



GRADUATION THESIS

---

**Approximating the Flow-Based Transport Capacity  
Constraints for the Day-Ahead Power Market.**

---

APRIL 22, 2016

FEMIA VAN STIPHOUT

**GRADUATION COMMITTEE**

UNIVERSITY OF TWENTE

PROF. DR. JOHANN HURINK

PROF. DR. MARC UETZ

DR. IR. ALBERT MOLDERINK

ENECO

VINCENT VISSER









# Abstract

Eneco trades power on several power markets, including the day-ahead power market, where power is exchanged for delivery the next day. On the day-ahead power market market parties of different countries participate, which makes it possible to import or export power from other countries. When making trades between market parties in different countries, transport capacity must be taken into account and can limit the amount of exchange between the countries. Consequently the transport capacity has influence on the market clearing prices in different countries. The transport capacity is determined in advance by the Transmission System Operators (TSOs) of the countries participating in the day-ahead power market. Since the 20<sup>th</sup> of May 2015 the TSOs use a so-called ‘flow-based’ model to determine the transport capacity. As the details of the used approach are not public, Eneco does not know in advance how the transport capacity resulting from the flow-based model is impacting the trades for the next day. This makes it harder for Eneco to forecast the day-ahead market clearing prices.

In this research, different methods are designed based on flow-based transport capacity data from the past, to predict the transport capacity constraints used in an upcoming day. The method Eneco currently uses to estimate the transport capacity is used as benchmark for the designed methods. The first designed method is a fairly naive method, which searches for one or more similar hours (similar in the sense that certain characteristics of these hours are equal) and uses this hour to estimate the transport capacity. Another designed method uses probabilities to determine which transport capacity constraints from the past are used to estimate the transport capacity. The last designed method also takes correlation between the constraints into account, when deciding which constraints from the past are used to estimate the transport capacity. The main results of this research is that the benchmark method works relatively well, but the probabilistic method provides also a quite good estimation, when comparing how close the forecast market clearing price is to the actual market clearing price, in mean absolute error. The similar hour method can be less accurate but in some specific cases it provides a relatively good estimation. It turns out that the probabilistic method in combination with correlation does not provide a good estimation for the transport capacity.

**Keywords:** *Flow-Based domain, Day-Ahead Power Market, PTDF-matrix, Flow-Based Transport Capacity Constraints.*







# Preface

After more than six and a half years of studying, it is time to conclude my study with this thesis. Six and a half year which contained solving problems, studying for exams, learning study materials, teach some basic mathematics to new students, being a board member of a study association, being a chairwoman of a sports association, getting to know new people, getting to know more new people, looking for a job, many highlights, also some low lights, living in Barcelona and so much more. Now it is time to finish my study applied mathematics at the University of Twente.

Studying mathematics has brought me much knowledge but also many skills which I used during the writing of this thesis. I started studying applied mathematics (partly) because of my mathematics teacher, whom was a very inspiring man. A quote related to him, which I would never forget is:

“Kies één leven, zonder terugleggen. Wat is de kans dat het geweldig word?”

I never regretted my decision to study applied mathematics in Enschede. I still think it is an amazing study from which I will profit the rest of my life. Writing this thesis is a good finishing piece of my study. I was able to do this research for Eneco, which provided me with insights in the possible applications of the mathematics I learned during my study. At the Eneco Energy Trade department I always enjoyed the ambiance, it was never a boring place to work, which is the perfect working environment for me. Although my research was very specific, it always felt useful for Eneco. Here I would like to thank Vincent Visser for being my supervisor at Eneco, for always thinking about the practical stuff I sometimes forgot about and for the weekly meetings which were always useful.

Something I learned during the writing of this thesis is that graduating consists of two parts, doing research and writing about it. Both parts were challenging to me. Writing is something I can do fast, but structuring the problem and writing it in such a way that someone else can understand it, can be a real challenge. I want to thank Johann and Marc for the many, many, many times they have read my report and gave comments. Without those comments I think it would be a lot harder to understand what I did for the last seven months. Also I want to thank Johann, Marc en Vincent for all the other inspirational moments, brainstorm sessions and other practical stuff.

Last but not least I want to thank my family and friends who supported me during the writing of my thesis. Especially I want to thank my boyfriend Marco who was always there for me and supported me no matter what. I hope you enjoy reading this thesis!



# Contents

Abstract

Preface

<b>Table of Contents</b>	<b>1</b>
<b>1 Introduction into Cross-Border Electricity Trade</b>	<b>3</b>
1.1 Background of Trading Power . . . . .	3
1.2 Research Questions and Solution Approach . . . . .	18
1.3 Background Literature of Energy Trade and Forecasting . . . . .	20
<b>2 The Parameters within the Flow-Based Model</b>	<b>23</b>
2.1 Branches Considered in Different Scenarios . . . . .	23
2.2 Estimation of Flows in the Electrical Network . . . . .	26
<b>3 Mathematical Problem Description</b>	<b>29</b>
3.1 Input Characteristics for the Flow-Based Domain Generation . . . . .	29
3.2 Function Used to Generate the Flow-Based Domain . . . . .	29
3.3 Mathematical Description of the Problem . . . . .	30
<b>4 Characteristics of Influence</b>	<b>32</b>
4.1 Comparing Flow-Based Domains . . . . .	32
4.2 Testing Different Methods on (a Subset of) Hours . . . . .	37
4.3 Likely Candidates to be Characteristics of Influence on the Flow-Based Domain .	38
4.4 Definition of Equivalence Classes . . . . .	39
4.5 Analyse Methods to find the Characteristic of Influence . . . . .	40
4.6 Results of the Characteristics Analysis . . . . .	45
4.7 Number of Hours per Equivalence Class . . . . .	48
4.8 Choice of a Set of Characteristics . . . . .	50
<b>5 Approximation Methods</b>	<b>51</b>
5.1 Using Similar Hours to Approximate the Flow-Based Domain . . . . .	51
5.2 Using Probabilities to Approximate the Flow-Based Domain . . . . .	52
5.3 Probabilistic Method Including Correlation . . . . .	56
<b>6 Results</b>	<b>59</b>
6.1 Error Calculation . . . . .	59
6.2 Visualisation and Interpretation of the Results . . . . .	60



6.3	Statistically Comparing Different Methods . . . . .	70
6.4	Validation and Verification . . . . .	72
<b>7</b>	<b>Conclusions, Recommendations and Discussion</b>	<b>74</b>
7.1	Main Conclusions of the Results . . . . .	74
7.2	Recommendations for Further Research . . . . .	75
7.3	Discussion and Remarks on this Research . . . . .	76
	<b>Symbols and Abbreviations</b>	<b>78</b>
	Symbols . . . . .	78
	Abbreviations . . . . .	80
<b>A</b>	<b>Appendix</b>	<b>82</b>
A.1	The Different Test-Sets . . . . .	82
A.2	Detailed Calculation of the RAM-value . . . . .	86
A.3	Results . . . . .	88
A.4	Visualisation of the Results Continued . . . . .	91
A.5	Results Statistical Tests . . . . .	100
	<b>Bibliography</b>	<b>101</b>



# Chapter 1

## Introduction into Cross-Border Electricity Trade

Eneco is an electrical power company which core business is providing its customers with electricity and gas. For the electricity part Eneco buys electricity from electricity producers in the Netherlands, but also from producers in other countries. In order to be a smart player on the electricity markets, it is good to know how the prices of electricity behave in the Dutch market, but also in the markets of other countries. The prices in the different countries are dependent on certain factors, for example electricity cannot be transported unlimited, so the amount of electricity that can be bought from a certain country is restricted. This restriction can influence the price in the Netherlands but also in the country where the electricity is produced. For Eneco it is valuable to know how the prices in the different markets are influenced, for this we have to know first what different markets there are, but also how the transport of electricity works and what the restrictions on this transport are. For this the following sections give some background about trading electricity, transport of electricity, the restrictions on transporting electricity and the market coupling between the countries in Europe. After this background description we are able to define the problem researched in this thesis and present the research questions used to tackle the problem. Also this chapter contains the relevant literature subject to the problem, which provides some insight in what is already researched about this subject.

### 1.1 Background of Trading Power

The underlying subject of this research is the power market. This section describes the economic background of the power market and explains the different power markets, which provides some necessary context about power markets relevant for the research of this thesis.

#### 1.1.1 Power Markets

Trading power is done on different power markets. The products traded on those markets differ in time up to delivery. For example, in the forward and futures market, the power that is traded has to be delivered at a date which is weeks, months or even years ahead. In Figure 1.1 an overview of the power markets and their time period is given. In [9] a more extensive description



of the electricity markets can be found, here we briefly discuss these markets.



Figure 1.1: The different power markets and their time period.

Electric power companies, such as Eneco, trade power on the power markets described in the following paragraphs. The electric power companies have an obligation to their customers to provide them with an uninterrupted flow of electricity. Normally the customers have a contract for a fixed electricity price for a year, but on the power market the price of electricity is not fixed, but time dependent. If now the price of electricity is higher than the fixed price the customers pay for the electricity, the electric power company loses money. To reduce the risk of loss, electric power companies buy and sell power on the forward and futures power market. Furthermore, electric power companies have often certain assets to generate power. The power markets also create the possibility to make as much money as possible with these assets.

#### 1.1.1.1 Forward and Futures Power Market

The forward and futures power market consists of products that have a delivery date, up to the next trading day, in the future. In the forward and futures power market the price movements are much smaller than in the other power markets. This is because the availability of power plants, the amount of demand and the weather are still unknown. The forward and futures power market provides market participants with the possibility to manage their long term risk of loss. Market participants can already buy contracts to purchase or sell electricity with a delivery date weeks, months, quarters, seasons or even years ahead.

#### 1.1.1.2 Day-Ahead Power Market

In the day-ahead power market, traders place their orders for delivery of power the next day. A platform for the exchange of power, the so-called *power exchange*, collects these orders for a certain group of market participants. Each power exchange only collects the orders of the market participants in a certain geographical area, the so-called *market area* (usually market participants of the same country belong to the same market area). In the Netherlands this power exchange is the *Amsterdam Power Exchange (APX)*.

Power  
Exchange  
  
Market Area  
  
Amsterdam  
Power  
Exchange  
(APX)

Special about the day-ahead market is that, at the time the orders are placed, the actual electricity demand and supply (of weather dependent power sources, such as wind turbines or solar panels) are still unknown. Electricity producers have to forecast the amount of electricity its weather dependent electricity generators are going to produce and electricity consumers have to forecast the amount of electricity its clients are going to consume.



Market  
Clearing  
Price (MCP)  
Matched  
Volume

An order in the day-ahead power market consists of a volume, a price and a time period. For demand orders the price is a maximum price the consumer is willing to pay and for sell orders the price is a minimum price for which the producer is willing to produce. The time period expresses during which time period the electricity is delivered. The individual orders of the market participants are combined to an order curve for each market area. Figure 1.2 gives an example of a combined order curve. The price at the intersection of the demand curve and supply curve corresponds to the *Market Clearing Price (MCP)*. The quantity at the intersection corresponds to the amount of volume for which suppliers are matched to consumers and is called the *matched volume*. The price producers and consumers respectively receive and pay is the market clearing price.

