NTNU
Norwegian University of
Science and Technology

Volume and Price in The Nordic Balancing Power Market

Polina Pires Ferreira

Master of Energy and Environmental Engineering
Submission date: June 2016
Supervisor: Magnus Korpås, ELKRAFT
Co-supervisor: Eirik Mo, Statkraft

Norwegian University of Science and Technology
Department of Electric Power Engineering

NTNU
Norwegian University of
Science and Technology

Volume and Price in the Nordic Balancing Power Market

Polina Pires Ferreira

Energy and Environmental Engineering
Sudmision date: June 2016
Superviser: Magnus Korpås, ELKRAFT

Norwegian University of Science and Technology
Department of Electrical Power Engineering

Abstract

The main condition for the normal functioning of the power system is the balance between the power production and the power consumption at any moment. The increasing amount of power that comes from the renewable sources of energy such as wind and sun, and more variable and uncertain power consumption have lead to that the imbalances in the power system occur more frequently.

If an imbalance occurs in the operating hour, it will be dealt with by activating the balancing power reserves. The activation of the reserves will cause an extra cost to the market actors that are responsible for the imbalance. The market actors have also the possibility to offer their production capacity to the Balancing power market, and gain an extra income if their bids will be activated. That is why having an accurate forecast of the balancing power volume and the balancing power price for an operating hour could have been an advantage for the market actors in order to reduce their imbalance costs and to possible gain an extra income.

In this master thesis the possibility of forecasting the balancing power volume and the balancing power price by using the Machine Learning algorithms has been examined. The Boosted Decision Tree regression model and the Decision Forest regression model provided by the Microsoft Azure Machine Learning Studio have been used to obtain the forecast of the balancing power volume and the balancing power price for an operating hour in the price area NO3 day ahead and closer to the operating hour.

The results obtained under this work have shown that it is impossible to get an accurate forecast of the balancing power volume without using the past values of the balancing power volume as one of the predictors, and without taken into consideration the influence of the events, which have occurred outside of the NO3, on the activation of the balancing power reserves in the NO3. An accurate enough forecast of the balancing power volume can be made a couple of hours ahead when using the past values of the balancing power volume as an explanatory variable. However, when taken into account the situation in the other price areas and forecasting the total imbalance in price area NO3, a forecast with a high accuracy can be conducted day ahead. The balancing power price can also be forecasted with a good accuracy, but an accurate forecast of the balancing power volume is required.

Sammendrag

Balansen mellom kraftproduksjonen og kraftforbruket til enhver tid er den viktigste forutsetningen for at kraftsystemet skal fungere uten forstyrrelser. Økende mengde av kraft som kommer fra fornybare energikilder, og mer variabel og usikker strømforbruk har ført til at flere ubalanser oppstår i kraftsystemet.

Ved en ubalanse i driftstimen, vil et behov for aktivering av regulerkraftreserver oppstå. Aktivering av reserver vil føre til en ekstra kostnad for markedsaktørene som er ansvarlige for ubalansen. Markedsaktørene har også mulighet til å tilby sin produksjonskapasitet i Regulerkraftmarkedet, og få en ekstra inntekt dersom deres bud vil bli aktivert. Derfor vil en nøyaktig prognose av regulerkraftvolumet og regulerkraftprisen i en driftstime være til en fordel for markedsaktørene med tanke på å redusere deres ubalansekostnader og å skaffe en ekstra inntekt.

I denne masteroppgaven har muligheten til å forutse regulerkraftvolumet og regulerkraftprisen ved hjelp av Machine Learning algoritmer blitt undersøkt. Boosted Decision Tree regresjonsmodellen og Decision Forest regresjonsmodellen fra Microsoft Azure Machine Learning Studio har blitt brukt til å lage prognosen av regulerkraftvolumet og regulerkraftprisen i prisområdet NO3 en dag fremover og noen timer før den faktiske driftstimen.

Resultatene som har blitt oppnådd under arbeidet, har vist at det er umulig å skaffe en nøyaktig prognose av regulerkraftvolumet for en time uten å ha regulerkraftvolumet fra tidligere timer som en av forklaringsvariablene, og uten å ta hensyn til hendelser som har skjedd utenfor NO3. En god nok prognose av regulerkraftvolumet kan bli skaffet et par timer før den faktiske driftstimen dersom regulerkraftvolumverdiene fra tidligere timer er brukt som en av forklaringsvariablene. Imidlertid, ved å ta hensyn til situasjonen i de andre prisområdene og å forutse den totale ubalansen i prisområdet NO3, kan en nøyaktig prognose for et døgn i forveien, bli skaffet. Regulerkraftprisen kan også bli forutsett med en god nøyaktighet, men en nøyaktig prognose for regulerkraftvolumet er påkrevd.

Acknowledgments

I would like to express my gratitude to my supervisor, Professor Magnus Korpås, for his guidance, encouragement and support throughout the work with this master thesis. I would also like to thank all of the employees that I have met and worked beside at Stakraft. And especially I would like to thank Eirik Mo and Mads Vilhelm Lindsjørn at Statkraft for all guidance, interesting discussions and help during the work.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

In 2007 EU legislated a so-called 20-20-20 target, which implies a 20% reduction of greenhouse gas emissions, 20% reduction of the power consumption and a 20% increase of the use of energy from the renewable sources.

The national targets of increasing the share of renewable energy vary for the different countries. In Norway, Sweden, Denmark and Finland the share of the renewable energy in each of the countries gross final energy consumption should be, respectively, 67.5%, 49%, 30% and 28% by 2020. (European Commission, 2016) Figure 1.1 shows the expected installed electricity production capacity from the different energy sources in some European countries in 2020. As it can be seen from the figure the electricity production capacity from the renewable energy sources will increase in many countries considerably.

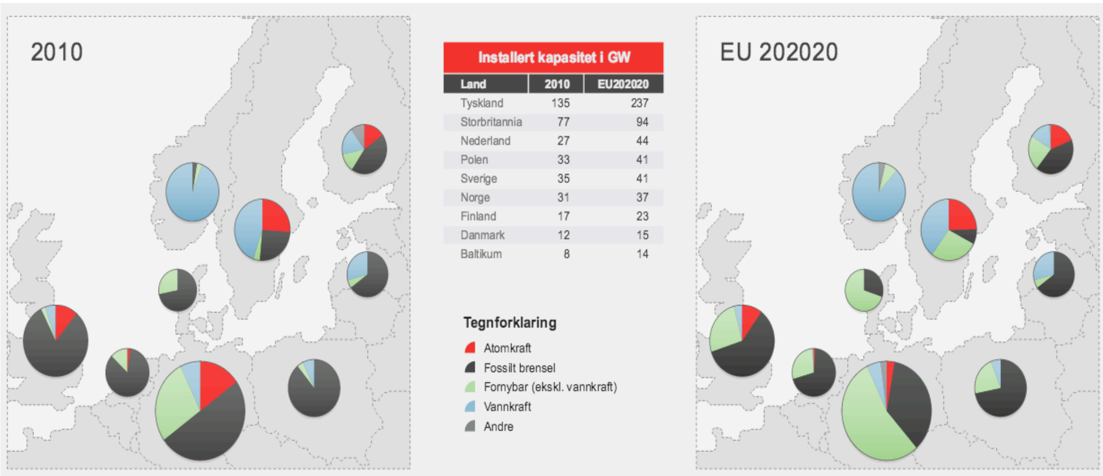


Figure 1.1. Expected installed production capacity in European countries in 2020. (Statnett, 2014)

Most of the power production from the renewable sources is variable and uncontrollable. This increases the value and the demand for the power production that is suitable for up- and down-regulation in the power system.

At the same time the power consumption has become more variable and uncertain due to a different reasons, such as implementation of measurers for improving energy efficiency of the buildings, integration of new technologies (electrical cars and solar panels) in daily life, and the climate change. More variable power production and consumption will lead to a greater need to balance the power system.

Activation of the balancing power reserves imposes an extra cost on those who is responsible for the imbalances in the power. Nevertheless, the power producers have a possibility to gain income when providing the balancing power reserves to the balancing power marked for up- or down-regulation. Since more frequently imbalances in the power system can both bring an extra expense and an extra income to the power producers, it is useful to know in advance about, if and when, the balancing power reserves will be activated and for which price. Knowing this, the power producers can decide how they will handle their imbalances and if it is beneficial for them to offer any production capacity to the balancing power marked.

1.2 Problem definition

Statkraft has a leading position in Europe when it comes to the power production from the renewable sources of energy. The core activity of Statkraft is mainly hydropower production and wind power production. The company has a big share of the controllable hydropower production, some uncontrollable hydropower production and some wind power production. The company has also plans to build out 1000 MW of wind power in Trøndelag (in price area NO3) in the upcoming years.

With increasing amount of wind power production in Statkraft's production portfolio, the company's expenses related to causing the imbalances in the power system will increase. This is one of the reasons why the Balancing power market has become increasingly more interesting for the company. The big share of the hydropower production in the company's production portfolio, gives Statkraft a high flexibility and a production that is suitable for a quick up- and down-regulation, which the company wishes to utilize on the most cost-efficient way.

In order to reduce the cost associated with causing the imbalances in the power system and the possibility of gaining an extra income from trading in the Balancing power market, having a robust forecast for the balancing power volume and the balancing power volume is an advantage. That is why Statkraft has an interest of investigating the possibility of obtaining an accurate forecast of the balancing power volume and the balancing power price.

The aim of this master thesis is to examine the possibility of making forecasts for the balancing power volume and the balancing power price by using a set of input factors/variables. Following subtasks are included:

1. Conduct a statistical analysis of the balancing power volume and the balancing power price. Investigate the relationship between the balancing power volume and the factors that influence the balance in the power system.
2. Conduct a statistical analysis of the balancing power price. Investigate the relationship between the balancing power price and the factors that can have an influence on it.

3. After identifying the factors that can indicate an up- or a down-regulation in the power system and the factors that affects the balancing power price, make a forecast of the balancing power volume and the balancing power price for the price area NO3 by using algorithms provided by the Microsoft Azure Machine Learning Studio and the factors mentioned above as a model input.
4. Evaluate and discuss the results.

1.3 Literature study

Before starting the work with the developing of the forecasting model for the balancing power volume and the balancing power price, the literature concerning the topics about forecasting the balancing power volume and the balancing power prices has been studied in order to get an overview of it.

All existing models can by distinguished between the models, which model the regulation states first and uses it as an input variable to forecast the regulating prices and the models that do not take into account the regulating volume forecasting the regulating prices.

Klæboe, Eriksrud & Fleten (2015) have formulated five different forecasting models in order to examine the possibility of forecasting the balancing power price (premium), which are based on the models that have earlier been developed by Jaehnert, Farahmand & Doorman (2009), Olsson and Soder (2008), Boomsma, Juul & Fleten (2014) and Conejo et al (2005).

Jaehnert, Farahmand & Doorman (2009) have developed a price-forecasting model that consists of a short-term model based and a long-term model. The short-term model determines the regulation state and the volume by using the SARIMA process. Further he uses statistical description of the regulating power volumes as an input variable into the long-term model where the linear relationship between the regulating volumes and the regulating prices is utilized to generate future price scenarios.

Olsson and Soder (2008) used the Markov model to determine the regulation state and the SARIMA process for forecasting the Regulating power market premium for either an up- or a down-regulation depending on the forecasted state.

Boomsma, Juul & Fleten (2014) have developed a model that predicts the regulating prices without taken into account the expected state of regulation. Boomsma, Juul & Fleten (2014) use in their forecasts an autoregressive time series and an external input in form of the spot prices.

Gro Klæboe et al. (2015) have concluded their work with that the balancing power volume and the balancing power premium in the Balancing power market are random.

However, they have pointed out that even though the models formulated in the work, it have not managed to forecast the balancing power premium precisely. It does not mean that the Balancing power market forecasting is futile, and that the models that take into consideration the expected regulation state, give a far better explanation of the future balancing power volume and the future balancing power premium, than the models without the balancing state information.

Chapter 2

The Nordic Power Market

2.1 The Nordic Power System

The power systems of Norway, Sweden, Finland and Denmark, except Jylland, create one synchronised Nordic Power System. Jylland is synchronised with the continent and is connected to the Nordic Power system via connections with Norway, Sweden and East Denmark. (Statnett, 2014)

The Nordic Power System can be defined as a hydrothermal power system, where share of hydropower production in the annual Nordic power production is 61%, share of thermal production is 30% and the remaining 10% is wind power production.

The Norwegian power production is dominated by hydropower, covering 96% of the annual power production in Norway. In Sweden, hydropower production make up around 52% and thermal power production constitutes about 39% of the annual Swedish power production. The Finish power production is dominated by thermal power, at 69%, the remaining 31% is divided between hydropower production, at 24%, and other types of production. Due to environmental policy, feed-in tariffs and investment incentives that makes it possible for expansion of power production from renewable sources of energy, the Danish power system, in contrast to the other Nordic countries, has a high share of wind power production, which is at 49%. The remaining production is covered by thermal power.

The annual production of 2015 per production type for each of the different Nordic countries is represented in the table 2.1. The information in this table is conducted from the ENTSO-E Transparency Platform.

Type of production	Actual production per production type per country [TWh]				Total production per production type [TWh]
	Norway	Sweden	Denmark	Finland	
Hydro	132.5	70.6	0.02	15.5	218.62
Nuclear	-	47.3	-	22.2	69.5
Other thermal	2.9	5.5	13.7	22.1	36.5
Wind	2.3	13.4	13.6	2.1	31.4
Other	0.3	-	0.7	2.1	3.1

Table 2.1. Annual production per production type of the year 2015 for each of the Nordic countries.

From table 2.1 one sees that the Nordic power system is dominated by the hydropower production that gives high production flexibility in combination with relatively low costs. It gives the possibility to store water in the reservoirs over the days and seasons, which enables them to produce the power when the price is high. The decision of the producers to produce or not will depend on the demands for electricity and current, and the expected reservoir levels. In the power system where the share of the uncontrollable production from renewable sources of energy, such as wind and solar, steady increases, the role of the hydropower production is crucial when it comes to the quick adjustment of the production in order to cover the electricity demand and to keep balance in the system.

2.2 The Nordic Power Market

The Nord Pool Spot is the Nordic market place where trade of physical contracts for electricity delivery takes place, owned by the Nordic and Baltic TSOs. The Nord Pool Spot runs Elspot Market, Elbas Market and N2EX Market and offers both the day-ahead and intraday trading to the customers. The N2EX Market applies to the UK, and is therefore not discussed in this paper. Power producer, Transmission System Operators (TSOs), power intensive industries, large consumers and power companies actively participates in the power markets and are defined as power market actors.

Elspot is a day-ahead market where the market actors enter into contracts for physical delivery of power for the next day. Within 12 a.m. the day before the actual delivery, market actors submit their bids to the market where they provide information about amount of how much power and at which price that they are willing to sell/purchase for, at each hour of the following day. At the market clearing the spot price for each price area will be determined based on the electricity demand and the electricity supply and transmission capacities between the areas. In addition, the system price will be calculated under the assumption that there are not any limitations in the transmission capacity between the price areas. The system price is used as a reference for trading in the Financial market.

Between the clearing of the Elspot and the actual delivery of the power, the consumption and the production can change so that an imbalance in the power system occurs. In order to adjust the imbalance, the market actors can take an advantage of using the Elbas market.

Elbas market is an intraday market in Norway, Sweden, Denmark, Finland, Germany, Latvia, Lithuania, Estonia, the Netherlands, the UK and Belgium where power producers and power consumers continually can adjust their imbalances up to one hour before the actual power delivery. As trading in the Elbas is continually, no clearing price will be establish meaning that trading in the Elbas works in the same way as any sales or purchases in a regular stock market. (Bye et al, 2010)

The TSOs have the responsibility for adjusting the imbalances in the power system that occurs in the operating hour. To obtain the reserves that are required in order to balance the system, the TSOs have established several markets including the Balancing power market and the Balancing power option market.

Participation in the Balancing power market is voluntary making it in some periods important for the TSOs to ensure sufficient reserves in the Balancing power market. The TSOs in the aforementioned countries have different ways to obtain necessary reserves. The Norwegian system operator, Statnett, uses the Balancing power option market in order to ensure enough reserves for the up-regulation in the Norwegian part of the Balancing power market in a certain period of time, mainly during the winter months. Statnett assesses the need for the reserves on the basis of the current energy situation, forecasts for consumption, production, potential bottlenecks and power exchanges with foreign countries. This is the reason why Statnett enters into contracts with power producers and consumers as a guarantee of their participation in the Balancing power market in a certain time period. (Statnett, 2013) Suppliers of the reserves get paid a certain price for providing the reserves in the disposal of Statnett, in addition to the price determined in the Balancing power market for the amount of energy that eventually may be used. The capacity that is accepted in the Balancing power option market can because of this no longer be offered in the Elspot.

Market participants uses the Financial market at Nasdaq OMX for risk hedging and risk management where they have the possibility to trade financial products (contracts) for up to six years in the future. The contracts entered in this market does not imply the physical delivery of power, meaning that any technical limitations in the power grid is not taken into consideration during the trading. The following financial products can be purchased in the market: Futures, Forwards, Electricity Price Area Differentials (EPAD) and options. (Olje- og energidepartementet, 2014)

Futures and Forwards are contracts that hedge the difference between the contract prices and the variations in the system price compared to the settled amount of power for a certain period of time. Futures can be settled in both the trading and the delivery period, while the settlement of the Forwards occurs only in the delivery period. Electricity Price Area Differentials are Forwards contracts, which provide the opportunity to hedge either the difference between the system price and area prices or spatial differences between area prices. (Wangensteen, 2012) Options sold at NASDAQ OMX are European options conferring rights to purchase or sell Forwards in the future for an agreed price. (Olje- og energidepartementet, 2014)

The figure 2.1 below shows the structure and the functioning of the Norwegian power market.

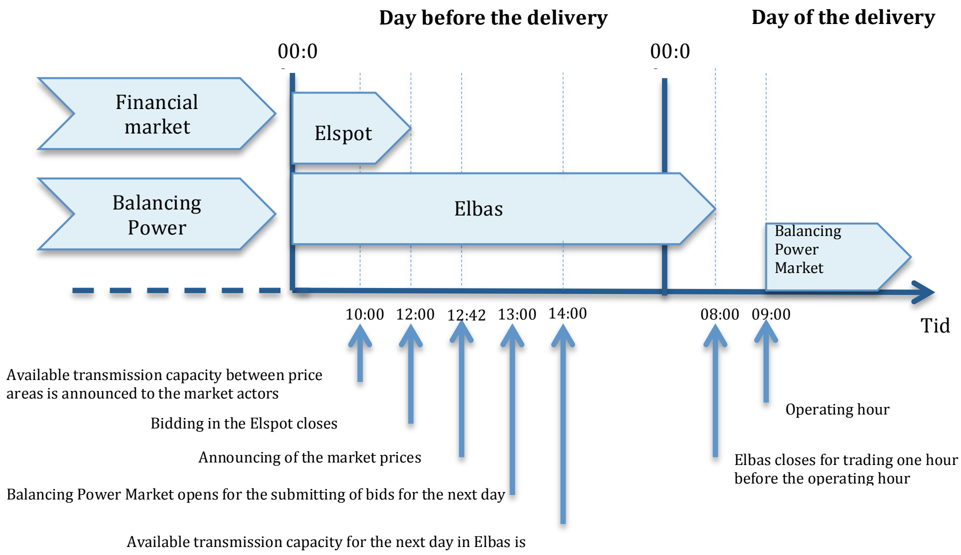


Figure 2.1. Power market structure in Norway.

2.3 Congestion management

The Nordic region is divided into 12 price areas that reflect the presence of physical limitations in the transmission capacity in the transmission grid. In order to achieve a secure operation of the power grid and to make use of the power resources on the most cost efficient way, the transmission capacity limitations between areas have to be taken into consideration at the power market clearing. This will result into different prices in different areas.

Congestion occurs if the market efficient power flow exceeds transmission capacity between the price areas. There are many different ways that the TSOs can use, in order to manage congestion between the areas. The Nordic TSOs use the market splitting to handle the bottlenecks. This method is based on splitting the power exchange into geographical areas that have a limited exchange capacity. Nord Pool Spot performs the market splitting. First the market price or the system price is found by calculating the supply and demand in the power exchange. Then the TSO will calculate the necessary power flow and determine bottlenecks in the system. If the bottlenecks are identified between the areas, a new spot price for each area will be calculated. (Wangensteen, 2012)

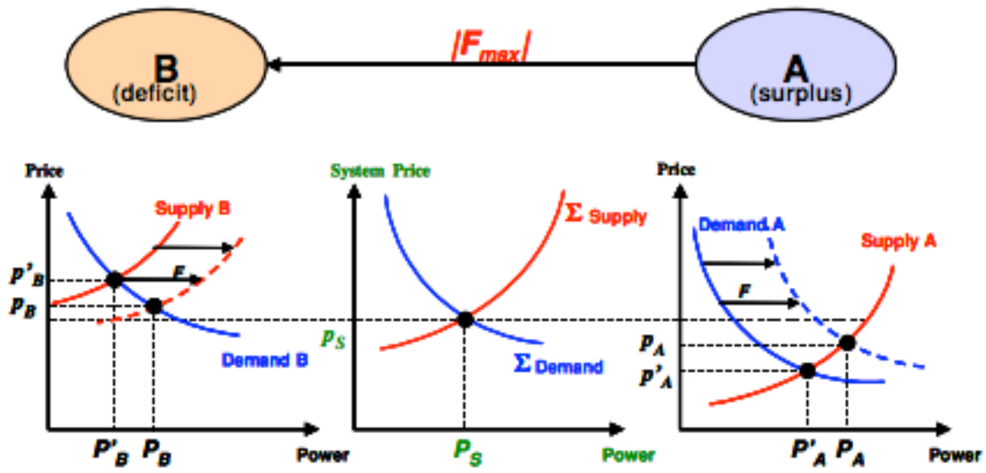


Figure 2.2. Market splitting. (T.Kristiansen, 2004)

The figure 2.2 shows an example of using the market splitting for two areas with insufficient exchange capacity. Area A is a surplus area and area B is a deficit area, which means the surplus power can be exported from area A to area B. This flow will lead to a decreased power price in area B and an increased power price in area A, and, if there are no limitations in exchange capacity between area A and B, the power price in both areas will be equal to the system price. With a limited exchange capacity the areas will still have two different prices: area A will have a lower price than area B.

2.4 Balancing the power system

The TSOs have the responsibility for keeping balance in the power system and restoring it, in case of which an unexpected event occurs. As mentioned before the market participants have the possibility to adjust their imbalances by trading in the Elbas, up to one hour before the actual power delivery. The immediate changes in the operating hour will be handled by the TSO. In this chapter the control mechanisms for balancing the power system used by the Norwegian TSO, Statnett, will be described.

Statnett has three levels of control reserves:

1. Primary or Frequency Containment Reserves that are used in order to deal with the system imbalances instantly by automatic activation of the production.
2. Secondary or Automatic Frequency Restoration Reserves that are used to deal with the system imbalances and bring the system frequency closer to the nominal value. Activation of the Secondary Reserves also releases the Primary reserves so that it can handle the new imbalances.

3. Tertiary Reserves or the Balancing Power Market is used for the release of the Primary and Secondary Reserves and for dealing with the regional bottlenecks.
4.

From the figure 2.3 that shows the principles for the activation of the reserves, one sees that if an imbalance occurs in the operating hour, Statnett will begin to activate its balancing reserves in order to restore the nominal frequency in the system. First the Primary reserves activate automatically and run up to two minutes before the Secondary reserves will be activated. After about 15 minutes with an imbalance in the system, Statnett will manually start to activate its Tertiary Reserves, so that the Secondary Reserves will be released.

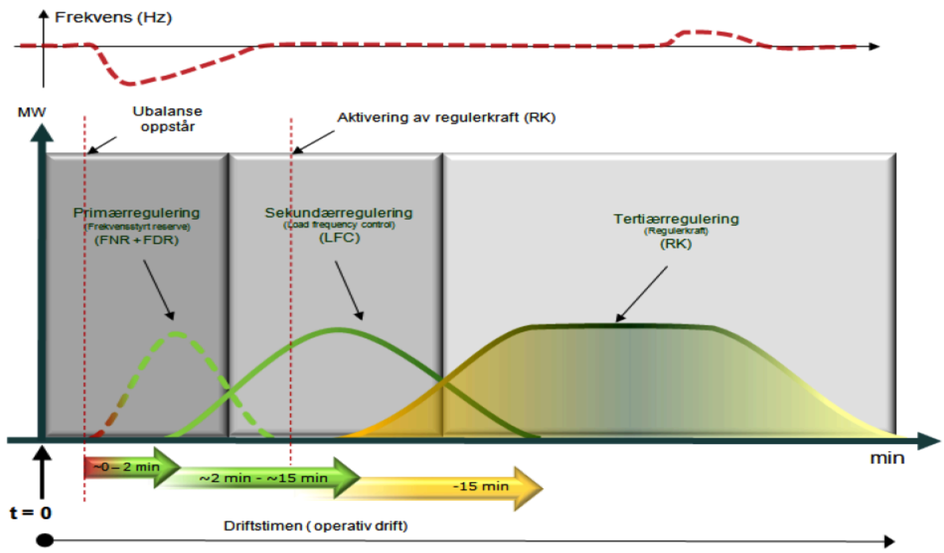


Figure 2.3. Principles for the activation of the reserves. (Bye et al, 2010)

2.5 The Balancing Power Market

The Balancing power market is a physical power market that is used by the Nordic TSOs to ensure frequency stability in the power system and balance between the power production and the power consumption in the operating hour. The Norwegian TSO, Statnett and the other Nordic TSOs drive this market, which are Svenska Kraftnät in Sweden, Energinet in Denmark and Fingrid in Finland.

The Balancing power market opens for submission of the balancing power bids for the next day at 1 p.m. the day before. At this moment the power producers get to know which one of their bids that got excepted in the day-ahead market and the spot price in the different price areas, so they can assess whether or not they have any production to offer in the Balancing power market. Only the producers that can adjust their production within 15 minutes notice, and large power consumers that can reduce their consumption at a short times notice can participate in this market. (Statnett, 2015)

Market participants submit their reserves for both up- and down-regulation separately. In the bids, they specify how much power and at which price it is offered for regulation in each hour of the following day. Preliminary bids have to be submitted before 9:30 p.m. the same day. Market participants can also change their bids or submit more bids, but no later than 45 minutes before the operating hour. (Statnett, 2015) All submitted bids gathers into a common list in the price merit order so that the TSOs can exchange their balancing reserves between each other.

The bids for an up- or down-regulation will be activated in the case of an imbalance in the power system that varies longer then 15 minutes. If the actual consumption is greater than the power production in the operating hour, a need for an up-regulation in the system will occur. Which means that the power producers will have to increase their production or the power consumers will have to decrease their consumption. If the actual power consumption is less than the power production in the operating hour, the system has to be down regulated, by decreasing the production. In principle, the consumption can also be decreased, but this is hardly relevant.

The balancing power volumes and prices will be announced at the end of the operating hour. In case the direction of the regulation changes in the same operating hour, the dominating direction will be determined based on the net regulated volume in the operating hour. In hours when no regulation occurs the balancing power volumes will be set to zero.

2.4 Pricing of the activated balancing power

Before offering reserves in the Balancing power market the market participants evaluates possible start and stop costs, maintenance costs etc. that can incur if their bids will be activated. The power consumers consider the costs that can be caused by the reduction of the consumption. The reduction of the consumption at a short notice is very expensive, so the TSOs use the consumption side for up-regulation only under certain conditions. (Bye et al, 2010)

If regulation is needed, the TSOs start to activate bid after bid in the Balancing Power Market. The last activated bid in the operating hour will determine the price of the balancing power in this hour. In the hour when an up-regulation takes place, the most expensive bid that has been activated will determine the balancing power price in this hour. In the hour with a down-regulation the cheapest activated bid will determine the price of the regulation.

The balancing power price under an up-regulation is usually equal or higher than the spot price in a price area, and it is equal or lower than the spot price in a price area under a down-regulation. The reason for this is that for an up-regulation, the power producers offer production, which has not been accepted in the Elspot, i.e. production with higher marginal cost than the spot price in a price area. So the price-determining unit in the Elspot market is the cheapest unit in the Balancing power market if an up-regulation occurs, in the power system. For a down-regulation, the power producers offer production units that are already in use, i.e. the units that have been accepted in the Elspot market. So the price-determining unit in the Elspot is the most expensive unit in the Balancing Power market under a down-regulation of the power system. It is cost efficient to reduce the production of the Elspot's price-determining unit first.
The balancing power price for each price area is determined at the end of each hour. In hours when no regulation takes place, the balancing power price is equal to the spot price of a price area. (Statnett, 2015)

The market actors that have caused an imbalance will pay for the balancing reserves that have been activated, in order to restore balance in the power system. The Norwegian TSO, Statnett, uses two different strategies when charging producers and consumers for their imbalances.

The table 2.2 shows how the power consumers and producers will be priced for causing imbalances in the power system. The power consumers will be charged with a one-price system, which means that they will always pay or get paid the balancing power price.

Statnett uses a two-price system in order to price imbalances from the power producers, so that any deviation in the power production in the opposite direction to the regulation direction will be penalized. If the power production exceeds the amount of power, which has been determined in the Elspot or Elbas market, and there is a need for an up-regulation in the system, the power producer will be paid the spot price for their surplus power. If the power producers produce less, they will have to pay the balancing power price in order to cover their imbalance.

It there is a need for a down-regulation in the system and the producers produce more power than they are supposed to, they will get paid the balancing power price. While producers that produce less will have to pay the spot price. Such system gives the power producers an initiative to remain in the balance. (Statnett, 2015)

System / Actor	Deficit – up-regulation	Surplus – down-regulation
Produces too much	Will be paid the spot price	Will be paid the balancing power price
Produces too little	Will be charged the balancing power price	Will be charged the balancing power price
Consumes too much	Will be charged the balancing power price	Will be charged the balancing power price
Consumes too little	Will be paid the balancing power price	Will be paid the balancing power price

Table 2.2 Pricing the deviation between the Elspot/Elbas obligation and actual power production or power consumption. (Bye et al, 2010)

2.6 Special regulation

The limitations in the transmission capacity in the central and regional grid, internal bottlenecks in a price area or fault situations, will be taken care of with by using the special regulation. Under the special regulation, Statnett activates bids from the common balancing power list. However, number of bids that can be used will be reduced due to limitations in the transmission capacity, so that Statnett will not always be able to activate the bids according to the price order. Statnett will cover the additional cost associated with the activation of bids outside of the price order under the special regulation, i.e. the price difference between bids which are denoted as special regulation and the current balancing power price in the operating hour. The market actors responsible for the imbalance in the system will be charged the remaining costs of the bid activation. (NVE, 2014)

The limitations in the transmission capacity between the price areas will be determined in the Elspot market, while internal bottlenecks in a price area will be handle under the special regulation. However, in some cases up- or down-regulation in a price area will be used as a special regulation:

- When power production in a price area exceeds the export's transmission capacity from this area, down-regulation in the area will be replaced with the special regulation.
- When power consumption in a price area exceeds the import's transmission capacity to this area, up-regulation in the area will be replaced with the special regulation.

In both cases the bottlenecks between the price areas will cause different balancing power prices in these areas.

The equal spot prices in each different price area indicate the availability of the transmission capacity between them. So that if there is any need for a regulation in one of the areas, Statnett can activate the balancing power bids in the price order and the regulation can take place in the price area with imbalances, as well as in any other price area. (NVE, 2104)

2.7 Influence factors on the balancing power volumes

One of the fundamentals of the power system is that the power production has to be equal to the power consumption at any time. Both the power production and the power consumption will be determined one day before the actual delivery. The actual situation in the operating hour may differ from what was planed a day earlier, due to different reasons, so that the TSO will need to activate its balancing power reserves in order to restore the balance in the power system. Some suggestions for reasons or factors that can influence the power system balance are listed below:

Power consumption

Various power consumers have different load patterns that vary within a year, during a week and over the course of a day or night. In recent years the power consumption has been more variable and more difficult to predict. There are numerous reasons why.

In countries where electricity primarily is used for space heating, power consumption strongly depends on the outside temperature. Climate change has led to milder winters in the Nordic region and thus lower power consumption at wintertime compared to earlier years.

Power consumption in buildings and households has also been influenced in recent years, by implementation of energy efficiency measures. This has resulted that less electricity is required for space heating in the winter season and more electricity is required for space cooling in the summer season compared to how it was before.
Another difficulty with power consumption is that there is a delayed response from the consumption side on the temperature change. Which means that if the temperature increases or decreases by a few degrees, it will take some hours before consumers will adjust their load.

Overall adaptations of the new technologies in everyday life e.g. different household appliances', computers, electrical cars, among others, increase electricity consumption in general, which makes the load increase and decrease impulsively.

All these aspects will influence the accuracy of the power consumption forecasts, and as a result, compliance between the predicted and the actual power consumption.

The deviation between the actual and the forecasted power consumption if the actual power consumption exceeds the predicted power consumption, will lead to an up-regulation in the power system, and vice versa.

Temperature

Temperature does not directly affect balance in the power system, but indirectly through having influence on the power consumption. The power consumption forecasts are conducted based on the temperature forecasts. So any deviations between the actual temperature and the temperature forecast will cause the deviation between the actual and the forecasted power consumption. Which means that if the temperature is milder than the forecasted temperature, the power consumption will be lower than expected, and therefore a down-regulation in the power system will be necessary. If the temperature is lower than the forecasted temperature, the power consumption will be higher than forecasted and consequently it will be a need for an up-regulation in the power system.

Wind power production

Expansion of the wind power production makes it challenging for the TSO to keep balance in the power system, due to the fact that the wind is a variable and an uncontrollable source of energy. Wind power producers estimate how much power they can produce the next day, based on wind forecasts that entail some uncertainty. This is why the actual wind power production will often deviate from the estimates in the operating hour and thereby will cause an imbalance in the power system. If the actual wind power production is higher than expected, down-regulation will take place in the power system. If the actual wind power production is lower than expected, the power system will be up-regulated.

Another disadvantage with the wind power production, considering balance in the power system, is the dependency of it on extreme weather conditions. So if the wind speed exceeds 25 m/s, the windmills will stop completely and a need for an up-regulation of the production will occur.

Power plant outage

Unexpected power plant outages can occur just as likely before the actual operating hour as well as in the operating hour. If an outage occurs before the Elspot market closes, the market actors will be able to take that into consideration when handling their bids for the next day. So the outage will influence balance in the power system only during the current day. However, if an outage occurs after the Elspot market has closed, it will influence the system balance both during the current and the next day. Whether or not the power plant outage will lead to an up-regulation of the production or a down-regulation of the consumption in the system, it will depend on the strategy that the power producer will use in order to solve the problem. They can choose between the handling of the outage by themselves, by replacing the failure unit with another one or, if possible, by increasing production of the units which are already in operation. Another option will be leaving the responsibility for dealing with the imbalance, due to the outage, to the TSO that will result in an up-regulation of the production or a down-regulation of the consumption.

Nuclear power plant outage
On the basis of information posted on ENTSO-E Transparency Platform, the installed capacity of nuclear power constitutes approximately 23% (8890 MW) in Sweden, and 18% (2782 MW) in Finland of the total installed capacity for different production types. (ENTSOE-E, 2016) The total installed capacity of nuclear power in Sweden is allocated to 8 reactors and in Finland it is allocated to 2 reactors.

So an outage of one of the nuclear reactors will cause a big impact on the power system balance. The outages of the nuclear power plant will also lead to an up-regulation of the production or a down-regulation of consumption, just as for outages of any power plant

Power line outages between price areas
Outages on power lines will limit the transmission capacity between price areas. So power line outages where power export was planed, will lead to a down-regulation in the export area. Power line outages where power import was planed, will lead to an up-regulation in the import area.

Sun radiation
As with the temperature, the sun radiation will influence balance in the power system by affecting the power consumption. It is not always easy to estimate how the intensity of the sun radiation will affect the power consumption. In hours where the sun radiation is high compared to the previous hours, it is difficult to fully estimate the affect of the heating of the sun on the consumption which would lead to a down-regulation of the power system.

