UPPSALA
UNIVERSITET

UPTEC STS 22029
Examensarbete 30 hp
Augusti 2022

Day-ahead Grid Loss Forecasting

A study of linear and non-linear models when modelling electrical grid losses

Alicia Söderlind

Civilingenjörsprogrammet i system i teknik och samhälle

UPPSALA UNIVERSITET

Day-ahead Grid Loss Forecasting

Alicia Söderlind

Abstract

Accurate day-ahead grid loss forecasts are, among other things, essential to determine the electricity price for the upcoming day. The more accurate forecast, the closer the trading on the 'day-ahead' electricity market can become the actual operation the next day, which decrease the need for correcting production on the balancing market. Followingly, the need for extra imbalance costs, which make the electricity price higher, is reduced with accurate forecasts. This project's purpose was to explore a wide-range of mathematical models to increase the energy company Fortum's day-ahead grid loss forecasting accuracy, and thereby contribute to lower the risk for high imbalance costs.

Two electrical grids located in Sweden, with different characteristics, were studied. One electrical grid was located in Dalarna and the other one was located in Värmland. Four different model types were tested for each grid. The linear models ARIMAX and SARIMAX were explored and the two artificial neural networks FNN (Feedforward Neural Network) and LSTM-RNN (Long-Short Term Memory Recurrent Neural Network) were explored. By constructing different model structures of each model type, as well as statically testing which predictors to include as input to the models, the most accurate model for grid loss forecasting was found. The models' forecasting accuracy were validated based on the MAPE (Mean Absolute Percentage Error).

Variables important as predictors were found to be power production, electricity prices and grid losses at previous time steps. For the grid in Dalarna ARIMAX(2,0,2) was the model generating most accurate day-ahead grid loss forecasts and for the grid in Värmland, SARIMAX(1,0,0)(0,0,1)[24] was the most accurate model. That is, different models were found as the most accurate one for grid loss forecasting, as the two studied electricity grids had different characteristics. Hence, this result implies that there is no universal model that is the most adequate at modelling all types of grid losses. To find useful models when forecasting grid losses day-ahead, an analysis of the particular grid losses being studied is therefore not irrelevant.

Teknisk-naturvetenskapliga fakulteten

Uppsala universitet, Utgivningsort Uppsala

Handledare: Gabriela Peña Ämnesgranskare: Per Mattsson

Examinator: Elísabet Andrésdóttir

Populärvetenskaplig sammanfattning

Elektricitet används på alla nivåer i samhället. I hemmen utförs bland annat mycket hushållsarbeten med elektriskt drivna apparater och elektrisk belysning möjliggör att vår dygnsrytm inte enbart styrs av solljuset. I industrin spelar elektricitet en viktig roll då tunga arbeten både underlättas och blir säkrare med elektricitetens hjälp. Även frakt- och transportmedel, såsom järnvägen, är beroende av elektricitet. Därmed skulle alternativet, att inte ha tillgång till el, innebära en helt annan vardag med ett fullständigt annorlunda sätt att leva på. Alltså är elektricitet en samhällsviktig resurs för att dagens samhälle ska fungera så som vi är vana vid.

Precis som med alla andra resurser är tillgängligheten på elektricitet en kostnadsfråga. Elpriserna bestäms på marknader med bud från producenter av el och konsumenter av el. Båda parter lägger bud på hur mycket el och till vilket pris de är villiga att sälja respektive köpa. Det finns olika typer av marknader, varav en är ’Dagen före marknaden’ (fritt översatt från engelskans ’Day-ahead market’). Som namnet antyder bestäms elpriset för nästkommande dag. Till exempel, elpriset på onsdagar bestäms på tisdagar. Likväl, under dygnets alla stunder måste produktionen av elektricitet vara precis lika mycket som konsumtionen. Om mängden produktion och konsumtion inte är lika kan elnätet bli instabilt och till och med kollapsa.

Dessvärre kan inte den på förhand överenskomna mängden el (på dagen före marknaden) alltid exakt överensstämma med det faktiska behovet av el dagen efter. En orsak är att konsumtionen ständigt varierar, bland annat baserat på årstid, veckodag och tid på dygnet. För att kompensera för avvikelser nästkommande dag anpassas produktionen för att matcha den rådande konsumtionen. Det sker på den så kallade balansmarknaden. Extra kostnader tillkommer till elpriset via balansmarknaden och dessa kallas obalanskostnader. Dessa kostnader är instabila och kan variera hastigt, vilket gör dem icke önskvärda eftersom elpriset riskerar att bli väldigt högt i och med tillskottet av dem. Därmed är det eftersträvansvärt att minimera behovet av obalanskostnader. För att minimera obalanskostnader behövs därför precisa prognoser av elkonsumtionen för nästkommande dag. Därtill är även prognoser för elnätsförluster nödvändiga.

När el transporteras i elnätet från producenter till konsumenter är uppkomsten av förluster oundviklig. Ur ett elnätsperspektiv är elnätsförluster samma sak som konsumtion, eftersom båda företeelserna innebär att energi ’plockas ut’ från elnätet. Följaktligen är även korrekta prognoser av elnätsförluster nödvändiga för att mer korrekt kunna handla på elmarknaden en dag i förväg, och därmed minska behovet av obalanskostnader.

I det här projektet har flera olika matematiska modeller undersökts med syfte att öka precisionen av energiföretaget Fortums prognoser för nästkommande dags

elnätsförluster. Förlusterna i två olika elnät i Sverige studerades, ett i Dalarna och ett i Värmland. För att hitta en passande modell som kunde beskriva förlusterna på ett korrekt sätt, för respektive elnät, valdes fyra olika modelltyper ut för att testas. Gemensamt för alla fyra modeller var att de konstruerades med samma metod. Metoden var att studera elnätsförluster som ett system med indata och utdata. Indata bestod av olika variabler som påverkar hur stora elnätsförlusterna blir och utdata var elnätsförlusterna.

Tillvägagångsättet var att anpassa de valda matematiska modellerna till uppmätt in- och utdata, eftersom modellerna därigenom dimensionerades för att beskriva elnätsförluster baserat på känd data, såsom väder-, elproduktions- och elprisdata. Följaktiglen kunde de anpassade modellerna användas för att prediktera framtida elnätsförluster baserat på de kända variablerna.

Genom det här projektet upptäcktes det att mängden producerad elektricitet i elnäten samt elnätspriserna var passande variabler för att beskriva de studerade elnätsförlusterna. Därtill visade det sig att mängden elnätsförluster vid tidigare tidpunkter kunde användas för att modellera framtida förluster än mer precist. Resultateten från det här projektet indikerar att mer exakta elnätsförlustprognoser är möjliga att generera med modellerna som testats i det här examensarbetet.

Table of Contents

1. Introduction

This introducing section contains general background of the subject, the purpose of the thesis, limitations and delimitations. Also, an overview of the report's disposition is presented.

1.1 General background

At all instances – every second, hour and day of the year – the exact amount of consumed electric energy must also be produced, meaning that power production coincide with consumption in time. In Sweden the main power resources generating electric energy are hydro power, nuclear power, wind power and thermal power [1]. The momentaneous occurrence of power production in time does generally not hold for the spatial occurrence. Power stations are often located far away from consumption, with power grids connecting the producers and consumers. Power grids are divided in two parts, transmission and distribution grids. In this thesis distribution grids refer to regional and distribution grids. The transmission grid transports electricity from power plants to substations and distribution grids transport power from substations to end consumers. In Sweden it is the state owned Transmission System Operator (TSO) Svenska Kraftnät (Svk) who alone operates the transmission grid, with voltage levels at 220 and 400 kV. There are many distribution grids which are operated by many different local Distribution System Operators (DSOs).

When transmitting electricity from producers to consumers, grid losses inevitably occur and some of the produced energy is 'lost' in the grid. Grid losses are the difference between the amount of power that is injected and the amount of power that is withdrawn from a grid. Injected power consists of two parts, local power production situated in the grid and imported power to the grid. Withdrawn power is also a two folded entity. Local consumption in the grid, as well as exported power from the grid [2]. Grid losses are undesirable since they reduce a grids' transmitting capacity. Moreover, losses also affect the electricity market and followingly has an impact on actors trading on the market.

1.1.1 The Swedish Electricity Market

The Swedish electricity market is divided into four areas: SE1, SE2, SE3 and SE4. The most northern area is SE1 and SE4 is the most southern (see Figure 1). The geographical borders of the four areas have been set where power transmission generally is constrained. If transmitting capacity across a border is constrained, a price difference occurs between the two areas. On the other hand, if capacity is sufficient, prices are equal in the two areas. Power flows from low price areas to high price areas [3]. In the northern zones SE1 and SE2 power production is normally larger than consumption, resulting in these areas being low price areas. The opposite holds for the

south, production is less than consumption, making the southern areas SE3 and SE4 high price areas. Hence, electricity is generally transmitted from north to south [4].

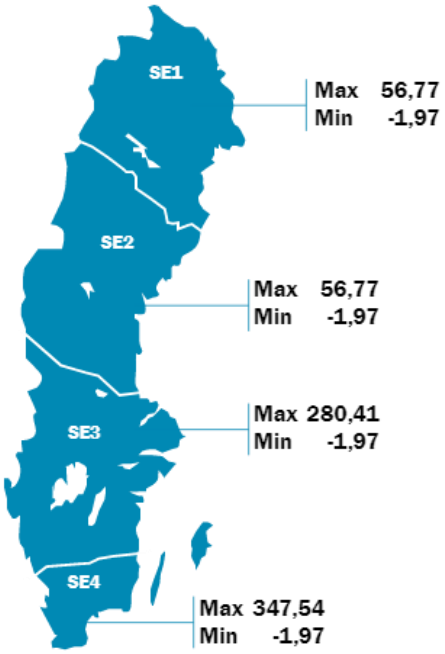


Figure 1: An illustration of the Swedish electricity price areas SE1, SE2, SE3 and SE4, located from north to south. Highest and lowest electricity prices in October 2021 are presented in euro, showing lower prices in the north and higher in the south [5].

In Sweden, as well as in all other Nordic countries, electricity prices are decided in an auction taking place on electricity markets. Nord Pool Spot is the market place for the day-ahead spot market Elspot and the intraday market Elbas [6]. On the day-ahead market, buyers and sellers place bids on how much electricity they can sell respectively buy the next, and to what price. Bids are placed every day before noon and concern the 24 hours in the upcoming day, hence, the name ‘day-ahead market’ [7]. When the auction has closed, hourly spot prices are determined at the point where selling and buying bids intersect, see Figure 2 for an illustration [8]. In comparison to the spot market, the intraday market enables bids on the operating day, which are placed one hour before the operating hour [9].

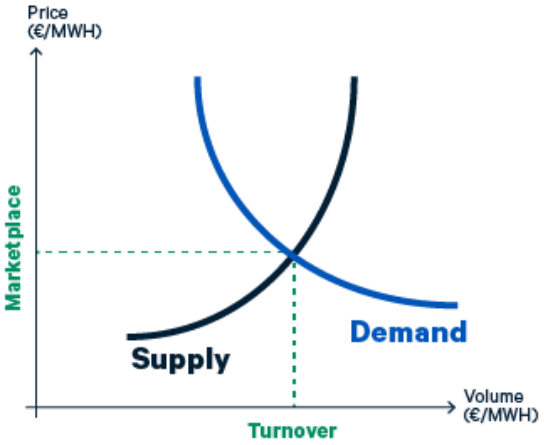


Figure 2: A schematic picture illustrating the concept of electricity prices being determined at the intersection of supplying and demanding bids [10].

Except for Elspot and Elbas there is also the balancing market operated by a country's TSO, who has the responsibility that there is balance in the electricity grid at all times [11]. In Sweden balance is achieved by keeping the grid frequency at 50 ± 0.1 Hz, which is accomplished when power production exactly matches consumption [1]. Consumption varies frequently due to seasons, weather conditions, outdoor temperatures, hour of the day, day of the week and holidays. If consumption is larger than production, production is not enough to meet the demand and grid frequency decreases. The other way around, consumption being less than production, means that too much electricity is produced in comparison to the demand and grid frequency increases. Therefore, to handle deviations in grid frequency, power production is altered to match consumption and in Sweden this is managed on the balancing market by Svk [12]. In addition to beforehand traded day-ahead and intraday prices extra costs arise on the balancing market. Imbalance costs are volatile and prices can fluctuate fast, making these costs undesirable and a reduction of them interesting to attain.

To reduce imbalance costs, accurate predictions of day-ahead power consumption is essential. The more precise consumption forecasts, the closer the day-ahead trading can become the actual operation the next day. Consequently, the need for extra imbalance costs at the operating hour decreases. In a power system perspective power consumption is not any different from grid losses. Consumption and losses are equal in the sense that power is withdrawn from the system. Therefore, it is not only forecasts of power consumption that are needed when deciding day-ahead prices on the day-ahead market. Accurate grid loss forecasts are also of importance to achieve adequate day-ahead trading. The responsibility to provide grid loss forecasts lays on all grid market actors who are a Balance Responsible Party (BRP). BRPs are responsible to report day-ahead grid loss forecasts to Nord Pool every day before noon. This means that the amount of expected grid losses at every hour of the upcoming day must be forecasted and presented. [2]

1.1.2 Technical Grid Losses

There are two types of losses that arise in power grids; non-technical and technical. Non-technical losses include theft, measurement errors and non-measured power consumption [13]. On the other hand, technical losses are caused by well-known physical constraints in components shaping power grids. These components are e.g. conductors, transmission and distribution line equipment, transformer windings and cores. Technical grid losses can both be fixed and variable. Fixed losses do not change as electric current changes, whereas variable losses change as current varies. In other words, variable losses depend on the amount of electricity being transmitted in the electricity grid [14].

In general there are five types of active power losses that appear when alternating current is transmitted. These are: 1) resistive heat losses, 2) leakage losses, 3) dielectric magnetizing losses, 4) dielectric polarization losses and 5) corona losses. Overall, these five types of losses depend on current, voltage and grid frequency in different forms [15]. Additionally, grid losses relate to the electrical network configuration, in terms of grid development, outages and maintenance. Market behaviour and weather conditions are also factors directly affecting grid losses [16].

Weather conditions such as solar radiation, wind speed and outdoor temperature have been found to have a complex effect on power production, consumption and losses. Hence, power transmission is a very dynamic process [16], [17], [18]. As seen in Figure 3, grid losses can have a seasonal behaviour, depending on the month, week, day and hour. Furthermore, as the share of intermittent generation – wind and solar power – increases, the more weather dependent and dynamic the grid losses become. Followingly, grid losses become an even more difficult phenomenon to forecast [18].

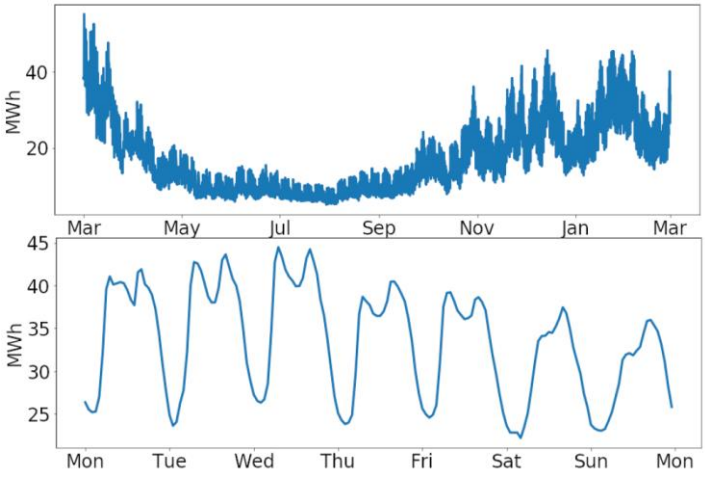


Figure 3: Two graphs illustrating a seasonal behaviour of grid losses. The top graph shows grid losses over one year. The bottom graph shows grid losses over one week [17].

1.2 Purpose

The purpose of this thesis project was to investigate mathematical models and predictors when forecasting day-ahead grid losses at the energy company Fortum. Fortum is a BRP and obliged to report adequate day-ahead grid loss forecasts to Nord Pool on a daily basis. In this project, a range of forecasting techniques were explored with the aim to increase Fortum's day-ahead grid loss forecasting accuracy and thereby contribute to lower the risk for producers and consumers paying high imbalance prices. To fulfil the purpose, following aims were set up as guidance.

Aim 1. Identify predictors that are of importance when forecasting grid losses.

Aim 2. Distinguish suitable model structures for grid loss forecasting.

Aim 3. Validate tested models and compare the forecasting accuracy with the forecasting tool used at Fortum.

1.3 Limitations & Delimitations

1.3.1 Limitations

Fortum's consumption data were available on a network and geographical scale only, which was not relevant to use when modelling grid losses. Hence, pure power consumption data were not used as a predictor when forecasting gird losses. Instead, other variables were used as a measure of consumption. More about that is addressed in Section 4.1.

It was not possible to get access to the forecasting models used by Fortum, causing the Fortum models appear as black-boxes. The lack of information about the model structures restrained a detailed analysis of the Fortum models. Consequently, the comparison of this thesis' models with Forum's forecasts could not righteously and evidently state if the explored models were better or worse at forecasting grid losses. Instead, the comparison gave an indication if the explored models were more or less accurate at forecasting losses compared to Fortum's models. In other words, the comparison was limited to suggest, and not doubtlessly declare, if tested models in this project were more adequate or not.

Another limitation consisted of the amount of time and computational capacity that was required for tested model types in this project. That is, time and computational capacity needed for parameter estimation of some of the regression models and parameter optimisation of Artificial Neural Networks were limited. These model types are described further in Section 2.2.1 and 2.2.2.

1.3.2 Delimitations

A necessary delimitation was to model grid losses in no more than two electrical grids. As time was limited, two grids had to be chosen instead of modelling losses in all grids. One of the chosen grids was located in the Swedish county Dalarna and the other one in Värmland. These two grids were most relevant to study due to their grid losses being largest in volume in Fortum's portfolio.

Furthermore, it was essential to delimit the types of models being explored. As there is a huge amount of mathematical models presented in research work, it was neither reasonable or feasible to test all of them. Hence, relevant model types for this project was distinguished by the literature review of related studies, conducted in this thesis.

1.4 Disposition

The report is structured in seven chapters, the first being the introduction presenting the outline of the project. In the second chapter a theoretical background is given, addressing forecasting techniques in general and related work regarding grid loss forecasting in particular. Thereafter, a data chapter follows and the chosen data processing procedure is presented. Chapters four and five include the method and results. The two last chapters are discussion and conclusions, summing up the findings and learnings from this project.

2. Theoretical Background

This section consists of existing theory and knowledge of forecasting methods. Text books and articles from previous studies in the field construct the base of knowledge needed for this project.

2.1 Models and Forecasting Techniques

Models are a common tools used to describe real systems and illustrate their behaviours. There can never be one 'true' model of a system since models are simplifications and not exact replicas of reality. Hence, no model is perfect but some are useful in certain situations. To find a useful model in a certain situation many different models should be tested and validated. Two types of modelling approaches are *physical modelling* and *system identification* [19]. For technical systems, physical modelling typically means that laws of nature are utilised to construct models, and the objective is to model fundamental physical properties in the system. For instance, Ohm's law can be utilised when modelling electrical systems. However, modelling vast and complex systems, such as electrical systems, by only physical laws can be extremely difficult. To correctly model a system with physical modelling, all physical phenomena involved in the system must be known. Therefore, the other modelling approach called system identification is often used as a complement to physical modelling, or entirely used by its own [19].

2.2 System Identification

The principal of system identification is to use measured observations from a system to adapt a model so it accurately represents the behaviour of the system. Based on measured input and output data from a system, a mathematical model can be found, describing how input values transfer in the system to become the output values. This means that models found by a system identification approach heavily depends on the data gathering process. Available data lay the foundation for the rest of the modelling procedure and followingly affects the outcome and final model's usefulness. Hence, this approach does not require extensive insight about how everything in the system works, as physical modelling does. Figure 4 illustrates the type of models, and typical input-output relation, usually identified within system identification. Input is denoted u, a disturbance signal is denoted e and output is denoted y [19].

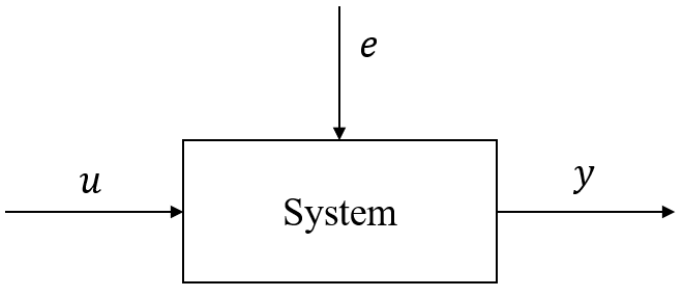


Figure 4: A schematic illustration of the principal of a system identification approach. Measured data (u) and disturbance (e) are inputs to a system which gives the measured output (y).

A highly general flowchart of the modelling process is presented in Figure 5. 'Step 1' in Figure 5 is to choose what data to gather from the system, which then must be processed and studied. For example, missing or erroneous samples, detected outliers or unwanted trends must be dealt with [19]. Thereafter, input variables (u) that are of importance to describe the output variable (y) should be identified. Statistical tests such as correlation, causation and regression analyses are common tools used to find variables that preferably should be chosen as input (u) to describe the modelled output (y) [20]. In this thesis, variables significant as input to describe the output variable are referred to as predictors.

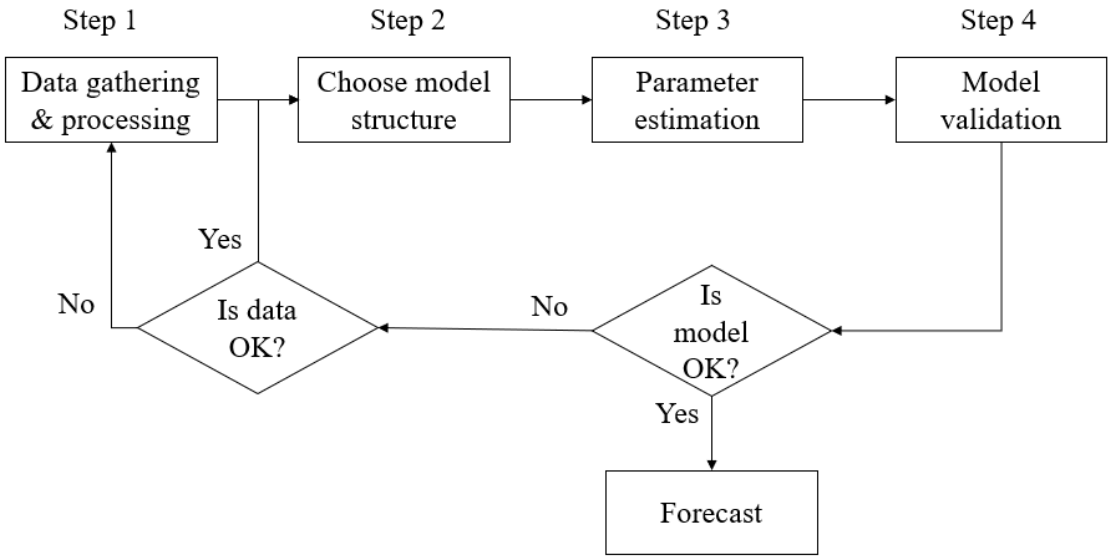


Figure 5: The general process of modelling regression models. Rectangles are processes and the diamonds are decision points.

With pre-processed data and chosen predictors, one moves to 'Step 2' in Figure 5, choosing a model structure to model the system. There are many different model structures that can be explored when modelling systems. Therefore, it is important to decide model structure in consideration of the specific system and purpose of the model. This report focuses on models found as key examples in related research work. Figure 6

shows a taxonomy tree of models reviewed in this project. For a more complete literature review of models, see e.g. [21]. Generally, models can be divided in physical and system identification models as stated above. In turn, system identification models can be categorised in linear and non-linear models. The theory of these two model types are presented in more detail below.

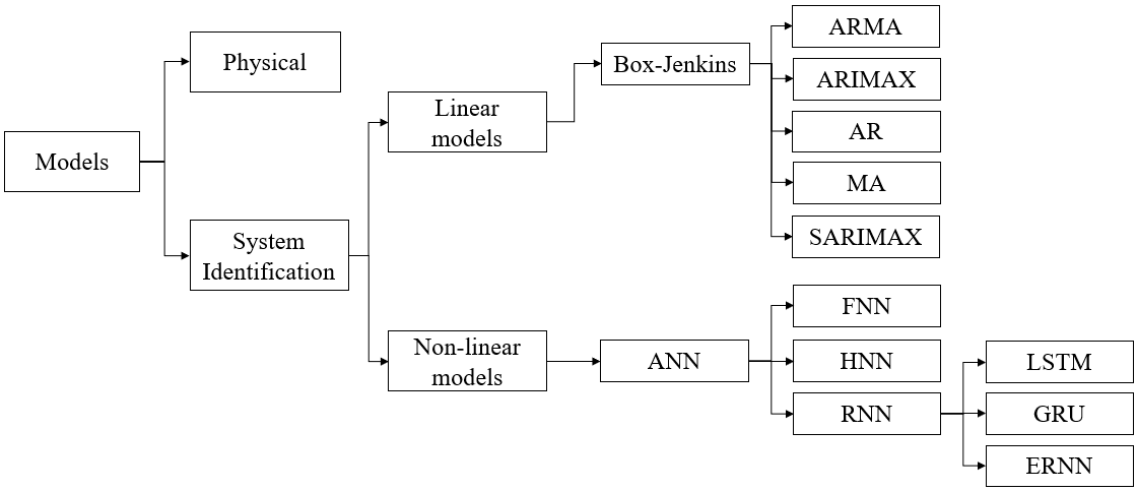


Figure 6: Taxonomy tree of models reviewed in this project.