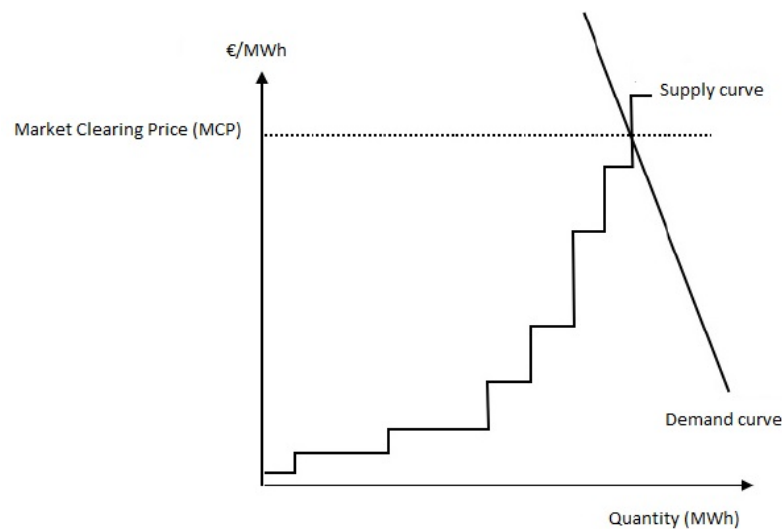


Figure 1.2: A combined order curve for a certain market area.

Source: [8].

Recall that, electric power companies have an obligation to their customers to provide them with an uninterrupted flow of electricity. Because of this, they are in principle willing to pay any price to fulfil this obligation, which results in an buy order with a high (maximum) price. On the other hand, wind turbines always generate power when the wind is blowing, so suppliers with wind turbines send in a supply order with a very low (even negative) (minimum) price. If the wind is blowing but the demand is low, wind turbines are still generating power which cannot be stored. So the suppliers with wind turbines are willing to give away the electricity generated by the wind turbines. In extreme cases the blades of the wind turbines can be pitched, so the turbine does not generate power any more. So in practice those extreme (minimum and maximum) prices almost never occur as market clearing price.

### 1.1.1.3 Intraday Trading

If the forecasts of demand and renewable energy generation, used to send in the orders on the day-ahead market, are not entirely correct, there arises a difference in supply and demand (also called *imbalance*). During the day (after the deadline of the day-ahead market) there is still the

Imbalance



possibility to trade power, called intraday trading. Up to one hour prior to delivery time, trades can be made. If there is still an imbalance after closing of the intraday market the maintainer of the balance in the electricity network, the so-called *Transmission System Operator (TSO)*, solves this problem on the imbalance power market.

Transmission  
System Op-  
erator  
(TSO)

#### 1.1.1.4 Imbalance Power Market

When after the day-ahead market and intraday trading there is still an imbalance between demand for and supply of electricity, further actions are taken by the TSO to maintain the balance on the electrical grid. Market parties can offer to decrease or increase their electricity production or electricity consumption on the imbalance power market, which is called reserve capacity. The price for which the market parties want to increase or decrease their production or consumption is known by the TSO in advance. The TSO may use reserve capacity to restore the balance between demand and supply. The party that causes the imbalance has to pay the price of restoring the balance. Sometimes this price can still be positive. For example, if there is an amount of wind energy which causes the imbalance, a supplier with a gas power plant can offer to turn down the gas power plant. In this case the supplier uses less gas (which saves money) and can pay a small price for the wind energy but still make money from the imbalance.

The power markets described above form the total of electricity trade. For this research we focus on the day-ahead power market. Especially on the day-ahead power market of the *Central Western Europe* (CWE) region, which to this prospect, consists of five countries: The Netherlands (NL), France (FR), Germany (DE), Belgium (BE) and Luxembourg (LU)<sup>1</sup>.

Central  
Western  
Europe  
(CWE)

### 1.1.2 Speciality of Trading Power

Trading electricity is a special kind of trading. Electricity has the character that it is not storable. Therefore, traders can sell their power only in the period where it is produced. On the other hand, they can turn power plants on or off if the price is high or low, but power plants need some time before they are fully launched or turned off. This makes the response time of assets in the power market slow. Besides the storage problem, electricity is a commodity that cannot be transported unlimited, but it is restricted by the capacity of transport cables, lines and transformers. These cables and lines on the one hand allow trading of power between countries, but on the other hand the capacity of those cables and lines may restrict this trading. To stay in line with the nomenclature that is used in power trading the collection of cables and lines, which are used to transport electricity in and between countries, is called *branches*. The branches form the main part of the electrical network. The next subsection gives some more background on this infrastructure.

Branches

### 1.1.3 The Crux of Transporting Power

Electricity is transported through the electrical grid, which is a connected network. The electrical grid is the collection of high voltage (e.g. 380, 110 kV) transmission lines that transport electric-

<sup>1</sup>Officially Luxembourg is part of the CWE region, but Luxembourg is not considered as a separate market area. This is because Luxembourg has two separate transmission networks which are not interconnected and each are fully integrated within the networks of the two neighbouring countries Belgium and Germany.



ity and distribution (e.g. 50, 10 kV) lines that connect consumers. The electrical system is the collection of the electrical grid and electricity generators that produce electricity. The TSOs of a country maintain the high voltage transmission lines and they have to provide a reliable and uninterrupted supply of electricity on the high voltage grid. Every country has its own TSO. In the Netherlands the TSO is TenneT, but a country can have more than one TSO. For example in Germany, there the four TSOs are: Amprion GmbH, TenneT, 50Hertz Transmission GmbH and TransnetBW.

To provide consumers with low voltage (i.e. 230 V) electricity, electrical substations are used to transform the electricity from high to medium and low voltage. The medium and low voltage electricity is transported through distribution lines to consumers. The distribution lines do not cross borders of countries. High voltage lines can cross borders of countries, which makes collaboration between TSOs of different countries possible. The transport of electricity across borders is referred to as cross-border transport.

When a producer of electricity and a consumer of electricity trade, the amount of this traded electricity does not flow on a direct path from the generation source to the consumer, but it spreads out over the network, according to Kirchhoff's law, [20]. This implies that the consumer does not necessarily get that electricity produced by the producer of its trade, it also can be a different producer. Figure 1.3 illustrates the flow of electricity in a grid. In this figure, the red lines symbolise the shortest path from the producer to the consumer and the dotted arrows describe the paths of the *physical flow* (in this case we assume that every line has the same resistance). The physical flow through the network is therefore different than a flow along the

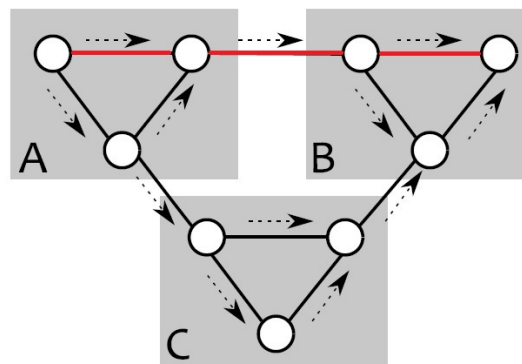


Figure 1.3: Difference between physical flow and commercial flow in an electrical network.

Source: [24].

Commercial  
Flow

shortest path, this shortest path is also called the *commercial flow*. If another flow in the network (for example, as consequence of another trade) is added on top of the already present flow, it has consequences for all the physical flows on the branches. The added flow also spreads out on all the available parallel paths, changing the flow that was already present. Consequently, flow on all branches might be influenced by trades in all market areas. An added flow in a certain market area can influence all flow present on the branches in this market area and also the flow present on the cross-border branches (and sometimes even flows on branches in other market areas). So the physical capacity of the cross-border branch is not directly the physical cross-border transport capacity between two market areas. As consequence of the physical behaviour of flows,



a trade within a market area can already influence the physical cross-border transport capacity. This makes calculating the cross-border transport capacity fairly complex. In the process of calculating the available cross-border capacity between two market areas, expected trades within those market areas and expected trades between other market areas should be taken into account.

The paragraphs above describe the crux of transporting power within a market area. When transporting power between market areas some more difficulties arise such as collaboration between TSOs, available capacity between market areas, communication between market parties of different market areas etc. In order to make this easier the North Western European countries decided to start working together within a market coupling. This market coupling is described in the following paragraphs.

### 1.1.4 North Western European Market Coupling

To optimise the use of the capacity of the cross-border transport branches between countries, the North Western European (NWE)<sup>2</sup> market coupling has been brought to life on the 4<sup>th</sup> of February 2014 by the TSOs of the NWE market areas. An extensive description of the NWE market coupling is given in [22] and [6]. A short summary is given here, starting with a description of the concept of market coupling.

#### 1.1.4.1 Advantages of Market Coupling

Market coupling is a frequently used method to integrate power markets in different areas, [1]. In this method cross-border capacity is implicitly made available to market participants without the need to explicitly purchase the corresponding transport capacity. When markets are coupled, trading between market areas becomes easier. If there is a market area with a high price, electricity can flow from a market with a low(er) price to the market with the high(er) price. In this case there is more electricity available for the high price market, so this price will decrease. In the low price market there will be more space for suppliers to sell their electricity, which will result in a somewhat higher price. In both cases some market participants will gain from the market coupling. Some individual market participants would gain less than in the uncoupled market, but the overall gain is higher than the sum of the individual losses. If electricity could be transported unlimited all the prices in the coupled market areas would be equal. However this is not always the case. The cross-border transport of electricity is limited by the capacities in the electrical network. The main goal of market coupling is price convergence and efficient allocation of electricity in the market areas. The cross-border capacity constrains the flow of electricity between market areas and consequently constrains the benefit that can be gained from the market coupling. If the cross-border capacity is large enough, electricity will flow from the market area with a high price to the market area with a low price, until the price in both market areas is equal. If the cross-border capacity is not large enough the prices cannot converge. An example of price convergence is shown in Figure 1.4. When the full amount of available transport capacity is used, we say that the cross-border cable is *congested*.

Congestion

For the NWE market coupling, a method is developed to calculate simultaneously the market

<sup>2</sup>The following countries are part of the NWE region: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.