The factors described above will affect the balance in the power system and lead to activation of the balancing power reserves. In addition to the balancing power market, it exists other ways/mechanisms that the market actors can use in order to deal with their imbalances.

Elbas market
Market actors have a possibility to trade their imbalances in the Elbas market during the day. Since trading in this market stops one hour before the actual delivery, it can be difficult for market actors to evaluate the actual situation in the operating hour. Market actors can trade too much power and cause imbalance with opposite sign in the operating hour or too little power so that part of the imbalance will occur in the operating hour. Therefore, trading in Elbas can either be a cause of an imbalance in the operating hour or it can contribute to removing the imbalance in the operating hour. In the Table 2.3 below the distribution of hours for each scenario is shown.

Number of hours / Direction of regulation		Year			Analysis period
Elbas market	**Balancing market**	**2013**	**2014**	**2015**	
Up	Up	1469	1878	1522	4869
Down	Down	2503	1796	1775	6074
Up	Down	2030	1740	1894	5664
Down	Up	1380	1513	1251	4144
Up	Zero	577	658	652	1887
Down	Zero	652	599	511	1762

Table 2.3. The Elbas market.

<u>*Special regulation*</u>
If an imbalance occurs, the TSO will normally activate bids from the balancing power list following the price order balancing the system. However, in the case of limited transfer capacity internally in a price area or when a failed situation occurs, the TSO will remain with a limited number of bids to restore the systems frequency. In this case, activating the balancing recourses the TSO will not take into consideration the prices of the bids and will activate the best suitable bid from the balancing power list in order to handle an imbalance/problem in the most possible efficient way.

Special regulation takes place in the operating time and the TSO activates bids from the balancing power list so that is why the special regulation will be considered as a part of the balancing power in this thesis. This will influence the value of the balancing power volumes by balancing some parts of the total imbalance in the system.

<u>*Deviation between the actual and the planed power flow between the price areas*</u>
If an imbalance occurs in one price area, it can be taken care of by activating balancing reserves either in the same area or in another price area. Since, as long as there is no need for a special regulation, the TSO will first make use of the least expensive reserves regardless on where these reserves are located. In this case the power flow between these price areas will be adjusted. This will cause a deviation between the planed and the actual power flow. That is why it is considered as an option to balance the system.

2.8 Influence factors on the regulating power prices

To be able to forecast the regulating power prices it is important to find the factors that the prices correlate with and that can be an indicator of which price to expect.

In principle the regulating price is volume dependent. Which means that the regulating price will increase with the increasing purchased volume. So one of the main factors affecting the regulating price is the activated regulating power volume. It should be noticed that in the case of an up-regulation, the regulating price will increase as more reserves will be activated, and in case of a down-regulation the regulating price will decrease, as more reserves will be activated due to reasons described in subsection 2.7.

However, by using only the relationship between the balancing prices and the size of the activated balancing power volume it is not possible to explain all of the prices that have or/and will occur in the Balancing Power market. So it has to be some other factors that influence the balancing power price. Some factors/ events that are supposed to explain some of the balancing power prices are described below.

Spot price

The balancing power price is limited by the spot price. The balancing power price for a down-regulation is always lower than the spot price, and the balancing power price for an up-regulation is always higher than the spot price. So, in hours with high prices in the Elspot market, the probability to get high regulating power prices in the case of an up-regulation increases, especially in combination with the high up-regulation volumes. In hours with low spot price the probability for getting low balancing power price for a down-regulation is high.

Spot bid curve

The power production bids that have not been accepted in the Elspot market can be offered to the Balancing power market. The less unaccepted bids that are left in the bid curve over the given price, the bigger the chance to getting higher regulating power prices under the up-regulation. The slope of the bid curve that will lay to the right from the system price can affect the balancing power price if the down-regulation will occur in the system. The less steeper the bid curve is, the bigger chance of getting the balancing power price closest to the system price is.

The regulating power prices can also be influence by the share of production units with a high degree of regulation that does not run. It is a high probability that the regulating power price will be high if most of the power plants with a high degree of regulation are running, and in the case of an up-regulation a power plant with a low degree of regulation has to be started up.

Inflow

The inflow is usually the highest in the spring when the snow melts; it is also high during the summer and autumn due to heavy rainfall. In these periods the power consumption is low and the power producers will store the biggest part of the inflow. In periods with a high inflow, especially in combination with a high reservoir level, the risk for the water loss or/and flood is high. Due to the fact that the spot price will be low, the balancing power price in the case of a down-regulation will also be low. In periods with a low inflow, especially in combination with a low reservoir level, the spot price will be high, which will lead to the high balancing power prices for the up-regulation.

It is also suspected that the regulating power prices can depend on the day of the week and the hour of day. For example, an up-regulation can be more expensive in the weekends when thermal power plants are less prepared for running.

Chapter 3

Time series forecasting

The process of developing a forecasting model for a variable of interest by using given data can be divided into number of steps shown in the figure 3.1.

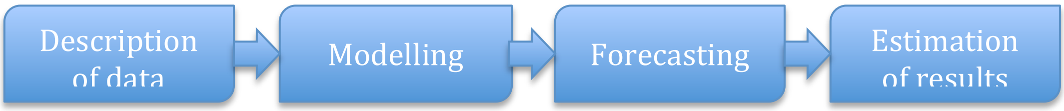


Figure 3.1. Developing a forecasting model.

First the data that will be used in the analysis need to be described by using summary statistics and/or graphical methods. Secondly, it is necessary to find a suitable statistical model to describe the data generating process. Then by using the statistical model, the future values of the variable of interest will be predicted. In order to evaluate the goodness and suitability of the data to the given dataset, the forecasting results need to be estimated. To fulfil each of these steps requires knowledge of the theoretical concepts.

In this chapter will be given a simple introduction to the some basic concepts of data description (subsection 3.1). Subsection 3.2 gives some overview over existing types of forecasting methods and subsection 3.3 describes the methods of estimating the forecasting results.

In the subsection 3.4, the concept of Machine learning as a forecasting tool will be also introduced. The parts of this subsection namely 3.4.2, 3.4.3 and 3.4.4 are devoted to an introduction to Microsoft Azure Machine Learning Studio and Microsoft Azure Machine Learning Studio regression models that has been used to develop a forecasting model for the balancing power volume and price.

3.1 Some basics of time series analysis

Chatfield (2001) defines a time series as “a set of observations measured sequentially through time” (Chatfield 2001, p.1). Depending on either the measurements of time series values have been made continuously or discrete set of time points, it can be distinguished between continuous and discrete time series respectively.
Before choosing a forecasting model and make forecasts for future values of a time series, it could be worthwhile to examine the series for historical patterns that can be

used in forecasting. While exploring historical patterns it can be convenient to distinguish between seasonal variation, trend, other cycle variation and irregular fluctuations of the series values.

Seasonal variation of a variable can be defined as similar patterns of behaviour of the variable observed at a certain time period of a year. An example is the power consumption pattern during the year that is always higher at wintertime.
Trend is an upward or a downward movement in time series values that can be identified over a long time periods. An example of trend is decreasing of power consumption over the last years due to implementation of energy efficiency measures in buildings.

Other cycle variations are upward or downward movements in time series values that can have duration from at least two years. Changing power demand of industry due to changes in world's economical situation can be an example on cycle variations. (Bowerman & O'Connell, 1993)

Irregular fluctuations are variations in the time series that do not follow any recognizable or regular patterns. It are movements in the data that can not be explained either by seasonal variation, trend or other cycle variation.
Time series that display seasonal variation (3.1(b)), trend (3.1(a)) and other cycle variation (3.1(c)) are shown in figure 3.1.

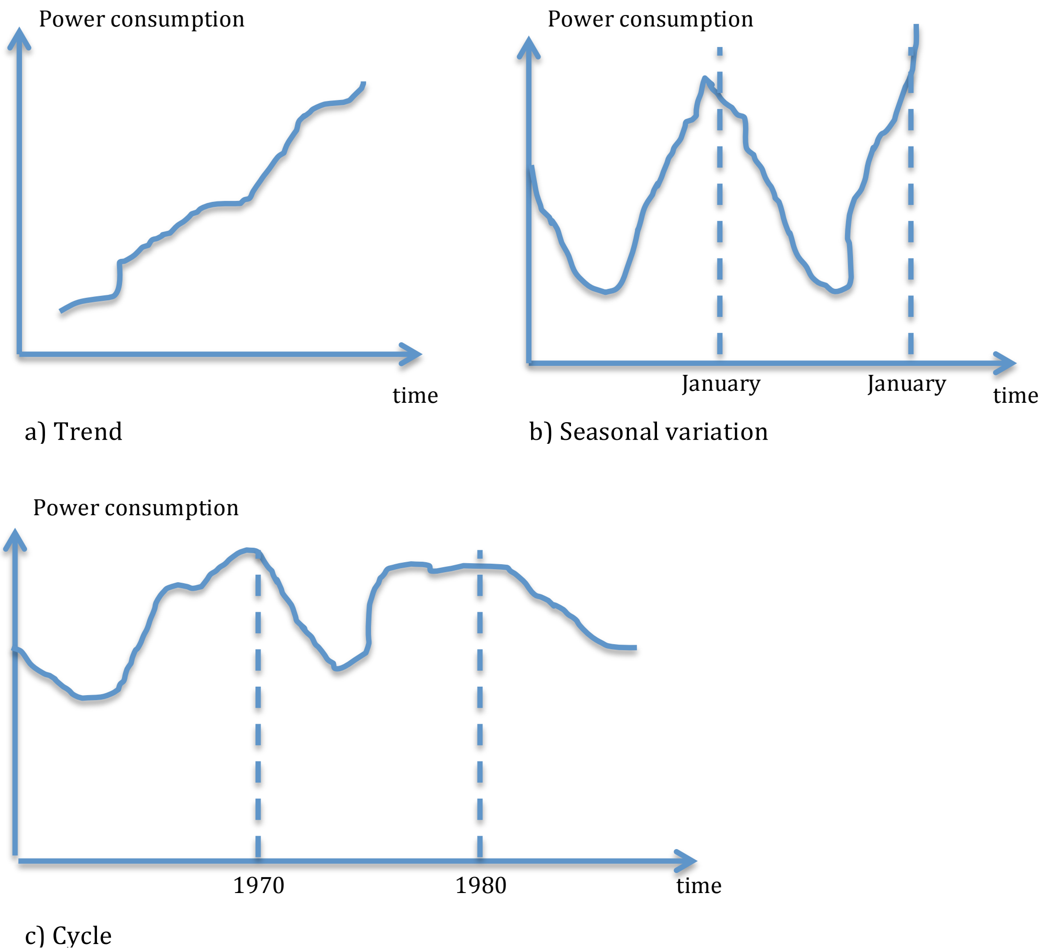


Figure 3.2. Patterns of the time series.

3.2 Type of forecasting method

Forecasting methods can be defined as approaches that can be used in order to make predictions of variable of interest from historical data. There are many different forecasting methods for forecasting of future events. In this chapter the brief overview over existing forecasting methods will be given.

Forecasting methods can be classified in different ways. Bowerman and O'Connell (1993) classify forecasting method into qualitative methods that are based on subjective expert opinions and quantitative methods in that statistical technics for analysis of historical data are used in order to forecast future values of a variable. And the quantitative methods are divided on causal and univariate methods. Qualitative technics includes such as subjective curve fitting, Delphi Method, technological comparisons, cross-impact methods etc.

Quantitative methods can be divided into univariate and causal methods. Univariate methods make forecasts for future values of a time series using the present or past values of this time series. When using an univariate method for time series prediction, the time series will be analysed for having any trends/patterns. These methods based on an assumption about that the existing trend would continue in the future.

When causal methods are used, the variables, which variable to be predicted could depend on, will be identified. These relationships will be used in order to predict future values of variable of interest.

Chatfield (2001) classifies forecasting methods as judgemental forecast, univariate methods and multivariate methods. This classification is very similar to Bowerman and O'Connell (1993) classification where qualitative methods and causal methods correspond to respectively to judgemental forecasts and multivariate methods in Chatfield (2001) classification.

Chatfield (2001) also distinguishes between non-automatic methods, which requires human intervention and automatic methods, which do not; and between simple and complicated methods. Univariate methods are defined as simple methods, and multivariate methods as complicated methods.

3.2.1 A brief introduction to univariate forecasting methods

In this section a short overview over some univariate methods will be given. Generally, all univariate methods can be divided into model-based forecasting methods and ad hoc forecasting methods.

When model-based forecasting methods is used, a particular univariate model for a particular time series will be build, then the parameters of the model will be estimate and the time series forecasts from the fitted model will be made. Univariate models can be divided into ARIMA models, state space models, growth curve models and non-linear models. The detail description of these types of models can be found at Chatfield (2001).

Ad hoc methods are forecasting techniques that are not based on explicitly on the probability models. These methods are based on assumption that analysed data has a compositional structure and involves breaking up and composing time series into supposed patterns. Ad hoc methods include simple exponential smoothing, Holt's linear trend method, the Holt-Winters forecasting procedure etc. more about these methods can be found at Chatfield (2001) and Talluri and Van Ryzin (2005).

It should be noticed that there are several univariate time series methods in addition to those that have been described above.

3.2.3 Multivariate forecasting methods

Multivariate forecasting methods utilize the relationships between a variable to be predicted, which is called dependent variable or response variable, and values of additional time series time series, which are called predictors. The response variable can depend totally or partly on the predictors. There are a various number of multivariate models that multivariate forecasting methods based on. In general it can be distinguished between single-equation model, vector AR and ARMA models etc. The single-equation models include all types of regressions and transfer function models. (Chatfield, 2001)

3.3 Estimation of forecast error

All forecasts contain some uncertainty that will lead to the forecast errors. In order to evaluate the forecast errors several types of error estimators can be used.

The forecast error e_i for a particular forecast i can be found by subtracting the forecasted value $\hat{y}_i$ of a variable of interest from the its actual value y_i and can be defined as

$$e_i = y_i - \hat{y}_i \quad \text{(Eq. 3.1)}$$

To measure the magnitude of the forecast errors over a time period the forecast errors e_i for a particular forecast i over the time period could be summed. However, the forecast errors e_i can take as positive as negative values, and sum of the errors can be near zero. So that wrong conclusions can be made about the forecast accuracy. In order to avoid it the absolute deviation of the forecasting error, which is defined in equation 3.2, can be considered.

$$Relative\ Absolute\ Error = |e_i| = |y_i - \hat{y}_i| \quad \text{(Eq.3.2)}$$

Known the absolute deviation, Mean Absolute Error (MAE) that measures the average of the absolute deviations $|e_i|$ for all forecasts n can be defined (se Eq.3.3).

$$MAE = \frac{\sum_{i=1}^{n}|e_i|}{n} = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{n} \quad \text{(Eq.3.3)}$$

Another option is to square the forecast error e_i as it is shown in equation 3.4.

$$Relative\ Squared\ Error = e_i^2 = (y_i - \hat{y}_i)^2 \quad \text{(Eq.3.4)}$$

Using the squared error the Mean Squared Error (MSE) can be define:

$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$ (Eq. 3.5)

So the Mean Squared Error is the average of the squared errors for all forecasts. (Bowerman & O'Connell, 1993)

Coefficient of Determination or R^2 can also be used in order to evaluate performance of a forecasting model. The Coefficient of Determination can be expressed by:

$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}$

(Eq. 3.6)
(Walpole, 2012)

3.4 Use of Machine Learning for time series forecasting

The Microsoft Azure Machine Learning Studio is used in this master thesis in order to develop a forecasting model for the balancing power volume and price. This subsection has been written in order to give an introduction to the machine learning in general (subsection 3.4.1) and to the Microsoft Azure Machine Learning Studio. Some common problems that can be solved by using the machine learning technics and the algorithms that are available with Microsoft Azure Machine Learning Studio are briefly described in the subsection 3.4.2. In section 3.4.3 the regression algorithms will be described in more details

3.4.1 Machine learning

Machine learning can be defined as " a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty" (P.Murphy, 2012, p.1). Machine learning based technology has been broadly used for different type of tasks: blant annet learning of a digital camera to detect faces, learning of anti-spam software to filter text massages, applications in smart-phones learn to recognize voice commands.

The role of tum for choosing the machine learning algorithms prior direct programming of a computer to perform a task is a problem with high complicity and a need for adaptivity. An example of problems that are to complex to program is task performed by humans/animals such as speech recognition and image understanding, or analysis that involve examination of large datasets. Compared to the programmed tools the machine learning has ability to adapt to the new input. All programmed tools have one limitation: once they are written down and installed, they will remain unchanged. Such tools are not able to adapt to the changes in the tasks over the time or from one user to another, i.e. they are static. The machine learning tools are able to adapt their behaviour to the new data input and to suggest a solution for such issues. The machine learning tools are, by nature, dynamic and are adaptive to the environment that they interact with. (Shalev-Shwartz & Ben-David, 2014)

There are many different types of machine learning. In general they can be classified as:

- Supervised versus Unsupervised learning. In supervised learning the data will be distinguished into the training data and test data. The training data is labeled with a correct answer. The labels provide extra information to the learner, which he will use under training in order to "gain expertise by using experience". By using the acquired expertise the learner can then predict that missing information for the test data. Classification and regression are the most common types of this type of learning.
  In unsupervised learning the data will not be characterized into the training data and the test data. The learner will process the data in order to summarize it or compress it on one or another way. Clustering and dimension reduction are typical examples of unsupervised learning.
- Active versus passive learners. The machine learning algorithms can be classified by the learner's role. An active learner interacts with the environment under the training time, while a passive learner only observes the information provided by the environment without influencing it.
- Helpfulness of the teacher. This criteria divides the machine learning algorithms into ones with an active teacher and ones with a passive teacher.
- Online versus Batch Learning Protocol. In case of Online Learning Protocol the learner has to respond throughout the learning process. In case of Batch Learning Protocol the learner has to engage the acquired expertise after processing large amount of data. (Shalev-Shwartz & Ben-David, 2014)

3.4.2 Types of machine learning problems

All algorithms provided by the Microsoft Azure Machine Learning Studio are represented in the Microsoft Azure Machine Learning Algorithm Cheat Sheet that is shown in the figure 3.3 and can by classified by the type of the problem to be addressed and divided into four categories:

- Regression algorithms are used in order to predict a numerical or continuous value of the variable of interest for the new data given in the dataset that contains one or more features or dependent variables.
- Classification algorithms address the problems of assigning the new unknown inputs to one or more labels or classes. If data needs to be assigned into two categories two-class classification will be used. The multi-class classification will be used for assigning data into tree or more categories.

- Clustering algorithms are used to discover a structure of the data by grouping similar objects together and separate dissimilar objects into the different groups.
- Anomaly detection algorithms are used in order to detect unusual data points in the dataset. (Barga, Fontama & Tok, 2014)

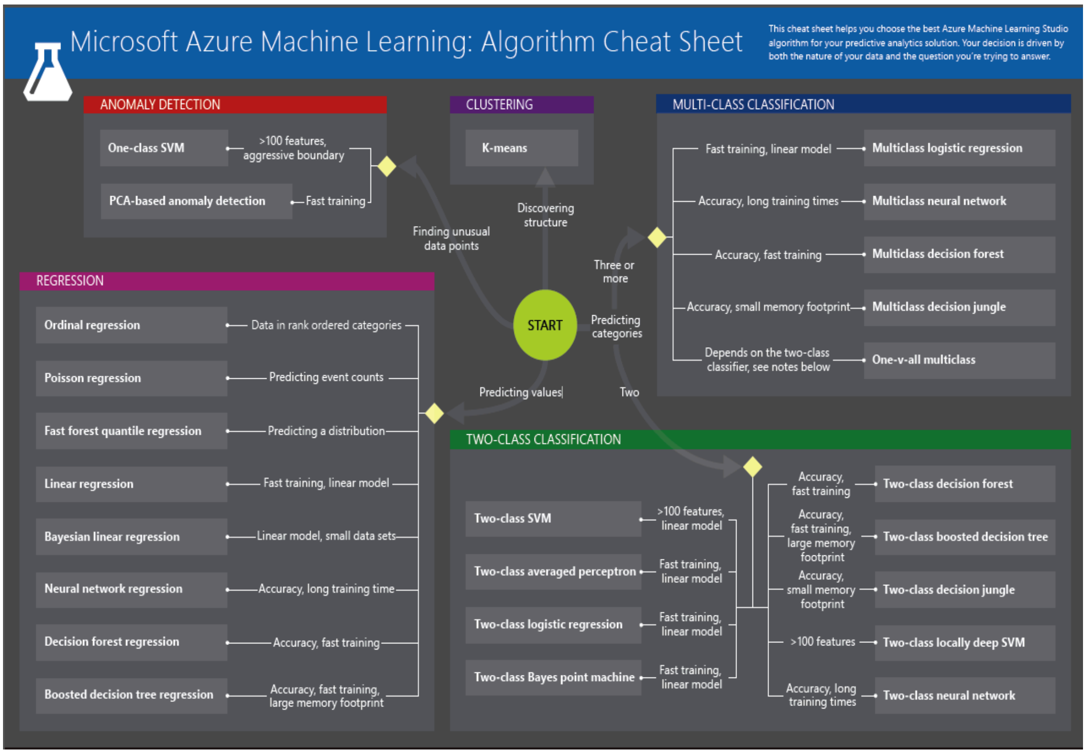


Figure 3.3. Algorithm cheat sheet. (Microsoft Azure A, 2016)

As it can be seen from the figure 3.3 the Microsoft Azure Machine Learning Studio supports various regression, classification, clustering and anomaly detection algorithms. However, the focus in this thesis will not be on all of these algorithms, but a two of them that will be used in order to develop a forecasting model for the balancing power volume and price. The description of the algorithms can be found on the Microsoft Azure Machine Learning Studio webpage (Microsoft Azure, 2015) and a rigorous introduction to the concepts underlying the machine learning algorithms are made by Shalev-Shwartz & Ben-David (2014).

The Boosted Decision Tree regression and the Decision Forest regression will be used to develop a forecasting model for the balancing power volume and price and therefore need to be presented in order to understand the concept that underlying these two regressions.

3.4.3 The Boosted Decision Tree regression

The Boosted Decision Tree regression is a supervised learning method that predicts the target variable by creating an ensemble of decision tree by using boosting. Each new tree created depends on the prior tree and learns by fitting the residuals of the tree that comes before it. (Microsoft Azure B, 2016)

The Decision Tree algorithm is a hierarchical technique that processes the data by splitting the dataset iteratively based on a certain criteria and thereby creates a decision tree. The goal of decision tree is to maximize the variance across the nodes in the tree and to minimize the variance within each node. This algorithm provides a tree-based approximation of a regression function for a given dataset.
(Barga, Fontama & Tok, 2014)

Each decision tree consists of a rote node, internal nodes and leafs or terminals.
The rote node has no incoming branches and has a zero or more outgoing branches. The node contains all the data in the dataset. Each of internal nodes has one incoming branch and two or more outgoing branches. The internal nodes and the root node contain test conditions that are used to separate the records in the dataset with different characteristics. Each of the leafs has one incoming branch and has no outgoing branches. Leafs contain scored labels of the target variable. Barga, Fontama & Tok, 2014)

The structure of a decision tree is shown in the figure 3.3 by a simple example of how decision tree algorithm can be used in order to predict the balancing power premium in an area. The decision tree starts from checking either the spot price is higher than 5 EUR. If the spot price is higher than 5 EUR, the tree turns to examine the hour of the day, and after determining the hour of the day it assigns a scored label to the leafs, i.e. predicts a possible balancing power premium. If the price is lower than 5 EUR, the tree start study the balancing power volume and then draws conclusions about the balancing power premium by assigning scored labels to each of the leafs.

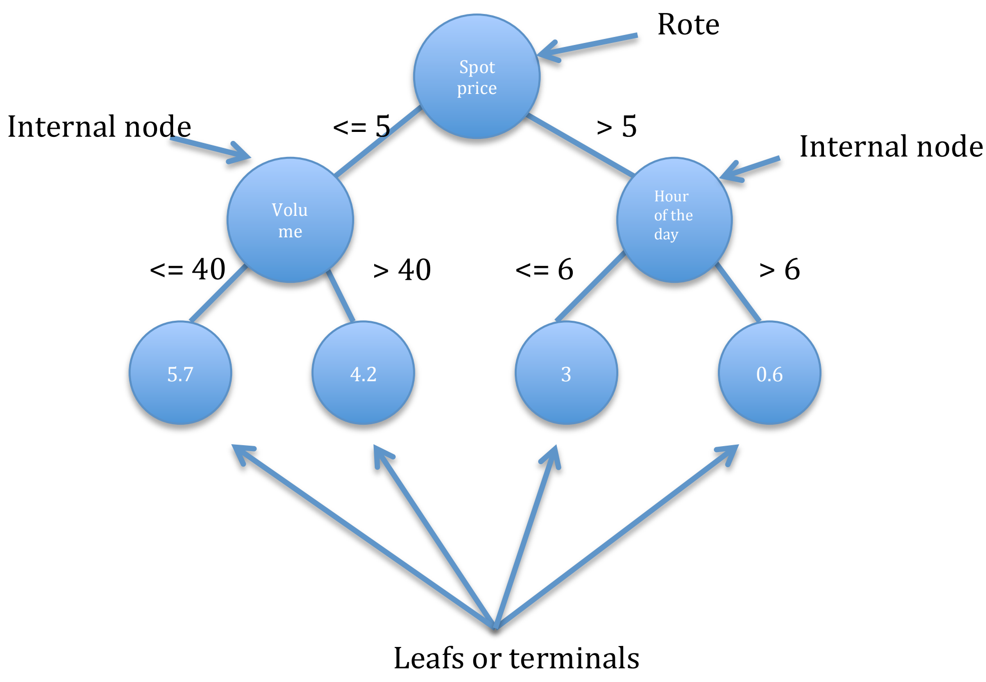


Figure 3.4. Decision tree structure.

In principle, there are exponentially many decision trees that can be constructed from a given dataset. These trees will have different accuracy and it is computationally infeasible to search for the optimum among them. Therefore, a number of efficient algorithms have been develop that have an aim to create a reasonably accurate, albeit suboptimal decision tree in a reasonable amount of time. There are different algorithms that are used in order to construct regression decision trees. Some of the commonly used algorithms are Iterative Dichotomizer 3 (ID3), C4.5 and C5.0 (successors of ID3), Automatic Interaction Detection (AID), and Classification and Regression Tree (CART). Barga, Fontama & Tok, 2014) A seminal work in the area of developing algorithms for decision tree constructing is done by Breiman et al (1984), Sonquist & Morgan(1964) and others. The differences of the algorithms will not be discussed in this thesis.

Decision tree based models have number advantages: they are easy to interpret, they are not effected by non-linear relationship between variables, they provide automatic parameter selection and computational efficiency and are able to handle unknown attributes and categorical data. However, the models have some disadvantages as well including the overfitting and instability due to at a small variation in data can lead to that different trees will be generated.
In order to address such an issue as overfitting boosting techniques can be used.

Boosting is one of the methods that are used in machine learning in order to create ensemble models. There is various numbers of boosting algorithms that can be used in machine learning. In the Microsoft Azure Machine Learning Studio the implementation of Multiple Additive Regression Trees (MART) gradient boosting algorithm is used. The algorithm constructs each regression tree step by step. Using the predefined loss function it measures the error in each step and correct it in the next one. The prediction model is an ensemble of the weaker prediction models. In the regression problem, boosting creates a series of trees in a step-wise fashion, and then using an arbitrary differentiable loss function, selects the optimal tree. (Microsoft Azure B, 2016)

3.4.4 The Decision Forest regression

The Decision Forest regression is a supervised learning method that creates a regression model consisting of an ensemble of randomly trained decision tree. It creates a number of decision trees that have the structure described in the subsection 3.4.3. Decision trees are non-parametric models that perform a series of simple tests for each instance, traversing a binary tree data structure until a leaf node is reached. The outputs of each tree in the decision forest is a Gaussian distribution by way of prediction. The algorithm then performs an aggregation over the ensemble of trees in order to find a Gaussian distribution closest to the combined distribution for all trees in the model. (Microsoft Azure C, 2016)
The output of the model in general can be represented as:

$$p(y|v) = \frac{1}{T}\sum_t^T p_t(y|v)$$
(Eq. 3.7)

where
$p_t(y|v)$ is posterior distribution obtained by the t th tree.
T is number of trees in the forest. (Criminisi & Shotton, 2013)

The decision tree in the Decision Forest regression have the same advantages and disadvantages as regular decision tree described in the subsection 3.4.3.

Chapter 4

Statistical analysis of data

In this thesis it will be attempted to predict the balancing power volume and the balancing power prices by the use of a regression model. As known, a regression model predicts the variable of interest by utilizing the relationships between it, and a set of predictors, i.e. the variables the variable of interest is depending on. In order to get a good prediction of the balancing power volume and the balancing power prices, by using a regression model, it is important to identify the variables that can be used as predictors. To do this the correlation analysis for the balancing power volume and price, and the factors that are suspected for having an influence on them, have been carried out and are described in this chapter.

Subsection 4.1 describes the correlation analysis for the balancing power volumes and the factors that influence the power system balance. This subsection is a continuation of the work that have been conducted in advance of this master thesis, so the new experiments that have been carried out and the important results from the previous work for this thesis will be represented in subsection 4.1.

In subsection 4.2 the statistical analysis of the balancing power prices is performed. This subsection is divided into different parts – the analysis of the balancing power prices for having daily, weekly or annual variation (subsection 4.2.3) and the correlation analysis for the balancing power prices and the factors that can influence it (subsection 4.2.4 – 4.2.8).

4.1 Statistical analysis of the balancing power volume

This subsection is based on the preliminary work that has been made in advanced of writing this master thesis. The aim was to conduct a statistical analysis for data of the balancing power volume and to test the relationship between the balancing power volume and the power system disturbance factors.

The results from this work that are important for this thesis and the extension of the analysis that have been made in order to conduct more accurate results is written in subsection 4.1.1. The description of the model that will be used in the analysis and the analysis objectives are described in subsection 4.1.2 and subsection 4.1.3 respectively. Subsection 4.1.4 and subsection 4.1.5 describes the work that have been done in order to carry out the correlation analysis between the balancing power volume and the set of the power system disturbance factors. In subsection 4.1.6, the dependence of the balancing power volume in an hour on the past values of the balancing power volume from the previous hours has been examined. The results of this analysis are summarized in subsection 4.1.7.

4.1.1 Improvement and expansion of the correlation analysis

The purpose of the preliminary work, which this section is based on, was to study the balancing power volume data for having daily, weekly or annual variation and to determine the possible relationship between the balancing power volume and the disturbance factors that influences the balance in the power system.

The results of the analysis were that the balancing power volume has some daily, weekly or annual variation that is shown in the figure 4.1.

The correlation between the balancing power volume and some balance disturbance factors, such as deviation between the forecasted and the actual temperature, power consumption and the wind power production on the weekdays is as follows:

- Correlation between the balancing power volume and the deviation between the actual and the forecasted temperature is -0.1
- Correlation between the balancing power volume and the deviation between the actual and the forecasted power consumption is 0.54
- Correlation between the balancing power volume and the deviation between the actual and the forecasted wind power production is 0.09.

From the results it can be concluded that the deviation between the actual and the forecasted wind power production and the deviation between the actual and the forecasted temperature have little influence on the balancing power volumes. One sees also that only the deviation between the actual and the forecasted power consumption has some significant impact on the balancing power volumes.

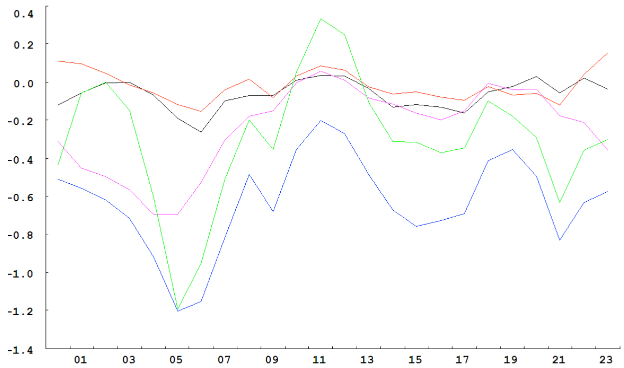
However, the balance in the power system can be influenced by one disturbance factor or more disturbance factors at the same time. These factors can contribute to the same regulation state (two or more factors indicate an up- or down-regulation) or can counteract each other (one or more factors indicate the opposite regulation state). Therefore, when considering just one disturbance factor at the time, one risks to underestimate or overestimate the net imbalance volume in the system. It is also important to remember the other options for balancing the power system such as power flow adjustments between price areas, trading in the Elbas and special regulation, which can be used in order to remove the imbalances in the system, partly or whole. The impacts of those mentioned above were not taken into consideration under the previous correlation analysis.

When studying the time series for the balancing power volume one noticed that the regulation state often tends to last longer than one hour. When analysing the duration of the regulation state in the Nordic region it was found that:

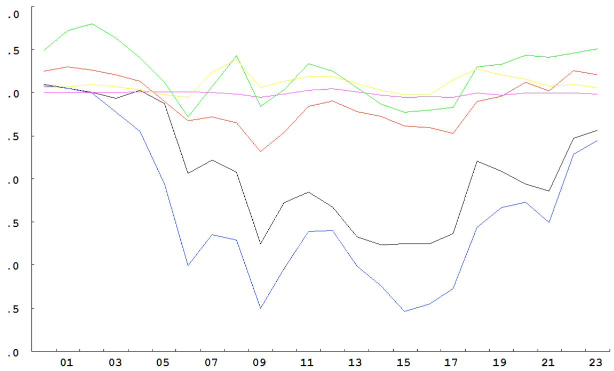
- If no balancing reserves have been activated in under 1 hour, the state is most likely to change
- If the up-regulation has lasted no longer than 3 hours, the regulation state will most likely continue in the next hour
- If the down-regulation has lasted no longer than 4 hours, the regulation state will most likely continue in the next hour

More detailed results can be found in Appendix A. The impact of the past values of the balancing power volume on the future ones has not been taken into consideration under the previous correlation analysis.

Due to these arguments it was decided to conduct a new correlation analysis for the balancing power volumes and the factors that influence the balance in the power system while taking into consideration the reflections above.

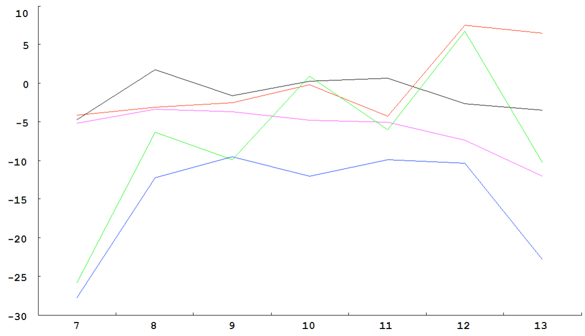


(a) Daily variation of the balancing power volume in the Norwegian price areas.

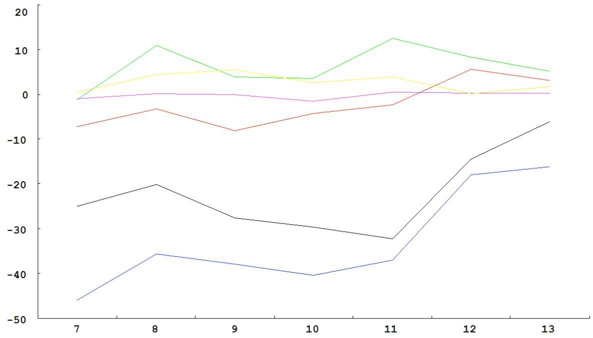


(b) Daily variation of the balancing power volume in the Swedish and Danish price areas.