2.2.1 Linear Models

There is a group of standardised linear model structures that are based on transfer functions, $G(q,\theta)$ and $H(q,\theta)$, mapping input data $u(t)$ and a disturbance signal $e(t)$ to the modelled output $y(t)$. The transfer functions contain unknown variables, which are constructing an n-dimensional column vector θ, generally called the *parameter vector*. When adapting a model to measured data, the aim is to find the unknown variables in the θ vector, so the output accurately can be modelled based on the input. All these models are variants of the Box-Jenkins model and can be described by the equation:

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t), \tag{1}$$

or written as

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t), \tag{2}$$

where

$$\frac{B(q)}{F(q)} = \frac{b_1 q^{-nk} + b_2 q^{-nk-1} + \cdots + b_{nb} q^{-nk-nb+1}}{1 + f_1 q^{-1} + \cdots + f_{nf} q^{-nf}} = G(q,\theta), \tag{3}$$

$$\frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}}{1 + d_1 q^{-1} + \cdots + d_{nd} q^{-nd}} = H(q,\theta), \tag{4}$$

and q^{-n} is the backward shift operator defined as $q^{-n}f(k) = f(k-n)$. For instance, multiplying a variable with q^{-3} is the same as taking the variable three timesteps ago, at t-3 [19]. As seen in (3)-(4) the two transfer functions $G(q,\theta)$ and $H(q,\theta)$ are made of polynomials $B(q)$, $F(q)$, $C(q)$ and $D(q)$, containing four sets of variables b_i, f_i, c_i and d_i. These sets of variables are the unknown elements, in the θ vector.

How many unknown coefficients there are of each variable is determined based on the orders of the polynomials. The polynomial orders are set by the user who chooses model structure by the five *hyperparameters* nk, nb, nf, nc and nd. That is, selected hyperparameters determine the polynomial orders and consequently the model structure. All model structures have different properties being either advantageous or disadvantageous when describing studied systems and their behaviours [19].

A common model structure is the ARMAX model, represented by setting the $F(q)$ and $D(q)$ polynomials equal to each other by setting $F(q) = D(q) = A(q)$. ARMAX stands for AutoRegressive Moving Average with an eXogenous variable. ARMAX models are described by writing (2) as

$$A(q)y(t) = B(q)u(t) + C(q)e(t). \tag{5}$$

The autoregressive (AR) term $A(q)y(t)$ involves modelling the output variable (y) based on its own values at previous time steps. The moving average (MA) term $C(q)e(t)$ models the disturbances. Lastly, the exogenous term $B(q)u(t)$ means that other variables (u) are used as input [19].

Models presented so far assume that time series data are stationary, generally meaning that time series have a constant mean and variance over time. The formal definition of stationarity is that the $A(q)$ and $C(q)$ polynomials contain no roots inside the unit circle, which ensures causality and invertibility [22]. However, a time series might have trends, such that the mean value and variance vary in time, called non-stationary data. It is not uncommon that data is non-stationary in real applications, but methods designed for stationary time series should not be used on non-stationary data. If non-stationary data is used there is a risk of the results giving erroneous answers [23]. However, non-stationary data can be transformed by differentiation to make it stationary. To differentiate non-stationary data with ARMAX models an integrating term 'I' is added, constructing ARIMAX models. Differentiation is selected based on successive unit-root tests and the order is determined once no unit root is found [24]. The order of differentiation is often 1 or 2. There could also be periodic patterns in data. To handle periodic patterns a seasonal term 'S' can be added to the ARIMAX model, creating the SARIMAX structure [25]. The value of the seasonal term 'S' is an integer m, set to the number of periods in one season. For example, m can be chosen as 12 to handle data with a monthly periodicity [26]. Mathematically SARIMAX models are described by writing (2) as

$$A_S(q^m)A(q)\Delta^d\Delta_S^D y(t) = B(q)u(t) + C_S(q^m)C(q)e(t), \tag{6}$$

where $A(q)$ and $C(q)$ are non-seasonal AR and MA components, $A_S(q^m)$ and $C_S(q^m)$ are seasonal AR and MA components, Δ^d and Δ_S^D are non-seasonal and seasonal differentiating operators, m is the seasonal term and $B(q)$ is the exogenous polynomial [27]. In Python [28] the polynomials and differentiating components are defined as

Non-seasonal:

$$A(q) = 1 - a_1 q^{-1} - \cdots - a_{na} q^{-na}$$
$$C(q) = 1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}$$
$$\Delta^d = (1 - q)^d$$

Seasonal:

$$A_S(q^m) = 1 - a_1^S q^{-m} - \cdots - a_{NA}^S q^{-NA*m}$$
$$C_S(q^m) = 1 + c_1^S q^{-m} + \cdots + c_{NC}^S q^{-NC*m}$$
$$\Delta_S^D = (1 - q^m)^D$$

Note that lowercase hyperparameters na, d and nc belong to the non-seasonal models and uppercase hyperparameters NA, D, NC and m are included in seasonal models. With Python notation, ARIMAX and SARIMAX models are expressed as ARIMAX(na, d, nc) and SARIMAX$(na, d, nc)(NA, D, NC)[m]$ [29]. This notation is used when referring to the models in the remaining parts of the report.

When selecting hyperparameters for models with an ARMA-type structure, plots of the Partial AutoCorrelation Function (PACF) and AutoCorrelation Function (ACF) can be used as guidance. This method is limited, but brings an indication of the orders needed when iteratively testing different combinations of model orders. The PACF plot gives an estimate of the $A(q)$ polynomial. The ACF plot is used for estimating the order of the $C(q)$ polynomial [25]. See Figure 7 for an example.

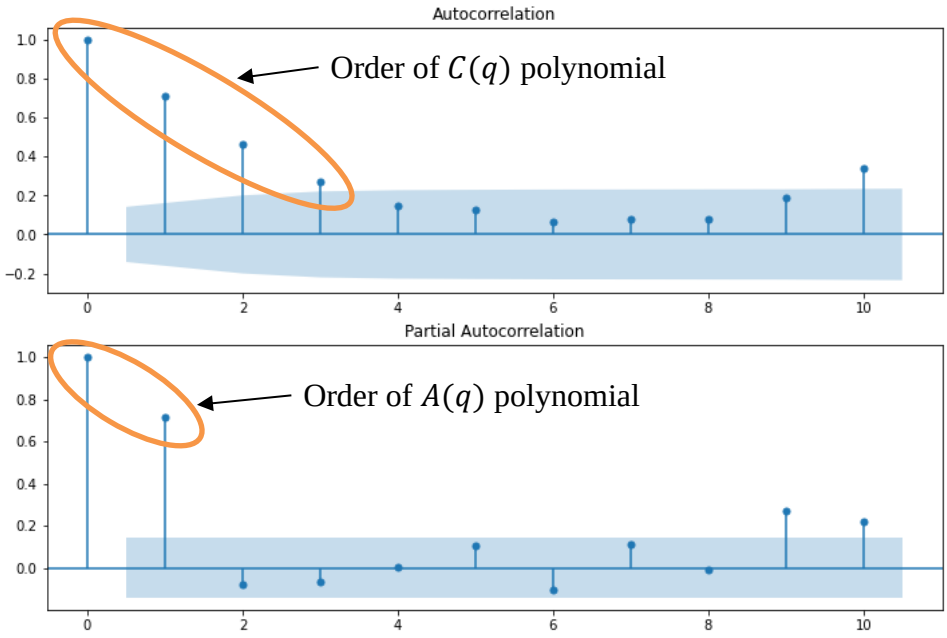


Figure 7: An example of an ACF plot (top graph) and a PACF plot (bottom graph). The plots indicate a 4:th order C(q) polynomial and a second order A(q) polynomial.

Furthermore, hyperparameters can be determined based on standard regression and statistical tests of parameter significance. Which order of the polynomials that are of

significance can be investigated with the statistical z- or t-test[1]. From the tests a p-value is produced for each parameter, showing the probability to reject the null hypothesis, insignificant parameters. If the p-value is less than or equal to the significance level 0.05 there is at least a 95 % chance the parameter affect the output variable. By this approach a model order can be found when all polynomial parameters are significant. By the same statistical approach, exogenous variables that are significant to use as input to the model can be found. [30]

In summary, choosing polynomial orders is 'Step 2', choose model structure, in Figure 5 for linear Box-Jenkins models. Further, linear models constitute the most simple type of models and may be explored first in a modelling process. According to the principle to describe a system as simple as possible but no simpler, it is a valid approach to test linear models first. However, if there are nonlinearities in a system, it is not certain that linear models are able to describe the fundamental system behaviours good enough. There is a risk of oversimplifying real world systems and in that case other types of models should be tested. How model structure is chosen for non-linear models are described next.

2.2.2 Non-linear Models

One of the most common non-linear model types is Artificial Neural Networks (ANN), in shorter terms known as Neural Networks (NN). In computer science NN modelling is a widely used technique for forecasting variables of linear and non-linear systems. NNs can detect important behaviours of systems that have complicated internal physics or complex relationships between variables, that otherwise are difficult to fully discover and understand.

The structure of a NN contains neurons (nodes), one input and one output layer, one or more hidden layers, activation functions, weights and biases/thresholds [31]. Figure 8 gives an illustration of a NN. Each parameter has its own effect on the input and these are described more thoroughly below.

[1] In statistics the t-test is a hypothesis testing method, investigating if explanatory variables have independent effects on the dependent variable. The t-test is called z-test for large sample sizes [31][31].

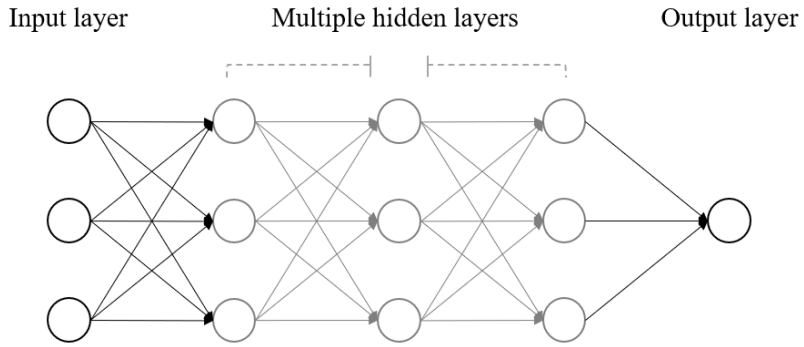


Figure 8: An illustration of a NN constructed of 3 hidden layers and 3 neurons in each layer, except for the one neuron in the output layer.

In practice the approach when modelling systems with NNs, is to decide the number of hidden layers, the number of neurons in each layer and type of activation functions. The number of neurons and number of layers are the hyperparameters of an NN, which the programmer choses to build the model [33]. Before adapting the constructed NN to measured data, data can be normalised to take values between 0 and 1. Moreover, a batch size and number of epochs must also be chosen. Instead of feeding NNs with the entire training dataset, a batch size divides the data to smaller batches and adapts the network by feeding the model pieces of data at a time. The number of epochs is a parameter defining how many times the learning algorithm goes through the NN [34].

Weights and biases are the unknown parameters in NNs. Thus, weights and biases are not parameters chosen by the programmer, but optimised by computations [35]. By referring to the description of linear models, presented in Section 2.2.12.2.2, weights and biases construct the θ vector containing the model's unknown parameters.

The computational procedure of NNs is the following. Each given input is assigned a weight, expressing the importance of the input to the output neuron. An important input variable will be assigned a larger weight, making its contribution to the output more significant compared to other input variables. The inputs are multiplied with their respective weights and summed. This weighted sum is then passed through the activation function. There are different types of activation functions that can be chosen, e.g. sigmoid (logistic), ReLu (rectified linear activation) or tanh (hyperbolic tangent) (for a complete revision of these, the reader may refer to Malekian et al. [31])

Activation functions are static functions determining the value of the output node based on the input. If the output value exceeds a given threshold (the bias) it activates the node and data is fed forward to the following layer. That is, if a node is activated, the output of the node becomes the input to nodes in the next layer. When data only travels in one direction, from one layer to the next, it is called a *Feedforward* Neural Network (FNN) [36]. How the output of one node is calculated, based on input values, weights, plus a bias and activation function, is demonstrated in Figure 9.

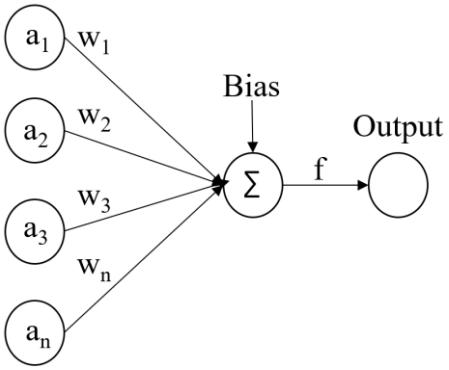


Figure 9: An example of how the output node is calculated based on input nodes a_n, their corresponding weighing factors w_n, a bias and an activation function f.

As an equation, the algorithm in Figure 9 is described as

$$output = f\left(\sum_{i=1}^{n} w_i a_i + b\right). \tag{7}$$

where f is an arbitrary activation function, w_i are weights, a_i are inputs and b is the bias.

Since data travels in one direction in FNNs they are static models, meaning that given the same input, the same output is always obtained. However, there are NNs that enable feedback loops and these are called Recurrent Neural Networks (RNNs). Activation of neurons might be determined by previous activations of that specific neuron or earlier input to the neuron [37]. More specifically, RNNs give feedback to a *hidden state* which is connected to the hidden layers in the RNN. Each hidden layer has its own hidden state. Consequently, the calculated output, at every node in the network, depends both on current input and the hidden state, which is based on previous input [38]. That is, the main difference between FNNs and RNNs is that RNNs have hidden states, allowing output to be calculated based on current and previous input.

RNNs have become a competitive alternative to other more established models for time series forecasting. One advantage with RNNs is that they can model seasonal behaviours directly from data containing homogenous seasonal patterns [39]. However, one challenge with RNNs is that the training process becomes more difficult. Gradients of the cost function are not only propagated backwards in the network layers, but also backwards in time. To solve this issue there are many different methods addressing the gradient problem when training RNNs [37]. The most common techniques used are the Long-Short Term Memory (LSTM), Grated Recurrent Unit (GRU) and Elman Recurrent Neural Network (ERNN) techniques. Hewamalage et al, found that LSTM was the best approach for forecasting purposes compared to the two other techniques GRU and ERNN [39].

That is, choosing type of NN, the number of hidden layers, the number of neurons in each layer, type of activation functions, batch size and number of epochs is 'Step 2' in Figure 5. The following step, after choosing a suitable model, is adapting the model to fit measured data from the studied system. This is 'Step 3', parameter estimation, in Figure 5.

2.2.3 Parameter Estimation

Regardless of chosen model structure, linear or non-linear models, estimating the unknown parameters in the θ vector is the same process. The aim of estimating the unknown parameters is to adapt a model as accurate as possible to known measured data. However, one risk when fitting models is overfitting. A model is overfitted when it also fits to the noise in the data. The risk of overfitting increases as models become more complicated with a large number of parameters. Hence, there is a principal to describe a model as simple as possible but no simpler [19].

The overall procedure is to find the unknown parameter vector θ that makes residuals $\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta)$, errors between measured output ($y(t)$) and predicted output ($\hat{y}(t|\theta)$), as small as possible [40]. The θ vector can be found by the principle of the *least squares method*, meaning that the loss function

$$V(\theta) = \sum_{t=1}^{N} \varepsilon(t,\theta)^2 = \sum_{t=1}^{N} (y(t) - \hat{y}(t|\theta))^2, \tag{8}$$

is to be minimized with respect to θ. The θ that minimizes the loss function is denoted $\hat{\theta}$ and contains the estimated parameters that are adapted to the data [40]. A technique frequently used to find the solution of such a minimization problem is called *gradient descent*. In short, the gradient of the loss function ($\nabla V(\theta)$) is calculated to find the loss function's minimum. The gradient show the direction in which the loss function decreases the most, and the gradient decent technique moves the estimated θ in that direction. As an iterative process, the gradient descent algorithm finds a local minimum of the loss function [36].

For linear, Box-Jenkins models the optimal one step prediction of the next coming output is given by

$$\hat{y}(t|\theta) = [1 - H^{-1}(q,\theta)]y(t) + H^{-1}(q,\theta)G(q,\theta)u(t), \tag{9}$$

where $\hat{y}(t|\theta)$ is the predicted output, $H(q,\theta)$ and $G(q,\theta)$ are the transfer functions containing the unknown sets of variables (b_i, f_i, c_i and d_i), $y(t)$ is the measured output and $u(t)$ is the measured input [19]. Hence, when adapting Box-Jenkins models to data, the set of variables b_i, f_i, c_i and d_i that minimises the loss function in (8) are sought to obtain.

Regarding NNs, the forecasted output $\hat{y}(t|\theta)$ is determined by the output of all nodes that are connected in the network. As NNs can be large, interconnections between all nodes can become highly complex, entailing that the calculated output $\hat{y}(t|\theta)$ is complicated to express with a general equation. However, as described in Section 2.2.2, it is all the weights and biases in the NNs that construct the unknown θ vector. Thus, when adapting NNs to fit data, it is desired to obtain the set of weights and biases that minimises the loss function.

When unknown parameters in the chosen model has been estimated, by adapting the model to measured data, it is important to validate the model's performance ('Step 4' in Figure 5). To adapt and thereafter validate a model, data can be divided in two datasets. A training dataset and one validation dataset. With two separated datasets, models are not validated on the same data they have been designed to fit to.

2.2.4 Model Validation

When validating a model, it is decided if gathered and processed data, as well as the choice of model structure, generated a model capable of describing the studied system. Validation of a model commonly consists of comparing the modelled output with the actual measured output, given in the validation dataset. The difference between the modelled and real output shows how accurately the model calculate future values, and this is called *prediction error* [19]. If the model is considered valid it can be used for forecasting. Otherwise, either the data gathering and data processing must be redone or another model structure must be tested. Hence, it is an iterative process to find an accurate model; this iterative process is seen as the loop in Figure 5.

One subtle but important difference when modelling the output, is the difference between simulating versus predicting the output. Simulations are produced if input data only consist of other variables and not the output variable, whereas predictions are obtained if input also contain values of the modelled output at previous time steps [41].

Apart from using the prediction error as validation, another essential part when validating linear models is to investigate the residuals. When creating a linear model it is generally assumed that the errors are independent, identically distributed random variables (i.i.d), coming from a normal distribution with zero mean and variance σ^2 [42]. This is the same properties as white noise have. Therefore, for residuals from a valid model the following should be considered.

- No autocorrelation. Residuals should be independent from previous time steps and from input variables.
- Residuals should follow a normal distribution.
- The residuals' variance should be constant over time. [42]

Residuals fulfilling these properties are considered valid and can be accepted. On the other hand, residuals not fulfilling these properties imply that there might be some

behaviour in the studied system that has not been captured in the chosen model. In other words; residuals not looking like white noise pose that another model can describe the system better than currently tested model [19].

2.3 Grid Loss Forecasting Models

Previous research work in the field has approached grid loss forecasting in various ways. Key examples of the grid loss forecasting research, available in the literature, is presented in this section.

Sulakov presents a solution to forecast hourly day-ahead losses in the Bulgarian transmission grid [16]. Based on a statistical approach it was investigated which parameters that correlate with grid losses. It was found that the amount of injected electricity in the grid correlated with resistive ohmic transmission losses. Further, hourly contracted net exports were detected to correlate linearly with transmission losses. Sulakov also identified that grid losses increased due to humid and icy weather conditions. Quadratic functions were found to describe the relation between weather conditions and corona losses. Quadratic functions were also found to describe the relation between grid losses and the amount of renewable energy production in the system, i.e. solar production and wind power production [16].

Hence, the forecasting model took hourly load forecasts, daily patterns of net export, forecasted meteorological conditions, forecasted hourly wind power and forecasted hourly solar power production as input. The linear and quadratic functions describing how the variables related to the different types of technical grid losses were solved with regression functions [16]. That is the type of modelling described in Section 2.2.1.

Another more recent study from 2020 by Dalal et al., used a machine learning technique to forecast grid losses for 24 hours the next day in a distribution network in Norway [17]. Historical measurements of grid losses, calendar features, temperature forecasts and load predictions were identified as key features when forecasting day-ahead grid losses. In the study four different algorithms were evaluated: 1) an FNN with five hidden layers, 2) a decision tree regressor, 3) a gradient boosting regressor from sci-kit learn[2] and 4) CatBoost, an opensource tool for gradient boosting on decision trees. CatBoost with minimal hyperparameter tuning was chosen as the best model because it had the smallest MAE (Mean Absolute Error) [17].

In another study from 2020 conducted by Tulensalo et al. a LSTM-RNN was suggested when forecasting losses in electricity grids [18]. Tulensalo et al. applied the LSTM-RNN model to forecast transmission losses in the Finnish transmission grid with the purpose to predict losses in the intraday market. A NN approach was chosen due to the complexity of power systems, the increasing share of renewable power generation and

[2] A module in Python.

followingly the weather dependence and nonlinear correlations related to grid losses. Moreover, an LSTM-RNN was chosen since RNNs can capture non-linear time-variate dependencies. [18]

Input data consisted of calendar data, which were indicators of weekdays, years and public holidays and a sinusoidal representation of the daily cycle. Historical weather data was another input, consisting of solar radiation, wind speed and outdoor air temperature. Historical electricity market data was also used as input, consisting of hourly wind power generation, electricity demand and power flows from other countries. Lastly, system-specific data of the Finnish power grid was used as input. The pre-processing of data made the input follow a uniform distribution on the interval [0, 1] and the output was scaled to have unit variance and zero mean [18]. This procedure belongs to the type of modelling described in Section 2.2.2.

In addition to these grid loss studies, other projects have investigated which models to use when forecasting losses in power grid components or when forecasting power production and consumption. In 2013 Bouzerdoum et al. successfully forecasted hourly production of a photovoltaic power plant a few days ahead [43]. A hybrid SARIMA-SVM[3] model gave the most accurate predictions among the tested models [43]. A study from 2020 by QiheLou et al. showed a similar result, but with forecasting hourly power consumption day ahead. A hybrid SARIMA-SVR[4] model was found to be the best performing model of the ones tested in the study [44].

It has also been found that NNs are useful in related fields. Saelee et al. [45] suggests a Hierarchy Neural Network (HNN) when predicting power losses in a 3-phase distribution transformer. In a project by Lopez-Martin et al. a NN was proposed for probabilistic short and medium term electric load forecasting [46]. Further, Hong and Fan point out that many types of NNs successfully have been used for electric load forecasting, and especially successful have been FNNs and RNNs [47]. Regarding RNNs, Hossen et al. found that the LSTM was more adequate than, the other popular RNN model, GRU when forecasting residential load day-ahead [48]. Lastly, Deb et al. have shown that hybrid models consisting of two or more model types, e.g. linear and non-linear models, are most effective at forecasting buildings energy consumption [49]. By using hybrid models the different model types compliments each other, strengthening the hybrid model's advantages. For more examples of commonly used NNs, please read the article by Gamboa [50].

[3] SVMs (Support Vector Machines) are a set of classification models.
[4] SVRs (Support Vector Regressions) are a set of regression models.

3. Data

This section begins with a visualisation and explanation of the data used in this project, followed by a description of the data pre-processing procedure. All practical steps in this study were implemented by the open source programming tool Python.

3.1 Available Data

Data used in this project were supplied by Fortum and consisted of weather, electricity spot prices, power production plans and grid losses. Weather data consisted of real measured values, spot prices were the realised values published by Nord Pool, production data was the day ahead planned production by power plants and grid losses were historical data provided by Ellevio.

To adapt and validate models, slightly more than six years of hourly data were available from 2016-01-01 at 00:00 to 2022-02-27 at 23:00. The first 5 ½ years were used for model adaptation and the last ½ year was used for validation. Grid loss forecasting was investigated in two grids; one in Dalarna[5] (DAL) and the other in Värmland[6] (VML), which correspond to 63% of the total annual grid losses in Fortum's portfolio at the moment of writing this thesis.

There are some distinctive differences between the DAL and VML grids. The DAL grid is characterised as a grid that distribute power and the VML grid is characterised as a grid that transmits power. In the DAL grid injected power is mainly local production, whereas in the VML grid the injected power is mostly imported power. Another difference is that withdrawn power from the DAL grid consists of both consumption and export, whereas withdrawn power from the VML grid foremost consists of exported power. That is, the main differences are that the DAL grid have local production and consumption, while the VML grid primarily works as a power transmitting infrastructure by importing and exporting power.

