CHALMERS
UNIVERSITY OF TECHNOLOGY

The impact of electricity price forecasts on production plans in the district heating system

A comparative study between different forecasting methods

Master's thesis in Master Sustainable Energy Systems

Miriam Bouizgarne
Clara Lindqvist

DEPARTMENT OF SPACE, EARTH AND ENVIRONMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

MASTER'S THESIS 2021

The impact of electricity price forecasts on production plans in the district heating system

A comparative study between different forecasting methods

Miriam Bouizgarne
Clara Lindqvist

Department of Space, Earth and Environment
Division of Energy Technology
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

The impact of electricity price forecasts on production plans in the district heating system
A comparative study between different forecasting methods
Miriam Bouizgarne
Clara Lindqvist

Supervisor: Joel Goop, Per Ringqvist, Sigholm
Examiner: Lisa Göransson, Department of Space, Earth and Environment

Master's Thesis 2021
Department of Space, Earth and Environment
Division of energy Technology
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

The impact of electricity price forecasts on production plans in the district heating system
A comparative study between different forecasting methods
Miriam Bouizgarne and Clara Lindqvist
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

This thesis covers the investigation of electricity price forecasts and the effect they have on production plans and production costs in the district heating system. Four different forecasts were tested on four different models representing district heating systems. The forecasts were investigated in terms of error between the forecast and the actual spot price and the DH-models were optimized in GAMS to find the minimum cost. The first forecast is the naive benchmark which copies previous events. The second forecast is a rolling average performed on the naive benchmark. These two simpler forecasts are compared with a linear regression forecast and a random forest forecast which both are constructed using machine learning-algorithms. The first model of the district heating system contains a combined heat and power plant coupled with a heat pump. In the second model heat storage was added since that is a fundamental equipment in many heat producing plants today. In the two last models an electric boiler and a bio-oil boiler were added respectively. The different units were added to make the system more reality-like and for the purpose of investigating different features and how they were affected by the forecasts. The simple models with few units are less dependent on the forecasts since they have less options of how to produce heat. Thus a limit in the electricity price decides if the electricity producer or the electricity consumer should produce the heat. The linear regression outperformed the other forecasts in terms of error and cost for all models of the district heating plant. Also, a rolling average on a benchmark forecast can make the outcome in form of production cost better. When thermal energy storage (TES) was implemented the production costs were decreased and the difference in cost between the forecasts were larger. The hypothetical future system, using danish electricity prices with more fluctuations which can be expected in electricity systems with more renewable electricity generation, was harder to predict.

Keywords: Forecast, reference, electricity price, production cost, production plan, naive, rolling average, regression, random forest

Acknowledgements

First and foremost, we want to thank our supervisors from Sigholm, Joel Goop and Per Ringqvist, for guiding us through this entire process. They have been really supportive and patient with us from the beginning. Also, we are really grateful for the opportunity to write our master's thesis at Sigholm during the circumstances of the pandemic. The employees have been welcoming and very help full since the start. They were really invested in our road to finish this thesis.

We also want to show our gratitude towards our examiner Lisa Göransson who let us have her as examiner even though she was busy holding courses during the same time. Thankfully, Joel and Lisa had worked together before and she could rely on him to give us the help we needed.

In general we feel very pleasant and happy for having the chance to write our thesis at Sigholm. The opportunity to write a master's thesis at a company during a pandemic was low but Sigholm greeted us and for that we are thankful.

Miriam Bouizgarne and Clara Lindqvist, Gothenburg, June 15, 2021

Contents

List of Figures

List of Tables

List of Concepts

AI Artificial Intelligence.

Benchmark simple forecast models for comparison purposes.

CHP Combined Heat and Power.
COP Coefficient of Performance.
Ctot The minimal total cost calculated in GAMS from the production plan [SEK].

DH District Heating.

EB Electric Boiler.

Fuelprice The price of fuel for running the CHP [SEK/MWh].
Fuelprice oil The price of fuel for running the OB [SEK/MWh].

HP Heat Pump.

k1th Abbreviation for load change rate up for the CHP [MWh].
k2th Abbreviation for load change rate down for the CHP [MWh].

Linear Regression A forecast method using ML-algorithm.
Load change cost is the cost of load change rate for the heat condenser in the CHP [SEK/MWh].

ML Machine Learning.

Naive A forecast method using the same electricity price from the week before. Usually used like a benchmark forecast.

OB Bio Oil boiler.

Pdv Abbreviation for the heat production in the direct condenser [MWh].
Pel Abbreviation for the electricity production [MWh].
PelEB the abbreviation of the electricity consumption of the EB [MWh].
Pelhp Abbreviation of the electricity consumption in the HP [MWh].
POB Abbreviation for the heat production from the oil boiler [MWh].
Pout Abbreviation of discharging of the heat storage [MWh].
Psin Abbreviation of charging of the heat storage [MWh].
Pvk Abbreviation for the heat production in the turbine condenser [MWh].

RF Random Forest A forecast method using ML-algorithm.

Rolling average A forecast method usually used as a benchmark method.

Spot price is the electricity spot price [SEK/MWh].
Start cost CHP is the cost for start-up of the plant [SEK].
Start cost EB the cost for start-up of the EB [SEK].
Start cost HP is the cost for start-up of the HP [SEK].
Start cost OB is the cost for start-up of the oil boiler [SEK].

TES Thermal Energy Storage.

VEB Abbreviation of the binary variable indicating start of EB.
Vhp Abbreviation of the binary variable indicating start of HP.
VOB Abbreviation for binary variable indicating of when the OB is starting.
Vth Abbreviation for binary variable indicating of when the CHP is starting.

1
Introduction

The district heating system is supplying heat through pipelines connected to heat producing plants. The system is a mix of combined heat & power plants (CHP), waste heat from the industry and heat plants consisting of boilers and heat pumps [1]. The idea of district heating is to move individual heating, originally fueled by fossil fuels, into a centralized network. In that way the production of heat can be optimized and the usage of fuel can be controlled. District heating, compared to traditional heating, lowers carbon dioxide emissions, provides high energy security and efficient use of resources [2]. Sweden, especially, fuels many of the heat producing plants with renewables and waste.

To minimize the cost and resources for the producers, is it important with a reliable production plan to decide which units to operate, when and how much. The production plan can be made from optimization models that takes into account several factors including the conditions and constraints of the plant, the heat demand and the electricity price [3]. The electricity market today is unpredictable and as the share of variable renewable energy is increasing the market gets even more volatile and thus harder to predict [4]. An inaccurate forecast can affect plants with different features in a varying amount. A missed opportunity to produce electricity at high electricity prices might for example mean a great loss in income for a CHP plant.

A possible way to improve electricity price forecasting is with artificial intelligence (AI) and Machine learning (ML), an approach within AI. ML-models are algorithms that learn from the input data in order to perform better in the future without external inputs and might be better at handling the non-linear fluctuations in the electricity price[5].

1.1 Aim & research questions

The aim with this thesis is to research forecasts computed with AI/ML. Will these compared to more simple forecasts used as benchmark, not only result in lower errors in forecast, but also in lower production costs for the plants in the district heating system? Premade forecasts are received and the project investigates the dependency of an accurate forecast by looking into the production plans resulting from different forecasts and the resulting costs.

- At what time periods do the forecasts perform well and why? Which variables are the main contributors to the biggest errors? Does external factors affect?
- What is the most difficult for the forecasts to predict? Is it amplitude, time or duration?
- What is the relation between forecast and production cost? Does an accurate forecast (accurate in terms of the average error) result in a low production cost?
- Does certain forecasting methods result in better production plans for certain plants in the district heating system?
- Can the usage of ML-models contribute to more accurate forecasts and more specialized production plans, adapted to the conditions of the plant?
- Does heat storage make it possible to be less dependent on the electricity price forecast?
- How will more renewables in the system affect the accuracy of the electricity price forecasts?

2

Background

This section aims to provide necessary knowledge and background information about the district heating system and electricity market in preparation to investigate how they are correlated.

2.1 District heating system

Heat is produced in thermal plants and distributed throughout the city through a network of pipes to industrial and commercial buildings, apartment blocks, and single-family households. A visualization of a hypothetical system is presented in Figure 2.1. The district heating system may also use waste heat from industrial plants or waste water.

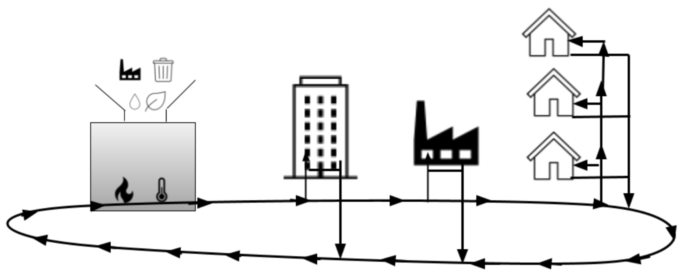


Figure 2.1: A simplified figure showing the principle of a district heating system where the plant is the figure to the left and the three other symbols in the systems presents where the plant provides and get fuel from.

District heating could be built using combined heat and power technology [3]. The figure is showing the production facility to the left. The distribution network is presented by the lines that distributes the heat to the customers.

2.1.1 Combined heat and power

Steam turbine based combined heat and power plants (CHP) are widely used in the industry and the district heating sector. They consist of boilers that burns fuel into heat that in turn heats water into steam. The water passes turbines that create power and then condensers that exchange heat with the water returning to the district heating system. Heat is always created as a by-product of electricity production. Combined heat and power transfer this heat to the district heating system, compared to traditional power generation who let the heat go to waste. Thus, the advantage of CHP is that it maximizes the energy utilization. A CHP plant have an efficiency of around 90-93% [6].

2.1.2 Thermal energy storage

Thermal energy storage, TES, is what it sounds like, a reservoir to store energy that could be used later. TES helps to reduce the time or rate mismatch between energy supply and demand, which is essential for better energy management [7]. By reducing waste energy, TES will save premium fuels and make a plant more cost-effective. Since the transition from fossil fuels to renewable energy sources are on going, TES becomes of more importance. Renewable energy sources such as wind is intermittent and TES will increase its importance with the energy source shift. There exists multiple versions of TES and is a component that has been implanted as a fundamental equipment in systems today. The TES works in that way that when electricity prices are high and the heat demand is high, it is possible to take energy from the heat storage to avoid buying electricity for a high price. When the electricity prices are high but the heat demand is low, it is possible to charge the heat storage again [8].

2.1.3 Heat pumps and electric boilers

The purpose of having a heat pump or an electric boiler is to have a electricity consumer in the system. A heat pump uses energy, or electricity, to produce heat. The ratio of heat output to work input is used to measure a heat pump's efficiency and is called the coefficient of performance, COP [9]. The typical value of COP for a heat pump is usually around 3. By adding a heat pump to the system provides more flexibility to the plant [10]. During low electricity prices, the plant can run the heat pump at low cost instead of producing electricity with low profit. Electric boilers, EB, use electricity to heat water. Since the conversion from electricity to thermal energy are made from resistance. All of the electricity goes through the heating device in an electric boiler which has a resistance on the way which will create the heat. This will make the transformation from electricity to heat almost 1:1 and the efficiency can reach 100% [11]. Electric boilers have lower investment costs than heat pumps.

2.1.4 Bio-oil boilers

Bio-oil boilers, OB, are used as peak boilers. Instead of using fossil fuels, it is possible to use bio-oil that emits less carbon dioxide [12]. Oil boilers can be a cheaper solutions than electric boilers at some occasions, example during high electricity prices and high heat demand. The oil boiler works in that way that the fuel is fired up and warms up the cold water. For an oil boiler to be in usage, they must have at least an efficiency at 86% [13]. This is an equipment in plants that turns on when the CHP is insufficient of meeting the demand [14].

2.2 Electricity market and link to district heating

The electricity price depends on the demand and supply of electricity. The supply in the Nordic countries consists mainly of hydro, nuclear, wind and thermal power plants. The producers supply electricity in the order of increasing production costs until the demand is met [15]. The supply mix depends on several factors such as hydrological conditions, weather, occasional shutdowns, imported electricity and fuel prices. The demand depend mostly on human activity level and the temperature. It varies in the short term from hour to hour and on longer term on season. Low electricity prices occur when there is a lot of wind, hydropower reservoirs are full and high temperatures. High electricity prices occurs during opposite conditions [16].

On the day-ahead market the hourly electricity spot price for the next 24 hours of the following day is decided every day at noon. The spot price is an aggregate where the bids and offers of electricity is matched together to form the market clearing price. Figure 2.2 explains the bidding process.

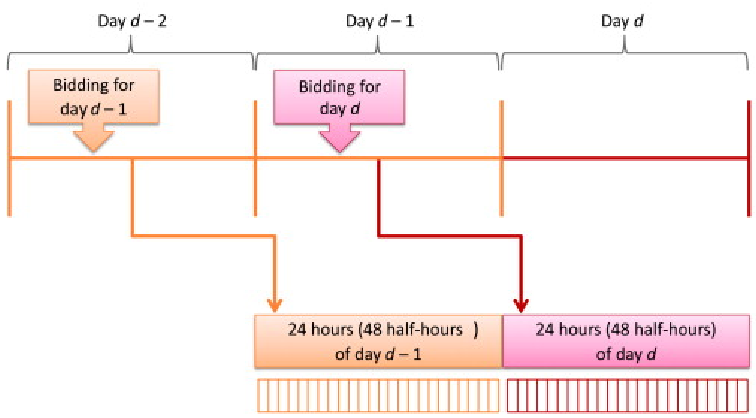


Figure 2.2: The bidding process for the electricity spot price on the day-ahead market [17].

Production plans in heat and power plants are made in advance, before knowing the spot price, for planning purposes. The planned production and consumption of electricity is a part of the buy- and sell bids that forms the basis for the prices that are set for the coming days on the market [17]. This means that the producers are obligated to produce the planned amount of electricity for the next day if the spot price is to apply. If they deviate from the plan, regulatory prices apply that are set on a shorter time scale on the intra-day market. This means that it is important to have a good electricity price forecast when the electricity plan is to be laid [18].

