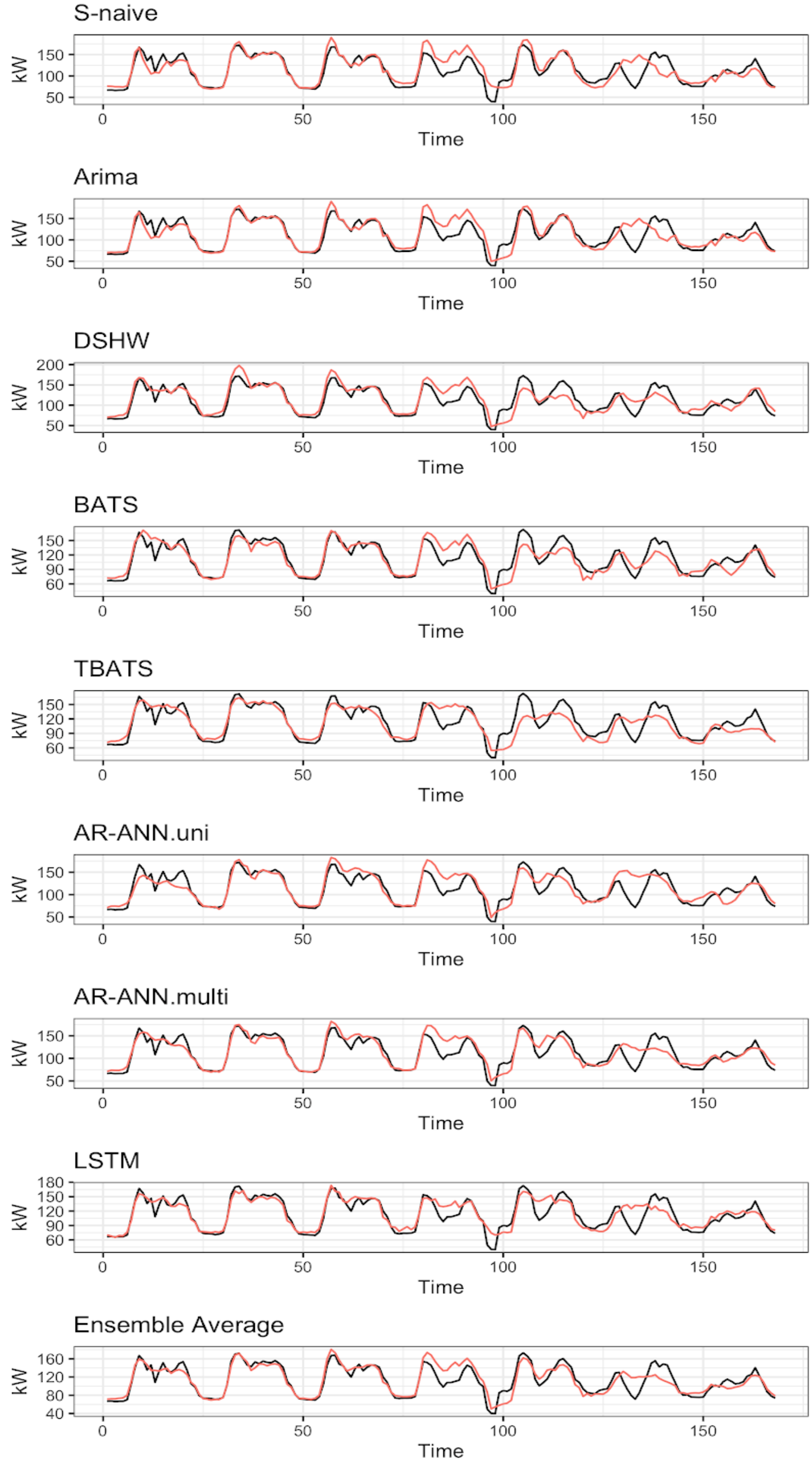

Figure A 12 Point Forecast for All Models