Figure 10 and Figure 11 show the historical measured grid losses in DAL and VML grids from January 2016 to February 2022. Autocorrelation plots (Figure 12) showed that grid losses correlated with themselves at previous time steps. There was a 24 hour seasonality trend in autocorrelation for both DAL and VML losses. Furthermore, as a result of the autocorrelation, future losses were chosen to be predicted instead of simulated. As predictions, in comparison to simulations, use output at previous timesteps as input, and thereby should provide better forecasts compared to simulations.

[5] Geographically located in central Sweden, in north-western Svealand. The grid is in market area SE3.
[6] Geographically located in central Sweden, in western Svealand. The grid is in market area SE3

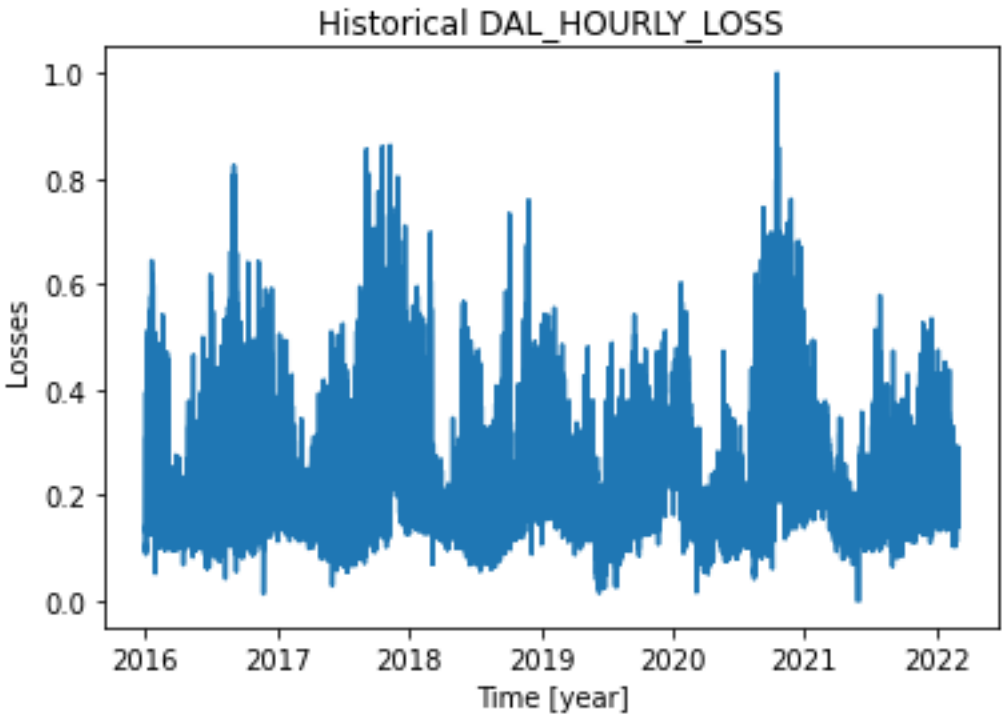


Figure 10: All historical data samples of DAL losses. Plotted data are normalised.

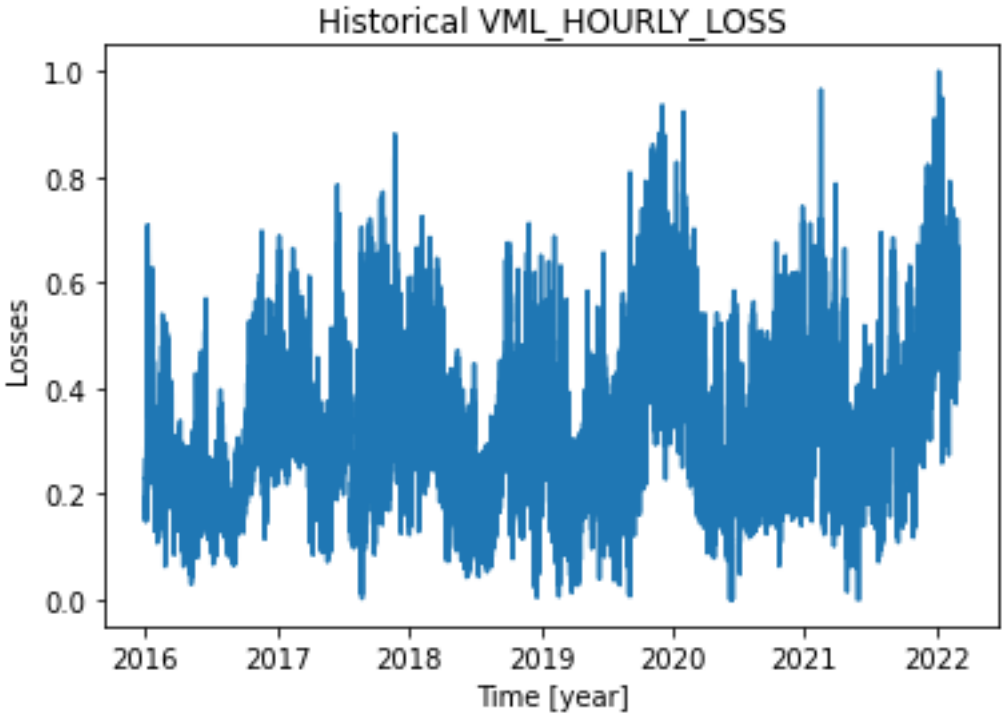


Figure 11: All historical data samples of VML losses. Plotted data are normalised.

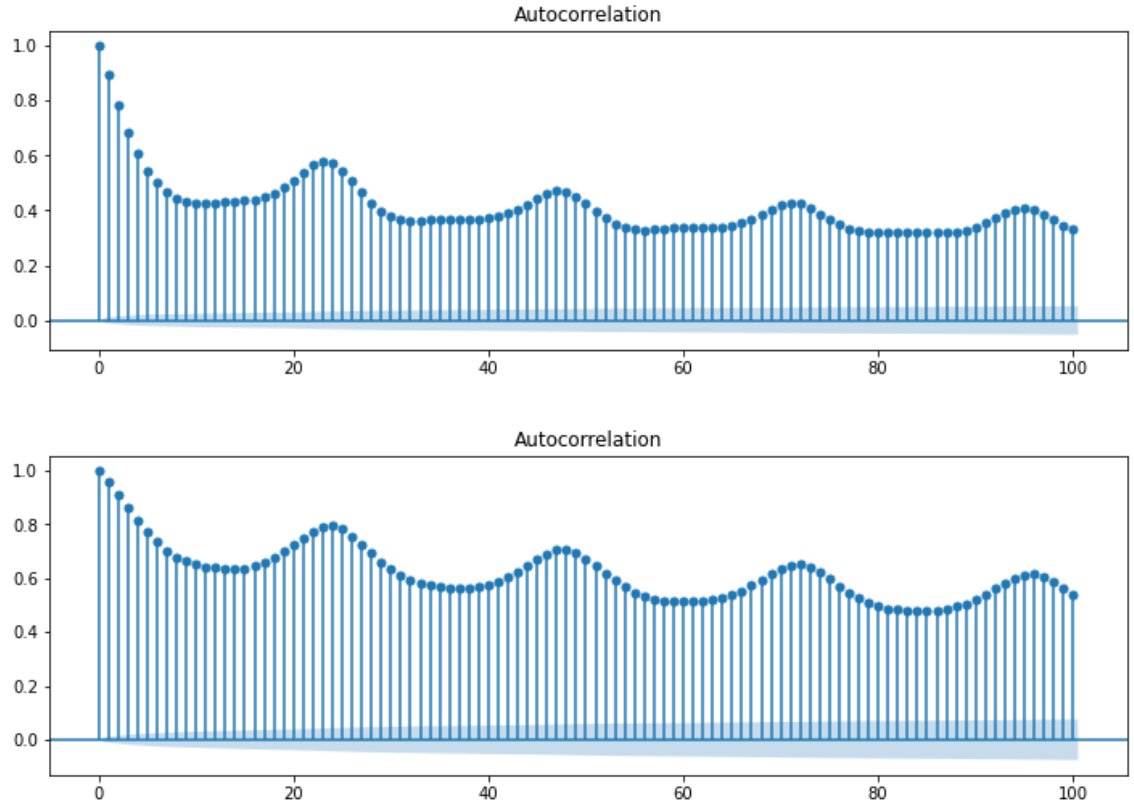


Figure 12: Top graph shows autocorrelation with 100 lags for DAL grid losses. Bottom figure shows autocorrelation with 100 lags for VML grid losses.

3.2 Data Pre-Processing

Data were first processed by removing the mean value from the time series. When removing the mean, the variance of the time series is modelled instead of the raw data. Mean values are removed to better deal with skewed data. Thereafter, outliers were removed and stationarity was checked.

3.2.1 Removing Outliers

Before using historical grid loss data in the modelling process outliers were detected and removed. It is important to identify outliers since including them, especially in parametric modelling, can cause faulty results leading to erroneous conclusions. One useful method to detect and remove outliers is the Interquartile Range method (IQR) [51]. When defining outliers, the method utilises the difference between the first and third sampled quartiles, Q_1 and Q_3. The method is based on inner fences f_1 and f_3 and outer fences F_1 and F_3, specified as:

$$f_1 = Q_1 - 1.5(Q_3 - Q_1),$$
$$f_3 = Q_3 + 1.5(Q_3 - Q_1),$$
$$F_1 = Q_1 - 3(Q_3 - Q_1),$$
$$F_3 = Q_3 + 3(Q_3 - Q_1).$$

Data outside inner fences f_1 and f_3 are considered mild outliers and data outside outer fences F_1 and F_3 are defined as extreme outliers [51]. Inner fences were applied in this project when identifying and removing outliers in grid loss data, since more samples are considered outliers when using inner fences. From DAL losses, 0.046 % of the data was removed and from VML, 0.014 % of the data were removed. Figure 13 illustrates detection and removal of outliers in DAL losses with the IQR method. The same approach was used when removing outliers in VML losses.

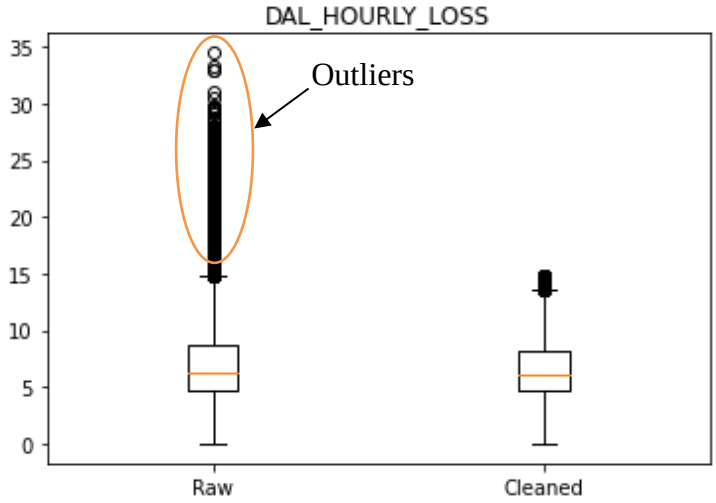


Figure 13: An illustration of detected and removed outliers with the IQR method. The plot show grid loss data from the electricity grid in Dalarna.

3.2.2 Stationary Data

With outliers removed from data it was tested if the grid loss time series were stationary. Stationarity means that a time series has a constant mean and constant variance over time. If data is not stationary, a transformation should be applied. One common technique is data differentiation. To statistically test stationarity, the widely used test of Augmented Dickey-Fuller (ADF) was used [52], [53]. The null hypothesis of the ADF test is that time series are non-stationary. If the p-value from the test is less than the significance level 0.05, the null-hypothesis can be rejected, implying that data are stationary. That is, a p-value less than 0.05 is the desired result [54].

The p-value for DAL losses was $2.30 * 10^{-26}$, far less than 0.05 and the null hypothesis of non-stationarity was rejected. The p-value for VML losses was $1.60 * 10^{-19}$, which also entails strong evidence against the null hypothesis of non-stationary data could be rejected. Therefore, the two time series of grid losses in DAL and VML did not have to be processed further to make them stationary.

4. Suitable Predictors & Models

The processes of identifying predictors and model types of relevance, for grid loss forecasting in the two studied grids, are described in this section.

4.1 Predictors

By definition, grid losses are the difference of injected and withdrawn power from a grid, therefore, ideal predictors when forecasting losses ought to be the produced, imported, consumed and exported power as; $Production - Consumption + Imports - Exports - Losses = 0$. However, the total production and consumption at grid level were unknown for this study, since several other utilities apart from Fortum, operate within each grid area.

To generate day-ahead grid loss forecasts, grid loss predictors require to be available for the upcoming day. Therefore, predictors must first be generated based on other predictors. Among other variables, forecasted power production is based on forecasted spot prices and forecasted consumption is based on weather predictions. This means that spot prices and weather forecasts could be relevant predictors for grid loss forecasting. Figure 14 illustrates how consumption and production forecasts are related to other predictors and the chosen exogenous variables in this project are coloured in grey.

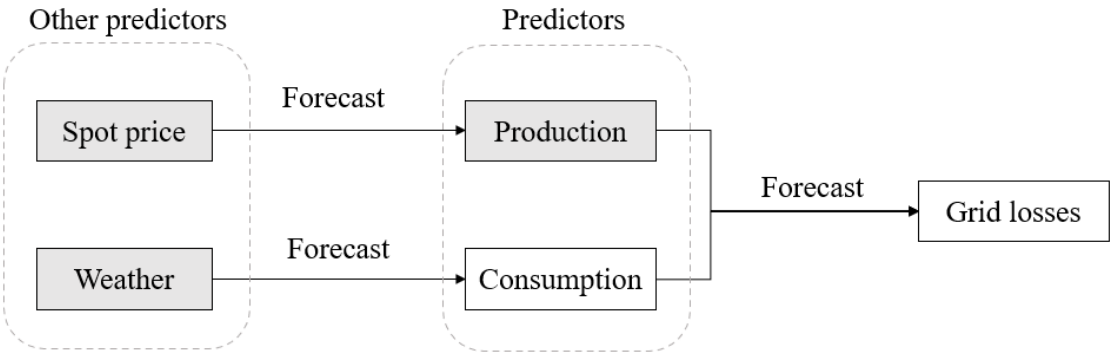


Figure 14: A flow chart showing that grid losses can be modelled based on forecasted spot prices and weather conditions. Grey coloured variables are exogenous variables chosen for this thesis.

In this thesis, historical data of weather (air temperature, solar radiation and wind speed) and spot prices (in the SE2 and SE3 price areas) were used together with day-ahead power production plans and historical grid losses in the training set. Due to consumption not only depending on weather, but also on consumption patterns over weekdays, weather data together with a day-type classification were treated as exogenous variables to predict consumption. Also, Tulensalo et al. used calendar day as an input variable as a way of day-type classification when forecasting grid losses [18]. Spot prices were used as exogenous variables to predict production, since these are key parameters to optimize hydro production. Additionally, day-ahead production plans for

key hydropower reservoirs located in the Dalälven and Värmland grids were included as predictors, since these capture other physical characteristics of the power system, hydrology water permits and market fundamentals, among others.

Based on the literature review presented in Section 2.3, several studies also found that weather data, historical grid losses, power production and consumption are good predictors to forecast grid losses. Table 1 summarises the datasets used in this thesis project.

Table 1: A summary of the datasets available as input for grid loss modelling and forecasting.

Dataset	Data used in training set	Day Ahead forecasts available?	Source
Historical grid losses	Historical data	No	Ellevio
Air temperature	Measured data	Yes	Weather forecast provider
Solar radiation	Measured data	Yes	Weather forecast provider
Wind speed	Measured data	Yes	Weather forecast provider
Power production	Historical data	Yes	Fortum
Spot prices	Historical data	Yes	Nord Pool and Fortum

Air temperature, wind speed and solar radiation were measured at the counties, Dalarna's and Värmland's, respective local weather stations. In Dalarna the weather station was located in Sala and in Värmland it was located in Sunne. Weekdays were implemented by constructing a variable 'Weekday', taking values 1-7, where 1 represented Monday and 7 Sunday. With the weekday variable, all hourly grid losses were categorised by an integer as a day of the week.

The most important hydropower producers in the DAL grid are the three hydropower plants of Trängslet (Trä), Gråda (Grå) and Mockfjärd (Mfj), all located in the Dalälven river network. In Värmland, power production is largely dominated by nine hydro plants: Röjdåfors (Rfs), Kymmen (Kmn), Gullspång (Gsg), Glava (Gla), Höljes (Hös), Tåsan (Tån), Letten (Ltn), Jössefors (Jfs) and Rottnen (Ron). Due to power plants in VML have stronger interdependences in production, a new variable 'PROD_TOT' was created by adding the production from the nine hydropower plants.

When investigating relations between grid losses and the other variables, correlation was measured with the Pearson's correlation coefficient r, taking values between 0 and ±1. Zero means no correlation, 1 implies a positive correlation and -1 a negative correlation [42]. Variables resulting in r being close to 0 had low correlation and were

not considered important to include as input to the models. High correlation was considered for variables showing either highly positive or negative r values, indicating that those variables could be necessary to include as predictors when constructing the models.

In the model adaptation and validation step of the linear models, predictors were investigated further with a statistical analysis. The z-statistics was used to investigate if chosen exogenous variables, previously identified with the Pearson coefficients r, were statistically relevant to include in the models. The common significant level 0.05 was chosen, entailing that variables with p-values smaller than 0.05 were accepted and variables with p-values greater than 0.05 were removed as input to the models.

4.1.1 Chosen Predictors

Predictors identified with Pearson's correlation test are shown in heatmaps, see Figure 15 and Figure 16. Based on Figure 15 it was considered relevant to include temperature, weekday, production in TRÄ, GRÅ, MFJ and spot prices in SE2 and SE3 as exogenous variables, when modelling DAL losses. Based on Figure 16 relevant exogenous variables for VML losses were temperature, weekday, total production and spot prices in SE2 and SE3.

All variables, apart from temperature and weekday, were chosen since they had highest positive correlation, as seen as lightest colours in the heatmaps. Temperature was chosen due to its negative correlation with grid losses, seen as a dark blue colour. Even though weekday neither showed high or low correlation it was tested if weekdays together with temperature gave more accurate forecasts, since the combination could correlated with power consumption.

When adapting and validating models for both DAL and VML losses it was found from the z-test that temperature and weekday were not significant variables and smaller MAPE was not achieved when temperature and weekday were used as inputs. Instead, it was found that power production and spot prices were the most valuable predictors when modelling losses in the DAL and VML grids. The predictors for each grid and their explicit r values are given in Table 2 and Table 3.

Table 2: Predictors for DAL losses and Pearson's correlation coefficient r.

DAL predictors	**Pearson correlation *r***
TRÄ	0.73
GRÅ	0.38
MFJ	0.14
SE2	0.17
SE3	0.24

Table 3: Predictors for VML losses and Pearson's correlation coefficient r.

VML predictors	**Pearson correlation *r***
PROD_TOT	0.41
SE2	0.12
SE3	0.27

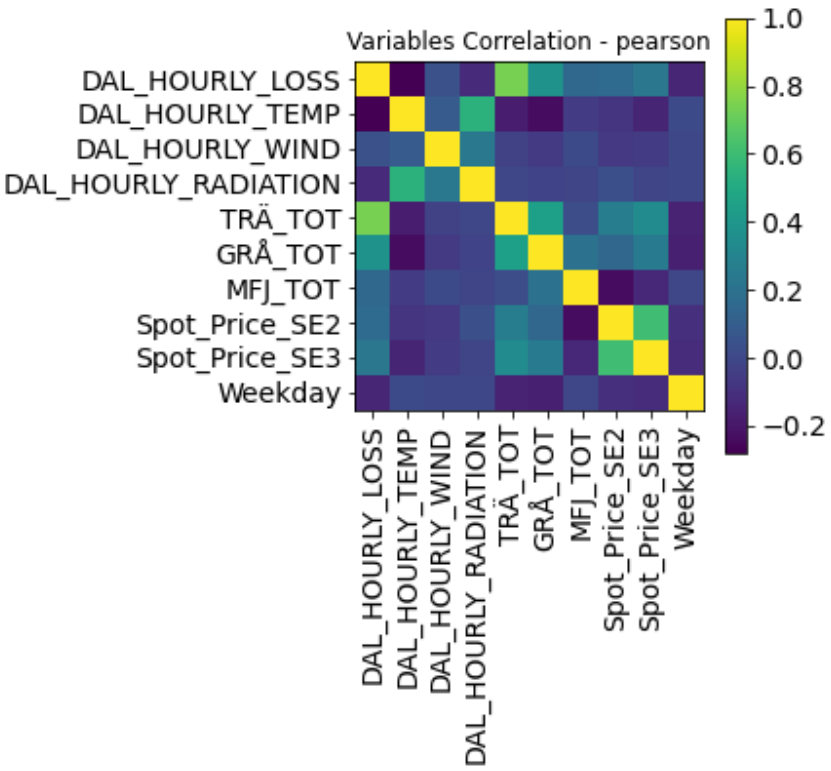


Figure 15: Heatmap based on Pearson's correlation for DAL losses and other possible variables affecting losses. Higher correlation is represented by a more yellow coloured tone.

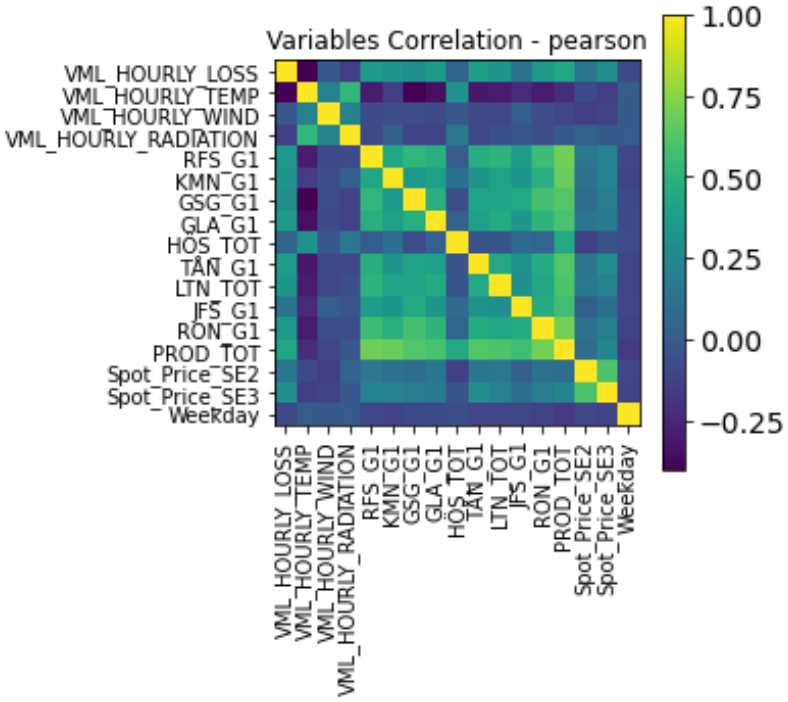


Figure 16: Heatmap based on Pearson's correlation for VML losses and other possible variables affecting losses. Higher correlation is represented by a more yellow coloured tone.

4.2 Proposed Models

Based on the literature review in Section 2.3, it was found that linear models are largely used in the literature and are suitable models for grid loss forecasting. Thereby, ARIMAX and SARIMAX models were considered relevant to test in this project. Moreover, as the nature of this data with a trend (moving average component) and an autoregressive component, as grid losses correlated with itself at previous time steps, it was appropriate to test models with a moving average and an autoregressive component. Hence, ARIMAX models were tested. The SARIMAX model structure was also suitable to test due to seasonality characteristics of the DAL and VML grid losses. Relevant predictors discussed in Section 4.1.1 were included as exogenous variables suitable to try.

Due to the complexity of grid losses, it was suitable to also test more complex models than linear models. As the non-linear NNs can capture non-linearities and complicated interactions appearing in systems, NNs were also tested in this project. Feedforward neural networks (FNNs) and recursive neural networks (RNNs) were considered relevant to explore due to evidence in the academic literature where these show to be the most successful NNs in electric load forecasting [47]. Regarding RNNs, the LSTM technique was of most interest to test, by reasons of the literature review in Section 2.3. In the literature review it was found that LSTM models were more suitable for time

series forecasting of grid losses, and predictions in related fields e.g. power consumption, compared to the two other popular RNN techniques GRU and ERNN.

Thus, the tested models were

- Linear models – ARIMAX
- Linear models – SARIMAX
- Non-linear models – FNN
- Non-linear models – LSTM

To find the most accurate model, the various tested models' forecasting performances had to be validated.

4.2.1 Model Validation

The models were validated based on the mean absolute percentage error (MAPE), a measurement of the difference between forecasted losses and actual losses, defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|,$$

where n is the number of data points, $\hat{y}_i$ is forecasted values and y_i is measured values [55].