2.3 Electricity price forecasting

Electricity price forecasting has been attempted using a large number of approaches and concepts. The electricity price depends on a wide set of unpredictable variables which lead to sudden changes in the price and makes it hard to predict. The electricity market is special since electricity is economically non-storable and a constant balance is required between supply and demand. Electricity price forecasting is an important tool in decision making for the companies in the heat and power market [17].

2.3.1 Statistical time series modeling

Statistical times series forecasting uses a set of historical data values from a given time period to predict future values. This is widely used within electricity price forecasting [19]. Statistical models are often chastised for their inability to model the usually non-linear behaviour of energy rates and associated fundamental variables. Yet, in practice, their results are comparable to nonlinear alternatives [17]. Linear regression is a common method that can be used within statistical times series modeling.

2.3.2 Artificial intelligence and machine learning models

Machine learning models are able to understand non-linear relations from data and can therefore be more flexible when it comes to electricity price forecasting. In this category, the forecasts models are described as "perfect" for short-term forecasting and at handling non-linear problems [17]. In machine learning you split the data set into a training set on which the algorithm can perform its training and a validation set.

3
Methodology

The methodology section aims to explain forecast modelling, optimization models, input data, the evaluation method and limitations made in this thesis. The fundamentals of the forecasts and how they were modelled are described, however, the main focus is the analysis of the forecasts and not the construction of them.

3.1 Forecast models in Python

When doing time series forecasting it is common to use a benchmark method. When using a benchmark, simple models are created for comparison purposes. The performance of the simple models are set as baseline and should be outperformed by the forecast of interest for it to be any good [20]. Four different forecasts are analyzed in this project, two benchmark forecasts and two forecasts modelled with ML-algorithms. The benchmark forecasts used are *the Naive* and *the Rolling average*. Naive is a common benchmark method along others [21]. The rolling average makes weekly mean values of the naive forecast. Since the naive lags in time compared to the actual spot price, the naive won't properly predict the peaks in time. An expectation is to see if a flattened curve of the naive (the rolling average) will result in a better outcome. These two models are created during this thesis. The other two forecasts used are a linear regression model and a random forest model. The modelling of the ML-forecasts are not performed by the authors since the focus of this thesis is not the construction of ML-forecasts.

3.1.1 Modelling of benchmark forecasts

The benchmark forecasts are constructed in *Jupyter Notebook* using Python.

3.1.1.1 Naive

The naive creates a prediction for day d by copying the electricity price from the same weekday one week before d-7.

3.1.1.2 Rolling average

The rolling average creates a forecast for day d and hour h by guessing on an average of the electricity price for 24 hours around that hour one week before (d-7) and repeats this every following hour ($h+1$, $h+2$...). As mentioned above the rolling average is implemented on the naive forecast.

3.1.2 Modelling of ML-forecasts

The ML-models were implemented using the Python framework Scikit-learn. The models used are ARX-models (Auto-regressive models with an exogenous variable). The ARX-model use lag-features (past values of the time-series data) and the temperature as input to predict future values. Lagged electricity prices and temperatures are added as features for every hour of the training data. The lag-features used are presented in Table A.1. These features are used in both the linear regression and the random forest forecasts. Different models apply to the different days of the week since they show different patterns. Indicator variables are used to add weight to each hour of the day since the price has daily seasonality. The regression is made in a recursive manner where the predicted output is used again as input to the next time step.

3.1.2.1 Linear regression

The linear regression makes an equation system for the price every hour, this can be seen in Equation 3.1.

$$price(t) = a_0 + a_1 * price(t-1) + a_2 * price(t-2) + ... + a_n * T18(t) \tag{3.1}$$

The regression finds values of a_0, ..., a_n to minimize the error presented in Equation 3.2. T18(t) is the corrected temperature *min(18°C-T, 0)*, this is used instead of the actual temperature to take into account that temperatures above 18°C results in a low heat demand that will have negligible impact on the produced amount of heat. Also, the variable can not be negative when constructed this way.

$$error = \sum(price(t) - (a_0 + a_1 * price(t-1) + a_2 * price(t-2) + ... + a_n * T18(t))^2 \tag{3.2}$$

This method assumes that the price the hour before is known. However, in the electricity market there is a lack of data for the previous 24 hours. This can result in a better prediction than reality.

3.1.2.2 Random forest

Random forest is a supervised machine learning algorithm made up of decision trees [22]. A random subset of the features described above are used for every decision tree to make predictions and then an average value of these results is used as prediction output.

3.1.3 Input data

Historical electricity price data from Nord pool was used as input data in the models. Data from year 2017 to 2020 was found in the historical market data database where it is possible to find hourly electricity prices in SEK/MWh. Data from year 2017, 2018 and 2019 was used as input data to the forecasts models and data from year 2019 and 2020 was used as validation. The analysis was made from the 1st of January to the 21st of October and every forecast period starts at 10:00. A forecast period consists of 168 hours and a new forecast period is made once every 48 hours. The time periods was chosen since the ML-forecasts were received in that format. The electricity system is changing as larger amounts of variable renewable energy takes place in the electricity mix [7]. A sensitivity analysis was made to investigate how well the results apply to a future system. In this analysis, Danish electricity prices were used as input to the GAMS-model. The Danish electricity system has a high penetration level of wind power and other renewables compared to the Swedish system [23].

3.2 GAMS models

The models representing the district heating system were modelled in *GAMS*, a linear programming and optimization modeling software. The purpose was to minimize the total cost, C_{tot}, of the system by optimizing the heat and electricity production. This results in a production plan including all units with an hourly resolution. The GAMS model makes an optimization for one week at a time using forecasts as input, but in the analysis the first 48 hours of every 168 hour period is picked out. The optimization is made on 168 hours because doing an optimization on 48 hours can cause unwanted effects. An example of that is that the system might decide to use all heat from the TES if the time period is limited.

Four different models with different features (electricity producer, electricity consumer, heat only producers) are tested to see how the features are affected by different forecasts. Model 1 contains a CHP plant with bypass and a heat pump. Bypass is the possibility to condense heat in a direct condenser without passing the turbine. Thus, the model contains units that are electricity producer and consumer. In model 2 TES was added to model 1 to test the flexibility of the system. In model 3 an electric boiler (electricity consumer) was added as a complement to the heat pump and CHP as a peak unit. Model 4 also has a peak unit but in this case it is a bio-oil boiler (heat producer). Standard models representing a CHP plant, a heat pump, TES and a bio-oil boiler was given to the authors, who in turn made different combination of these models. Parameters and constraints were chosen by the authors as well. The standard models contained equations stating relationships between parameters, as well as energy and mass balances. The different units are described more in detail in later sections. The models are presented in Table 3.1.

Table 3.1: Table containing all the models created in GAMS.

Model 1	CHP+heat pump
Model 2	CHP+heat pump+TES
Model 3	CHP+heat pump+electric boiler
Model 4	CHP+heat pump+oil boiler

The models use *Mixed Integer Programming* (MIP), a type of optimization problems which include integer variables [24]. A solver named Cplex is added that solves linear problems by using several alternative algorithms. The objective function is the cost of the system C_{tot} and is the sum of the cost variables given by Equation 3.3 to 3.13.

Equation 3.3 calculates the fuel price [SEK] for the CHP by multiplying the fuel price [SEK/MWh] with the production of heat in the direct condensers (Pdv [MWh]), turbine condensers (Pvk [MWh]) as well as the produced electricity from the turbines (Pel [MWh]).

$$Fuelprice_{CHP} = fuelprice \cdot (Pdv(I) + Pvk(I) + Pel(I)) \tag{3.3}$$

Equation 3.4 calculates the start cost of the CHP [SEK] by multiplying the start cost [SEK/MWh] with the binary variable Vth that indicates if the CHP is producing or not.

$$Startcost_{CHP} = startcostchp \cdot Vth(I) \tag{3.4}$$

Equation 3.5 calculates the sold electricity [SEK]. The produced electricity (Pel [MWh]) is multiplied with the spot price [SEK/MWh].

$$Electricity_{sold} = -(Pel(I) \cdot spotprice(I)) \tag{3.5}$$

Equation 3.6 calculates the load change cost [SEK] by multiplying the load change cost [SEK/MWh] with either the increase change rate (k1th [MWh]) or the decrease change rate (k2th [MWh]) of the CHP.

$$Loadchangecost = loadchangecost \cdot (k1th(I) + k2th(I)) \tag{3.6}$$

Equation 3.7 have the same principle as Equation 3.4. It multiplies the start cost for the HP [SEK/MWh] with the binary variable Vhp that indicates if the HP is running or not.

$$Startcost_{heatpump} = startcosthp \cdot Vhp(I) \tag{3.7}$$

Equation 3.8 is the opposite of Equation 3.5. The cost of electricity consumed in the HP [SEK] is calculated by taking the spot price [SEK/MWh] multiplied with electricity consumption in the HP (Pelhp [MWh]).

$$Electricity_{consumed,heatpump} = spotprice(I) \cdot Pelhp \tag{3.8}$$

Equation 3.9 calculates the charge cost of the TES. This cost is calculated by multiplying the 0.2 [SEK/MWh], which is the cost of charge and discharge of the TES, with either the charge rate (Psin [MWh]) or the discharge rate (Psout[MWh]).

$$Chargecost_{TES} = 0.2 \cdot (Psin(I) - Psout(I)) \tag{3.9}$$

The start cost of the EB [SEK] is calculated by taking the start cost of the EB [SEK] multiplied with the binary variable indicating the start of the EB (VEB) seen in Equation 3.10.

$$Startcost_{EB} = StartcostEB \cdot VEB(I) \tag{3.10}$$

As in Equation 3.8, the electricity consumption is calculated in equation 3.11. Where in this case the spot price is multiplied with the electricity consumption in the EB (PelEB [MWh]).

$$Electricity_{consumed,EB} = spotprice(I) \cdot PelEB \tag{3.11}$$

The start cost for the oil boiler is calculated by multiplying the start cost of the oil boiler [SEK] with the binary variable indicating the start of the HOB in Equation 3.12.

$$Startcost_{OB} = Startcost_{OB} \cdot VOB(I) \tag{3.12}$$

The production cost of the heat only boiler is calculated by multiplying the fuel price [SEK/MWh] with the amount of produced heat from the oil boiler [MWh] seen in Equation 3.13. Where 0.9 indicates the efficiency of the conversion from fuel to heat.

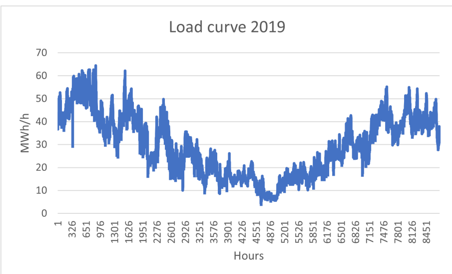
$$Productioncost_{OB} = \frac{fuelpriceoil \cdot POB(I)}{0.9} \tag{3.13}$$

The equations considering the TES, EB and OB are only added to the full equation in the models they are included.

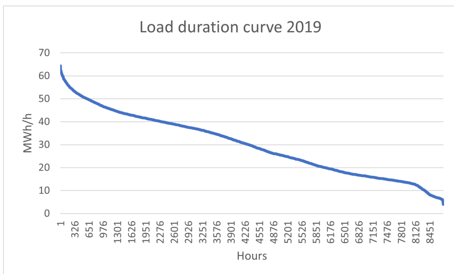
3.2.1 Input data

The electricity price, the heat load and the forecasts that were modeled in Python are used as input data in all the models as well as the electricity tax, that is applied when electricity is consumed from the grid. The electricity tax is set to 350 SEK [25]. A numerous of parameters are also set for the individual GAMS models. The ones of most importance are described below.

The heat load for the system was used to plot the load curve and load duration curve shown in Figure 3.1a and 3.1b for year 2019. The heat load has a large impact on the system. The heat load was used in the design of the GAMS models when choosing many of the parameters.



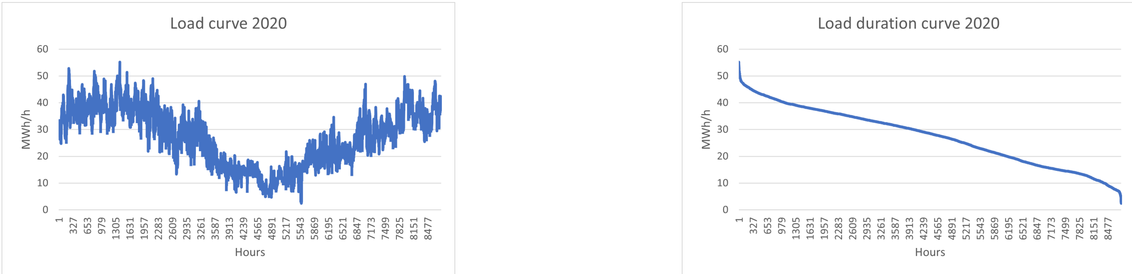
(a) Load curve 2019.



(b) Load duration curve 2019.

Figure 3.1: Load curve and load duration curve for the heat demand year 2019.

The same load curve and load duration curve were made for year 2020 and are shown in Figure 3.2a and 3.2b.



(a) Load curve 2020. **(b)** Load duration curve 2019.

Figure 3.2: Load curve and load duration curve for the heat demand year 2020.

These load curves shows the demand of heat during the year. The load curves show how much the load fluctuates during the year and the load duration curves shows the load during the year in a decreasing order and for many hours that load is needed.

