# Short-term Spatiotemporal Load Forecasting for Norwegian Bidding Zones

Atle Følid

Master's thesis in Software Engineering

Department of Computer Science, Electrical

Engineering and Mathematical Sciences,

Western Norway University of Applied Sciences

Department of Informatics,

University of Bergen

Supervisor: Prof. Reza Arghandeh

June 2021

# Abstract

Short-term load forecasting is vital for electric utility companies. The objective of this thesis is the short-term load forecasting of the five bidding zones in the Norwegian electrical grid. This master thesis proposes a novel method of approaching short-term load forecasting problem called *Lagged SpatioTemporal Features Short-Term Load Forecasting* (LSTF STLF) using LSTM. LSTF STLF is based on a spatiotemporal feature selection approach.

The dependencies between the five of Norway's bidding zones in the Nord Pool power market are discovered using tools such as correlation and mutual information to find the best spatiotemporal features from all bidding zones to better perform electricity demand forecasting in each given zone. By applying the proposed spatiotemporal feature extraction approach, forecasting accuracy improved significantly for the five bidding zones on a 48-hour forecasting horizon.

# Acknowledgements

I would first like to thank my supervisor Prof. Reza Arghandeh and the Ci2Lab research group (Michele Gazzea, Sindre Aalhus, and Dr. Mojtaba Yousefi) for the suggestions and guidance in the time working on this thesis. I would also like to thank my classmates for making the year a great experience and of course my family for their support.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Electricity load forecasting is the prediction of power consumption. For electricity utility planning, load forecasting is the fundamental business problem[1]. It is also known as consumption prognosis and vital for both the transmission system operators and the power suppliers. These stakeholders have to plan the distribution and generation of electricity on a constant basis. Electricity is not easily stored. Therefore there must be a balance between generation and consumption of power to reduce the risk of undersupply or oversupply issues. In the case of under-supply of power, there will be power outages. In the case of oversupply, the excess energy has to be disposed of or sold at a lower price, both actions lead to financial loss [2].

Electric utilities run the power grid, known as the most complex man-made system on earth, to deliver electricity to more than six billion people around the globe. Electric utilities also play a critical role in Norway as the main energy deliveries. Electricity is the main source of all industries in Norway. The Norwegian power grid system is geographically divided into five bidding zones; NO1, NO2, NO3, NO4 and NO5. The power engross NordPool [3] provides historical power consumption data for these zones. These are hourly resolution consumption data, provided in the metric megawatt-hour(MWh), which will be the basis of the data for this thesis. Alongside all the weather variables collected from the respective zones.

This master thesis will experiment on the use case of short-term forecasting of electricity consumption for the five bidding zones in Norway. The main objective is to explore the spatiotemporal relation of data between these zones and discover if the data from the zones can be utilized for improving the short-term load forecasts. *Spatiotemporal* means data that is collected across both space and time. The naming comes from the spatial and temporal domains. The spatiotemporal weather variables are also considered in the case study. To compare the impact of these features, a baseline vanilla LSTM will be implemented. The models are only as good as their data input, so the feature selection for the model is crucial in load forecasting. By applying the best possible input to the model makes it more likely to learn the data, understand the underlying trends and forecast more accurately.

The data-driven spatiotemporal feature selection methodology applied in this thesis is tailored for the case study. By first performing data analysis of the spatiotemporal data collected for the bidding zones, the feature selection will be narrowed down based on the results. The processing of the features will reveal if applying lag to the spatiotemporal features is beneficial. The step-by-step approach is validated through testing the vanilla LSTM implementation with the features as input.

## 1.1 Motivation

Statnett is the main transmission operator for the Norwegian electrical grid. One of their main responsibilities is to balance the power consumption with the power production [4]. To accomplish this, they rely heavily on good load forecasts.

More accurate forecasts lead to less excess electricity being produced. In Norway, the majority of our power generation comes from green sources such as hydropower, wind and solar power. In other countries and especially the underdeveloped countries, power generation is relying on coal and other polluting sources. If the load forecasts are improved, especially the peak load forecasts, this could lead to less strain on the environment.

For the generation companies who provide the electricity, the more accurate forecasts for their region, the better they can plan their generation. In Norway, they send in a production scheme that contains what they are willing to produce at a given price. The generation companies with better forecasts can get an advantage in the market.

Dr. Hong estimates that a improvement of 1 % to short-term load forecasting can save 300 000 USD for a year with 1-gigawatt peaks [5]. In 2020, the peak load for Norway was 21.86 GWh. Adapted to the Norwegian use case means approximately 6.5 million USD saved a year by improving the short-term load forecast by 1 percentage.

## 1.2 Thesis contribution and Research Questions

The knowledge gap this thesis aims to fill, is the exploration and utilization of the spatiotemporal relations in a large case study like the Norwegian bidding zone use case. This thesis will implement a baseline vanilla neural network and perform an extensive data-driven feature selection on the available spatiotemporal data. This data is both the historical electricity consumption and the weather variables collected. The objective is to select these features based on spatiotemporal data analysis and process them for the five Norwegian bidding zones and discover a novel method for approaching short-term load forecasting. This use case is also scalable to other countries, especially the Nordic countries. Provided that the areas have historical spatiotemporal data for electricity consumption data. The following research questions represent the red line through this master thesis:

**RQ 1** How are the spatiotemporal relationships for the data five bidding zones of electricity demand in Norway?

**RQ 2** How to utilize spatiotemporal dependencies in short-term load forecasting methods?

## 1.3 Thesis Structure

The outline of the thesis where each chapter is shortly introduced.

**Chapter 1** presents the introduction and motivation for this thesis, and lastly the research questions.

**Chapter 2** provides a theoretical background for the relevant subjects of the thesis. The literature review of load forecasting as a topic and popular methods in load forecasting for this thesis will be presented.

**Chapter 3** presents the the use case for the thesis.

**Chapter 4** presents the research method and methodology.

**Chapter 5** describes the results and discussion for the experiments of the thesis.

**Chapter 6** concludes the thesis and present ideas for further work.

# Chapter 2

# Background

In this chapter, an introduction to the relevant subjects in this thesis is provided. First, an overview of the load forecasting topic and the literature review conducted for this is presented. Then the popular methods and models for load forecasting are described.

## 2.1 An Overview of Electricity Load forecasting

Load forecasting is the prediction of electricity power to be consumed. A good comprehensive introduction of the basics of load forecasting is provided in the white paper on load forecasting [1]. The white paper has great pinpoints on how to approach the load forecasting problem. Many methods, methodologies and proven algorithms are presented. The authors provide case studies, which give the reader insights into how a load forecasting case could be solved. The terminology for load forecasting is set along with the general history of load forecasting and the advancements made. Furthermore, they explain what pitfalls to avoid when approaching the load forecasting problem.

In figure 2.1.1, the terms for electric load forecast horizons are shown in a rough overview. These terms vary in their temporal definition from paper to paper, and there is no commonly accepted understanding for the duration of the electric load forecasting horizon. The only repeating information is that the cut-off between short-term and medium-term forecasts is set at two weeks. We can observe that there are four categories that belongs to the short-term horizon [6]. Energy purchasing is done on the market of Nord pool for Norway, where generation companies and electricity vendors buy and sell electricity. The reason for the energy trading inclusion in the medium- and long-term horizons, is because the stakeholders can buy and sell futures to guarantee a spot price in the future. This is done as a risk management and hedge against fluctuating prices. Transmission and distribution planning is mainly done by the transmission system operator(TSO). Demand side management is done by the TSO and the generation companies. The short-term forecasts are vital for this horizon. Good forecasts enables these stakeholders to plan the the electricity generation for the coming electricity consumption. The operations and maintenance have to be planned on a short-term aspect. For instance, the parts of a hydropower plant wear and loses effect over time. Due to fluctuating spot prices and electricity demand, the generation companies have to plan these operations in a short time frame to minimize financial loss. The maintenance work means down-periods for the electricity generation.

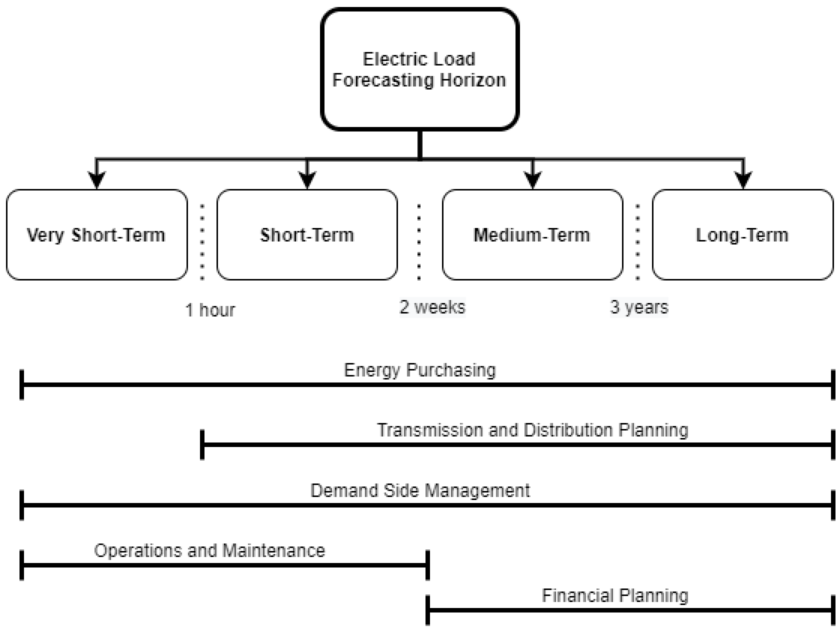



**Figure 2.1.1:** An overview of the terms used for forecasting horizons. Inspired by: [7].

To understand the Nord Pool market and the different powers that control it, a master thesis [8] gives a good overview of the different factors and stakeholders in the power market. Statnett is the main transmission system operator in Norway. It is funded by the government and they operate the grid and the electricity transmission flow. Statnett have an own data science team. An article from their data science team [4] tells us about their approach to creating and deploying short-term load forecasting models for live applications. They also disclose their results. Their LSTM model yields a Mean Average Percentage Error(MAPE) of 4.2 % for 48 hours forecast horizon for the 5 individual Norwegian bidding zones. Their result provides this thesis a benchmark and comparison basis for my methodology and results. This is the most comparable result found in the literature review of this topic and case study of Norway. Statnett's data science team made two different models, one Ridge regression model and one LSTM recurrent neural network model. The LSTM model performed the best out of these. However, these results will not be directly comparable of the results for the experiments in this thesis. The reason for this is that the Statnett team used the electricity consumption data for 2018 in their results, while this thesis will consider data from 2020 and 2021.

In 1978, the U.S Congress passed a bill to deregulate the power market. This was done to promote greater use of renewable energy due to environmental concerns, high inflation and increased fuel prices. An overview of factors that affects load forecasting is presented in a research paper [9], which gives the reader insights into how important load forecasting is in a deregulated economy. Since the Norwegian energy market was one of the first in the world to deregulate, it is interesting to delve deeper into this evolution of the market. The paper also provides introductions to the different methods and algorithms which are widely used and approved by the scientific community for the different forecast horizons. The authors study which factors affect energy consumption, such as the weather variables temperature, humidity and wind. These spatiotempporal weather variables will be collected and tested in this thesis as well.

The wind variables are tough to incorporate in a load forecasting use case, while the temperature variables are deemed important and easier to forecast on due to the daily pattern cycle. The demographic factors are mentioned in the paper as important features for medium and long-term load forecasts. Since this master thesis will only consider the short-term forecast horizon, the demographic factors will not be considered as features of the models. However, the similarities between the bidding zones in demographics are relevant in understanding the spatiotemporal relation for the bidding zones.

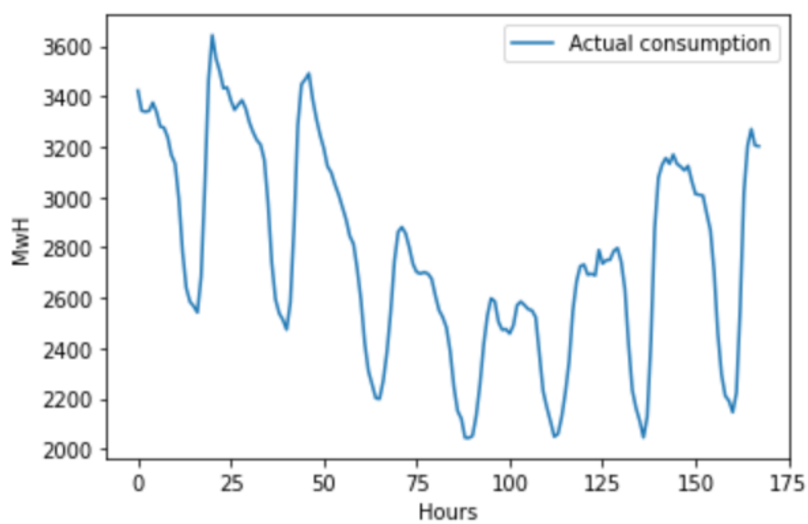



**Figure 2.1.2:** A random sample of 7 days of electricity consumption in NO1 region

The authors have acknowledged that the load profile is different for weekdays and weekends in a paper about using neural networks [10] for short-term load forecasting. In the figure 2.1.2, we can see that during a 7 day period of a random NO1 zone consumption load profile, there are approximately 72 hours with lower mean values. This indicates a weekend. The authors trained one model for weekdays, and one for weekends. In their case study they achieved good results even without applying weather parameters.