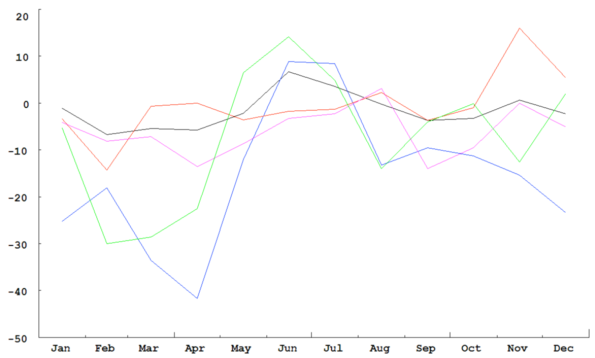
Figure 4.1. Daily,



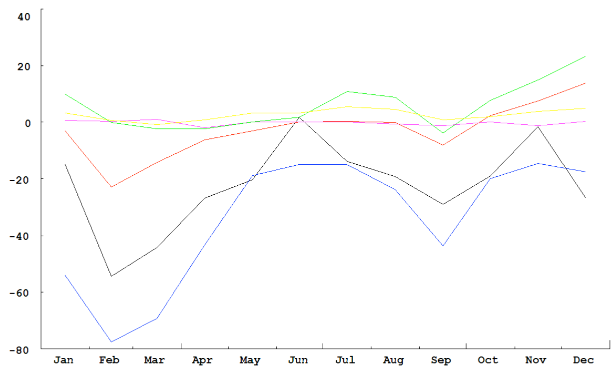
(c) Weekly variation of the balancing power volume in the Norwegian price areas.



(d) Weekly variation of the balancing power volume in the Swedish and Danish price areas.



(e) Annual variation of the balancing power volume in the Norwegian price areas.



(f) Annual variation of the balancing power volume in the Swedish and Danish price areas.

weekly and annual variation of the balancing power volume.

4.1.2 Modell description

For the analysis, the same model that was used in the preliminary work will be used.

The whole Nordic region, except Jylland, is one joint synchronised area. This means that the power system in all the Nordic countries has the same frequency, and therefore an imbalance in one price area will affect the balance of the whole system. (Statnett, 2014) The Nordic TSOs also exchange their balancing power resources with each other through the common balancing power list, where the only one criterion for activating a bid is its price (do not include situations when special regulation takes place). If an imbalance occurs in one price area, activating the balancing reserves in any other price area and adjusting the power flow between the areas can balance it.

That is why, when carrying correlation analysis between the balancing power volumes and the factors described above, it will be natural to look at the balancing power volumes of the whole Nordic region and not only on the balancing power volumes in each price area separately.

When conducting the correlation analysis it was chosen to consider Norway, Sweden and Denmark together as one common area with aggregated balancing power volumes, consumption, production etc. It was decided to keep Finland outside of the analysis for a number of reasons:

- Spot price in Finland is usually higher compared to the spot price in the other Nordic price areas. That is why the power often flows towards Finland.
- Statkraft has a low activity in Finland and that is why they are not particularly interested in the balancing power forecasts there.
- Hydropower is not a dominating type of production in Finland, so the Finish power system is not particularly flexible.
- There is one big actor in the Finish power market that has the major share of both the power consumption and the power production so the imbalances caused by this actor often are regulated internally.

Possible exchange of the balancing power resources between Finland and each of the three countries, which are Norway, Sweden and Denmark, will be taking into account, by considering the deviation between the actual and the spot power flow between the price areas NO4-FI, SE1-FI and SE3-FI.

The deviation between the actual and the spot power flow between price areas in Norway, Sweden and Denmark and countries outside the Nordic region, whom they have connections with, will be considered as the balancing power reserves exchange, in order to balance the power system in the Nordic region or outside of it.

Since Norway, Sweden and Denmark will be treated as one price area, for each variable, which will be considered in the analysis, aggregated time series will be used. The detailed overview over the time series used in the analysis is provided in Appendix B.

4.1.3 Analysis objectives

The main purpose of this analysis is to investigate possible dependences between the balancing power volume and the balance disturbance factors, such as deviation of the actual consumption, wind power production or/and temperature from the forecasted values. In the correlation analysis other options that can be used to balance the power system, such as adjustment of the power flow between the price areas, special regulation and trading in the Elbas market, will be taking into account. So that, instead of looking at the correlation between the balancing power volumes and each of the balance disturbance factors individually, in this correlation analysis it will be looked for a correlation between a set of disturbance factors and a set of options that market actors can choose between in order to balance the power system. The correlation analysis will be carried out for a time period from 2013 to 2015. The choice of the length of the analysis period was based on how long back in time the records goes in the time series, and also on that the power market is dynamic and that it continually undergoes changes in form of changing the trading rules, the price areas boarders and construction of new power lines etc. So the data that lays too long back in time will not be that relevant in order to estimate the present situation in the power system. All experiments under the correlation analysis will be carried out with the help of Statkraft's database called Fame.

Under the analysis following will be tested out:

- Correlation between a set of balancing options and a set of balance disturbance factors
- Dependence of the balancing power volume on the temperature fluctuation
- Correlation between the balancing power volume in an hour and the balancing power volume in the previous hours

It should be noticed that some of the influencing factors described in chapter 2 will not be taken into account when conducting the analysis due to data unavailability.

In this analysis the effect of the power plants outages on the balancing power volumes would not be taken into account due to a lack of data of the power plant outages. However, it should be noticed that the overview over the power plant outages could be conducted from the Urgent Market Massages (UMM), but it has not been done due to two reasons. Firstly, conducting information from the UMMs for several years is time consuming. Secondly, power producer reports about their power plants outages independent on if the imbalances, due to these outages, were handled by the power producers themselves by activating other production units or if they have decided to handle the responsibility for balancing the system over to the TSO. Which means power plant outages will not always lead to the activation of the balancing reserves.

The influence of the sun radiation will not be taken into account as well. The reason for that is a lack of data for the whole Nordic region, so the result conducted under the experiment using only the sun radiation data for Norway are not necessarily significant, and it can only be elaborated if these results can give any additional information about the balancing power volume in the whole Nordic region.

4.1.4 Correlation between the set of balancing options and the set of disturbance factors

In order to analyse the relationship between a set of balancing options and a set of possible disturbance factors, it was decided to take the starting point in one of the fundamentals of the power system. To maintain the power system frequency continually at 50 Hz, the power production in the system must always be equal to the consumption. The balance between the power production and the consumption in general can be expressed as:

$$\textit{Power production} = \textit{Power consumption} + \textit{Export} - \textit{Import} \qquad \text{(Eq. 4.1)}$$

By any disturbances in the system, the balancing reserves will be activated in order to restore the system balance. The equation (1) can be rewritten as:

$$\begin{aligned} &\textit{Power production} + \textit{Regulating power} = \\ &= \textit{Power consumption} + \textit{Export} - \textit{Import} + \textit{Imbalance} \qquad \text{(Eq. 4.2)} \end{aligned}$$

The equation (2) can be simplified and rewritten as:

$$\textit{Regulating power} = \textit{Imbalance} \qquad \text{(Eq. 4.3)}$$

By placing the deviation between the actual and the forecasted power consumption, deviation between the actual and the forecasted wind power production on the right hand side (further RHS) of equation 3 and the balancing power volumes, special regulation and the deviation in the power flow on the left hand side of equation 3 (further LHS), the correlation between a set of balancing options and a set of possible disturbance factors can be tested by testing the correlation between the RHS and the LHS of equation 4.

$$\begin{aligned} &\textit{Balancing power volume} + \Delta \textit{Power flow} + \textit{Special regulation} = \\ &= \Delta \textit{Wind power production} + \Delta \textit{Consumption} \qquad \text{(Eq. 4.4)} \end{aligned}$$

where

ΔPower flow = Actual power flow – Spot power flow

ΔWind power production = Forecasted wind power production – Actual wind power production

ΔConsumption = Actual consumption – Forecasted consumption

As one can see, volumes traded in the Elbas have not been assigned to either the left hand side or the right hand side of the equation. As it has been mentioned before in chapter 2.1.1, trading in the Elbas market can contribute to the power system balance as well as being a reason for the imbalance in the power system. To decide on which side of the balance equation (4) to place the Elbas volumes, the following hypothesis will be tested:

H_0: Trading in the Elbas is a balance disturbance factor in the power system

H_1: Trading in the Elbas contributes to balance in the power system.

To determine which of these two hypotheses one should reject, the correlation between the RHS and the LHS of the following equations will be tested:

$$\begin{aligned}&\textit{Balancing power volume} + \Delta \textit{Power flow} + \textit{Special regulation} + \textit{Elbas volume} = \\ &= \Delta \textit{Wind power production} + \Delta \textit{Consumption} \qquad \text{(Eq. 4.5)}\end{aligned}$$

$$\begin{aligned}&\textit{Balancing power volume} + \Delta \textit{Power flow} + \textit{Special regulation} = \\ &= \Delta \textit{Wind power production} + \Delta \textit{Consumption} + \textit{Elbas volume} \qquad \text{(Eq. 4.6)}\end{aligned}$$

If the correlation between the RHS and the LHS is stronger when it comes to equation 5, the hypothesis H_1 will be rejected. If not, the hypothesis H_0 will be rejected.

The correlation between the RHS and the LHS of equation 5 is 0.50. The correlation between the RHS and the LHS of equation 6 is 0.59. Since the correlation between the RHS and the LHS is higher when it comes to equation 6, the hypothesis H_1 has been rejected, and trading in the Elbas will from now on be considered as a balance disturbance factor of the power system.

Activation of volumes in Balancing power market, special regulation and the deviation between the actual and the spot power flow are considered as a set of balancing options. The deviation between the actual and the forecasted power consumption, the deviation between the actual and the forecasted wind power production and the Elbas volumes are considered as a set of balance disturbance factors. The correlation between a set of balancing options and a set of the balance disturbance factors is 0.59. Figure 4.2 shows a scatter plot of a set of balancing options and a set of the balance disturbance factors.

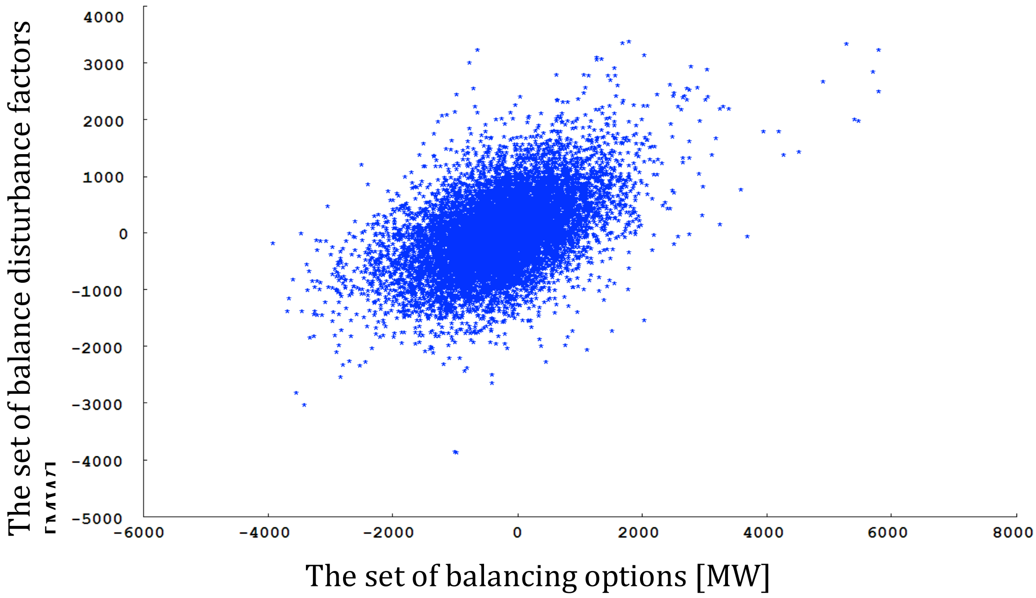


Figure 4.2. Scatter plot of a set of balancing options versus a set of balance disturbance factors.

By removing a variable one by one on both sides of equation 6 and by testing the correlation between the RHS and LHS of each new variation for equation 6, the importance of taking into consideration all the factors and aspects which the balancing power volume can depend on, has been shown. The results for all of the possible variable combinations on the RHS and LHS of equation 6 are shown in Appendix C. The results of the correlation between the LHS and the reduced RHS of equation 6 are shown in table 4.1.

LHS of the equation 6	RHS of the equation 6	Correlation coefficient
Balancing power + ΔPower flow + Special regulation	*ΔWind power production + Elbas volume*	0.31
Balancing power + ΔPower flow + Special regulation	*ΔConsumption + Elbas volume*	0.58
Balancing power + ΔPower flow + Special regulation	*ΔWind power production + ΔConsumption*	0.54
Balancing power + ΔPower flow + Special regulation	*ΔConsumption*	0.52
Balancing power + ΔPower flow + Special regulation	*ΔWind power production*	0.20
Balancing power + ΔPower flow + Special regulation	*Elbas volume*	0.35

Table 4.1. Results of the correlation test

Correlation between the set of the balancing options and the set of the disturbance factors of 0.59, applies for all of the weekdays and also the weekends and the holidays in the analysis period. However, it is believed that the correlation between them in the weekdays will be stronger than in the weekends. This assumption is based on, that in the weekends and the holidays the market actors have a much less overview over the situation in the power system due to less people at work and more events takes place on the consumer side that the consumption forecasts not always are able to capture. This will lead to more disturbances in the data, which reduces the correlation.

To test this hypothesis the correlation between the RHS and the LHS of equation 6 was tested separately for the weekdays, and the weekends and the holidays. The correlation between the set of the balancing options and the set of disturbance factors on the weekdays, and the weekends and the holidays are respectively 0.62 and 0.51 that confirms the assumption made earlier.

From figure 4.2, it can be seen that as bigger the imbalance in the power system is, the more balancing reserves will be activated in order to balance the system. This yields for both an up- and a down-regulation. The values of the balancing power volume depends both on the total imbalance in the power system and the other options that can be used in order to deal with the imbalance, i.e. activation of the balancing reserves in an other price area and special regulation. The importance of taken into consideration all these factors when carrying out the correlation analysis for the balancing power volume, and the power system disturbance factors can be seen in the table 4.1 and Appendix C. That is where the sensitivity of the correlation between the set of the balancing options and the set of the disturbance factors to the substitution of the different factors is shown. The correlation is highest when all the factors are taken into consideration, and will decrease when the number of factors is reduced. This correlation will also vary for the weekdays and the weekends/holidays, and it will be highest during the weekdays.

4.1.5 Dependence of the balancing power volumes on the temperature

Testing the correlation between the sum of the balancing power volumes, special regulation volumes, the deviation between the actual and the spot power flow and the sum of the deviation between the actual and the forecasted power consumption, wind power production and the Elbas volumes, the influence of the deviation between the actual and the forecasted temperature has not been taken into account due to two reasons. The first reason is low correlation between the balancing power volumes and the deviation between the actual and the forecasted temperature (subsection 4.1.1). The second reason is that the error in the temperature forecast most likely affects the balancing power volumes indirectly by affecting the power consumption, and by this, causing the deviation between the actual and the forecasted consumption. So the influence of the deviation between the actual and the forecasted temperature on the balancing power volumes will be taken into account via the deviation between the actual and the forecasted power consumption.

However, it was decided to investigate whether or not the correlation between the balancing power volumes and the disturbances factors will differ depending on the size of the deviation between the actual and the forecasted temperature.
The starting point was taken in equation 6 and the correlation between the RHS and the LHS of the equation was tested for different temperature intervals.
The deviation between the actual and the forecasted temperature, *ΔTemperature* is defined as:

$$\Delta Temperature = Forecasted\ temperature - Actual\ temperature \qquad \text{(Eq. 4.7)}$$

Zero was chosen as a starting point for the intervals and the length of each interval was sett to 1 degree. The maximum and minimum deviation between the actual and the forecasted temperature, as it is defined in equation 4.7, is around 2 and -3 degrees respectively. The results of the test were summarized in figure 4.3. From the figure, one sees that the correlation between the RHS and the LHS of the equation differs for different intervals of the deviation between the actual and the forecasted temperature. The correlation is quite constant when the actual temperature does not exceed the forecasted temperature by more than 2 degrees, and when the forecasted temperature does not exceed it by more than 1 degree. The more the actual temperature exceeds the forecasted temperature, the more the correlation increases. The correlation between the RHS and the LHS of equation 4.6 will be the lowest in the case when the forecasted temperature exceeds the actual temperature by 2 degrees.

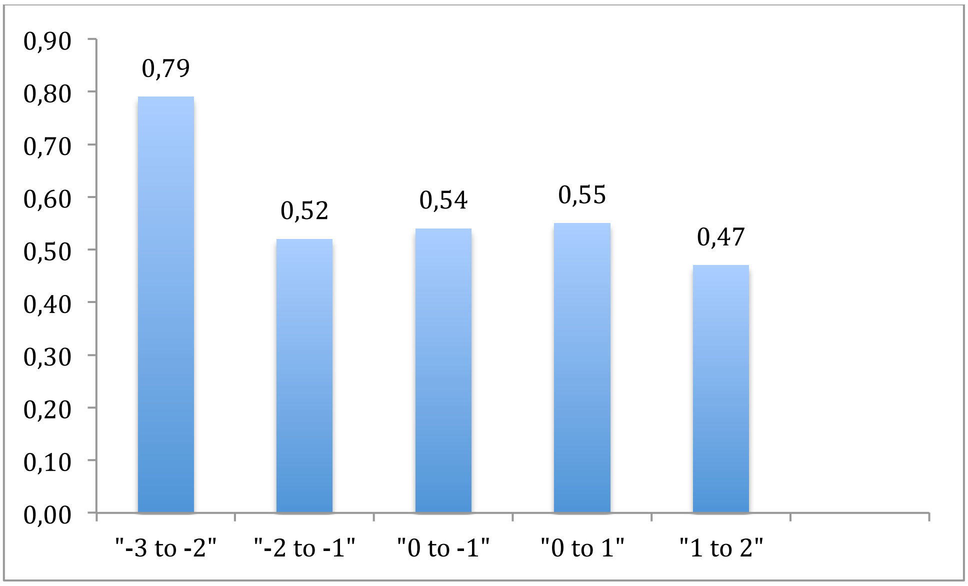


Figure 4.3. Sensitivity of the correlation between the LHS and the RHS of the equation 4.6 to the temperature deviation.

4.1.6 Correlation between the balancing power volume in an hour and the balancing power volume in the previous hours

The correlation between the balancing power volume in an hour and the past values of the balancing power volume has been examined. For the paste values the records of the balancing power volume that go 1, 2, 3, 4, 5, 6 and 7 hours back in time, have been taken. The correlation coefficients for each of the cases are represented in table 4.2, and the scatter plots for the values of the balancing power volume in an hour versus the past values of the balancing power volume are shown in figure 4.4.

Hours back in time	Correlation coefficient
1	0.88
2	0.72
3	0.62
4	0.53
5	0.47
6	0.42
7	0.38

Table 4.2. Results of the correlation test

From the scatter plots and the table one sees that the values of the balancing power volume in an hour have a linear dependence on the past values of the volume, and as bigger the value of the activated balancing power volume in the previous hours, the bigger the value of the volume in an hour will be. However, as longer back in time the past values of the balancing power volume goes, the lower the correlation coefficient will be. So the values of the balancing power volume in an hour will be mostly affected by the past values of the balancing power volume that does not go too long back in time.

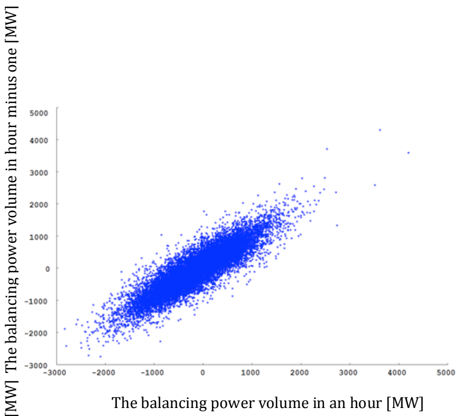
4.1.7 Summarizing the results

In this subsection the correlation analysis for the balancing power volume and the power system disturbance factors has been carried out. The results of the analysis have shown the importance of taken into consideration all the options that are available for balancing the power system and all balance disturbance factors in order to draw the right conclusions about the dependence of the balancing power volume on the factors influencing the system balance. It can be seen that when testing the dependence between the balancing power volume and each of the disturbance factors, a low or no correlation will be obtained between the balancing power volume and each of the factors. The correlation between the balancing power volume and the deviation between the actual and the forecasted consumption is 0.40, the deviation between the actual and the forecasted wind power production is 0.20 and the Elbas volumes is 0.01. However, by looking at the sum of all the imbalances in the system, the correlation of 0.49 between the balancing power volume and the net imbalance in the system can be obtained.

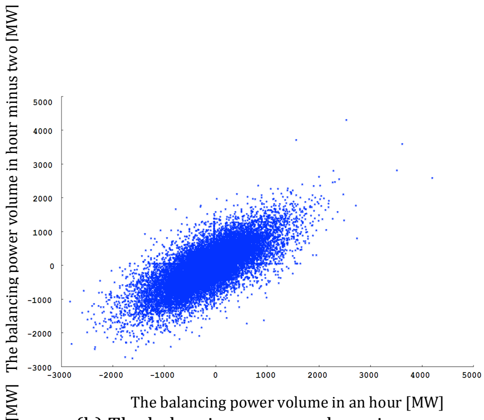
The correlation between the balancing power volume and the net imbalance in the system will be higher (0.59), when taking into consideration the other options that the market actors can make use of to balance the system such as special regulation and power flow adjustment between the price areas. However, the special regulation does not have a big impact when it comes to dealing with the total imbalance in the system, since it is used in order to deal with the local bottlenecks. So if it is not taken into consideration the correlation between the net balancing power and the net imbalance in the system will remain the same, i.e. 0.59.

The deviation between the actual and the forecasted temperature has no direct influence on the balancing power volume. However, the correlation between the net balancing power and the net imbalance in the system varies for the different values of the temperature forecast error. It will also vary depending on the day of the week, and is the highest, at 0.62, on the weekdays.

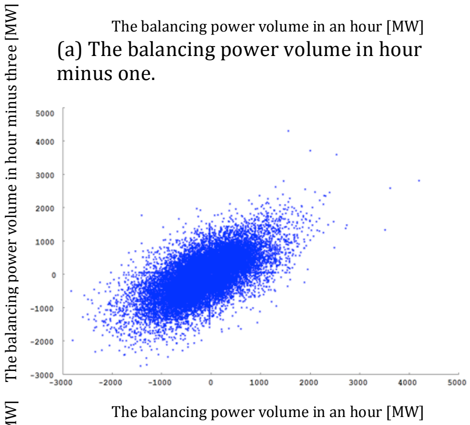
The balancing power volume in an hour is also influenced by the values of the balancing power volume in the previous hours. The highest correlation coefficient of 0.88 is obtained when testing the correlation between the balancing power volume in an hour and the balancing power volume from the previous hour. The correlation between the balancing power volume in an hour and the past values of the balancing power volume will be lower as the past values of the balancing power volume, which are used in the analysis, goes further back in time.



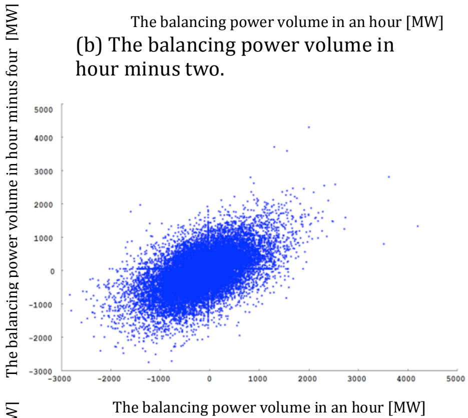
(a) The balancing power volume in hour minus one.



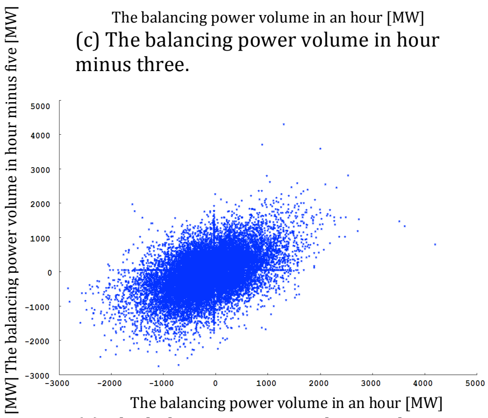
(b) The balancing power volume in hour minus two.



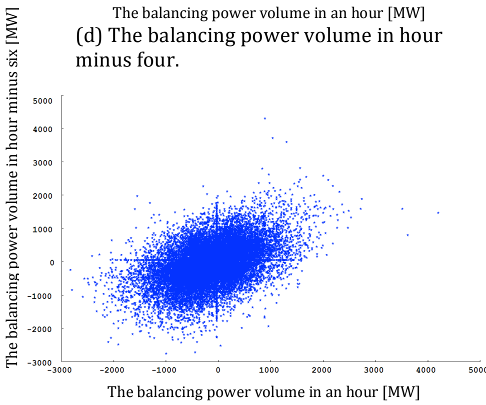
(c) The balancing power volume in hour minus three.



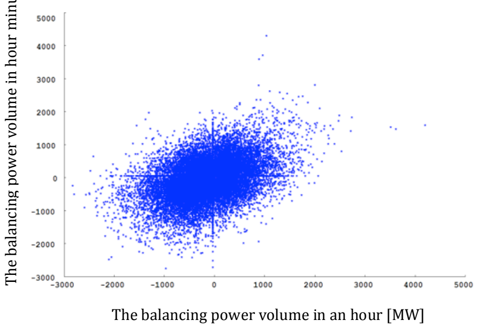
(d) The balancing power volume in hour minus four.



(e) The balancing power volume in hour minus five.



(f) The balancing power volume in hour minus six.



(g) The balancing power volume in hour minus seven.

Figure 4.4. Scatter plot of the balancing power volume in an hour versus the balancing power volume in the previous hours.

4.2 Statistical analysis of the balancing power price

In this subsection the statistical analysis of the balancing power price is performed. In sections 4.2.1, and 4.2.2 the model description and the objectives of the analysis are given. The analysis of the balancing power price having any daily, weekly or annul patterns is conducted in subsection 4.2.3. Various experiments that have been done within the correlation analysis for the balancing power premium and the different factors that are supposed to influence it, are presented in subsections 4.2.4-4.2.8.

4.2.1 Model description

When analysing the regulating power prices, it is natural to analyse the prices for each price area separately. It is also intuitive to consider the balancing power prices for up- and down-regulation separately.

When conducting the analysis and further developing of the forecasting model, the concept of the regulating premium, which is stated below in equation 4.8, will be utilized

$$\Delta p = \begin{cases} p_{up} - p_{spot} & if\ up - regulation \\ p_{down} - p_{spot} & if\ down - regulation \\ 0 & if\ no\ regulation \end{cases} \qquad \text{(Eq. 4.8)}$$

Equation 2.7 defines the premium for activating regulating reserves Δp as a difference between the regulating price p_{up}/p_{down} and the spot price p_{spot}.

When using such a concept, it is important to define what is a high premium and what is a low premium. When it comes to the premium for up-regulation, everything is straight forward, and as high the difference between the balancing power price for up-regulation and the spot price is, the higher the premium is.

With premium for down-regulation it is not that simple and defining the premium as low or high can easily be misunderstood. The balancing power price for a down-regulation is always lower than the spot price, so when subtracting the spot price from the balancing power premium, as it is shown in equation 4.8, the last one will be negative. Which means the balancing power premium for a down-regulation actually cannot be high. However, the balancing power premium for a down-regulation will be defined as high, when it is closest to zero and it will be defined as low, as it decreases.

4.2.2 Analysis objectives

The purpose of the analysis is to investigate the balancing power premium for having a daily, a weekly or an annual variation and to find out if the premium will be influenced by the factors described in the chapter 2 and by the past values of the balancing power premium from the previous hours.

Following experiments will be conducted:

- Analysis of the balancing power premium for having a daily, a weekly or an annual variation
- Test of the correlation between the balancing power premium and the balancing power volume
- Test of the correlation between the balancing power premium and the spot price
- Test of the correlation between the balancing power premium and the slope of the bid curve
- Test of the correlation between the balancing power premium and the inflow
- Test of the correlation between the balancing power premium in an hour and the past values of the balancing power premium from the previous hours

The analysis period for all experiments is from 2013 to 2015, and the overview over the time series used in the analysis can be found in Appendix D. All the experiments will also be conducted with the help of Fame.
In order to capture the possible dependence of the value of the balancing power premium on the day of the week and the hour of the day, when examining the correlation between the balancing power premium and the balancing power volume, the spot price or the slope of the bid curve, the correlation coefficient will be obtained for five different cases:

- For the whole analysis period
- For weekdays
- For weekends
- Day hours
- Night hours

When searching for the balancing power premium variation, the premium in the price areas in Norway, Sweden and Denmark will be taken into consideration in order to be able to analyse the patterns in all of the price areas and in order to find some common attributes.

For the correlation analysis, it was decided to use only the price area NO3, since the forecasting model will be developed based on the data from the NO3.

The NO3 is not isolated and it has interconnections with the neighbour areas. So, the balancing power premium in the NO3, can also be influenced by the other areas. However, it is impossible to know which area, at which situation, that can influence the premium in the NO3, so it would be very time consuming to carry out the correlation analysis taking into consideration all the possible combinations of how the different areas can influence the balancing power premium in the NO3. So when carrying out the correlation analysis, the influence on the balancing power premium in the NO3 by the other areas will not be considered.

4.2.3 Examining daily, weekly and annual variation of the balancing power price

When analysing the time series, it is important to examine whether or not the time series follows some seasonal or any other types of variation, since the knowledge about the time series variation can be helpful in order to predict the future values. In this chapter the daily, weekly and the annual variation of the balancing power premium will be examined. The variation of the balancing power premium in each price area in Norway, Sweden and Denmark will be examined for following any specific patterns. Then each pattern will be compared to each other to find some common attributes. The balancing power premiums for up- and down-regulation will be tested for having a variation separately.

In order to determine if the balancing power premium has any daily variation, the values in the time series have been organized in the matrix that has the hour of the day as column indexes and the days of the year as row indexes. The matrix is shown in figure 4.5 below. The prices in each column were added to each other, and their sum has been divided by the number of rows in the matrix. The obtained result represents an average value of the balancing power premium per hour of the day for the whole analysis period.

$$\xrightarrow{\text{Hour of the day}}$$

$$\begin{array}{ccc}
 & & \begin{matrix} 01 & 02 & 03 & 04 & 05 & 06 & \cdots & \cdots & 18 & 19 & 20 & 21 & 22 & 23 & 24 \end{matrix} \\
\text{Date}\Big\downarrow & \begin{matrix} 01.01.12 \\ 02.01.12 \\ \vdots \\ \vdots \\ 31.12.14 \end{matrix} & \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \cdots & \cdots & x_{18} & x_{19} & x_{20} & x_{21} & x_{22} & x_{23} & x_{24} \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & \cdots & \cdots & y_{18} & y_{19} & y_{20} & y_{21} & y_{22} & y_{23} & y_{24} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \cdots & \cdots & a_{18} & a_{19} & a_{20} & a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}
\end{array}$$

Figure 4.5. Matrix for testing the daily variation of the balancing power premium.

To examine the balancing power premium for having a weekly and an annual variation, the similar procedure to the one described above has been done. The difference is in indexing of columns and rows in the matrixes. When searching for a weekly variation the constructed matrix has the day of the week as column indexes and the week of the year as row indexes (see figure 4.6). The values in the matrix represent an average value of the balancing power premium per day in the analysis period.
When examining the balancing power premium for the annual variation the constructed matrix has the month of the year as column indexes and the year as row indexes (see figure 4.7). The values in the matrix represent an average value of the balancing power premium per month in the analysis period.

The Fame code that has been used to conduct results can be found in the Appendix E.

Day of the week

Week

$$\begin{array}{cc} & \begin{array}{ccccccc} Sun & Mon & Tue & Wed & Thu & Fri & Sat \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ \vdots \\ 156 \end{array} & \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \end{bmatrix} \end{array}$$

Figure 4.6. Matrix for testing the weekly variation of the balancing power premium.

Month of the year

Year

$$\begin{array}{cc} & \begin{array}{cccccccccccc} Jan & Feb & Mar & Apr & May & Jun & Jul & Aug & Sep & Oct & Nov & Dec \end{array} \\ \begin{array}{c} 2012 \\ 2013 \\ 2014 \end{array} & \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \\ z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 & z_8 & z_9 & z_{10} & z_{11} & z_{12} \end{bmatrix} \end{array}$$

Figure 4.7. Matrix for testing the annual variation of the balancing power premium.

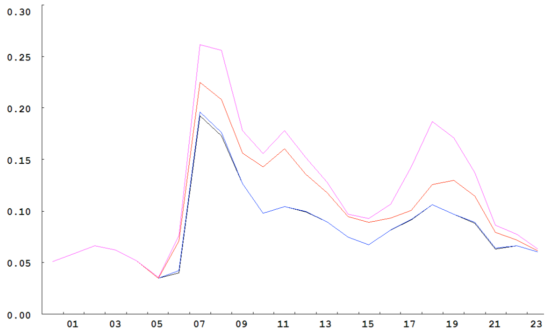
The daily variation of the balancing power premium is shown in figure 4.8, and the weekly variation and the annual variation are shown in figures 4.9 and 4.10, respectively.

The daily variation of the balancing power premium for an up-regulation for the different price areas and the daily variation of the balancing power premium for a down-regulation for the different prices is presented in figures 4.8(a), 4.8(b), 4.8(c) and figures 4.8(d), 4.8(e), 4.8(f) respectively. From the figures, it can be seen that the balancing power premium for the up-regulation has an evident pattern: it has a peak between 6 a.m. and 10 a.m. and between 4 p.m. and 8 p.m., which approximately corresponds to the spot price peaks. The balancing power premium for the down-regulation follows a certain daily pattern as well. It is highest between 10 a.m. and 4 p.m. and during the night hours. Comparing the premium for an up- and a down-regulation one sees that in hours when the premium for the up-regulation is high the premium for the down-regulation is low, and vice versa.

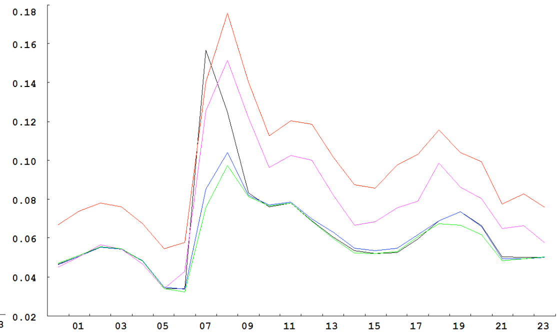
In figures 4.9(a), 4.9(b), 4.9(c) the weekly variation of the balancing power premium for the up-regulation is shown. The patterns for the different price areas differ, but it is possible to highlight some common attributes. It can be seen that the premium for the up-regulation is highest on Mondays. It decreases during the week, it is quite low on Fridays and Saturdays and it starts to increase again on Sundays. The weekly variation of the balancing power premium for the down-regulation, which is shown in the figures 4.9(d), 4.9(e), 4.9(f), is similar to the weekly regulation of the premium for the up-regulation, but has one difference- it has two peaks, one on Mondays and one on Fridays.

The annual variation of the balancing power premium for the up- and down-regulation is shown in the figures 4.10(a), 4.10(b), 4.10(c) and 4.10(d), 4.10(e), 4.10(f) respectively. In annual patterns of the premium for both the up- and down-regulation one sees some variations for the different price areas. However, the high premium for the up-regulation can be identified from April until July, and then it decreases before taken off again around October-November. When it comes to the annual pattern of the premium for the down-regulation, it can be seen that it also varies for the different price areas, but three peaks in the premium during the year in all areas can be determined. The balancing power premium for the down-regulation is highest between May and September; it is also high between November and April.

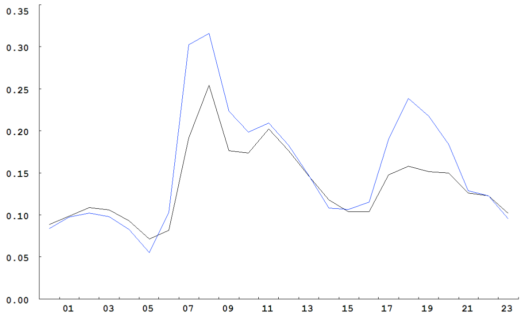
It should be noticed that these patterns does not provide any information about the value of the regulating power premium, but about how the balancing power premium changes during a day, a week or a year.