3.2.2 CHP

Some of the parameters were introduced in the equations in section 3.2. Combinations of the parameters described below were tested to achieve a feasible solution in GAMS and reasonable production in the heat producing units. An optimization was not made where all parameters were regulated and assessed. Some of the parameters are standard values. The thermal heat from the CHP boiler, in this thesis called *Pth* was given the values of minimum 10 MW and maximum 90 MW. The CHP is a base-load unit and is therefore expected to operate at most hours, hence the minimum load. The power to heat ratio, the *alpha* value is an important parameter to decide for the CHP plant. Alpha decides how much electricity that can be converted into heat in the system and is set to 0.4 in this thesis. This was set since it is a standard value in CHP. The heat production in the turbine condensers, shortened *Pvk*, has a limitation of minimum and maximum production of heat set to 0 and 65 MW. These effects are decided based on the heat load mentioned in the previous chapter. In this way the CHP plant is able to deliver the full heat load if necessary.

Next parameter is the heat production in the direct condensers, *Pdv*. Production of heat in the direct condensers means that the steam is bypassing the turbines and only produces heat instead of both heat and power. The minimum and maximum value of the bypass is set to 0 and 80 MW. The maximum value says that the direct condenser is able to produce more heat than the turbine produce electricity due to less losses. The *load change rate* in the CHP decides how fast the plant can shift load. This parameter is set to 10 MW, a reasonable value of this is up to 25% of the maximum load. Another parameter is *minimum up time and down time*. Those values are set to 3 hours and 10 hours respectively. The minimum down time represents the time the unit, if taken out of operation, needs to stay out of operation before it can be available for production again. The minimum up time is the time the unit has to stay in operation once it is started.

The *maximum and minimum load* of the turbine is automatically decided as a result of the constraints considering the maximum load of the boiler. All the parameters mentioned above are presented with their set value in Table 3.2.

Table 3.2: Table containing all the input parameters for the CHP used in the GAMS models.

Parameters	Chosen values [unit]
Fuel price	200[SEK]
Alpha	0.4
Pvk min./max.	0/65[MW]
Pdv min./max.	0/80[MW]
Pth min./max.	10/90[MW]
Load change rate	10 [MW]
Min. up time/downtime	3/10 [h]
Turbine max. load/min. load	10/40 [MW]

3.2.3 Heat pump

The COP mentioned in section 2.1.3 is set for the heat pump to be as sufficient as possible. This value is set to three which mean that it is possible to create three times more heat than the electricity input. A COP value of three is a standard value. All the values for the heat pump are presented in Table 3.3. The physical constraints which limits the maximum and minimum heat production, are decided in a similar way as the parameter presented for the CHP, shortened Php.

Table 3.3: Table containing all the input parameters for the heat pump used in the GAMS models.

Parameters	Chosen values [unit]
COP	3
Php min./max.	0/30 [MW]
Load change rate	10 [MW]
Min. uptime/downtime	1/1 [h]

3.2.4 Thermal energy storage

For the model with TES all parameters previously mentioned still apply. The volume of the storage is decided as the *maximum and minimum storage in tank*. The flow of energy into and out of the storage is limited by the maximum charge and discharge parameters. In Table 3.4 are all parameters for the heat storage listed with their values. These values were decided after being tested in GAMS to obtain a reasonable usage of TES compared to the load of the other units.

Table 3.4: Table containing all the input parameters for the TES used in the GAMS models.

Parameters	Chosen values [unit]
Max. storage level	800 [MWh]
Min. storage level	200 [MWh]
Max charge	50 [MW]
Max discharge	50 [MW]

In [26] large-scale hot water tanks capacity is set to around 175 MWh for one unit, the large capacity was decided so that there would be no limiting constraint. The levels in TES after running GAMS never reached as high as 800 MWh.

3.2.5 Electric boiler

The electric boiler is modelled in the same way as the heat pump but with a COP value of 1 and a reduced capacity to 15 MW since the boiler is not a base-load unit.

3.2.6 Bio-oil boiler

The bio-oil boiler is not dependent on the electricity spot price, but instead pays a fuel price. Since the oil boiler is producing when the CHP and heat pump don't have the capacity to cover the demand, the maximum production does not have to be large and an investment in a relatively small oil boiler is reasonable. It also needs to be flexible to turn on and off so the minimum heat production can be set to zero. This is to avoid the start-up cost from zero to minimum production that can occur. Table 3.5 shows all the parameters decided for the oil boiler.

Table 3.5: Table containing all the input parameters for the oil boiler used in the GAMS models.

Parameters	Chosen values [unit]
Fuel price	700 [SEK/MWh]
Max./min. heat prod.	10/0 [MW]
Load change rate	500 [MW]

Another way to make the oil boiler flexible is the load change rate. By having a high value on this parameter, the boiler can adapt to the fluctuation that occurs in demand.

3.2.7 Output

The output variables of interest are used in the *the objective function*, which provides the minimal cost and optimal running plan, the heat production in the direct condenser, the turbine condenser and the condenser in HP. Also the electricity production of the turbine and the electricity consumption of the heat pump. These and other performance outputs depends on the electricity price.

3.3 Evaluation of forecast models

The most common way of evaluating electricity price forecasts is to look at the error between the forecasted electricity price and the actual spot price set on Nord Pool. For the purpose of this thesis the evaluation also considers how the forecasted price affected the resulting production plan and the total production cost over the year.

3.3.1 Model runs

The results were analysed in Jupyter Notebook using Python. All the forecasts with electricity price for 168 hours where manually saved for every other date starting at the 1st of January. The following data set was iterated 48 hours at a time and so on until the 21st of October with a script that loops through all the dates. The loop takes the different forecasts and insert them into the different GAMS models, the optimization is made and result files are opened in Python.

3.3.2 Forecast error

An error measure was used to evaluate the prediction accuracy. It is called the *Mean absolute error* or MAE. The forecast electricity price Pf_h is subtracted from the actual electricity price P_h. The absolute value of this is then divided by the amount of hours of interest (T). The formula is shown below in equation 3.14.

$$MAE = \frac{\Sigma|P_h - Pf_h|}{T} \tag{3.14}$$

3.3.3 Total production cost

The real electricity price was used as input in GAMS to obtain the reference cost of the plant. This is the total production cost that results from following a production plan if knowing the spot price. However, plants that use forecasting today do not have access to the real electricity price in advance. The forecasted electricity prices are used to obtain production plans for all the cases where it it possible to see which units to run and on which effect. The objective function of these optimizations presents the cost that would result if the forecasted electricity prices were correct with no errors. To obtain the real cost the actual electricity price must be used in a cost equation (same as the objective function) together with the electricity consumption, production and heat production obtained from the production plan.

3.4 Limitations

- A hypothetical DH system consisting of different plants was constructed. The hypothetical plant does not completely reflect on reality.

- the rolling average was made on the naive forecast. The method could be implemented on the regression and random forest as well in future studies.

- The modelling of ML-forecasts is not performed in this study.

- The project handles Swedish district heating and Swedish data from Nord pool. SE3 was chosen as the electricity price area to investigate since this is the section in Sweden with the highest variability in electricity price and therefore the most interesting.

- The forecasts were modelled and analyzed for year 2019 and 2020. 2019 is seen as a common year and 2020 is analyzed as a special case that does not represent the usual patterns in electricity prices.

- The assumption is made that the production plants follow the production plan made from the forecast and therefore costs from the intra-day market are not added.

- The temperature used in the modelling of forecasts is the actual temperature and not the forecasted temperature.

- The electricity demand is not used as input to GAMS, since the electricity system is much larger and more complex, the effect of the electricity demand is negligible.

4

Results and discussion

In this section the results regarding the different forecasts and the effect they have on production plans and costs are presented and discussed. The different forecast methods are compared using prediction errors and production costs. The production plans are investigated to learn more about how the different features of the DH system are affected by the different forecasts. All the forecasts are compared after being used as input to the four different models explained in Table 3.1.

4.1 Electricity price forecasts overview

In this section the resulting electricity price forecasts are plotted together with the actual electricity price for year 2019. In Figure 4.1 is the naive forecast plotted with the actual price.

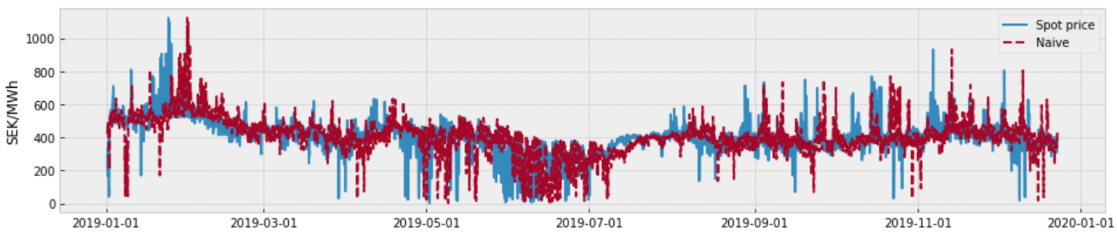


Figure 4.1: The figure shows the naive forecast plotted together with the actual electricity spot price for year 2019.

The naive forecast copies previous events and therefore a peak in the electricity price always reappears as a peak in the naive forecast one week later. In Figure 4.2 is the rolling average forecast plotted with the reference electricity price.

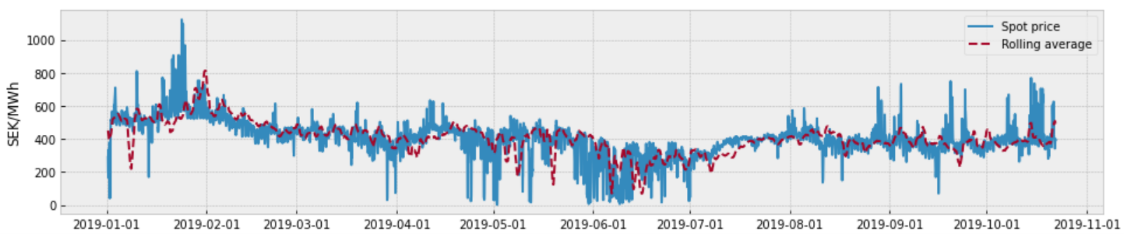


Figure 4.2: The figure shows the rolling average forecast computed plotted together with the actual electricity spot price for year 2019.

The rolling average forecast is the flattened curve compared to the actual price last week, since it is a mean value of the previous electricity prices which also results in less data points. The rolling average is also the mean value of the naive forecast for the week before. This forecast do not properly predict the altitude of the peaks nor the exact position of the peaks. These two forecasts are the benchmark forecast and have fundamental constructions compared to the next two forecasts. In Figure 4.3 is the regression forecast plotted with the reference price.

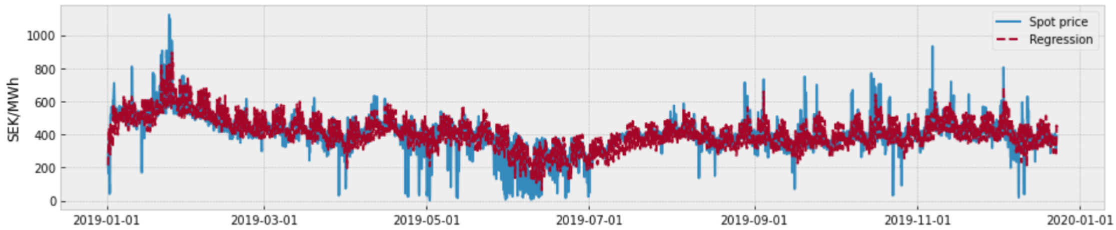


Figure 4.3: The figure shows the regression forecast plotted together with the actual electricity spot price for year 2019.

The regression works on a smaller time interval (linear regression on points in time close to one another) than the benchmark forecasts. This forecast does not manage to properly predict the altitude of the peaks and off-peaks. However, the timing of the predicted peaks is relatively better for the regression than for the naive and rolling average. The regression has access to electricity price data from the same day, which is not the case in reality. Since the electricity data from the past hours says a lot about the current situation, the forecast might be a bit more precise than it would be if only using data from the day before and longer back in time. In Figure 4.4 is the last forecast method, the random forest, plotted with the actual price.

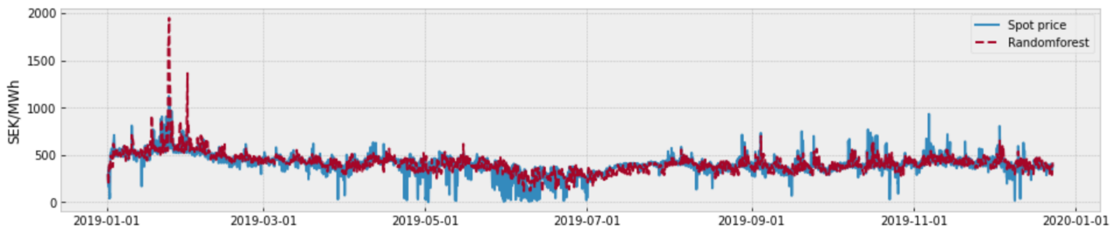


Figure 4.4: The figure shows the random forest forecast plotted together with the actual electricity spot price for year 2019.

The random forest forecast has almost the same pattern as the regression forecast. In the beginning of the year there is a high peak. This peak is the consequence of lack of capacity in hydropower, higher price of raw material and little wind [27]. During this peak, the random forest is predicting an even higher peak then the actual spot price but during the right time.

In Table 4.1 is the average electricity price presented for the reference and the four forecasts.

Table 4.1: The average electricity price for the reference (which is the electricity price for year 2019) and the forecasts in the unit SEK/MWh.

Forecast	Electricity price (SEK/MWh)
Reference	407.2
Naive	409.2
Rolling average	407.1
Regression	408.8
Random forest	417.6

It is clear from Table 4.1 that there are not a lot of differences between the reference price and forecasted prices when looking at the average for a whole year. However, the average price might not be that eloquent when looking at the performance of forecasts. When looking at specific time periods there are significant differences between them. The result of some specific time periods from year 2019 are presented in section 4.3.1.2.

4.2 Errors in forecast overview

In order to see how well the different forecasts performed compared to one another the errors between the actual electricity prices and the predicted electricity prices were calculated. Time periods with large and small errors were compared to see at what time the forecasts performed well and when they didn't. Table 4.2 shows the mean absolute errors calculated for 2019.