## 2.2 Popular Methods for Load Forecasting

This section presents the literature review and past work for popular methods for load forecasting. The two load forecasting methods chosen for this thesis' experiments are presented in their own subsections.

There are four main methodologies for load forecasting presented in the white paper for load forecasting [1]. *Variable selection* or *feature selection* is a methodology for determining which variables are useful and improve the load forecasts. *Similar day* approach was one of the first methodologies in load forecasting. The similar day approach is looking at the consumption of similar days historically, with weather factors being the main contributor in finding these days. *Weather station selection* is a methodology for load forecasting, due to the impact weather has on the electricity consumption. *Hierarchical forecasting* is a more novel methodology, because of the new smart electrical grid and the way this measures the energy consumption. More live data can be accessed, and aggregated consumption data on smaller levels such as households or city-level can be utilized.

The white paper also mentions three main methods for short-term load forecasting. These are Multiple linear regression(MLR), Auto-Regressive Integrated Moving Average(ARIMA), and artificial neural network(ANN). The MLR is described as easy to implement and achieves good accuracy, but needs explanatory variables and at least two years of history. The ARIMA achieves good accuracy on short term forecasts and it functions well with less historical data and few variables. The ANN achieves good accuracy during normal daily patterns for the load, and minimum domain knowledge of load forecasting is required to implement. The downside for ANN is that it needs heavy computing power and it is difficult to interpret and understand the results.

A comparative research paper [11] provides a case study of methods for load forecasting in Turkey. The paper concludes that artificial neural networks(ANN) and Least-squares support-vector machines(LS-SVM) perform better than the multiple linear regression(MLR) model.

For one hour ahead forecasts [12], the authors compare the methods of multilayer perceptron(MLP), ANN, and a support-vector machine(SVM). The ANN method achieves the lowest MAPE score, but they conclude that the SVM is preferred since it exhibits repeatability to always find the global minimum. However, the models were unable to accurately perform load forecasts if there were any erratic load patterns or missing data. This can be mitigated through outlier detection and missing data imputation, which this thesis's experiments will have as a pre-processing step for the data. In a case study for 24-hour load forecasts on weekends with an ANN, the results were accomplished [13].

The Long Short-Term Memory(LSTM) model is an recurrent neural network(RNN) architecture. This allows the model to recognize and predict sequences data. By applying the LSTM model with long sequences [14], the authors achieve good results in their use case. A comparative study was performed in a use case of Estonia [15]. The results further demonstrate the LSTM's ability to provide accurate short-term load forecasts. It surpasses the support-vector machines(SVM) model in this study.

The similar day approach is used with an XGBoost model in a case study [16]. The approach is to classify the main influencing factors of electricity consumption so that a feature map is constructed to select a similar day. The results improve, showing promise as a modern take on the similar day methodology. To model in weather parameters, a paper [17] proposes using Fuzzy modeling to incorporate these parameters.

The auto-regression methods are widely used in load forecasting. In a comparison with the SVM, the Auto-Regressive Integrated Moving Average(ARIMA) model scores slightly worse [18]. However, the ARIMA seems to detect the trends better. So for repeatability and the robustness of the model, the ARIMA can be considered preferable. In a case study of the Karnataka State electrical load, the author applies the ARIMA model for one-hour forecasts throughout a month [19]. The results show favorable short-term load forecasting accuracy.

A method for improving the features and achieving better forecasts is by applying lag to the variables. Either for imputing variables that are not up to date, or for improving the feature input to the models. Lagging the temperature variable and feature extract a average temperature are considered the two most popular approaches. Lagging the weather variables such as the temperature have documented an effect in day-ahead load forecasts [20]. To help data scientists in discovering how many time-steps a variable should be lagged, two analysis tools called autocorrelation function(ACF) and partial autocorrelation function (PACF) can be applied. A paper on short-term load forecasting [21] using regression analysis presents these tools as efficient. Lagged values will be tested and applied in the feature processing in this thesis' experiment. Lagging the spatiotemporal electricity consumption in a large case study like the Norwegian bidding zones has not been found in the literature review. Therefore, it will be a new approach when considering a large electrical grid.

Exploring the spatiotemporal dependency between different zones has been tried in case studies earlier. In a research paper [22], the results justify using this approach in the prepossessing of the data. Applying spatiotemporal analysis improves the load forecast accuracy in a paper where the use case has smart grid data [23]. A dynamic Spatio-temporal(DST) algorithm [24] was implemented in a paper, and the authors experienced a great performance increase when applying this algorithm to their use case. It outperformed a vector autoregressive (VAR) model.

The case study for this thesis is large in the sense of electricity volume and the spatial distances of the bidding zones in Norway. The analysis and experiments in the case study will explore the spatiotemporal approach. As this literature review has been conducted, many methods and algorithms have been unveiled. As there is no possibility to apply them all, it has been narrowed down to using the LSTM as the main model. The LSTM model has shown great results for several different short-term load forecasting use cases. The ARIMA models will be implemented as a basis for comparison. Both are presented in further depth in the following sections.

### 2.2.1 Long Short-Term Memory

The most popular method when applying artificial neural networks in load forecasting problems is by using a recurrent neural network(RNN). The LSTM is the most popular and widely used model of the RNNs, and in the following subsection, it will be explained. This model is the base model for this thesis's experiments. There are several reasons for this choice. The LSTM achieves great results on various load forecasting use cases found in the literature review. Statnett's data science team applied it with success in the same use case for the Norwegian bidding zones. In the initial experiments with the data for electricity consumption in Norway, it outperforms the other models tried without any tuning of the model.

The LSTM white paper was released in 1997 and has set record-breaking results in various applications [25]. It is widely used in regression tasks such as time series and load forecasting due to its feedback connections, which differ from the standard feedforward neural networks such as convolutional neural networks. The LSTM network also performs well on classification applications such as image processing.

First, the LSTM unit will be explained. Then the overall architecture of the autoencoder-LSTM network applied in the thesis' experiment is presented. An LSTM unit has a cell, an input gate, an output gate and a forget gate [26]. These gates determine what information in a unit to update, forget and output. The figure 2.2.1 visualizes the flow of information through an LSTM unit. The input to a cell is the $h_{t-1}$, the hidden state from the previous time step, and the $x_t$ which is the new information.

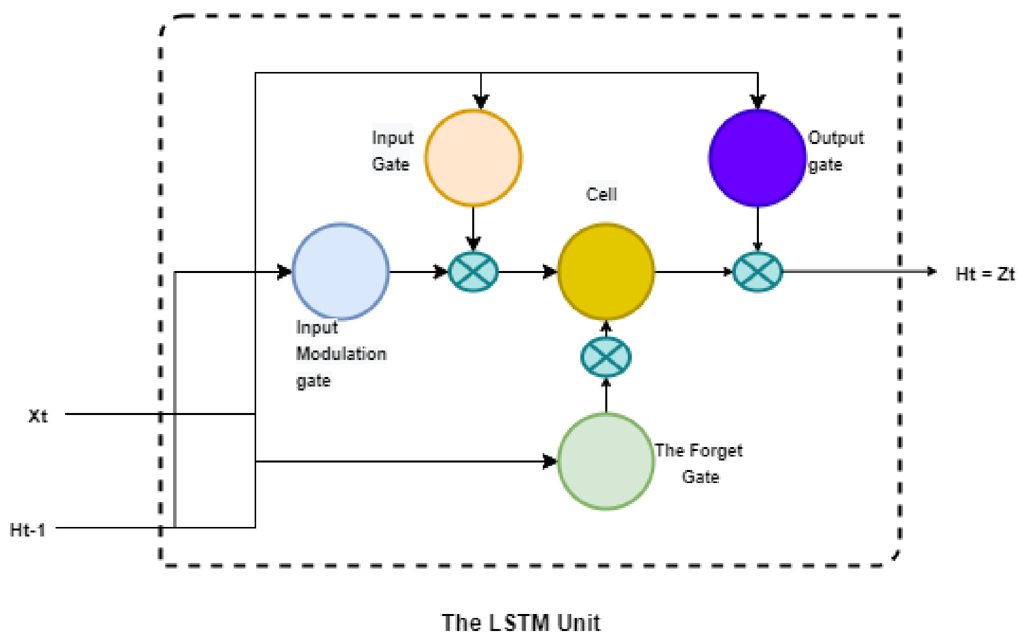



**Figure 2.2.1:** Overview of the LSTM unit. Inspired by fig. 3: [26]

Cell state is that the recurrent information flow from previous time-steps can be stored. The cell has memory to remember information, while the three gates control the flow of information to and from the cell. It does not store every bit of information, since the forget gate forces it to lose information. This is done by the forget gate's sigmoid activation function, which multiplies 0 to a position in the cell states matrix if it is to be deleted, or 1 if it is supposed to be remembered.

The input gate and the input modulation gate have a shared name called the save vector and determines what information should be allowed to flow to the cell state. There are two activation functions. For the input gate, it is a sigmoid function within a range from 0 to 1. This activation function will only add memory and not forget information. However, the input modulation gate applies a tanh activation function which ranges from -1 to 1. This activation function provides the possibility for the cell state to remove information.

The output gate decides what information should be allowed to flow through to the next cell with the sigmoid activation function. $H_t$ in figure 2.2.1 is the hidden state output from the LSTM unit.

When choosing which LSTM architecture to implement, an Autoencoder-LSTM(AE-LSTM) showed promise. An AE-LSTM architecture for load forecasting use case is applied in a paper[27]. They achieved improved results from the standard LSTM implementation. The standard network is a regular LSTM architecture. The AE-LSTM network has an encoder-decoder part and an repeat vector which differs from the regular LSTM. A paper with the

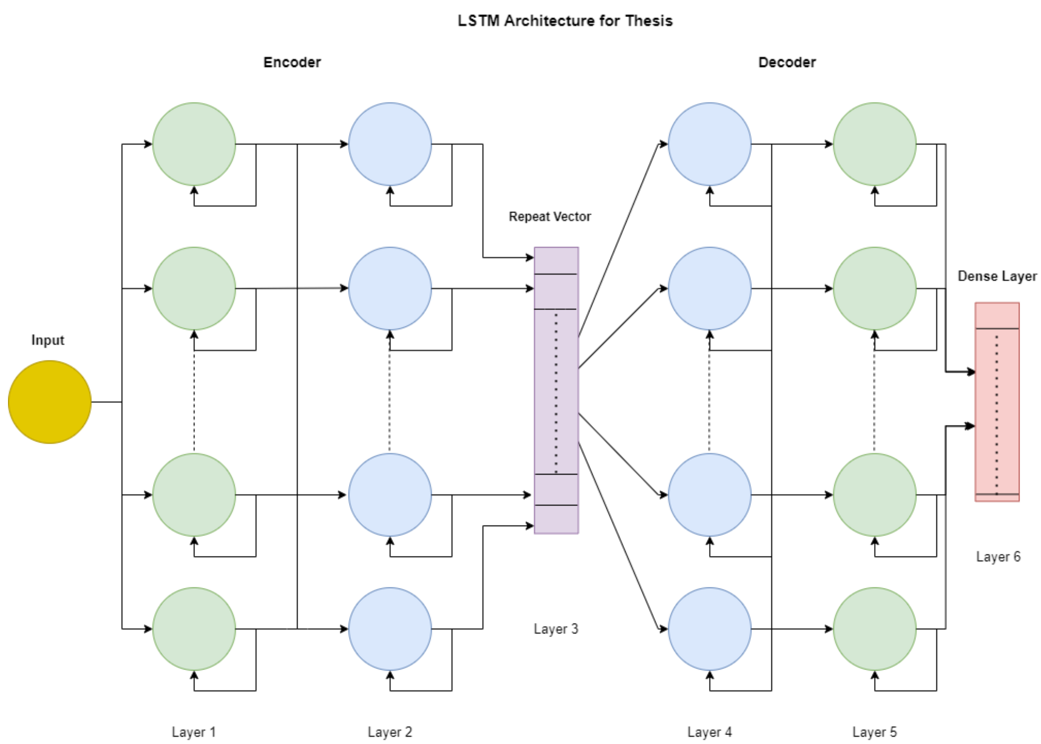



**Figure 2.2.2:** The Autoencoder-LSTM network architecture applied in this thesis.

AE-LSTM used for forecasting solar power production for the next 24-hours showed that the AE-LSTM outperformed the regular LSTM [26]. They used a wide range of weather variables, and incorporated these in their model through a data-driven feature selection method. The AE-LSTM showed that it was more equipped with for handling complex weather conditions.

In figure 2.2.2, the schematics of the architecture of the AE-LSTM network applied in this thesis is presented. The autoencoder consists of an encoder and decoder. There is input to the network, which can be univariate or multivariate. This is sent through the hidden layers where the LSTM units work out what information to store in the cell states and what to send further through the model. Between layer 2 and layer 4 there is a repeat vector layer. This acts as a bridge between the encoder and decoder sections of the network. The repeat vector duplicates the values for the decoder part of the network. Layer 4 and 5 are the mirrored layers of the layer 1 and 2. These equal the encoder part, with the dense layer as the last layer. The prediction is made after this.