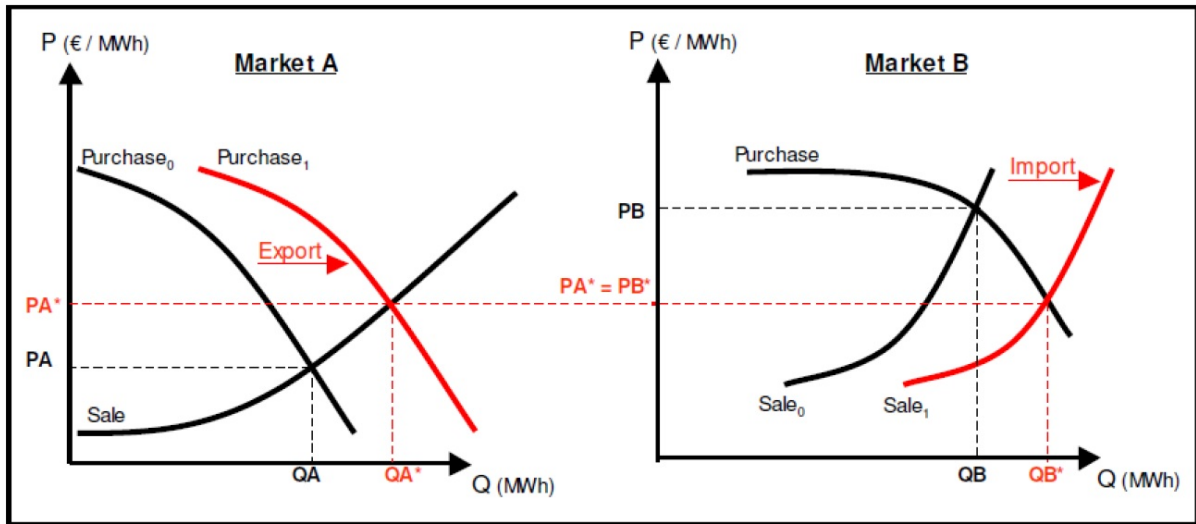


Figure 1.4: An example of price convergence through market coupling for a certain hour for two market areas.

Source: [3].

Net Position

clearing prices, *net positions* (difference between the matched demand volume and the matched supply volume) and flows of electricity in and between market areas in the NWE region. The market coupling method takes into account the transport capacity to ensure the flows of electricity between market areas do not exceed this capacity. Section 1.1.5 describes this method more extensively.

#### 1.1.4.2 Using the Flow-Based Model to Constrain the Transport Capacity

Available  
Transfer  
Capacity  
(ATC)

The transport capacity which is necessary for the market coupling method, is determined by the TSOs. For the market areas in the CWE region, this is done with the so-called flow-based model, for the remaining NWE countries this transport capacity is determined with the so-called *Available Transfer Capacity (ATC)* model. This section gives a superficial description of both methods. The flow-based model has been introduced and described in [3] and [5], and is explained in detail in Section 1.1.6.3. The project to implement the flow-based model for the CWE market areas went live on the 20<sup>th</sup> of May 2015.

Both the flow-based model and the ATC-model have as starting point the physical maximum capacity for each branch in the electrical grid. Also per branch it is calculated how much the flow on this branch is influenced when the amount of electricity transported between market areas is increased or decreased. When the capacities of all branches in the network are respected, the network will remain secure, which means that there will be no outages due to lack of capacity. The flow on each branch in the electrical grid is influenced differently when changing the amount of electricity transported between market areas. This influence in combination with the physical maximum capacity of a branch determines whether a certain change in transport of electricity between market areas is possible or not, this is represented by a (transport capacity) constraint. The constraints of the branches in a certain market area together represent a feasible area for the change in transport of electricity between this market area and other (adjacent) market areas.



This feasible domain is called the *security domain*.

Security Do-  
main

In the ATC-model only the cross-border branches are considered when determining the cross-border capacity to transport electricity between market areas. When two market areas have one or more cross-border branches, those branches are merged into one restriction on the available transfer capacity. Consequently this restriction is set to the capacity of one of the cross-border branches or to the capacity of an other internal branch, depending on which of those capacities is the lowest. Figure 1.5 shows the merged branches which determine the available transfer capacity. Per modelled cross-border branch the TSOs can restrict the import of electricity and export

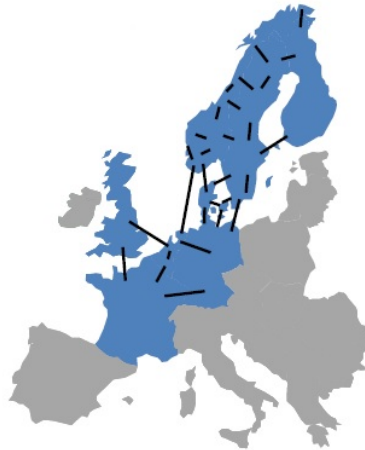


Figure 1.5: The branches between countries in part of Europe.

Source: [12].

of electricity, which results in a square within the security domain. Figure 1.6 contains a simplified example of a security domain. The area within the blue lines is the security domain and the red square denotes the ATC-restrictions. The example in Figure 1.6 contains three market areas,  $A$ ,  $B$  and  $C$ . A cross-border transport of electricity from  $A$  to  $B$  which is negative represents a cross-border transport in the opposite direction, thus from  $B$  to  $A$ . This results in the positive vertical axis to be the transport from market area  $A$  to market area  $C$ . The negative vertical axis represents the transport from  $C$  to  $A$ . Similarly, the positive horizontal axis represents the transport from  $A$  to  $B$  and the negative horizontal axis represents the transport from  $B$  to  $A$ . The transport directions  $B$  to  $C$  and  $C$  to  $B$  are not included, since the cross-border transport in those directions can be calculated when taking into account that the sum of all the net positions of the three market areas should be zero to maintain the balance in the electrical network. In the ATC-model every border has one restriction on the cross-border capacity for import and one for export. The TSOs want to find the largest square which is still contained in the security domain, because then they make as much capacity available to the market without jeopardising the stability of the electrical network. The ATC-model is simple and consequently conservative. In the flow-based model cross-border transport of electricity is allowed within the entire security domain. Within the flow-based model the security domain is called the *flow-based domain*. This expands the allowance of the transport of electricity between market areas, without decreasing the security.

Flow-Based  
Domain

With the NWE market coupling the market participants in Europe aim to trade electricity



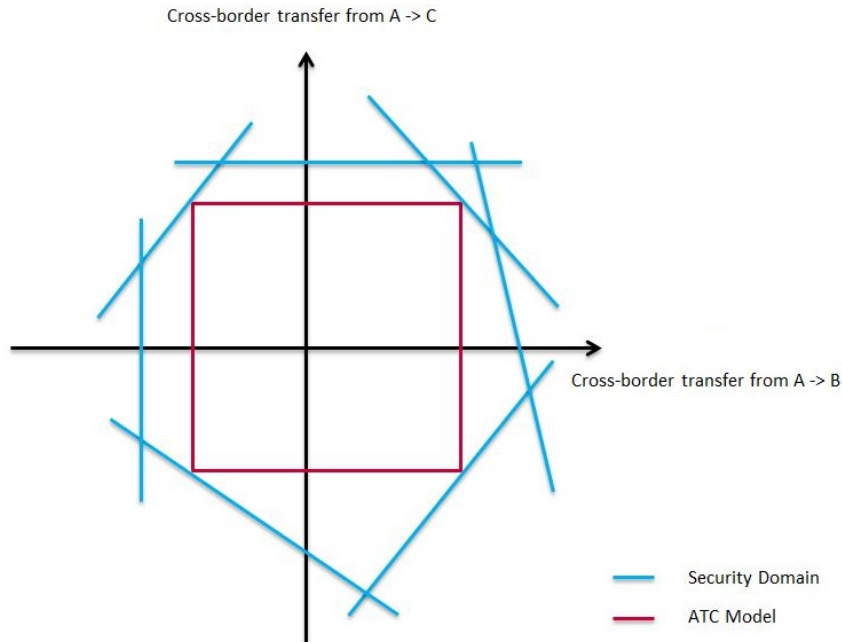


Figure 1.6: Visualisation of the security domain.

not only within their market area but also in other market areas. To make this possible cross-border branches are available. The paragraphs above give some more insight in the complexity of the transport of electricity between market areas and the calculation of the transport capacity. The market areas in the NWE region developed a method to make optimal use of the transport capacity. This method is described in the next section.

### 1.1.5 The Method of the NWE Market Coupling

Recall that in the NWE market coupling, a method is used to calculate simultaneously the market clearing prices, the net positions and the flows of electricity between market areas. This method has been developed especially for this market coupling and is described comprehensively in [22]. Also [24] gives a somewhat more mathematical description. The name of the method is Euphemia. The following paragraphs describes the method, starting with the objective, second the constraints are described and last the output is given.

#### 1.1.5.1 Objective Function of the Market Coupling Method

A goal of the NWE market coupling is an optimal allocation of electricity. When electricity is transported from a market area with a high price to a market area with a low price the TSO transporting the electricity profits from this. The profit made by the TSO is called the *congestion rent*. As example, if  $p_A$  denotes the price in the high price market  $A$ ,  $p_B$  the price in the low price market  $B$  and  $f_{B-A}$  the flow of electricity from the low price market  $B$  to the high price market  $A$ , the congestion rent is given by:

$$\text{congestion rent} = (p_A - p_B) \cdot f_{B-A}.$$

Congestion  
Rent



With the transport of electricity between market areas also producers are able to sell more electricity, but for a somewhat lower price. Producers have a minimum price they want for a certain amount of electricity. The difference between the minimum price the producer wants for its supply and the price it eventually receives for its supply (the market clearing price) multiplied by the volume is called: *producers surplus*. Also with the transport of electricity between market areas more consumers are able to buy electricity, but for a somewhat higher price. Consumers have a maximum price they are willing to pay for a certain amount of electricity. The difference between the maximum price the consumer is willing to pay and the price it eventually pays, also multiplied by the volume is called: *consumers surplus*. Figure 1.7 visualises the calculation of the consumers and producers surplus. Here the red and blue area describe the sum of the consumers and producers surplus. The sum of the congestion rent, the producers surplus and the consumers surplus is called: *social welfare*. In the market coupling method, the social welfare is maximised by matching demand volumes and supply volumes of producers and consumers. With the transport of electricity between market areas more consumers and producers can be coupled but their surplus is somewhat lower, the market coupling method optimises the social welfare, so in general more (or equal) surplus is gained than in the scenario without transport of electricity between market areas. The model that the algorithm, corresponding to the market coupling method, solves is an Integer Linear Program (ILP)<sup>3</sup>.

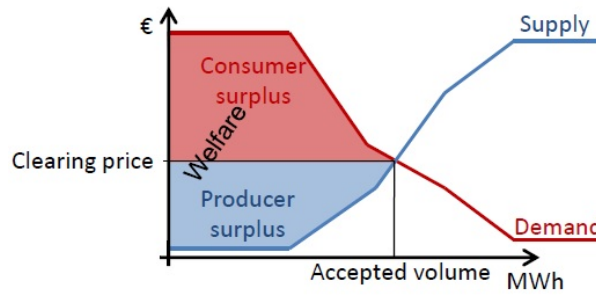
Producers  
SurplusConsumers  
SurplusSocial  
Welfare

Figure 1.7: Visualisation of consumers and producers surplus.

Source: [23].