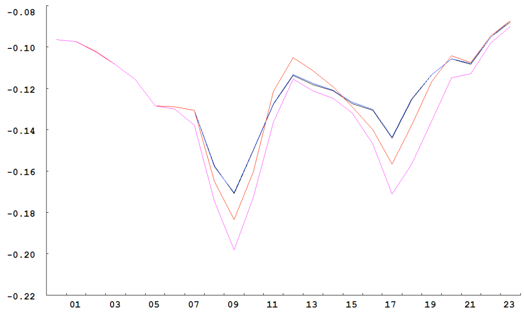
(a) Daily variation of the balancing power premium for up-regulation in the Swedish price areas



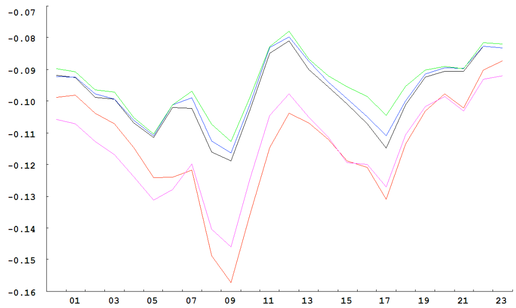
(b) Daily variation of the balancing power premium for up-regulation in the Norwegian price areas



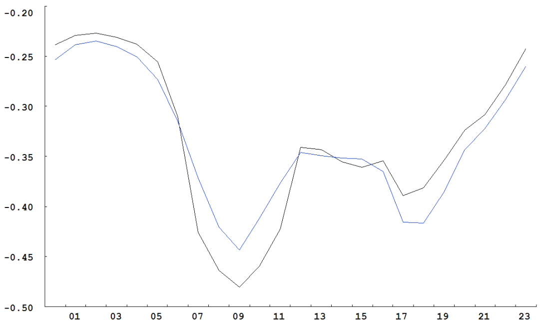
(c) Daily variation of the balancing power premium for up-regulation in the Danish price areas



(d) Daily variation of the balancing power premium for down-regulation in the Swedish price areas

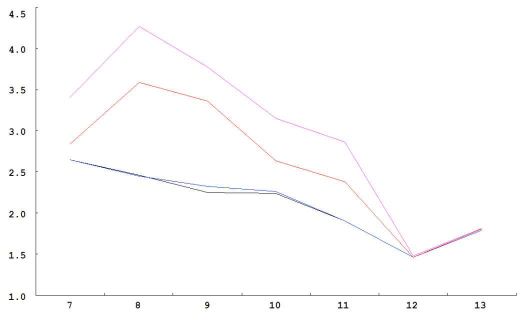


(e) Daily variation of the balancing power premium for down-regulation in the Norwegian

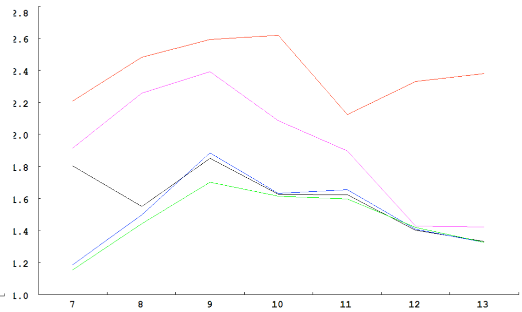


(f) Daily variation of the balancing power premium for down-regulation in the Danish price areas

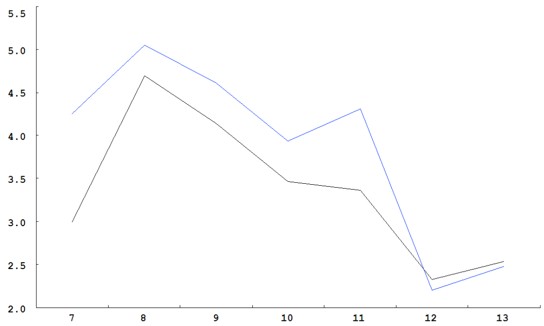
Figure 4.8. Daily variation of the balancing power premium.



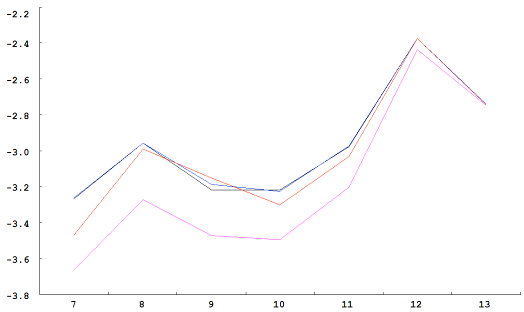
(a) Weekly variation of the balancing power premium for up-regulation in the Swedish price areas



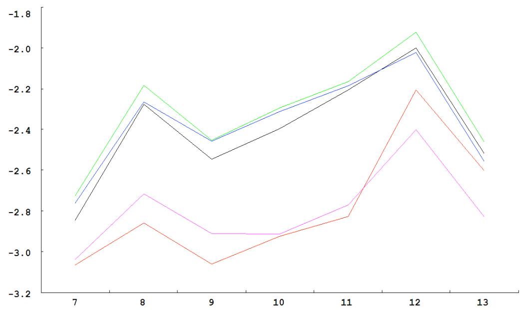
(b) Weekly variation of the balancing power premium for up-regulation in the Norwegian price areas



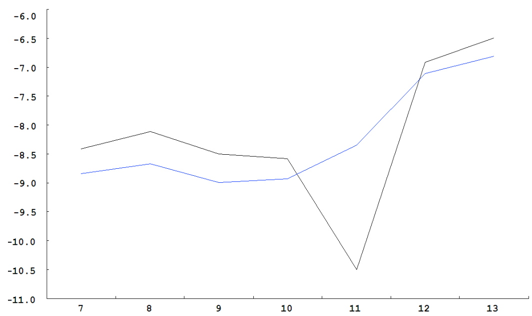
(c) Weekly variation of the balancing power premium for up-regulation in the Danish price areas



(d) Weekly variation of the balancing power premium for down-regulation in the Swedish price areas

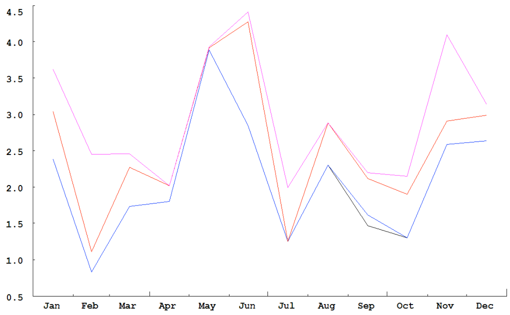


(e) Weekly variation of the balancing power premium for down-regulation in the Norwegian price areas

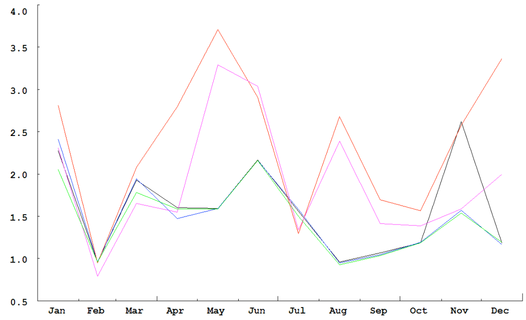


(f) Weekly variation of the balancing power premium for down-regulation in the Danish price areas

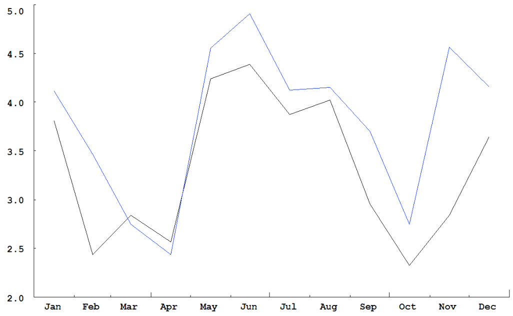
Figure 4.9. Weekly variation of the balancing power premium.



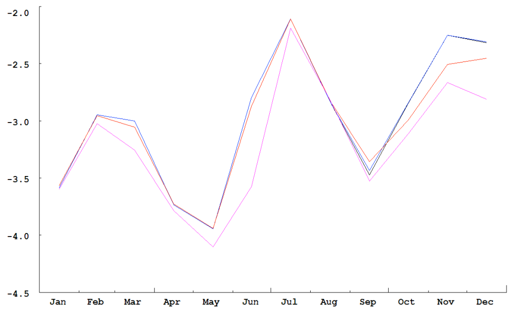
(a) Annual variation of the balancing power premium for up-regulation in the Swedish price areas



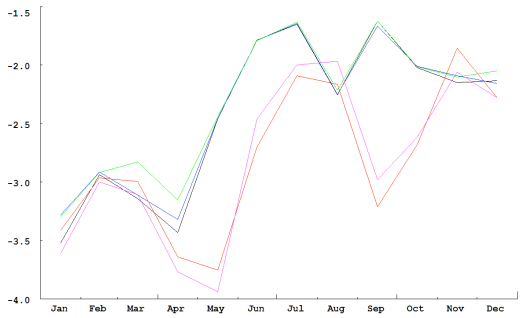
(b) Annual variation of the balancing power premium for up-regulation in the Norwegian price areas



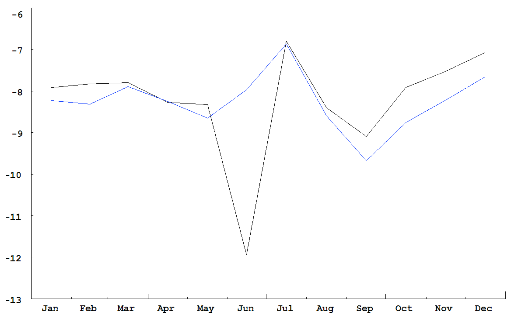
(c) Annual variation of the balancing power premium for up-regulation in the Danish price areas



(d) Annual variation of the balancing power premium for down-regulation in the Swedish price areas



(e) Annual variation of the balancing power premium for down-regulation in the Norwegian price areas



(f) Annual variation of the balancing power premium for down-regulation in the Danish price areas

Figure 4.10. Annual variation of the balancing power premium.

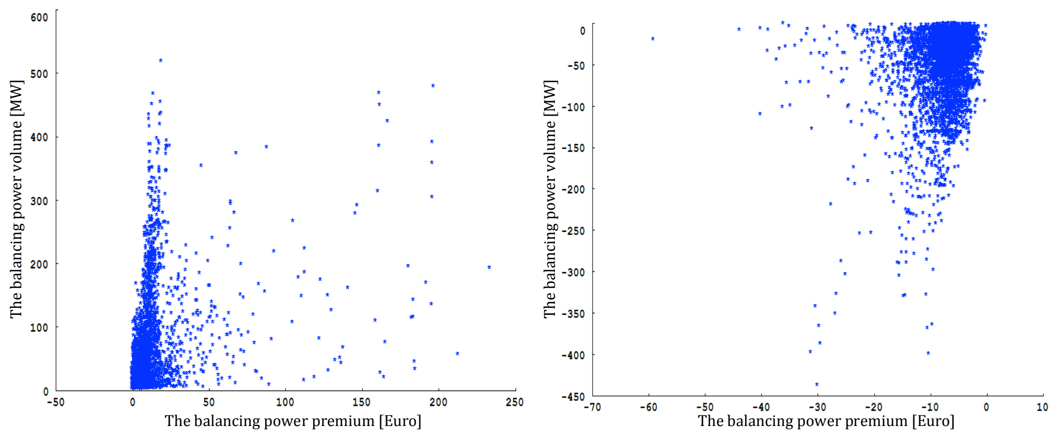
4.2.4 Correlation between the balancing power premium and the balancing power volume

If an imbalance occurs in the system the TSO starts to activate the balancing reserves. As more reserves will be activated, the higher the cost of the regulation will be. In order to test this dependence, the correlation between the balancing power premium and the balancing power volume has been tested. First, the correlation between the balancing power premium and the balancing power volume has been tested for the whole analysis period. Then the analysis period has been divided into working days/weekends and day/night hours, and the correlation test has been carried out one more time for each of the cases.

The results of the analysis and the scatter plots of the balancing power volume and the balancing power price under both the up- and down-regulation are shown in table 4.2 and in figure 4.11, respectively.

Filter	Correlation coefficient	
	Up-regulation	**Down-regulation**
All hours	0.28	0.23
Weekdays	0.24	0.20
Weekends	0.44	0.30
Day	0.33	0.19
Night	0.53	0.32

Table 4.3. Results of the correlation analysis.



(a) Up-regulation (b) Down-regulation

Figure 4.11. Scatter plot of the balancing power premium versus thebalancing power volume.

Figure 4.11(a) and figure 4.11(b) shows the scatter plot for the balancing power premium versus the balancing power volume in all hours of the analysis period for the up- and down-regulation receptively. One sees that even though the correlation between the balancing power premium and the balancing power volume for both the up- and down-regulation is quite low, (0.28 for up-regulation and 0.23 for down-regulation), the balancing power volume and premium have some linear dependency. As bigger the balancing power volume under the up-regulation is, the higher the premium, and as bigger the balancing power volume for the down-regulation is, the lower the premium. However, numerous outlying points that can be seen in the figure make the correlation between the balancing power volume and the balancing power premium lower. From table 4.3 it can also be seen that the correlation between the balancing power volume and the balancing power premium varies for the different hours of the day and for the different days of the week.

4.2.5 Correlation between the balancing power premium and the spot price

As it was mentioned in chapter 2, the balancing power price is limited by the spot price, and the balancing power price is higher and lower then the spot price for the up- and down-regulation respectively. However, it is not suitable to say that the balancing power premium is limited by the spot price and that it has the same dependence on it as the balancing power price. Since the balancing power premium is a remaining value when subtracting the spot price from the balancing power price, and it represents the extra cost of activating the balancing reserves. So, for example, even though the spot price is low, the balancing power premium for the up-regulation can be really high, if there is a need for a lot of regulation, and the bids that have been submitted to the Balancing power market have a high price. The balancing power premium for the down regulation can be low even though the spot price is high if it is a great need for a down-regulation in the system.

It is difficult to say precisely which dependence there is between the balancing power and the spot price. So the correlation between the balancing power premium and the spot price has been examined. The correlation between them in the whole analysis period and in weekdays/weekends and day/night hours is shown in Table 4.4. The scatter plot for the balancing power premium versus the spot price in the whole analysis period can be seen in figure 4.12

Filter	Correlation coefficient	
	Up-regulation	**Down-regulation**
All hours	0.34	-0.36
Weekdays	0.36	-0.49
Weekends	0.17	-0.09
Day	0.24	-0.08
Night	0.38	-0.45

Table 4.4. Results of the correlation analysis.

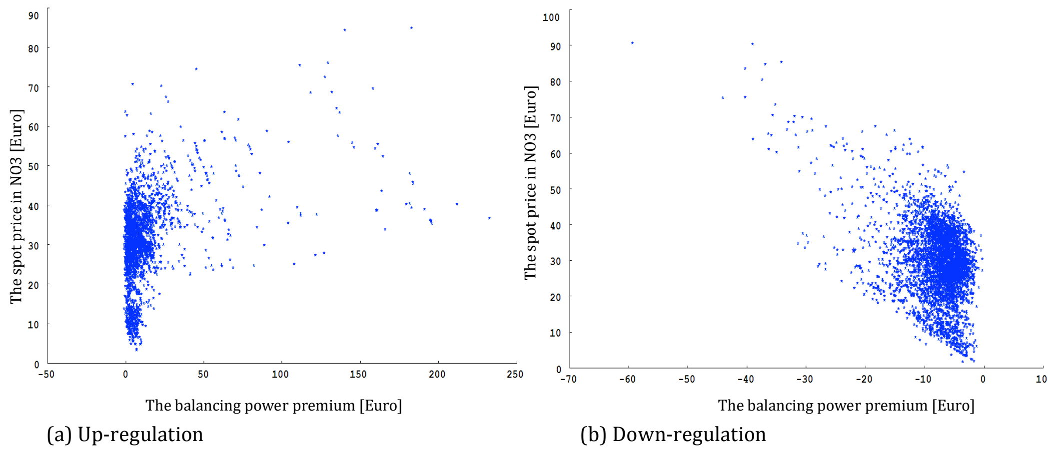


Figure 4.12. Scatter plot for the balancing power premium versus the spot price.

From the results in table 4.4 and figure 4.12(b), it can be seen that the balancing power premium for the down-regulation is negatively correlated with the spot price: as lower the spot price is, the higher the balancing power premium will be and as higher the spot price is, the lower the premium will be. The balancing power premium for the up-regulation is positively correlated with the spot price: the higher the spot price is, the higher the premium will be. However, from figure 4.12(a), it can be noticed that there are many outlying points, and in some hours the premium was really high even though the spot price is relatively low and vice versa.

The balancing power premium for both the up- and down-regulation is not strongly correlated with the spot price, and the correlation coefficient between the premium and the spot price is 0.34 in the case of an up-regulation, and -0.36 in the case of a down-regulation. The correlation coefficient also varies for the different days of the week and for the different hours of the day. The balancing power premium and the spot price are correlated the most at the night hours and the weekdays in case of both an up- regulation and a down-regulation.

4.2.6 Correlation between the balancing power premium and the slope of the bid curve

As it was mentioned before in chapter 2, the fewer bids that are left after the Elspot clearing, that have not been accepted, the bigger the chance is to get a high regulating price in case of an up-regulation, and it is a bigger chance to get the balancing power price that are high under a down-regulation, when many bids that has been accepted in the Elspot have the same or almost the same price. The balancing power premium is most likely to have the same dependence on the bids that have been accepted in the Elspot and their price.

In order to investigate the relationship between the balancing power premium and the bids that have been accepted in the Elspot in case of a down-regulation and that have not been accepted in the Elspot in case of an up-regulation, it was decided to look at the relationship between the premium and the slope of the Elspot bid curve.

The Elspot bid curve has been generated for each hour in the analysis period by using the algorithm developed by Statkraft. The curve has been aggregated from both the bids for demand and the bids for supply, due to that many market actors submit their bids for both supply and demand, and under the Elspot clearing these bids will be cleared as a net value.

All bids in the bid curve that will lie to the left of the system price are assumed to be submitted to the Balancing Power market for the same price as they were submitted in the Elspot. It should be noticed that in reality, bids that have not been accepted in the Elspot will be submitted to the Balancing Power market, since the market actors will evaluate the suitability of the remaining units for a quick up- or down-regulation and the profitability of the bidding in the market, and then decide whether or not to offer their reserves to the Balancing Power market. The price of the submitted bids will also be higher due to start-up cost, cost of running only in few hours and so on. So the bid curve for the Balancing Power market will have a steeper slope and will lie higher up along the price-axis than the Elspot curve.

When calculating the slope of the bid curve, the expression stated below is used:

$$Slope = \frac{P_2 - P_1}{V_2 - V_1} \qquad \text{(Eq. 4.9)}$$

where V_1 and V_2 are volumes that lie 1000MW below and over in relation to the last accepted bid in the Elspot V, and P_1 and P_2 are prices that corresponds to them as shown in figure 4.13.

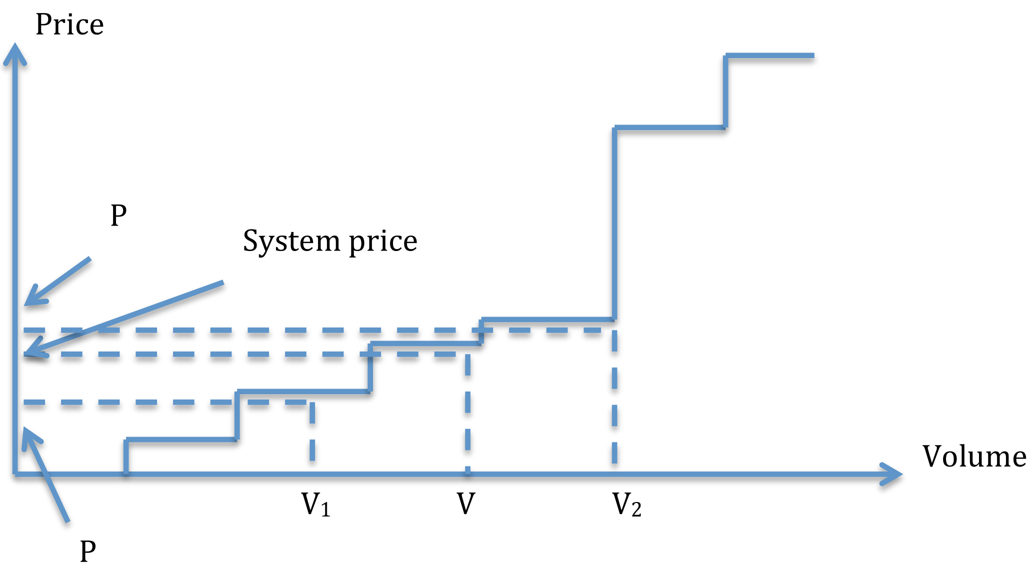


Figure 4.13. The Elspot bid curve.

The Fame code that has been used to obtain the slope of the Elspot bid curve for each hour of the analysis period can be found in Appendix F.

After finding the slope, the correlation between the slope of the bid curve and the balancing power premium has been determined. The results are shown in table 4.5.

Filter	Correlation coefficient	
	Up-regulation	**Down-regulation**
All hours	0.35	-0.44
Working days	0.35	-0.51
Weekends	0.16	-0.32
Day	0.19	-0.37
Night	0.39	-0.48

Table 4.5. Results of the correlation analysis.

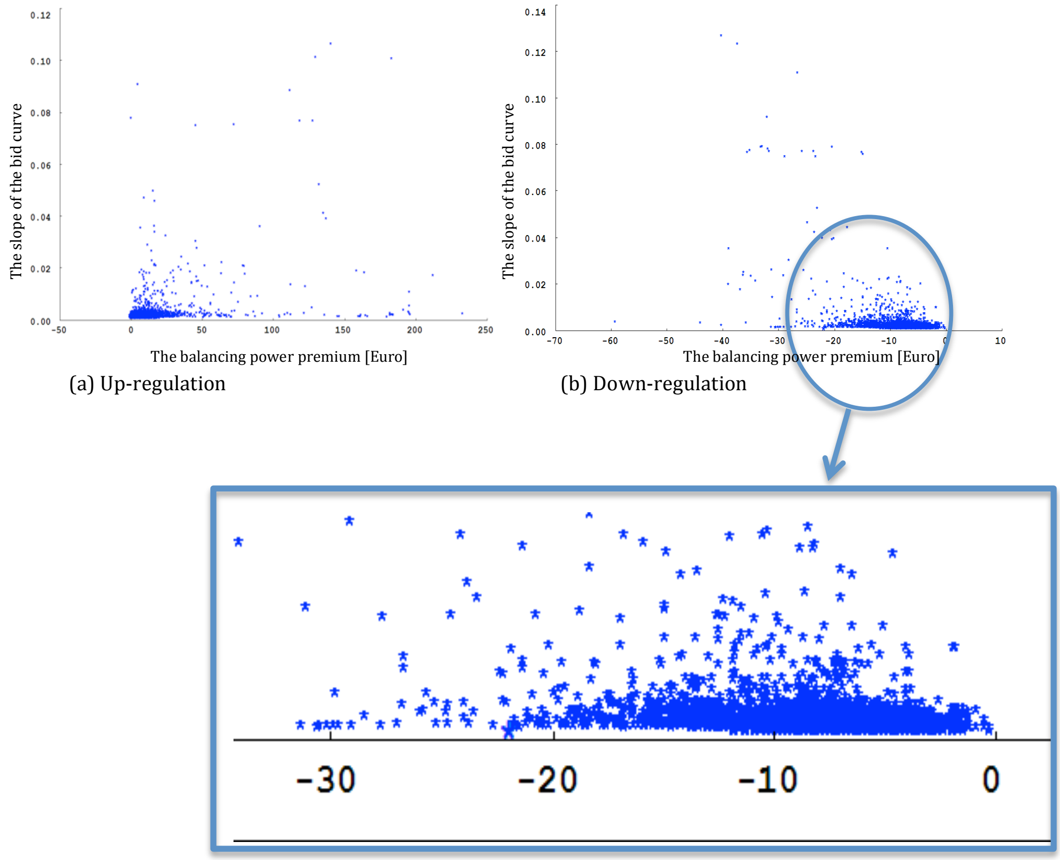


Figure 4.14. Scatter plot of the balancing power premium versus the slope of the.

As it can be seen from table 4.5, the balancing power premium is correlated with the slope of the bid curve: in the case of an up-regulation, the correlation coefficient is 0.35, and in the case of a down-regulation, the correlation coefficient is -0.44. The premium for the down-regulation will be lower as the slope of the bid curve will be steeper, and the premium for the up-regulation will be higher as the slope of the curve will be steeper. However, as one sees from figures 4.14(a) and 4.14(b) there are many outlying points that cannot be described with this relationship. The balancing power premium is more dependent on the slope of the bid curve on the working days and at night hours.

4.2.7 Correlation between the balancing power premium and the inflow

In chapter 2 the possible dependence between the balancing power price and the inflow has been described. When it comes to the relationship between the balancing power premium and the inflow, it most likely is the same. So in periods with a low inflow and a low reservoir level the premium for the up-regulation can be high, and in periods with a high inflow, especially in combination with a risk for a flood, the premium for the down-regulation can be really low.

When testing the correlation between the inflow and the balancing power premium, the analysis period has been divided into periods with a high and a low inflow:

- Winter period (from December to middle of April), when inflow is low and consumption is high
- Spring (from middle of April to middle of July), when inflow is very high due to snow melting
- Summer and autumn (from middle of July to November), when inflow can be high due to rainfall

The results of the analysis and the scatter plot of the balancing power premium and the inflow for the different periods are shown in table 4.6 and figure 4.15, respectively.

Period	Correlation coefficient	
	Up-regulation	**Down-regulation**
Winter	-0.06	-0.04
Spring	-0.04	0.16
Autumn and Summer	0.09	-0.10

Table 4.6. Correlation between the balancing power premium and the inflow.

As it can be seen from table 4.6, the correlation between the balancing power premium and the inflow for the different periods is close to zero. However, the scatter plots 4.15(c) and 4.15(f) show some relation between the inflow and the premium in the winter period and the scatter plot 4.15(d) shows some relation between the balancing power premium for an up-regulation and inflow in the summer/autumn period. It is, however, impossible to distinguish any clear trends from the figures.

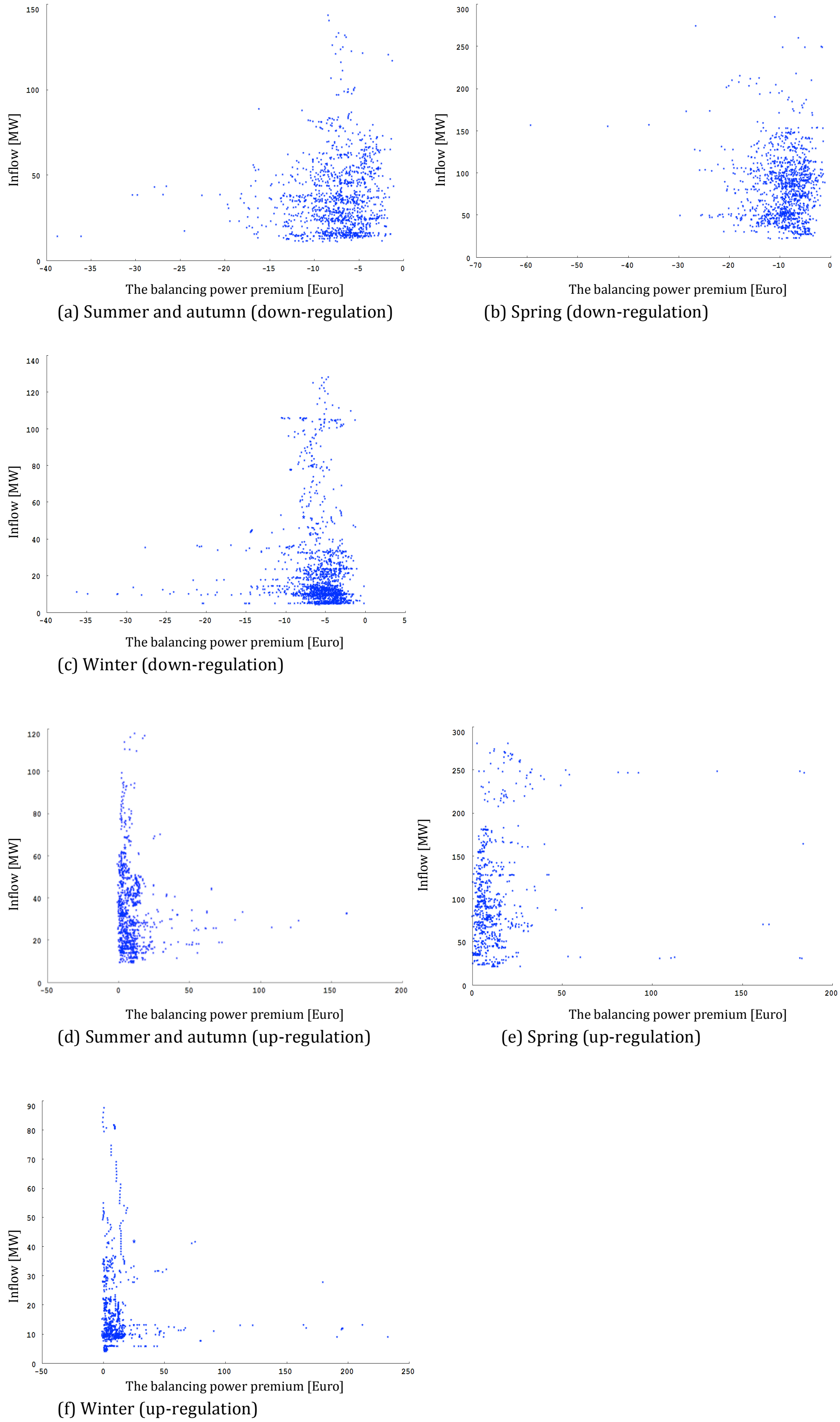


Figure 4.15. Scatter plot of the balancing power premium versus inflow.

4.2.8 Power premium in an hour and the balancing power premium in the previous hours

It is assumed that the value of the balancing power premium in an hour can be influenced by the past values of the premium. In order to prove or disapprove this assumption, the correlation between the values of the balancing power premium and the past values of the premium has been examined.

In the previous subsection, when carrying out the experiments, the balancing power premium for an up-regulation and the balancing power premium for a down-regulation have been considered separately. However, in this subsection when testing the correlation between the balancing power premium in an hour and the balancing power premium from the previous hours, one will look at an aggregated time series of the balancing power premium in price area NO3. The time series has been aggregated from the time series of the balancing power premium for an up-regulation and a down-regulation. The hours with zero values in the series correspond to the hours when the balancing power reserves have not been activated.

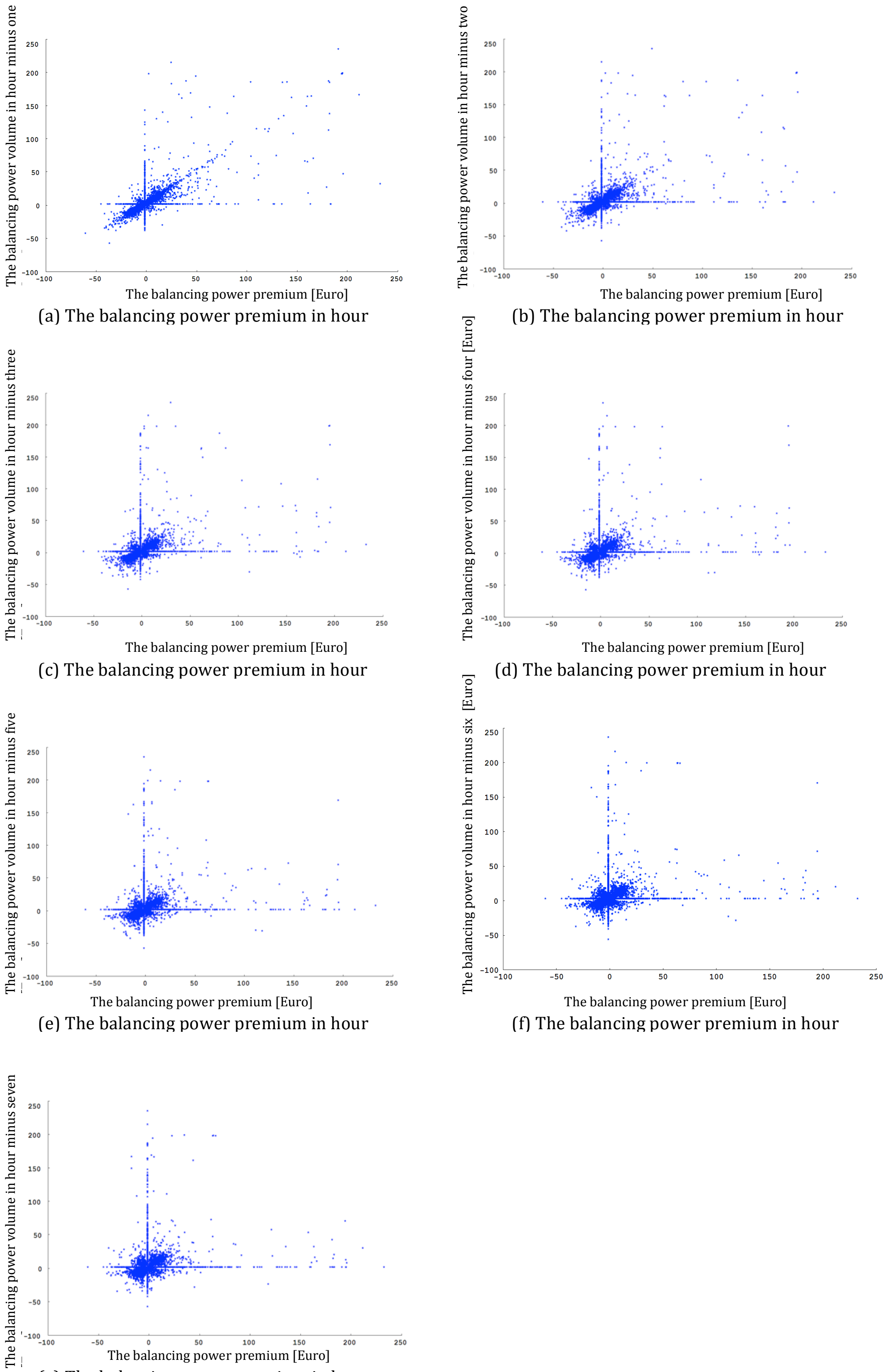
For the past values, the records of the balancing power premium that go 1, 2, 3, 4, 5, 6 and 7 hours back in time, have been taken into consideration. The correlation between the past and the future values of the balancing power premium for an up- and a down-regulation has been tested separately. The obtained correlation coefficients for each of the cases are shown in table 4.7, and the scatter plots for the values of the balancing power premium in an hour versus the past values of the balancing power volume are shown in figure 4.16.

Hours back in time	Correlation coefficient
1	0.72
2	0.48
3	0.35
4	0.27
5	0.22
6	0.20
7	0.16

Table 4.7. Results of the correlation analysis.

Both from the table and from the figures it can be concluded that the balancing power premium in an hour is correlated with the values of the balancing power premium from the previous hours. The correlation between the balancing power premium in an hour and the past values of the balancing power premium from the previous hour and one hour before the previous hour is the strongest. However, the correlation coefficient between the premium in an hour and the past values of the premium from the previous hours decreases as the past values go further back in time.

In general the relationship between the premium in an hour and the past values of the premium can be described as follows: as higher the balancing power premium in the previous hours is, the higher the balancing power premium in an hour will be, and as lower the balancing power premium in the previous hours is, the lower the balancing power premium in an hour will be.