### 2.2.2 ARIMA

This section will cover the basics of the ARIMA model, which will be implemented to have validation results to compare the final method's performance against.

The Autoregressive moving average(ARMA) models are a widely used for load forecasting use cases [28]. They are especially popular in time series forecasting problems. It comes in many variances such as the seasonal auto regressive integrated moving average exogenous (SARIMAX) model and the previously mentioned ARIMA model [18]. SARIMAX takes multivariate input, while ARIMA only takes univariate data. The AR-family are popular forecasting models due to their simplicity and the ability to generalize for non-stationary series. In the case of seasonal data and medium- to long-term forecast horizons, the use of SARIMAX can be applied. However, the ARIMA model is implemented in this thesis due to the good results it shows in short-term load forecasting use cases.

$$Y_t = \beta_1 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} \tag{2.1}$$

The Auto-Regression(AR) equation is displayed is shown in 2.1. An Auto Regressive model means that the variable input into the model depends on the past data values of itself. This highly depends on the fact that those previous values are correlated with the last data point. For the model to decide the weight of the past values for future predictions, the past values are checked for a correlation between the last data point and the past values. This is done by partial auto-correlation. Auto-correlation is an automatic detection for how many past values have a relatively high correlation with the current time-step.

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \cdots + \omega_q \varepsilon_{t-q} + \varepsilon_t \tag{2.2}$$

The Moving Average(MA) equation is shown in 2.2. The model analyzes the errors in the previous time steps. To be able to predict better for the current time step, the model incorporates these errors from earlier to perform better for later time-steps. The auto-correlation affects how much these errors should be weighted for later predictions.

As the AR and the MA have been covered for the ARIMA model, the Integrated(I) will be disclosed. If the time series to be predicted or forecasted on is not stationary, it is necessary to transform the data by differencing the data set [19]. Stationarity means that the time series data does not trend upwards or downwards over time, but is stationary with the approximately same values during a certain period of time.

To summarize, the ARIMA has three main calibrations:

- Auto-Regression(AR), which is the lag order. This is determined by finding the past values where the correlation for current observation are still significant.
- Integrated(I), which is the order of differencing. This is determined through analysis into the stationarity of the data.
- Moving Average(MA), which is the size of the moving average window. This is decided by checking how many past observation errors are significant.

# Chapter 3

# Use case

The use case for this thesis work is forecasting the electricity consumption for the Norwegian electricity bidding zones. There are five zones, with logical geographical separation. As can be seen in figure 3.0.1, the zones vary in size and shape. The NO1 and NO2 has generally the highest consumption, being the zones with the largest population. The five zones have their own bidding market, with vendors trading and providing electricity to the market.

There are transmission lines that go across the zones, and also from other countries such as Sweden, Denmark and Finland. If there is a surplus of production versus the consumption of electricity in Norway, there is an opportunity for export to other countries. The transmission operator Statnett governs the grid, maintaining the order of the electricity flows for the transmission lines. They also have the mandate to halt production for the power plants, to avoid surplus production they can not transfer anywhere. The lines

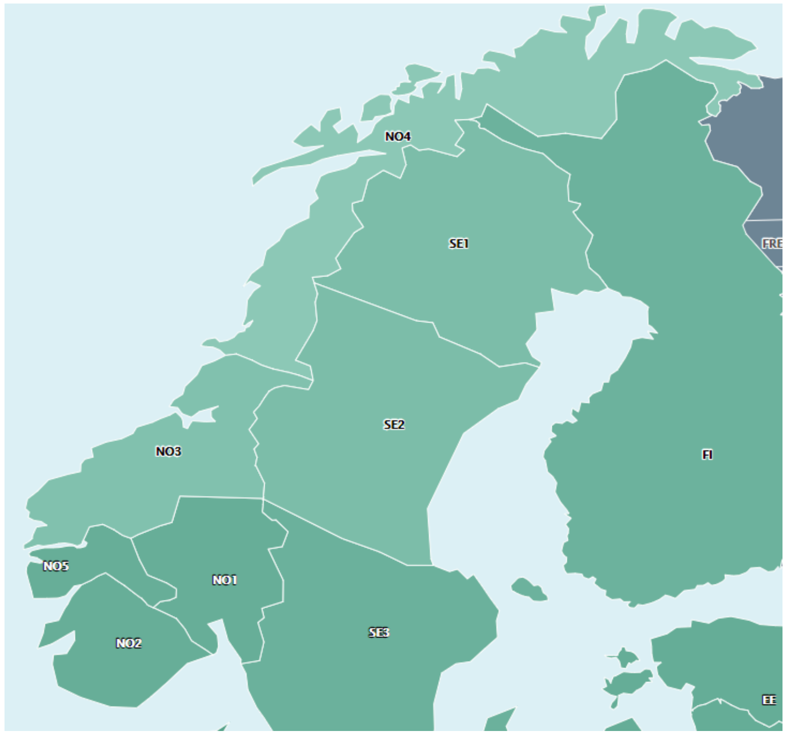



**Figure 3.0.1:** The NO bidding zones which the Transmission System Operator Statnett governs [29]

have limitations in form of the amount of electricity that can be transported at any given moment.

What affects the spot price for electricity is in large degree the production and consumption of electricity. Is there a surplus production in the market; for instance when there is lots of wind in southwestern Sweden, which has a huge windmill park that is not as easy to stop the production of as a hydropower plant. The market gets overwhelmed with the surplus, the prices drop, and the hydropower plant companies in Norway may halt their production to save their reservoirs of water to more profitable times. In the winter season the water reservoirs are low and the production is not at its peak. The consumption is also higher for the scandinavian countries. The generation companies want higher prices for using their reservoirs. There are many unregulated hydropower plants in Norway with no water reservoirs. They use the river as their fuel for production. These often run dry during the winter, while in the spring and summer they run full capacity because of the yearly snowmelt. This causes overflow of electricity production and it is often timed with warmer weather and lower electricity consumption; leading to lower spot prices.

The day-ahead market is the primary market for power trading in the Nordic region and is where the largest volumes are traded on NordPool[30]. Making the intra-day and day-ahead market the most important part for the power suppliers and vendors. For this thesis, the forecast horizon is set to 48 hour since this is deemed as the most important horizon for short-term load forecasting in the generation and consumption aspect.

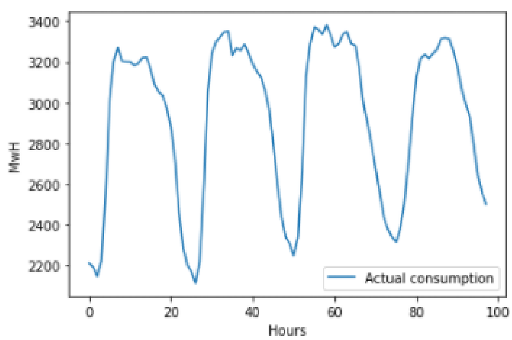
## 3.1 Data set description

The data with the hourly frequency of electricity consumption from the five zones are retrieved from Nord pool data [3]. The data is continually updated, and may be fetched through an API or downloaded manually. Analyzing one full year of data from 2020 gives insights into the variations between the zones. The table 3.1 holds the values for mean, minimum and maximum for electricity consumption for the 5 bidding zones in MWh and three of the weather variables from the NO5 zone. The values for the NO bidding zones are in megawatt-hour (MWh). The Temp column is the temperature in Celcius degrees, humidity is in percentage, and wind in meters per second. All these weather variables are collected from weather stations in the city of Bergen. As can be observed, the mean values have differences. The more populated areas of NO1 and NO2 use far more electricity than the less populated areas such as NO5 and NO4.

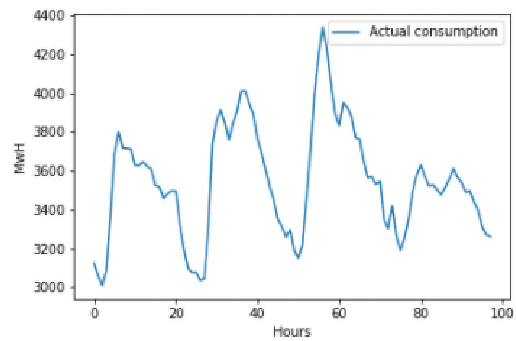
**Table 3.1:** The electricity consumption zonal data and weather variables insight for 2019.

| | NO1 | NO2 | NO3 | NO4 | NO5 | Temp | Humidity | Wind |
|---|---|---|---|---|---|---|---|---|
| Mean | 3948 | 4141 | 3023 | 2127 | 1874 | 8.5 | 75.2 | 5.3 |
| Min | 1812 | 2759 | 1751 | 1288 | 1128 | -5.2 | 18 | 0 |
| Max | 6846 | 5883 | 4262 | 3170 | 3218 | 32.1 | 97 | 15.8 |

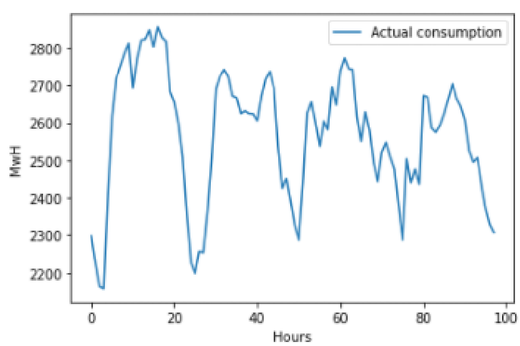
In figure 3.1.1, the load profiles for four of the bidding zones are shown. These are from the same period in June 2020 for 96 hours. As we can observe, the load profiles differ greatly. In figure 3.1.1 (a), the NO1 load profile is presented. The daily patterns are very similar, while the daily peak load for NO2 in figure 3.1.1 (b) fluctuates from 4350 MWh to 3600 MWh. This could lead to difficulties for the load forecasts with such varying daily peaks. The NO3 zone in figure 3.1.1 (c) has similar daily trends as the NO1 zone, while the NO4 zone presented in figure 3.1.1 (d) is more volatile as can be seen in the bottom right corner.



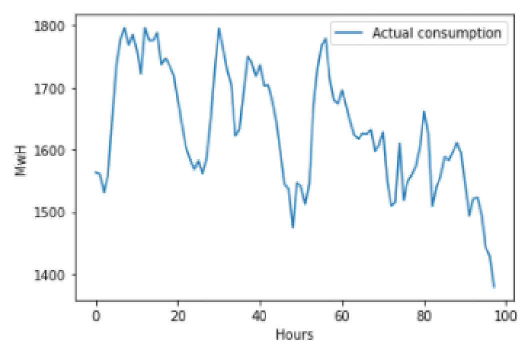

**(a)** NO1 Load




**(b)** NO2 Load




**(c)** NO3 Load

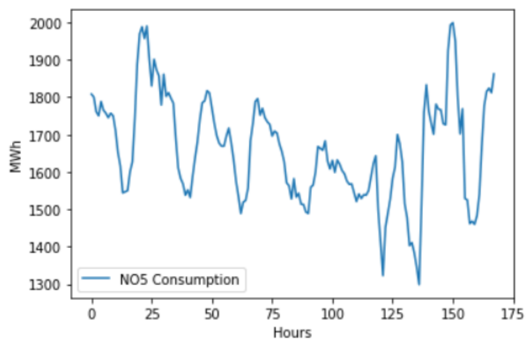



**(d)** NO4 Load

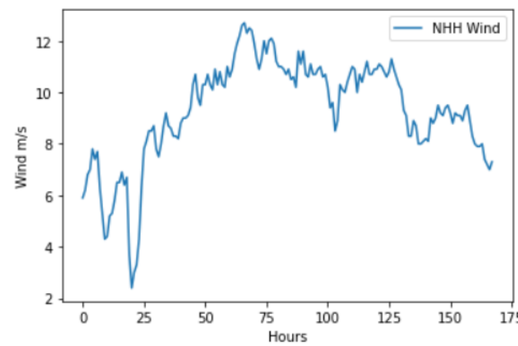
**Figure 3.1.1:** Load profile for 96 hours in June 2020 for 4 of the bidding zones

The other data for this use case are weather variables collected from the FROST API [31]. These are hourly values from weather stations throughout Norway. The individual variables were chosen and collected after investigating the data integrity from many weather stations in the five regions.