For the market coupling method a few things are known in advance, for example the market areas of which the NWE region consists. Those market areas are named  $m$  and  $M$  is the set of all market areas in the NWE region. Per market area  $m$  the set of market orders is known, namely  $O_m$ . For each order  $o \in O_m$  it is known whether it is a supply- or demand order and what the corresponding price  $p_{o,m}$  (in €/MWh) and quantity  $q_{o,m}$  (in MWh) are. The price  $p_{o,m}$  is a maximum price in case of a demand order and to a minimum price in case of a supply order. The goal of the market coupling method is to maximise the social welfare. The way to do so is by deciding which order is (partly) accepted and which order is declined. The decision variables that correspond to this decision problem are the variables  $x_{o,m}$ , which give the fraction of acceptance of order  $o$  in market area  $m$ . For all orders  $o$  and all market areas  $m$  it holds that  $0 \leq x_{o,m} \leq 1$ . The volume  $q_{o,m}$  is considered negative for supply orders and positive for demand orders. The above results in the mathematical form of the simplified objective of the

<sup>3</sup>Due to interpolated orders, which we do not cover here, this program is actually quadratic. For simplicity this is not included, this is explained in Section 7.3.



social welfare maximisation problem of the market coupling method:

$$\text{maximise}_{x_{o,m}} \sum_{m \in M} \sum_{o \in O_m} x_{o,m} \cdot p_{o,m} \cdot q_{o,m},$$

### 1.1.5.2 Relevant Network Data for the Market Coupling Method

In order to respect the design of the electrical grid, some network data has to be contained in the market coupling method. This is done with the so-called network constraints. The method has to respect those constraints when finding a feasible solution to the problem. The description of the electrical grid is necessary to ensure that the security of the electrical grid is maintained in the found solution.

1. Market Areas: The set  $M$  is the set of all market areas in the NWE region. All the orders of a market area are collected at the power exchange of the corresponding market area, which combines them in one order curve and provides this curve as the set of market orders  $O_m$  to the market coupling method.
2. Network Representation: The network representation is given by the branches, generators, customers, etc. The set of branches is denoted by  $B$ .
3. Balance Constraints: The generation and consumption of electricity should always be in balance to prevent outages. To model this for each market area  $m$ , we introduce a variable  $NEP_m$  which represents the net position of market area  $m$ . This leads to the following constraints:

$$\sum_{o \in O_m} q_{o,m} \cdot x_{o,m} + NEP_m = 0 \quad \forall m \in M. \quad (1.1)$$

When the  $NEP_m$  is positive this means market area  $m$  is exporting electricity. When  $NEP_m$  is negative this means market area  $m$  is importing electricity. Note, that with this constraint the net position of each market area is calculated. To ensure that over all market areas we get a proper balance of generation and consumption, we have to add also the constraint:

$$\sum_{m \in M} NEP_m = 0.$$

Those constraints are denoted by the so-called balance constraints,  $\mathbf{C}_{balance}$ .

4. Transport Capacity Constraints: The transport capacity constraints represent the fact that electricity cannot be transported unlimited. The transport capacity constraints are generated by two models: the flow-based model and the ATC-model. The combination of those models is called the hybrid network model. The specific transport capacity constraints of both models are described in Section 1.1.6. The general description of the transport capacity constraints are as follows:

$$\begin{aligned} -\tilde{f}_b &\leq f_b \leq \tilde{f}_b & \forall b \in B, \\ f_b &= g(NEP_1, \dots, NEP_m) & \forall b \in B, \end{aligned}$$

where  $f_b$  is the flow on branch  $b$  and  $\tilde{f}_b$  is the maximum capacity of branch  $b$  available for the market, both in MW. Function  $g$  describes the dependency of the flow  $f_b$  on branch  $b$ , on the net positions  $NEP_m$  of the market areas  $m \in M$ . The net position  $NEP_m$  is



implied by the decision variables  $x_{o,m}$  and thereby function  $g$  links the decision variables to the flow on the branches.

5. Losses: When electricity is transported, a small part of the electricity is consumed by the branch. The amount of electricity that is consumed by the branch can be calculated in advance. For simplicity reasons we do not further use nor explain these losses.

### 1.1.5.3 Different Types of Orders

The orders of a market area  $m \in M$  are collected in the set  $O_m$ . There are different types of orders. Those orders generate extra constraints for the method.

1. Hourly Orders: Hourly orders are buy or sell orders. For every order the minimum or maximum price (depending on the type of the order, sell or buy) and the hour  $h$  during which the delivery of electricity takes place, is given. For those orders holds: orders with a supply at price lower than the MCP or demand at a price higher than the MCP (so-called orders *in-the-money*) are fully accepted. Orders with a supply at price higher than the MCP or demand at a price lower than the MCP (so-called orders *out-of-the-money*) are fully rejected and orders which have a price equal to the MCP (so-called orders *at-the-money*) can be partly accepted/rejected. This also follows from the objective, because the acceptance of an order out-of-the-money would produce negative surplus.
2. Block Orders: There are different types of block orders. For all those types it holds that they are buy or sell orders and have one single price. A block order can only be fully accepted or fully declined, so-called fill-or-kill.
  - (a) *Regular Block Orders*: A regular block order consists of a number of periods. The simplest case is a block order for a consecutive number of periods, with the same volume.
  - (b) *Profiled Block Orders*: The only difference between a regular block order and a profiled block order is that for a profiled block order the volume is allowed to differ for each time period.
  - (c) *Linked Block Orders*: The linked block order is also called the family block order. It consists of different block orders that are dependent on each other in a ‘parent-child’ relationship. The acceptance of a child block order is conditional to the acceptance of the parent block order. A parent block which is out-of-the-money can be accepted if the child blocks compensate for the loss of the parent block. A child block which is out-of-the-money cannot be accepted (unless it is also a parent block and the children generate enough surplus to compensate).
  - (d) *Exclusive Group Block Orders*: This order contains a set of block orders for which holds that only one of those block orders can be accepted.

in-the-  
money  
out-of-the-  
money  
at-the-  
money

Block orders play an important role in the market coupling method. Those orders complicate the model by turning the linear program into an integer linear program. Also block orders often consist of a large volume (in comparison with the hourly orders) because the block orders extend over several hours. Large volumes around the market clearing price can withhold the market clearing price from large jumps.



#### 1.1.5.4 Output Generated by the Algorithm Corresponding to the Market Coupling Method

As output the algorithm returns the following:

- The market clearing price per market area  $m$ ;
- the corresponding matched volumes, ' $q_{o,m} \cdot x_{o,m}$ ';
- the net position of each market area,  $NEP_m \forall m \in M$ ;
- the flows  $f_b$  through the branches  $b \in B$  in and between the market areas;
- the selection of orders that is accepted,  $x_{o,m} \forall o \in O_m \& \forall m \in M$ .

#### 1.1.6 Transport Capacity Constraints

Section 1.1.5.2 describes the different constraints that are relevant for the market coupling method. One set of those constraints are the transport capacity constraints, which play an important role in this research. The collection of the transport capacity constraint is denoted by  $C_{transport}$ . In this section those constraints are explained in more detail.

Calculating the transport capacity constraints can be done in different ways (see [8]). The method used in the non-CWE countries of the NWE region is the ATC-model. For the CWE region the flow-based model is used. An alternative for the calculation of the transport capacity constraints is the Nodal method, used in some parts of the United States of America. We first explain the nodal method in order to make the flow-based model and ATC-model easier to understand.

##### 1.1.6.1 Nodal Model to Generate Transport Capacity Constraints

In the nodal model the whole electrical grid is considered in detail. Every physical branch, generation unit and consumption point is taken into account as a node (generation units and consumption points) or edge (branches), leading to a set  $V$  of nodes and  $B$  of branches. This way of modelling is the most accurate way. Every node is considered as a different market area. This means that every order is addressed to a specific consumption point or generation unit. Consequently every node has its own market clearing price. The net position of a node  $v \in V$  is the amount of production or consumption at that node, represented as  $NEP_v$ . The net position  $NEP_v$  per node  $v \in V$  is implied by the decision variable  $x_{o,v}$ , which is the fraction of acceptance of order  $o \in O_v$  in node  $v \in V$ , hereby  $O_v$  is the set of all orders for node  $v$ . The calculation of the net position is done with Equation (1.1), as every node  $v \in V$  is a market area.

In the nodal model the capacity constraints are specified as follows:

$$\begin{aligned} -\bar{f}_b &\leq f_b \leq \bar{f}_b \quad \forall b \in B, \\ f_b &= \sum_{v \in V} PTDF_{b,v} \cdot NEP_v \quad \forall b \in B, \end{aligned}$$

where parameter  $PTDF_{b,v}$  is the *Power Transfer Distribution Factor* for node  $v$ , the  $PTDF_{b,v}$  denotes the influence of the change of 1 MW in the  $NEP_v$  of node  $v$  on the flow on branch  $b$ , variable  $f_b$  is the flow that is implied by the PTDF-values and the net positions. Note, that



because of the difference between the physical flow and commercial flow of electricity, a change in net position of market area  $v \in V$ , which is not incident to  $b$ , can also influence the flow on branch  $b$ .

This model of calculating transport capacity constraints is quite accurate. However, if we want to use this model for Europe it has a few problems, among which one is that the demands orders have to be given per node. Nonetheless, in Europe the orders are per market area and not per node. Another problem with this model is that it also allows different prices at different generation units or consumption points and not per country, which is undesirable in Europe.

### 1.1.6.2 Available Transfer Capacity Model to Generate Transport Capacity Constraints

In the ATC-model the electrical grid is simplified. A complete market area is represented as one node and all cross-border branches per two market areas adjacent to each other are modelled as one cross-border branch, so in this case the set  $B$  only contains one cross-border branch per two market areas adjacent to each other. For the ATC calculation the TSOs estimate the network conditions, generation and load patterns. With this estimation the TSOs simulate the flow on the branches in the network, which results in the security domain, from which the ATC restrictions can be calculated. Every cross-border branch  $b \in B$  has a restriction  $ATC_b$ , leading to the following transport capacity constraints:

$$\begin{aligned} ATC_b^{min} &\leq f_b \leq ATC_b^{max} \quad \forall b \in B, \\ \sum_{o \in O_m} q_{o,m} \cdot x_{o,m} &= \sum_{b \in B} f_b \cdot r_{b,m} \quad \forall m \in M, \end{aligned}$$

where  $ATC_b^{min}$  is the ATC-bound in the ‘export’ direction and  $ATC_b^{max}$  the ATC-bound in ‘import’ direction. The parameter  $r_{b,m}$  indicates whether a cross-border branch  $b$  is starting at a certain market area ( $r_{b,m} = 1$ ), ending there ( $r_{b,m} = -1$ ) or not connected to this market area ( $r_{b,m} = 0$ ). Note again, that the flow on branch  $b$  is implied by the decision variables  $x_{o,m}$ .

The non-CWE countries in the NWE region are currently using this model. The problem with this model is that the calculation of the ATC bounds is not very transparent and the model is not very accurate, this is because the simulation of the flows on the branches is based on an estimation of the network conditions, generation and load pattern. To assure that the security of the electrical grid is maintained, it turns out that in its application this model leads to a quite conservative model.