Based on the models' purpose to make accurate predictions for the next coming day, the models made a 24 hour forecast for all days in the 6 months validation dataset. One hour at a time was forecasted during the 24 hour forecast. The Python method *forecast()* was used and input to the models were updated every 24 hour. That is, the input was updated but the models were not refitted to added available data for every new 24 hour forecast. This validation method is called backtesting without refitting and the concept is illustrated in Figure 17. Re-estimating models as more data become available is also an option, meaning that models are re-adapted to the latest available data before every new forecast [56]. This approach was not chosen as backtesting without refitting is a faster strategy when having large sets of data, as the models are fitted to data once. Additionally, since a forecast was generated for every day in the 6 month validation dataset (that is 180 days), the models had to be refitted 180 times with the re-estimating strategy instead of only once, which takes more time.

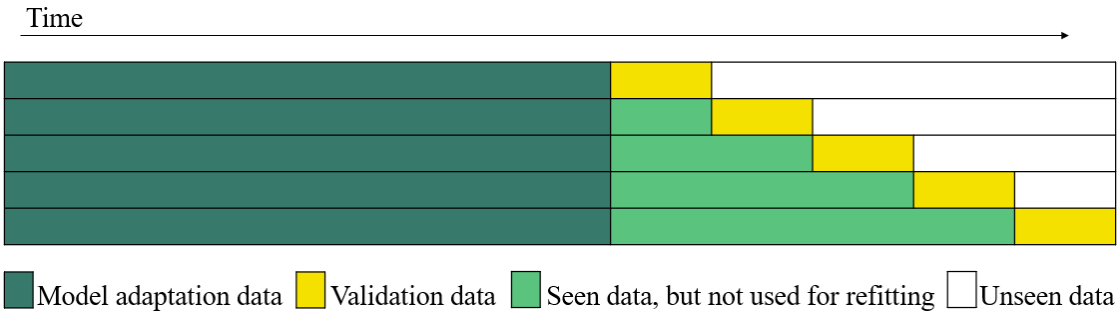


Figure 17: An illustration of the concept backtesting without refitting when more data become available.

The MAPE was first calculated on the entire validation dataset of 6 months, to get an overall indication of the most accurate model. With this approach one single value was generated for each model, which made it easy to distinguish the model with smallest MAPE. However, since the models should be accurate at forecasting 24 hours day-ahead, and not 6 months, the MAPE for every hour from 0-23 were relevant to examine. Based on all hours in the forecasted 6 months, the hourly MAPE was calculated and plotted. Thereby, the forecasting precision during each of the 24 hours in a day were presented. An hourly MAPE was only calculated for the models with smallest 6 month MAPE.

Apart from evaluating the ARIMAX and SARIMAX models with MAPE, the linear models' validity was investigated in terms of the analysis of residuals, to examine if these had white noise properties. The residual analysis consisted of four plots generated by the Python function *plot_diagnostics()*. If the plots showed that residuals resembled white noise the model was considered valid. The four plots were as follows.

1. Standardised residuals over time – showing patterns, which were used for checking if variance was constant over time.
2. Histogram of standardised residuals – showing normality.
3. Normal Q-Q plot with normal reference line – showing normality.
4. Correlogram (also known as ACF plot) – showing autocorrelation.

5. Modelling Procedure & Results

The process of finding the most accurate model structure of each model type (ARIMAX, SARIMAX, FNN, LSTM) is described in this section. First the ARIMAX model is addressed, followed by the SARIMAX model and the two NNs (FNN and LSTM). Additionally a sensitivity test of the most accurate model structures, as well as a comparison with the forecasting tool used at Fortum, are also presented. Lastly, a final result is presented. Based on the models' MAPE, sensitivity test and comparison with Fortum's forecasts, it is concluded which model that is most accurate at forecasting DAL grid losses and which model that is best for VML grid losses.

5.1 ARIMAX

The linear ARIMAX model was explored first, as it is the most simple model type out of the tested models. To construct and fit linear models to data in Python, the *statsmodel* module in Python was utilised, specifically the class *statsmodels.tsa.statespace.-sarimax.SARIMAX*. When fitting both ARIMAX and SARIMAX models to data, three input variables were needed:

1. grid losses from the model fitting dataset,
2. model orders (na, d, nc) respectively $(na, d, nc, NA, D, NC, m)$ and
3. exogenous variables from the model fitting dataset.

Note that this Python class does not enable one to choose the order of the $B(q)$ polynomial, the polynomial is just equal to 1. This means that n number of lags of the exogenous variables $u(t-n)$ could not be used as input. Only the value $u(t)$ could be used at each time instant t. However, this was not an issue as the losses did not correlate to lagged exogenous variables. In Appendix A cross-correlation plots show that lagged values of the exogenous variables did not correlate more to grid losses in comparison to current values.

To find out which ARIMAX model that gave most accurate predictions, different model orders of na and nc were tested based on PACF and ACF plots (seen in Figure 18 and Figure 19). For both the DAL and VML models na and nc were set to intervals and all combinations of the na and nc values in the chosen intervals were tested. The lower bound of the intervals were 0 for both na and nc in both DAL and VML models. The upper interval of na was set to 2 and the upper limit of nc could take high values for both the DAL and VML models. Considering that one should not make models more complex than necessary, the upper limit of the nc-interval was set to 4 for both DAL and VML models. The d was set to 0 for both grid loss models, since DAL and VML grid loss data were stationary, as described in Section 3.2.2. Table 4 summarises the chosen intervals of the model orders. In all combinations a total of 15 ARIMAX models were explored for each grid.

Apart from testing polynomial orders based on PACF and ACF plots, the decision was complemented by a statistical analysis. The z-test was used to investigate if all parameters in the na and nc polynomials were statistically significant. Polynomial parameters with a p-value less than 0.05 were accepted and parameters with p-values greater than 0.05 were rejected. Thereafter, the most accurate model was refitted on the entire dataset, to find the beforehand unknown parameters in the θ vector.

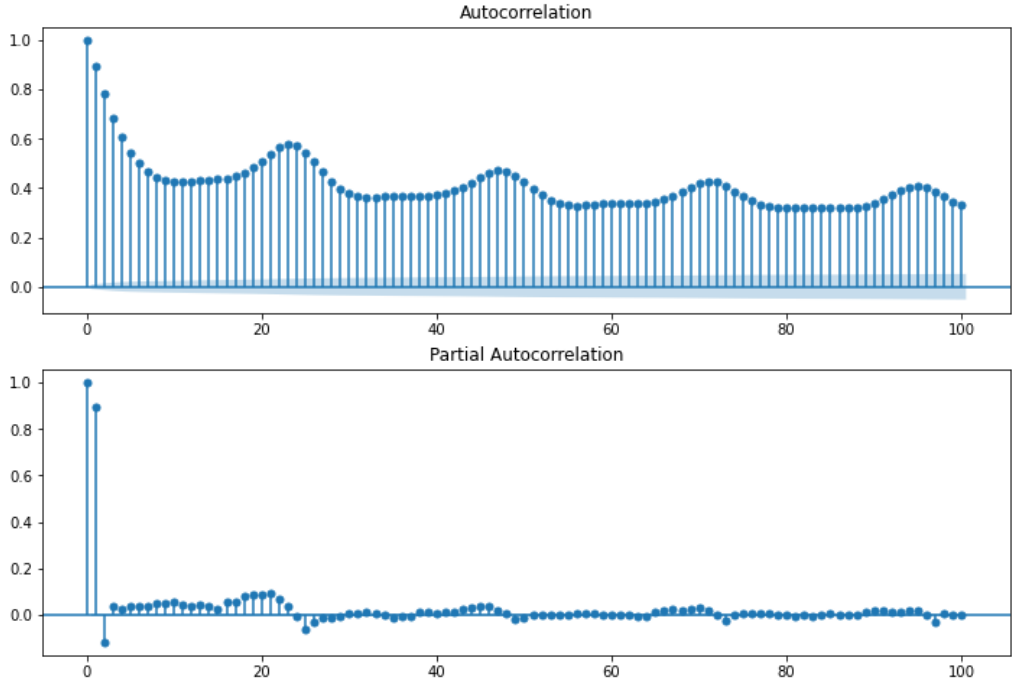


Figure 18: Top graph: ACF plot of DAL grid losses. Used for deciding highest relevant nc to test. At least the first 4 lags were relevant to test. Bottom graph: PACF plot for DAL grid losses. Used for deciding highest relevant na to test. No more than $na = 2$ was relevant to test.

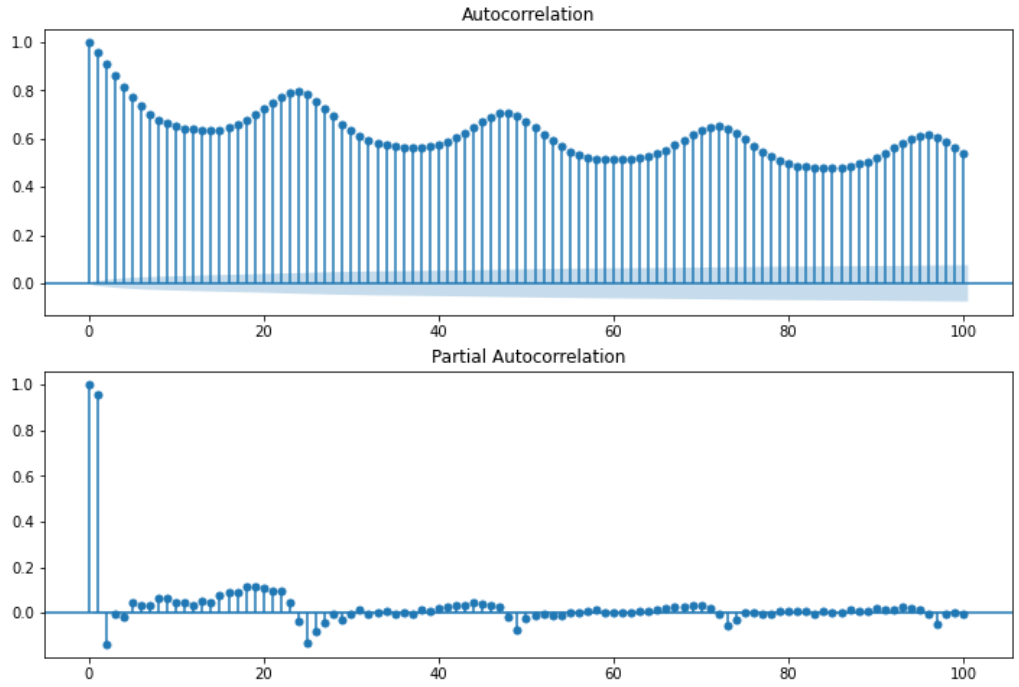


Figure 19: Top graph: ACF plot of VML grid losses. Used for deciding highest relevant nc to test. At least the first 4 lags are relevant. Bottom graph: PACF plot for VML grid losses. Used for deciding highest relevant na to test. No more than $na = 2$ was relevant to test.

Table 4: Intervals for hyperparameters of adapted and validated ARIMAX models.

Hyperparameters	DAL	VML
$\boldsymbol{na}$	[0,1,2]	[0,1,2]
$\boldsymbol{d}$	[0]	[0]
$\boldsymbol{nc}$	[0,1,2,3,4]	[0,1,2,3,4]
Number of ARIMAX	15	15

All MAPE values for the 15 tested ARIMAX models, for both grids, are presented in Appendix B. Following two sections presents the most accurate ARIMAX model found for the DAL and VML grids. (A summary of the most accurate ARIMAX model for each grid is presented in Appendix C, including results from the z-statistics and estimated parameters in the beforehand unknown θ vector.)

5.1.1 Most Accurate ARIMAX - DAL

The best ARIMAX was ARIMAX(2,0,2), with MAPE=0.1808. From the z-statistics it was found that all p-values were less than 0.05, meaning that the (2,0,2) polynomial orders were significant.

The residual analysis presented in Figure 20 shows that there are some yearly behaviour in the system that is not captured by the model. This is seen as the periodical pattern

with 6 fluctuations (6 years of data) in the top left graph. The histogram and Q-Q plot (top right and bottom left graphs) show that residuals are not exactly following, but are quite close to a normal distribution. The autocorrelation plot (bottom right graph) shows that residuals have no autocorrelation with lagged values, which is desired. Hence, it was considered that residuals had a sufficient resemblance to white noise and the ARIMAX(2,0,2) was acknowledged as the best structure. When fitting the model on all available data, the parameters in the θ vector were found and as an equation the model was described as

$$y(t) = [1.1568q^{-1} - 0.2127q^{-2}]y(t) + [0.0155 \quad 0.0149 \quad 0.0299 \quad 0.0121 \quad 0.0111]\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \\ u_5(t) \end{bmatrix} + [1 \quad -0.4473q^{-1} - 0.0734q^{-2}]e(t),$$

where $y(t)$ is grid losses, $[u_1(t), u_2(t), u_3(t), u_4(t), u_5(t)]$ is [TRÄ, GRÅ, MFJ, SE2, SE3] and $e(t)$ is the prediction error. To illustrate the model, the fitted model and forecasted losses is presented in Figure 21. To enable visualisation of the model, five days of historical data and seven forecasted days are plotted. The seven days is the first week in the 6 month validation dataset. To widen the illustrated result, two plots of seven other forecasted days are presented in Appendix D. (This portion of visualised data holds for all presented models.)

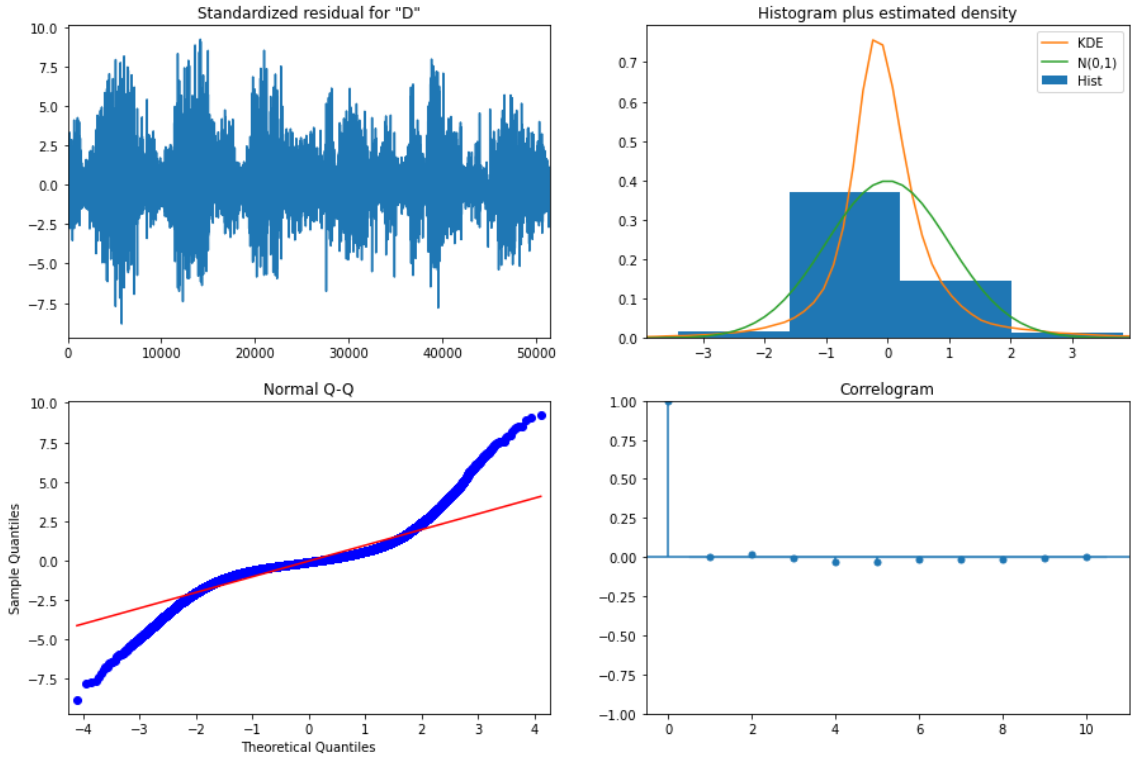


Figure 20: Residual analysis of ARIMAX(2,0,2) when modelling DAL losses.

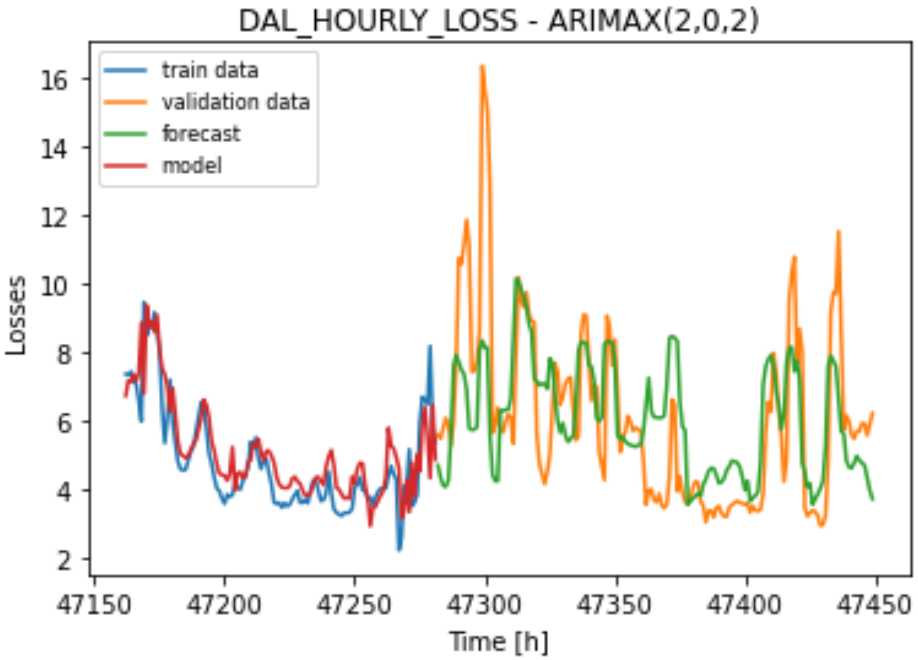


Figure 21: Plot of forecasted DAL losses with ARIMAX(2,0,2).

5.1.2 Most Accurate ARIMAX - VML

The model with smallest MAPE was ARIMAX(1,0,4) with MAPE=0.1839. However, the 4:th order q-polynomial was not significant. Instead, the model with smallest MAPE and significant polynomial orders was ARIMAX(1,0,0), with MAPE=0.1840.

In Figure 22 the residual analysis show that all four plots except for the histogram and Q-Q plot resembles properties of white noise. No other model of the tested ARIMAX models had a smaller MAPE and showed better results on the residual analysis. Therefore, this model was chosen as the best one, even though the residuals did not follow a normal distribution closely. Thus, ARIMAX(1,0,0) was accepted as the most accurate model structure and the model was described as

$$y(t) = 0.9559q^{-1}y(t) + \begin{bmatrix}0.0149 & 0.0211 & 0.0099\end{bmatrix}\begin{bmatrix}u_1(t)\\u_2(t)\\u_3(t)\end{bmatrix},$$

where $y(t)$ is grid losses and $[u_1(t), u_2(t), u_3(t)]$ is [PROD_TOT, SE2, SE3]. To illustrate the model, the fitted model and the forecasted losses is presented in Figure 23.

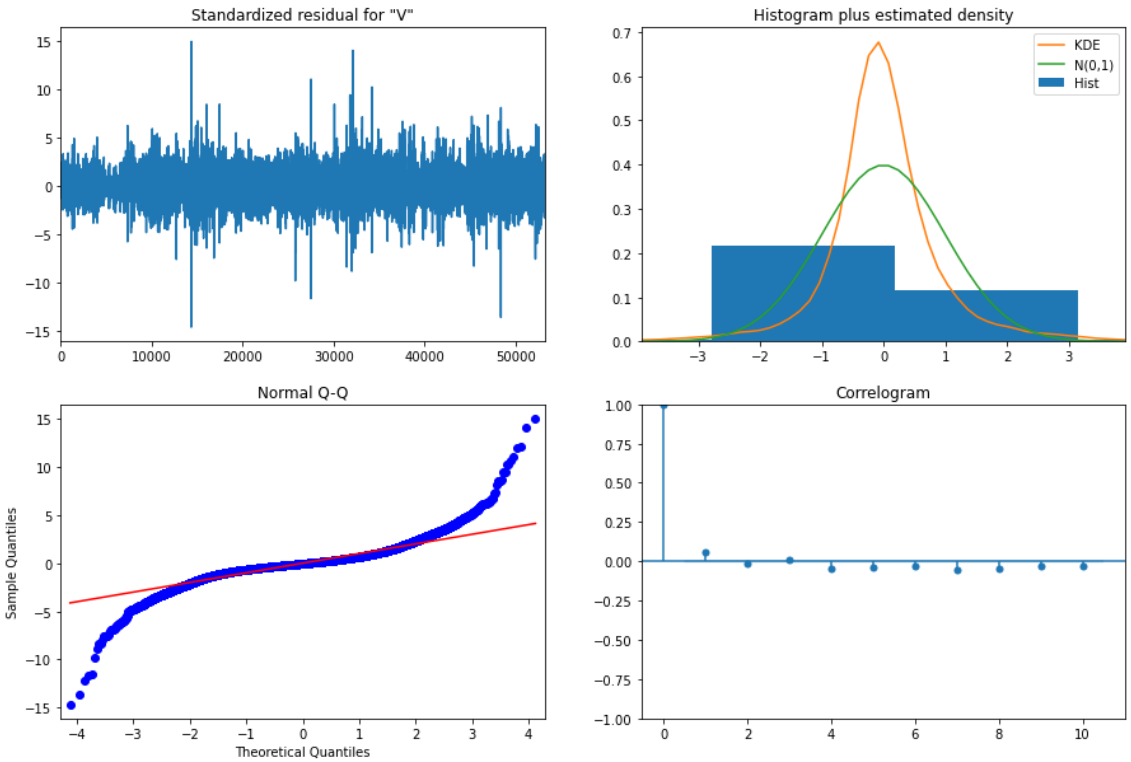


Figure 22: Residual analysis of ARIMAX(1,0,0) when modelling VML losses.

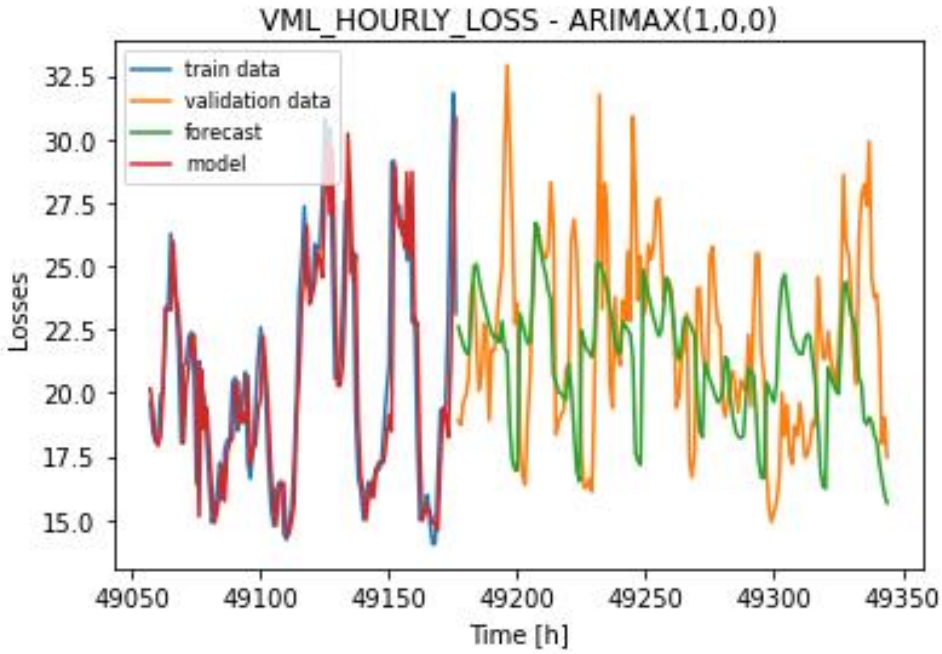


Figure 23: Plot of forecasted VML losses with ARIMAX(1,0,0).

5.2 SARIMAX

Regarding the 7 parameter SARIMAX model, fewer models than the ARIMAX were explored due to the model being more computationally demanding. The na and nc were set to the na and nc orders found from the best ARIMAX model. For the polynomial orders for DAL $na = nc = 2$ and for VML $na = 1$ and $nc = 0$. The differentiating terms d and D were set to 0 and NA and NC were either 0 or 1. No more than order 1 was tested for the seasonal polynomials, because the models became time consuming to fit to data and validate for higher seasonal polynomial orders. As the computational demand and time was a constraint for this project, simplifications was a necessity. Thus, a total of three SARIMAX models were tested for each grid. For all SARIMAX models the seasonal term m was set to 24 as data had a 24 hour periodic behaviour, mentioned in Section 3.1. Table 5 summarises the tested SARIMAX models.

Additionally, the decision of polynomial orders was complemented by a statistical analysis. The z-test was used to investigate if all parameters in the na, nc, NA and NC polynomials were statistically significant to describe the forecasted losses or not.

Table 5: Intervals for hyperparameters of adapted and validated SARIMAX models.