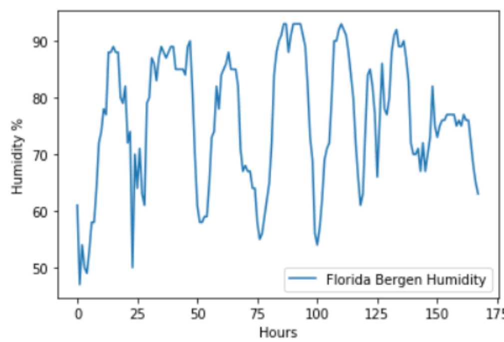
In figure 3.1.2 (a), the weather data is shown as a time series alongside the



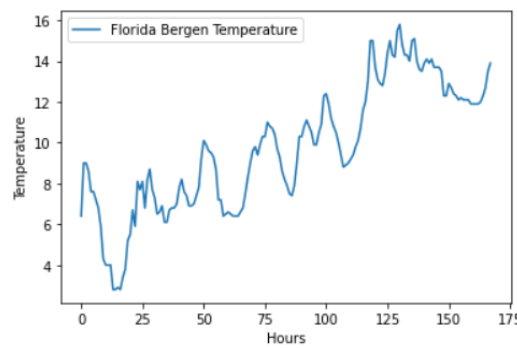

**(a)** NO5 load




**(b)** NHH wind




**(c)** Florida humidity




**(d)** Florida temperature

**Figure 3.1.2:** Consumption data, wind, humidity and temperature for 7 days in the NO5 zone.

consumption data for NO5. This is from the same 7 days, with weather variables from the NO5 zone. The wind data is from the Norges Handelshøgskule (NHH) weather station in Bergen, and as can be observed in figure 3.1.2 (b) was quite a windy week in Bergen. The humidity is shown in figure 3.1.2 (c) and fluctuates with daily peaks from 90 % to daily lows in the 50s percentage. Both the humidity and the temperature data are collected from the same weather station in Florida, Bergen. The temperature is shown in figure

3.1.2 (d) and it sees a constant daily rise during this week in June 2020.

In table 3.2 all the weather variables from Norway are displayed. They are categorized from in which NO zone the weather station data is located in. The weather variables vary from only 5 in the Northern Norway region NO4, to 13 in the Southern Norway NO2 zone. The reason for this is that some zones have many weather stations with good data integrity, while the NO4 zone has scarcity in both weather stations and quality of the data. The variables are wind, humidity and temperature, as these have shown positive effects for other load forecasting use cases [20].

**Table 3.2:** All weather variables collected through the FROST API [31]

| NO1 | NO2 | NO3 | NO4 | NO5 |
|---|---|---|---|---|
| BlindernTemp | SømskTemp | RisvollTemp | BodøTemp | FloridaHumid |
| HaugenTemp | VålandTemp | RisvollHumid | BodøHumid | NHHWind |
| IlsengTemp | OksøyHumid | MoldeTemp | TromsøTemp | FossmarkWind |
| IlsengHumid | VålandHumid | MoldeHumid | AltaTemp | FossmarkHumid |
| IlsengWind | TorungenTemp | MoldeWind | AltaWind | LundebotnWind |
| FredrikTemp | TorungenWind | ÅlesundTemp | | MjølfjellHumid |
| FredrikHumid | LandvikTemp | ÅlesundHumid | | FloridaTemp |
| | LandvikWind | Sverreumid | | VossevangenTemp |
| | LandvikHumid | SverreTemp | | OddaTemp |
| | SandnesTemp | | | |
| | LysebotnTemp | | | |
| | LysebotnWind | | | |
| | ÅlesundHumid | | | |

# Chapter 4

# Methodology

## 4.1 Design Science Research

Design science is the research method for the thesis and it is an information technology method of approaching research work. The research essay *Design science in Information Systems Research* [32] lists 7 guidelines for adopting the design science research method. Here are the 6 guidelines this thesis will incorporate:

- Guideline 1: Design as an Artifact
- Guideline 2: Problem relevance
- Guideline 3: Design evaluation
- Guideline 4: Research contributions

- Guideline 6: Design as a search process
- Guideline 7: Communication of research

For the first guideline, a viable artifact must be produced [32]. For my thesis, it is finding a novel method for approaching short-term load forecasting. An artifact can also be a model, but the models that are applied in this thesis' experiments are well-tested and already created. However, the tuning of the parameters and the network to fit the use case and data can be considered a smaller artifact.

For the second guideline, the research must seek to solve a relevant business problem. It has to provide an information technology-based solution to existing problems. Short-term load forecasting has a high degree of problem relevance, and a new method of approaching this is potentially very cost-saving [5]. The third guideline means to evaluate the artifact through well-executed evaluation methods. To adapt this to this thesis' work, the evaluation has to be a standard set in the research community and have multiple angles.

The fourth guideline reads research contributions. For this thesis, it means what the research and utilization of applying spatiotemporal features to a short-term load forecasting use case could provide for the research community. Furthermore, what the possible improvements could mean for the actual application of this for both transmission system operators, electricity generation companies and vendors.

Design as a search process is that the researcher should utilize all available means in the domain of the use case. For this thesis, it is gathering of all available and relevant data for the load forecasting use case of the Norwegian bidding zones. Furthermore, what load forecasting models to apply to the use case. The last guideline is the communication of the research work. The research should be presented in an effective way to both the domain experts and to the peers who are novel to the domain.

## 4.2 Short-term Spatiotemporal LSTM Forecasting

As for the methodology for this thesis, I decided on a data-driven spatiotemporal feature selection approach. A data-driven spatiotemporal approach is built on data analysis of the spatiotemporal data. The data-driven part is that further actions are determined based on the results of the data analysis and the validation results for the load forecasts. A paper using a data-driven feature selection method for solar power forecasts with an AE-LSTM model improved the results compared to other models [26]. The feature selection methodology is one of the main methodologies in short-term load forecasting [1].

The flow diagram in figure 4.2.1 shows the methodology pipeline for this case study. Firstly, the collected hourly resolution electricity consumption

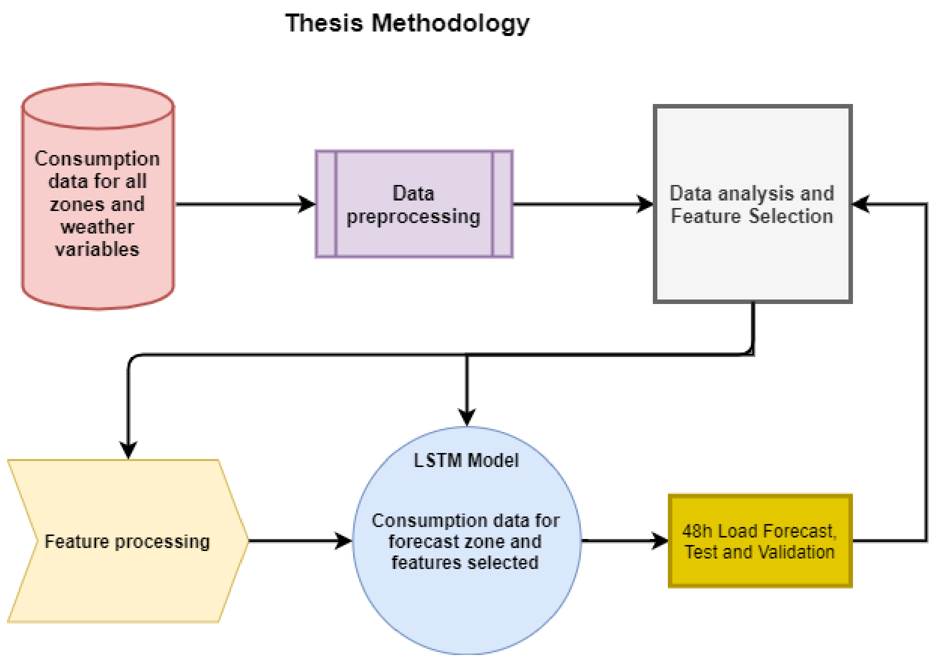



**Figure 4.2.1:** Overview of the Lagged Spatiotemporal Feature Short-Term Load Forecasting

data from all zones and weather variables are sent to the data pre-processing block. The integrity of these spatiotemporal data will be checked, through outlier and missing data detection. Then the data flow moves to the data analysis and feature selection block. With data analysis tools such as mutual information and correlation will reveal what features have the potential for improving the model.

The promising features are first sent to the LSTM model for test and validation. This is done to verify the feature selection based on the spatiotemporal data analysis performed. The feature processing block is meant to investigate if there are some tweaks or improvements that can be done for the spatiotemporal features. In this case it is testing the lagging of the features. The LSTM is chosen as the main model for the thesis work, so the inclusion of this in the methodology of the thesis is considered important. Then the LSTM network runs and produces a 48-hour load forecast to be tested and validated against the actual consumption. The workflow goes back to data analysis and feature selection until the results are deemed acceptable, and a method has been found. The following sections will unveil what actions happen in the different blocks.

### 4.2.1 Data Pre-processing Block

When applying a data set to a model, there has to be a thorough check to review the data integrity. The usual checks are outlier detection and missing values. Outlier detection is for detecting data points that are skewed and non-logical. For instance, a data point of 20 000 MWh for region NO2 which has a mean of 1874 MWh, or a 0 MWh without there being a power outage for the entire bidding zone. The second check is for missing data points. Most models will simply stop calculation and get errors if a Null or Not-a-Number(NaN) value is sent into the network.

The consumption data set is checked for outliers, but none were detected. There are also no missing data through the last 3 years of electricity consumption data sets. However, in the weather variables, there are several instances of missing data. These are mitigated using interpolation, where the adjacent value is applied to the missing data point.

Data formatting for this dataset, when sending it through an LSTM network, is using a scaler to normalize the values. We set the range to be between -1 and 1. This is a normal operation for data pre-processing because having features with different ranges can cause a neural network to weigh the features wrong. The normalized values are sent through the network and are converted back to the original domain space for testing and validation after the forecast.

### 4.2.2 Feature Selection and Processing Block

Feature selection in load forecasting is essential. The model performance is only as good as the input data. As a data-driven spatiotemporal methodology, there has to be a reasoning for the selection of what feature inputs to have in a model. The features to be selected from can be found in section 3.1. These are historical energy consumption data for the 5 bidding zones, and 44 weather variables.

The spatial distances between the zones are shown in figure 3.0.1. The zones are not all adjacent to each other. The temporal relation between the zones is that the data is collected across the same time-space with the same hourly resolution. The spatiotemporal relation is the data analysis of the two parts combined. The weather variables are also considered spatiotemporal. They are collected across time and space, with the same hourly resolution, and will be examined against their respective bidding zone. To determine if there is a spatiotemporal dependency for the bidding zone features and weather variables, we need to perform an analysis of the historical spatiotemporal data with tools such as correlation and mutual information.

Correlation checks reveal if and how variables are related. The range for the output of the check is from -1 to 1. If the value is above 0, it has a positive correlation. This means that the two variables move in the same direction. If x goes up, it is likely that y goes up at the same time. For negative correlation less than 0, this means through the time series when the value in variable x trends upwards, the variable y tends to trend downwards. They move in the opposite direction. The equation in 4.1 is the most widely applied correlation equation called Pearson correlation. Where $n$ is the number of observations, $(x_i - \bar{x})$ is the sum of scores for x, and $(y_i - \bar{y})$ is the sum of scores for y.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.1}$$

Mutual information reveals if there is some explanatory resemblance between the variables. It measures how much information can be acquired from one feature given another. When mutual information was applied as the main feature selection tool, the authors experienced improved results for their short-term load forecasting use case [33]. The mutual information can be equivalently expressed as displayed in 4.2 [34]. Where the marginal entropies are $H(X)$ and $H(Y)$, the conditional entropies are $H(X|Y)$ and $H(Y|X)$, and X and Y are represented as joint entropy by the $H(X, Y)$.

$$
\begin{aligned}
I(X;Y) \equiv H(X|Y) &\equiv H(X|Y) - H(Y|X) \\
&\equiv H(X) + H(Y) - H(X,Y) \\
&\equiv H(X,Y) - H(X|Y) - H(Y|X)
\end{aligned}
\tag{4.2}
$$

The results were also improved in the experiments for day-ahead pricing forecasts when the authors applied mutual information scores in their feature selection pipeline [35].

For the peak load forecasting problem, the authors in a research paper [36] concludes that using mutual information for feature selection could lead to improvements for peak load forecasting use cases. In another case study, the correlation and mutual information were combined as feature selection tools for weather input [37]. The authors improved their results on the short-term load forecasting horizon. This paper enforces the application of the data-driven spatiotemporal feature selection methodology disclosed earlier in this section.

In this thesis' experiments the correlation and mutual information checks will be applied to all features shown in 3.1, as part of the data analysis in feature selection. This will reveal which features have the potential to affect the short-term load forecast positively for this use case.

For the feature processing block, the features selected in the feature selection block shall be investigated first. A widely applied method of improving and achieving better forecasts is using lagged features. Either for imputing variables that are not up to date or by applying lag to the features applied to the model for forecasting. Lagging the weather variables such as temperature have documented an effect in past works [21]. Lagging the electricity consumption feature itself is applied as a feature input in a case study [38]. The authors achieved a 31.6 % Mean Absolute Percentage Error(MAPE) improvement when applying this method to their load forecasting use case. The spatiotemporal data in their study is the electricity consumption gathered from 1708 households.

Lagged values will be tested and applied for the features that show promise in the feature selection. The NO bidding zones electricity consumption features can be lagged with itself as a feature extraction process and applied as a feature to its own forecast. Or, the electricity consumption bidding zones can be applied as lagged spatiotemporal features in the forecast of another bidding zone.