### 1.1.6.3 Flow-Based Model to Generate Transport Capacity Constraints

The flow-based model is a combination of the nodal model and the ATC-model. Every ‘critical’ transmission unit (which means all branches and generation units) is taken into account (the definition of critical is determined by the TSO of the corresponding market area). The following



transport capacity constraints are defined<sup>4</sup>:

$$-RAM_b \leq f_b \leq RAM_b \quad \forall b \in B, \quad (1.2)$$

$$f_b = \sum_{m \in M} PTDF_{b,m} \cdot NEP_m \quad \forall b \in B, \quad (1.3)$$

Remaining  
Available  
Margin  
(RAM)

where  $RAM_b$  is the *Remaining Available Margin*, which denotes how much capacity there is left on branch  $b$  for trades, in MW. The calculation of the net position is again done with Equation (1.1). The net position is implied by the decision variables  $x_{o,m}$  which links the decision variables to the flow on branch  $b$ . An extensive description of the calculation of the  $PTDF_{b,m}$ -values and  $RAM_b$ -values is given in Chapter 2. The PTDF-values and RAM-value together are called the *parameters within the flow-based model*.

Parameters  
Within the  
Flow-Based  
Model

The flow-based model is implemented for the CWE countries in the NWE region. This model is less conservative than the ATC-model and results in more use of the transport capacity. Also the flow-based model is more accurate and reduces the risk of overloading of the electrical network.

#### 1.1.6.4 Summary of the NWE Market Coupling Method

All aspects of the market coupling method are described in the previous sections. This section mathematically summarises the market coupling method, when using the flow-based transport capacity constraints and ignoring the ATC transport capacity constraints. The objective of the market coupling method is to maximise the social welfare. In the process of maximising the social welfare, the above formulated constraints are respected. In the given formulation, some decision variables are integer, such as the acceptance of the regular block orders, where it is only allowed to fill-or-kill these orders. The resulting mathematical method represents an ILP model.

As the computation time for solving an ILP may get very large, there are different stopping criteria for the algorithm corresponding to the market coupling method. A possible straightforward criterion is: stop if the optimal solution is found. Another stopping criterion is a time limit. If the algorithm stops at this time limit, it is possible that the algorithm did not find the optimal solution but the best feasible solution so far.

The summary below contains a simplified version of the ILP.

$$\begin{aligned} & \text{maximise} && \sum_{o \in O_m, m \in M} x_{o,m} \cdot p_{o,m} \cdot q_{o,m} \\ & \text{subject to} && \sum_{o \in O_m} q_{o,m} \cdot x_{o,m} + NEP_m = 0 && \forall m \in M, \\ & && f_b = \sum_{m \in M} PTDF_{b,m} \cdot NEP_m && \forall b \in B, \\ & && -RAM_b \leq f_b \leq RAM_b && \forall b \in B, \\ & && 0 \leq x_{o,m} \leq 1 && \forall m \in M, \quad \forall o \in O_m. \end{aligned}$$

In this formulation the constraints defined by the specific orders and some other constraints are not taken into account. This research focuses on the transport capacity constraints. In order to

<sup>4</sup>The flow-based transport capacity constraints are in the flow-based model of the form:  $f_b \leq RAM_b \quad \forall b \in B$ , if the other side of the equation is also necessary, it is implemented as an extra constraint.



see more directly what the influence is of the flow-based transport capacity constraints on the decision variable some constraints are re-written and combined.

$$-RAM_b \leq \sum_m PTDF_{b,m} \cdot \left( - \sum_{o \in O_m} q_{o,m} \cdot x_{o,m} \right) \leq RAM_b,$$

this reformulated equation shows that  $x_{o,m}$  is indirectly restricted by the  $PTDF_{b,m}$  values and  $RAM_b$ -value and underlines the importance of the accurate approximation of the transport capacity constraints. The flow-based transport capacity constraints represent a feasible region for the outcome of the net positions of the CWE countries, which is the flow-based domain as already introduced in Section 1.1.4.2.

Previous sections explain the power markets, the transport of electricity, the market coupling between the market areas in the NWE region and the different models for calculating transport capacity constraints. The next step is to define the problem researched in this thesis. The following section describes the approach in this research and verifies the structure of the report. Also here a superficially description of the solution approach is given.

## 1.2 Research Questions and Solution Approach

This section starts with a short summary of the essential parts of the background which are important for the problem researched in this thesis, this results in the motivation for this research and the relevance for this research for Eneco. The last sub-sections describe the research questions and the chosen solution approach, which provide the structure of this report.

### 1.2.1 Relevant Background to the Problem Researched in this Thesis

Section 1.1 described how the market coupling for the day-ahead power market in the NWE region works. Every day the market participants of the NWE region hand in their orders for delivery of electricity the next day. After that, the market coupling method calculates, among other things, the market clearing price. In this process the capacity of cross-border transport of electricity between the market areas in the NWE region is taken into account. Unfortunately the cross-border transport capacity is limited, which sometimes prevents the market clearing prices in different market areas from converging. The transport capacity constraints are calculated with the flow-based model (for the CWE part of the NWE region) and the ATC-model (for the remaining part of the NWE region).

TSOs have enough information and experience of the behaviour of their network to make a model which can simulate the actual flows in the electrical network. With those flows, the flow-based transport capacity constraints are calculated. Most of the market parties do not have this information nor experience, which makes it harder for them to estimate the behaviour of the flows in the electrical network. Besides that, the TSOs use forecasts to establish the production and consumption levels, which are necessary to simulate the actual flows. As this forecasting can be done in many different ways, it is hard to reproduce the forecasts the TSOs use. As a consequence, without the information about the forecasts, experience and the information about the network it is hard to reproduce the flow-based transport capacity constraints. This is where



the problem considered in this thesis arises. The market participants have an interest to approximate the flow-based transport capacity constraints with good reliability in order to properly place orders on the market.

When Eneco is able to manage their position in the electricity market smartly, this leads to higher profits. Forecasting the market clearing prices for the CWE market areas contributes to this. As a consequence, forecasting the market clearing price is important. To do this, knowledge on the input to the market coupling method is of crucial importance and should be forecast. Eneco is already able to forecast the set of orders in an accurate way. The balance constraints and market areas are also known. It only remains to approximate the transport capacity constraints and that is exactly where the problem researched in this thesis arises.

### 1.2.2 Research Questions

When we are able to approximate the transport capacity constraints, we can use them as input for the market coupling method and forecast the market clearing prices, which helps Eneco with making decision when trading electricity. In order to give some structure to the research in which we aim to approximate the transport capacity constraints we define some research questions.

The main research question is:

**How can we make a good approximation of the flow-based transport capacity constraints?**

To answer this main research question, a few sub-questions are stated. The answers to these sub-questions together answer the main research question.

1. What are the parameters within the flow-based model and how are they calculated?
2. What are the characteristics that influence the flow-based transport capacity constraints?
3. When is an approximation a ‘good’ approximation?
4. Can we find different strategies to approximate the flow-based transport capacity constraints?

### 1.2.3 Overview Solution Approach

The research is carried out as follows:

1. **Initial analysis to identify the characteristics of influence:** In order to be able to approximate the flow-based transport capacity constraints we need to find out which characteristics have influence on the constraints. Some characteristics have influence on the physical capacity of a specific branch. Other characteristics influence the trades that are done within and between the different market areas. Different methods to find dependence of characteristics are investigated, those methods are described in Section 4.5.



2. **Method of finding similar hours:** One way to approximate the flow-based transport capacity constraints is to look in the past for hours which are similar to the hour we want to approximate (similar for example in terms of renewable energy forecast or demand forecast) and use the flow-based transport capacity constraints of that hour as approximation. We evaluate different ways and criteria to look for one or more similar hours in the past.
3. **Probabilistic sample sets of flow-based transport capacity constraints:** In a next step, a more elaborated way to approximate the flow-based transport capacity constraints will be investigated. The idea is somewhat more sophisticated compared to the similar hours approach. A possible method can be to collect the transport capacity constraints of similar hours and assign a probability to each constraint. This probability can be dependent on how often this specific constraint appears in the set of all constraints of the similar hours. With these constraints and corresponding probabilities, multiple sample sets of flow-based transport capacity constraints can be created. All those sample sets are used as approximation of the flow-based transport capacity constraints. The market clearing prices generated with this approximation of constraints can be used as forecast of the market clearing price.

## 1.3 Background Literature of Energy Trade and Forecasting

In the past few years a lot is changed in the electricity markets, but forecasting the market clearing prices of different market areas is always been an important activity for the market participants. The following section gives an overview of the research which is already done about forecasting market clearing prices but also about the calculation of the transport capacity constraints with the flow-based model.

### 1.3.1 Foundation of the Flow-Based Model

In Europe, many countries started working together in politics through the European Union and the electricity markets cannot be left behind. In order to ensure a secure, affordable, competitive and sustainable flow of electricity for European citizens and enterprises, a transformation in the power market is ongoing. The European Union has an energy union which aims to coordinate, support and stimulate this transformation, [10]. A first step to unite the European power markets went live on 9 November 2010 and was a market coupling between the CWE countries. Another important step to integrate the power markets was the NWE market coupling, which went live on 4 February 2014, [6], [7]. To guarantee the transparency of the collaboration of the NWE countries, a lot of documentation is provided about the market coupling method. The method of the NWE market coupling, is extensively described in [22] and more brief in [23]. Before the CWE flow-based model was introduced, all the countries used the ATC-model to calculate the available cross-border capacity constraints. The most recent step to further integrate the European power market is the implementation of the flow-based market coupling of the CWE countries on 20 May 2015, [4]. If this flow-based market coupling turns out to be



successful, it is very likely that it will be implemented in the other countries of the NWE as well.

### 1.3.2 Try-Out of the Flow-Based Model

During two years (2013 & 2014), the CWE countries monitored the behaviour of the flow-based model by means of a parallel run. Likewise to the NWE documentation, in order to ensure transparency, also a lot of documentation is available about the CWE flow-based model. For example all the results of the parallel run are available on [11]. The working of the model is described in [5] and [3]. In [8] and [24], a more extensive explanation is given and a mathematical description is provided.

### 1.3.3 Remaining Research about the Flow-Based Model

Besides the formal description and the extra explanation provided by [8] and [24], very little documentation, research nor (recent) publications are available about the flow-based model. During the time of development of the flow-based model a few parties researched the effect of flow-based market coupling. The writers of [2] researched the impact of the flow-based model compared to the ATC-model. Their conclusion is that the flow-based model performs better than the ATC-model in terms of available capacity to the market and consequently in terms of social welfare and price convergence. In [26] a similar research is done, however this paper focuses more on the practical implementation and their consequences. A result from [26] is that the flow-based model generates a more efficient use of the network which leads to lower generation cost and higher social welfare. However, in some situations sensitive monitoring is required. In [17] the influence of different factors on the market clearing price and social welfare is researched. The result of this research is that smaller market areas lead to less uncertainty and that various factors lead to considerable impact on the model.

### 1.3.4 Forecasting Market Clearing Prices in the Electricity Branch

Section 1.2 describes the problem researched in this thesis. We aim to forecast the market clearing prices by using an approximation of the flow-based transport capacity constraints. A lot of research is done about forecasting market clearing prices. For whoever is new to forecasting, [15] provides a good base for starters. In this book the principles of forecasting are explained. In [29] an overview is given about the methods which are developed in the last 15 years to forecast market clearing prices. In this review article the used methods to forecast market clearing prices are superficially explained and the strengths and weaknesses of the different procedures are described. Many methods aim to forecast the day-ahead market clearing price directly, while the focus in this research is approximating an input to the market coupling method, in order to forecast the market clearing price. Another angle of forecasting the market clearing price is brought up in [29], which forecasts the market clearing price based on the forecasting of the load, [28].