(a) The balancing power premium in hour

(b) The balancing power premium in hour

(c) The balancing power premium in hour

(d) The balancing power premium in hour

(e) The balancing power premium in hour

(f) The balancing power premium in hour

(g) The balancing power premium in hour

Figure 4.16. Scatter plot of the balancing power premium in an hour versus the balancing power premium in the previous hours.

4.2.9 Summarizing the results

In this subsection the balancing power premium has been examined for having a daily, a weekly and an annual variation and for having a dependence on one or more factors that are supposed to influence it. The number of conclusions has been made.

The balancing power premium for the up- and down-regulation has a daily, a weekly and an annual variation. The premium in different price areas does not always follow completely the identical daily, weekly or annual patterns, but it is still possible to indicate some common attributes between the patterns.

The balancing power premium in the NO3 is correlated with the balancing power volume in the NO3, the slope of the Elspot bid curve and the spot price in the NO3. The correlation coefficient between the premium and each of these factors varies depending on the day of the week (weekday/weekend) and hour of the day (day/night) and it depends on the state of the regulation. The results of the correlation analysis have been summarized in the table below:

	Correlation coefficient					
Factor	**Balancing power volume**		**Slope of the Elspot bid curve**		**Spot price**	
Regulation state / Filter	**Up**	**Down**	**Up**	**Down**	**Up**	**Down**
All hours	0.28	0.23	0.35	-0.44	0.34	-0.36
Weekdays	0.24	0.20	0.35	-0.51	0.36	-0.49
Weekends	0.44	0.30	0.16	-0.32	0.17	-0.09
Day	0.33	0.19	0.19	-0.37	0.24	-0.08
Night	0.53	0.32	0.39	-0.48	0.38	-0.45

Table 4.8. Summarized results.

As one sees from the table, the balancing power premium does not have a strong correlation with any of the factors due to the many outlying points, but it can still be indicated some dependences from the scatter plots 4.11, 4.12 and 4.14:

- As bigger the balancing power volume is, the higher the balancing power premium in case of an up-regulation will be, and as bigger the volume activated under a down-regulation is, the lower the balancing power premium will be;
- As higher the spot price is, the higher the balancing power premium in case of an up-regulation will be, and as lower the spot price is, the higher the balancing power premium for a down-regulation will be;
- As steeper the slope of the Elspot bid curve is, the higher the balancing power premium in case of an up-regulation will be, and as steeper the slope of the bid

curve is, the lower the balancing power premium for a down-regulation will be.

The balancing power premium does not have any linear dependence on the inflow, but from the scatter plot 4.15, the relationship between the premium and the inflow can be identified.

The balancing power premium in an hour is also influenced by the past values of the premium in the previous hours. The impact of the past values of the premium on the future ones decreases as the past values go further back in time. Thus, the correlation coefficient between the balancing power premium in an hour and the balancing power premium in the previous hour is 0.72, the correlation coefficient between the balancing power premium in an hour and the balancing power premium in one hour before the previous hour is 0.48 and so on.

Chapter 5

Forecasting model for the balancing power volume and price

The aim of this master thesis is to examine the possibility of using a regression model in order to forecast the balancing power volume and the balancing power premium. When developing the forecasting model for both the balancing power volume and the premium, the regression models provided in the Microsoft Azure Machine Learning Studio will be used. In this chapter the model objectives, the stepwise development of the model and the results obtained from the model testing are represented. In section 5.1 the aim of the model is described.

The reasons for choosing the Machine Learning algorithms and the specific regression model for making the forecast of the balancing power volume and the balancing power premium are given in subsections 5.2 and 5.3, respectively. Section 5.4 describes the choice of the variables for the set of the predictors that will be used as an input into the forecasting model.

The development of the experiment/model for forecasting the balancing power volume and the balancing power premium and the results obtained under the model running are represented in subsections 5.5, 5.6 and 5.7.

5.1 Model objectives

In this chapter the possibility of forecasting the balancing power volume and the balancing power premium will be examined. As mentioned before, there are 12 price areas that exchange the balancing power reserves between each other, and the balancing power volume in a price area can be influenced by the imbalances that have occurred in the other price areas. The same comes to the balancing power premium: the balancing power premium in a price area is also influenced by the events that have occurred in the other price areas. So, in principle, when forecasting the balancing power volume and the balancing power premium, the forecast needs to be made for all of the price areas. If not, the influence of the events that takes place in the other price areas on the balancing power volume and the balancing power premium in a price area needs to be taken into consideration. However, it is too time consuming. That is why it has been decided to limit the work by forecasting the balancing power volume and the balancing power premium only in one price area.

When selecting the price area for making the forecast, the choice fell on price area NO3 due to two main reasons:

- The NO3 has some wind power production and in a couple of years the share of the power production from the wind power will increase
- Statkraft will build out more capacity for the wind power production.

When it comes to the selection of the time when the forecasting of the balancing power volume and the balancing power premium in an hour will be made, it has been decided that the forecast can earliest be carried out after the closing of the Elspot market.

One of the conclusions that Gro Klæboe (2015) have drawn in her work, is that it is impossible to forecast the balancing power price before the closing of the Elspot market. When testing the correlation between the balancing power volume and the disturbance factors that influence the balance in the power system and the correlation between the balancing power premium and the factors that influence it, the dependence of the balancing power volume and the balancing power premium on the variables that are known after the closing of the Elspot market, has been identified. If one wants to use those variables as an input into the forecasting model, the balancing power volume and the balancing power premium can be forecasted after the Elspot market closes.

A day-ahead forecast of the balancing power volume and the balancing power premium will be made. The possibility of improving the accuracy of the forecast for an hour, when having new information that will be available during the day, will also be examined.

In this thesis, the balancing power volume and the balancing power premium in an hour, which is forecasted by using the information available a day before (or by using the information that is available some hours before) in relation to the hour of the forecast are referred to as a day-ahead forecast (or X-hour-ahead forecast) of the balancing power volume and the balancing power premium. However, the output of the forecasting model in reality is an extrapolation of the balancing power volume and the balancing power premium. The reason for this is that the forecast of the volume and the premium will be made for the year 2015 and the model will be trained by using the information from the previous years. When making an actual day-ahead forecast (or X-hour-ahead forecast), the model will be trained again every day (or every hour).

5.2 The choice of the method for the balancing power volume and the balancing power premium forecast.

In order to forecast the balancing power volume and the balancing power premium in NO3, it has been decided to use the Machine Learning algorithms, more particular, the algorithms provided by the Microsoft Azure Machine Learning Studio. The choice fell on the Machine Learning algorithms, due to several reasons.

One of the reasons is an availability of the Machine Learning algorithms to adapt to new conditions and rules. The situation in the price areas continuously undergoes changes, such as new interconnection, changes in the power consumption and power production changes in the boarder of the price areas. For example, a new power line between the NO3 and the NO5 has been set into operation at the end of the 2015, and the wind power expansion in the NO3 in the coming years will cause “new conditions”, which the Machine Learning algorithms will be able to take into consideration.

Another reason is that the Machine Learning algorithms can solve complex problems and deal with non-linear relationships between variables.

The last reason is that it is possible to retrain the model that are based on the Machine Learning algorithms by taking into consideration the errors that have been made in the previous forecasts. It will not be done in this thesis, but it is a possibility for improving the forecast, when one re-runs the model every day or every hour. The forecast errors that have been made in the previous day or the previous hour can be added into the model as an extra input variable.

5.3 The choice of the regression model for the balancing power volume and the balancing power premium forecast.

When developing a forecasting model for the balancing power volume and the balancing power premium, algorithms, which are available in the Microsoft Azure Machine Learning Studio, are used. The Microsoft Azure Machine Learning Studio offers several implemented regression models. The choice of the model can be based on time required in order to conduct result, accuracy of the model performance or objectives of using the model.

The desired output of the future forecasting model is a value of the balancing power volume and the balancing power premium. Knowing the desired model output, the opportunities one can choose from between the models, can be narrowed down by not taken into consideration the regression models, and only consider the models, which have as a result, a numerical value of a variable of interest. From the algorithms cheat sheet, one sees that the Ordinal regression model is used to predict the ranked values, the Poisson regression model is used to predict event counts and the Fast Forest Quantille regression model is used to predict the distribution of the predicted variable. Which is why these models will not be considered.

The Bayesian regression model suits best for the small datasets. The dataset, which will be used in order to forecast the balancing power volume and the balancing power premium, contains over x variables, in which all of them have 15000 records each, is considered as a big one. This is why the Bayesian regression model has been removed from the choices as well.

The Linear regression model that uses the ordinary least squares method or the online gradient descent method, has been found to be too simple in order to catch the dependences between the balancing power volume/price and the explanatory variables that can be non-linear.

The Neural Network regression model is used widely for the deep learning and the modelling of the complex problems. This model is, for example, used for the spot price forecasting. The Neural Network regression uses a complex algorithm and the performance of the regression is sensitive to the chosen tuning parameters. So it was decided to not consider this algorithm in this master thesis.

The Decision forest regression model and the Boosted decision tree regression model will be used in order to forecast the balancing power volume and the balancing power premium. It is difficult to make a choice between these models, since both models provide a good accuracy of the forecasts and have fast training time. So it was decided that both of the models will be run parallel for the same dataset and their performance will be evaluated and compared.

5.4 Choice of predictors

In chapter 4, the correlation between the balancing power volume/premium and the different factors, which are supposed to have an influence on the volume/premium in the Balancing power market, was examined. It was done in order to find variables that can be used as explanatory variables for the balancing power volume/premium when developing the forecasting model.

However, not all factors that have been considered in the correlation analysis can be used as predictors in the regression model.

5.4.1 Predictors for the balancing power volume

Under the correlation analysis for the balancing power volume the variables such as Elbas volume, special regulation, deviation between the actual and the planed power flow, deviation between the actual and the forecasted consumption, and deviation between the actual and the forecasted wind power production in an operating hour were claimed to have influence on the balancing power volume.

When making forecast of the balancing power volume in an hour the values of special regulation, the actual power flow, the actual power consumption and the actual wind power production in this hour will be unknown. So such variables as the special regulation, the deviation between the actual and the planed power flow, the deviation between the actual and the forecasted consumption, and the deviation between the actual and the forecasted wind power production can not be used as explanatory variables in the analysis. However, the power consumption forecast, the wind power production forecast and the planed power flow (spot power flow) are known when forecasting the balancing power volume for the next day or for the next hour, that is why it has been decided to include these variables in the set of predictors.

The Elbas volume in an operating hour can only be used as a predictor in really short-term forecasts, i.e. one hour before the operating hour. Since the trading in the Elbas closes one hour before the operating hour and bids in the Balancing Power Market can be adjusted at latest, 45 minutes before the operating hour (see figure 5.1), and the market actors would have a possibility to make a new forecast of the balancing power volume with the Elbas volume taken into consideration.

Due to the fact that the Elbas volume can be used only in a short-term forecast of the balancing power volume, and its influence on the balancing power volume is controversial, it was decided to not take it into consideration when making the forecast.

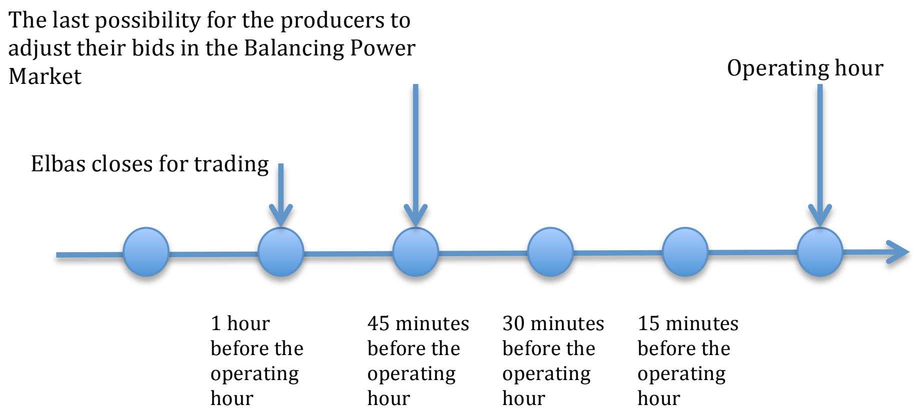


Figure 5.1. Structure of the power market.

Some implicit dependence between the balancing power volume and the deviation between the actual and the forecasted temperature was identified as well in chapter 4. The actual temperature will not be known at the moment of making the forecast, the deviation between the actual and the forecasted temperature cannot be used as a predictor, but the temperature forecast will be known and that is why it can be used as an explanatory variable.

The past values of the balancing power volume can also be used as a predictor for the balancing power volume in an hour. However, in subsection 4.1.6, it has been found out that the past values of the balancing power volume that do not lay too long back in time, has the strongest correlation with the balancing power volume in an hour. That is why it is quite uncertain if it is possible to use the past values of the balancing power volume to make the forecast of the balancing power volume in an hour for many hours ahead or not. Due to this issue the past values of the balancing power volume will be considered when making the forecast, but not in first place.

5.4.2 Predictors for the balancing power price

The correlation analysis that has been carried out in order to find any relationships between the balancing power price and the factors, which are supposed to influence it, has shown dependence of the balancing power premium in an area on the balancing power volume, spot price in the price area and the slope of the Elspot bid curve. All these variables, except the balancing power volume are known at the moment of making the balancing power price forecast, and can be used as predictors.

The actual balancing power volume in an operating hour is not be known at the moment of the forecast, but the forecast for the volume will be available before forecasting the balancing power price and it can be used as a predictor.

The inflow in price area NO3 has not shown to have any linear relationship with the balancing power premium in the area, but as it can be seen from figure 4.15 in subsection 4.2.7, a sort of a non-linear relationship between them is possible. That is why it has been decided to include the inflow in the NO3 into the set of predictors and examine whether or not it can be used in order to explain the values of the balancing power premium.

The balancing power premium in an hour has also been found to correlate with the balancing power premium from the previous hours. However, similar to the balancing power volume, the correlation between them will be lower as the past values of the balancing power premium that goes too long back in time are used, in order to predict the values of the balancing power premium in an hour. So if by using the past values of the balancing power volume in the set of predictors, it will, most likely, be possible to conduct only a short-term forecast of the premium. That is why the time series containing the past values of the balancing power premium will be considered as an explanatory variable, but only after examining the possibility of forecasting the balancing power premium without using its past values as a predictor.

5.4.3 Variable transformation

The balancing power volumes claim to have a daily, weekly and annual variation. In order to be able to reflect these variations in the model, this periodicity needs to be presented as a variable. Three variables will be created, which are: hour of the day, hour of the week and hour of the year. These variables have been transformed into polar coordinates via (sin;cos) pair of variables.

Following formula for the transformation has been used:

$$x \rightarrow [y_1; y_2] = [\sin\left(\frac{2\pi x}{\omega}\right); \cos(\frac{2\pi x}{\omega})] \qquad \text{(Eq. 5.1)}$$

where x is a variable to be transformed and ω is a period of variation. (Busseti, Osband & Wong, 2012) The period for the daily variation is 24 hours, the weekly variation is 168 hours and the annual regulation is 8760 hours. The result of the transformation is shown in figure 5.2.

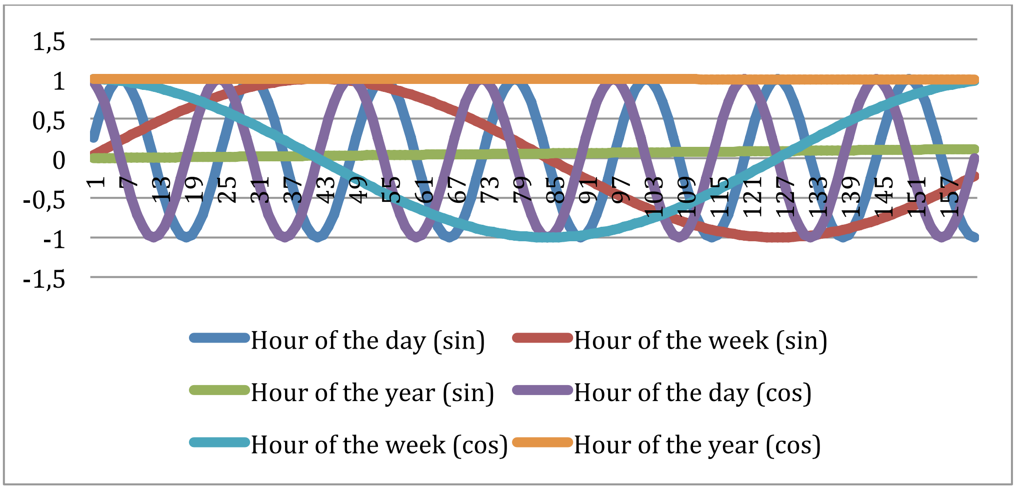


Figure 5.2. Transformation to the polar coordinates.

It should be pointed out that by transforming the variables the piecewise linear representation of the hour of the day, week and year will be avoided. In figure 5.3, the variables, which have not been applied on the transformation to the polar coordinates, are shown. Having variables in the linear form will lead to additional splits in the model, in the places where it can be avoided. For example, the hours of the day denoted from 0 to 23. In the model, the balancing power volume in hour 0 and hour 23 will be considered as having different characteristics and will be split, even though the characteristics of the volumes in these hours are most likely the same and could actually be treated together. By representing hour of the day, hour of the week and hour of the year as continuous variables, unnecessary splits in the model will be avoided.

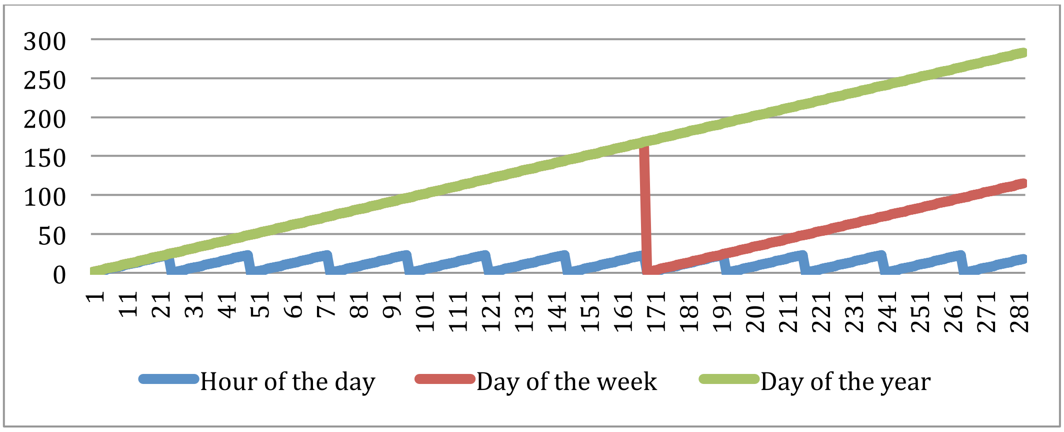


Figure 5.3. The variables before the transformation.

[*]The transformation of the hour of the day, hour of the week and hour of the year into the polar coordinates will be used as input variables for the model, and in the following subsections it will be referred to as the hour of the day, hour of the week and hour of the year.

5.4.4 Final set of predictors

The predictors that will be used in order to forecast the balancing power volume and the balancing power price are gathered in the table below:

Predictors for the balancing power volume	Predictors for the balancing power price
Hour of the day	Hour of the day
Hour of the week	Hour of the week
Hour of the year	Hour of the year
Wind power production forecast NO3	Balancing power volume NO3
Power consumption forecast NO3	Inflow NO3
Spot power flow NO3 NO4	Slope of the bid curve
Spot power flow NO3 SE2	Spot price NO3
Temperature forecast NO3	Balancing power premium in hour t-1
Balancing power volume in hour t-1*	Balancing power premium in hour t-2
Balancing power volume in hour t-2	Balancing power premium in hour t-3
Balancing power volume in hour t-3	Balancing power premium in hour t-4
Balancing power volume in hour t-4	

Table 5.1. The set of predictors.

[*] t is a designation of the hour for which the forecast is made.

In order to be able to use those variables in the Microsoft Azure Machine Learning Studio, the variables need to be gathered into the CSV-file together with the balancing power volume and the balancing power premium. It has been decided to create two separate CSV-files for forecasting the balancing power volume and the balancing power premium.

The CSV-file that will be used in order to forecast the balancing power volume contains the predictors for the balancing power volume from the table above, and the actual values of the balancing power volume in an hour. The CSV-file that will be used in order to forecast the balancing power premium contains the predictors for the balancing power premium from the table above, and the actual values of the balancing power premium in an hour.

5.4.5 Missing values

Data of the power consumption forecast and the wind power production forecast has 31% of missing values. These two variables are used as predictors for the balancing power volume forecast. So when developing the model for the balancing power volume, it is necessary to handle the missing values in these time series. In order to pre-process data, the Clean Missing Data module will be used.

Clean Missing Data gives the possibility to either remove the missing values, to replace them with mean, placeholder or other value, or to completely remove rows with the missing values. When using the Clean Missing Data, the initial dataset will not be changed. The module creates a new dataset in the workspace and will use it in the subsequent workflow. Clean Missing Data also gives an option to save the cleaned dataset and reuse it. (Microsoft Azure D, 2016)

It was decided to clean the missing values in the data of the power consumption and the wind power production forecast by removing all the rows in the dataset that contains missing values. The decision was made based on that the dataset has 26280 hours of registered data; so removing 31%, i.e. 8256 hours, of the records will not influence the quality of model training and the model scoring in large extent.

5.5 Hyper Parameter Optimization

One of the conditions for getting the best predictive model is to choose the right values for the algorithms tuning parameters. There are no fixed rules or guidance that would help in advance to determine the best parameters. So in order to find the parameters that will give the best forecast, a hyper parameter optimization has to be done manually or by the use of the Tune Model Hyperparameters module.

When doing the parameter optimization manually the n models with m parameter set has to be tested. It is also necessary to keep track on the testing results in order to be able to compare them and to make a decision about which parameter set will ensure the best model performance.

On the other hand, the Tune Model Hyperparameters module gives the possibility to build and train the model by using different combinations of settings and to find out optimum model parameters for the given dataset. The module supports two different methods in order to obtain the optimum parameters: Integrated train and tune that uses a parameter sweep to train a model and Cross validation with tuning that divides the data into some numbers of folds, and then build a model for each fold and then tests them to identify the best parameters for each fold. Finally, when the optimum parameter settings are found, they will be used to train the entire training data. (Microsoft Azure E, 2016)

In order to find the optimum parameters for the model the Tune Model Hyperparameters module will be used. The random sweep of the parameters will be performed. The maximum number of sweeps is set to 30.

5.6 Forecasting the balancing power volume

5.6.1 Constructing the experiment for the balancing power volume forecast

The experiment constructed in order to forecast the balancing power volume is shown in figure 5.4. The dataset is divided into two sets by using the Split Data module. The dataset is divided by rows, and in order to ensure that the data that has been used for the model training will not be used for the model testing, the Randomized split option has been disabled. The first set, which is 70% of the original dataset will be used to train the model, the second set that represents the remaining 30% of the original dataset will be used to generate the predictions. The data from 2013 to 2014 will be used for the model training and the data from 2015 will be used for the model testing.

The Decision Forest regression model and the Boosted Decision Tree regression model are used in the experiment. Training and testing of each model will be run independently and the performance of each model will be evaluated. The result from the model evaluation will then be used in order to compare these models and identify the most suitable model for the given prediction task and dataset.

The Score Model module generates predictions using a trained model, and the dataset, which has been put aside for testing. The output of this module for the Decision Forest regression model is a scored label standard mean and a scored label standard deviation. For the Boosted Decision Tree regression the output of this module is scored labels.

The Evaluate Model module is used to determine the accuracy of the models and to compare their performances. The module calculates Negative Log Likelihood, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Relative Squared Error and Coefficient of Determination. The description of these values can be found in subsection 3.3.

The Select Column in the Dataset module is added to the experiment in order to be able to choose which of the columns in the dataset that will be used in the experiment. The Edit Metadata module is used to change the metadata associated with the dataset, if necessary.

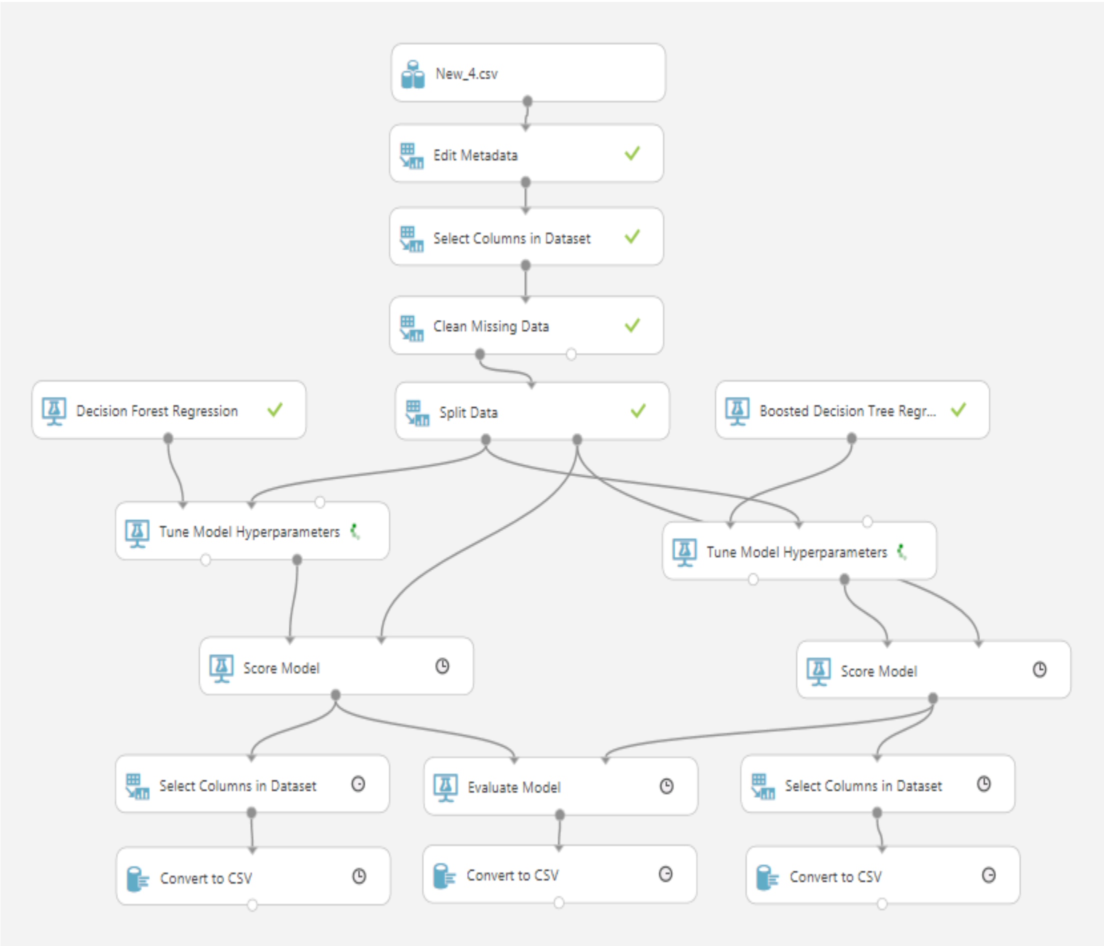


Figure 5.4. Structure of the experiment.

5.6.2 Forecasting the balancing power volume without considering the past values of the balancing power volume

The balancing power volume has been forecasted using the experiment that is shown in figure 5.4 and the dataset that contains the following attributes: the hour of the day, the hour of the week, the hour of the year, the power consumption forecast, the wind power production forecast, the spot power flow between NO3 and NO4, the spot power flow between NO3 and SE2 and the temperature forecast.

The accuracy of the models has been evaluated by using the Mean Absolute Error and Root Mean Squared Error. The correlation coefficient between the forecasted and the actual values of the balancing power volume has also been considered when estimating the quality of the model performance.

When training both the Decision Forest regression model and the Boosted Decision Tree regression model, 30 different sets of the model parameters have been tested in order to find the optimal parameter settings, which will provide the best model output.

The three best parameter settings and the corresponding Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) for each of the regression models are given in tables 5.2 and 5.3. All parameter sets that have been considered under the training of the models are represented in Appendix G.

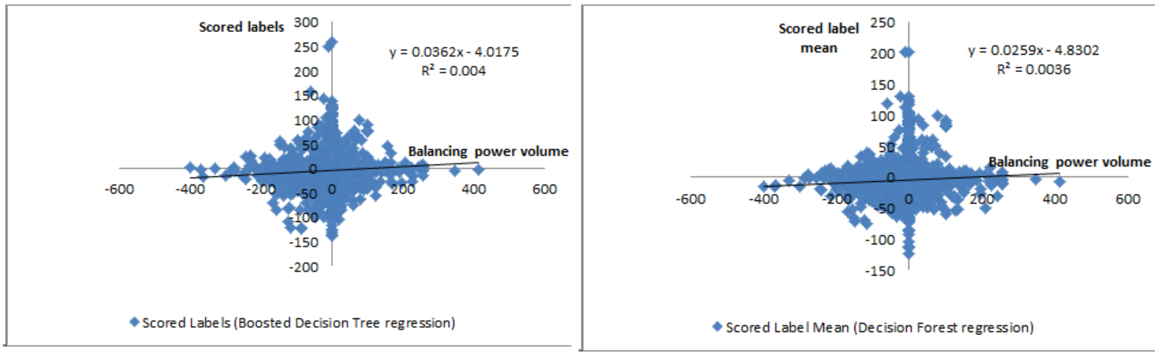
Boosted Decision Tree regression						
Parameter set	**Number of leaves**	**Minimum leaf instances**	**Learning rate**	**Number of trees**	**MAE**	**RMSE**
1	117	2	0.22	482	12.13	24.40
2	68	3	0.13	258	12.49	24.79
3	124	4	0.22	94	12.90	25.19

Table 5.2. Parameter settings for the Boosted Decision Tree regression model.

Decision Forest regression						
Parameter set	**Minimum number of samples per leaf node**	**Number of random splits per node**	**Maximum depth of the trees**	**Number of trees**	**MAE**	**RMSE**
1	2	200	32	19	13.79	28.03
2	2	627	36	21	13.73	28.08
3	4	155	29	14	14.52	29.36

Table 5.3. Parameter settings for the Decision Forest regression model.

Figure 5.6 is the graph of the forecasted values of the balancing power volume for a one-week period, and figure 5.5 represents the scatter plot for the actual values versus the forecasted values of the balancing power volume.



a) Boosted Decision Tree regression b) Decision Forest regression

Figure 5.5. Scatter plot of the actual values of the balancing power volume versus the forecasted ones.

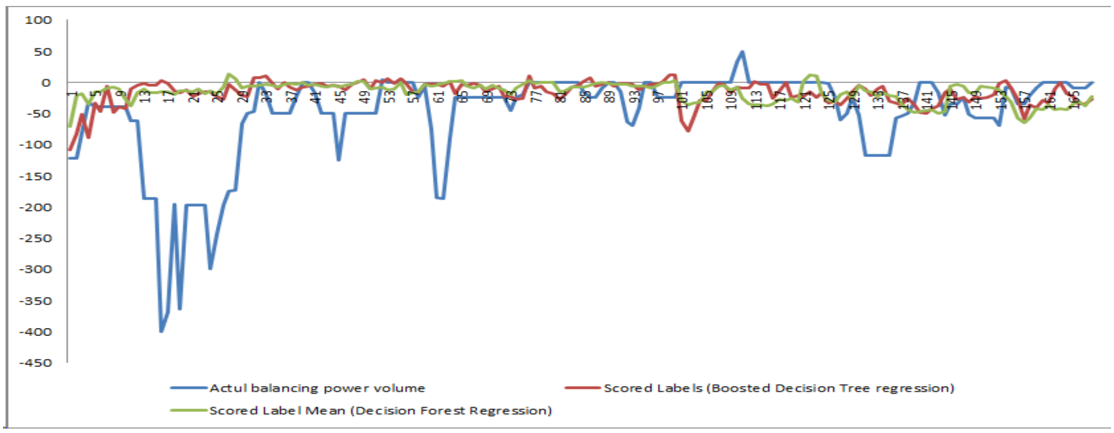


Figure 5.6. Forecast of the balancing power volume for a week.

From both figures one sees that neither the Boosted Decision Tree regression model nor the Decision Forest regression model have managed to make a good prediction of the balancing power volume day ahead, by using the given set of predictors.

When evaluating the performance of the models, two aspects in the forecasted results have been studied: the capability to determine the right regulation state, and the capability to capture the magnitude of the balancing power volume.

The balancing power volume has been forecasted in 5357 hours, and only in about 20% of the cases the models managed to predict the right regulation state. The models failed to predict the up-regulation in 94% of the cases and the down-regulation in 25% of the cases, and the models did not distinguish the balance in the system as a regulation state at all.

When it comes to the capability of the models to make a correct forecast on the size of the activated balancing reserves, it can be seen from figure 5.6 that both of the models that are used are insufficient to do this. Even though it is not required from the models to predict an exact value of the balancing power volume, but to be able to indicate an approximate magnitude of the expected balancing power volume, i.e. to indicate the peaks in the balancing power volumes for both the up- and down-regulation, they will fail to do this. The models do not manage to indicate hours where the balancing power volume under the up- or down-regulation will have it peaks. In hours with no regulation the models, in many cases, predict high values for the up- or down-regulation.

The insufficiency of the model is reflected in the high Mean Absolute Error (MAE), the high Root Mean Squared Error (RMSE) and the low correlation coefficient between the actual and the forecasted values of the balancing power volume, shown in table 5.4.

Estimator / Model	MAE	RMSE	Correlation coefficient
Boosted Decision Tree regression	25.90	47.14	0.01
Decision Tree Forest regression	27.73	49.58	-0.01

Table 5.4. Forecast error.

5.6.3 Forecasting the balancing power volume while considering the past values of the balancing power volume

In the previous subsection, the values of the balancing power volume for the next day were forecasted by using the set of predictors that are available at the moment of making the forecast. It can be noticed that when making the forecast of the balancing power volume in an hour, the models do not take into consideration the forecasts that have been made for the hours earlier. Both the Boosted Decision Tree regression model and the Decision Forest regression model, when forecasting the target variable, do not consider the past values of the target variable as long as they are not added explicit as variables in the dataset, since the models will process the dataset as separate rows both under the model training and the generating of the forecast.

However, in subsection 4.1.6, it was shown that the values of the balancing power volume from an hour correlate with the values of the balancing power volume from the previous hours. So that having the values of the balancing power volume from the previous hours as one of the explanatory variables, will most likely give a better forecast result. It has been clear that the set of the predictors used in the forecast does not manage to explain the values of the balancing power volume and it is impossible to make the day-ahead forecast of the volume by using only these predictors.

To examine the effect of the past values of the balancing power volume on the quality of the forecast the times series containing the values of the balancing power volume from the previous hours, have been added to the set of the predictors. The dataset in this experiment contains the following attributes: hour of the day, hour of the week, hour of the year, power consumption forecast, wind power production forecast, spot power flow between NO3 and NO4, spot power flow between NO3 and SE2, temperature forecast and the past values of the balancing power volume.
When using the past values of the total imbalance as an explanatory variable, it is important to find out which of the time series containing the past values of the total imbalance, is relevant to use as a predictor, as they go further back in time. From the correlation analysis in subsection 4.1.6, it can be seen that as longer back in time the past values of the balancing power volume goes, the less influence they will have on the future ones. So the values of the volume that do not go more than a couple of hours back in time will provide the best information about the future volume values. When examining the influence of the past values of the balancing power volume on the forecast accuracy, it was decided to forecast the balancing power volume in an hour, by using the past values of the time series from the previous hour first, then using the values from one hour earlier than the previous hour, and then using the values from two hours earlier than the previous hour and finally, by using the values from three hours earlier than the previous hour.

The model shown in figure 5.4 has been run four times, each time the past values of the balancing power volume from different hours were considered alternately. From tables 5.5 and 5.6, which contains the results from the testing, one sees that the balancing power volume can be forecasted with a good enough accuracy only 1-2 hours ahead. In other words, only the adding of the past values of the balancing power volume that lays no longer than two hours back in time, improves the accuracy of the forecast.