### 4.2.3 LSTM Forecasting Block

Showing the best results both in the initial studies and in the general literature review, the LSTM was chosen as the model for my case study. The LSTM model is not an out-of-the-box solution to all use cases and data. It has to be tailored, tuned and optimized to perform well and adopt the trends in the data. To optimize the LSTM, the tuning and tweaking of the hyperparameters, training length and settings have to be performed.

To have a fair and academically correct comparison of the results, a baseline vanilla model has to be made. In an LSTM model, there are several parameters and settings that affect the performance:

- Training data length. How much of the data should be used for training.
- How many past values the LSTM unit should hold in memory.
- Number of epochs to train and validate the data on.
- The batch size. How many values to be sent through the network at the same time.
- The number of layers in the network.
- The optimizer and loss function choice.
- Drop out-filters to mitigate overfitting.

To approach this scientifically, there has to be some trade-offs. One solution is to grid-search with different values for all of the parameters and settings at the same time. This will create too many iterations. A grid search approach is a way of testing many combinations of different parameters at the same time, to find the best result. The other solution is to choose step-by-step which parameters to be tested against each other. For instance setting the grid-search with 1, 3, 6, 9, 12, and 24 months of training data against batch sizes of 8, 16, 32, 64, and 128. This creates 30 iterations for the model to execute.

With the testing of the parameters through a grid search approach, a baseline vanilla model has been established. This was performed using univariate electricity consumption data for the LSTM model. This yielded the following settings shown in table 4.1 .

**Table 4.1:** The tuned values.

| Parameter tuned | Tuned value setting |
|---|---|
| Training data length | 6 months |
| Past values | 48 hours/time steps |
| Epochs | 13 |
| Batch size | 16 |

The optimizer ADAM has shown that it performs well and is chosen as the optimizer for this model in the experiments [26]. The loss function and the number of layers were also predetermined before optimization. The drop-out filter only skewed the results, so this was discarded.

The LSTM network chosen for this thesis' experiments is an autoencoder sequence-to-sequence model. A part of the objective for this thesis is to achieve as good results as possible. However, the more important scientific contribution lies in the attempt of discovering a novel method and approach for short-term load forecasting in a large case study like the Norwegian bidding zones. Therefore, the experiments in this thesis are mainly focused on spatiotemporal data analysis, feature selection, and feature processing. The comparison of the results between the vanilla model and the model with added features is an integral part for the validation of the thesis work.

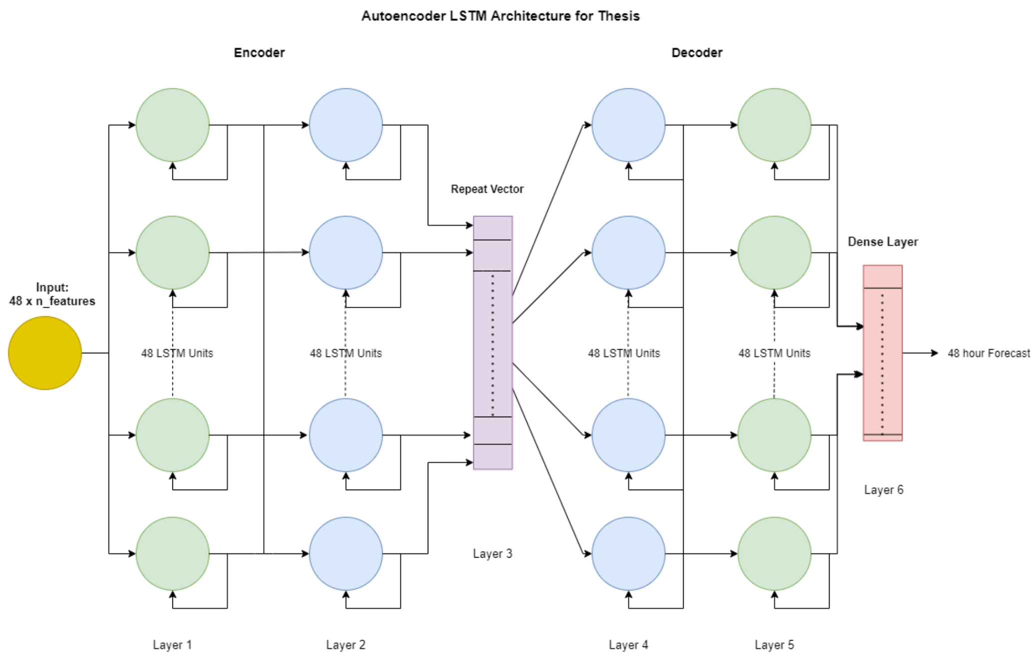



**Figure 4.2.2:** The LSTM architecture for the case study

In figure 4.2.2, the architecture of the final network is presented. The input to the model is 48 hours/data points x number of features. The number of units of the hidden layers is determined by how many time-steps are set. In this case, it is set to 48. The first layer has the return sequence parameter set to true, so this layer receives the input data and outputs 128 features with all 48 time-steps to the next layer [39]. The second layer receives the input and reduces the feature size to 64. This layer outputs to the repeat vector a vector with only 1 time-step because the return sequence is set to false. This output is called an encoded feature vector. The repeat vector duplicates the feature vector 48 times, once for every time step. This makes up the bridge between the encoder and decoder sections in the network. The decoder section reverses the order of the encoder section. The fourth layer has the same parameter settings and feature size as layer 3, while the fifth layer is the same as layer 2. The last layer in the network is the Time Distributed Dense layer. Layer 5 outputs a vector of the 48 time-steps x 128 features. The Time distributed layer duplicates the number of features from this vector equals the number of input features into the network. In the end the prediction of the model is performed, creating the 48 hour short-term load forecasts based on what the model has learned.

# Chapter 5

# Results and Discussion

For the results achieved in the use case of short-term load forecasting for the Norwegian bidding zones, the data explained in section 3.1 is applied to methodology shown in figure 4.2.1.
The results in this use case is executed on a Nvidia GeForce GTX 1060 6GB, a Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz and 16.0 GB ram.

## 5.1 Feature selection results

Here we disclose the feature selection results, where the correlation and mutual information scores are presented and discussed. They will point us in the direction of which of all the previously mentioned features should be added as an input to the LSTM model.

In regards to the spatiotemporal relationships between the bidding zones, the data analysis points to the fact that they are highly positively correlated. As can be seen in figure 5.1.1, with two years of spatiotemporal electricity consumption data from the full calendar years 2019 and 2020, zone NO2 scores the highest average correlation with the other zones. While NO4 has the lowest average score. The NO4 zone is the one with the most spatial distance from the other zones, and this could be the reason for the slightly lower correlation score. The spatiotemporal bidding zones correlation scores range from 0.88 to 0.98. The maximum positive score for correlation is 1. This tells us that if one of the zones trends upwards in consumption, most likely the others will as well. Furthermore, if one of the zones trends downwards, there is a strong case for the others zones will trend at approximately the same time.

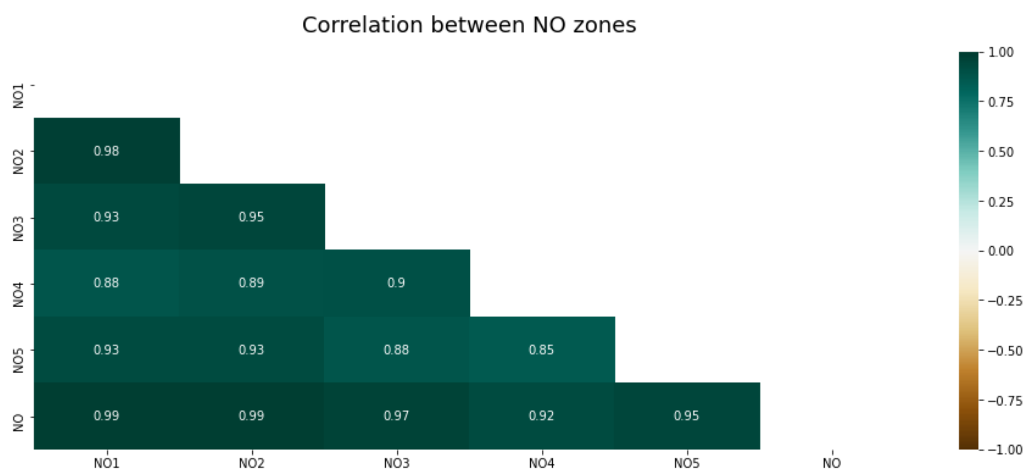



**Figure 5.1.1:** The correlation between the 5 bidding zones and the total consumption for Norway variable NO.

The mutual information scores can be observed in figure 5.1.2. The naming of the columns represents which NO zone is shown against which other NO zone. The first 4 columns represent the NO5 results in mutual information against the others, with the last number being the zone from which the check is performed. The takeaway is that NO1 and NO2 have the highest average scores and NO3, NO4 and NO5 have similar scores. The results from the mutual information analysis performed in the range of 4 to 6. A score of 6 tells us that information from one feature can explain a great deal from the other feature. This further strengthens the case that the zones have a spatiotemporal dependency.

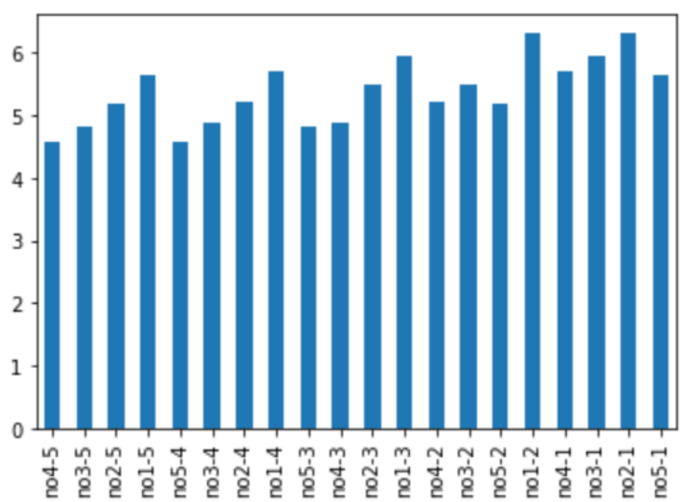



**Figure 5.1.2:** The Mutual Information scores between the 5 bidding zones.

The spatiotemporal relation is not equal for the zones concerning data analysis. It is not a given fact that the highest correlated or highest-scoring mutual information spatiotemporal zones perform the best in load forecasting use cases. When applied as a feature to the other zones, the results can either strengthen the case for the zones with the highest scores or debunk the differences. There could be other factors playing in.

| | NO1 |
|---|---|
| NO1 | 1.000000 |
| BlindernTemp | -0.816408 |
| BlindernWind | -0.190971 |
| BlindernHumidity | 0.198503 |
| HaugenstuaTemp | -0.789767 |
| IlsengTemp | -0.807515 |
| IlsengHumidity | 0.313041 |
| IlsengWind | -0.060585 |
| FredrikstadTemp | -0.826307 |
| FredrikstadHumidity | 0.236818 |
| FredrikstadWind | -0.066424 |

| | NO2 |
|---|---|
| NO2 | 1.000000 |
| SømskleivaTemp | -0.803169 |
| VålandTemp | -0.724228 |
| OksøyHumid | 0.041626 |
| VålandHumidity | -0.015817 |
| TorungenTemp | -0.813373 |
| TorungenWind | 0.096498 |
| TorungenHumid | 0.054785 |
| LandvikTemp | -0.745850 |
| LandvikWind | -0.072333 |
| LandvikHumidity | 0.079887 |
| SandnesTemp | -0.718712 |
| LysebotnTemp | -0.745638 |
| LysebotnHumidity | -0.187307 |
| LysebotnWind | 0.146905 |

| | NO3 |
|---|---|
| NO3 | 1.000000 |
| RisvollanTemp | -0.790667 |
| RisvollanHumid | 0.076550 |
| MoldeTemp | -0.719414 |
| MoldeHumid | -0.011115 |
| MoldeWind | -0.162502 |
| ÅlesundTemp | -0.755523 |
| ÅlesundHumid | -0.135433 |
| SverreborgHumid | 0.076550 |
| SverreborgTemp | -0.790667 |

| | NO4 |
|---|---|
| NO4 | 1.000000 |
| dfBodøTemp | -0.833860 |
| dfBodøHumid | -0.192555 |
| dfTromsøTemp | -0.822757 |
| dfAltaTemp | -0.844310 |
| dfAltaWind | 0.317668 |

| | NO5 |
|---|---|
| NO5 | 1.000000 |
| dfflorida_humidity | 0.023270 |
| dfNHWind | -0.275886 |
| dfFossmarkWind | 0.073145 |
| dfFossmarkHumidity | 0.034110 |
| dfLundebotnWind | -0.256850 |
| dfMjølfjellHumidity | 0.028847 |
| dfFlorida_temp | -0.745745 |
| dfVossevangen_temp | -0.738421 |
| dfOddaTemp | -0.757066 |

**Figure 5.1.3:** The correlation scores between the NO zones and the weather variables.

The weather variables have been checked for correlation and mutual information as well over two full calendar years 2019 and 2020. In figure 5.1.3, the correlation between the NO zones and the weather variables in their respective zones is displayed. All the temperature variables vary in the range from -0.7 to -0.84. This inclines a strong negative correlation. If the temperature decreases, the most likely will outcome for the electricity consumption is an increase. There seems to be no consistency in the results of the wind variables, with varying scores from 0.31 to -0.27. The humidity scores show that there is not much explanation to be found for electricity consumption in the historical humidity data. Except for the NO1 zones, where the humidity correlation ranges from 0.19 to 0.31.