Hyperparameters	DAL	VML
na	[2]	[1]
d	[0]	[0]
nc	[2]	[0]
NA	[0,1]	[0,1]
D	[0]	[0]
NC	[0,1]	[0,1]
m	[24]	[24]
Number of SARIMAX	3	3

The MAPE values for the 3 tested SARIMAX models, for both grids, are presented in Appendix B. Following two sections presents the most accurate SARIMAX model found for the DAL and VML grids. (A summary of the most accurate SARIMAX model for each grid is presented in Appendix C, including results from the z-statistics and estimated parameters in the beforehand unknown θ vector.)

5.2.1 Most Accurate SARIMAX - DAL

The best SARIMAX was SARIMAX(2,0,2)(1,0,1)[24], with MAPE=0.1927. All polynomial parameters were statistical significant. The residual analysis in Figure 24 show notable similarities to the residual analysis of the ARIMAX model. A yearly pattern is not captured by the model, but a resemblance of a normal distribution and no autocorrelation is satisfactory. Thus, SARIMAX(2,0,2)(1,0,1)[24] was accepted as the best SARIMAX structure. The model is described by

$$
\begin{aligned}
&y(t)\\
&= [1.1332q^{-1} \quad -0.1976q^{-2} \quad 0.7047q^{-24} \quad 0.7906q^{-25} \quad -0.1392q^{-26}]y(t)\\
&+ [0.0156 \quad 0.0095 \quad 0.0324 \quad 0.0081 \quad 0.0118]\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \\ u_5(t) \end{bmatrix}\\
&+ [1 \quad -0.4618q^{-1} \quad -0.0857q^{-2} \quad -0.5891q^{-24} \quad 0.2720q^{-25} \quad 0.0504q^{-26}]e(t),
\end{aligned}
$$

where $y(t)$ is grid losses, $[u_1(t), u_2(t), u_3(t), u_4(t), u_5(t)]$ is [TRÄ, GRÅ, MFJ, SE2, SE3] and $e(t)$ is the prediction error. The fitted model and forecasted losses is presented in Figure 25.

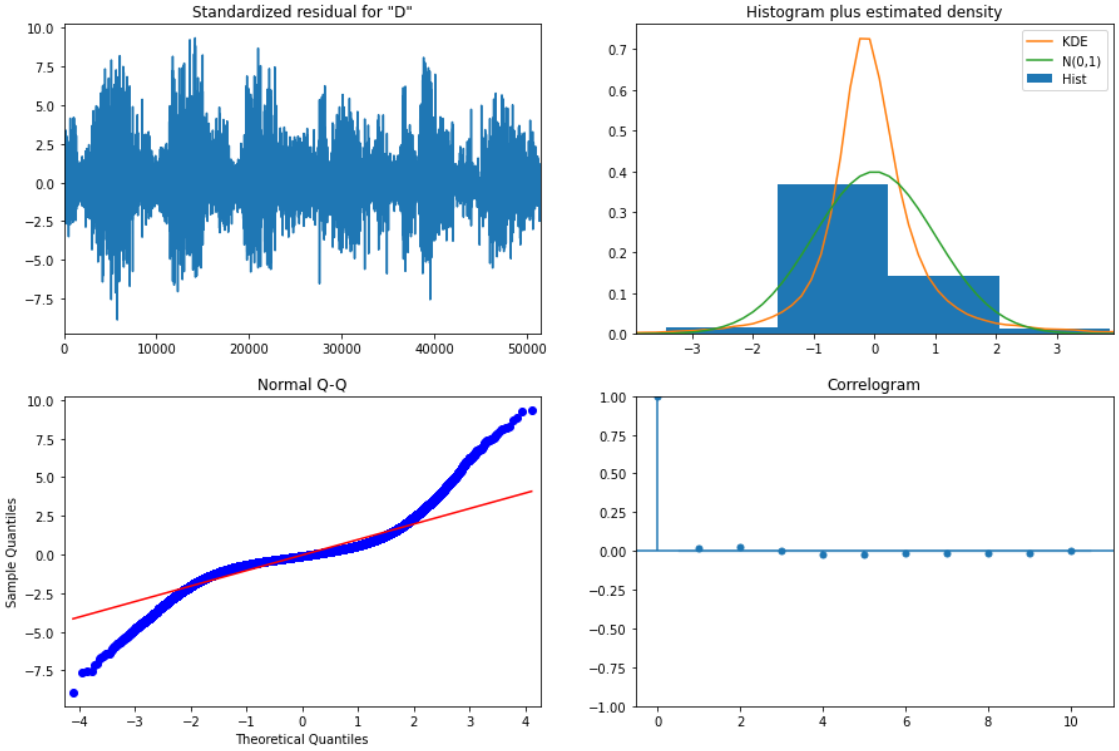


Figure 24: Residual analysis of SARIMAX(2,0,2)(1,0,1)[24] when modelling DAL losses.

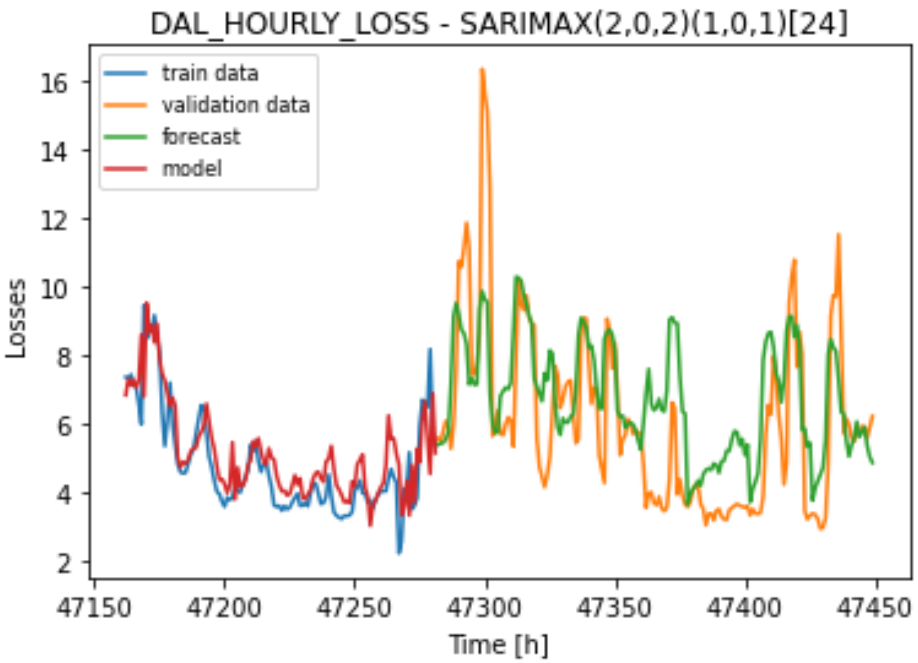


Figure 25: Plot of forecasted DAL losses with SARIMAX(2,0,2)(1,0,1)[24].

5.2.2 Most Accurate SARIMAX - VML

It was found that SARIMAX(1,0,0)(0,0,1)[24] had smallest MAPE, MAPE=0.1736, and all parameters were significant. The residual analysis in Figure 26 show that the variance and autocorrelation follow the properties of white noise. As the histogram and Q-Q plot show, residuals are not exactly following a normal distribution. Nevertheless, the distribution resembled a normal distribution to a slightly larger extent compared to the ARIMAX model and was therefore considered valid. The model was described by

$$y(t) = [0.9467q^{-1}]y(t) + [0.0.0136 \quad 0.0186 \quad 0.0092]\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} + [1 \quad 0.1652q^{-24}]e(t),$$

where $y(t)$ is grid losses, $[u_1(t), u_2(t), u_3(t)]$ is [PROD_TOT, SE2, SE3] and $e(t)$ is the prediction error. The fitted model and forecasted losses is presented in Figure 27.

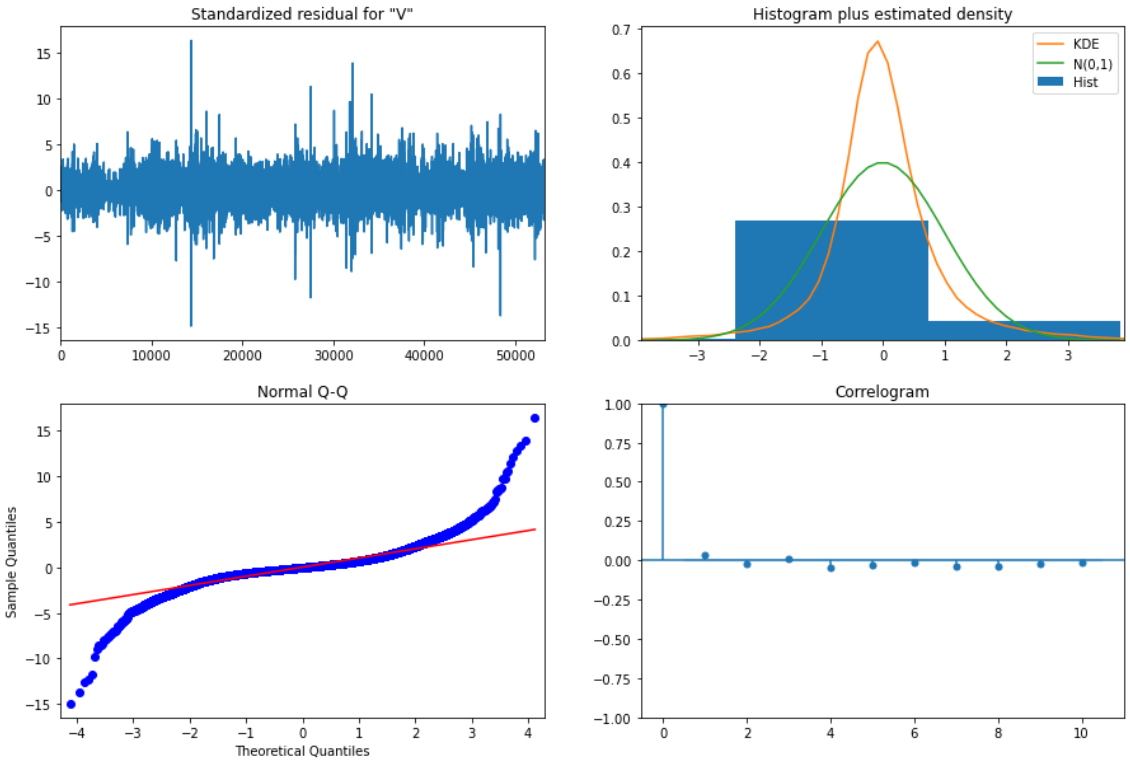


Figure 26: Residual analysis of SARIMAX(1,0,0)(1,0,1)[24] when modelling VML losses.

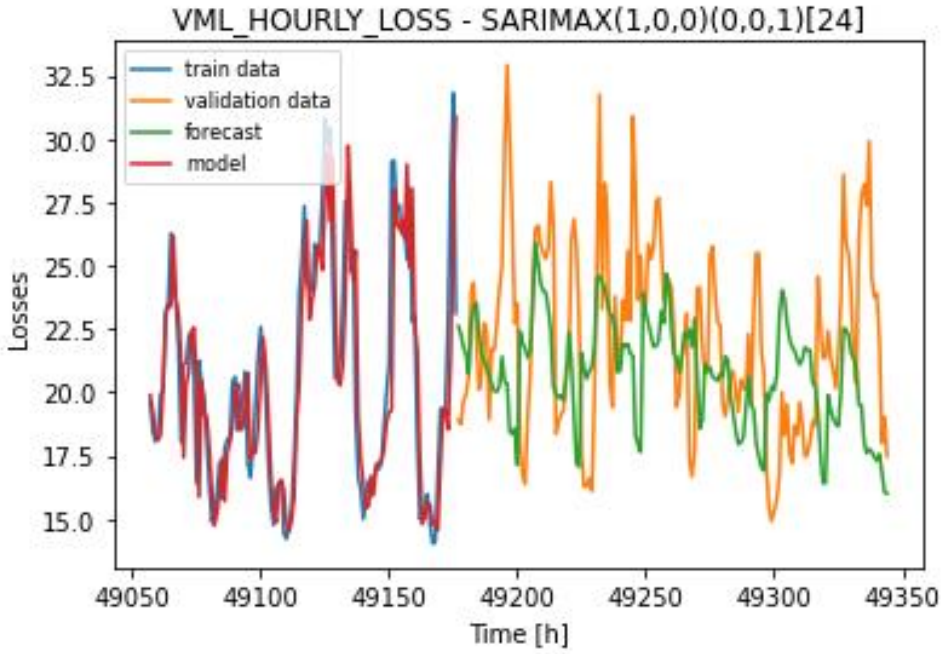


Figure 27: Plot of forecasted VML losses with SARIMAX(1,0,0)(0,0,1)[24].

5.3 Neural Networks

After testing linear models, the neural networks FNN and LSTM models were explored. When using NNs for time series forecasting, both input and output variables were scaled to take values between 0 and 1 before the NNs were adapted to data. Since all data were transformed by normalisation it was not necessary to remove the mean values from the data, as it was done for the linear models.

To build NNs, the activation functions, batch size, number of epochs, number of layers and neurons were chosen. Activation functions used for hidden layers were sigmoid, ReLu and tanh. According to a study by Brownlee, different activation functions are used depending on the type of NN [57]. FNNs should have ReLu as activation function for hidden layers and RNNs should have sigmoid or tanh. Thereby, ReLu was chosen for FNNs and both sigmoid and tanh were tested for LSTM models. Activation function for the output layer was determined based on the purpose of the NN. In this case of time series forecasting it is suggested by Brownlee that the activation function should be linear for the output layer [57]. Thus, a linear activation function was chosen for both FNNs' and LSTMs' output layer.

Common batch sizes that has been used in previous studies are 32, 64 or 128 samples and 100, 500, 1000 or a larger number of epochs are commonly tested [58]. Therefore, when training NNs, batch sizes of 32, 64 and 128 samples and 100, 500 and 1000 number of epochs were tested. There is no systematic approach to tune the number of layers and neurons, and therefore the testing included some randomness. Considering losses correlated most to losses at one lag and to exogenous variables at lag 0, grid losses at $t-1$ and exogenous variables at time t was chosen as input. That is, the NNs were trained to predict grid losses one timestep ahead (at time t) based on grid losses at $t-1$ and exogenous variables at time t.

To forecast a 24 hour period, the NNs forecasted one hour at a time in a for-loop. Thereby, all forecasts except for the first, were based on a forecasted grid loss value at $t-1$, calculated in the previous iteration in the for-loop. The models found to have the smallest difference between modelled losses and actual losses (smallest MAPE) were considered the best FNN respectively LSTM model. Approximately 50 FNNs and 50 LSTMs were explored by varying the number of epochs, batch sizes, number of layers, nodes and input variables. Following two sections presents the most accurate FNN and LSTM models for the DAL and VML grids.

5.3.1 Most Accurate FNN & LSTM - DAL

The model with smallest MAPE was a 3 layered FNN with 35 neurons in each layer. It was found that a number of 100 epochs and a batch size of 32 samples gave the most accurate forecasts, with MAPE=0.2010. The trained model and forecasted values is plotted in Figure 28.

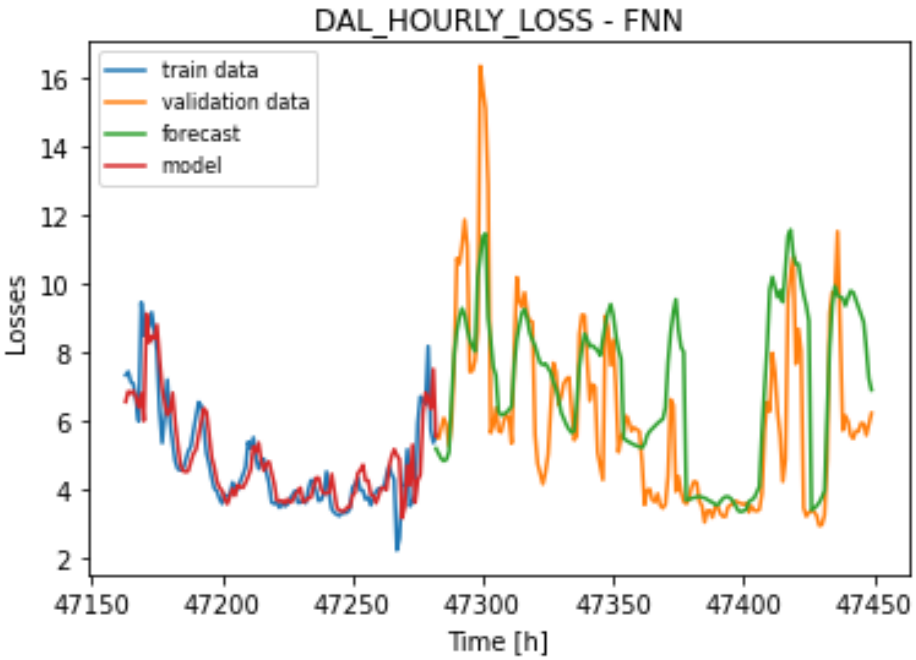


Figure 28: Plot of forecasted DAL losses with the best FNN model found.

Regarding the LSTM, it was found that the LSTM with smallest MAPE was a 4 layered network with 10 neurons in the first two layers and 15 neurons in the last two layers, 100 number of epochs, a batch size of 32 and tanh as activation function. MAPE was 0.1886. The trained model and forecasted values is presented in Figure 29.

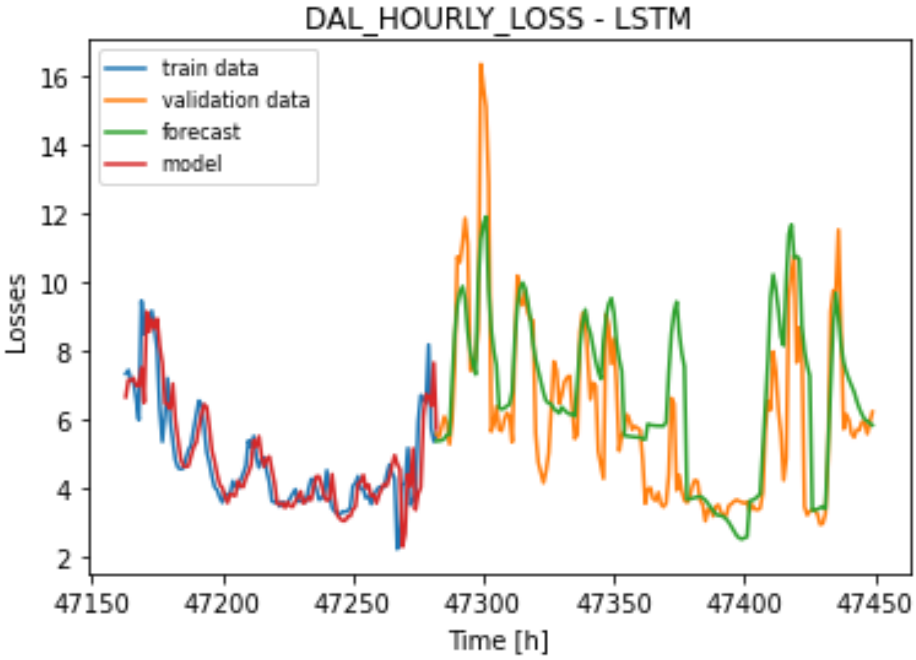


Figure 29: Plot of forecasted DAL losses with the best LSTM model found.

5.3.2 Most Accurate FNN & LSTM - VML

The model with smallest MAPE was a 3 layered FNN with 10 neurons in each layer. It was found that a number of 100 epochs and a batch size of 64 samples gave the most accurate results, which was MAPE=0.2328. The trained model and forecasted losses is presented in Figure 30.

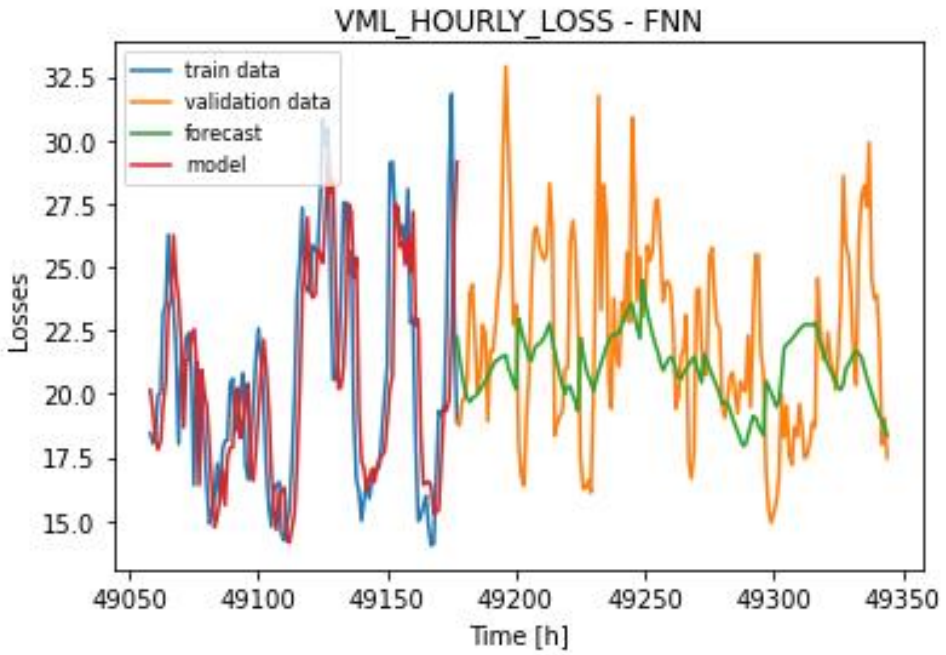


Figure 30: Plot of forecasted VML losses with the best FNN model found.

The best LSTM model was a 4 layered network with 5 neurons in each layer. It was found that 100 number of epochs, a batch size of 32 and sigmoid as activation function gave the LSTM model with smallest MAPE, MAPE=0.2202. The trained model and forecasted values from the LSTM is presented in Figure 31.

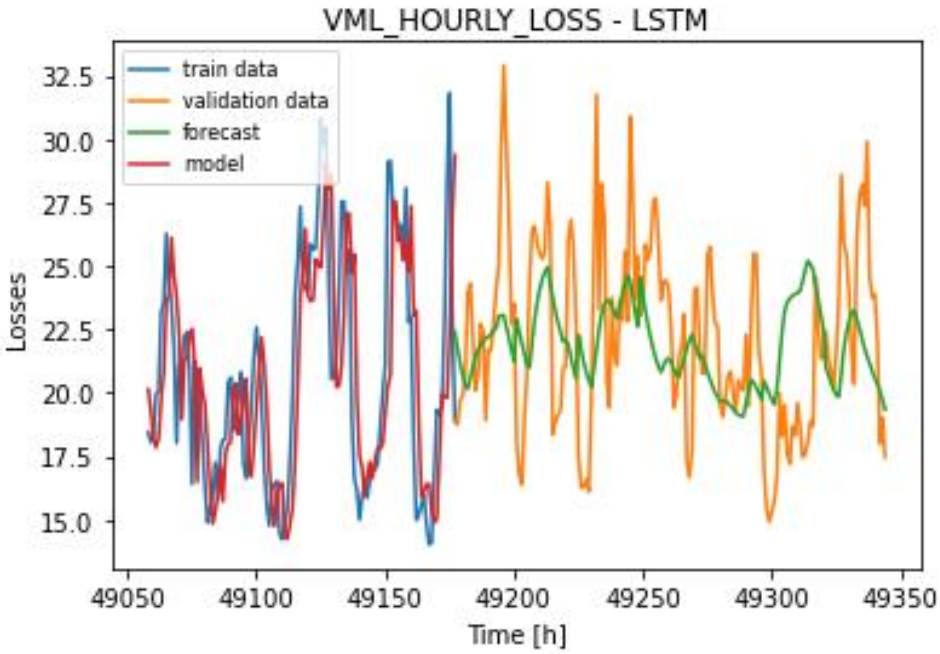


Figure 31: Plot of forecasted VML losses with the best LSTM model found.

5.4 Summary of ARIMAX, SARIMAX, FNN & LSTM - MAPE

In Table 6, the MAPE for all four tested models are presented for the DAL and VML grids. The MAPE being presented is calculated on the entire validation dataset of ½ year of data. For the DAL grid, ARIMAX had smallest MAPE, meaning that it is the most accurate model, and the FNN had highest MAPE, making it the least accurate model. Regarding the VML grid, SARIMAX had smallest MAPE and the FNN had highest. Hence, the SARIMAX model was most accurate and the FNN was least accurate at forecasting VML grid losses. However, it was not only smallest MAPE that was of relevance when investigating which model that was most useful for day-ahead grid loss forecasting. Broadening the model evaluation by exploring how sensitive the models were to changes in input was also of great importance.

Table 6: The MAPE for the most accurate models found in the four tested model types.