The model gives the best forecast for an hour, when having the past values of the balancing power volume from the previous hour and the values of the balancing power volume from one hour before the previous hour. In figures 5.7 and 5.8 the forecasted values of the balancing power volume for a week have been plotted. It can be seen that when having the past values of the balancing power volume from the previous hour, the performance of the model improves significantly, compared to when having the set of the predictors that was used to predict the balancing power volume in the previous subsection. The model gives a good forecast of both the magnitude of the balancing power volume and the expected regulation state in an hour. The correct regulation state has been predicted at about 72% of the cases and in around 77% of the cases (a small deviation of ±2MW is allowed between the actual and the predicted values) when having the past values of the balancing power volume from the previous hour and one hour before the previous hour respectively.

From tables 5.5 and 5.6 it can be noticed that the error of the forecast of the balancing power volume is lower, when having the values of the balancing power volume from the previous hour compared to when having the values of the balancing power volume from one hour before the previous hour in relation to the hour the forecast has been made for. However, when using the past values of the volume from one hour before the previous hour, the model predicts the correct regulation state more frequently. A higher forecast error when using the past values of the volume from one hour before the previous hour is caused by that the model gives a worse indication of the magnitude (the size) of the balancing power volume in this case.

It can be noticed that the forecasted values lag the actual values of the balancing power volume by one hour when having the time series containing the past values of the balancing power volume from one hour back in time as one of the predictors. This trend also has been noticed in cases when one predicts the balancing power volume in an hour, by using the past values of the volume that goes two, three and four hours back in time. Only in these cases the forecasted values have lagged the actual values by two, three and four hours respectively.

Decision Forest regression			
Estimator / **Additional variable**	**MAE**	**RMSE**	**Correlation coefficient**
-	27.73	49.58	-0.01
The balancing power volume in hour t-1	11.37	25.55	0.81
The balancing power volume in hour t-2	16.29	34.48	0.61
The balancing power volume in hour t-3	25.42	44.99	0.09
The balancing power volume in hour t-4	25.84	45.46	0.09

Table 5.5. Forecast error for the Decision Forest regression model.

Boosted Decision Tree regression			
Estimator / **Additional variable**	**MAE**	**RMSE**	**Correlation coefficient**
-	25.90	47.14	0.01
The balancing power volume in hour t-1	12.09	25.89	0.80
The balancing power volume in hour t-2	21.08	39.27	0.45
The balancing power volume in hour t-3	29.20	48.68	0.05
The balancing power volume in hour t-4	28.37	48.04	0.06

Table 5.6. Forecast error for the Boosted Decision Tree regression model.

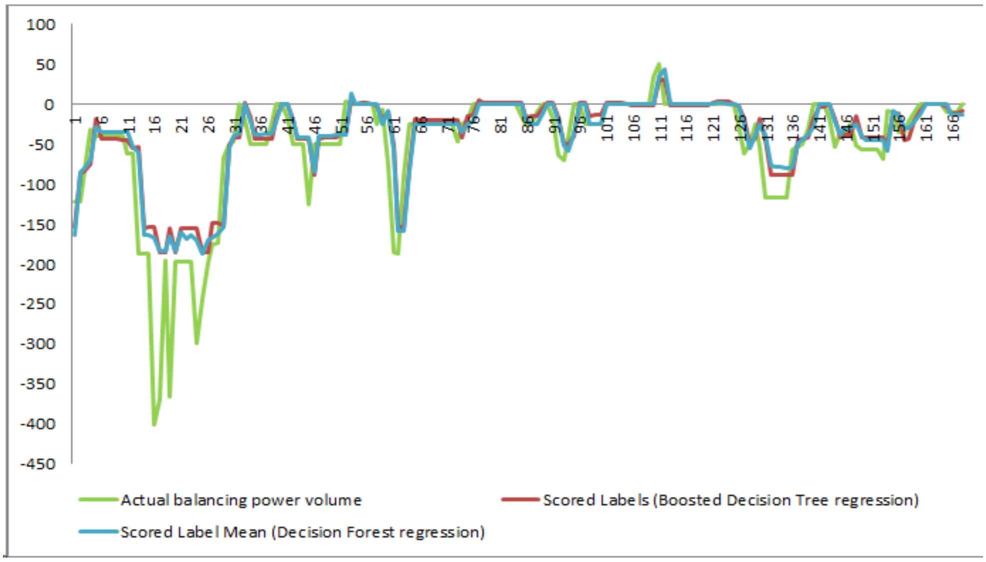


Figure 5.7. Forecast of the balancing power volume for a week with the balancing power volume from t-1 in the set of the predictors.

It can be seen that when having the time series of the past values of the balancing power volume that goes one or two hours back in time, in the set of predictors, the accuracy of the forecast has been significantly improved. So most likely, the past values of the balancing power volume from the previous hours have the biggest impact on the forecasted value of the balancing power volume in an hour, and it can probably be used alone when making the forecast. In order to investigate this possibility and study the importance of each of the predictors to the models, the impact of each variable from the set of predictors that has been used to make the forecast, has been examined. One by one, the predictors have been removed from the set, and the Mean Absolute Error and the Root Mean Squared Error have been obtained for each of the cases.

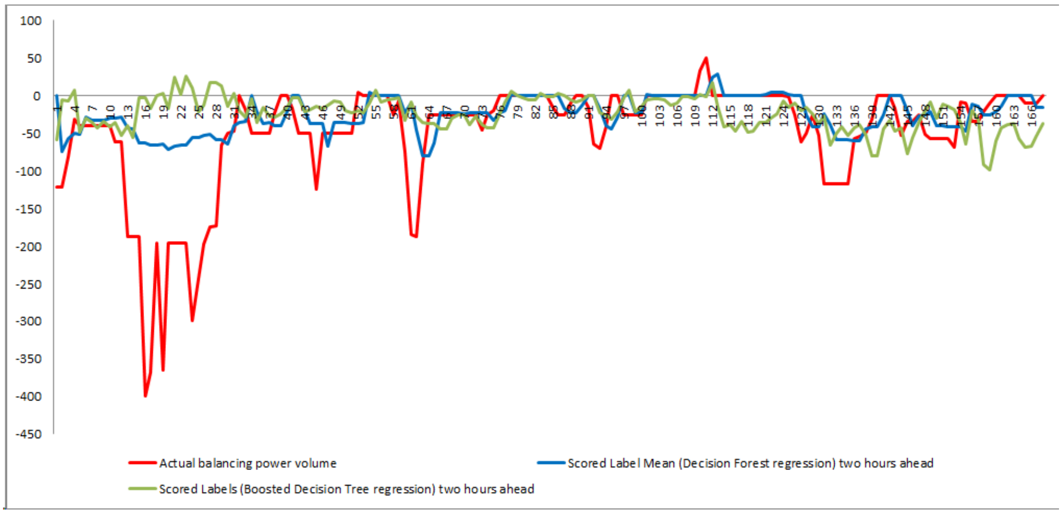


Figure 5.8. Forecast of the balancing power volume for a week with the balancing power volume from t-2 in the set of the predictors.

Tables 5.7 and 5.8 show how the Mean Absolute Error and the Root Mean Squared Error for each of the regression models changes when removing the temperature forecast from the set of the predictors first, then removing the wind power production forecast and so on. The result can be seen in the tables below:

Regression		**Decision Forest regression**			**Boosted Decision Tree regression**		
Number of variables in the set of predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
9	-	11.38	25.55	0.81	12.09	25.89	0.80
8	Temperature forecast	11.37	25.57	0.81	12.05	25.89	0.80
7	Wind power production forecast	11.41	25.66	0.81	12.50	26.63	0.79
6	Power consumption forecast	11.40	25.61	0.81	12.39	26.66	0.79
4	Spot flow NO3-NO4 and NO3-SE2	11.39	25.69	0.81	11.76	25.87	0.80
1	Daily, weekly and annual variation	11.21	25.58	0.81	11.26	25.73	0.81

Table 5.7. Sensitivity of the forecast error when having the balancing power volume from t-1 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
9	-	16.29	34.48	0.61	21.08	39.27	0.45
8	Temperature forecast	16.32	34.40	0.62	20.40	39.16	0.45
7	Wind power production forecast	16.32	34.48	0.62	21.49	39.34	0.46
6	Power consumption forecast	16.32	34.39	0.62	20.40	38.46	0.48
4	Spot flow NO3-NO4 and NO3-SE2	16.01	34.70	0.61	21.13	40.88	0.42
1	Daily, weekly and annual variation	16.10	34.22	0.62	16.28	34.41	0.61

Table 5.8. Sensitivity of the forecast error when having the balancing power volume from t-2 as a predictor.

From the table it can be concluded that the Decision Forest regression model gives an almost equally accurate forecast when only having the past values of the balancing power volume as a predictor and when having the past values of the balancing power volume as one of the predictors in the set of predictors. Hence the past values of the volume that go no longer than two hours back in time will dominate, and other variables will not have any impact on the model output.

When it comes to the Boosted Decision Tree regression model, the over-fitting of the model occurs when having the past values of the balancing power volume along with the other variables in the set of predictors. The Boosted Decision Tree regression model gives a more accurate forecast of the balancing power volume in an hour when only having the past values of the balancing power volume from the previous hours in the set of predictors.

5.6.4 Forecasting the total imbalance in the NO3 by using the set of the predictors

In the two previous subsections the possibility of making the forecast for the balancing power volume activated in the NO3 for the next day by using the regression models and the set of the predictors has been examined. The accuracy of the obtained forecast when having the set of the predictors that does not contain the past values of the balancing power volume is rather low. However, when making the prediction of the balancing power volume in price area NO3, the influence of the imbalances that have occurred outside of the price area, have not been taken into consideration. The quality of the forecast can possible be improved if the influence of the events that have occurred in the other price areas, will be taken into account. The influence of the factors from the other price areas on the balancing power volume in the NO3 can be taken into account by considering the deviation between the actual and the spot power flow between the NO3 and its neighbour price areas.

Based on that, it was decided to examine the possibilities of forecasting the total imbalance in the NO3 by utilizing the same regression models and the same set of the predictors that have been used in order to forecast the balancing power volume.

The total imbalance has been defined as the sum of the balancing power volume activated in the NO3 and the deviation between the actual and the planed power flow between the NO3 and the neighbour areas: NO1, NO4 and SE2. The power flow between the NO3 and the NO5 will not be taken into consideration since the time period that is used for the training and the testing of the model; the line between them has not been in operation yet. The special regulation volume will not be taken into consideration as well since it has been shown in subsection 4.1.4 that it contributes little to the removal of the imbalance. So the total imbalance in the NO3 has been defined as follows:

The total imbalance = The balancing power volume
+ Actual power flow NO3-NO1
+ Actual power flow NO3-NO4
+ Actual power flow NO3-SE2
- Planed power flow NO3-NO1
- Planed power flow NO3-NO4
- Planed power flow NO3-SE2

Using the model/experiment shown in figure 5.4 and the set of the predictors described containing the following attributes: hour of the day, hour of the week, hour of the year, power consumption forecast, wind power production forecast, spot power flow between the NO3 and the NO4, spot power flow between the NO3 and the SE2 and temperature forecast; the total imbalance in the NO3 has been forecasted.

The accuracy of the models has been evaluated by using the Mean Absolute Error and the Root Mean Squared Error. The correlation coefficient between the forecasted and the actual values of the total imbalance in the NO3 will also be considered when estimating the goodness of the forecast.

In tables 5.9 and 5.10 the three best parameter settings and the corresponding Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for each of the regressions, are shown. All parameter sets that have been considered under the training of the models are represented in Appendix H.

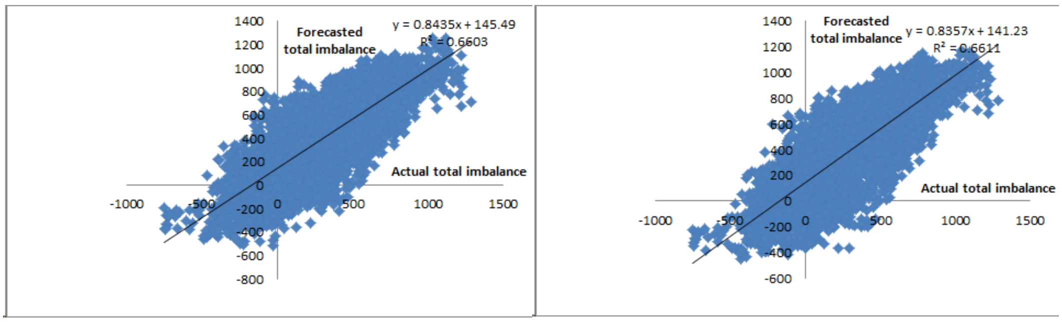
Boosted Decision Tree regression						
Parameter set	Number of leaves	Minimum leaf instances	Learning rate	Number of trees	MAE	RMSE
1	108	31	0.06	376	87.92	119.83
2	117	2	0.22	482	86.06	119.96
3	124	4	0.22	94	87.62	120.43

Table 5.9. Parameter settings for the Boosted Decision Tree regression model.

Decision Forest regression						
Parameter set	Minimum number of samples per leaf node	Number of random splits per node	Maximum depth of the trees	Number of trees	MAE	RMSE
1	2	200	32	19	86.76	121.39
2	2	627	36	21	87.27	122.54
3	4	318	16	25	89.22	123.51

Table 5.10. Parameter settings for the Decision Forest regression model.

By using the most optimal parameter settings, the models have obtained the forecast of the total imbalance in the NO3. In figures 5.10 and 5.11 the scatter plot of the actual versus the forecasted total imbalance in the NO3 and the graph of the forecasted values of the total imbalance for a one-week period are respectively shown.



a) Boosted Decision Tree regression b) Decision Forest regression

Figure 5.10. Scatter plot for the actual versus the forecasted values of the balancing power volume.

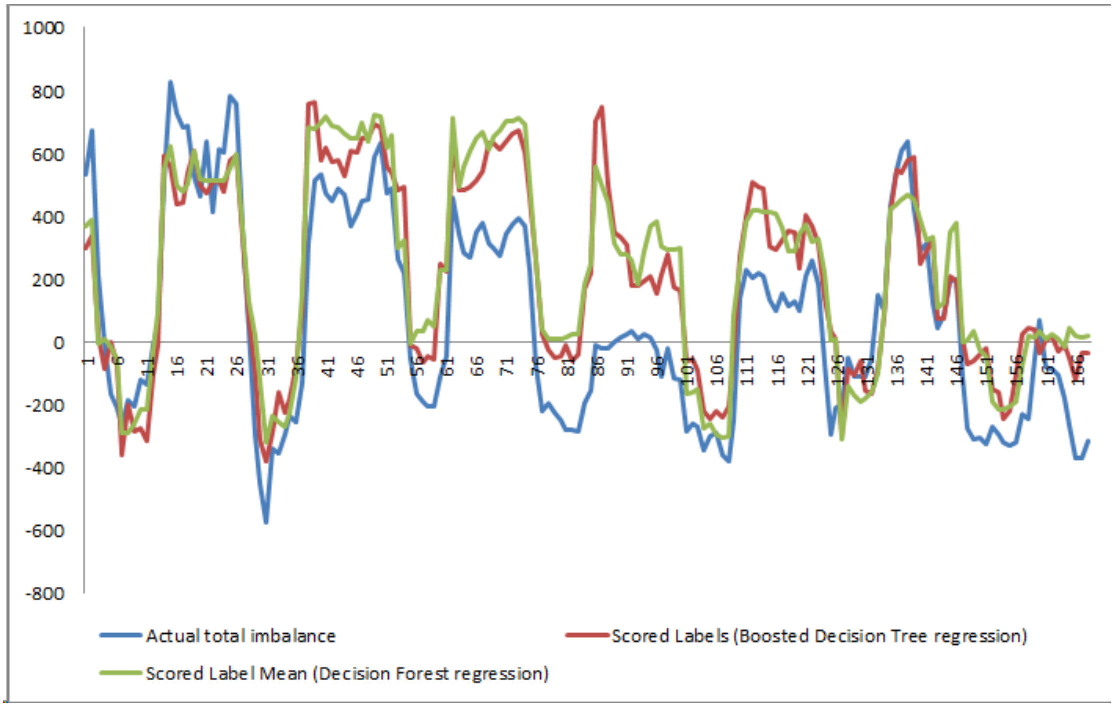


Figure 5.11. Forecast of the balancing power volume for a week.

From both figures it can be seen that both the Boosted Decision Tree regression model and the Decision Forest regression model have manage to make a good enough day ahead prediction of the total imbalance in the NO3 by using the given set of predictors. The models fail to predict the exact values of the total imbalance, but the output of the models gives a good indication of the imbalance direction, i.e. the need for an up- or a down-regulation in the area, and of the expected magnitude of the imbalance. The total imbalance in the NO3 has been predicted for 5357 hours, and the correct state of the regulation has been indicated in 87% of the cases.

However, it can be noticed that the Decision Forest regression in many hours gives a more accurate result than the Boosted Decision Tree regression. That is also reflected in the value of the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) for each of the regression and the correlation coefficient between the actual and the forecasted total imbalance in the NO3, that are shown in table 5.11. It can be seen that the values of the errors are quite high, even though the correlation coefficient between the actual and predicted imbalance is high. The reason for this is that the models have managed to capture the change in the magnitude of the imbalance, but in many cases the forecasted value will lay some MW below or above the actual ones.

Despite of the deviation between the actual and the forecasted values of the imbalance, the models are able to capture the peaks in the total imbalance and predict the right direction of the imbalance in around 86% of the cases. It should be noticed that in all hours which the forecast have been made for, there have been some imbalances in the NO3, and the models had to distinguish only between the need for the up-regulation and the need for the down-regulation.

Estimator Model	MAE	RMSE	Correlation Coefficient
Boosted Decision Tree regression	186.32	230.93	0.81
Decision Tree Forest regression	183.58	226.86	0.81

Table 5.11. Forecast error.

The importance of each of the predictors in order to obtain the forecast result has been examined. One variable at a time will be removed from the set of the predictors, and the remaining variables in the set will be used to forecast the balancing power premium. By changing place with the first variable, one by one of the variables will be taken out of the set, and consequently the variable that was already taken out, will be taken back into the set. Each time a variable changes place, a forecast of the balancing power premium is done and the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the correlation coefficient between the actual and the predicted values of the total imbalance that are shown in table 5.12 are obtained.

From the table it can be concluded that each of the variables that have been included in the set of the predictors when making the forecast, has an influence on the forecasted values and the accuracy of the forecast. The removal of the temperature forecast and the spot flow between the NO3-NO4 and the NO3-SE2 has the largest impact on the accuracy of both the Boosted Decision Tree regression model and the Decision Forest regression model and makes the value of the errors considerably.

Regression	Decision Forest regression			Boosted Decision Tree regression		
Estimator / Variable removed from the set	MAE	RMSE	Correlation coefficient	MAE	RMSE	Correlation Coefficient
-	183.58	226.86	0.81	186.32	230.93	0.81
Temperature forecast	191.52	236.56	0.81	190.04	234.20	0.81
Wind power production forecast	184.90	229.20	0.80	186.25	229.84	0.80
Power consumption forecast	186.00	228.58	0.80	185.36	228.01	0.80
Spot flow NO3-NO4 and NO3-SE2*	298.69	369.92	0.33	296.46	366.24	0.33
Daily, weekly and annual variation**	184.14	227.52	0.81	186.32	230.93	0.81

Table 5.12. Sensitivity of the results to the different predictors.

5.6.5 The influence of the past values of the total imbalance on the forecast's quality

In subsection 5.5.3 it was shown that the forecasting results for the balancing power volume have been improved by adding, into the set of the predictors, the past values of the balancing power volume. Only the past values that go no longer than 2-3 hours back in time in relation to the hour for which the forecast is made, will contribute to the most accurate forecast of the balancing power volume. Knowing this, it was decided to examine whether or not the adding of the past values of the total imbalance in the set of the predictors will improve the forecast's result for the total imbalance in the NO3.

The forecast for the total imbalance in the NO3 in an hour will be made for the following cases: when adding the values of the total imbalance from the previous hour, when adding the values of the total imbalance from one hour earlier than the previous hour, when adding the values of the total imbalance from two hours earlier than the previous hour and finally when adding the values of the total imbalance from three hours earlier than the previous hour. The model shown in figure 5.4 will be used in order to conduct the results.

The accuracy of the models has been evaluated by using the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The correlation coefficient between the forecasted and the actual values of the total imbalance has also been considered when estimating the model performance.

The results of the model evaluation are shown in tables 5.13 and 5.14. From the table one sees that the forecast of the total imbalance for an hour is more accurate when having the value of the total imbalance from the previous hour as one of the predictors. The accuracy of the forecast decreases as the past values of the total imbalance goes longer back in time in relation to the hour of the forecast.

In figures 5.12 and 5.13, two day-ahead and two-hour-ahead forecasts from the Boosted Decision Tree and the Decision forest regression of the total imbalance in the NO3 for a week are shown. To make the hour ahead forecast, the past value of the total imbalance hour prior to the hour of the forecast has been used as one of the predictors. The day-ahead forecasts are the forecasts from the previous subsection.

When using the past values of the total imbalance from the previous hour and one hour before the previous hour as one of the predictors, the accuracy of the forecast has been improved the most. In the first case the correct regulation state has been predicted in 92% of the hours for which the forecast has been made, and in the second case the correct regulation state has been predicted in 89% of the hours.

Decision Forest regression			
Estimator / Additional variable	MAE	RMSE	Correlation coefficient
-	183.58	226.86	0.81
Total imbalance in hour t-1	90.75	123.09	0.93
Total imbalance in hour t-2	127.28	163.31	0.89
Total imbalance in hour t-3	144.82	182.66	0.86
Total imbalance in hour t-4	155.42	193.79	0.85

Table 5.13. Sensitivity of the forecast error when having the balancing power volume from t-1 as a predictor.

Boosted Decision Tree regression			
Estimator / Additional variable	MAE	RMSE	Correlation coefficient
-	186.32	230.93	0.81
Total imbalance in hour t-1	102.19	135.02	0.93
Total imbalance in hour t-2	135.18	172.76	0.88
Total imbalance in hour t-3	148.13	186.34	0.86
Total imbalance in hour t-4	155.79	194.92	0.85

Table 5.14. Sensitivity of the forecast error when having the balancing power volume from t-2 as a predictor.

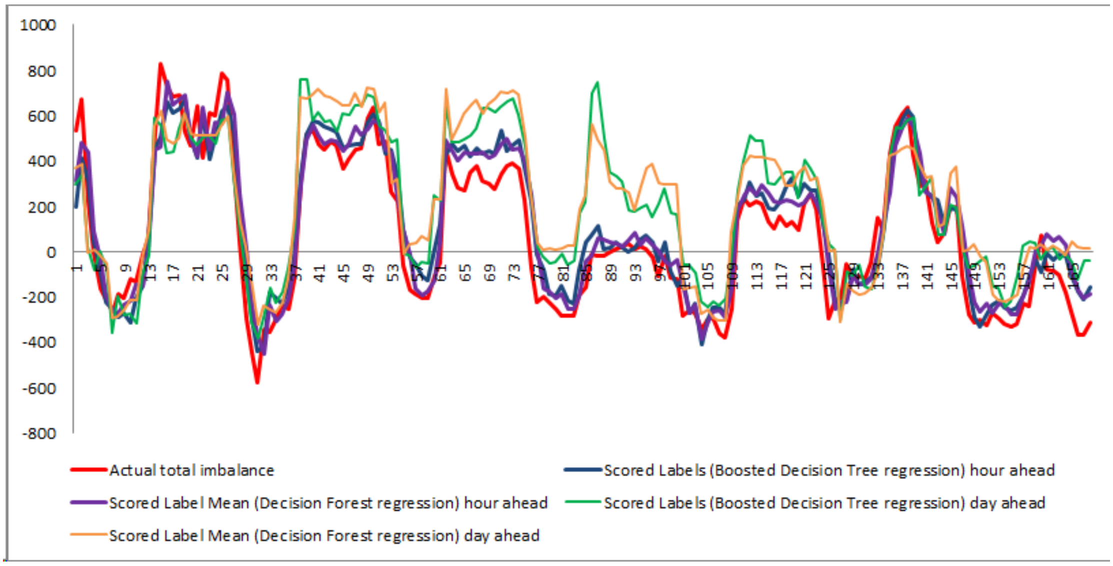


Figure 5.12. Forecast of the total imbalance for a week with the total imbalance from t-1 in the set of the predictors.

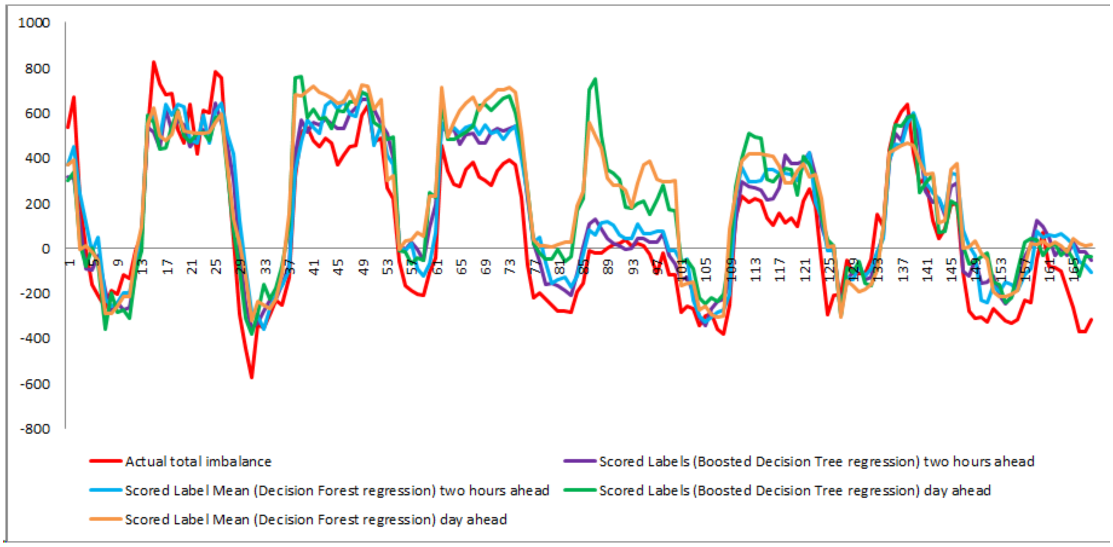


Figure 5.13. Forecast of the total imbalance for a week with the total imbalance from t-2 in the set of the predictors.

When using the past values of the total imbalance that lays a few hours prior to the hour, in which the forecast has been made for, in the set of the predictors, the accuracy of the forecast increases. So the time series with the past values of the total imbalance is a dominating variable among the other predictors when it comes to making a short-term forecast. This raises the question about the possibility to predict the total imbalance by using only the past values of the total imbalance as an explanatory variable and about the impact of the other predictors on the accuracy of the forecast. To examine this possibility, one by one of the variables has been removed from the set of the predictors and the models have been re-run for each of the cases. The Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the correlation coefficient between the forecasted and the actual values of the total imbalance that have been obtained are shown in the tables below:

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
9	-	90.75	123.09	0.93	102.19	135.02	0.93
8	Temperature forecast	90.43	123.83	0.93	97.34	130.33	0.93
7	Wind power production forecast	89.06	122.78	0.93	92.65	124.95	0.94
6	Power consumption forecast	86.49	120.34	0.94	93.02	127.81	0.93
4	Spot flow NO3-NO4 and NO3-SE2	69.75	106.38	0.95	70.36	106.77	0.95
1	Daily, weekly and annual variation	69.32	105.93	0.95	70.36	106.77	0.95

Table 5.15. Sensitivity of the forecast error when having the balancing power volume from t-1 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
9	-	127.28	163.31	0.89	135.18	172.76	0.88
8	Temperature forecast	130.77	168.05	0.88	136.01	172.97	0.88
7	Wind power production forecast	125.03	162.17	0.89	125.42	161.24	0.89
6	Power consumption forecast	131.55	171.90	0.87	134.08	175.45	0.87
4	Spot flow NO3-NO4 and NO3-SE2	115.74	164.29	0.87	116.60	165.15	0.87
1	Daily, weekly and annual variation	115.17	163.89	0.87	116.60	165.15	0.87

Table 5.16. Sensitivity of the forecast error when having the balancing power volume from t-2 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
9	-	144.82	182.66	0.86	148.13	186.34	0.86
8	Temperature forecast	146.18	184.78	0.86	149.37	187.47	0.87
7	Wind power production forecast	140.65	178.73	0.87	140.40	177.46	0.87
6	Power consumption forecast	158.46	202.96	0.83	158.43	201.01	0.83
4	Spot flow NO3-NO4 and NO3-SE2	152.75	206.77	0.79	153.39	207.04	0.79
1	Daily, weekly and annual variation	152.45	206.33	0.79	153.39	207.04	0.79

Table 5.17. Sensitivity of the forecast error when having the balancing power volume from t-3 as a predictor.

From the tables it can be concluded that an over-fitting of both the Decision Forest regression model and the Boosted Decision Tree regression model occurs, when having the past values of the total imbalance that go 1-2 hours back in time in relation to the hour of the forecast, along with other explanatory variables in the set of the predictors. The accuracy of the models output in this case is lower than, when having the past values of the total imbalance as an only explanatory variable. However, when having the past values of the balancing power volume that goes more than 2 hours back in time in relation to the hour of the forecast in the set of the predictors, the models will give a better output when having the time series with the past values in the set of the predictors among the other variables.

5.7 Forecasting the balancing power price

5.7.1 Constructing the experiment for the balancing power premium forecast

In order to forecast the balancing power premium for the next day an experiment that is similar to the one from subsection 5.6, has been constructed. The difference in the experiment structure of the experiment, which is shown in figure 5.14, consists in number and placement of the Split Data modules. In the experiment, two Split Data modules are used due to the structure of the time series of the balancing power volume that will be used in order to forecast the balancing power premium. The actual values of the balancing power volume in the time series in hours, for which the balancing power volume has been forecasted, are overwritten with predicted values. The dataset then is divided into the dataset that will be used for the training and into the dataset that will be used in order to generate the predicted values. The first dataset contains the actual values of the balancing power volume and the second one contains the forecasted values of the balancing power volume that have been obtained in the previous subsection 5.6.

In order to ensure that the actual values of the balancing power volume will not be used to generate the predicted values of the balancing power premium, the dataset has been divided by using a relative expression. From the hour 21501 (count of the hours starts from the 1st. of January 2013 at 1 a.m.) the actual values of the balancing power volume are overwritten by the forecasted values, so the dataset will be split at hour 21501. Still the data from the years of 2013 to 2014 will be used for the model training and data from the year of 2015 will be used for the model testing. Since belonging of the records is determined by the hour in which they are made, and the variable that identifies the hour will not be used in order to train the model and to generate predictions (the projection of this variable into the polar coordinates is used), the Split Data modules have been placed on the top of the experiment. This gives the possibility to divide the dataset by using the hour, and then to avoid of having it as one of the variable that will be used in order to train the model and to generate predictions. The relative expression in the first Split Data module (the left one) is Hour < 21501, and the relative expression in the second Split Data module is Hour ≥ 21501 (the right one).

Besides this, the structure of the experiment is the same as the experiment described earlier in subsection 5.6.

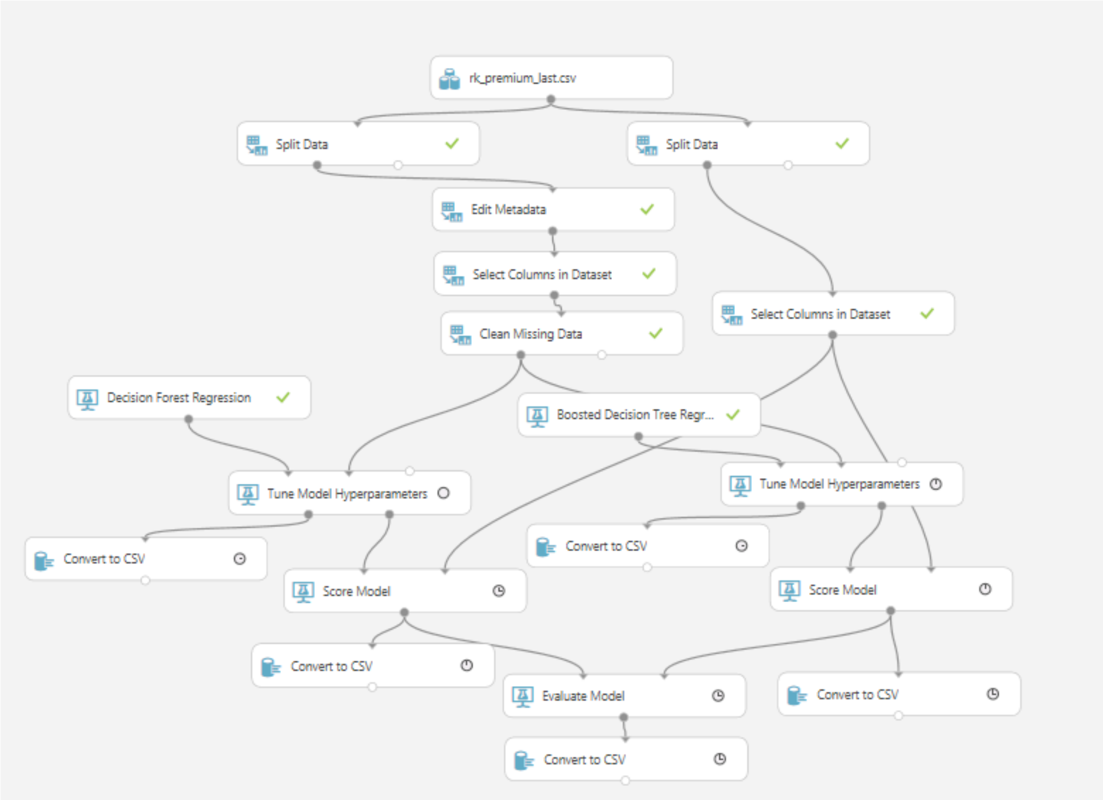


Figure 5.14. Structure of the experiment.

5.7.2. The choice of the form of forecasting the balancing power premium

From the structure of the experiment one sees that that the model will be trained by using the dataset containing the hours in which either an up-regulation, a down-regulation or a balance in the power system has taken place.

However, in the hours when it is balance in the power system, the balancing power premium is equal to zero. If knowing that it will be a balance in the power system before making the forecast of the balancing power premium, there will be no need in the forecast of the premium.

When knowing if an up-regulation or a down-regulation occurs in the system, one can run two separate models in order to predict the balancing power volume. This can give a more accurate forecast of the balancing power premium.
In order to find out the most suitable way for forecasting the balancing power premium, the accuracy of the forecast of the balancing power premium has been tested for the following cases:
Case 1: The model is trained by using the dataset that contains both the hours with an up-regulation, with a down-regulation and with a balance in the power system. The prediction is generated only for the hours in which a down-regulation has occurred.

Case 2: The model is trained by using the dataset that contains only hours with a down-regulation in the power system. The prediction is generated only for the hours in which a down-regulation has occurred.
Case 3: The model is trained by using the dataset that contains both hours with an up-regulation, with a down-regulation and with a balance in the power system. The prediction is generated only for the hours in which an up-regulation has occurred.
Case 4: The model is trained by using the dataset that contains only hours with a down-regulation in the power system. The prediction is generated only for the hours in which a down-regulation has occurred.
Case 5: The model is trained by using the dataset that contains both hours with an up-regulation, with a down-regulation and with a balance in the power system. The prediction is generated only for the hours in which an up-regulation or a down-regulation has occurred.
Case 6: The model is trained by using the dataset that contains both hours with an up-regulation and with a down-regulation in the power system. The prediction is generated only for the hours in which an up-regulation or a down-regulation has occurred.
Case 7: The model is trained by using the dataset that contains both the hours with an up-regulation, with a down-regulation and with a balance in the power system. The prediction is generated for the hours in which an up-regulation, a down-regulation or a balance has occurred in the power system.