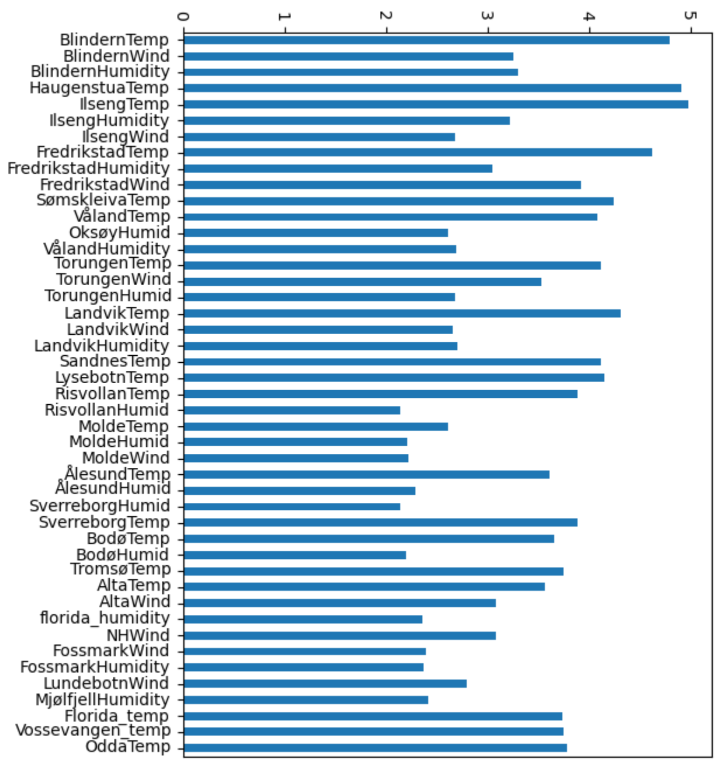



**Figure 5.1.4:** The Mutual Information scores between the NO zones and their weather variables.

In figure 5.1.4, we observe the 44 weather variables checked for mutual information against the respective bidding zones where the weather station is located. For instance, the variable 'BlindernTemp' is the historical temperature data from the weather station Blindern in Oslo, which is in the NO1 zone. What can be discovered in the data, is that the temperature variables score the highest on mutual information. The humidity variables score the lowest, while the wind variables being slightly higher. The data analysis for the spatiotemporal weather variables reveals that the temperature variables are most likely the most applicable features to be included in the short-term load forecasts. All the temperature variables score high throughout the data analysis tests. There seems to be a spatiotemporal dependency.

After the spatiotemporal data analysis has been conducted, some features stand out. By following the data-driven feature selection methodology shown in figure 4.2.1, the next step is testing and validate these features in the LSTM model. We must uncover which features improve the model before testing out modifications to the features. First, we need a way of scoring the short-term load forecasts for a comparison basis.

The most common and standard way of scoring load forecast results is by using Mean Absolute Percentage Error(MAPE). Since Mean Absolute Error(MAE) could confuse as to what is an acceptable result, which would differ from the sole amount of load in the given use case. In this thesis, the mean values of electricity consumption have differences between the zones, meaning that the most reasonable way to evaluate the results is to use MAPE scoring. This yields a percentage average score based on the error in the load forecasts. The MAPE equation is shown in 5.1. For the load forecasting domain; a 48 hour forecast means 48 $T$ observations. The $y_t$ is the electricity consumption at time $t$. The $f_t$ is the forecast of the actual consumption $y_t$.

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T} 100 \left| \frac{y_t - f_t}{y_t} \right| \tag{5.1}$$

The MAPE value is a good indicator for the evaluation of the model's performance. The lower the percentage score is, the more accurately the forecast is for the horizon set. If the score is 0 %, the model predicts the exact value over the future unknown 48 data points in this experiment. A MAPE score of 5 % indicates a 5 % average error for the 48 data points.

The initial baseline results were performed for each of the five bidding zones. This was done with the vanilla LSTM model disclosed in section 4.2.3, using univariate data from the region to forecast on. For instance, the bidding zone NO1 baseline result is achieved by running the LSTM network with the electricity consumption from NO1. The output is a 48-hour short-term load forecast, to have as a comparison basis and validation of the feature selection. The same forecast time is applied to the feature selection testing and validation part. Let us move to multivariate data and feature selection. The best scores in correlation and mutual information in the data analysis part were the electricity consumption bidding zones. By first applying these NO features to the baseline model, 3 out of 5 zones showed better results than the baseline vanilla model. This was done through a grid search approach. For instance, the NO1 zone was tested individually with the NO2 zone, then NO3, and so forth. Adding all of the 4 other NO zones at once caused too much noise for the model, and the results worsened.

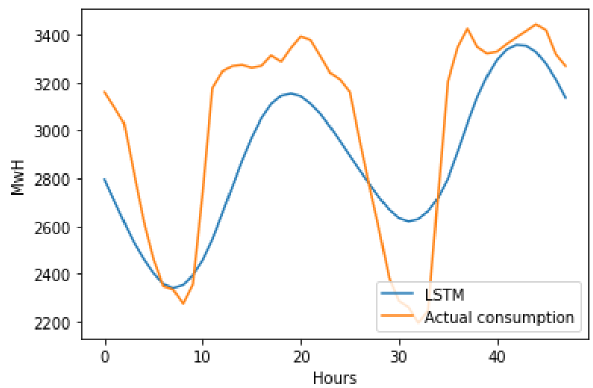



**Figure 5.1.5:** NO1 bidding zone forecast with all weather variables as input, MAPE score 8.25 %

The temperature variables showed the most promise of the weather variables in the spatiotemporal data analysis section. These were tested both individually and then incrementally if they showed promise. When applying all the weather variables for a region into the model, the results were poor, as can be seen in figure 5.1.5. This achieved a MAPE score of 8.25 % for the NO1 zone forecast. The blue line is the short-term load forecast and the orange line is the actual electricity consumption. When implementing the temperature variables into the model, none of them improved the baseline result. In figure 5.1.6, the NO1 forecast with the 'Haugenstua' temperature variable is shown. This variable had one of the highest scores in the data analysis and achieved a MAPE score of 3.85 %. However, it still did not improve from the univariate vanilla LSTM result for the NO1 region. This forecast shows that the model learns the daily trends better than the forecast in figure 5.1.5.

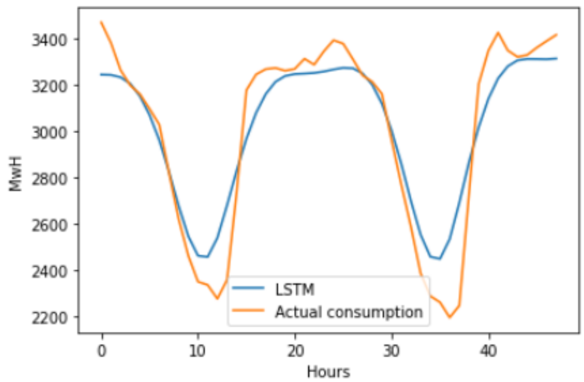



**Figure 5.1.6:** NO1 bidding zone forecast with best temperature feature 'HaugenstuaTemp', MAPE score 3.85 %

After checking all the temperature variables with their inherent zone, none of them improved the vanilla baseline result. To be thorough in the experiment, we checked all the 44 weather variables for their respective bidding zone. These are listed in table 3.2. Alas, only 2 out of the 44 weather variables improved the baseline model. When these two were run together in the model, the results worsened from the baseline result again.

The next step was the feature processing explained in section 4.2.2. In the literature review of this thesis, the majority of papers on lagged variables consider lagging the weather variables. By first lagging the temperature variables for the NO1 region with a range from 8 to -8, one feature improved the baseline model's MAPE score. This was with -4 lag, meaning that the time series was shifted 4 hours back in time. By having discovered that a temperature variable improved when applying -4 lag, the rest of the temperature variables for the other bidding zones were tested. Only one other temperature variable improved from the baseline model's result.

Recalling that the spatiotemporal bidding zones features had a high positive correlation and that they most likely trend in one direction at approximately the same time. The real discovery of the experiments in this thesis was made when lagging the spatiotemporal bidding zones. With a grid search approach, we applied it to the spatiotemporal NO electricity consumption features. Testing a lag range from 8 to -8 again for the bidding zones, to discover if shifting the added NO zone feature back or forth in time was the most effective for all of the zones. This would possibly tell us if a zone has to be lagged to function as a "trendsetter" for the rest of the NO zones. It was quickly discovered that either -2 or -4 was the optimal lag for the spatiotemporal features.
Forecasts for all zones were improved beyond all-time best MAPE score when applying a lagged spatiotemporal bidding zone feature to the forecast zone.

To find out which of NO zones are the best ones to apply as a lagged spatiotemporal feature, several tests were carried out. By running the forecasts for different periods in time in April, June, and September of 2020, it was clear to see which spatiotemporal features distinguished themselves from the others. In figure 5.1.7, we can see the spatiotemporal connectivity graph for what feature was applied to a NO zone forecast. The NO2 feature

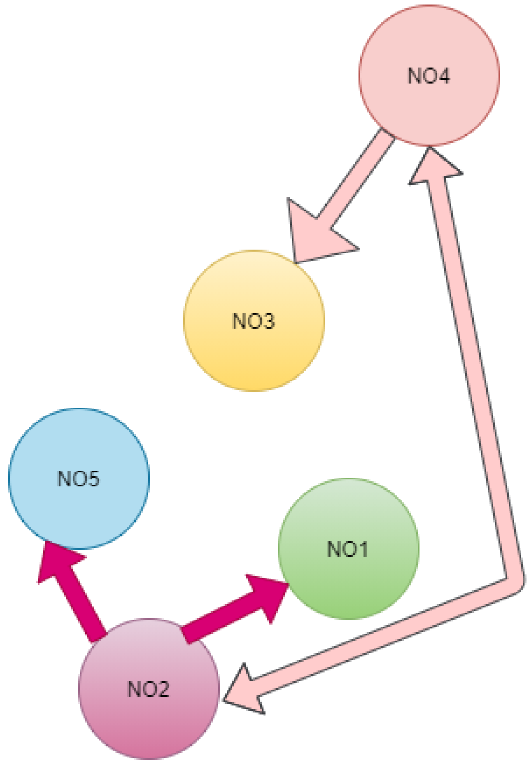



**Figure 5.1.7:** The spatiotemporal connectivity graph of the Norwegian bidding zones

repeatedly had the best MAPE score when applied to the NO1, NO4, and NO5 zone short-term load forecasts. The NO4 lagged feature had the best MAPE scores when applied to the NO3 and NO2 zone. Both the NO2 and NO4 spatiotemporal zones performed best when lagged with -2 either or -4. When checking the correlation before and after lagging the time series by -2 and -4, the correlation declines. However, the mutual information score remains around the same values.

Let us first consider the first scenario; the baseline vanilla LSTM results with just the electricity consumption of NO2 as input. Figure 5.1.8 shows the forecast in the blue line and the actual consumption in the orange for 48 hours in June 2020 for the NO2 zone. The MAPE result for the model is 1.63 %.

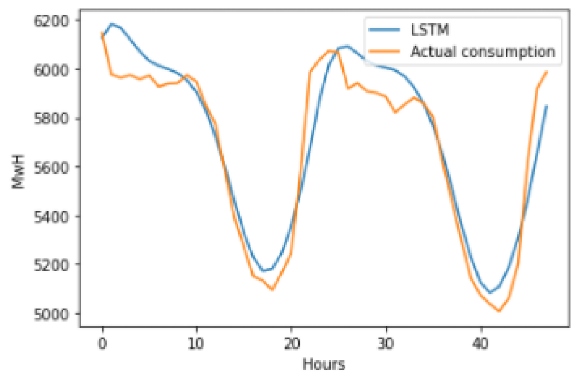



**Figure 5.1.8:** Vanilla LSTM forecast

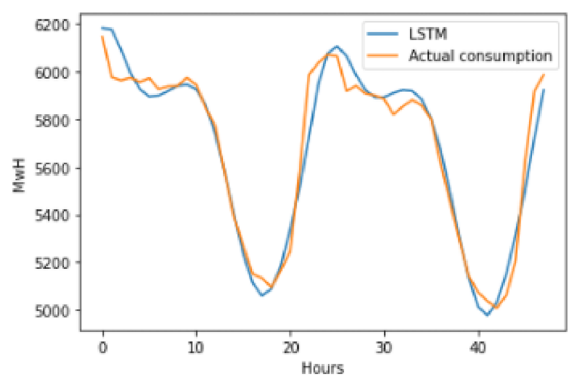



**Figure 5.1.9:** Lagged spatiotemporal LSTM forecast

Considering the second scenario, applying the NO4 as a lagged spatiotemporal feature for the NO2 forecast. This is executed for the same time and forecast horizon as the vanilla model. The figure 5.1.9 shows the forecast value versus actual consumption. The MAPE result for adding the lagged spatiotemporal feature to the baseline model is 0.97 %. Compared to the vanilla baseline model result, the improvement is a relative 65 %.