Table 4.2: The MAE for all forecasts calculated for year 2019.

Forecast	Average MAE
Naive	59.1
Rolling average	63.3
Regression	41.5
Random forest	48.2

The table shows that the regression forecast has the smallest mean absolute error compared to the rest. The rolling average has the largest MAE, as mentioned before it is a flattened curve of the reference price, therefore it is not a surprise. Figure 4.5 and 4.6 displays the MAE calculated for 48h at a time for the forecasts during the year. This confirms what was shown in Table 4.2, the naive and the rolling average shows significantly larger error measures over the year compared to the regression and random forest forecasts.

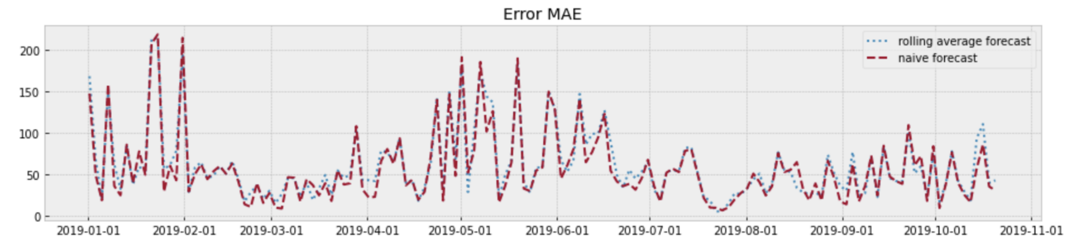


Figure 4.5: The MAE for naive and rolling average shown for the entire year.

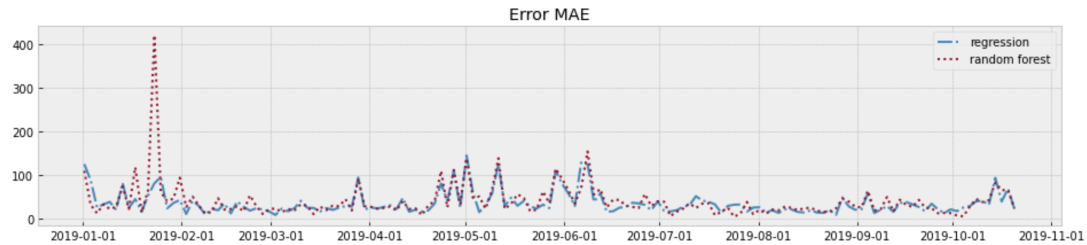


Figure 4.6: The MAE for regression and random forest shown for the entire year.

These figures explains the theory about how the benchmark forecasts have larger deviations from the reference price compared to the ML-models presented in Figure 4.6. The random forest forecast has a large MAE value in the beginning of the year. This is when a large electricity price is occurring. However, it is yet to see what this error does for the production cost.

The figures above shows that overall the largest errors occur during summer and winter. In summer the actual spot price is really low, as can be seen from the figures of the reference and the forecasts in section 4.1, while the forecasts are predicting some higher values. This result can occur since a lot of external factors are decisive during summer such as temperature and vacation. The summer of 2019 was a warmer summer than usual [28]. The pattern during summer vary from year to year which makes it hard to predict with historical data.

4.3 Production costs overview model 1

The forecasts have been investigated in terms of error. The goal here was to see the effect of such errors on the production costs after the optimization in GAMS. Figure 4.7 and 4.8 shows the total production costs resulting from the reference and all the forecasts.

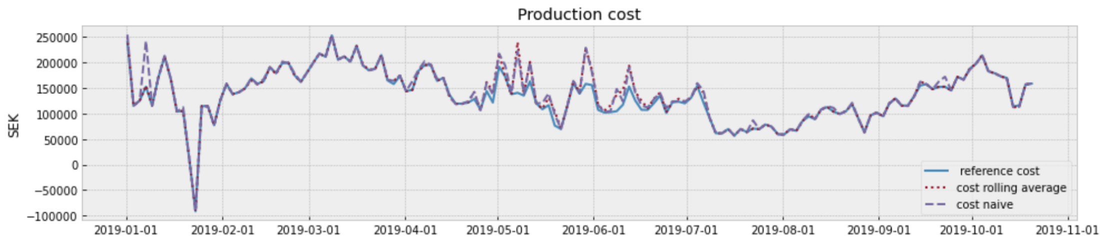


Figure 4.7: The production cost for the reference, naive and rolling average forecasts for the entire year 2019.

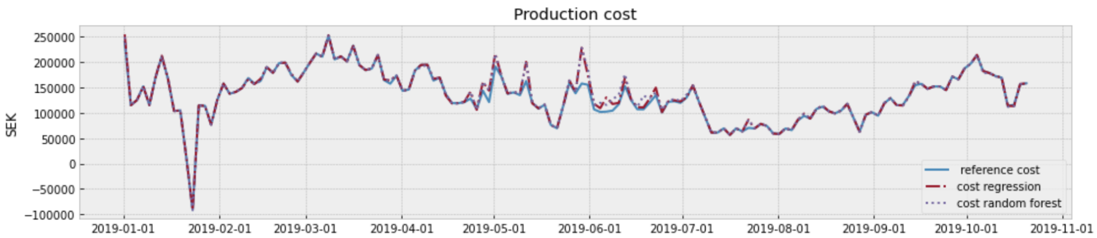


Figure 4.8: The production cost for the reference, regression and random forest forecasts for the entire year 2019.

Even though the naive and the rolling average showed high errors in forecast throughout the year, the production costs are the similar to the reference for large parts of the year. There is a visible cost difference during the summer between the reference and the benchmark forecasts, which also was where the largest MAE values occurred. The reasons for the difference in costs are explained further in section 4.3.1.2. The same patterns can be seen for the ML-forecasts, the error measurements of the forecasts didn't seem to effect the production cost as much as expected. In Figure 4.9 are the naive and the rolling average heat production plotted together with the reference.

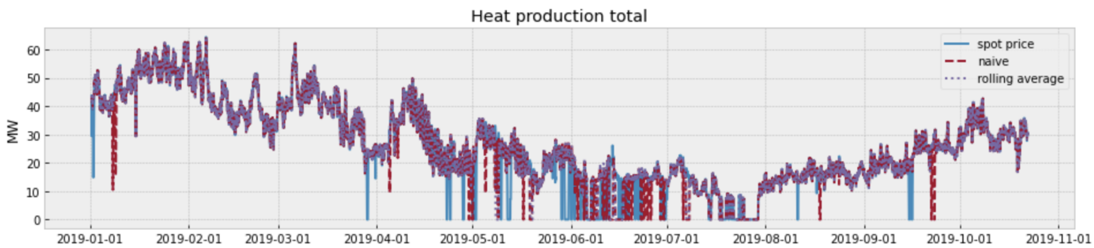


Figure 4.9: The total heat production from CHP for the reference, naive and rolling average forecasts for the entire year 2019.

The heat production from the CHP has a clear pattern, where the production decreases when it gets to the warmer months. In Figure 4.10 are the reference heat production plotted with the regression and random forest.

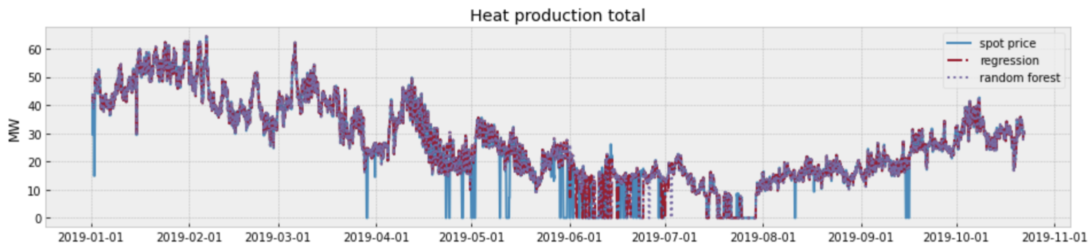


Figure 4.10: The total heat production from CHP for the reference, regression and random forest forecasts for the entire year 2019.

The production of heat has some differences from the forecast during the summer, when also the production cost has it biggest differences. The heat production comes mostly from bypass during summer sine the turbine is not operating during the low electricity price period. The total production costs values for the entire year are presented in Table A.2 in Appendix A.3. The difference in total production cost between the reference and the forecasts for model 1 is presented in Figure 4.11.

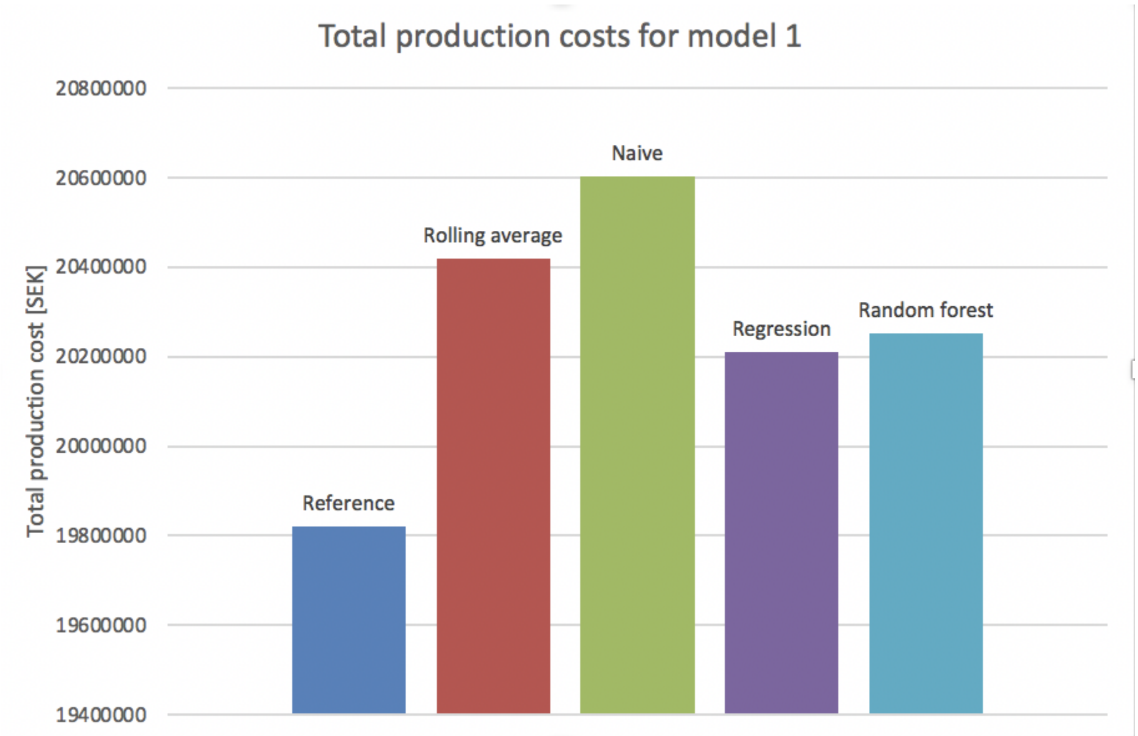


Figure 4.11: The total production costs for model 1.

In model 1 where the CHP and the heat pump are the only units, the difference in cost between the forecasts and the reference is significant. The regression forecasts (the purple staple) is the closest one in cost compared to the reference (the blue staple), Here is it also possible to see that the cost resulting from the rolling average forecast has less deviation from the reference cost than the naive forecast. One reason for testing the rolling average forecast was to see if it would improve the outcome of the naive forecast. These results correspond well with the errors in forecast results. In this case it seems like, overall during the year, a large error in forecast results in a larger production cost.

4.3.1 Production patterns analysis

To get a better insight of how the production plan and production cost depends on the electricity price and the forecasts, some time periods have been picked to explain the results further with some examples. The results shown are examples of general patterns that can be seen throughout the year.

4.3.1.1 Period with same production costs

Even though large differences in errors between forecasts occur during large parts of the year, the production costs are the same during many of these periods. Figure 4.12 shows the production costs resulting from the reference and all the forecasts for a short period in the end of January 2019.

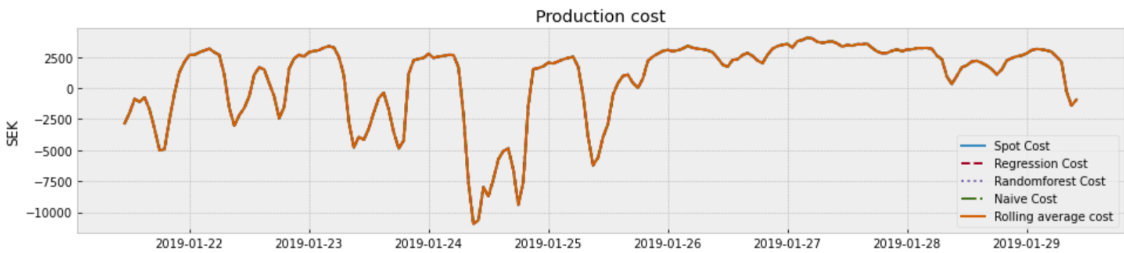


Figure 4.12: Production costs for the reference and all the forecasts for a period in January.

The production cost is the same for all the cases, during this period, even though the errors in forecast that can be seen in Figures 4.5 and 4.6 are not the same. An especially large difference can be seen during the beginning of the year. The cost between the 24th and 26th of January is a negative cost. Since the electricity price has a high peak at this period, the system will not run the heat pump. The production plan is to produce as much electricity as possible and make profit. In Figure 4.13 is the electricity production for the reference and all the forecasts.