### 1.3.5 Rise of Probabilistic Forecasting Methods

An ongoing trend in the electricity price forecasting is that probabilistic forecasting starts playing a larger role, [25]. This is because probabilistic forecasts do not only forecast the market clearing price but they also provide a certain risk to this price. This trend indicates that probabilistic forecasting can be a good approach. Also [29] denotes that the similar hour approach is often used as benchmark for forecasting the day-ahead market clearing price, [28], [21]. In order to qualify whether our approximation is a good approximation we have to give a quantification for good. In [19] a proposal is done for qualification of forecasting.



## Chapter 2

# The Parameters within the Flow-Based Model

This chapter describes the parameters within the flow-based model and how TSOs calculate them, this answers the first sub-question: What are the parameters within the flow-based model and how are they calculated? In [3], it is described how TSOs calculate the parameters within the flow-based model and in [8], an alternative description of the calculation of these parameters is given. This chapter starts with a more precise definition of the branches, which is necessary for further explanation of the calculation of the parameters within the flow-based model. The sections after that describe the tools that the TSOs. The information in this chapter motivates certain choices made for the methods to solve the problem researched in this thesis.

### 2.1 Branches in Different Scenarios

The TSOs use an extended definition of branches, which is described in this section. This definition explains the calculation of the flow-based transport capacity constraints by the TSOs more clearly. In a normal scenario all branches are functional, certain power plants are producing and certain consumers are consuming. However, if one of the branches in the electrical grid is not functioning or under maintenance, the flow of electricity on many other branches is influenced compared to the normal scenario. Also, if a power plant is under maintenance or is producing less than in the normal scenario, this changes the flow of electricity compared to the normal scenario. As the TSOs have to maintain the stability of the network and want to prevent blackouts, also in case of an outage, they not only look at the normal scenario but also at the scenario with an outage. The scenario in which exactly one branch is not functioning is called a ' $|B| - 1$ '-scenario, where  $|B|$  is the number of branches in the network. Thus, there are exactly  $|B|$  different  $|B| - 1$ -scenarios (every branch  $b$  can be nonfunctional). Also  $|B| - 2$ ,  $|B| - 3$ , etc. scenarios can be considered, those are respectively the scenarios where two or three branches are nonfunctional. The TSOs have the right to chose the scenarios they take into account, based on their experience. Usually scenarios  $|B|$ ,  $|B| - 1$ ,  $|B| - 2$  and some other scenarios depending on the risk policy of a specific TSO are included in the set  $N$  of considered scenarios.



In order to calculate the parameters within the flow-based model, the TSOs calculate the influence of change in net position of a market area on the flow on all branches. This influence differs for every outage scenario. In order to explain this, Figure 2.1 gives a simplified example of a market area with two power plants, one consumer and different branches. Assume that in a normal scenario all power plants are working, the consumer

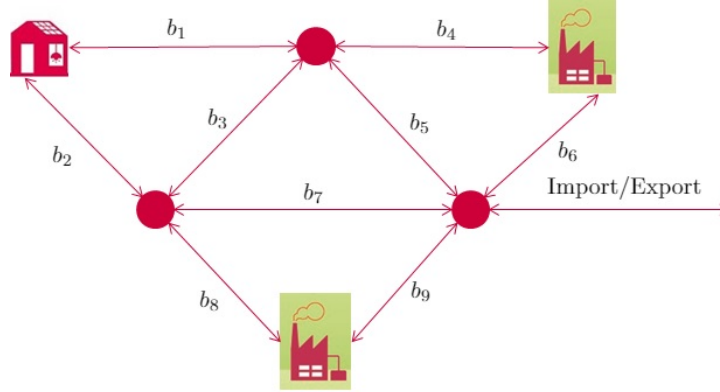


Figure 2.1: A simplified electrical network.

is consuming, all branches are functional and there is no import nor export. If branch  $b_1$  is nonfunctional and the consumer is still consuming the same amount of electricity, the flow on branch  $b_2$  increases. Furthermore, in case of a change in the import or export, the influence on the flow on branch  $b_2$  is different in the scenario without branch  $b_1$  than in the scenario with branch  $b_1$ . This is why the TSOs not only consider a scenario where all branches are functional but also different scenarios where one or more branches are nonfunctional.

Let  $b(n)$  denote the branch  $b \in B$  in scenario  $n \in N$ . All branches  $b \in B$  that are available in scenario  $n \in N$  are collected in the set  $B_n$ . For example if  $n$  is the normal scenario, all branches belong to the set  $B_n$ . The total set of branches (in combination with a scenario) that are considered by the TSOs is:  $B^{total} = B_1 \cup B_2 \cup \dots \cup B_{|N|}$ . Note that the set  $B^{total}$  contains several copies of the same branch  $b \in B$ , only considered in different scenarios.

For every  $b(n) \in B^{total}$  the TSOs calculate how the flow on this branch changes in scenario  $n$ , when the net position of a CWE market area changes by 1 MW. This is because the TSOs want to know the dependence of the flow on branch  $b(n)$  as a function of the net position of the market areas  $m$ . It is assumed that this dependence is linear and the TSOs estimate the coefficient, which is called the  $PTDF_{b(n),m}$ -value for branch  $b(n)$  and market area  $m$ . For example, if the flow on branch  $b(n)$  changes 0.1 MW when changing the net position of market area  $m$  by 1 MW in scenario  $n$ , this results in  $PTDF_{b(n),m} = 0.1$ .



### 2.1.0.1 The Definition of Critical Branches

The entire electrical network consists of many branches, which are all influenced by change in net position of a market area. However, this influence differs and branches which are little affected by the change in net position are not really relevant as they (almost) never constrain the net position in the price calculation of the NWE region. Consequently, the TSOs do not consider those branches as ‘critical’. Only the branches that are critical are used as constraints for the NWE market coupling method. The procedure for determining whether branches are critical or not, is done as follows: a branch  $b(n)$  is called critical if the change in net position of a CWE market area ‘significantly’ impacts the branch  $b(n)$ . To determine if a certain branch  $b(n) \in B^{total}$  is critical the concept ‘sensitivity’ is defined.

$$Sen(b(n)) = \max\{PTDF_{b(n),m_1}, \dots, PTDF_{b(n),m_{|M|}}\} - \min\{PTDF_{b(n),m_1}, \dots, PTDF_{b(n),m_{|M|}}\}.$$

Recall that the  $PTDF$ -values denote how much the flow on branch  $b(n)$  is impacted when changing the  $NEP_m$  of market area  $m$  with 1 MW. The sensitivity is the difference between the maximum and minimum impact on the flow on branch  $b(n)$ , which gives an indication on how sensitive the flow on branch  $b(n)$  is for changes in the net positions of the market areas. When a branch  $b(n)$  is very sensitive for changes in the net positions of the market areas this branch has a high risk of being overloaded and the change in net positions must be constraint by the flow-based transport capacity constraint corresponding to this branch  $b(n)$ . The TSOs have defined the following rule to decide whether a branch in a certain scenario is significantly impacted by the change in net position of a CWE market area or not. If  $Sen(b(n))$  exceeds the threshold of 5%, then branch  $b(n)$  is called critical and its corresponding constraint (as defined in Equation (1.2) and Equation (1.3)) is added as flow-based transport capacity constraint in the set  $\mathbf{C}_{transport}$ . This threshold is in constant observation to make sure the threshold is still up-to-date. The TSOs have the right to define a branch  $b(n)$  as critical even if it does not reach the threshold. The branches  $b(n)$  that are critical are collected in set  $B^{crit} \subseteq B^{total}$ .

### 2.1.0.2 Translation of Transport Capacity into Linear Constraints

Constraints given as flow-based transport capacity constraints and the parameters of these constraints are input for the NWE market coupling method. The NWE market coupling method takes into account these constraints when calculating the acceptance of orders which imply the market clearing price. The linear constraints are mathematically described in the form  $\mathbf{Ax} \leq \mathbf{b}$ . In our case the matrix  $\mathbf{A}$  is given by the  $PTDF$ -matrix  $\mathbf{PTDF}_{h,d}^{(|B^{crit}| \times |M|)}$ , for a specific hour  $h$  on day  $d$ . The number of market areas involved is  $|M|$ , in our case 4. The  $\mathbf{b}$ -vector for the same hour on the same day, is the  $\mathbf{RAM}$ -vector  $\mathbf{RAM}_{h,d}^{(1 \times |B^{crit}|)}$ . This leads to the following system:

$$\begin{bmatrix} PTDF_{1,1} & \dots & PTDF_{1,|M|} \\ \vdots & & \vdots \\ & \ddots & \\ \vdots & & \vdots \\ PTDF_{|B^{crit}|,1} & \dots & PTDF_{|B^{crit}|,|M|} \end{bmatrix} \times \begin{bmatrix} NEP_1 \\ \vdots \\ NEP_{|M|} \end{bmatrix} \leq \begin{bmatrix} RAM_1 \\ \vdots \\ RAM_{|B^{crit}|} \end{bmatrix},$$



this set of constraints defines a feasible domain for the outcome of the net positions in the optimal solution for the CWE countries. Solving the ILP leads to the optimal values of  $x_{o,m}$  for which the social welfare is maximal and for which the implied net positions of the market areas are within the boundaries of the domain, as defined by the above system. Recall that this domain is called the flow-based domain.

## 2.2 Estimation of Flows in the Electrical Network

As already stated, the flow-based domain represented by the flow-based transport capacity constraints is an input to the NWE market coupling algorithm. The TSOs calculate the flow-based transport capacity constraints by first changing the net position of each market area  $m \in M$  with 1 MW and calculating the influence this change has on the flow on branch  $b(n)$  (which results in the  $PTDF_{b(n),m}$ -value). When the TSOs want to calculate this influence, they have to know in advance how much flow is already present on branch  $b(n)$ . This flow is called the reference flow  $f_{b(n)}^{ref}$  and can be a flow as a consequence of other (internal) trades. The reference flow is not known at the time of running of the algorithm corresponding to the market coupling method (which takes place one day ahead of delivery). This is why the TSOs need to estimate the reference flow. The reference flow is the results of which power plants are generating, which renewable energy sources are generating and which consumers are consuming. Thus first the TSOs have to estimate these things in order to be able to calculate the reference flow. The collection of this estimated information is called the *base case*. The base case is specific for a specific hour on a specific day. The reference flow on branch  $b(n)$  and the physical capacity of branch  $b(n)$  together are used to calculate the RAM-value (see Section 2.2). Each TSO constructs a base case for its own market area, also each TSO has its own assumptions and methods to construct the base case. Finally the base cases of the different market areas are merged into one base case.