## 5.2 Validation Results and Discussion

For the validation technique of the short-term load forecasts for the use case of the Norwegian bidding zones, the most fitting approach is analysis. For this thesis' experiment it means analyzing the results from the same model and forecast horizon, but with different input to the model. Testing and validating the short-term load forecasts on the four seasons to have seasonal results is another extra validation step made. The evaluation will be presented in comparative tables with an explanation. To evaluate the quality of the best method found in this use case lies in comparing the results with the univariate LSTM model, the ARIMA model, and Statnett performances. The method which outperformed the other methods and models is to apply a lagged spatiotemporal feature to the baseline LSTM model.

In table 5.1, the seasonal test results are presented using the final method: *Lagged Spatiotemporal Feature Short-Term Load Forecasting* (LSTF STLF) using LSTM. These results were achieved by adding a lagged spatiotemporal feature alongside the consumption data for the forecast region. Which zonal features that were applied to a zone can be viewed in the spatiotemporal connectivity graph in 5.1.7. The 4 forecast test times are chosen at random given a season from summer 2020 to spring 2021. The forecast horizon is 48 hours, and those data points are not seen by the model. The training data ends at the given data point, and the predictions are made by the lagged spatiotemporal feature LSTM model based on the training.

**Table 5.1:** The MAPE results in percentage with the LSTF LSTM.

| Test time | NO1 | NO2 | NO3 | NO4 | NO5 |
|---|---|---|---|---|---|
| Spring 2021 | 1.54 | 0.97 | 1.25 | 1.42 | 1.55 |
| Winter 2020 | 2.36 | 2.12 | 2.08 | 1.76 | 1.86 |
| Fall 2020 | 2.32 | 1.85 | 1.76 | 1.82 | 1.62 |
| Summer 2020 | 1.75 | 1.98 | 1.75 | 1.31 | 1.74 |
| Average MAPE | 1.93 | 1.73 | 1.73 | 1.57 | 1.69 |

As can be discovered in table 5.1, region NO4 has the lowest average MAPE score at 1.57 %, while NO1 has an average score of 1.98 %. The results for Summer and Spring achieve the best performing short-term load forecasts. While the winter season is the toughest one to forecast.

**Table 5.2:** The average MAPE model scores in percentage.

| Model | MAPE % |
|---|---|
| LSTF LSTM | **1.73** |
| Vanilla LSTM | 1.98 |
| ARIMA | 3.20 |

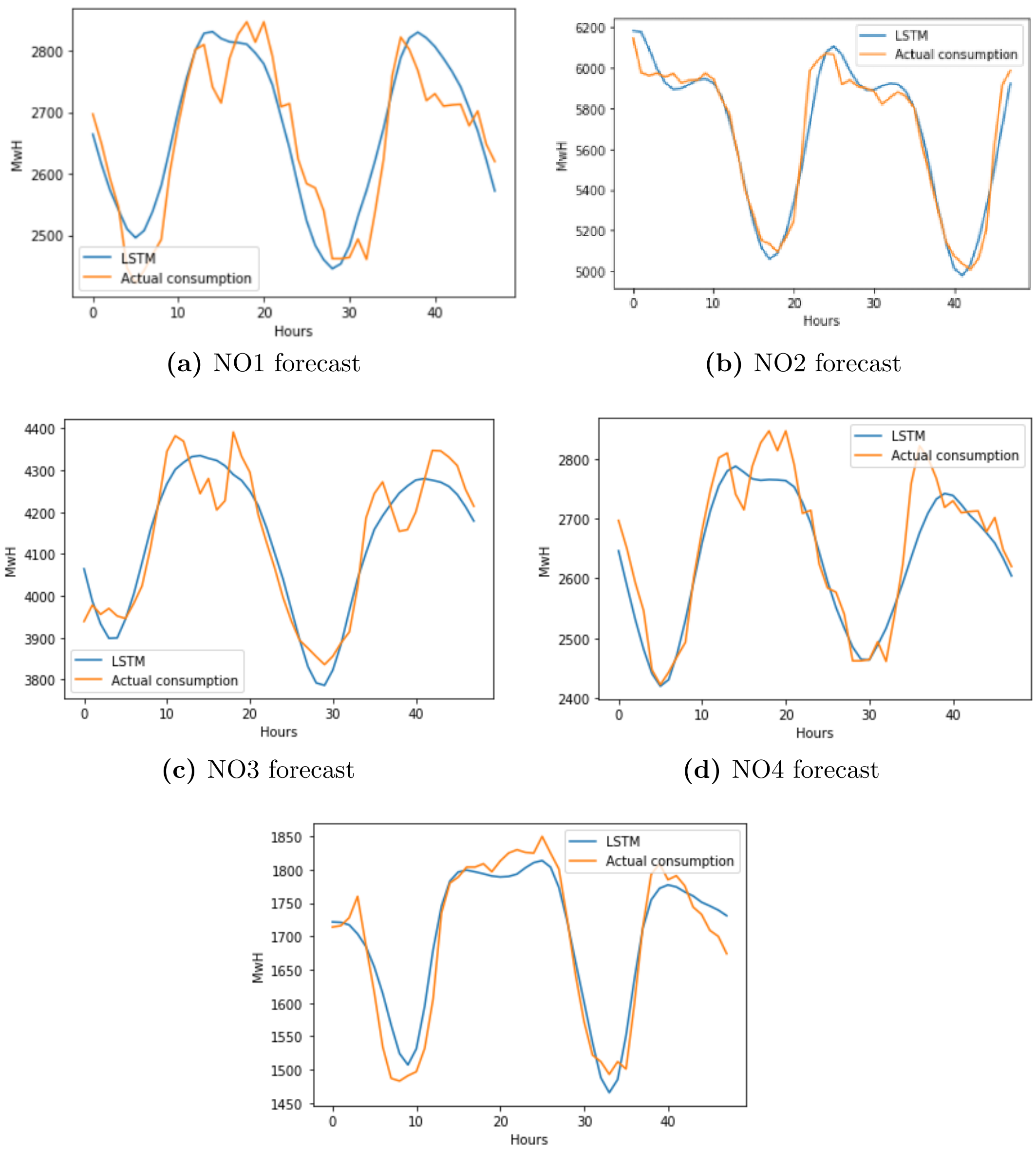
Table 5.2 shows the average MAPE results for the models; the Lagged Spatiotemporal Feature using LSTM(LSTF LSTM), the Vanilla LSTM, and the ARIMA model is tested on the same short-term load forecasting use case data. Applying only electricity consumption for the forecast zone in the Vanilla LSTM network, the results are accomplished. The average MAPE

score over all bidding zones from the seasonal results is 1.98 %. The LSTF LSTM MAPE score over all bidding zones from the seasonal results is 1.73 %. The ARIMA model performs an average MAPE of 3.20 % with results from the 5 bidding zones in Summer 2020 using univariate data.

Statnett's data science team achieved a score of 4.2 % in their approach to the short-term load forecasting of the bidding zones [4]. It is comparable in the fact they used the same recurrent neural network model and use case. However, their test data was from 2018 and can not be directly compared with the data from 2020 and 2021 in these tables. Statnett is the transmission system operator for the Norwegian electrical grid and their work relies on the accurate forecast. Their results were the benchmark we had set before the experiments.

When comparing the results with the vanilla model, the LSTF LSTM performs 14.4 % better. The ARIMA model had an average MAPE of 3.20 % for the 5 regions in summer 2020. The LSTF LSTM performs a relative 85 % better than the ARIMA implementation. The ARIMA model has not been as thoroughly tuned and tested as the LSTM baseline model.

In figure 5.2.1, the graphs presents five of the the validation results. The actual consumption lines are the orange ones, while the blue lines are the short-term load forecasts for the method LSTF LSTM. Every zone is represented in the collage. In figure 5.2.1 (a) the NO1 load forecast for Spring 2021 is displayed. The figure 5.2.1 (b) is the NO2 load forecast for Spring 2021.




**(a)** NO1 forecast

**(b)** NO2 forecast

**(c)** NO3 forecast

**(d)** NO4 forecast

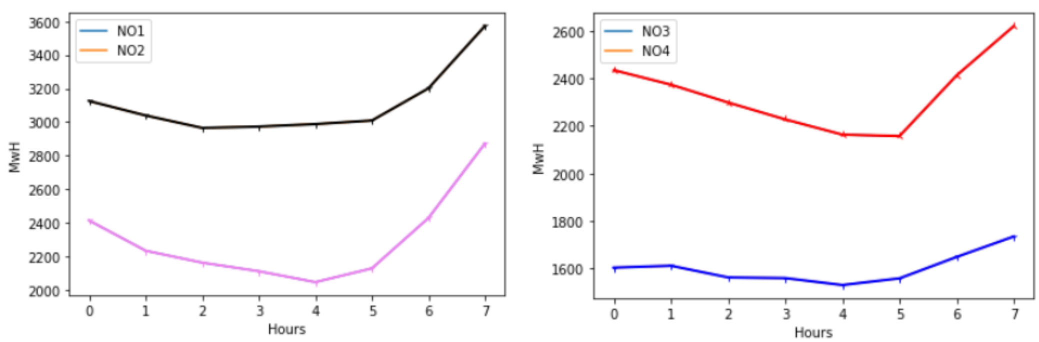
**(e)** NO5 forecast

**Figure 5.2.1:** A selection of the short-term load forecasts for the five bidding zones.

This was the best scorer of all results, with a MAPE of 0.97 %. Translated into MWh, each hour misses on average by approximately 56 MWh. The NO3 load forecast is for summer 2020, shown in figure 5.2.1 (c). In figure 5.2.1 (d), the NO4 load forecast for the summer of 2020 is displayed. Lastly, the NO5 forecast for Fall 2020 is shown in 5.2.1 (e).

When first applying the spatiotemporal bidding zones as features to the forecasts for another zone, the results showed some improvements from the vanilla model. The major milestone was when we tried lagging the spatiotemporal features. The short-term load forecasts improved. The reason for this is still up for debate. NO2 and NO4 have shown that they are the best spatiotemporal features to apply for the short-term load forecasting use case of the Norwegian bidding zones. The distinction in the results between the zones grew when lag was applied to the features as well. This underlines the fact that these two lagged spatiotemporal features are the best to utilize.

To explain why we could argue that the two zones score differently in the spatiotemporal data analysis. The NO2 zone has the highest average correlation with the other zones, while the NO4 zone has the lowest. For mutual information, the NO2 zone has the second-highest average, while NO4 has one of the lowest. Both of the zones have a relatively high score in both tests. There may be some explanation in this. The NO4 zone is the one zone that behaves slightly differently, and this could benefit the forecasts in regions NO2 and NO3. The correlation tells us that these spatiotemporal time series trend together. When we applied either -2 or -4 lag to these two spatiotemporal features and used them in the forecasts, all 5 zones achieved new all-time best MAPE scores for this experiment.




**(a)** NO1(pink line) and NO2(black line) **(b)** NO3(red line) and NO4(blue line)

**Figure 5.2.2:** Comparing morning electricity consumption for the zones.

In figure 5.2.2, the electricity consumption from midnight to 08:00 in the morning is displayed. As can be observed in figure 5.2.2 (a), the NO1 zone with the pink line starts the increase in consumption 1 hour before the NO2 zone with the black line. As the best result for NO1 forecasts is achieved when we apply lag -2 to the spatiotemporal NO2 feature, this makes sense. The NO2 data is shifted 2 hours back in time, meaning that the NO2 data starts the pick up in consumption for the morning before the NO1 zone. This gives the model a daily trend to learn. The same can be observed in figure 5.2.2 (b). The NO3 zone with the red line starts a major uptake in consumption 3 hours before the NO4 zone with the blue line increases much. The best results for the NO3 forecasts are with -4 lag applied to the NO4 spatiotemporal feature.

If we apply Dr. Hong's estimate [5] of what could be saved a year with 1 % improvement of short-term load forecasting to the Norwegian use case, this number reaches approximately 6.5 million USD based on the maximum peak load of 21.86 GWh for Norway in 2020. The average MAPE score of 1.73 % for the LSTF LSTM from this thesis' results compared with the MAPE result of 4.2 % from Statnett is a 2.47 % improvement. Following Dr. Hong's estimate, this thesis' final method could lead to savings of approximately 16.19 million USD for a full year for the Norwegian bidding zone use case. However, Statnett's article is from 2018, and they most likely have improved their short-term load forecasts by now.

As can be seen in figure 5.2.1, where we compare the forecasted values with the actual consumption, the model struggles with some of the peak loads. This is a common issue in load forecasting when using machine learning models and is difficult to mitigate. To avoid power outages these peaks have to be predicted or at least taken into account in the actual handling of the power grid.