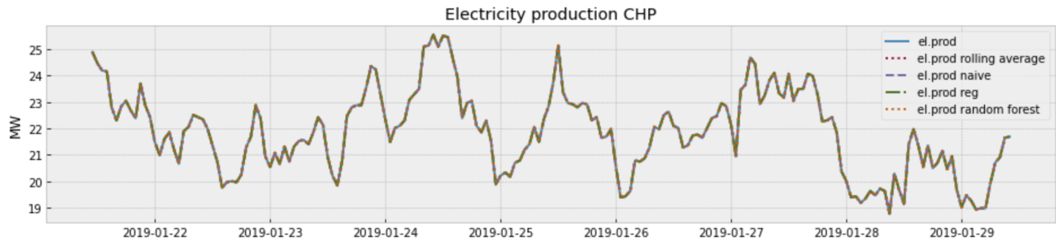


Figure 4.13: Electricity production for the reference and all the forecasts during a period in January.

This figure states that all the forecasts are producing electricity at the same amount as the reference. The variations in production are determined by the heat demand. By looking at the electricity price during this period these results can be explained. In Figure 4.14 is the spot price plotted together with the naive and the rolling average forecast during this period.

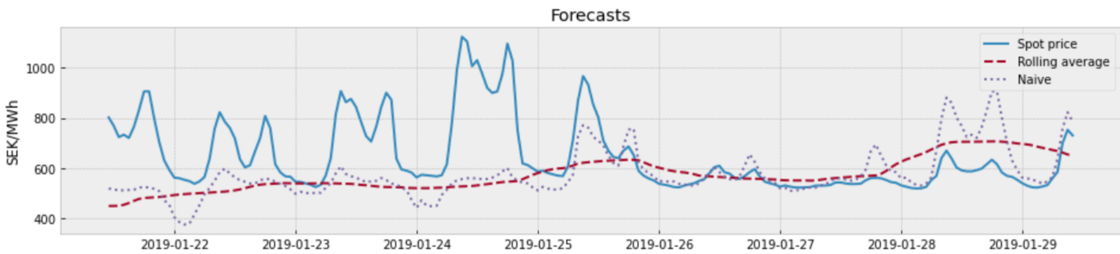


Figure 4.14: Forecast for the reference, naive and rolling average for a period in January.

This period has a high peak between 24th and 25th which is the same time as the production cost was negative. However, the naive and the rolling average do not reach the altitude of the peak but still predict relatively high electricity price. In Figure 4.15 is the reference plotted with the regression and random forest.

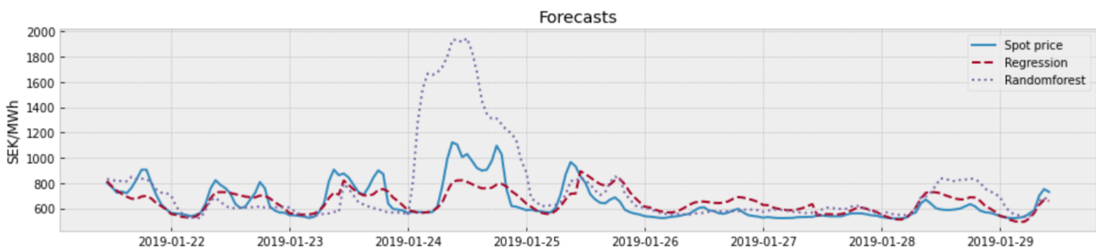


Figure 4.15: Forecast for the reference, regression and random forest for a period in January.

Here both the regression and random forest predicted electricity prices higher than for the benchmark methods. However, the random forest forecast is predicting a higher price than the reference. This distribution of different prices do not seem to affect the production cost for any of the cases. It is also clear from the figure that the regression and random forest forecasts predict the daily seasonality of the price well. From this period it is possible to state that, during a high price peak, the important quality of the forecast is to time the peak, not to reach the peak in altitude. The system consists of only two plants and therefore the options of how to produce heat are not that many. Since the profit occurs from production of electricity all the forecasts and the reference results in running the CHP. Also, all of the forecasts resulted in a negative cost during this period which is auspicious for the system.

4.3.1.2 Periods with different production costs

There where also periods with difference in production costs. The biggest errors were seen during the summer months. Figure 4.16 shows the production costs resulting from the reference, rolling average and the naive forecast during a period in the beginning of June.

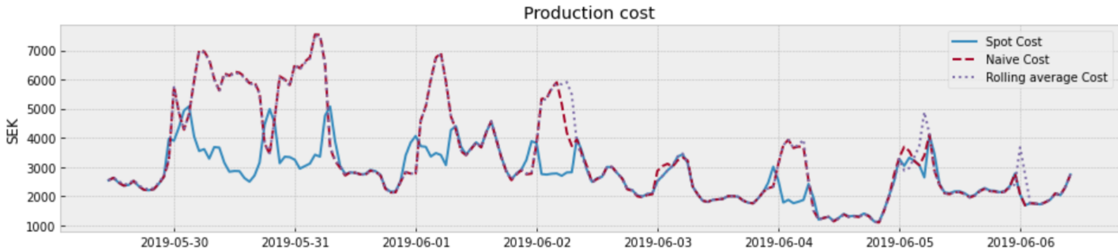


Figure 4.16: Production costs resulting from the naive and rolling average forecast at a period in June.

This is a period when the forecasts have higher production costs than the reference. In Figure 4.17 is the same period plotted for the reference, regression and the random forest forecasts.

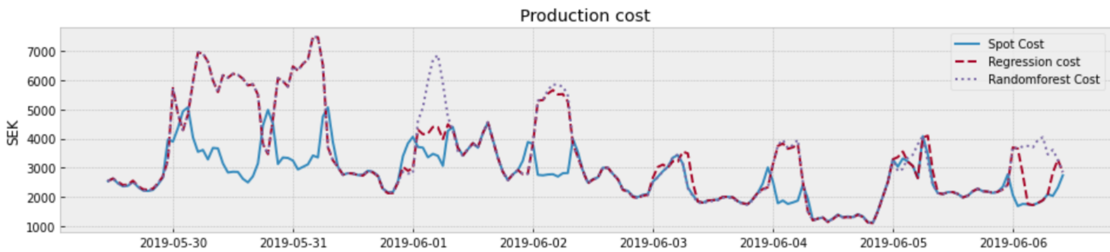


Figure 4.17: Production costs resulting from the regression and random forest forecasts at a period in June.

Also here the costs are lower for the reference case. These high costs can be explained by looking at the production patterns resulting from the forecasts. Figure 4.18 shows the electricity consumption in the heat pump resulting from the reference, rolling average and naive forecasts.

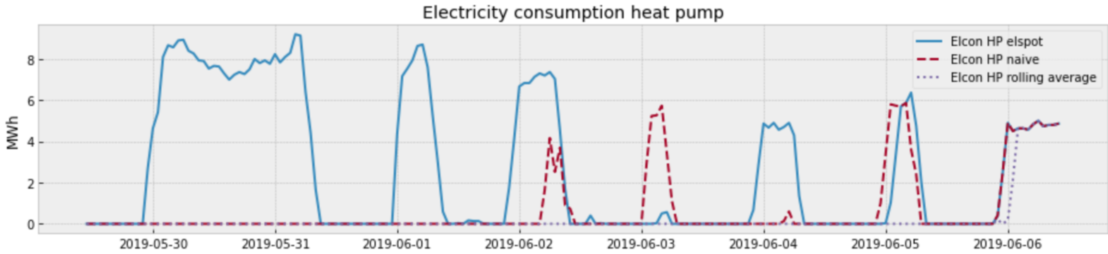


Figure 4.18: Electricity consumption in heat pump resulting from the naive and rolling average forecasts at a period in June.

Here is a visible image showing that different forecasts results in different production patterns. When the spot price is low the heat pump consumes electricity, but this is not the case for the naive and rolling average. Similar patterns are shown in Figure 4.19 where the electricity consumption is plotted for the reference, regression and random forest forecasts.

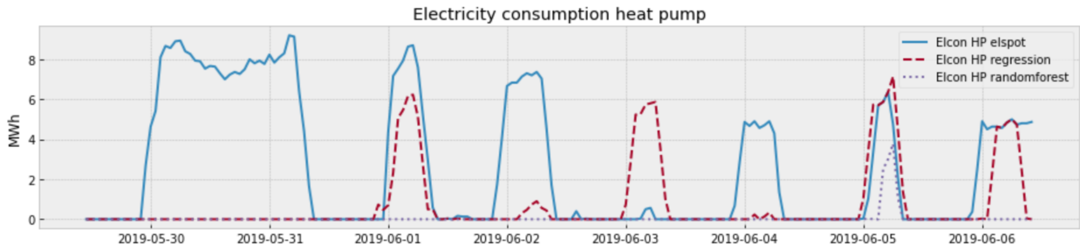


Figure 4.19: Electricity consumption in heat pump resulting from the regression and random forest forecasts at a period in June.

This mismatch can in turn be explained by looking at the electricity price forecasts during this period. Figure 4.20 and 4.21 shows how all the forecasts fail to predict the low electricity price.

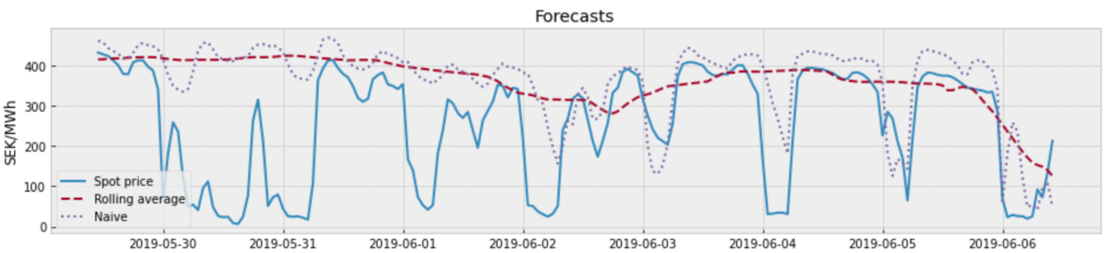


Figure 4.20: Forecasts for the reference, naive and rolling average for a period in June.

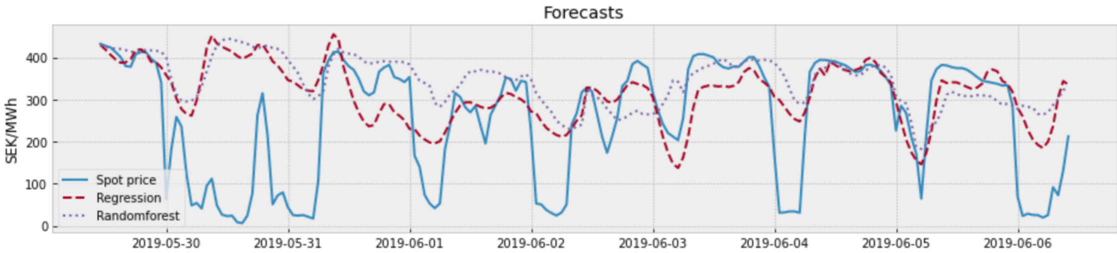


Figure 4.21: Forecasts for the reference, regression and rolling average for a period in June.

The actual electricity price is lower than the prediction for all the forecasts. This is why, in the reference case, the heat pump produces more than the cases with the forecasts. The production costs gets higher than the reference since the production of electricity is expensive during this period and the consumption of electricity is cheap. In this example it is possible to state that to not mange to predict low electricity hours can be crucial for the production cost.

4.4 Production costs overview model 2

In model 2 TES was added to the system to see how it would affect the production plan and total production cost compared to model 1. Figure 4.22 shows the production cost for the reference, rolling average and the naive forecast.

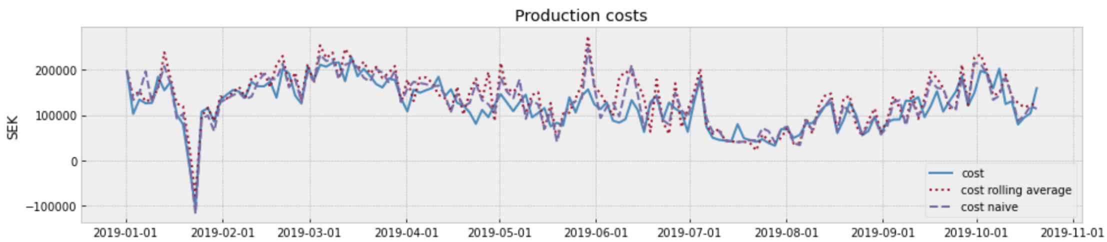


Figure 4.22: Production cost for naive and rolling average with TES for year 2019.

In Figure 4.23 are the production cost for the reference, regression och random forest plotted together when TES is added to the system.

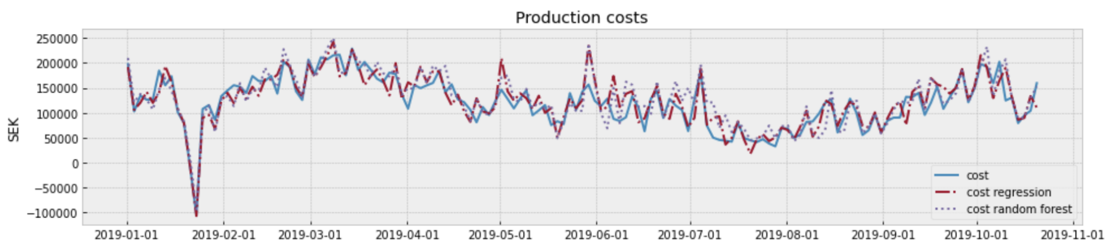


Figure 4.23: Production cost for regression and random forest with TES for year 2019.

It is clear that there is a larger variation in costs between the reference and the different forecasts when TES is added to the system. From the figures of the MAE in section 4.2 the largest errors in forecast (besides the peak in the early year) occur during summer. This is also when the production cost result in larger deviations from the reference, which can be seen in Figures 4.22 and 4.23. In Figure 4.24 are the total production costs resulting from the reference and the forecasts.