Model	**MAPE validation**	
	DAL	**VML**
ARIMAX	0.1808	0.1840
SARIMAX	0.1927	0.1736
FNN	0.2010	0.2328
LSTM	0.1886	0.2202

5.5 Testing Sensitivity

An essential part when evaluating the models was to test the models' sensitivity to changes in input. In the training dataset, spot prices and weather conditions were historical measured values. However, only forecasted prices and weather conditions are available when predicting grid losses in reality, meaning that the input variables may differ compared to the measured values. Nonetheless, inaccurate input should, for a robust model, not affect the grid loss forecasting accuracy to a large extent. Testing robustness was done by first varying spot prices and weather conditions ±10 %, ±15 % and ±20 %. Thereafter, the models robustness was tested by changing input ±100 % and ±200 %, to investigate how the models would react to outliers or erroneous data. Consequently, by investigating robustness a more thoroughly result was obtained of the models' actual usefulness, since it simulated forecasted spot prices and weather conditions. It was not necessary to vary production in this way since production data already consisted of planned production.

5.5.1 Sensitivity Test - DAL

From the sensitivity test it was found that all four models were robust when changing SE2 and SE3 with ±10 %, ±15 % and ±20 %. Considering all models the change in MAPE was between 1.2-6.4 %, which is smaller than the 10-20 % change in input. When changing SE2 and SE3 with ±100 % and ±200 %, all models change in MAPE was between 4.5-109 %. Most robust model was the FNN, with an average of 2.3 % change in MAPE when varying input 10-20 % and an average of 18% when changing input 100-200 %. Least robust model was SARIMAX, with an average of 4.6 % change in MAPE when varying input 10-20 % and 49 % change in MAPE when varying input 100-200 %. All models' MAPE and percental changes when varying the input is summarised in Table 7 and Table 8.

Table 7: Summary of the sensitivity test for all DAL models. The MAPE is presented and in parenthesis is the percental change in MAPE for DAL models when varying SE2 and SE3 ±10, 15 and 20 %.

Model DAL	**MAPE (change in %)**						
	Validation	**-10 %**	**+10 %**	**-15 %**	**+15 %**	**-20 %**	**+20 %**
ARIMAX	0.1808	0.1760 (2.7)	0.1863 (3.0)	0.1739 (3.8)	0.1892 (4.6)	0.1720 (4.9)	0.1923 (6.4)
SARIMAX	0.1927	0.1868 (3.1)	0.1997 (3.6)	0.1842 (4.4)	0.2033 (5.5)	0.1818 (5.7)	0.2072 (5.2)
FNN	0.2010	0.1976 (1.7)	0.2034 (1.2)	0.1958 (2.6)	0.2050 (2.0)	0.1942 (3.4)	0.2067 (2.8)
LSTM	0.1886	0.1852 (1.8)	0.1938 (2.6)	0.1848 (2.0)	0.1967 (4.3)	0.1853 (1.7)	0.1998 (5.9)

Table 8: Summary of the sensitivity test for all DAL models. The MAPE is presented and in parenthesis is the percental change in MAPE for DAL models when varying SE2 and SE3 ±100 and 200 %.

Model DAL	MAPE (change in %)				
	Validation	**-100 %**	**+100 %**	**-200 %**	**+200 %**
ARIMAX	0.1808	0.1726 (4.5)	0.2547 (40)	0.2393 (32)	0.3537 (96)
SARIMAX	0.1927	0.1835 (5.0)	0.2841 (47)	0.2622 (36)	0.4032 (109)
FNN	0.2010	0.1796 (11)	0.2403 (20)	0.2201 (9.5)	0.2644 (32)
LSTM	0.1886	0.2232 (18)	0.2565 (36)	0.3387 (80)	0.3070 (63)

5.5.2 Sensitivity Test - VML

From the sensitivity test it was found that all four models were robust when changing SE2 and SE3 with ±10 %, ±15 % and ±20 %. Considering all models, the absolute change in MAPE was 0.2-3.6 %, which is considerably smaller than the 10-20 % change in input. When changing SE2 and SE3 with ±100 % and ±200 %, all models change in MAPE was between 0.9-49 %. Most robust model was SARIMAX, with a mean of 0.3 % change in MAPE when varying input 10-20 % and an average of 13 % when changing input 100-200 %. Least robust model was the LSTM, with a mean of 2.1 % change in MAPE when varying input 10-20 % and an average of 25 % when changing input 100-200 %. All models' MAPE and percental changes when varying the input is summarised in Table 9 and Table 10.

Table 9: Summary of the sensitivity test for all VML models. The MAPE is presented and in parenthesis is the percental change in MAPE for VML models when varying SE2 and SE3 ±10, 15 and 20 %.

Model VML	MAPE (change in %)						
	Validation	**-10 %**	**+10 %**	**-15 %**	**+15 %**	**-20 %**	**+20 %**
ARIMAX	0.1840	0.1836 (0.2)	0.1853 (0.7)	0.1835 (0.3)	0.1860 (1.2)	0.1835 (0.3)	0.1867 (1.5)
SARIMAX	0.1736	0.1737 (0.1)	0.1742 (0.3)	0.1737 (0.1)	0.1745 (0.5)	0.1738 (0.1)	0.1749 (0.7)
FNN	0.2328	0.2297 (1.3)	0.2350 (0.9)	0.2280 (2.1)	0.2361 (1.4)	0.2263 (2.8)	0.2371 (1.8)
LSTM	0.2202	0.2192 (0.5)	0.2250 (2.2)	0.2247 (2.0)	0.2266 (2.9)	0.2167 (1.6)	0.2282 (3.6)

Table 10: Summary of the sensitivity test for all VML models. The MAPE is presented and in parenthesis is the percental change in MAPE for VML models when varying SE2 and SE3 ±100 and 200 %.

Model VML	MAPE (change in %)				
	Validation	**-100 %**	**+100 %**	**-200 %**	**+200 %**
ARIMAX	0.1840	0.1930 (4.9)	0.2063 (12)	0.2210 (20)	0.2453 (33)
SARIMAX	0.1736	0.1808 (4.1)	0.1873 (7.9)	0.2011 (16)	0.2153 (24)
FNN	0.2328	0.2248 (3.4)	0.2623 (13)	0.2486 (6.8)	0.3132 (35)
LSTM	0.2202	0.2221 (0.9)	0.2558 (16)	0.3280 (49)	0.2907 (32)

5.6 Comparing Models with Fortum

The models in this project were compared to the forecasting model used at Fortum. It was investigated if suggested models performed better – in terms of smaller MAPE – compared to the forecasting technique that is in use at the moment of writing this thesis. The hourly MAPE graphs were compared, since the forecasting performance over 24 hours was the most relevant accuracy to examine.

At Fortum, grid loss forecasts are calculated with the forecasting tool Aiolos from the software provider Vitec. Aiolos generates weekly forecasts for the day ahead every day, being one day ahead forecast used for the daily bids. In principle, this is the same forecasting horizon (1 day ahead) as for the models in this thesis. Therefore, comparison of the models was possible. However, Aiolos use weather data, calendar data and historical grid losses as model inputs, which are less variables used as input compared to this thesis. Further, the model structure used by Aiolos was not known, which made a truly objective comparison of the models impossible. Nonetheless, the comparison could indicate and suggest if the tested models with predictors could perform better than the modelling method and predictors currently used by Aiolos.

5.6.1 Comparison with Aiolos - DAL

The hourly MAPE for all four models and Aiolos is plotted and presented in Figure 32. Note that the Aiolos graph is not following the same behaviour as the tested models. This is a result of Aiolos being a different model, using another method for grid loss forecasting (for a complete model description the reader my refer to Vitec's Aiolos documentation). All four tested models have an upward going trend over the 24 hours, meaning that the first forecasted hour is most accurate and the last is the least accurate. This is the expected pattern since all models use previous hourly losses as input and have been validated by backtesting – only using real measured values when forecasting the first hour and previous forecasted hours when forecasting the 23 coming hours.

Hence, the model comparison cannot solidly state if the tested models are better or worse than Aiolos based on the hourly MAPE graph. However, graphs lower than the Aiolos graph indicate that tested models accuracy could be better and graphs above the Aiolos graph indicate that tested models could be worse than Aiolos. Likewise, this holds for the hourly MAPE comparison when modelling VML losses.

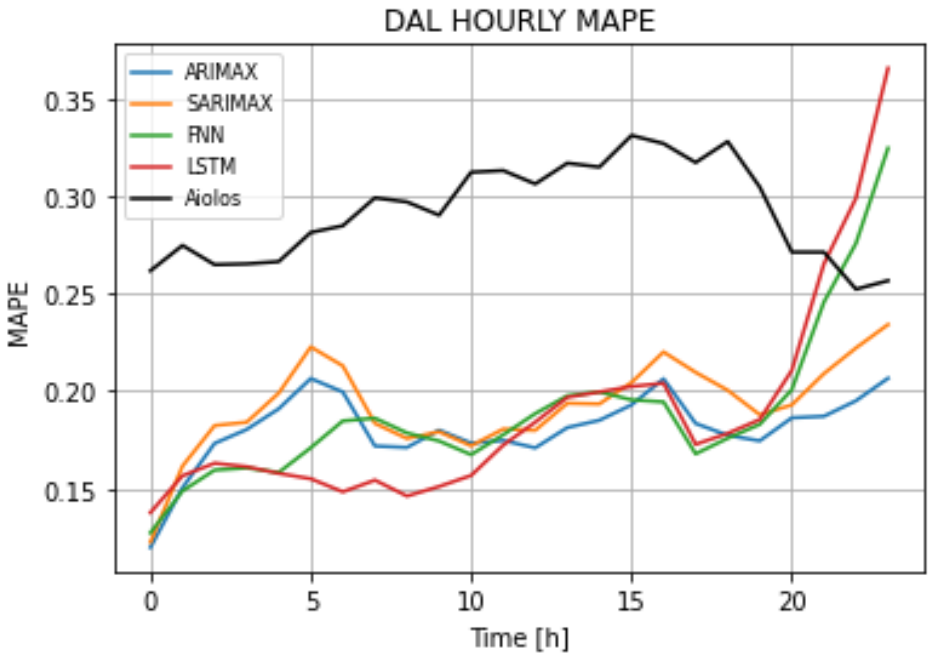


Figure 32: Model comparison with hourly MAPE when forecasting DAL losses.

As seen in Figure 32 all tested models have comparable hourly MAPEs until forecasted hour 20, seen as overlapping graphs starting below 0.15 and increasing to around 0.2. At hour 20 and onward to hour 24 all models MAPE increase, but for the NNs the increase is substantially higher. In other words, ARIMAX and SARIMAX are more stable at a 24 hour prediction compared to the FNN and LSTM which are more unstable. Compared to Aiolos, there is a clear indication that ARIMAX and SARIMAX could be better than Aiolos and the FNN and LSTM might be better.

5.6.2 Comparison with Aiolos - VML

The hourly MAPE for all four models and Aiolos is plotted and presented in Figure 33. SARIMAX had smallest hourly MAPE of all tested models and the FNN had highest. However, both ARIMAX and SARIMAX generated notably more stable forecasts compared to the NNs. Unstable forecasts by the NNs is seen as the rapid increase of the MAPE for the FNN and LSTM at hour 20-24. Compared to Aiolos, there is only a clear indication of ARIMAX and SARIMAX being more accurate than Aiolos. As a result of the FNN and LSTM being close to, and occasionally overlapping, the Aiolos curve there is no clear indication of them being more adequate than Aiolos.

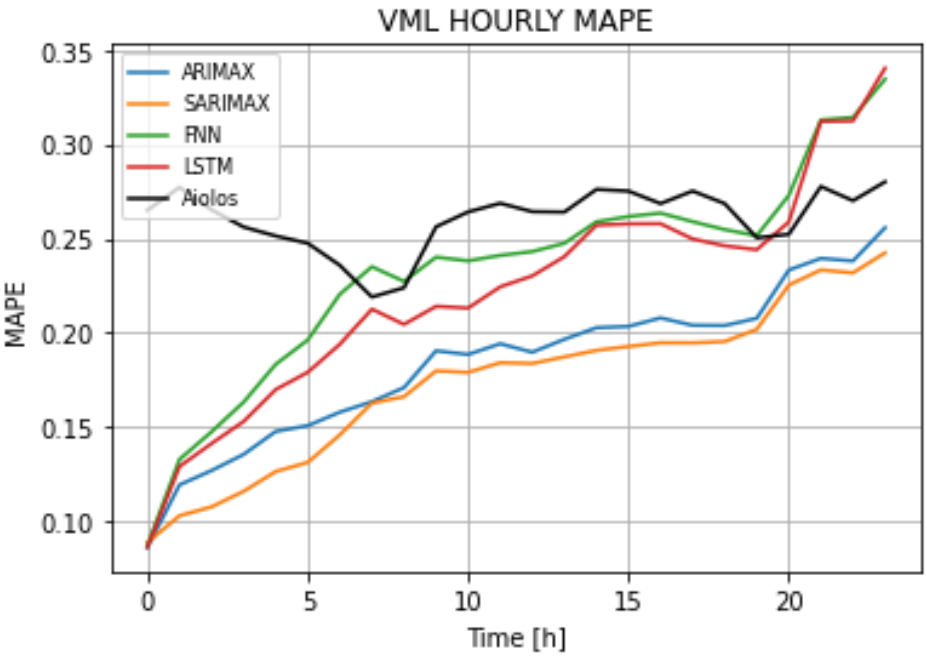


Figure 33: Model comparison with hourly MAPE when forecasting VML losses.

5.7 Total Results - Best Models for DAL and VML

By summing up the results from the calculated MAPE, sensitivity test and Aiolos comparison, the most suitable model for forecasting grid losses day-ahead in the DAL and VML grids were acknowledge.

5.7.1 Dalarna

Considering that ARIMAX had the smallest MAPE on the entire validation dataset, was robust to changes in input, and had stable and smallest hourly MAPE, ARIMAX(2,0,2) was considered to be the most accurate model for day-ahead grid loss forecasting of DAL losses. Compared to Aiolos the hourly MAPE was well below Aiolo's during all 24 hours, indicating that more accurate day-ahead forecasts are very likely to be achieved with the ARIMAX model, and by implementing the predictors, spot prices, power production plans and previous grid losses, when forecasting DAL losses day-ahead.

5.7.2 Värmland

Considering that SARIMAX had the smallest MAPE on the entire validation dataset, was most robust to changes in input, had the smallest hourly MAPE and a stable hourly MAPE, SARIMAX(1,0,0)(0,0,1)[24] was chosen as the most accurate model for day-ahead grid loss forecasting of VML losses. Compared to Aiolos SARIMAX's hourly MAPE was below Aiolo's during all 24 hours, indicating that more accurate day-ahead forecasts could be achieved with the SARIMAX model and predictors chosen in this project.

6. Discussion

The discussion is structured by addressing predictors, model structures and model validation in three separate sections. Each section relates to one of the three aims.

6.1 Identified Predictors

Aim 1. Identify predictors that are of importance when forecasting grid losses.

As grid losses are the difference of injected and withdrawn power from an electrical grid, predictors for modelling losses ought to relate to injected and withdrawn power. It was found that withdrawn power, in terms of consumption, was not the most crucial predictor to generate grid loss forecasts in neither the DAL nor the VML grid. This was found as temperature and weekday were not significant predictors. Additionally, when including them as input, higher MAPE was obtained compared to when excluding them from the input. Thus, consumption as an exogenous variable was not needed, as was hypothesised prior the modelling phase of the project.

On the other hand, it was found that injected power in terms of production and spot prices were significant predictors for losses in both grids. Spot prices in SE3 had higher correlation for both grids, which is reasonable considering both grids are geographically located in SE3. However, a distinct difference between the DAL and VML losses was that production had a much higher correlation with DAL losses compared to VML losses. This means that power production was a better predictor for DAL losses, than it was for VML losses. Considering that the VML grid is characterised as a grid that transmits power, with injected power mostly consisting of imported power and not produced power, it is reasonable that production was not an equally applicable predictor for VML losses as it was for DAL losses. However, how could imported and exported power, as predictors, contribute to forecast the VML losses?

As the residual analysis shows, there is no pattern in residuals for the VML regression models. Residuals follow a normal distribution quite accurately and they have no autocorrelation. Consequently, there is nothing major the models cannot capture, indicating that nothing seems to be absent in the VML models. Thus, imported and exported power as input would possibly not be needed to generate more adequate forecasts of VML losses. On the other hand, seeing to the residual analysis of the DAL losses, there is a clear yearly pattern in the residuals. This shows that something in the DAL losses is not entirely captured by the tested models. Another predictor, with yearly periodicity, might be needed as input to consider this yearly behaviour in the DAL losses.

In conclusion, the results show that differences between electrical grid types affect which predictors are more suitable, and which are not, when modelling grid losses. Therefore, testing many different variables and evaluate their contribution to describe losses is a necessary part of a grid loss forecasting process. This has proven to be extra

important since not all beforehand hypothesised predictors were statistically significant to describe grid losses in this project. In more general terms this means that an analysis of possible predictors should be conducted for every new studied grid. Predictors found to be important to model grid losses in one grid does not imply that the same predictors are suitable for another grid. Nonetheless, as power production, spot prices and grid losses at previous timesteps were significant predictors for the different grids, DAL and VML, these three variables could preferably be tested for future grid loss forecasting in other grids as well.

Additionally, finding suitable predictors is important to avoid overfitting of models and thereby erroneous results. As this project has shown, more adequate predictors tends to lead to more accurate predictions. However, at one stage more variables will not improve the forecasting accuracy to a large extent, as the dataset is descriptive enough to define the system. Hence, one of the challenges when modelling systems lies in the process of deciding when enough data has been gathered.

6.2 Model Structures

Aim 2. Distinguish suitable model structures for grid loss forecasting.

The most accurate model at forecasting DAL losses day-ahead was ARIMAX(2,0,2) and the best model at forecasting VML losses was SARIMAX(1,0,0)(0,0,1)[24]. By considering the properties of the electrical grids and features of the ARIMAX and SARIMAX model structures, these are not surprising results. This thesis period of 24 hours meant that losses 24 hours ago were included in the SARIMAX models. The autocorrelation for DAL losses at lag 24 was approximately 0.55, seen in the ACF plot. Compared to other exogenous variables, this was not the highest correlation. For instance, production in TRÄ had the highest correlation of 0.7, suggesting that 24 hour lagged losses should not need to be included to obtain better forecasts. Thus, the result of ARIMAX being a better model structure for DAL losses, compared to the SARIMAX model, is highly reasonable.

On the contrary, when forecasting VML losses a SARIMAX model performed best. VML losses at lag 24 had a high autocorrelation, around 0.8 as seen in the ACF plot. Thus, it is plausible that inclusion of losses at 24 hour lag would result in better models compared to models not including losses 24 hours earlier. Furthermore, all other exogenous variables for VML losses had notably lower correlation than 0.8. PROD_TOT had highest correlation of all exogenous variables and it was 0.4, meaning that lagged losses was the most important variable to describe current losses. Hence, the result of SARIMAX being a better model structure for VML losses, compared to the ARIMAX model, is reasonable.

For both grids the NNs were approximately equally good as the ARIMAX and SARIMAX models. Although, the NNs did not outperform the linear models when comparing the MAPE. Considering previous studies that found NNs to be exceptionally

more adequate models, this is a slightly deviating result. One possible explanation comes from the fact that there is no systematic testing of NNs. NNs found to be most accurate in this project does probably not have optimised hyperparameters, meaning that their performances could most likely have been better if more time had been available for hyperparameter tuning.

That is, the model structure that produce most accurate forecasts depend on the type of grid losses that is being modelled. Therefore, distinguishing adequate model types and model structures is an equally important part of modelling as identifying suitable predictors.

6.3 Model Validation

Aim 3. Validate tested models and compare the forecasting accuracy with the forecasting tool used at Fortum.

For the DAL losses, ARIMAX had the smallest MAPE over the entire 6 months and was most stable at forecasting all 24 hours day-ahead. However, the sensitivity test showed that the FNN was most robust to changes in input and the LSTM second most robust, not ARIMAX or SARIMAX. Furthermore, the hourly MAPE graph for the NNs were similar or below ARIMAX's during forecasted hours 0-20. This suggests that the NNs' hyperparameters possibly were quite adequately tuned for grid loss forecasting, even if they were not optimised.

The exception for this suggestion is the rapid increase seen in the last 4 hours in the hourly MAPE, which moreover ought to have worsen the total 6 month MAPE. Nonetheless, the 6 month MAPE for the NNs were close to the linear models' MAPE. In fact, the LSTM had lower MAPE than SARIMAX. Hence, it is arguable that the NNs were quite accurate since their total MAPE, despite the rapid increase in hourly MAPE, were close to the accuracy of the ARIMAX and SARIMAX models.

Regarding the VML losses, SARIMAX was the model with smallest MAPE. SARIMAX was also most robust, had lowest hourly MAPE graph and was most stable. The fact that SARIMAX had smallest MAPE has been discussed in terms of the model's structure, addressed in the previous Section 6.2. The SARIMAX model being most robust might be explained by the relevantly added 24 hour lagged grid losses. With more grid losses as input to the model, it is possible that the extra variables made the model more robust to changes in two of the other variables.

Intuitively, this reasoning should hold for the DAL losses as well, but it does not. The NNs were most robust when forecasting DAL losses, and SARIMAX was least robust. One possible explanation could be the fact that SARIMAX was not an equally adequate model for DAL losses, as it was for VML losses. Another potential explanation is that better NNs were found for DAL losses, than for VML losses. The MAPE for the NNs were in the same range as the linear models' MAPE, both in terms of the entire 6 month

MAPE and by looking at the hourly MAPE graph from hour 0-20. Further, the NNs were the most robust models when modelling DAL losses.

The NNs for the VML losses had higher 6 month MAPE compared to the linear models, they were the least robust models and had hourly MAPE graphs well above the linear models. Both NNs' hourly MAPE graphs were grouped on a higher level and the linear models followed each other at a lower level. This is a distinct difference compared to the NNs for the DAL losses, which suggests that NN-hyperparameters for DAL losses were more adequate than the hyperparameters found for the models describing VML losses.

As of the MAPE graphs, they showed that all models were most accurate at forecasting few hours ahead compared to forecasting many hours ahead. This is reasonable due to the forecasts being based on previous forecasts. However, there was a rapid increase in MAPE during the last 4 hours, especially for the two NNs. This pattern was seen both for the DAL and VML models. A suggestion is that hours 20-24 could be much more difficult to forecast. Alternately, the forecasting error could have been aggregated for the last 4 forecasted hours, due to the forecasts being based on previous forecasts.

By comparing the DAL and VML models' accuracy, VML models had generally slightly lower MAPE and were marginally more robust. However, note that plotted forecasts with VML models did not follow the validation dataset as accurate as DAL models (seen in the plots illustrating the models with seven forecasted days in Sections 5.1, 5.2 & 5.3). By solely studying the plots of seven days forecasted losses, considerably more accurate models seems to have been obtained for DAL losses. This result is counterintuitive to the MAPE results. One explanation for this contradiction could simply be that the particular seven plotted days happened to be easier to model for all DAL models and more difficult for all VML models. However, this coincidence is highly unlikely since forecasted DAL losses also followed actual losses at other time periods (seen in Appendix D).

A more probable explanation is that the yearly behaviour in DAL losses lowered the accuracy in terms of a higher MAPE, since the DAL models could not entirely consider that behaviour. Nonetheless, the non-modelled yearly behaviour in DAL losses was presumably not noticeable on the daily basis accuracy, illustrated in the plots.

6.3.1 Aiolos Comparison

The model comparison with Aiolos indicated that more accurate grid loss forecast could be obtained with the suggested models in this thesis. However, model structure as a reason for this result cannot fully be analysed, since the model structure of Aiolos was unknown.

Another contributing factor, which can be discussed, originate from the fact that different variables were used as input to the models. Forecasts made by Aiolos used

weather predictions and historical grid losses as input. As found from the correlation analysis, weather variables did not have as high correlation to losses as power production and spot prices. Since the suggested models in this project included production and spot prices as input, it is very likely that production and spot prices as exogenous variable was needed to generate more accurate forecasts.

By observing the hourly MAPE graphs of the DAL and VML models against each other, DAL models were lowered from Aiolos's graph to a greater extend compared to the VML models. Consequently, it is expected that most improvement in grid loss forecasting accuracy at Fortum comes from implementing the DAL models, rather than the VML models, in place of Aiolos's forecasts. However, the cause of this result cannot be definitely stated. Either, Aiolos generated better forecasts for VML losses and worse for DAL losses, making it reasonable that the most accurate models found in this project would have greatest improvement for DAL losses. Alternatively, Aiolos generated equivalently accurate forecasts for both grids, and instead, the greater improvement in DAL losses originated from the fact that DAL losses were easier to model with the chosen predictors and model structures, compared to VML losses.

7. Conclusions

To conclude the findings from this thesis, three conclusions are stated below, each one answering one of the project's three aims (presented in Section 1.2).

Conclusion 1. – important predictors

Predictors identified as important exogenous variables to include when forecasting grid losses were power production and spot prices. These predictors were important for the grid in Dalarna and for the grid in Värmland. Furthermore, grid losses at previous time steps were also found to be important to include as input to the models.

Conclusion 2. – suitable model structures

The model structure found to be most accurate at forecasting grid losses day-ahead in a grid in Dalarna, was ARIMAX(2,0,2). The model found to be the most accurate at forecasting grid losses in a grid in Värmland, was SARIMAX(1,0,0)(0,0,1)[24].