The spatiotemporal weather variables did not improve the forecasts for this use case. All of the 44 weather variables collected for this thesis were applied to their respective zones, and only 2 of them improved the baseline vanilla model forecast result. When applying lag to the temperature variables, two zones improved slightly from the baseline result. This could be explained by the size of the bidding regions. Temperature is considered an important factor for electricity consumption. In other case studies it has shown improvements for the short-term load forecasts. For smaller use cases like household load forecasting, the temperature is a very important and sensitive feature to include.

As an artifact for the design science research method in this thesis was mainly to find a novel approach for short-term load forecasting, the proposed novel method is called: Lagged SpatioTemporal Features Short-Term Load Forecasting(LSTF STLF) using LSTM.

# Chapter 6

# Conclusion

Through a data-driven spatiotemporal feature selection methodology in this thesis, a novel method for approaching short-term load forecasting for a large use case been created. The solution I have found for the use case of the Norwegian bidding zones is applying lagged electricity consumption features as input to the LSTM model. The novel method proposed after following the thesis methodology presented in figure 4.2.1 is called:

*Lagged SpatioTemporal Features Short-Term Load Forecasting (LSTF STLF).*

The spatiotemporal data analysis and the grid search for the zones and lags revealed the two bidding zones NO2 and NO4 as the best applicable spatiotemporal features for the other zones. This selection can be substantiated with high positive correlation and mutual information scores. NO4 scored lowest in the correlation score, and this can be explained by the spatial distance from the other zones. Furthermore, that the NO2 and NO4 zones seem to start their daily morning increase in electricity consumption later than the other zones. The 44 weather variables collected for this experiment did not have any positive impact on the short-term load forecasts on regions of this size.

By using univariate electricity consumption data on the same optimized LSTM network, the average short-term load forecast MAPE for the 5 bidding zones is 1.98 %. With an added lagged spatiotemporal feature the average is 1.73 %. Both are good results for a forecast horizon of 48 hours. In real-world application, however, the 14.4 % relative improvement could mean a substantial amount of money saved in several integral parts of the energy sector. Demand-side management is crucial For the Norwegian TSO Statnett. They need to have accurate short-term load forecasts to operate the grid and control the electricity flow. The stakeholders who have the best load forecasts have a competitive edge in the energy trading market. Maintenance work for the generation companies has to be planned on a short-term aspect to minimize financial loss, and knowing when it is most cost-effective to plan these operations is dependent on the short-term load forecasts.

## 6.1 Answers to Research Questions

**Research Question 1: How are the spatiotemporal relationships for the data in the five bidding zones of electricity demand in Norway?**

The spatiotemporal relationships between the zones are highly positively correlated. The five bidding zones seem to trend up or down at approximately the same time. The mutual information score is also high, which strengthens the case that they are highly spatiotemporal dependent on each other. The spatiotemporal temperature variables had high negative correlation and high mutual information scores for their regions.

**Research Question 2 How to utilize the spatiotemporal dependencies in short-term load forecasting methods?**

When adding the spatiotemporal bidding zone features to the LSTM, 3 out of 5 zones improve. However, when we tried lagging these features in a grid search approach, the best results for this use case of short-term load forecasting for the Norwegian bidding zones were revealed. With a lag of either -2 or -4, the NO2 and NO4 zone were the optimal spatiotemporal features. When adding several of these features at once to the forecast region, the results showed that it caused too much noise. So by applying one lagged spatiotemporal feature, as can be discovered in the spatiotemporal connectivity graph in figure 5.1.7, the best way to utilize the dependencies was revealed. The temperature variables did not improve the short-term load forecasts until we applied lag. Two of the bidding zones slighly improved from the baseline result.

## 6.2 Future work

For further work on the same use case, the work would contain more optimization for the tuning of the hyperparameters, settings, and architecture of the LSTM model. The main focus for this thesis has been to find a novel method for short-term load forecasting, and the architecture could be scaled more. The novel method this thesis presents could be transferable to other countries and regions. First and foremost to Sweden, who has a similar separation for bidding zones as Norway. It should be scalable to other regions as well if there is enough historical and regional zonal data to analyze, train and validate on.

# Bibliography

[1] Tao Hong and Mohammad Shahidehpour. "Load Forecasting Case Study". In: *EISPC, U.S. Department of Energy* (Jan. 2015).

[2] Redjol Resulaj. "Smart Meter Based Load Forecasting for Residential Customers Using Machine Learning Algorithms". MA thesis. University of Stavanger, Norway, 2019.

[3] Nordpool. *Historical Market Data*. URL: `https://www.nordpoolgroup.com/historical-market-data/`. Accessed: 28.04.2021).

[4] Data Science @ Statnett. *Developing and deploying machine learning models for online load forecasts*. URL: `https://datascience.statnett.no/2018/12/20/developing-and-deploying-machine-learning-models-for-online-load-forecasts/`. Accessed: 08.03.2021).

[5] Dr. Tao Hong. *Crystal Ball Lessons in Predictive Analytics*. URL: `https://energycentral.com/c/um/crystal-ball-lessons-predictive-analytics`. Accessed: 07.05.2021).

[6] Kasım Zor, Oguzhan Timur, and Ahmet Teke. "A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting". In: June 2017, pp. 1–7. DOI: 10.1109/IYCE.2017.8003734. URL: https://www.researchgate.net/publication/319367655_A_state-of-the-art_review_of_artificial_intelligence_techniques_for_short-term_electric_load_forecasting.

[7] Kasım Zor, Oguzhan Timur, and Ahmet Teke. *A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting.* URL: https://www.researchgate.net/figure/Electric-load-forecasting-applications-and-classification_fig1_319367655. Accessed: 07.05.2021).

[8] Eirik Braaten Brose and Andreas Sandal Haugsbø. "Flow-Based Market Coupling in the Nordic Power Market". MA thesis. Norwegian School of Economic, Norway, 2019.

[9] Vikas Gupta. "An Overview of Different Types of Load Forecasting Methods and the Factors Affecting the Load Forecasting". In: *International Journal for Research in Applied Science and Engineering Technology* V (Apr. 2017), pp. 729–733. DOI: 10.22214/ijraset.2017.4132.

[10] Saurabh Singh, Shoeb Hussain, and Mohammad Abid Bazaz. "Short term load forecasting using artificial neural network". In: *2017 Fourth*

*International Conference on Image Information Processing (ICIIP)*. 2017, pp. 1–5. DOI: 10.1109/ICIIP.2017.8313703.

[11] Fazil Kaytez et al. "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines". In: *International Journal of Electrical Power Energy Systems* 67 (2015), pp. 431–438. ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2014.12.036. URL: https://www.sciencedirect.com/science/article/pii/S0142061514007637.

[12] Glen Mitchell et al. "A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types". In: *2017 IEEE Manchester PowerTech*. 2017, pp. 1–4. DOI: 10.1109/PTC.2017.7980814.

[13] N. A. Salim et al. "Case study of Short Term Load Forecasting for weekends". In: *2009 IEEE Student Conference on Research and Development (SCOReD)*. 2009, pp. 332–335. DOI: 10.1109/SCORED.2009.5443006.

[14] Chang Liu et al. "Short-term load forecasting using a long short-term memory network". In: *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 2017, pp. 1–6. DOI: 10.1109/ISGTEurope.2017.8260110.

[15] Qiang Jiang et al. "Electricity Power Load Forecast via Long Short-Term Memory Recurrent Neural Networks". In: *2018 4th Annual Inter-*

*national Conference on Network and Information Systems for Computers (ICNISC)*. 2018, pp. 265–268. DOI: 10.1109/ICNISC.2018.00060.

[16] Xiaoqun Liao et al. "Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days". In: *2019 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS)*. 2019, pp. 675–678. DOI: 10.1109/ICITBS.2019.00167.

[17] P. Mukhopadhyay et al. "Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter". In: *2017 7th International Conference on Power Systems (ICPS)*. 2017, pp. 812–819. DOI: 10.1109/ICPES.2017.8387401.

[18] M. Abdullah Al Amin and Md. Ashraful Hoque. "Comparison of ARIMA and SVM for Short-term Load Forecasting". In: *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. 2019, pp. 1–6. DOI: 10.1109/IEMECONX.2019.8877077.

[19] Shilpa Gn. "Short-Term Load Forecasting Using ARIMA Model For Karnataka State Electrical Load". In: (Dec. 2019). URL: https://www.researchgate.net/publication/337858487_Short-Term_Load_Forecasting_Using_ARIMA_Model_For_Karnataka_State_Electrical_Load.

[20] M Vetri Selvi and Sukumar Mishra. "Investigation of Weather Influence in Day-Ahead Hourly Electric Load Power Forecasting with New Ar-

chitecture Realized in Multivariate Linear Regression Artificial Neural Network Techniques". In: *2018 8th IEEE India International Conference on Power Electronics (IICPE)*. 2018, pp. 1–6. DOI: 10.1109/IICPE.2018.8709498.

[21] Jennifer Hinman and Emily Hickey. "Modeling and Forecasting Short-term Electricity Load Using Regression Analysis". In: (Jan. 2009). URL: https://www.researchgate.net/publication/242580638_Modeling_and_Forecasting_Short-term_Electricity_Load_Using_Regression_Analysis.

[22] Liufeng Du, Linghua Zhang, and Xu Wang. "Spatiotemporal Feature Learning Based Hour-Ahead Load Forecasting for Energy Internet". In: *Electronics* 9.1 (2020). ISSN: 2079-9292. DOI: 10.3390/electronics9010196. URL: https://www.mdpi.com/2079-9292/9/1/196.

[23] Gangui Yan et al. "Short-Term Load Forecasting of Smart Grid Based on Load Spatial-Temporal Distribution". In: *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*. 2019, pp. 781–785. DOI: 10.1109/ISGT-Asia.2019.8881259.

[24] Jie Shi, Yang Liu, and Nanpeng Yu. "Spatio-temporal modeling of electric loads". In: *2017 North American Power Symposium (NAPS)*. 2017, pp. 1–6. DOI: 10.1109/NAPS.2017.8107311.

[25] Alex Sherstinsky. "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network". In: *CoRR* abs/1808.03314

(2018). arXiv: 1808.03314. URL: http://arxiv.org/abs/1808.03314.

[26] Yue Zhang et al. "Data-Driven Day-Ahead PV Estimation Using Autoencoder-LSTM and Persistence Model". In: *IEEE Transactions on Industry Applications* 56.6 (2020), pp. 7185–7192. DOI: 10.1109/TIA.2020.3025742.

[27] Ljubisa Sehovac and Katarina Grolinger. "Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention". In: *IEEE Access* PP (Feb. 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020.2975738.

[28] Ratnadip Adhikari and R. K. Agrawal. "An Introductory Study on Time Series Modeling and Forecasting". In: *CoRR* abs/1302.6613 (2013). arXiv: 1302.6613. URL: http://arxiv.org/abs/1302.6613.

[29] Nordpool. *Day-ahead overview*. URL: https://www.nordpoolgroup.com/maps/#/nordic/. Accessed: 13.05.2021).

[30] Energy Facts Norway. *The Power Market*. URL: https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet%5C. Accessed: 12.04.2021).

[31] Meterologisk Institutt. *What is Frost?* URL: https://frost.met.no/index.html. Accessed: 23.05.2021).

[32] Alan Hevner et al. "Design Science in Information Systems Research". In: *Management Information Systems Quarterly* 28 (Mar. 2004), pp. 75–100.

[33] Tanwalai Panapongpakorn and David Banjerdpongchai. "Short-Term Load Forecast for Energy Management System Using Neural Networks with Mutual Information Method of Input Selection". In: *2019 SICE International Symposium on Control Systems (SICE ISCS)*. 2019, pp. 9–15. DOI: 10.23919/SICEISCS.2019.8758714.

[34] Wikipedia contributors. *Mutual information — Wikipedia, The Free Encyclopedia*. [Online; accessed 31-May-2021]. 2021. URL: https://en.wikipedia.org/w/index.php?title=Mutual_information&oldid=1025370848.

[35] Nima Amjady and Farshid Keynia. "Day-Ahead Price Forecasting of Electricity Markets by Mutual Information Technique and Cascaded Neuro-Evolutionary Algorithm". In: *IEEE Transactions on Power Systems* 24.1 (2009), pp. 306–318. DOI: 10.1109/TPWRS.2008.2006997.

[36] Zhiyong Wang and Yijia Cao. "Mutual Information and Non-fixed ANNs for Daily Peak Load Forecasting". In: *2006 IEEE PES Power Systems Conference and Exposition*. 2006, pp. 1523–1527. DOI: 10.1109/PSCE.2006.296526.

[37] L. N. Silva et al. "Impact Evaluation of Feature Selection to Short-Term Load Forecasting Models considering Weather Inputs and Load History". In: *2019 54th International Universities Power Engineering Conference (UPEC)*. 2019, pp. 1–6. DOI: 10.1109/UPEC.2019.8893598.

[38] Jin Xu et al. “Spatial-temporal load forecasting using AMI data”. In: *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2016, pp. 612–618. DOI: 10.1109/SmartGridComm.2016.7778829.

[39] Chitta Ranjan. *Step-by-step understanding LSTM Autoencoder layers*. URL: https://towardsdatascience.com/step-by-step-understanding-lstm-autoencoder-layers-ffab055b6352. Accessed: 21.04.2021).

references.bib