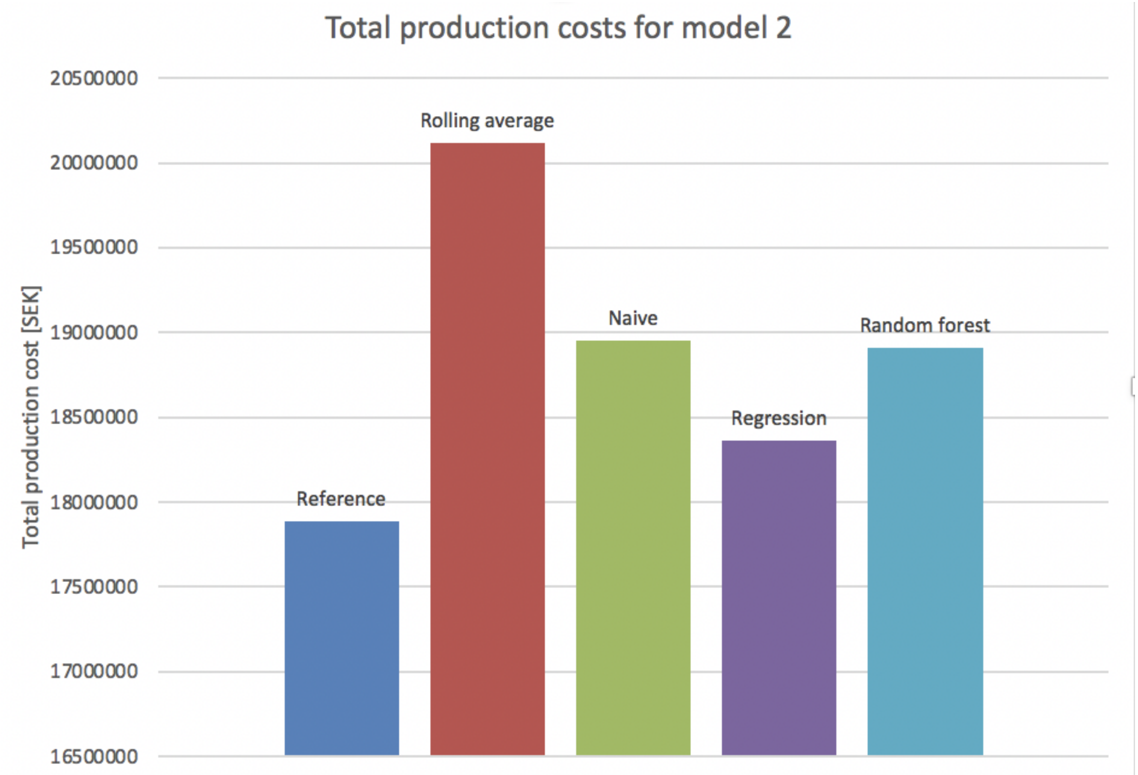


Figure 4.24: Total production costs for the reference and all forecasts for model 2.

The figure shows a remarkable decrease in the total production cost compared to model 1. This is because the TES makes the system more flexible and can help avoid running the heat pump at high electricity prices when the heat demand is high. However, the rolling average forecast now results in a higher cost than to the naive forecast compared to the previous model. This is because flattening of the forecast makes it harder to plan when to use the TES. The cost resulting from the regression forecast is outperforming all other forecasts here as well. The values of the total production cost is given in Table A.3 together with the differences in costs from the reference.

4.4.1 Production patterns analysis

Figure 4.25 shows the results for the same period in January that were described in section 4.3.1.1 for model 1. Without TES the resulting total cost for the system was the same for the reference and all the forecasts. But for model 2 there were significant differences during this period. This can be seen in Figure 4.25 where the total production cost is shown for the reference and all the forecast.

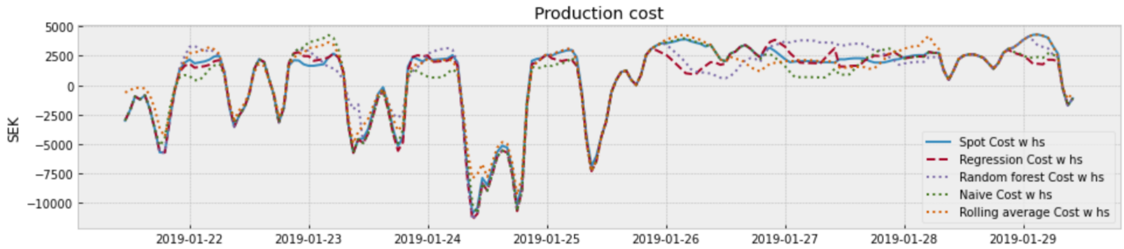


Figure 4.25: Production cost for the reference and all the forecasts, resulting from model 2, for a period in January.

It is clear that there are deviations in costs between the reference and the forecasts during this entire period. This makes it even more clear how the TES makes the system more flexible and not just dependent on the electricity price level. The system can now choose to operate the TES and not just the CHP and the heat pump. Since most district heating plants and systems are equipped with TES today, the results from model 2 are important and shows even more clearly the importance of forecasts. Figure 4.26 shows the electricity production resulting from the reference, naive and the rolling average forecast.

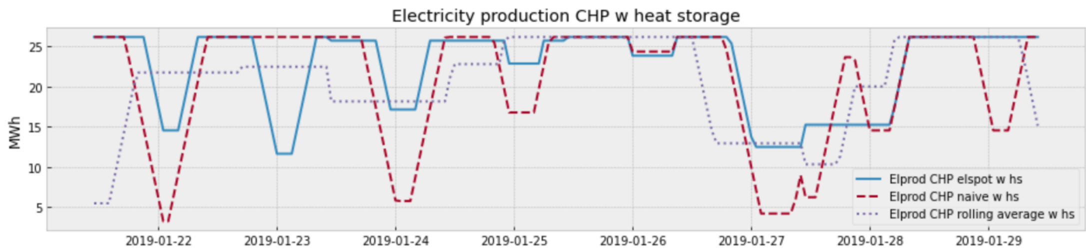


Figure 4.26: Electricity production in the CHP resulting for the reference, naive and rolling average for a period in January.

In Figure 4.27 are the same results but for the reference and the ML-forecasts.

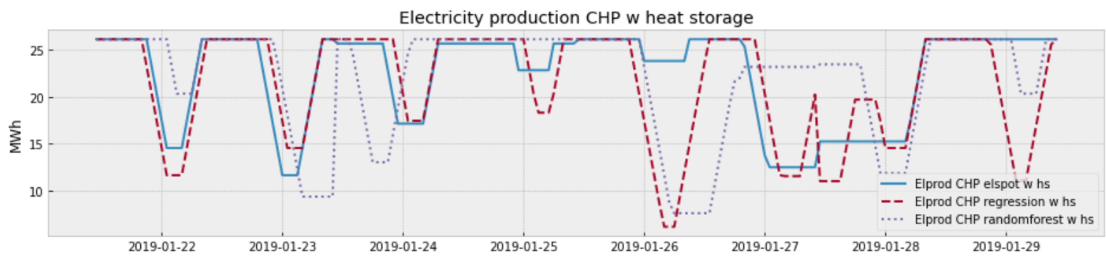


Figure 4.27: Electricity production in the CHP resulting for the reference, regression and random forest for a period in January.

These figures shows significant differences in production which is the cause of the difference in production costs as stated earlier as well. Figure 4.28 shows the energy output from the tank plotted for the reference, the naive and the rolling average.

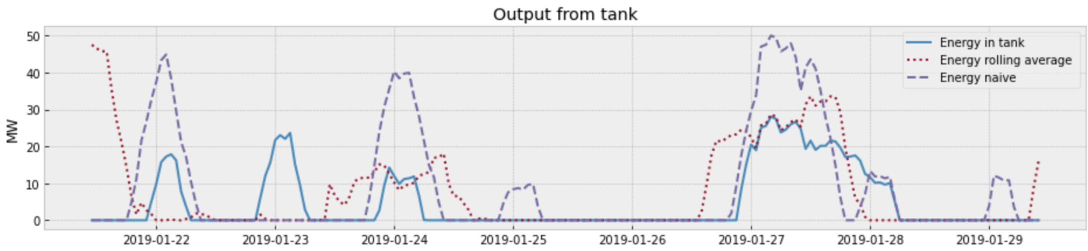


Figure 4.28: Energy level in the TES for the reference, naive and rolling average for a period in January.

Figure 4.29 shows the same results for the reference, regression and random forest.

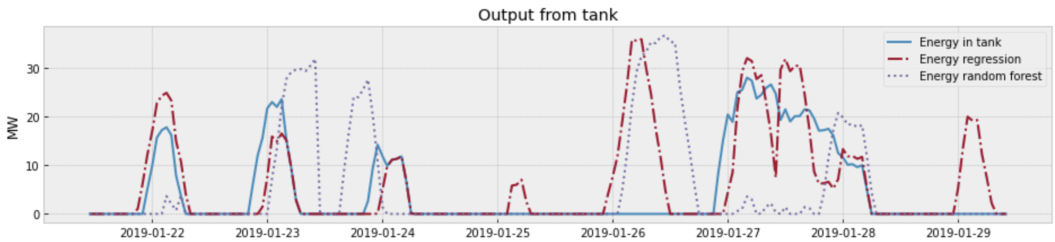
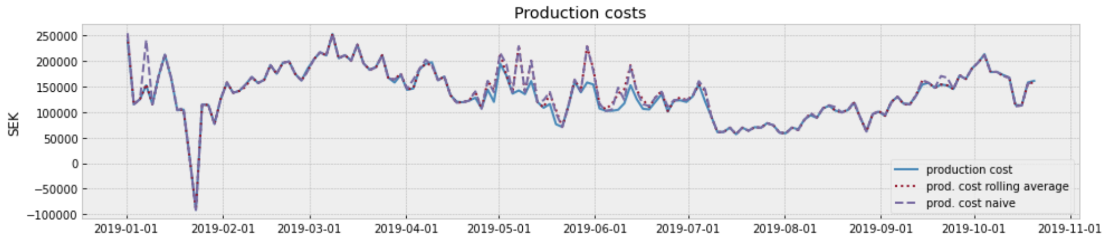


Figure 4.29: Energy level in the TES for the reference, regression and random forest for a period in January.

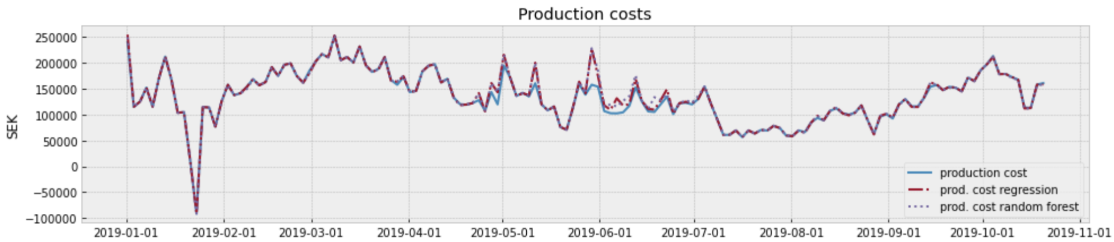
The usage of TES is also a significant factor of the differences in production cost. By using the TES, it is possible to produce heat without having to buy electricity. But by discharging too much will lead to longer time to charge the storage again and it might not have enough capacity during those hours it is needed. This creates a more complex system and production plans. More electricity is produced in the CHP for the reference case, while the forecast cases rely more on TES. Since the TES can manage large fluctuations, there are larger differences in costs during summer. This is when the TES can take advantage of the fact that it can use stored heat. In this thesis TES is used in a short time frame, and the levels vary from hour to hour. In reality TES is important when it comes to long-term storage (seasonal). It is more difficult for the forecasts to predict the electricity price on longer time-scales than just 168 hours ahead. More investments in electricity price forecasts on a longer time frame might be important when it comes to implementing TES.

4.5 Results for electric boiler and oil boiler

In this section the results for model 3 and 4 are presented. These models where constructed with the purpose of adding a peak-load unit and hence have a more complex system with more units to choose from in the optimization. Figure 4.30a and 4.30b shows the production costs plotted for the reference and the forecasts for model 3.



(a) Production cost for the reference, the naive and the rolling average for model 3.

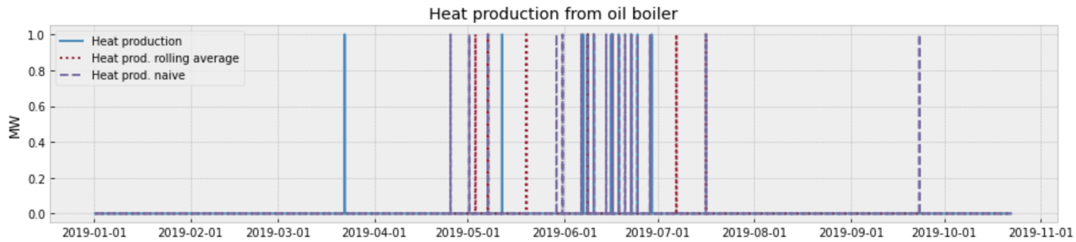


(b) Production cost for reference, regression and random forest for model 3.

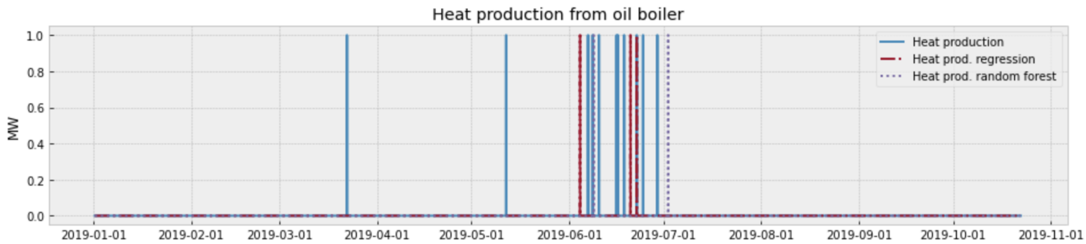
Figure 4.30: The production costs for the reference and all forecasts for model 3.