Conclusion 3. – compared forecasting accuracy

The results indicate that more accurate forecasts can be achieved with the two models found in this thesis, compared to the forecasting tool that is in use at Fortum at the time of writing this thesis. Most likely is that the ARIMAX(2,0,2) will improve forecasting accuracy in Dalarna more than SARIMAX(1,0,0)(0,0,1)[24] will in Värmland, if implemented and used instead of the currently used forecasting method at Fortum.

By putting this thesis findings into a broader perspective of grid loss forecasting, the following can be concluded. As the two studied grids in this thesis had different characteristics, but still had the same predictors, the results suggest that power production, spot prices and grid losses at previous time steps are of interest to consider as possibly important predictors for other grid losses as well. However, the level of importance of the predictors were slightly different for the two grids, which implies that a data analysis, to find suitable predictors, always is necessary before modelling and forecasting grid losses. That is, power production, spot prices and grid losses at previous time steps are likely to be important predictors for other grids as well, considering that those predictors were important for the two different grids studied in this thesis. However, there is no guarantee that those predictors will be important for all other grids, which makes a data analysis crucial to find important predictors for other grids.

Furthermore, different model structures were found as the best model at forecasting grid losses for the two grids, which is a result of the grid losses relating to the characteristics of the grids. In accordance to previous studies, which have studied different types of grids, different models and predictors have been found to be the most accurate model for grid loss forecasting. Consequently, when forecasting grid losses it is essential to explore different predictors and model structures to find an adequate model for the

particular grid losses being studied. No universal model is adequate at forecasting all types of grid losses. Hence, a throughout analysis, concerning the particular grid and grid losses, is of great necessity to find a valid and useful model for all day-ahead grid loss forecasting.

Nevertheless, both linear models and NNs performed better than the forecasting method that the models were compared to in this project. This means that these two model types could be of interest to consider when forecasting grid losses day-ahead in other grids as well. However, it cannot be concluded that suggested models in this project definitely can improve forecasting accuracy for other grids. This is because potential improvement in forecasting accuracy depends on the method that the models are compared to. Forecasting methods may be substantially different between different electrical companies, conducting day-ahead grid loss forecasts. Therefore, it cannot be concluded if the models explored in this thesis could increase, or decrease, forecasting accuracy in other grids in general.

7.1 Future Work

For future work, backtesting with refitting could be tested when validating the models. In this project, backtesting without refitting was implemented due to the method being faster. However, by refitting, models are adapted to data every time more data become available and new predictions are generated. This is a more real-world-like scenario, refitting models every now and again, instead of using the same model for 6 months (as in this thesis's validation phase). The models' forecasting accuracy after refitting would be an interesting result to compare to Aiolos's forecasts, and to the forecasts generated with this thesis's models.

Another proposal regarding validation, originates from current operations at Fortum. At Fortum, data is quarantined for about a week when doing forecasts, meaning that there is one week of data less in the training dataset. This was not directly tested. As all weeks in the validation dataset was forecasted based on the same training dataset, it was not completely applicable to regard the one week of missing training data. However, if backtesting with refitting is tested, one week less data in the training dataset can be implemented in future work to evaluate accuracy.

A suggestion for future work regarding NNs, consists of testing more NNs to optimise hyperparameters. It would be of great interest to examine the forecasting accuracy of the NNs if the hyperparameters are more adequately configurated before training the models.

Another possibility would be to explore hybrid models, instead of testing one model at a time as done in this thesis. Testing many different hybrid models and investigate their day-ahead forecasting accuracy is of significance, since research work in the literature suggest that hybrid models can be more accurate at time series forecasting compared to using one model.

The final suggestion is to test other variables as predictors. Especially relevant is to identify a variable with yearly behaviour in the grid in Dalarna. Due to a yearly pattern in the residual analysis, tested models and predictors for losses in the Dalarna grid could not capture all yearly behaviour in the grid losses. Therefore, a highly attractive research question constitutes of investigating how another, additional exogenous variable affects the MAPE, as the finding possibly could lower the DAL MAPE even further.

Acknowledgements

A first and foremost big thank you to the outstanding Gabriela Peña, my supervisor at Fortum, for guiding me throughout the project and making this thesis possible. Also, a big thank you to Per Mattsson, supervisor at Uppsala University, for assisting with in-depth knowledge of mathematical modelling. I would also like to give my expression of gratitude to Hans Bjerhag, thesis coordinator at Fortum, for being a great support in many aspects and for the immense devotion to all thesis students at Fortum. Furthermore, a sincere thanks to Henrik Wahlström, for much appreciated opposition.

Last but not least, thank you to all Fortum employees that I have been in contact with, for sharing your knowledge and giving me valuable insights to the field of grid loss forecasting. It has been a pleasure working with everyone involved in this project.

References

[1] Svk, (2022), Kontrollrummet. Available online: https://www.svk.se/om-kraftsystemet/kontrollrummet/ (2022-01-24)

[2] Energiföretagen Sverige, (2022), Svensk Elmarknadshandbok, Edition 22A, Version 2022-04-01 Available online. https://www.elmarknadshandboken.se-/Dokumentation/Texter/NEMHB.pdf (2022-04-05).

[3] Svk. (2021), Om elmarknaden. Available online. https://www.svk.se/om-kraftsystemet/om-elmarknaden/ (2022-02-10).

[4] Svk. (2021), Översikt av kraftsystemet. Available online. https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/ (2022-02-10).

[5] Fried, D. (2021), Elpriset 2021 – faktorer som påverkar. Energimarknadsinspektionen. Available online. https://www.ei.se/konsument/el/sa-har-fungerar-elmarknaden/elpris?sv.target=12.66185b2817dc26036872e83&sv.12.66185b2817dc26036872e83.route=/settings (2022-02-10).

[6] Nord Pool. (2020), Trading. Available online. https://www.nordpoolgroup.com/trading/ (2022-01-25).

[7] Svk. (2021), Operations and Electricity Markets. Available online. https://www.svk.se/en/national-grid/operations-and-electricity-markets/ (2022-01-25).

[8] Schavemaker, P., Van der Sluis, L. (2008), Electrical power system essentials. John Wiley & Sons Ltd: Chichester. Page 212.

[9] Nord Pool. (2020), Intraday trading. Available online. https://www.nordpoolgroup.com/trading/intraday-trading/ (2022-01-25).

[10] Nord Pool. (2020), Day ahead market. Available online: https://www.nordpoolgroup.com/the-power-market/Day-ahead-market/ (2022-02-01).

[11] Khodadadi, A., Herre, L., Shinde, P., Eriksson, R., Söder, L., Amelin, M. (2020), Nordic Balancing Markets: Overview of Market Rules. DOI: 10.1109/EEM49802.2020.9221992

[12] Svk. (2021), Balansmarknaden. Available online. https://www.svk.se/om-kraftsystemet/om-elmarknaden/balansmarknaden/ (2022-01-27).

[13] Manito, A. R. A., Bezerra, U. H., Soares, T. M., Vieira, J. P. A., Nunes, M. V. A., Tostes, M. E. L., & de Oliveira, R. C. (2019), Technical and non-technical losses calculation in distribution grids using a defined equivalent operational impedance. *IET Generation, Transmission & Distribution*, Vol. 13, No. 8, 1315-1323. DOI: 10.1049/iet-gtd.2018.5334

[14] Wang, W., Cai, D., Liu, H., Zhou, C., Rao, Y., Cao, K., Chen, Q., Yu, D., Zhang, S. (2021), Analysis of Technical Losses in Hubei Power Grid. IEEE. DOI: 10.1109/CIEEC50170.2021.9510450

[15] Wu, A., Ni, B. (2016), Line loss analysis and calculation of electric power systems, 2nd Edition. John Wiley & Sons Singapore Pte Ltd. Available online. https://go.exlibris.link/ndxbfMnW

[16] Sulakov, S. I. (2017), Forecasting software for hourly transmission losses. *2017 15th International Conference on Electrical Machines, Drives and Power Systems (ELMA)*, 110-114. DOI: 10.1109/ELMA.2017.7955412

[17] Dalal, N., Mølnå, M., Herrem, M., Røen, M., Gundersen, O. E. (2020), Day-Ahead Forecasting of Losses in the Distribution network. *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 34, No. 8, 13148-13155. DOI: 10.1609/aaai.v34i08.7018

[18] Tulensalo, J., Seppänen, J., Ilin, A. (2020), An LSTM model for power grid loss prediction. *Electric Power Systems Research*, Vol. 189, 106823. DOI: 10.1016/j.epsr.2020.106823

[19] Ljung, L., Glad, T. (2004), Modellbygge och simulering, 2nd Edition. Studentlitteratur AB: Lund.

[20] Wu, C. F. J., Hamada, M. S. (2021), Basic Concepts for Experimental Design and Introductory Regression Analysis. *Experiments - Planning, Analysis and Optimization.* 3rd Edition. John Wiley & Sons.

[21] Maçaira, P.M, Tavares Thomé, A. M., Cyriono Oliviera, F. L., Carvalho Ferrer, A. L. (2018), Time series analysis with explanatory variables: A systematic literature review. *Environmental Modelling & Software*, Vol. 107, 199-209. DOI: 10.1016/j.envsoft.2018.06.004

[22] Brockwell, P.J., Davis, R.A. (1991), Time Series: Theory and Methods. 2nd Edition. Springer-Verlag, New York.

[23] Nason, G. P. (2006), Stationary and non-stationary time series, Book chapter in Statistics in Volcanology, Mader H. M., Coles, S. G., Connor, C. B., Connor, L. J. 129-142, DOI: 10.1144/IAVCEI001.11

[24] Hydman, R. J., Khandakar, Y. (2008), Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, Vol. 27, No. 3. DOI: 10.18637/jss.v027.i03

[25] Auffarth, B. (2021), Machine Learning for Time-Series with Python: Forecast, Predict, and Detect Anomalies with State-Of-the-art Machine Learning Methods. Packt Publishing, Limited, Birmingham. Available online. https://go.exlibris.link/mDcTy8yS

[26] Perktold, J., Seabold, S., Taylor, J. (2022), statsmodels.tsa.statespace.sarimax.SARIMAX. Statsmodels. Available online.

https://www.statsmodels.org/devel/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html (2022-03-14).

[27] Brockwell, P.J., Davis, R.A. (1991), Time Series: Theory and Methods, Edition 2nd, Springer New York, New York, NY. DOI: 10.1007/978-1-4419-0320-4

[28] Perktold, J., Seabold, S., Taylor, J. (2022), SARIMAX: Introduction. Statsmodels. Available online. https://www.statsmodels.org/dev/examples/-notebooks/generated/statespace_sarimax_stata.html (2022-05-03).

[29] Perktold, J., Seabold, S., Taylor, J. (2022), statsmodels.tsa.statespace.sarimax.ARIMAX. Statsmodels. Available online. https://www.statsmodels.org/devel/generated/statsmodels.tsa.arima.model.ARIMA.html (2022-03-25).

[30] Dangeti, P. (2017), Statistics for Machine Learning, Packt Publishing, Limited, Birmingham. Available online. https://go.exlibris.link/CwDc4CK4

[31] Malekian, A., Chitsaz, N. (2021), Chapter 4 – Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. *Advances in Streamflow Forecasting*. Elsevier Inc. 115-147. DOI: 10.1016/B978-0-12-820673-7.00003-2

[32] Freedman, D.A. (2009), Statistical models: theory and practice. Cambridge University Press, New York. Available from: ProQuest Ebook Central. [2022]. DOI: 10.1017/CBO9780511815867

[33] Florea, A., Andonie, R. (2019), Weighted random search for hyperparameter optimization. *International Journal of Computers, Communications & Control*, Vol. 14, No. 2, 154-169. DOI: 10.15837/ijccc.2019.2.3514

[34] Purkait, N. (2019), Hands-On Neural Networks with Keras: Design and Create Neural Networks Using Deep Learning and Artificial Intelligence Principles, Packt Publishing. Available online. https://go.exlibris.link/TjjSlk31

[35] Tiwari, M. K., Deo, R. C., Adamowski, J. F. (2021), Chapter 10 – Short-term flood forecasting using artificial neural networks, extreme learning machines, and M5 model tree. *Advances in streamflow forecasting*. Elsevier Inc. 263-279. DOI: 10.1016/B978-0-12-820673-7.00012-3

[36] Nielsen, M. A. (2019), Chapter 1: Using neural nets to recognize handwritten digits. In *Neural Networks and Deep learning*, Determination Press, 2015. Available online. http://neuralnetworksanddeeplearning.com/chap1.html (2022-02-04).

[37] Nielsen, M. A. (2019), Chapter 6: Deep learning. In *Neural Networks and Deep learning*, Determination Press, 2015. Available online. http://neuralnetworksanddeeplearning.com/chap6.html (2022-02-04).

[38] Lei, C. (2021), RNN. In *Deep Learning and Practice with MindSpore*, Singapore: Springer Singapore, 83-93. DOI: 10.1007/978-981-16-2233-5_6

[39] Hewamalage, H., Bergmeir, C., Bandara, K. (2021), Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, Vol. 37, 388-427. DOI: 10.1016/j.ijforecast.2020.06.008

[40] Söderström, T., Stoica, P. (1989), System Identification. Prentice Hall. Cambridge: University Press.

[41] Mathworks. (2022), Simulate and Predict Identified Model Output. Available online. https://se.mathworks.com/help/ident/ug/definition-simulation-and-prediction.html (2022-05-07).

[42] Rydén, J. (2014), Stokastik för ingengörer. 1st Edition. Studentlitteratur AB: Lund. 19-23.

[43] Bouzerdoum, M., Mellit, A., Massi Pavan, A. (2013), A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, Vol. 98, 226-235. DOI: 10.1016/j.solener.2013.10.002

[44] QiheLou, QiLyu, Na, Z., Ma, D., Ma, X. (2020), Short-term electric power demand forecastingusing a hybrid model of SARIMA and SVR. *IOP Conference Series. Earth and Environmental Science*, Vol. 619, No. 1, 12035. DOI: 10.1088/1755-1315/619/1/012035

[45] Saelee, V., Sinsukudomachai, C., Khluabwannarat, P. (2012), Losses Prediction in the Distribution Transformer Using Hierarchy Neural Networks. *Journal of International Council on Electrical Engineering*, Vol. 2, No. 4, 384-390. DOI: 10.5370/JICEE.2012.2.4.384

[46] Lopez-Martin, M., Sanchez-Esguevillas, A., Hernandez-Callejo, L., Arribas, J. I., Carro, B. (2021), Additive ensemble neural network with constrained weighted quantile loss for probabilistic electric-load forecasting. *Sensors (Basel, Switzerland)*, Vol. 21, No. 9, 2979. DOI: 10.3390/s21092979

[47] Hong, T., Fan, S. (2016), Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, Vol. 32, No. 3, 914-938. DOI: 10.1016/j.ijforecast.2015.11.011

[48] Hossen, T., Nair, A. S., Chinnathambi, R. A., Ranganathan, P. (2018), Residential Load Forecasting Using Deep Neural Networks (DNN). *2018 North American Power Symposium (NAPS).* DOI: 10.1109/NAPS.2018.8600549

[49] Deb, C., Zhang, F., Yang, J., Lee, S. E., Shah, K. W. (2017), A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, Vol. 74, 902-924. DOI: 10.1016/j.rser.2017.02.085

[50] Gamboa, J. C. B. (2017), Deep Learning for Time-Series Analysis. *arXiv:1701.01887*. DOI: 10.48550/arXiv.1701.01887

[51] Carter, N. J., Schwertman, N. C., Kiser, T. L. (2009), A comparison of two boxplot methods for detecting univariate outliers which adjust for sample size

and asymmetry. *Statistical Methodology*, Vol. 6, No. 6, 604-621. DOI: 10.1016/j.stamet.2009.07.001

[52] Zhang, R., Guo, Z., Meng, Y., Wang, S., Li, S., Niu, R., Wang, Y., Guo, Q., Li, Y. (2021), Comparison of ARIMA and LSTM in Forecasting the Incidence of HFMD Combined and Uncombined with Exogenous Meteorological Variables in Ningbo, China, *International journal of environmental research and public health*, Vol. 18, No. 11, 6174. DOI: 10.3390/ijerph18116174

[53] Xu, D., Zhang, Q., Ding, Y., Huang, H. (2020), Application of a Hybrid ARIMA–SVR Model Based on the SPI for the Forecast of Drought—A Case Study in Henan Province, China, *Journal of applied meteorology and climatology*, Vol. 59, No. 7, 1239-1259. DOI: 10.1175/JAMC-D-19-0270.1

[54] Nokeri, T.C. (2021), Forecasting Using ARIMA, SARIMA, and the Additive Model in *Implementing Machine Learning for Finance*, Apress, Berkeley, CA, 21-50. DOI: 10.1007/978-1-4842-7110-0_2

[55] de Myttenaere, A., Golden, B., Le Grand, B., Rossi, F. (2016), Mean Absolute Percentage Error for regression models, *Neurocomputing (Amsterdam)*, Vol. 192, 38-48. DOI: 10.1016/j.neucom.2015.12.114

[56] Bravo, J. M., Coelho, E. (2019), Forecasting Subnational Demographic Data using Seasonal Time Series Methods, *CAPSI 2019 Proceedings.* 24. Available online: https://aisel.aisnet.org/capsi2019/24 (2022-05-18).

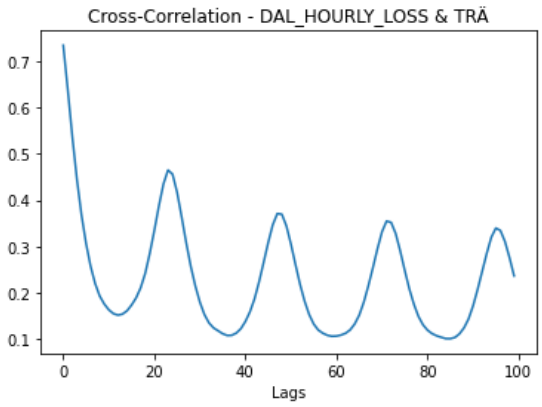
[57] Brownlee, J. (2022), How to Choose an Activation Function for Deep Learning. Machine Learning Mastery. Available online. https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/ (2022-03-28).

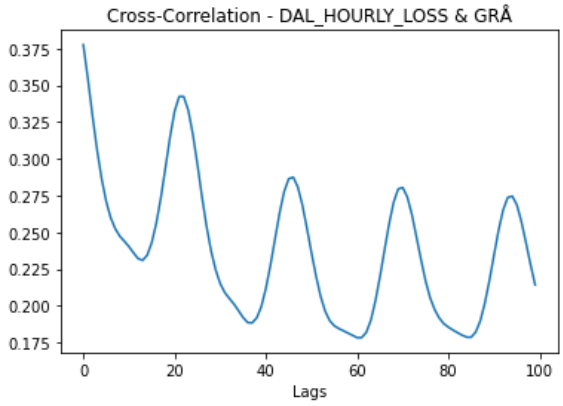
[58] Brownlee, J. (2019), Difference Between a Batch and an Epoch in a Neural Network. Machine Learning Mastery. Available online. https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/ (2022-03-28).

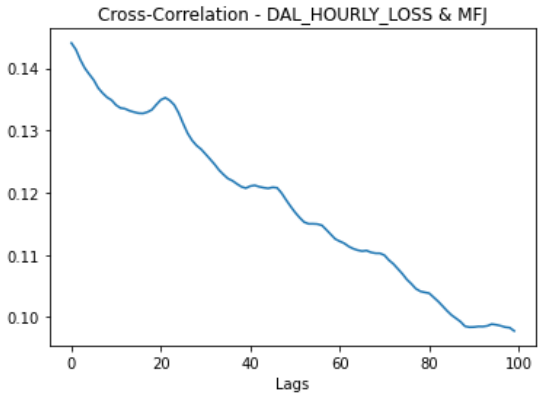
Appendix A

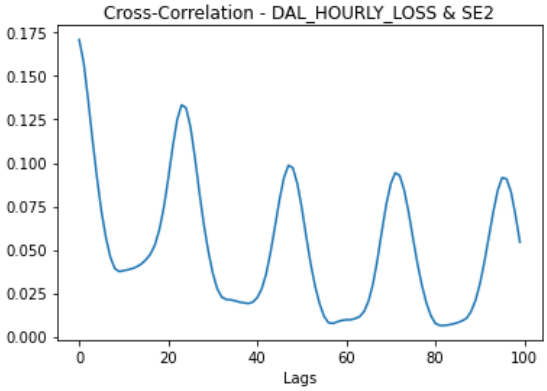
Cross-correlation plots Dalarna.

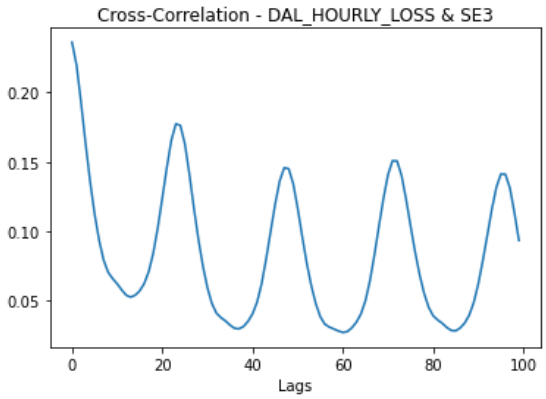
The 5 pots below show that historical values of exogenous variables u(t-n) does not correlate as much as the current value u(t), at lag 0. Hence, it was not necessary to include previous values of the exogenous variables and a B(q) polynomial equal to 1 was reasonable for DAL losses.





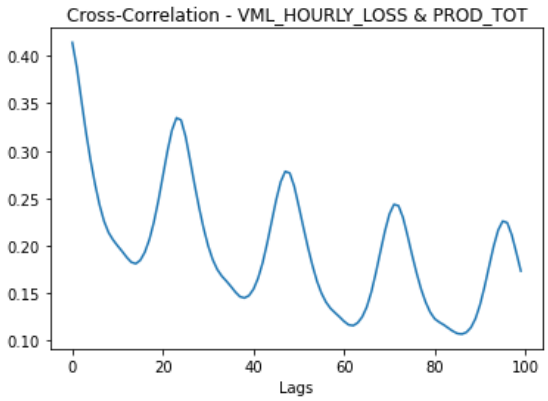


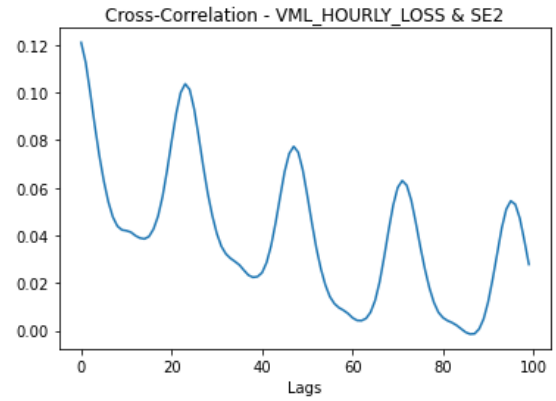


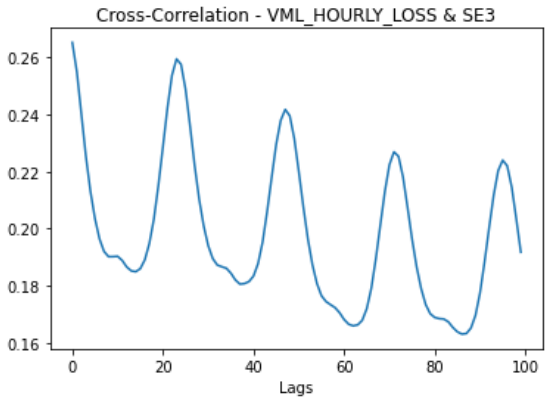


Cross-correlation plots Värmland.

The 12 pots below show that historical values of exogenous variables u(t-n) does not correlate as much as the current value u(t), at lag 0. Hence, it was not necessary to include previous values of the exogenous variables and a B(q) polynomial equal to 1 was reasonable for VML losses.







Appendix B

Tables present all tested ARIMAX and SARIMAX models.

Table 11 and Table 12 regard DAL losses.
Table 13 and Table 14 regard VML losses.

Table 11: ARIMAX models with TRÄ, GRÅ, MFJ, SE2 and SE3 as input. Model with smallest MAPE is bolded.

Model – DAL	**MAPE**
ARIMA(0,0,0)	0.2244
ARIMA(0,0,1)	0.2554
ARIMA(0,0,2)	0.2731
ARIMA(0,0,3)	0.2708
ARIMA(0,0,4)	0.2668
ARIMA(1,0,0)	0.2379
ARIMA(1,0,1)	0.2209
ARIMA(1,0,2)	0.2134
ARIMA(1,0,3)	0.2063
ARIMA(1,0,4)	0.1973
ARIMA(2,0,0)	0.2266
ARIMA(2,0,1)	0.1935
ARIMA(2,0,2)	**0.1808**
ARIMA(2,0,3)	0.2148
ARIMA(2,0,4)	0.2035

Table 12: SARIMAX models with TRÄ, GRÅ, MFJ, SE2 and SE3 as input. Model with smallest MAPE is bolded.