The forecast has been made using as an input in the dataset containing the following variables: the hour of the day, the hour of the week, the hour of the year, the inflow in the NO3, the slope of the Elspot bid curve, the balancing power volume in the NO3 (actual values for both the training of the model and for the generating of the predictions) and the spot price in the NO3.
The results of this examination are shown in the table below:

	Decision Forest regression		**Boosted Decision Tree regression**	
	MAE	**RMSE**	**MAE**	**RMSE**
Case 1	2.85	3.81	2.89	3.93
Case 2	2.72	3.50	2.89	3.97
Case 3	7.86	11.78	6.46	13.56
Case 4	6.73	11.72	6.49	13.75
Case 5	4.43	7.33	4.02	8.28
Case 6	3.59	6.99	4.13	8.47
Case 7	2.01	4.51	1.61	5.11

Table 5.18. The results of the examination.

From the table it can be seen that the error of the balancing power premium is minimal when having the model that makes the forecast of the balancing power premium for an up-regulation, a down-regulation or no regulation. However, when having a model that forecasts the balancing power premium in the hours with either an up-regulation or a down-regulation, i.e. when knowing that it will be no regulation in the system, the balancing power premium is set to zero and no forecast is necessary, the error of the forecast increases. This indicates that a lower forecasting error in case 7 is caused by that the approximation line will go through the zero point, and that in the time series of the balancing power premium there are many hours in which the balancing power premium are equal to zero.

When having two separate models for forecasting the balancing power premium for an up-regulation and a down-regulation, the error of the forecast is lower in the case of a down-regulation and it is higher in the case of an up-regulation compared to cases 5 and 6. The reason why the comparison to cases 5 and 6 has been done is that when approximation line goes through the zero point the forecast error will be lower compared to all of the other cases. However, if the zero point is not included in the forecast (case 5 and 6), the forecast error will be higher. In order to decide whether it is more adventitious to have a model forecast the balancing power premium for any regulation state or two models that forecasts the balancing power premium for an up-regulation and a down-regulation separately, it is more intuitive to compare the forecast error cases 1-4 with the forecast error in cases 5-6. This issue is shown in figure 5.15.

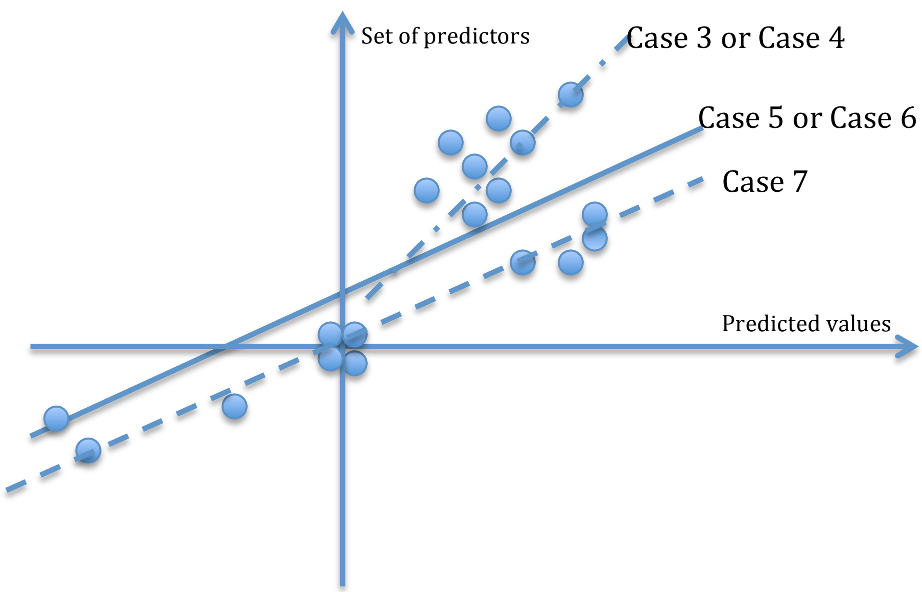


Figure 5.15. Linear approximation.

Taking into consideration the results represented in table 5.18, and that in case of having two separate forecasting models for the premium in the case of an up-regulation and in the case of a down-regulation, one is dependent on having a very accurate forecast of the regulation state, it has been decided to have one model that makes a forecast of the balancing power volume in the hours in which either an up-regulation, a down-regulation or no regulation has occurred.

5.7.3 Forecasting the balancing power premium using the balancing power volume as one of the predictors

One of the predictors that are supposed to be used in order to forecast the balancing power premium is the balancing power volume. This requires the availability of an accurate forecast of the balancing power volume. In subsection 5.6 it was concluded that the most accurate forecast of the balancing power volume in an hour can be conducted when having the values of the balancing power volume from the previous hour. In this case, only a short-term forecast (1 hour ahead) of the balancing power premium can be made.

The balancing power premium has been forecasted by using the set of the predictors which contains the following attributes: the hour of the day, the hour of the week, the hour of the year, the inflow in the NO3, the slope of the Elspot bid curve, the balancing power volume in the NO3 (actual values for the training of the model and the forecasted values for the generating of the predictions), the spot price in the NO3.

Under the training of both the Boosted Decision Tree regression and the Decision Forest regression model, 30 iterations with parameter settings switching have been done in order to find an optimum parameter set. All parameter settings that have been considered under training can be found in Appendix I, and the best three parameter sets and the corresponding Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) are shown in the table below:

Boosted Decision Tree regression						
Parameter set	**Number of leaves**	**Minimum leaf instances**	**Learning rate**	**Number of trees**	**MAE**	**RMSE**
1	17	10	0.211194	299	1.708919	6.799584
2	23	10	0.253417	140	1.687187	6.820208
3	4	20	0.331017	117	1.769563	6.93115

Table 5.19. Parameter settings for the Boosted Decision Tree regression model.

Decision Forest regression						
Parameter set	**Minimum number of samples per leaf node**	**Number of random splits per node**	**Maximum depth of the trees**	**Number of trees**	**MAE**	**RMSE**
1	7	998	56	31	2.213226	7.529644
2	5	722	61	31	2.271652	7.548914
3	10	850	44	18	2.291713	7.617313

Table 5.20. Parameter settings for the Decision Forest regression model.

The results of the evaluation of the models are shown in table 5.21. The graph of the forecasted and the actual values of the balancing power volume for a week is shown in figures 5.16.

From the table it can be concluded that the Boosted Decision tree regression gives a more accurate forecast compared to the Decision Forest regression model. However, the overall performance of both models is rather poor. The models struggle to predict the correct sign and the size of the premium. The balancing power premium has been forecasted in 5367 hours. The correct sign of the balancing power premium (a small deviation of +-1 Euro, from zero has been allowed) has been predicted in 58% of the cases by the Boosted Decision Tree regression and in 20% of the cases by the Decision Forest regression.

In most of the cases the prediction of an incorrect sign of the premium is caused by the failed prediction of the regulation state when forecasting the power volume. The accuracy of the forecast of the balancing power volume has also an impact on the accuracy of the prediction of the balancing power premium size. So a forecast of the balancing power volume that has a low accuracy will cause a low accuracy when forecasting the balancing power premium.

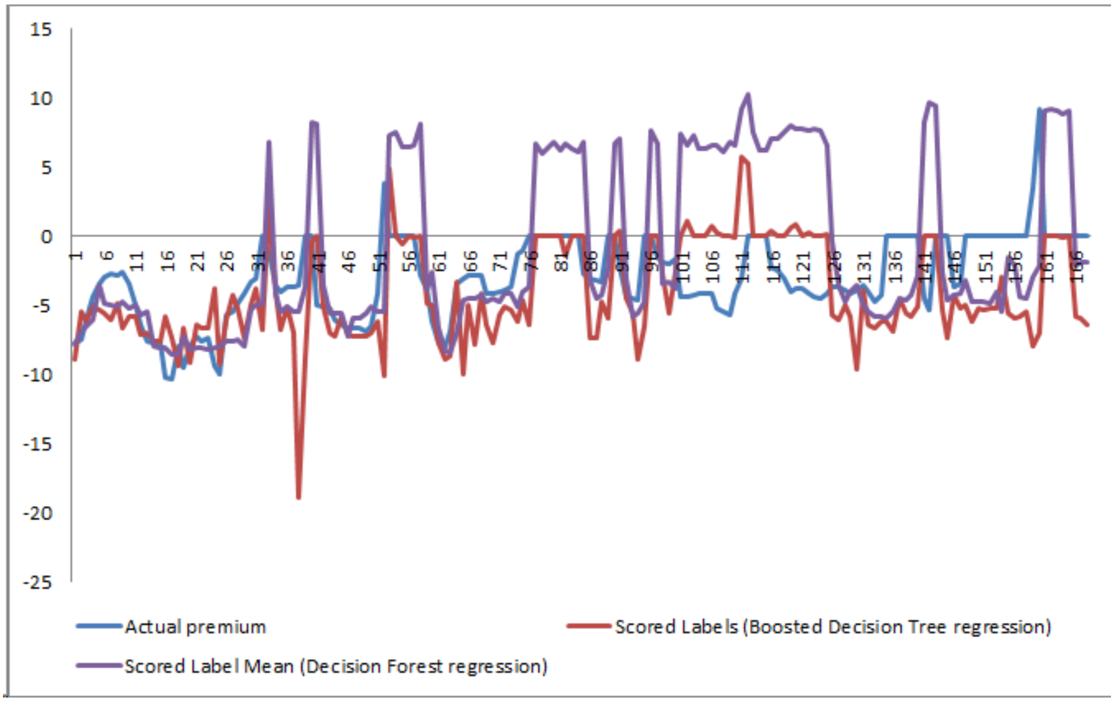


Figure 5.16. Forecast of the balancing power premium for a week.

Estimator / Model	MAE	RMSE	Correlation Coefficient
Boosted Decision Tree regression	4.84	8.15	0.06
Decision Tree Forest regression	9.36	11.16	0.05

Table 5.21. Forecast error.

To examine how the availability of a very accurate forecast of the balancing power volume will affect the accuracy of the forecast of the balancing power premium it has been assumed that a perfect forecast of the balancing power volume is available when making the forecast of the balancing power premium. The actual values of the balancing power volume will be used in order to predict the premium. The results of the accuracy evaluation of the forecasted values are shown in table 5.22, and figure 5.17 shows the actual and the forecasted values of the balancing power premium for the same week as in figure 5.16.

Estimator Model	MAE	RMSE	Correlation Coefficient
Boosted Decision Tree regression	1.61	5.11	0.66
Decision Tree Forest regression	2.01	4.51	0.70

Table 5.22. Forecast error when using the actual balancing power volume.

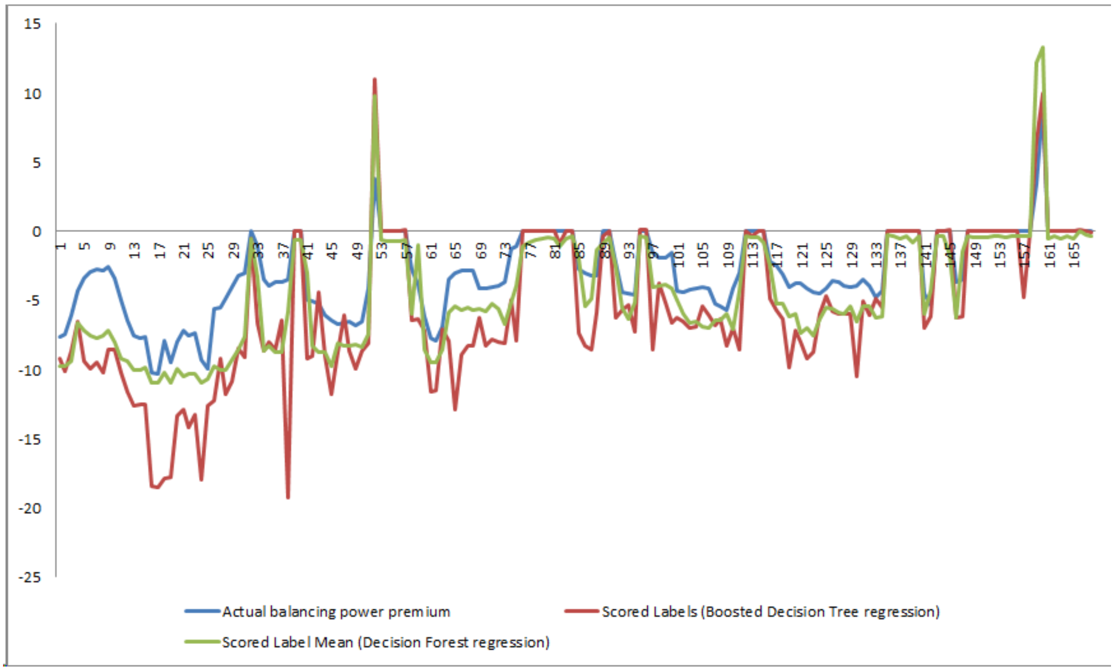


Figure 5.17. Forecast of the balancing power premium for a week when using the actual balancing power volume.

From the table and the figure it can be concluded that when having an "ideal" forecast of the balancing power volume, an accurate enough forecast of the balancing power premium can be obtained. In this case both of the models manage to predict the correct sign of the premium almost in all hours (95%) that the forecast have been made for, to indicate the hours when the premium is zero (a small deviation of +-1 Euro, from zero has been allowed), and to indicate an increase and a decreases in the value of the balancing power premium. However, the Boosted Decision Tree regression model makes a more accurate forecast compared to the Decision Forest regression model.

The accuracy of the forecast of the balancing power premium depends a lot on the quality of the time series of the balancing power volume. However, there are several variables in the set of the predictors that has been used in order to forecast the balancing power premium and that influence the accuracy of the forecast as well. To investigate the impact of each of the predictors on the forecast result, the sensitivity of the forecast accuracy to the removal of the different variables from the set of the predictors have been tested.

One variable at a time will be removed from the set of the predictors, and the remaining variables in the set will be used to forecast the balancing power premium. By changing place with the first variable, one by one of the variables will be taken out of the set, and consequently the variable that was already taken out, will be taken back into the set. Each time a variable changes place, a forecast of the balancing power premium is done. The results of this examination can be found in the table below:

Regression	**Decision Forest regression**			**Boosted Decision Tree regression**		
Estimator / Variable removed from the set	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
-	2.01	4.51	0.70	1.61	5.11	0.66
Balancing power volume in NO3	3.02	6.19	0.02	3.47	6.47	0.02
Spot price in the NO3	1.84	4.51	0.72	1.85	5.38	0.69
Slope of the Elspot bid curve	1.91	4.44	0.70	1.82	5.51	0.69
Inflow in the NO3	1.98	4.53	0.70	1.76	4.83	0.70
Daily, weekly and annual variation	1.98	4.49	0.70	1.82	5.34	0.69

Table 5.23. Sensitivity of the result when using the actual balancing power volume.

As it can be seen from the table, the accuracy of the forecast made by both the Boosted Decision Tree regression model and by the Decision Forest regression model reduces the most when removing the balancing power volume from the set of the predictors. When removing other variables from the set of the predictors, the accuracy of the forecast made by the Boosted Decision Tree regression model reduces, but the accuracy of the forecast from the Decision Forest regression model increases. However, since the Boosted Decision Tree regression model gives a better forecast of the balancing power premium, the sensitivity of this model to the removal of the different variables from the set of the predictors will be considered. So it can be concluded that in order to get the most accurate forecast of the balancing power premium, it is important to take into consideration both the balancing power volume in the NO3, the spot price in the NO3, the slope of the Elspot bid curve, the inflow in the NO3, the hour of the day, the hour of the week and the hour of the year.

5.7.4 Forecasting the balancing power premium without the balancing power volume in the set of the predictors

From the previous subsection it can be concluded that accuracy of the forecast of the balancing power premium depends a lot on the quality of the time series of the balancing power volume. At the moment of making the forecast of the balancing power premium in an hour, the actual balancing power volume in this hour is not known, so one is depended on having an accurate forecast of the balancing power volume. The most accurate forecast of the balancing power volume that has been conducted, is not good enough in order to use it to forecast the balancing power premium. That is why it has been decided to investigate the possibility of forecasting the balancing power premium without having the time series of the balancing power volume as one of the predictors.

When examining the sensitivity of the forecast accuracy to the removal of the different variables from the set of predictors in the previous subsection, it has been found out that it is impossible to get an accurate forecast of the balancing power premium using the set of the predictors that contains the spot price in the NO3, the slope of the Elspot bid curve, the inflow in the NO3, the hour of the day, the hour of the week and the hour of the year. So it is necessary to have additional explanatory variables in order to get a more accurate forecast.

The balancing power premium in an hour has shown to have a correlation with the past values of the balancing power premium. It is, most likely, that in the similarity to the balancing power volume, the premium in an hour will be better explained by the past values of the balancing power premium from the previous hour and from one hour before the previous hour. The reason for this is that the correlation between the balancing power premium in an hour and the past values of the premium decreases as the past values that is used under the correlation goes further back in time, just like in the case of the balancing power volume. So the past values of the balancing power premium that will be used in order to forecast the balancing power premium in an hour goes no longer than four hours further back in time in relation to the hour in which the premium is predicted.

The balancing power premium in an hour has been forecasted by using the model shown in figure 5.14 and the set of the predictors containing the following variables: the past values of the balancing power premium, the spot price in the NO3, the slope of the Elspot bid curve, the inflow in the NO3, the hour of the day, the hour of the week and the hour of the year.

The results are shown in the table below:

Decision Forest regression			
Estimator / Additional variable	MAE	RMSE	Correlation coefficient
Balancing power premium in hour t-1	2.01	4.53	0.67
Balancing power premium in hour t-2	2.57	5.42	0.45
Balancing power premium in hour t-3	2.80	5.81	0.32
Balancing power premium in hour t-4	2.83	5.89	0.27

Table 5.24. Forecast error for the Decision Forest regression model.

Boosted Decision Tree regression			
Estimator / Additional variable	MAE	RMSE	Correlation coefficient
Balancing power premium in hour t-1	1.67	4.50	0.68
Balancing power premium in hour t-2	2.42	5.45	0.45
Balancing power premium in hour t-3	2.70	5.57	0.40
Balancing power premium in hour t-4	2.83	5.89	0.27

Table 5.25. Forecast error for the Boosted Decision Tree regression model.

From the tables it can be seen that the forecast of the balancing power premium in an hour is most accurate when having the past values of the balancing power premium from the previous hour as one of the predictors. And the accuracy of the forecast decreases, when the past values of the balancing power premium that have been used in the set of the predictors, goes further back in time. Both the Boosted Decision Tree regression model and the Decision Forest regression model gives almost equally accurate forecast of the balancing power premium, but it can be seen that the forecast made by the Boosted Decision Tree regression model has a slightly lower forecast error.

When comparing the accuracy of the forecast that is made using the balancing power volume (actual values) as an explanatory variable, with the accuracy of the forecast that is made using the past values of the balancing power premium as one of the predictors, it can be seen that the forecast error in the first case (MAE=1.61, RMSE=5.11) and the lowest forecast error in the second case (MAE=1.67, RMSE=4.5) are almost equal to each other. This holds when the balancing power premium from the previous hour in relation to the hour the forecast is made for is one of the explanatory variables. When the balancing power premium in an hour has been forecasted by using the past values of the balancing power premium that goes more than one hour back in time in relation to the hour the forecast is made for, the forecast error increases. The forecast error in this case will be higher than the forecast error when having the balancing power volume (actual values) in the set of the predictors.

Figures 5.18 and 5.19 show the balancing power premium forecast for a week when having the past values of the balancing power premium from the previous hour and from one hour before the previous hour in relation to the hour, which the forecast has been made for, respectively.

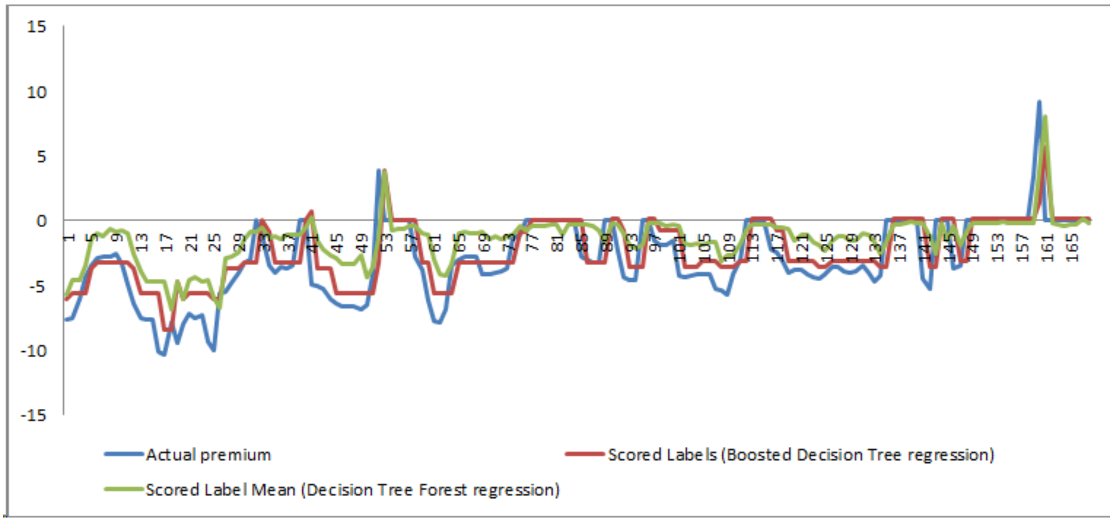


Figure 5.18. Forecast of the balancing power premium for a week with the balancing power premium from t-1 in the set of the predictors.

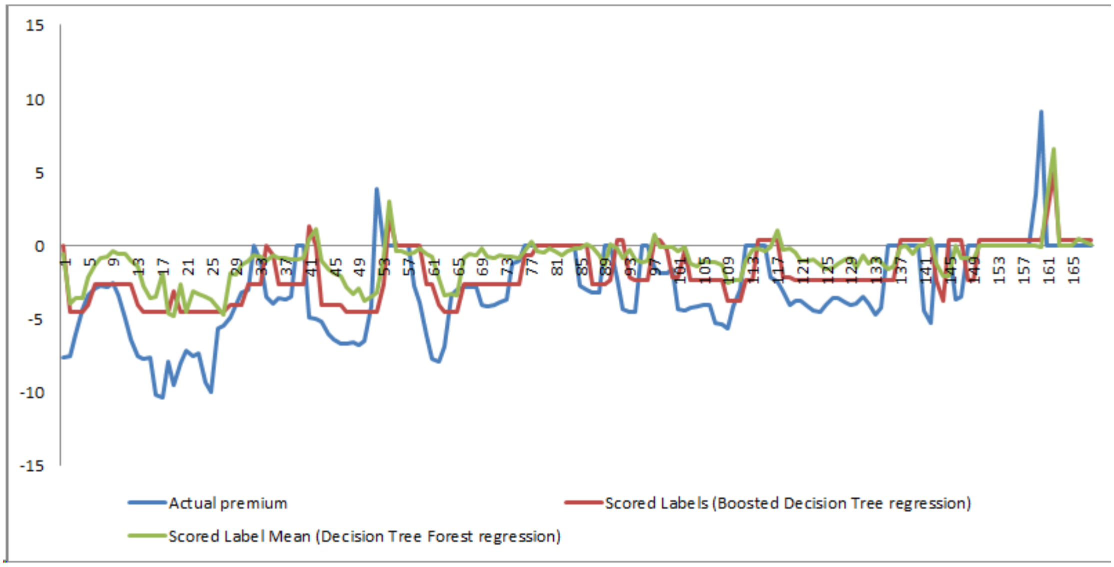


Figure 5.19. Forecast of the balancing power premium for a week with the balancing power premium from t-2 in the set of the predictors.

From the figures it can be seen that both the Boosted Decision Tree regression model and the Decision Forest regression model manage to indicate the correct sign and the magnitude of the balancing power premium. However, the predicted values have a tendency to lag the actual values of the balancing power premium with one and two hours when having the past values of the balancing power premium that goes one and two hours further back in time in relation to the hour the forecast is made for, respectively. When having the past values of the balancing power premium from three and four hours further back in time in relation to the hour the forecast has been made for, the forecasted values lag the actual ones with three and four hours respectively.

The balancing power premium has been predicted in 5357 hours. When forecasting the balancing power premium in an hour using the past values of the balancing power premium from the previous hour, the correct sign of the regulation premium (a small deviation of ±1 Euro has been allowed) has been predicted in around 86% of the cases. The number of hours with the correct predicted sign of the premium decreases as the past values of the balancing power premium that are used in order to predict the premium in an hour, go further back in time. So when using the past values of the premium that lays two hours back in time in relation to the hour the forecast has bee made for, the correct regulation state has been predicted in 78% of the cases. When using the past values of the premium that lays three hours back in time and four hours back in time in relation to the hour the forecast has been made for, the correct regulation state has been predicted in 74% and 52 % of the cases, respectively.

When using the past values of the balancing power premium that lays a few hours prior to the hour, which the forecast has been made for, in the set of the predictors, the accuracy of the forecast is higher compared to the case when the balancing power premium have been predicted using the set of the predictors that does not contain either of the past values of the premium or the values of the balancing power volume. So the time series with the past values of the balancing power premium is, most likely, a dominating variable among the other predictors when it comes to making a forecast using the set of the predictors in which the values of the balancing power volume are included. This raises a question about the possibility to predict the balancing power premium by using only the past values of the balancing power premium as an explanatory variable and about the impact of the other predictors on the accuracy of the forecast. To examine this possibility, the same procedure that has been used in order to test the possibility of forecasting the balancing power volume (subsection 5.6.2) and the total imbalance (subsection 5.6.4) by using only its past values has been done. One by one of the variables has been removed from the set of predictors and the models have been re-run for each of the cases. The Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the correlation coefficient between the forecasted and the actual values of the balancing power premium that have been obtained are shown in the tables below:

Regression		**Decision Forest regression**			**Boosted Decision Tree regression**		
Number of variables in the set of the predictors	**Estimator** / **Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
7	-	2.01	4.53	0.67	1.67	4.50	0.68
6	Inflow	2.09	4.56	0.66	1.68	4.50	0.68
5	Spot price	1.80	4.49	0.68	1.68	4.50	0.68
4	Slope of the bid curve	1.71	4.58	0.66	1.65	4.48	0.68
1	Daily, weekly and annual variation	1.73	4.60	0.66	1.65	4.48	0.68

Table 5.26. Sensitivity of the forecast error when having the balancing power premium from t-1 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of the predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
7	-	2.57	5.42	0.45	2.42	5.45	0.45
6	Inflow	2.49	5.39	0.46	2.45	5.46	0.45
5	Spot price	2.51	5.42	0.46	2.45	5.46	0.45
4	Slope of the bid curve	2.50	5.61	0.42	2.44	5.45	0.45
1	Daily, weekly and annual variation	2.51	5.60	0.42	2.44	5.45	0.45

Table 5.27. Sensitivity of the forecast error when having the balancing power premium from t-2 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of the predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation coefficient**	**MAE**	**RMSE**	**Correlation coefficient**
7	-	2.80	5.81	0.32	2.70	5.57	0.40
6	Inflow	2.68	5.62	0.38	2.91	5.80	0.35
5	Spot price	2.82	5.65	0.38	2.91	5.80	0.35
4	Slope of the bid curve	2.88	6.07	0.28	2.81	5.74	0.36
1	Daily, weekly and annual variation	2.89	6.08	0.28	2.81	5.74	0.36

Table 5.28. Sensitivity of the forecast error when having the balancing power premium from t-3 as a predictor.

Regression		Decision Forest regression			Boosted Decision Tree regression		
Number of variables in the set of the predictors	**Estimator / Variable removed from the set**	**MAE**	**RMSE**	**Correlation Coefficient**	**MAE**	**RMSE**	**Correlation Coefficient**
7	-	2.83	5.89	0.27	2.83	5.89	0.27
6	Inflow	2.81	5.78	0.31	3.07	5.96	0.28
5	Spot price	2.98	5.82	0.31	3.07	5.96	0.28
4	Slope of the bid curve	3.04	6.18	0.22	2.98	5.91	0.29
1	Daily, weekly and annual variation	3.05	6.19	0.22	2.98	5.91	0.29

Table 5.29. Sensitivity of the forecast error when having the balancing power premium from t-4 as a predictor.

From the tables it can be concluded that when having the past values of the balancing power premium that go no longer than two hours further back in time in relation to the hour, which the forecast is made for, the accuracy of the forecast from both the Boosted Decision Tree regression model and the Decision Forest regression model will approximately be same independent of if the past values of the premium is the only variable in the set of the predictors or not. This indicates that besides the past values of the balancing power premium, other variables in the set of the predictors will not have any effect on the results of the forecast.

In case of using the past values of the balancing power premium that goes more than 2 hours further back in time in relation to the hour, which the forecast has been made for, in the set of the predictors, the results of the examination are different. The accuracy of the forecast is lower when having only the past values of the balancing power premium than when having more variables in the set of the predictors. So as further back in time the past values of the balancing power premium goes in relation to the hour, which the forecast is made for, the less they will explain the forecasted value of the balancing power premium.

Chapter 6

Discussion

In this master thesis the possibility of forecasting the balancing power volume and the balancing power premium using a regression model and the set of the predictors has been examined. It has been chosen to use the regression model based on the Machine Learning algorithms, and more precisely the model provided by the Microsoft Azure Machine Learning Studio. Among many different models that can be found in the Microsoft Azure Machine Learning Studio, the Boosted Decision Tree regression model and the Decision Forest regression model have been chosen in order to forecast the balancing power volume and the balancing power premium.

In chapter 4 a statistical analysis of the time series of the balancing power volume and the time series of the balancing power premium has been presented in order to possibly identify the variables that can be used as input factors for the regression model. Not all of the variables that the balancing power volume and the balancing power premium have shown to depend on have been used as input variables for the model. The reason for this is the unavailability of the values for some of the variables at the moment of making the forecast. For example, the actual power consumption and the actual wind power production in an operation hour that have an influence on the balancing power volume, are not known before the operational hour ends. The time series for both the balancing power volume and the balancing power premium has also been claimed to have a daily, a weekly and an annual variation that has been used as one of the explanatory variables.

In chapter 5 the balancing power volume and the balancing power premium in the price area NO3 have been forecasted. In subsection 5.6.2, the possibility of forecasting the balancing power volume by using the variable that are known at the moment of the opening of the Balancing power market for the submission of bids, as an input into the regression model. The forecast obtained in this case has a lower accuracy; so one could have concluded that it is impossible to predict the balancing power volume day ahead. However, when making the forecast, only the factors influencing the balance in the power system in price area NO3 have been taken into consideration. The balancing power volume in one price area can be influence by the events that take place as in this price area, as well as in the other price areas since the TSOs in Norway, Sweden, Denmark and Finland exchange their balancing power reserves between each other. By taken into consideration the situation in the other price areas, the accuracy of the forecast can be significantly improved.

It is possible to obtain an accurate enough forecast of the balancing power volume in an hour by using the past values of the balancing power volume from the previous hours as an explanatory variable. A forecast with a good accuracy has been conducted when using the values of the balancing power volume that lays one or two hours further back in time in relation to the hour, in which the forecast is made for. The forecasted values of the balancing power volume in an hour has the best accuracy when having the values of the balancing power volume from the previous hour in the set of the predictors. However, due to the market structure, in reality the value of the balancing power volume from the previous hour is not known when the market actors have their last chance to adjust their bids for the balancing power in the operating hour. The market actors can adjust their bids at latest 45 minutes before the operating hour, and the balancing power volume that has been activated in the previous hour in relation to the operating hour at this moment is unknown. So the last value of the balancing power volume that is known 45 minutes before the operating hour lies two hours further back in time in relation to the operating hour. In this case the obtained forecast is less accurate compared to when using the value of the balancing power volume that lies one hour forward in time, but still the forecast in this case has a better accuracy than in the case when using the past values of the balancing power volume in the set of the predictors. In this case the balancing power volume can be forecasted two hours ahead.

It is worth to point out that as further back in time the past values of the balancing power volume that has been used in the forecast goes, the lower the accuracy of the forecast will be. The past values of the volume that lies further than two hours back in time in relation to the hour, which the forecast has been made for, are not sufficient in order to explain the values of the balancing power volume that have been forecasted.

The influence of the situation in other price areas on the balancing power volume in the NO3 can be taken into account in the model by considering the flow of the deviation between the actual and the spot power flow. If the balancing power reserves in the NO3 have been activated due to imbalances in other price areas, or the imbalance that has occurred in the NO3, has been regulated by reserves activation outside the NO3, the flow between the NO3 and the neighbour areas will be adjusted and a deviation between the actual and the spot flow will occur. In subsection 5.6.4, the deviation between the actual and the spot power flow for the NO3 has been taken into consideration by adding it to the balancing power volume, and then making the forecast for the total imbalance in the NO3 a day ahead. The forecast results that have been obtained, have a high accuracy. The correct direction of the total imbalance has been predicted in 87% of the hours, in which the forecast has been made. The model has also managed to indicate the peaks in the magnitude of the total imbalance. The forecast of the total imbalance can also be improved, as more information will be available towards the operating hour.

The values of the total imbalance do not give per se information about the balancing power volume in the NO3. However, if one manages to predict the total imbalance in all of the 12 price areas with an equally good accuracy, the obtained forecast can be used to calculate the value of the balancing power volume that has been activated. Some of the imbalances will be regulated by the other imbalances with an opposite sign. In order to regulate the remaining imbalance, the balancing power reserves will be activated. In this case the value of the balancing power volume for Norway, Sweden, Denmark and Finland will be obtained. Exactly in which price area the balancing power volume will be activated will depend on how the market actors can and are willing to offer their production capacity into the Balancing power market. The activity of the market actors in the Balancing power market can change over time. In order to identify in which price area the regulation will take place, the power flow between the price areas, in principle, can be modelled as long as the total imbalance in all of the price areas is known.

However, the total imbalance in the NO3 can directly be used by the market actors in order to reduce their imbalance costs. Knowing the total imbalance in the area, market actors can evaluate whether or not their imbalances have the same direction with the total imbalance, and then they can make a decision about if the imbalance should be traded in the Elbas market or if it can be left for the regulation in the Balancing power market. Having their imbalances in the same direction with the total imbalance in the price area is essential for the market actors in order to avoid to be penalized for having their imbalance in the opposite direction of the system imbalance. So knowing the value and the direction of the total imbalance, the market actors can reduce their imbalance costs and probably, insure some income by trading their imbalances in the Elbas market.

When forecasting the balancing power premium, it has been evident that the accuracy of the forecast of the balancing power premium depends on the quality of the time series of the balancing power volume that is used as an input into the model. In other words, an accurate forecast of the balancing power volume is required in order to obtain an accurate forecast of the balancing power premium. The most accurate forecast of the balancing power volume that has been conducted in this thesis is not sufficient enough to provide an accurate forecast of the balancing power volume. However, as it is mentioned above, when having a forecast of the balancing power volume that has a high accuracy, for all of the price areas, an accurate forecast of the balancing power volume per price area can be obtained.

An accurate enough forecast of the balancing power premium can be obtained without using the balancing power volume as one of the explanatory variables. In this case the past values of the balancing power premium that lay more than one hour further back in time in relation to the hour, in which the forecast is made for, have been used in the set of the predictors. However, the best accuracy of the forecast that has been obtained in this case is still lower compared to the accuracy of the forecast of the balancing power premium when the balancing power volume has been used as an explanatory variable. The forecast of the balancing power premium can also not be obtained more than for two hours ahead.

Chapter 7

Conclusion

The aim of this thesis was to evaluate the possibility of forecasting the balancing power volume and the balancing power premium in price area NO3 by using the regression model based on the Machine Learning algorithms and a set of the predictors. From the results obtained in this work, the following conclusions can be made:

- It is impossible to obtain an accurate forecast of the balancing power volume day ahead in the price area without taking into consideration the events that occur outside of the area.
- An accurate enough two-hours-ahead forecast of the balancing power volume in an hour can be obtained when using the values of the balancing power volume that have occurred to hours earlier.
- It is possible to obtain a day-ahead forecast of the total imbalance in the price area with a high accuracy.
- In order to get an accurate forecast of the balancing power premium, a very accurate forecast of the balancing power volume is required.
- The balancing power premium in an hour can be predicted two hours ahead with a good enough accuracy when using the values of the balancing power premium that have occurred two hours earlier.