There were no difference in costs when comparing these results with model 1. This is simply because the optimization model did not choose to run the electric boiler at all. The system could have chosen not to run the EB at all due to the high capacities of the CHP and the HP. This is not surprising since investment costs is not included in the calculations. EB is mainly motivated by offering a low investment cost. If there is not enough capacity to meet the heat load at certain hours or not enough capacity of the HP to take advantage of low electricity prices, the EB comes in handy. However, with a CHP unit which can cover all the heat load and a HP on top of this, the systems has larger investments in capacity that is common today. This means that the system that is investigated has more flexibility to the expense of higher investment cost. This is different parameters were tested in GAMS but no composition was found that choose to include the electric boiler. In model 4, where an oil boiler was added instead, there was a slight difference in costs. The total production costs were slightly increased compared to model 1 and can be seen in Table A.4 in the appendix.

Figure 4.31a and 4.31b shows the heat production in the oil boiler, the boiler starts at about six occasions throughout the year.



(a) Heat production from oil boiler for reference,naive and rolling average where 1 MW indicates that the boiler is running, for model 4.



(b) Heat production from oil boiler for reference, regression and random forest, where 1 MW indicates that the boiler is running, for model 4.

Figure 4.31: The amount of times the oil boiler is operating for the reference and all the forecasts for model 4.

The oil boiler is only used at a few occasions where it is more profitable than running only the CHP and the HP. This can happen during hours when the electricity price is low but the heat demand is high, it is not profitable to run the CHP at low prices and the heat pump is not enough to meet the demand. As seen from the figures, the forecasts has trouble with the timing of the operations. Since few occasions occur where it is profitable to run the oil-boiler it is even more important that the forecasts manage to predict these occasions. Differences in production of heat in the oil boiler between the forecasts and the reference will lead to high differences in total production costs between the cases. Because of high fuel prices, the oil boiler also increases the total production cost for the system.
In order to get more reliable results and make more conclusions considering the electric boiler and the oil boiler more work has to be put into the details of the models in GAMS and what parameters to choose in the different units.

4.6 Sensitivity analysis

The forecasts and production plans both have multiple factors that could change the outcome of the results. Two new data sets were used to see if the results would change when using different data as input when making the forecasts.

4.6.1 Year 2020

The first sensitivity analysis was to use data from 2020. The patterns in electricity prices were different this year compared to usual years. Figure 4.32, shows the rolling average and the naive forecasts plotted with the actual electricity price for year 2020 and figure 4.33 shows the same results for the ML-forecasts.

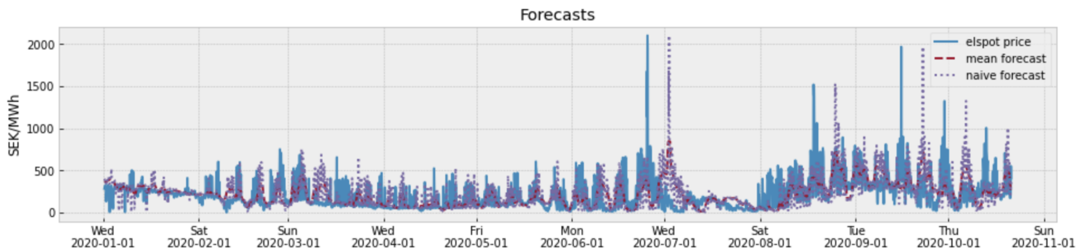


Figure 4.32: The forecast price for the reference, naive and rolling average forecasts for the entire year 2020.

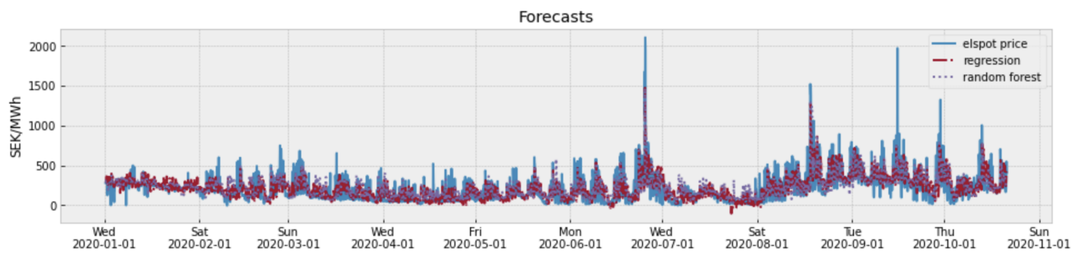


Figure 4.33: The forecast price for the reference, regression and random forest forecasts for the entire year 2020.

The benchmark-forecasts only use data from 2020. A change in data set is therefore expected to change the results more than for the ML-forecasts. The regression and random forest forecast trains on data from previous years (not only 2020) and are therefore expected to be less affected by this change in data. It is hard to see much details in the figures showing the forecasts over the year but figure 4.33 seems to shows how the regression and random forest forecasts successfully follows the reference electricity price. Compared to 2019, where the electricity prices were really high in the beginning of the year, the electricity prices for January 2020 are remarkably low. The mean electricity price for 2020 was the lowest in a long time and negative prices also occurred. The high levels of water was the main contributor. This was also a consequence of high temperatures in the beginning of the year which decreased the demand of heat and high wind production [29].

However, the price has a peak in the middle of the summer which is rare. This peak is a consequence of the low prices from January and little wind [30].

The MAE values for the forecasts during 2020 are presented in Table 4.3.

Table 4.3: The average value of the MAE for the forecasts.

Forecast	average MAE
Naive	102.9
Rolling average	112.2
Regression	77.0
Random forest	89.1

Compared to 2019 in Table 4.2, the average MAE values are significantly higher for all forecasts. However, in comparison between the forecasts the regression forecast has the lowest MAE in 2020 as well as in 2019. The rolling average has the highest MAE as it did for year 2019. However, a shift between the rolling average and the naive occurs in which of them that had the largest errors and largest total production cost throughout the different models for year 2019.

The production costs for 2020 can seen in Figure 4.34 and in Figure 4.35.

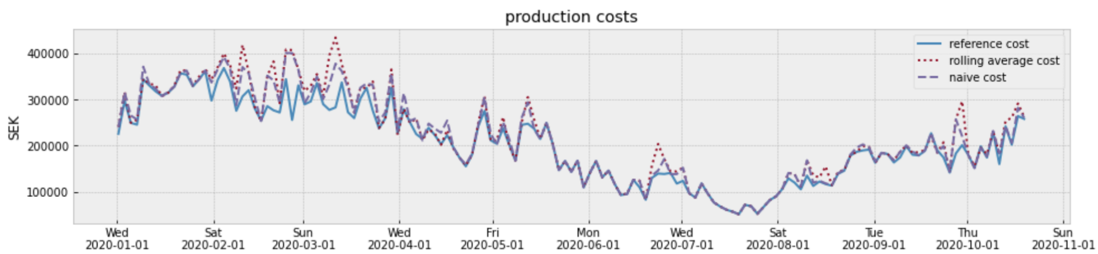


Figure 4.34: The production cost for the reference, naive and rolling average forecasts for the entire year 2020.

are the reference cost plotted with the regression and random forest for year 2020.

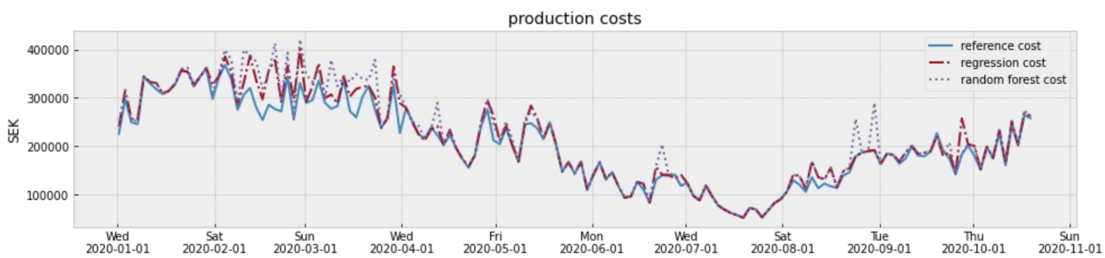


Figure 4.35: The production cost for the reference, regression and random forest for the entire year 2020.

All the forecasts resulted in larger deviations in cost from the reference compared to 2019, this can be seen throughout the year. The forecasts seem to predict higher prices than the reference during the year. The figures show that the forecasts predicts the best during summer and show the largest errors during march. March is also the month where the Covid-19 pandemic started to affect Sweden and the living patterns of people drastically changed.

The total production costs for 2020 are presented in Table 4.4.

Table 4.4: The total production costs for the reference and the four forecasts and the difference between the reference and the forecasts for year 2020.

Forecast	Total production cost [SEK]	Difference in cost
Reference	30 123 561	0
Naive	32 001 408	+1 877 857
Rolling average	32 503 184	+2 379 623
Regression	31 676 612	+1 553 051
Random forest	32 450 547	+2 326 986

The table shows that all forecasts were far above the reference cost for year 2020. The differences are much larger than in any of the previous models. The regression forecast is the closest one in cost ,compared to the reference, in all cases investigated and also for this sensitivity analysis. This reinforce that ML-forecasts has the ability to predict a forecast even though the patterns in data that exists are disturbed. The reference cost is also significantly higher than for all cases in 2019. This was due to the irregularities in electricity prices mentioned earlier.

4.6.2 Future electricity system

As more and more renewables are introduced to the electricity system, the electricity prices tend to fluctuate more. This can affect the performance of the forecasts. Another sensitivity analysis was therefore made with Danish electricity prices. In Figure 4.36 are the Danish electricity prices plotted against the Swedish ones for year 2019.

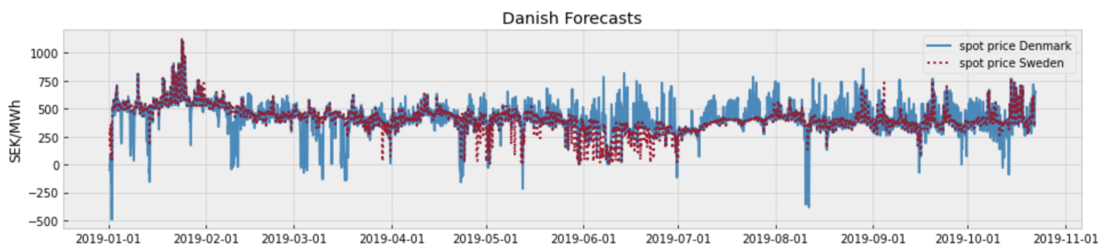


Figure 4.36: The Danish electricity prices are presented as the blue line and the Swedish electricity prices as the red dotted line for year 2019.

The Danish system has more renewable in the system and as a result the electricity prices vary more than Swedish prices and negative prices occur at multiple occasions. In Table 4.5 the average electricity prices for Sweden and Denmark presented.

Table 4.5: The average electricity price for Sweden and Denmark year 2019.

Country	Average electricity price [SEK/MWh]
Sweden	407.2
Denmark	409.3

Even though Denmark has some negative prices, the average price is higher than in Sweden. The figure shows that during summer time Denmark has higher prices than Sweden. The production costs resulting from the Danish prices and the Swedish prices are shown in Figure 4.37.

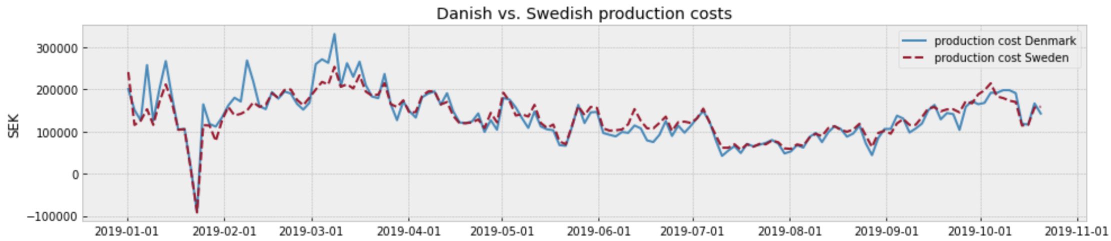


Figure 4.37: The Danish production cost for the Danish reference and the Swedish production cost for the Swedish reference.

The forecast methods has similar patterns for the Danish production cost as for the Swedish one. However, the forecasts do not have the same amount of errors during summer than for the Swedish system but more problems in the beginning of the year. In Table 4.6 are the total production cost for the reference and the benchmark forecasts presented. Also the difference between the reference and the forecasts.

Table 4.6: The total production costs for the reference and the two benchmark forecasts for danish production cost.

Forecast	Total production cost [SEK]	Difference in cost
Reference	19 993 849	0
Naive	22 349 383	+2 355 534
Rolling average	22 058 709	+2 064 860

The total production costs for the reference are slightly higher than the Swedish ones from Table A.2. However, the benchmark forecasts has significantly larger differences in the Danish case than for the Swedish case. Also, the naive forecasts has higher difference in cost than the rolling average as it did for all the Swedish cases except model 2.

5

Conclusions

From the result and discussion above there are some conclusions that now can be drawn. The conclusions aim to answer the research questions.

The performance of the forecasts does shift between the different models with different units, but there is a clear winner: the regression forecast that outperformed the other forecasts in terms of errors and production costs throughout all of the models and data sets. This states that the forecast with the lowest errors, results in the best production plan. The rolling average forecast resulted in the lower cost than the naive forecast for model, but when TES was added this shifted.

Large differences in production costs occurred around a limit in electricity price, if the forecast predicted a price below this limit the result was production of heat in the heat pump. This only occurred at low electricity prices. If other forecasts predicted prices over the limit, the heat pump would not be part of the mix and a price difference occur. This is where the correct forecast matter, if the forecasts can predict high and low electricity prices at the right time, then the production plan will run the most profitable unit and can avoid high production costs. This also answers the question of when forecast perform well and why, periods with low electricity prices (mostly summer) resulted in larger differences in production costs. This also applies when an oil boiler was connected to the system. Since the fuel price is high compared to regular electricity prices it will highly affect the outcome of the total production cost.