Model – DAL	**MAPE**
SARIMA(2,0,2)(1,0,0)[24]	0.1988
SARIMA(2,0,2)(0,0,1)[24]	0.2108
SARIMA(2,0,2)(1,0,1)[24]	**0.1927**

Table 13: ARIMAX models with PROD_TOT, SE2 and SE3 as input. Model with smallest MAPE is bolded.

Model – VML	**MAPE**
ARIMA(0,0,0)	0.3092
ARIMA(0,0,1)	0.2907
ARIMA(0,0,2)	0.2699
ARIMA(0,0,3)	0.2583
ARIMA(0,0,4)	0.2536
ARIMA(1,0,0)	0.1840
ARIMA(1,0,1)	0.1871
ARIMA(1,0,2)	0.1865
ARIMA(1,0,3)	0.1882
ARIMA(1,0,4)	**0.1839**
ARIMA(2,0,0)	0.1872
ARIMA(2,0,1)	0.1858
ARIMA(2,0,2)	0.1865
ARIMA(2,0,3)	0.1880
ARIMA(2,0,4)	0.1877

Table 14: SARIMAX models with PROD_TOT, SE2, SE3 and TEMP as input. Model with smallest MAPE is bolded.

Model – VML	**MAPE**
SARIMA(1,0,0)(1,0,0)[24]	0.1748
SARIMA(1,0,0)(0,0,1)[24]	**0.1736**
SARIMA(1,0,0)(1,0,1)[24]	0.1950

Appendix C

Figures summarising the ARIMAX and SARIMAX models. The left most column contain all independent variables used to calculate the DAL and VML losses. In the same column a variable 'sigma2' is given, which is the variance of the residuals. The coef column give the fitted parameters in the θ-vector and the P>|z| column show p-values from the z-statistics. That is, P>|z| less than 0.05 entails that the null hypothesis of non-significant parameters can be rejected, meaning that the variables should be used in the model.

Figure 34 and Figure 35 regard DAL losses.
Figure 36 and Figure 37 regard VML losses.

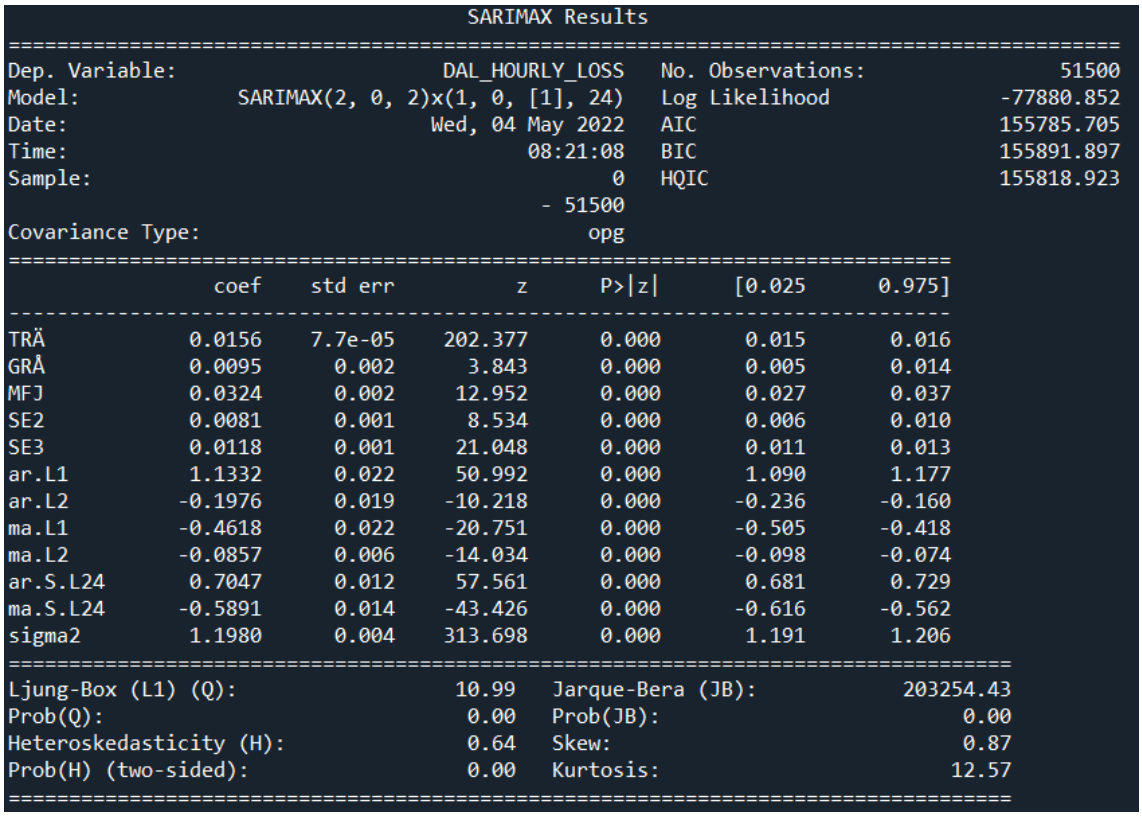
SARIMAX Results

Dep. Variable:	DAL_HOURLY_LOSS		No. Observations:		51500	
Model:	SARIMAX(2, 0, 2)		Log Likelihood		-78453.910	
Date:	Mon, 02 May 2022		AIC		156927.820	
Time:	09:52:14		BIC		157016.314	
Sample:	0		HQIC		156955.502	
	- 51500					
Covariance Type:	opg					

	coef	std err	z	P>|z|	[0.025	0.975]
TRÄ	0.0155	7.73e-05	201.198	0.000	0.015	0.016
GRÅ	0.0149	0.002	5.984	0.000	0.010	0.020
MFJ	0.0299	0.002	12.268	0.000	0.025	0.035
SE2	0.0121	0.001	13.330	0.000	0.010	0.014
SE3	0.0111	0.001	20.390	0.000	0.010	0.012
ar.L1	1.1568	0.025	47.016	0.000	1.109	1.205
ar.L2	-0.2127	0.022	-9.682	0.000	-0.256	-0.170
ma.L1	-0.4473	0.025	-18.092	0.000	-0.496	-0.399
ma.L2	-0.0734	0.006	-12.059	0.000	-0.085	-0.062
sigma2	1.2374	0.004	305.055	0.000	1.229	1.245

Ljung-Box (L1) (Q):	0.41	Jarque-Bera (JB):	220364.97
Prob(Q):	0.52	Prob(JB):	0.00
Heteroskedasticity (H):	0.64	Skew:	0.84
Prob(H) (two-sided):	0.00	Kurtosis:	12.99

Figure 34: Summary of ARIMAX(2,0,2), modelling DAL losses.



SARIMAX Results

Dep. Variable:	DAL_HOURLY_LOSS	No. Observations:	51500
Model:	SARIMAX(2, 0, 2)x(1, 0, [1], 24)	Log Likelihood	-77880.852
Date:	Wed, 04 May 2022	AIC	155785.705
Time:	08:21:08	BIC	155891.897
Sample:	0	HQIC	155818.923
	- 51500		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
TRÄ	0.0156	7.7e-05	202.377	0.000	0.015	0.016
GRÅ	0.0095	0.002	3.843	0.000	0.005	0.014
MFJ	0.0324	0.002	12.952	0.000	0.027	0.037
SE2	0.0081	0.001	8.534	0.000	0.006	0.010
SE3	0.0118	0.001	21.048	0.000	0.011	0.013
ar.L1	1.1332	0.022	50.992	0.000	1.090	1.177
ar.L2	-0.1976	0.019	-10.218	0.000	-0.236	-0.160
ma.L1	-0.4618	0.022	-20.751	0.000	-0.505	-0.418
ma.L2	-0.0857	0.006	-14.034	0.000	-0.098	-0.074
ar.S.L24	0.7047	0.012	57.561	0.000	0.681	0.729
ma.S.L24	-0.5891	0.014	-43.426	0.000	-0.616	-0.562
sigma2	1.1980	0.004	313.698	0.000	1.191	1.206

Ljung-Box (L1) (Q):	10.99	Jarque-Bera (JB):	203254.43
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.64	Skew:	0.87
Prob(H) (two-sided):	0.00	Kurtosis:	12.57

Figure 35: Summary of SARIMAX(2,0,2)(1,0,1)[24], modelling DAL losses.

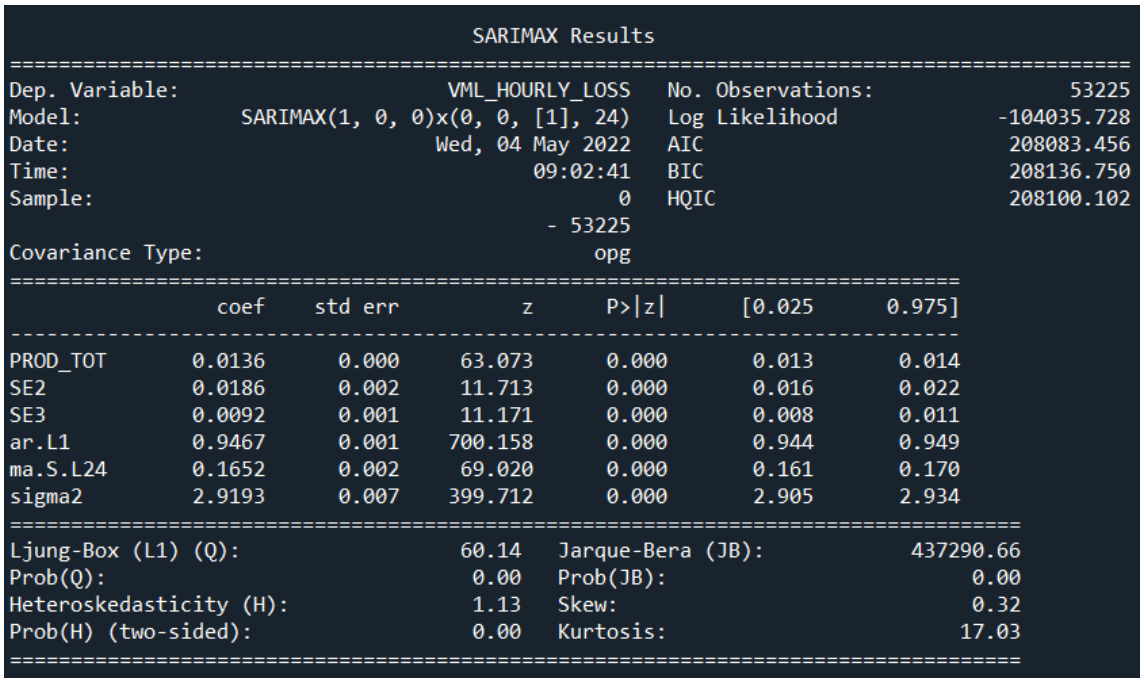
SARIMAX Results

Dep. Variable:	VML_HOURLY_LOSS	No. Observations:	53225
Model:	SARIMAX(1, 0, 0)	Log Likelihood	-104895.663
Date:	Mon, 02 May 2022	AIC	209801.326
Time:	11:02:30	BIC	209845.738
Sample:	0	HQIC	209815.198
	- 53225		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
PROD_TOT	0.0149	0.000	70.634	0.000	0.015	0.015
SE2	0.0211	0.002	14.004	0.000	0.018	0.024
SE3	0.0099	0.001	12.169	0.000	0.008	0.012
ar.L1	0.9559	0.001	759.605	0.000	0.953	0.958
sigma2	3.0153	0.008	397.669	0.000	3.000	3.030

Ljung-Box (L1) (Q):	169.30	Jarque-Bera (JB):	363153.61
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	1.15	Skew:	0.31
Prob(H) (two-sided):	0.00	Kurtosis:	15.78

Figure 36: Summary of ARIMAX(1,0,0), modelling VML losses.



SARIMAX Results

Dep. Variable:	VML_HOURLY_LOSS	No. Observations:	53225
Model:	SARIMAX(1, 0, 0)x(0, 0, [1], 24)	Log Likelihood	-104035.728
Date:	Wed, 04 May 2022	AIC	208083.456
Time:	09:02:41	BIC	208136.750
Sample:	0	HQIC	208100.102
	- 53225		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
PROD_TOT	0.0136	0.000	63.073	0.000	0.013	0.014
SE2	0.0186	0.002	11.713	0.000	0.016	0.022
SE3	0.0092	0.001	11.171	0.000	0.008	0.011
ar.L1	0.9467	0.001	700.158	0.000	0.944	0.949
ma.S.L24	0.1652	0.002	69.020	0.000	0.161	0.170
sigma2	2.9193	0.007	399.712	0.000	2.905	2.934

Ljung-Box (L1) (Q):	60.14	Jarque-Bera (JB):	437290.66
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	1.13	Skew:	0.32
Prob(H) (two-sided):	0.00	Kurtosis:	17.03

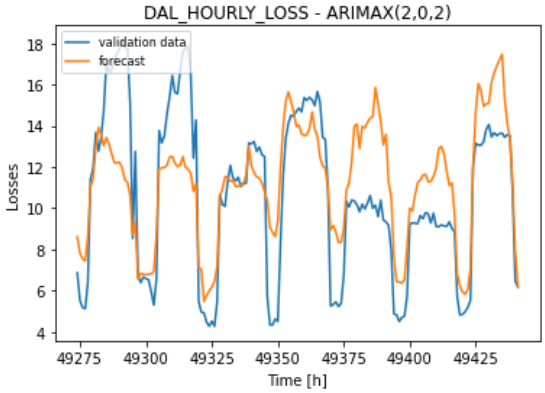
Figure 37: Summary of SARIMAX(1,0,0)(0,0,1)[24], modelling VML losses.

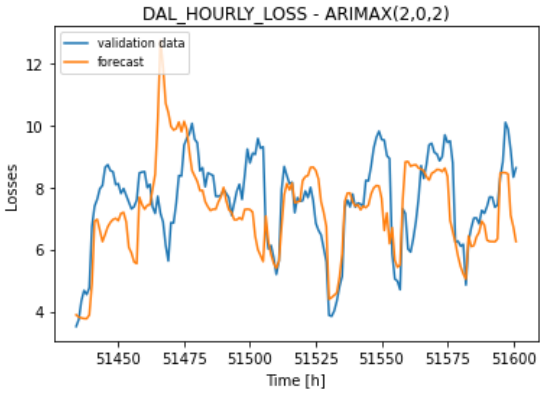
Appendix D

Forecasted losses - Dalarna

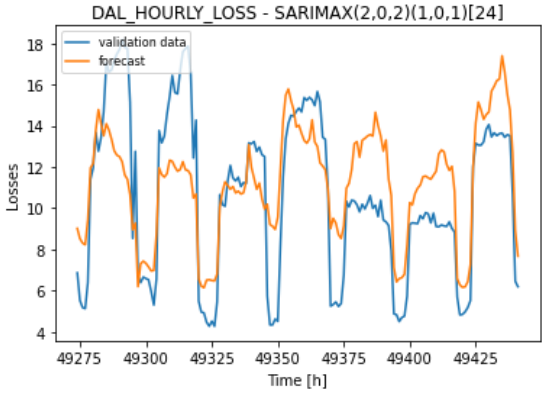
Seven days forecasted DAL losses. Left column: one week in the middle of the validation dataset. Right column: last week in validation dataset.

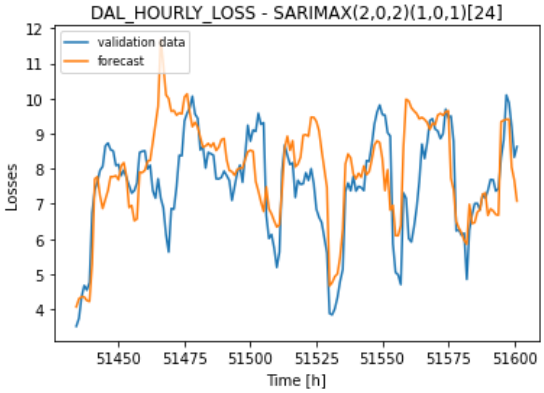
ARIMAX



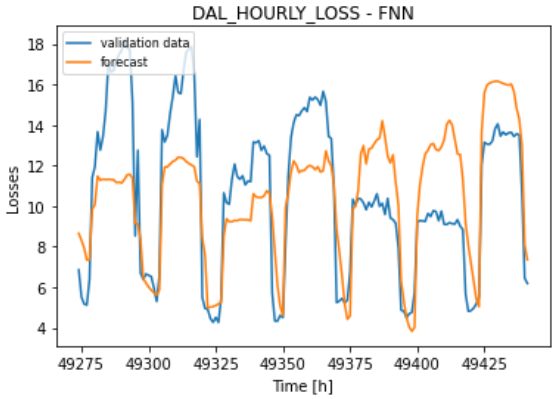


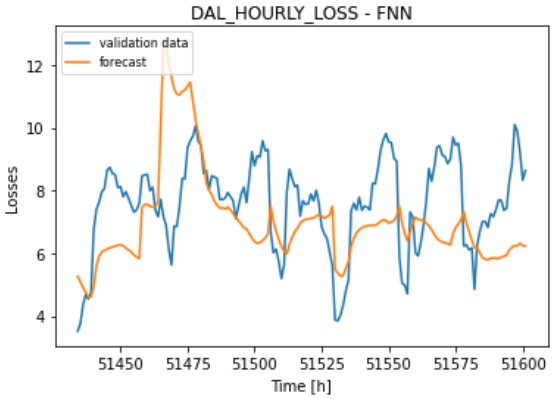
SARIMAX



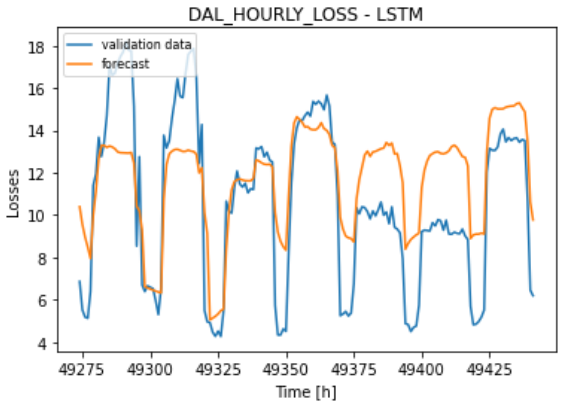


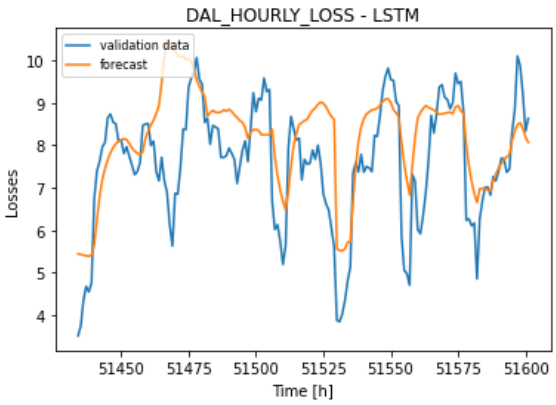
FNN





LSTM

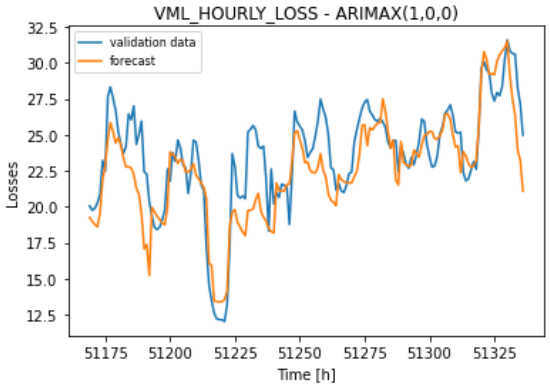


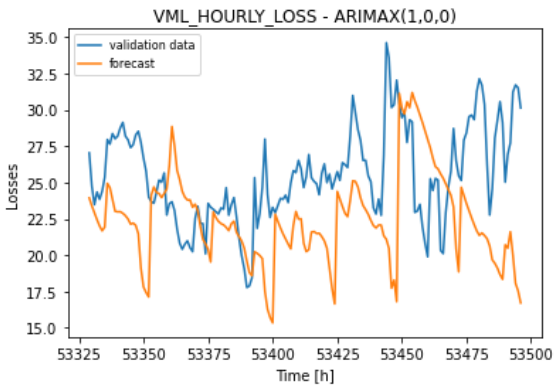


Forecasted losses - Värmland

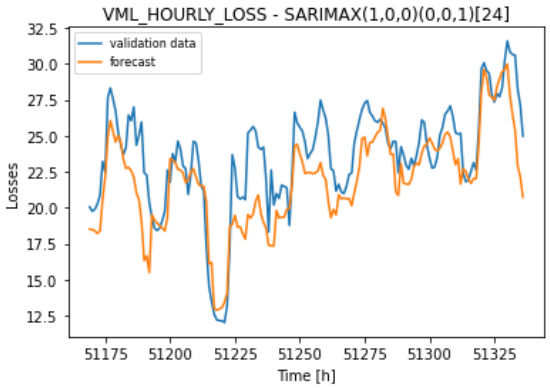
Seven days forecasted VML losses. Left column: one week in the middle of the validation dataset. Right column: last week in validation dataset.

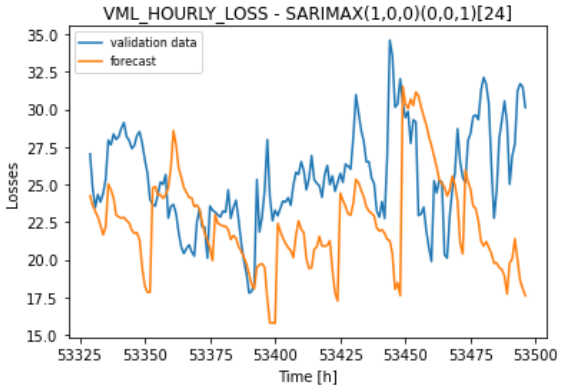
ARIMAX



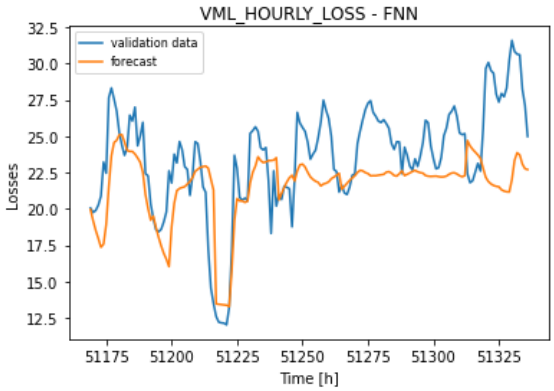


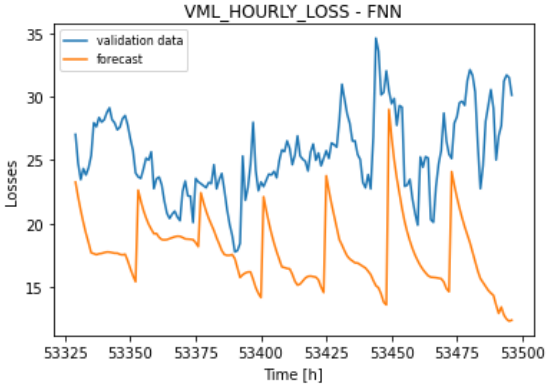
SARIMAX



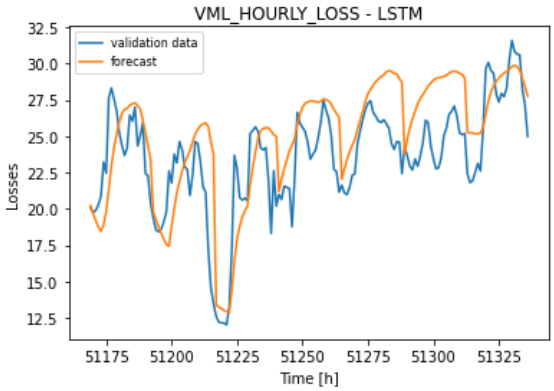


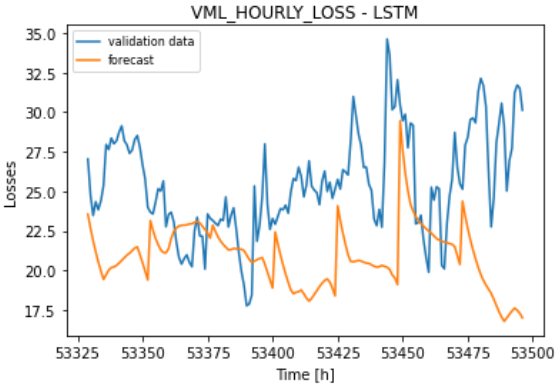
FNN


VML_HOURLY_LOSS - FNN
validation data
forecast
Losses
Time [h]


VML_HOURLY_LOSS - FNN
validation data
forecast
Losses
Time [h]

LSTM


VML_HOURLY_LOSS - LSTM
validation data
forecast
Losses
Time [h]


VML_HOURLY_LOSS - LSTM
validation data
forecast
Losses
Time [h]