Chapter 8

Further work

Despite of the fact that a very accurate forecast of the balancing power volume and of the balancing power premium has not been conducted, a very promising results that can be a basis for further work have been obtained.

An accurate forecast of the total imbalance in price area NO3 has been obtained. If the corresponding forecast of the total imbalance will be conducted in the other price areas, it will be, most likely, possible to obtain the balancing power volume. The accuracy of the values of the balancing power volume in this case will depend on the accuracy of the forecast of the total imbalance. When having an accurate forecast of the balancing power volume, it can be used in order to get a robust forecast of the balancing power premium.

So the model needs to be extended for all of the price areas in Norway, Sweden, Denmark and Finland. The accuracy of the forecast of the balancing power volume and the balancing power premium can be possibly improved by taken into consideration the other factors that can have an impact on them. An example on such factors can be the power production outages, sun radiation etc.

Bibliography

BARGA, R., FONTAMA, V. & TOK, W. H. (2014) *Predictive Analytics with Microsoft Azure Machine Learning: Build and Deploy Actionable Solutions in Minutes.* Apress.

BOOMSMA, T. K., JUUL, N., & FLETEN, S.-E. (2014) Bidding in sequential electricity markets: The Nordic case. *European Journal of Operational Research,* 238(3):797-809.

BOWERMAN, B. L. & O'CONNELL, R. T. (1993) *Forecasting and time series: An applied approach.* 3th Ed. Belmont: Duxubry Press.

BREIMAN, L. et al. (1984) *Classification and Regression Trees.* New York: Chapman & Hall.

BUSSETI, E., OSBAND, I. & WONG, S. (2012) *Deep Learning for Time Series Modeling.* CS 229 Final Project Report.

BYE, T., et al (2010) *Flere og riktigere priser – Et mer effektiv kraftsystem.* Oslo

CHATFIELD, C. (2001) *Time-series forecasting.* Chapman & Hall/CRC.

CONJEO, A. J., CONTRERAS, J., ESPINOLA, R. & PLAZAS, M. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting,* 21:435-462.

CRIMINISI, A. & SHOTTON, J. (2013) *Decision Forests in Computer Vision and Medical Image Analysis.* Springer.

EUROPEAN COMISSION. (2016) *Europe 2020 in your country*. [Online] Available at: http://ec.europa.eu/europe2020/europe-2020-in-your-country/index_en.htm [Accessed: 10th May 2016]

JAEHNERT, S., FARAHMAND, H., & DOORMAN, G. L. (2009) *Modelling of prices using the volume in the Norwegian regulating power market.* [Online] Available from: https://www.sintef.no/globalassets/project/balance-management/paper/modelling-of-prices-using-the-volume-un-the-norwegian-regulation-power-market_-jaehnert_2009.pdf [Accessed: 3th April 2016]

KLÆBOE, G. (2015) *Stochastic Short-term Bidding Optimisation for Hydro Power Producers.* NTNU.

KLÆBOE, G., LUND ERIKSRUD, A. & FLETEN, S., (2015) *Benchmarking time series based forecasting models for electricity balansing market prices*. [Online] Available from: http://www.gwu.edu/~forcpgm/2013-006.pdf [Accessed: 3th April 2016]

KRISTIANSENT, T. (2004) *Risk Management in Electricity Markets Emphasizing Transmission Congestion.* NTNU.

MICROSOFT AZURE (2015) *Machine Learning/Initialize Model.* [Online] Available from: https://msdn.microsoft.com/en-us/library/azure/dn905812.aspx [Accessed: 7th April 2016]

MICROSOFT AZURE A (2016) *Machine learning algorithm cheat sheet for Microsoft Azure Machine Learning Studio.*[Online] Available from: https://azure.microsoft.com/nb-no/documentation/articles/machine-learning-algorithm-cheat-sheet/ [Accessed: 7th April 2016]

MICROSOFT AZURE B (2016) *Boosted Decision Tree Regression.* [Online] Available from: https://msdn.microsoft.com/en-us/library/azure/dn905801.aspx [Accessed: 7th April 2016]

MICROSOFT AZURE C (2016) *Decision Forest Regression.* [Online] Available from: https://msdn.microsoft.com/en-us/library/azure/dn905862.aspx [Accessed: 7th April 2016]

MICROSOFT AZURE D (2016) *Clean Missing Data.* [Online] Available from: https://msdn.microsoft.com/en-us/library/azure/dn906028.aspx [Accessed: 7th April 2016]

MICROSOFT AZURE E (2016) *Tune Model Hyperparameters.* [Online] Available from: https://msdn.microsoft.com/en-us/library/azure/dn905810.aspx [Accessed: 7th April 2016]

MURPHY, K. P. (2012) *Machine Lerning. A Propabilistic Perspective.* London: The MIT Press.

NVE. (2014) *Driften av kraftsystemet 2013.* [Online] Available from: http://publikasjoner.nve.no/rapport/2014/rapport2014_38.pdf [Accessed: 5th April 2016]

OLJE- OG ENERGIDEPARTEMENTET. (2014) *Fakta 2015. Energi- og vannressurser i Norge.* [Online] Available from: https://www.regjeringen.no/contentassets/fd89d9e2c39a4ac2b9c9a95bf156089a/1108774830_897155_fakta_energi-vannressurser_2015_nett.pdf [Accessed: 5th April 2016]

OLSSON, M. & SODER, L. (2008) Modelling real-time balancing power market price using combined SARIMA and Markov processes. IEEE *Transactions on Power Systems,* 23(2):443-450.

SHALEV-SHWARTZ, S. & BEN-DAVID, S. (2014) *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press.

SONQUIST, J. A. & MORGAN, J. N. (1964) *The detection of interaction effects.* Ann Arbor: Institute of Social Research, University of Michigan.

STATNETT. (2013) *Om regulerkraftopsjoner (RKOM).* [Online] Available from: http://statnett.no/Drift-og-marked/Markedsinformasjon/RKOM1/Om-RKOM/ [Accessed: 1th January 2016]

STATNETT. (2014) *Systemdrifts- and markedsutviklingsplan 2014-20.* [Online] Available at: http://www.statnett.no/Global/Dokumenter/Kraftsystemet/Systemansvar/SMUP%20oppdatert%20060514.pdf [Accessed: 15th March 2016]

STATNETT. (2015) *Vilkår for anmelding, håndtering av bud og prissetting i regulerkraftmarkedet (RKM).* [Online] Available from: http://statnett.no/Drift-og-marked/Markedsinformasjon/RKOM1/Om-regulerkraftmarkedet-RKM/ [Accessed: 5th April 2016]

TALLURI, K. T. & VAN RYZIN, G. J. (2005) *The theory and practice of revenue management.* USA: Springer.

WALPOLE, R. E. et al. (2012) *Probability & Statistics for engineers and scientists.* 9th. Ed. Boston: Pearson

WANGENSTEEN, I. (2012) *Power System Economics-the Nordic Electricity Market.* 2th Ed. Trondheim: Tapir Academic Press.

Appendix A

The duration of the regulation state and the possibility of that a regulation state in an hour will remain the same of will change with another regulation state in the next hour or in the future hours has been examined. The examination has been done for the period from 2013 to 2015. Three regulation states are possible: a down-regulation, an up-regulation or balance. These three regulation states can change each other or remain the same during the time. Following is possible:

- If there is a down-regulation in an hour, it will be a down-regulation in the next hour.
- If there is a down -regulation in an hour, it will be an up-regulation in the next hour.
- If there is a down-regulation in an hour, it will be balance in the system in the next hour.
- If there is an up-regulation in an hour, it will be an up-regulation in the next hour.
- If there is an up -regulation in an hour, it will be a down-regulation in the next hour.
- If there is an up-regulation in an hour, it will be balance in the system in the next hour.
- If there is balance in the system in an hour, it will be balance in the system in the next hour.
- If there is balance in the system in an hour, it will be a down-regulation in the next hour.
- If there is balance in the system in an hour, it will be an up-regulation in the next hour.

When examining the aggregated times series of the balancing power volume a count of these events has been hold track on. In addition the number of times when one regulation state has been changed with another one or remained the same after x hours has been counted. For example, it has been counted how many times a down-regulation has changed an up-regulation in the system after 5 hours with an up-regulation in the system.

Then the possibility for that regulation state will remain the same in the several hours has been calculated by using the following formula:

$$P(RS\text{->}RS)=\frac{Number\ of\ cases\ when\ regulation\ state\ has\ remained\ the\ same\ after\ x\ hours}{Number\ of\ hours\ with\ down-regulation}$$

In order to calculate the possibility for that a regulation state in the power system will be changed by another regulation state after x hours, the Bayes' theorem has been used:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where

A-regulation state in hour x

B-regulation state in hour x

P(A)-probability for regulation state A

P(B)-probability for regulation state B

$P(A|B)$- probability for observing regulation state A given that B is true

$P(B|A)$- probability for observing regulation state B given that A is true

The results of this examination are shown in the table below:

Hour	The possibility for that a regulation state will change or will continue								
	Down->Down	Down->Up	Down->Balance	Up->Up	Up->Down	Up->Balance	Balance->Balance	Balance->Down	Balance->Up
1	0,894	0,069	0,036	0,863	0,013	0,048	0,559	0,232	0,209
2	0,735	0,168	0,097	0,656	0,057	0,126	0,328	0,355	0,317
3	0,606	0,210	0,184	0,530	0,054	0,215	0,208	0,407	0,385
4	0,501	0,302	0,197	0,432	0,059	0,273	0,138	0,504	0,358
5	0,433	0,288	0,279	0,351	0,055	0,393	0,089	0,516	0,396
6	0,378	0,268	0,355	0,288	0,059	0,417	0,055	0,600	0,345
7	0,316	0,421	0,263	0,232	0,076	0,413	0,033	0,322	0,644
8	0,273	0,498	0,229	0,191	0,067	0,469	0,022	0,544	0,435
9	0,241	0,358	0,401	0,163	0,073	0,384	0,014	0,448	0,538
10	0,208	0,417	0,375	0,135	0,089	0,376	0,010	0,424	0,566
11	0,187	0,502	0,311	0,123	0,037	0,482	0,005	0,284	0,711
12	0,164	0,544	0,293	0,108	0,046	0,535	0,003	0,499	0,499
13	0,147	0,518	0,335	0,094	0,069	0,415	0,002	0,000	0,998
14	0,127	0,462	0,411	0,081	0,063	0,501	0,002	0,000	0,000
15	0,114	0,548	0,337	0,066	0,040	0,747	0,002	0,000	0,000
16	0,105	0,559	0,336	0,058	0,046	0,579	0,002	0,000	0,000
17	0,092	0,495	0,413	0,046	0,127	0,381	0,001	0,000	0,999
18	0,082	0,516	0,401	0,041	0,053	0,533	0,001	0,000	0,999
19	0,075	0,569	0,356	0,037	0,030	0,688	0,001	0,000	0,000
20	0,068	0,645	0,287	0,033	0,034	0,691	0,001	0,000	0,000
21	0,063	0,268	0,669	0,031	0,053	0,000	0,001	0,000	0,000
22	0,057	0,600	0,343	0,027	0,041	0,695	0,001	0,000	0,000
23	0,052	0,237	0,711	0,024	0,024	0,781	0,001	0,000	0,000
24	0,046	0,174	0,781	0,021	0,026	0,734	0,001	0,000	0,000
25	0,042	0,479	0,479	0,020	0,029	0,654	0,001	0,000	0,000
26	0,040	0,720	0,240	0,017	0,095	0,393	0,000	0,000	1,000
27	0,039	0,000	0,961	0,015	0,038	0,000	0,000	0,000	0,000
28	0,037	0,642	0,321	0,014	0,039	0,000	0,000	0,000	0,000
29	0,033	0,322	0,645	0,013	0,123	0,000	0,000	0,000	0,000
30	0,030	0,323	0,647	0,012	0,047	0,000	0,000	0,000	0,000
31	0,027	0,195	0,779	0,011	0,000	0,989	0,000	0,000	0,000
32	0,023	0,559	0,419	0,010	0,000	0,990	0,000	0,000	0,000
33	0,020	0,245	0,735	0,008	0,124	0,000	0,000	0,000	0,000
34	0,019	0,000	0,981	0,008	0,071	0,000	0,000	0,000	0,000
35	0,017	0,655	0,328	0,007	0,000	0,993	0,000	0,000	0,000
36	0,017	0,000	0,983	0,007	0,000	0,993	0,000	0,000	0,000
37	0,017	0,000	0,000	0,005	0,000	0,995	0,000	0,000	0,000
38	0,015	0,492	0,492	0,005	0,111	0,000	0,000	0,000	0,000
39	0,015	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000

Hour	The possibility for that a regulation state will change or will continue								
	Down->Down	Down->Up	Down->Balance	Up->Up	Up->Down	Up->Balance	Balance->Balance	Balance->Down	Balance->Up
40	0,014	0,493	0,493	0,004	0,000	0,996	0,000	0,000	0,000
41	0,014	0,986	0,000	0,004	0,000	0,000	0,000	0,000	0,000
42	0,012	0,659	0,329	0,004	0,000	0,000	0,000	0,000	0,000
43	0,011	0,989	0,000	0,004	0,000	0,996	0,000	0,000	0,000
44	0,011	0,000	0,989	0,003	0,000	0,997	0,000	0,000	0,000
45	0,009	0,495	0,495	0,002	0,200	0,000	0,000	0,000	0,000
46	0,009	0,000	0,991	0,002	0,000	0,998	0,000	0,000	0,000
47	0,008	0,496	0,496	0,002	0,000	0,000	0,000	0,000	0,000
48	0,007	0,497	0,497	0,002	0,000	0,000	0,000	0,000	0,000
49	0,007	0,000	0,000	0,001	0,333	0,000	0,000	0,000	0,000
50	0,007	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
51	0,006	0,000	0,994	0,001	0,000	0,000	0,000	0,000	0,000
52	0,006	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
53	0,006	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
54	0,006	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
55	0,005	0,000	0,000	0,001	0,500	0,000	0,000	0,000	0,000
56	0,005	0,995	0,000	0,001	0,000	0,000	0,000	0,000	0,000
57	0,005	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
58	0,005	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
59	0,005	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
60	0,004	0,498	0,498	0,001	0,000	0,000	0,000	0,000	0,000
61	0,003	0,000	0,997	0,001	0,000	0,000	0,000	0,000	0,000
62	0,002	0,000	0,998	0,001	0,000	0,000	0,000	0,000	0,000
63	0,002	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
64	0,002	0,998	0,000	0,001	0,000	0,000	0,000	0,000	0,000
65	0,002	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
66	0,002	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
67	0,001	0,999	0,000	0,001	0,000	0,000	0,000	0,000	0,000
68	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
69	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
70	0,001	0,999	0,000	0,001	0,000	0,000	0,000	0,000	0,000
71	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
72	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
73	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
74	0,001	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000
75	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
76	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
77	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
78	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
79	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
80	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
81	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
82	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Hour	The possibility for that a regulation state will change or will continue								
	Down->Down	Down->Up	Down->Balance	Up->Up	Up->Down	Up->Balance	Balance->Balance	Balance->Down	Balance->Up
83	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
84	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
85	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
86	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Appendix B

Following time series have been used in the statistical analysis of the balancing power volume:

- Balancing power volume contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the balancing power volume for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2 and for each regulation state, i.e. up-regulation and down-regulation. Source is Statkraft.
- Actual power consumption contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the actual power consumption for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2. Source is Statkraft.
- Forecasted power consumption contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the Forecasted power consumption for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2. Source is Statkraft.
- Actual wind power production contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the actual wind power production for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2. Source is Statkraft.
- Forecasted wind power production contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the Forecasted wind power production for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2. Source is Statkraft.
- Actual temperature contains weighted hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the actual temperature for cities Oslo, Bergen, Trondheim, Tromsø, Stockholm, Malmo, Göteborg, Luleå and Sundsvall. Source is Statkraft.
- Forecasted temperature contains weighted hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the forecasted temperature for cities Oslo, Bergen, Trondheim, Tromsø, Stockholm, Malmo, Göteborg, Luleå and Sundsvall. Source is Statkraft.
- Actual power flow contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the actual power flow between the price areas DEU-SE4, POL-SE4, DEU-DK1, DEU-DK2, NLD-NO2, FIN-SE3, FIN-SE1. Source is Statkraft.
- Spot power flow contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the spot power flow between the price areas DEU-SE4, POL-SE4, DEU-DK1, DEU-DK2, NLD-NO2, FIN-SE3, FIN-SE1. Source is Statkraft.
- Special regulation contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the special regulation for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2 and for each regulation state, i.e. up-regulation and down-regulation. Source is Nordpool.

- Elbas volume contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been aggregated from the time series of the Elbas volume for each of the price areas price area, i.e. NO1, NO2, NO3, NO4, NO5, SE1, SE2, SE3, SE4, DK1 and DK2 and for each regulation state, i.e. up-regulation (volume bought in the Elbas market) and down-regulation (volume sold in the Elbas market). Source is Nordpool.

Appendix C

Left hand side of balance equation	Right hand side of balance equation	Correlation between the left and the right hand side of equation
Balancing power + ΔPower flow + Special regulation	*ΔWind power production + ΔConsumption + Elbas volume*	0.59
Balancing power + ΔPower flow	*ΔWind power production + ΔConsumption + Elbas volume*	0.59
Balancing power + Special regulation	*ΔWind power production + ΔConsumption + Elbas volume*	0.57
Balancing power	*ΔWind power production + ΔConsumption + Elbas volume*	0.49

Left hand side of balance equation	Right hand side of balance equation	Correlation between the left and the right hand
Balancing power + ΔPower flow + Special regulation	*ΔWind power production + Elbas volume*	0.31
Balancing power + ΔPower flow + Special regulation	*ΔConsumption + Elbas volume*	0.58
Balancing power + ΔPower flow + Special regulation	*ΔWind power production + ΔConsumption*	0.54
Balancing power + ΔPower flow + Special regulation	*ΔConsumption*	0.52
Balancing power + ΔPower flow + Special regulation	*ΔWind power production*	0.20
Balancing power + ΔPower flow + Special regulation	*Elbas volume*	0.35

Left hand side of balance equation	Right hand side of balance equation	Correlation between the left and the right hand
Balancing power + ΔPower flow	*ΔWind power production + Elbas volume*	0.34
Balancing power + ΔPower flow	*ΔConsumption + Elbas volume*	0.57
Balancing power + ΔPower flow	*ΔWind power production + ΔConsumption*	0.54
Balancing power + ΔPower flow	*ΔConsumption*	0.49
Balancing power + ΔPower flow	*ΔWind power production*	0.20
Balancing power + ΔPower flow	*Elbas volume*	0.36

Left hand side of balance equation	Right hand side of balance equation	Correlation between the left and the right hand
Balancing power + Special regulation	*ΔWind power production + Elbas volume*	0.16
Balancing power + Special regulation	*ΔConsumption + Elbas volume*	0.47
Balancing power + Special regulation	*ΔWind power production + ΔConsumption*	0.50
Balancing power + Special regulation	*ΔConsumption*	0.49
Balancing power + Special regulation	*ΔWind power production*	0.17
Balancing power + Special regulation	*Elbas volume*	0.01

Left hand side of balance equation	Right hand side of balance equation	Correlation between the left and the right hand
Balancing power	*ΔWind power production + Elbas volume*	0.19
Balancing power	*ΔConsumption + Elbas volume*	0.44
Balancing power	*ΔWind power production + ΔConsumption*	0.51
Balancing power	*ΔConsumption*	0.49
Balancing power	*ΔWind power production*	0.20
Balancing power	*Elbas volume*	0.01

Appendix D

Following time series have been used in the statistical analysis of the balancing power premium:

- Balancing power premium in the price area NO3 contains hourly observation from 1.01.2013 to 31.12.2015. The time series has been conducted by substitution of the values of the spot price from the values of the balancing power price. There are different time series for each regulation state, i.e. up-regulation and down-regulation. Source is Statkraft.
- Balancing power volume in the price area NO3 contains hourly observation from 1.01.2013 to 31.12.2015. There are different time series for each regulation state, i.e. up-regulation and down-regulation. Source is Statkraft.
- Inflow in the price area NO3 contains hourly observation from 1.01.2013 to 31.12.2015. Source is Statkraft.
- Spot price in the price area NO3 contains hourly observation from 1.01.2013 to 31.12.2015. Source is Statkraft.
- The slope of the bud curve of the Elspot market contains hourly observation from 1.01.2013 to 31.12.2015. The values of the slope of the bud curve have been calculated from the bud curve of the Elspot market provided by Statkraft.

Appendix E

```
----------------------------------------------------
procedure $daily_variation
argument %series_list
 ----------------------------------------------------
local scalar %serie_index : precision = 0
local scalar %snitt_list : namelist = {}
local scalar %daglengde : precision = 27000


  block
  frequency hourly
  date <frequency daily> 1Jan13                    --colculate the average value of the balancing power volume/price per hour of the day over the analysis period
  loop for %serie in %series_list
    set %serie_index = %serie_index + 1
    local series id("%snitt_" + name(%serie)) : precision by date
    set id("%snitt_" + name(%serie)) = series(0, date)
    loop for %skift = 0 to %daglengde - 1
      set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie)) + overlay(shift(%serie,24*%skift), series(0, date))
    end loop
    set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie))/%daglengde
    set %snitt_list = %snitt_list + {id("%snitt_" + name(%serie))}

  end loop
    type %snitt_list  --plot the average value
  plot #2 color blue, solid
  plot #3 color red, solid
  plot #4 color magenta, solid
  plot #5 color green, solid
  plot #6 color yellow, solid

  legend off
  graph %snitt_list
end block
end procedure
```

```
----------------------------------------------------
procedure $weekly_variation
 argument %serie_list
 ----------------------------------------------------
  local scalar %serie_index : precision = 0
  local scalar %snitt_list : namelist = {}
  local scalar %daglengde : precision

  block
  frequency weekly(sunday)    --colculate the average value of the balancing power volume per week day over the analysis period
  date <frequency annual> 2013 to 2015
  set %daglengde = lengthdate
  date firstdate
  frequency daily
  disp @date
  loop for %serie in %serie_list
    set %serie_index = %serie_index + 1
    local series id("%snitt_" + name(%serie)) : precision by date
    set id("%snitt_" + name(%serie)) = series(0, date)
    loop for %skift = 0 to %daglengde - 1
      set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie)) + overlay(shift(convert(%serie, daily), 7*%skift), series(0, date))
    end loop
    set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie))/%daglengde
    set %snitt_list = %snitt_list + {id("%snitt_" + name(%serie))}
  end loop
  plot #1 color black, solid    --plot the average value
  plot #2 color blue, solid
  plot #3 color red, solid
  plot #4 color magenta, solid
  plot #5 color green, solid
  plot #6 color yellow, solid
  legend off
  graph %snitt_list
  end block
end procedure
```

```
----------------------------------------------
procedure $annual_variation
 argument %serie_list
-----------------------------------------------
    local scalar %serie_index : precision = 0
    local scalar %snitt_list : namelist = ()
    local scalar %daglengde : precision

    block
    frequency annual         --colculate the average value of the balancing power volume per month over the analysis period
    date 2013 to 2015
    set %daglengde = lengthdate
    date firstdate
    frequency monthly
    loop for %serie in %serie_list
    type name(%serie)
      set %serie_index = %serie_index + 1
      local series id("%snitt_" + name(%serie)) : precision by date
      set id("%snitt_" + name(%serie)) = series(0, date)
      loop for %skift = 0 to %daglengde - 1
        set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie)) + overlay(convert(shiftyr(%serie, %skift), monthly, constant, averaged, daily, on), series(0, date))
        end loop
      set id("%snitt_" + name(%serie)) = id("%snitt_" + name(%serie))/%daglengde
      set %snitt_list = %snitt_list + (id("%snitt_" + name(%serie)))
    end loop
    plot #1 color black, solid       --plot the average value
    plot #2 color blue, solid
    plot #3 color red, solid
    plot #4 color magenta, solid
    plot #5 color green, solid
    plot #6 color yellow, solid
    plot #7 color cyan, solid
  legend off
  graph %snitt_list
  end block
end procedure
```

Appendix F

```
-----------------------------------
procedure $sell_buy_curve
-----------------------------------
  local series <frequency hourly> %st_tall.h: precision by date
  local scalar %tall: precision
  local scalar %ind_1: precision
  local scalar %ind_2: precision
  local scalar %ind_3: precision
  local scalar %sys_price: precision

  block
    load file("Eirik_data\buy_sell.pc")
    local series %price_total.c, %volume_total.c : precision by case
    local scalar <fre h> %hourly_date : date
    frequency hourly
    date 01Jan13:00 to 31Dec15:23
    loop for %hourly_date in date
      block
        block
          case *
          which not missing(%price_total.c)
          if lengthcase gt 0
            set %price_total.c = series(ND, case)
            set %volume_total.c = series(ND, case)
          end if
        end block
        frequency hourly
        date %hourly_date
        $buy_sell_total %hourly_date, %price_total.c, %volume_total.c
        if not missing(%price_total.c[1])
          set %sys_price = PRICE.SPOT.EL.SP1._.NPS.EUR.MWH.H[%hourly_date]
          block
          which %price_total.c le %sys_price
          set %ind_1 = lastcase
          end block
            block
            which %volume_total.c le lsum(%volume_total.c[%ind_1],-1000)
            set %ind_2 = lastcase
            end block
              block
            which %volume_total.c le lsum(%volume_total.c[%ind_1],1000)
            set %ind_3 = lastcase
          end block
          set %tall = (%price_total.c[%ind_3]-%price_total.c[%ind_2])/(%volume_total.c[%ind_3] - %volume_total.c[%ind_2])
          set %st_tall.h[%hourly_date] = %tall
        end if
      end block
    end loop
    $save_rk %st_tall.h
  end block
end procedure
```

Appendix G

Parameters of the Decision Forest regression model:

Minimum number of samples per leaf node	Number of random splits per node	Maximum depth of the trees	Number of trees	MAE	RMSE	Coefficient of Determination
15	582	4	9	19.97	45.42	0.16
6	1015	3	22	19.88	46.98	0.10
11	1018	20	2	19.10	37.54	0.43
6	24	7	28	18.97	41.80	0.29
15	80	33	5	18.66	36.79	0.45
15	18	42	9	18.64	37.79	0.42
13	628	7	24	18.62	39.90	0.35
7	495	14	3	17.57	35.57	0.49
14	31	33	30	17.52	36.06	0.47
7	955	9	22	17.40	36.47	0.46
1	370	54	2	17.37	37.43	0.43
15	602	49	11	17.37	35.11	0.50
11	913	42	8	17.09	34.15	0.53
14	558	50	16	16.93	34.41	0.52
13	658	37	18	16.79	34.12	0.53
13	582	41	17	16.74	34.09	0.53
8	59	18	16	16.44	33.26	0.55
10	850	44	18	16.21	32.96	0.56
4	136	57	6	16.10	32.01	0.58
8	653	21	12	16.05	32.61	0.57
6	535	18	16	15.42	31.45	0.60
4	485	13	28	15.42	31.77	0.59
1	398	52	7	15.41	31.16	0.61
7	998	56	31	15.29	31.52	0.60
3	189	39	8	15.22	30.68	0.62
4	155	29	14	15.14	30.59	0.62
4	318	16	25	14.93	30.66	0.62
5	722	61	31	14.73	30.31	0.63
2	200	32	19	14.37	28.93	0.66
2	627	36	21	14.29	28.98	0.66

Parameters of the Boosted Decision Tree regression model:

Number of leaves	Minimum leaf instances	Learning rate	Number of trees	MAE	RMSE	Coefficient of Determination
2	18	0.34	49	22.55	47.03	0.10
4	20	0.33	117	19.54	37.60	0.43
86	49	0.14	50	17.41	31.79	0.59
46	49	0.04	352	16.92	31.21	0.60
37	35	0.38	494	16.78	27.32	0.70
54	48	0.36	486	16.69	26.88	0.71
60	24	0.11	54	16.61	32.21	0.58
18	31	0.23	334	16.57	28.61	0.67
123	29	0.31	175	16.45	27.38	0.70
91	44	0.27	142	16.33	27.41	0.69
79	41	0.28	293	16.11	26.70	0.71
52	46	0.07	355	16.05	28.54	0.67
113	27	0.32	266	15.84	26.57	0.71
127	28	0.05	146	15.69	29.44	0.65
108	28	0.27	267	15.67	26.38	0.72
63	32	0.14	197	15.66	27.63	0.69
104	32	0.24	285	15.53	26.46	0.72
50	26	0.13	263	15.48	27.71	0.69
23	10	0.25	140	15.38	28.62	0.67
35	24	0.10	446	15.29	27.53	0.69
34	7	0.36	105	15.21	27.64	0.69
17	10	0.21	299	15.17	27.74	0.69
108	31	0.06	376	14.82	26.77	0.71
34	16	0.12	396	14.78	26.89	0.71
31	8	0.19	236	14.40	26.59	0.71
124	4	0.22	94	13.46	26.33	0.72
127	1	0.27	153	13.12	26.73	0.71
44	2	0.06	441	13.08	26.03	0.72
68	3	0.13	258	12.88	25.64	0.73
117	2	0.22	482	12.83	26.08	0.72

Appendix H

Parameters of the Decision Forest regression model:

Minimum number of samples per leaf node	Number of random splits per node	Maximum depth of the trees	Number of trees	MAE	RMSE	Coefficient of Determination
6	1015	3	22	195.34	246.84	0.66
15	582	4	9	169.51	213.40	0.74
6	24	7	28	130.37	165.62	0.84
13	628	7	24	127.34	162.99	0.85
7	955	9	22	112.11	146.05	0.88
15	18	42	9	109.22	142.90	0.88
11	1018	20	2	106.83	145.29	0.88
15	80	33	5	104.19	138.26	0.89
1	370	54	2	103.89	147.44	0.88
14	31	33	30	103.80	136.91	0.89
7	495	14	3	101.31	138.69	0.89
15	602	49	11	101.02	135.10	0.90
14	558	50	16	99.46	133.00	0.90
13	582	41	17	98.75	132.26	0.90
13	658	37	18	98.53	132.66	0.90
11	913	42	8	98.52	133.30	0.90
10	850	44	18	96.15	130.49	0.90
8	59	18	16	96.08	129.54	0.91
8	653	21	12	94.61	129.15	0.91
4	485	13	28	93.35	126.76	0.91
4	136	57	6	92.89	128.37	0.91
7	998	56	31	92.15	126.54	0.91
6	535	18	16	92.06	126.71	0.91
1	398	52	7	91.02	127.64	0.91
3	189	39	8	90.40	126.15	0.91
4	155	29	14	89.94	124.14	0.91
5	722	61	31	89.75	123.94	0.91
4	318	16	25	89.22	123.51	0.91
2	627	36	21	87.27	122.54	0.92
2	200	32	19	86.76	121.39	0.92

Parameters of the Boosted Decision Tree regression model:

Number of leaves	Minimum leaf instances	Learning rate	Number of trees	MAE	RMSE	Coefficient of Determination
2	18	0.34	49	138.10	174.46	0.83
4	20	0.33	117	117.92	151.96	0.87
60	24	0.11	54	101.23	132.72	0.90
46	49	0.04	352	97.99	130.04	0.90
86	49	0.14	50	97.92	129.86	0.90
23	10	0.25	140	97.81	130.14	0.90
17	10	0.21	299	96.64	128.78	0.91
34	7	0.36	105	95.93	128.77	0.91
18	31	0.23	334	95.27	127.66	0.91
31	8	0.19	236	93.37	125.48	0.91
52	46	0.07	355	92.88	124.61	0.91
44	2	0.06	441	92.54	124.05	0.91
127	28	0.05	146	92.42	124.15	0.91
37	35	0.38	494	92.15	125.18	0.91
34	16	0.12	396	91.63	123.54	0.91
35	24	0.10	446	91.61	123.77	0.91
50	26	0.13	263	91.24	123.20	0.91
63	32	0.14	197	91.23	122.77	0.91
54	48	0.36	486	91.06	124.05	0.91
91	44	0.27	142	90.85	123.08	0.91
79	41	0.28	293	89.59	122.10	0.92
127	1	0.27	153	89.24	123.44	0.91
113	27	0.32	266	89.15	122.38	0.92
123	29	0.31	175	88.61	121.90	0.92
108	28	0.27	267	88.59	121.88	0.92
68	3	0.13	258	88.47	120.73	0.92
108	31	0.06	376	87.92	119.83	0.92
104	32	0.24	285	87.75	120.57	0.92
124	4	0.22	94	87.62	120.43	0.92
117	2	0.22	482	86.06	119.96	0.92

Appendix I

Parameters of the Decision Forest regression model:

Minimum number of samples per leaf node	Number of random splits per node	Maximum depth of the trees	Number of trees	MAE	RMSE	Coefficient of Determination
6	24	7	28	3.55	10.22	0.00
15	18	42	9	3.53	10.23	0.00
14	31	33	30	3.38	10.08	0.03
8	59	18	16	3.12	9.89	0.07
1	370	54	2	2.86	8.68	0.28
4	136	57	6	2.80	8.78	0.26
7	495	14	3	2.78	8.58	0.30
15	80	33	5	2.77	9.38	0.16
15	582	4	9	2.74	9.51	0.13
11	1018	20	2	2.73	8.38	0.33
6	1015	3	22	2.71	9.48	0.14
8	653	21	12	2.70	8.42	0.32
3	189	39	8	2.69	8.60	0.29
7	955	9	22	2.68	8.74	0.27
4	155	29	14	2.68	9.08	0.21
13	582	41	17	2.67	8.34	0.34
2	200	32	19	2.66	9.05	0.22
6	535	18	16	2.65	8.63	0.29
4	318	16	25	2.65	9.07	0.21
4	485	13	28	2.63	8.89	0.24
13	628	7	24	2.59	9.15	0.20
15	602	49	11	2.55	8.30	0.34
14	558	50	16	2.55	8.28	0.34
13	658	37	18	2.52	8.29	0.34
2	627	36	21	2.51	8.42	0.32
10	850	44	18	2.43	8.15	0.37
1	398	52	7	2.41	8.43	0.32
11	913	42	8	2.39	8.15	0.36
5	722	61	31	2.29	8.02	0.39
7	998	56	31	2.24	8.02	0.39

Parameters of the Boosted Decision Tree regression model:

Number of leaves	Minimum leaf instances	Learning rate	Number of trees	MAE	RMSE	Coefficient of Determination
2	18	0.34	49	1.93	7.97	0.39
54	48	0.36	486	1.88	7.68	0.44
37	35	0.38	494	1.88	7.68	0.44
113	27	0.32	266	1.86	7.71	0.43
108	28	0.27	267	1.86	7.67	0.44
104	32	0.24	285	1.85	7.64	0.44
117	2	0.22	482	1.84	7.83	0.41
123	29	0.31	175	1.84	7.61	0.45
79	41	0.28	293	1.84	7.53	0.46
4	20	0.33	117	1.83	7.25	0.50
18	31	0.23	334	1.83	7.18	0.51
63	32	0.14	197	1.79	7.40	0.48
91	44	0.27	142	1.79	7.32	0.49
34	7	0.36	105	1.79	7.18	0.51
31	8	0.19	236	1.78	7.26	0.50
52	46	0.07	355	1.77	7.31	0.49
46	49	0.04	352	1.77	7.36	0.48
108	31	0.06	376	1.77	7.37	0.48
17	10	0.21	299	1.77	7.16	0.51
23	10	0.25	140	1.76	7.15	0.51
50	26	0.13	263	1.75	7.26	0.50
124	4	0.22	94	1.74	7.43	0.47
35	24	0.10	446	1.74	7.22	0.50
86	49	0.14	50	1.74	7.44	0.47
68	3	0.13	258	1.74	7.48	0.46
34	16	0.12	396	1.71	7.17	0.51
44	2	0.06	441	1.71	7.43	0.47
127	28	0.05	146	1.69	7.21	0.50
60	24	0.11	54	1.67	7.21	0.50
127	1	0.27	153	1.57	8.31	0.34