Base Case

At the time of construction of the base case, very little information about the possible flows is known. Also some information about maintenance is unknown. Likewise it is unknown which power plants will be active, because this is an outcome of the market coupling method. On the other hand, a few characteristics of a certain hour can be forecast quite well, such as consumption and renewable energy generation, as those are not very price dependent. The base case consists of the following information:

- An estimation of the net position for each market area (estimated on the basis of a reference day);
- the grid outages (as planned);
- a consumption forecast (demand for electricity from consumers);
- a renewable energy generation forecast;
- the generator outages (as planned);
- a production forecast of generators (dependent of outage planning, forecast consumption and net position).

For some of these estimations the TSOs use a reference day (e.g. for the estimation of the net position). This reference day is in general for Tuesday through Friday the day before.



For Monday, the Friday before. For Saturday, the previous Saturday and for Sunday, the previous Sunday. The TSOs use those reference days as a starting point. From this starting point they can, based on experience, adjust the flows, net positions and other aspects of the electrical network according to information about forecasts of generators, renewable energy forecasts, planned outages, etc.

The next paragraphs describe a few additional tools the TSOs use to calculate the parameters within the flow-based model and tools to adjust the flow-based domain such that the feasible net positions guarantee an electrical network with (almost) no chance of failure through overloading of branches.

### Generation Shift Key (GSK)

To determine the PTDF-values the TSOs need a so-called Generation Shift Key (GSK). The generation shift key is defined as the change in generation output of a specific generation unit when the net position of a market is increased with 1 MW. In practice, the generation shift keys define how the change of the net position of a market is mapped to the change in generation of the available generators in the concerning market area. Each TSO has its own way of forecasting the generation shift keys.

### Remedial Action (RA)

The problem with using the base case to calculate the parameters within the flow-based model is, that the base case is mostly based on estimations and forecasts. If these estimations or forecasts are too optimistic this can result in an overload of a branch when the net positions of the market areas are within the feasible area. In this case a Remedial Action (RA) can be performed to make more capacity available on the branch that would otherwise overload. The remedial actions are implemented in the model via a flow-based transport capacity constraint and are of influence on a specific branch  $b(n)$ . A few examples of remedial actions are: changing the tap position of a phase shifter transformer, opening or closing of a branch or transformer, changing the output of some generators or demand. The remedial actions that cannot be modelled in the PTDF-matrix are included in the Final Adjustment Value ( $FAV_{b(n)}$ ) (explained in Appendix A.2).

### External Constraint (EC)

Another problem with the calculation of the flow-based transport capacity constraints, is that the constraints do not prevent flows that are too far from the reference flow. Those flows can cause stability problems to the grid. A way to solve this problem is by the introduction of the so-called External Constraints (ECs). The external constraints are added to the set of flow-based transport capacity constraint.



### Remaining Available Margin (RAM)

For the calculation of the RAM-value also the base case is used, or more specific the reference flow implied by the base case is used. The formula to calculate the RAM-value is:

$$RAM_{b(n)} = f_{b(n)}^{Max} - f_{b(n)}^{ref} - FRM_{b(n)} - FAV_{b(n)},$$

where  $f_{b(n)}^{Max}$  is the maximum allowable flow on branch  $b(n)$ ,  $FRM_{b(n)}$  is the flow reliability margin (explained in Appendix A.2) and  $FAV_{b(n)}$  is the final adjustment value (explained in Appendix A.2).

The previous sections described the definition of branches the TSOs use and the construction of the base case, from which the TSOs can calculate the reference flow on branch  $b(n)$ . After the calculation of the reference flow, the TSOs calculate the *PTDF*-values and *RAM*-value for branch  $b(n)$  which are the parameters of the flow-based transport capacity constraint of branch  $b(n)$ . This calculation of the flow-based transport capacity constraints by the TSOs does not immediately provide a way to approximate the flow-based transport capacity constraints, but it provides us with more insight in the calculation of the constraints which we can use to make a smart approximation of the flow-based transport capacity constraints.



## Chapter 3

# Mathematical Problem Description

In the previous chapters the background of the problem, the research description and some insight in the calculation process of the parameters within the flow-based model are given. In order to fully understand the logic behind the possible methods to solve the problem, it is necessary that the problem is mathematically clear. This chapter gives a formal description of the problem which is researched in this thesis.

### 3.1 Input Characteristics for the Flow-Based Domain Generation Process

First we present a mathematical description of the characteristics that are used as input to generate the flow-based transport capacity constraints. Every hour of every day has certain characteristics. Those characteristics are given<sup>1</sup> for each hour  $h$  and denoted by  $Ch_1^h, \dots, Ch_k^h$ , where  $k$  is the number of characteristics. This means that every hour  $h$  has its own characteristics-vector  $\mathbf{Ch}^h = (Ch_1^h, \dots, Ch_k^h)$ . Every characteristic  $Ch_i^h$  lies in a domain, specific for that characteristic,

$$Ch_i^h \in \mathcal{D}_i \subseteq \mathbb{R}^+.$$

With this characteristics vector (and some other characteristics which are unknown to Eneco) the TSOs calculate the flow-based transport capacity constraints which represent the flow-based domain. The main problem is that, it is not known how the TSOs do this exactly and that we do not know exactly of which characteristics  $\mathbf{Ch}^h$  exists.

### 3.2 Function Used by the TSOs to Generate the Flow-Based Domain

The TSOs calculate the flow-based transport capacity constraints which represent the flow-based domain, using a certain internal procedure. Hereby, every transport capacity

---

<sup>1</sup>Most of the characteristics are not exactly known, but can be forecast quite good. In this case ‘given’ means that the forecast value is given.



constraint  $b(n)$  is of the following form:

$$C_{b(n)} = PTDF_{b(n),m_1} \cdot NEP_{m_1} + PTDF_{b(n),m_2} \cdot NEP_{m_2} + \dots + PTDF_{b(n),m_{|M|}} \cdot NEP_{m_{|M|}} \leq RAM_{b(n)},$$

where the  $PTDF_{b(n),m}$ -values and  $RAM_{b(n)}$ -value are the parameters within the flow-based model and  $NEP_m$  is the net position of market area  $m$ . Branch  $b(n)$  with corresponding constraint  $C_{b(n)}$  can be critical or not critical. The TSOs pursue a certain threshold in order to decide whether branch  $b(n)$  was critical or not (see Section 2.1.0.1). The precise procedure the TSOs use to do this, is unknown to Eneco. The information we have, is which branches were critical in the past for specific characteristic vectors  $\mathbf{Ch}^{h_1}, \dots, \mathbf{Ch}^{h_\gamma}$ . More precisely, we know the result of the function:

$$g_{tso} : \mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_k \rightarrow \{0, 1\}^t,$$

for a given set of characteristic vectors  $\{\mathbf{Ch}^{h_1}, \dots, \mathbf{Ch}^{h_\gamma}\}$ , where the  $b(n)^{th}$  component of  $g_{tso}(\mathbf{Ch}^j)$  is 1 if branch  $b(n)$  is critical with corresponding constraint  $C_{b(n)}$  for characteristics vector  $\mathbf{Ch}^j$ . Besides that,  $t$  is the number of different branches that appear in the data-set and  $\gamma$  denotes the number of different hours in the data-set.

The result of the function  $(g_{tso}(\mathbf{Ch}^h))_{b(n)}$  denotes only whether branch  $b(n)$  is critical or not, but when branch  $b(n)$  is critical, the parameters of the constraint  $C_{b(n)}$  can still vary. This means that every constraint corresponding to a critical branch  $b(n)$  can have different parameters ( $PTDF_{b(n),m}$ -values and  $RAM_{b(n)}$ -value). This can be a problem when approximating the outcome of function  $g_{tso}$  for a certain hour  $h$  in the future, because we do not only have to approximate whether a certain branch  $b(n)$  is critical or not for this hour  $h$  in the future, but we also have to approximate the parameters for which constraint  $C_{b(n)}$  is realised.

Another way to look at the problem is not by only looking at specific constraints but looking at the set of all constraints for a specific hour  $h$ . More specific, looking at the flow-based domain which is represented by the constraints that are critical in hour  $h$ . The motivation behind this view is that it is known that there are multiple ways to describe a similar flow-based domain with different constraints. However, for Eneco, only the feasible domain is important to forecast the market clearing price. In this case we consider a formulation of the following form:

$$\tilde{g}_{tso} : \mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_k \rightarrow \mathcal{P}_j,$$

where  $\mathcal{P}_j$  is the polytope represented by the set of constraints  $S_j = \{C_{b(n)} | (g_{tso}(\mathbf{Ch}^j))_{b(n)} = 1\}$ .

### 3.3 Mathematical Description of the Problem

As mentioned in the previous sections, the main problem is that both  $g_{tso}$  or  $\tilde{g}_{tso}$  are unknown and that only the outcome of the functions  $g_{tso}(\mathbf{Ch}^h)$  and  $\tilde{g}_{tso}(\mathbf{Ch}^h)$  for a certain set of hours in the past and the characteristics vectors  $\mathbf{Ch}^{h_1}, \mathbf{Ch}^{h_2}, \dots, \mathbf{Ch}^{h_\gamma}$  corresponding to these hours in the past are known.



The goal of this research is to approximate the function  $g_{tso}(\mathbf{Ch}^h)$  or  $\tilde{g}_{tso}(\mathbf{Ch}^h)$ , using the information that is available from the past. In principle we have the following two different methods, which are described in more detail in Chapter 5.

- Option one:** Find an approximation method which can generate, for a given hour  $h$  (with corresponding characteristics vector  $\mathbf{Ch}^h$ ), the outcome of  $g_{tso}(\mathbf{Ch}^h)$ , i.e. find the constraints  $C_1, \dots, C_j$  which are critical together with the parameters to suit these constraints.
- Option two:** Find an approximation method which can generate, for a given hour  $h$  (with corresponding characteristics vector  $\mathbf{Ch}^h$ ), the outcome of  $\tilde{g}_{tso}(\mathbf{Ch}^h)$ , i.e. the polytope  $\mathcal{P}_h$  which is the flow-based domain for the situation characterised by  $\mathbf{Ch}^h$ .

This approximated set of constraints  $C_1, \dots, C_j$  or the approximated polytope  $\mathcal{P}_h$  can be used by Eneco to forecast the market clearing prices per market area of the day-ahead market.



## Chapter 4

# Characteristics of Influence on the Flow-Based Domain

As mentioned, one problem which arises when researching the main problem is that we do not know exactly which characteristics are input for the functions  $g_{tso}$  and  $\tilde{g}_{tso}$ . Based on the content in Chapter 2, a few likely candidates are selected. This chapter describes the methods that are used to find out which characteristics are of influence on the flow-based domain. More precise: it describes the methods to find the domain  $\mathcal{D}$ , which is the input to the functions of the function  $g_{tso}$  and  $\tilde{g}_{tso}$ . Furthermore, this chapter contains the results of the developed methods to find the domain  $\mathcal{D}$ . Section 4.4 gives the definition of equality, which defines when two (or more) characteristics vectors are called equal.

### 4.1 Comparing Flow-Based Domains

To find out which set of characteristics allows us to approximate the flow-based domain, we aim to find out which set of characteristics generates the most ‘comparable’ domain for a certain reference hour. For this we need a definition of ‘comparable’, in sense of comparing two (or more) flow-based domains.