The forecasts has proved that timing of the peaks is more important than reaching the altitude of the peak. This might be why the naive forecast was improved when the rolling average was performed on it. Off-peak altitudes can be of more importance since low electricity prices were decisive in the choice of which unit to operate. The different forecast performs differently on different kinds of events.

Using TES decreased the total production cost remarkably. It also resulted in larger differences in production costs between the forecasts. TES therefore makes the system more flexible but also more dependent on the electricity price forecast.

To obtain more results regarding model 3 with the electric boiler, more efforts on how to design the system in terms of deciding the size of the units and other important parameters must be made in future work. This also applies to the results regarding the other peak-load unit (bio-oil boiler).

The comparison between the Swedish and the Danish system with more renewables resulted in higher production costs for the Danish system. Fluctuations in price made it harder for the benchmark forecasts to predict the electricity price. ML-forecasts might have a larger role to play in predicting the electricity price in a system with more renewable energy since they performed better overall. Future work should investigate how the ML-forecasts perform on this kind of data.

The electricity price has a clear correlation with external factors. Seen from the sensitivity analysis, the world events occurring will affect the prices, which in turn affect the production cost. From year 2020, the irregular events drastically disturbs the forecast methods that had trouble with prediction and results in higher differences in total production cost.

Bibliography

[1] "District heating," *Swedish energy agency*, 2015. [Online]. Available: https://www.energimyndigheten.se/en/sustainability/households/heating-your-home/district-heating/.

[2] S. Werner, "District heating and cooling in sweden," 2017. [Online]. Available: (http://www.sciencedirect.com/science/article/pii/S0360544217304140).

[3] A. Elnaz, "Optimization of energy production of a chp plant with heat storage," 2015. [Online]. Available: https://www.researchgate.net/publication/281735295_Optimization_of_energy_production_of_a_CHP_plant_with_heat_storage/.

[4] M.Negrete-Pinetic, "The value of volatile resources in electricity markets," *ScienceDirect*, July 2017. DOI: https://doi.org/10.1016/j.segan.2017.07.001.

[5] K. Vajapey, "What's the difference between ai, ml, deep learning, and active learning?" *datanami*, 17 sep 2019. [Online]. Available: https://www.datanami.com/2019/09/17/whats-the-difference-between-ai-ml-deep-learning-and-active-learning/#:~:text=Deep%5C%20learning%5C%20is%5C%20a%5C%20more,improved%5C%20version%5C%20of%5C%20the%5C%20AI..

[6] Energiföretagen. (February 2021). "Kraftvärme," [Online]. Available: https://www.energiforetagen.se/energifakta/kraftvarme/. (accessed: 13.05.2021).

[7] Celsius, "Thermal energy storage," *Celsius*, 17 august 2020. [Online]. Available: https://celsiuscity.eu/thermal-energy-storage/#:~:text=Thermal%5C%20energy%5C%20storage%5C%20(TES)%5C%20is%5C%20one%5C%20form%5C%20of%5C%20energy%5C%20storage.&text=TES%5C%20can%5C%20also%5C%20reduce%5C%20peak,is%5C%20in%5C%20solar%5C%20thermal%5C%20systems..

[8] Ucsusa. (2015). "How energy storage works," [Online]. Available: https://www.ucsusa.org/resources/how-energy-storage-works#:~:text=When%5C%20demand%5C%20is%5C%20greater%5C%20than,later%5C%20when%5C%20it%5C%20is%5C%20needed.. (accessed: 22.05.2021).

[9] H. Zheng, "Coefficient of performance," *ScienceDirect*, 2017. DOI: https://doi.org/10.1016/B978-0-12-805411-6.00008-7.

[10] N. J.Hong, "Assessing heat pumps as flexible load," *Power and Energy*, February 2013. DOI: 10.1177/0957650912454830.

[11] G. Match. (2021). “Electric boilers: An in-depth guide,” [Online]. Available: https://www.greenmatch.co.uk/boilers/electric. (accessed: 04.05.2021).

[12] HVP. (June 2020). “What role can bio-oil lay in the future of home heating?” [Online]. Available: https://www.hvpmag.co.uk/What-role-can-bio-oil-play-in-the-future-of-home-heating/11929. (accessed: 24.05.2021).

[13] Viessmann, “How does an oil fired boiler work?” *Viessmann*, [Online]. Available: https://www.viessmann.co.uk/heating-advice/how-does-an-oil-fired-boiler-work, (accessed: 24.05.2021).

[14] Å. Marbe, “Biofuel gasification combined heat and power—new implementation opportunities resulting from combined supply of process steam and district heating,” *ScienceDirect*, DOI: https://doi.org/10.1016/j.energy.2004.01.005, (accessed: 24.05.2021).

[15] Energimarknadsinspektionen, “Elmarknader och elhandel,” 2016. [Online]. Available: https://www.ei.se/sv/for-energikonsument/el/Elmarknader-och-elhandel/.

[16] K. Lindholm, “Negativa elpriser - energiföretagen förklarar,” 2020. [Online]. Available: https://www.energiforetagen.se/pressrum/nyheter/2020/februari/negativa-elpriser--energiforetagen-forklarar/.

[17] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *ScienceDirect*, 18 September 2018. DOI: https://doi.org/10.1016/j.ijforecast.2014.08.008.

[18] Nordpoolgroup. (2021). “See what nord pool can offer you,” [Online]. Available: https://www.nordpoolgroup.com/. (accessed: 08.02.2021).

[19] E. A. M. Mulaosmanovic, “Short-term electricity price forecasting on the nord pool market,” *Diva poratl*, [Online]. Available: http://mdh.diva-portal.org/smash/get/diva2:1169409/FULLTEXT01.pdf, (accessed: 25.05.2021).

[20] A. Vidhya. (April 2020). “Benchmarking methods for time series forecast,” [Online]. Available: https://medium.com/analytics-vidhya/benchmarking-methods-for-deep-learning-based-time-series-forecast-ec45f78b61e2. (accessed: 13.05.2021).

[21] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, DOI: https://doi.org/10.1016/j.ijforecast.2014.08.008, (accessed: 24.05.2021).

[22] D. David. (August 2020). “Random forest classifier tutorial: How to use tree-based algorithms for machine learning,” [Online]. Available: https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/. (accessed: 13.05.2021).

[23] H. Ritchie, “Electricity mix,” *Our world in data*, [Online]. Available: https://ourworldindata.org/electricity-mix, (accessed: 18.05.2021).

[24] B. Chachuat. (2020). “Mixed-integer linear programming (milp): Model formulation,” [Online]. Available: http://macc.mcmaster.ca/maccfiles/chachuatnotes/07-MILP-I_handout.pdf. (accessed: 27.05.20).

[25] M. Demirian, “Sweden energy market energy tax on electricity,” *NUS consulting group*, [Online]. Available: https://www.nusconsulting.com/energy-blog/sweden-energy-market-energy-tax-on-electricity#:

~:text=All%5C%20electricity%5C%20consumed%5C%20in%5C%20Sweden,%5C%2C6%5C%20Swedish%5C%20cents%5C%2FkWh., (accessed: 24.05.2021).
[26] D. E. Agency and Energinet, "Hot water tanks," *Technology Data - energy storage*, p. 53, [Online]. Available: http://www.ens.dk/teknologikatalog.
[27] Bostadsrättnytt. (February 2019). "Elpriset i januari – det högsta på åtta år," [Online]. Available: https://bostadsrattsnytt.se/nyheter/energi-och-miljo/2019-02-06-elpriset-i-januari-det-hoegsta-pa-atta-ar. (accessed: 21.05.2021).
[28] SMHI. (2019). "Sommaren 2019 - varmare än normalt," [Online]. Available: https://www.smhi.se/klimat/klimatet-da-och-nu/arets-vader/sommaren-2019-varmare-an-normalt-1.150462. (accessed: 20.05.2021).
[29] E. Rydegran, "Rekordlåga elpriser sammanfattar elåret 2020," *Energiföretagen*, [Online]. Available: https://www.energiforetagen.se/pressrum/pressmeddelanden/2020/rekordlaga-elpriser-sammanfattar-elaret-2020/, (accessed: 25.05.2021).
[30] K. Lundin, "Elpriset rusar i södra sverige," *Sydsvenskan*, 2019. [Online]. Available: https://www.sydsvenskan.se/2020-06-25/elpriset-rusar-i-sodra-sverige, (accessed: 25.05.2021).

A

Appendix

A.1 Lag-features used in the modelling of ml-forecasts

Table A.1: Table containing all the features used in the ML-modelling.

$price_{lasthour}$
$price_{lasttwohours}$
$price_{lastthreehours}$
$price_{lastday}$
$price_{lasttwodays}$
$price_{lastthreedays}$
$price_{lastweek}$
$price_{lasttwoweeks}$
$price_{lastthreeweeks}$
$price_{lastfourweeks}$
$price_{meanlastfourweeks}$
$price_{meanlasteightweeks}$
$T18(t)$
$T18(t)_{last_hour}$
$T18(t)_{last_two_hours}$

A.2 Historical electricity spot prices

In this section is some historical electricity spot data presented. From year 2017 to year 2020. It makes it more clear how much the prices vary depending on external occasions, also some small patterns are visible to see from year to year.

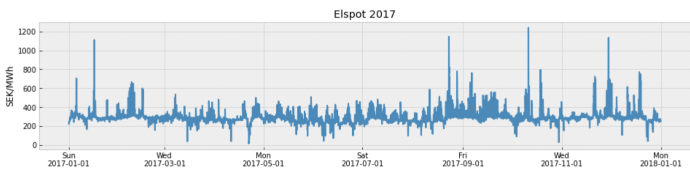


Figure A.1: The electricity spot price for year 2017.

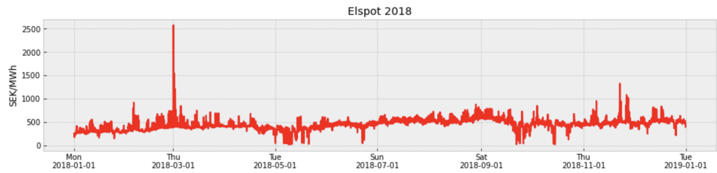


Figure A.2: The electricity spot price from year 2018.

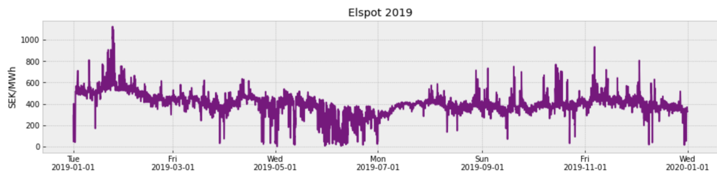


Figure A.3: The electricity spot price from year 2019.

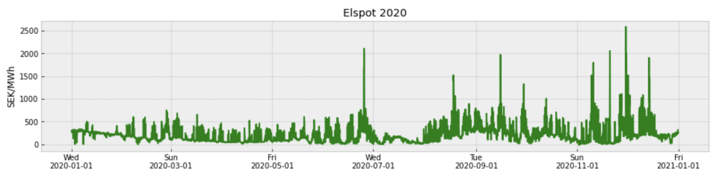


Figure A.4: The electricity spot price from year 2020.

A.3 Production cost and differences

In this section are the values of all forecasts in all four models presented. This is to see the values clearly and the differences from model to model and from forecast to forecast.

Table A.2: The total production costs for the reference and the four forecasts and the difference between the reference and the forecasts for model 1.

Forecast	Total production cost [SEK]	Difference in cost
Reference	19 820 656	0
Naive	20 602 239	+781 583
Rolling average	20 419 934	+599 278
Regression	20 211 512	+390 856
Random forest	20 304 677	+484 021

Table A.3: The total production costs for the reference and the four forecasts and the difference between the reference and the forecasts for model 2.

Forecast	Total production cost [SEK]	Difference in cost
Reference	17 886 504	0
Naive	18 954 213	+1 067 709
Rolling average	20 116 951	+2 230 447
Regression	18 362 899	+467 395
Random forest	18 912 630	+1 026 126

Table A.4: The total production costs for the reference and the four forecasts and the difference between the reference and the forecasts for model 4.

Forecast	Total production cost [SEK]	Difference in cost
Reference	19 887 572	0
Naive	20 691 515	+ 803 943
Rolling average	20 510 475	+622 903
Regression	20 261 663	+374 091
Random forest	20 304 677	+417 105

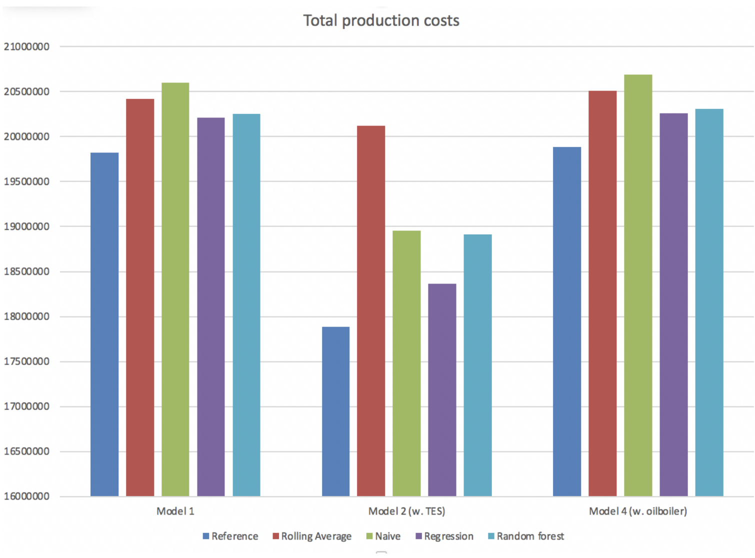


Figure A.5: The total production costs for three of the models.