We first start to clarify what our goal is for this comparison. For Eneco it is important that the electricity market analysts are able to forecast the market clearing prices in the CWE region. In the NWE market coupling method the flow-based domain restrains the outcome of the net positions, which in turn determine the possible trades, which eventually can have influence on the market clearing prices. The main influence on the market clearing prices are the orders, so when we compare the forecast market clearing price with the actual market clearing price the orders of the day for which we make the approximation, can have an influence on how large this difference can be.

The objective of the market coupling method is to maximise the social welfare, sometimes this social welfare becomes higher when transporting electricity from one market to another. This transport between market areas can influence the market clearing prices in these market areas. In some cases no electricity is transported between market areas in the



optimal solution. For this specific case the transport capacity constraints are unnecessary. When the transport capacity constraints are unnecessary, it does not matter how good our approximation of the transport capacity constraints is, because in terms of forecasting the market clearing prices the forecast is still good, even when our approximation of the transport capacity constraints is very bad. The consequence is that we cannot draw a conclusion about the accurateness of the approximation of the flow-based domain solely based on the outcome of the market clearing prices. So finding a method to determine whether two domains are ‘comparable’ or not, is twofold: First we want to be able to say something about the accurateness of the approximation of the flow-based domain when the transport capacity constraints are unnecessary and second we want to be able to find out which set of characteristics implies the most ‘comparable’ domain.

#### 4.1.1 Comparing Polytopes

One possible method to qualify an approximation of a domain is to calculate the difference of the approximated and actual domain. For this, a distance function for domains (in our case polytopes) is needed. Possible distance functions may rely on volume, size, mean distance to the corner points, number of faces, etc. This section describes multiple ways to compare polytopes in context of our problem and the pros and cons of these comparison methods.

##### 4.1.1.1 Mean Distance to Corner Point Method

The first method to compare polytopes is calculating the mean distance from the origin to all the corner points of the polytope. This results in a single value for a polytope. Denoted by  $Dis(\mathcal{P})$ .

$$Dis(\mathcal{P}) = \frac{\sum_{cp \in CP} |cp|}{|CP|},$$

where  $CP$  is the set of corner points of polytope  $\mathcal{P}$

An advantage of this comparison method is that it is an easy method and results in a single value for each polytope. This makes it easy to compare the outcome of this distance function for two polytopes. A disadvantage of this method is that this method is not very precise. Figure 4.1 shows two similar polytopes (a reference polytope  $\mathcal{P}_{ref}$  and a candidate polytope  $\mathcal{P}_{can}$ ) which have a very different  $Dis(\mathcal{P})$ . Also Figure 4.2 shows two polytopes which have the same  $Dis(\mathcal{P})$ , but are very different and can give a very different outcome in net positions. As both situations are not desirable, this comparison method is not used in this research.

##### 4.1.1.2 Common Volume Method

Another method to compare two polytopes is the common volume comparison method, which uses the volume of the polytopes and calculates the amount of volume two polytopes



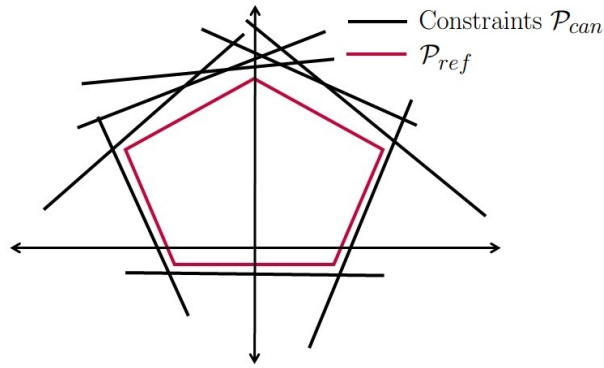


Figure 4.1: Two similar polytopes which have a very different  $Dis(\mathcal{P})$ .

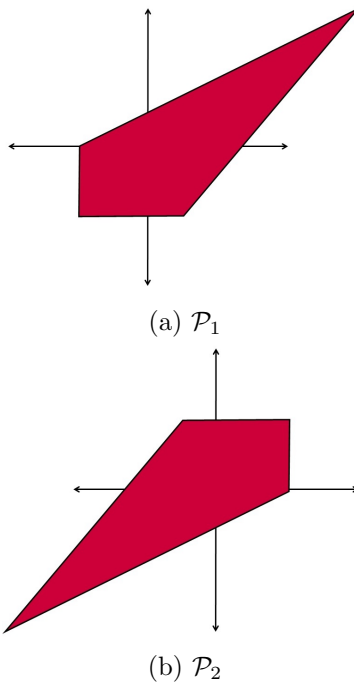


Figure 4.2: Two different polytopes which generate equal  $Dis(\mathcal{P})$ .

have in common, relative to the size of the candidate polytope. In order to calculate the volume both polytopes have in common, both constrain sets are combined. Figure 4.3 visually illustrates this method.

The advantage of this method is that it gives a good indication of the similarity of the polytopes. A disadvantage of this method is that the calculation of the percentage is in favour of small polytopes. If a candidate polytope is entirely contained in the reference polytope, the similarity percentage has an outcome of 100%, which suggest a very similar domain, Figure 4.4 gives an example of this case. As this method of calculating the percentage can be quite distorted, this comparison method is not used in this research.



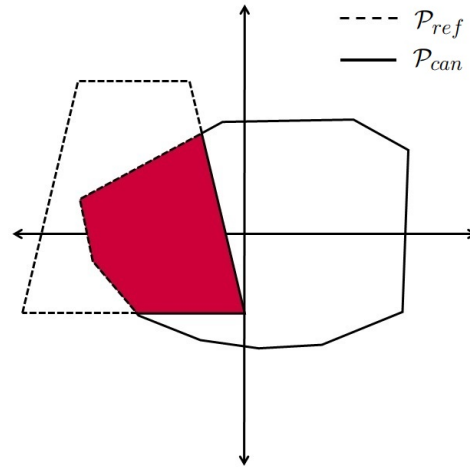
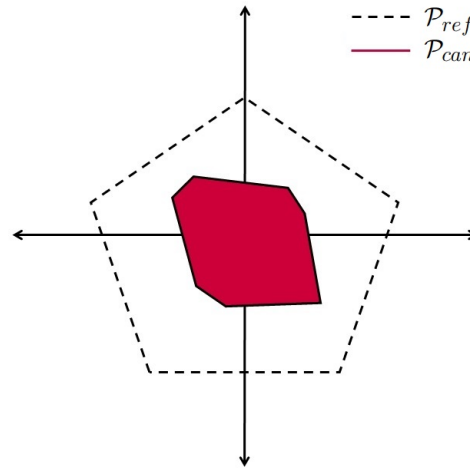


Figure 4.3: An example of the Compare Common Volume Method.

Figure 4.4: An example of a 100% match for  $\mathcal{P}_{can}$ .

#### 4.1.1.3 Separate Volume Method

Another method to compare polytopes is the separate volume comparison method, where not the common volume is used, but the volume two polytopes do not have in common is used, relative to the size of both polytopes ‘combined’ (combined in sense of  $\mathcal{P}_{ref} \cup \mathcal{P}_{can}$ ). Figure 4.5 illustrates this method. This method has the same advantages as the previous method. However, this method is not favourable for small polytopes.

One thing we should not let out of our sight is the problem. When calculating the common or separate volume we do not look at the outcome of the net positions in the problem. This might result in a distorted outcome. For example, if we have a polytope which does not look like the reference polytope at all, but it has the right constraint which in the outcome of the market coupling method constraints the net positions, the market clearing prices can be quite similar. So leaving out the market clearing prices in our validation process is not desirable.



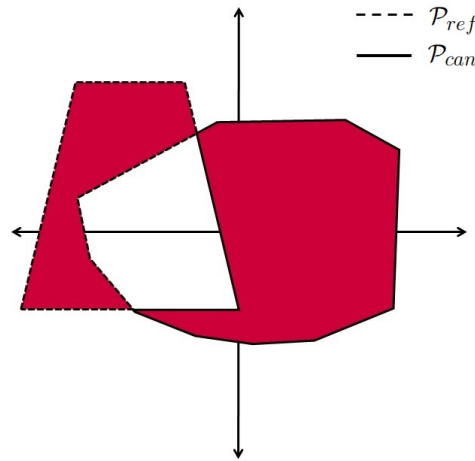


Figure 4.5: An example of the Compare Separate Volume Method.

#### 4.1.1.4 Direction of Net Positions Method

The last method to compare polytopes does not look at the volume of the polytope at all, but only to the shape of the polytope. More specific, we only consider the shape of the polytope at a specific place. This method calculates the distance between the boundary of the candidate polytope and the outcome of the net positions. Figure 4.6 visualises this method.

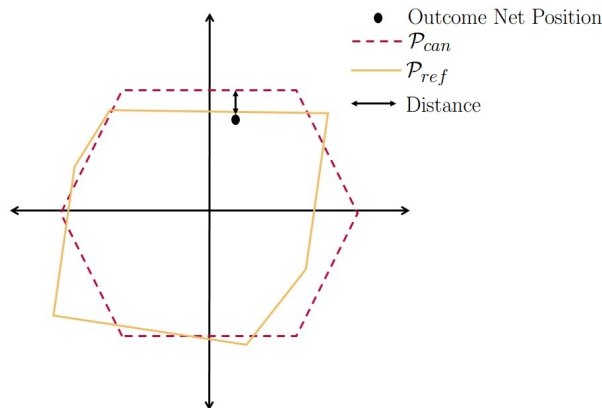


Figure 4.6: An example of the Compare in Direction of Net Positions Method.

A big advantage of this method is that it takes into account the outcome of the net positions in the polytope, which gives a better connection to the problem. A disadvantage is that this method is useless when the transport capacity constraints are unnecessary.

#### 4.1.1.5 Used Method to Compare Polytopes in this Thesis

The most important criteria to identify whether our approximation is a good approximation, is comparing directly the forecast and actual market clearing prices. Therefore, we



use this measure in this thesis. The main reason for this decision is that the goal of Eneco is to forecast the market clearing prices, so the flow-based domain approximation should work good in that sense.

However, we also want to be able to say something about the correctness of the approximated flow-based domain, even when the flow-based domain is not of influence on the market clearing prices (when the transport capacity constraints are unnecessary). For this we use the “separate volume” method. This method is chosen because it takes into account volume and location of the polytopes that we compare and does not favour smaller polytopes. We do not use the “mean distance to corner points” method because there are too many cases for which this is not a good comparison method, as it is too much dependent on the shape of the polytope. We do not use the “compare in direction of net positions” method because this method is only interesting when the net positions are constraint by the flow-based domain, but in some cases there is no ‘direction’ of the net positions. However, when there is a direction of the net positions, a more direct way to say something about the correctness of the domain is to directly compare the outcome of the market clearing price and the actual market clearing price. So the comparison in direction of the net positions method does not have added value.

In order to find out which characteristics have influence on the flow-based domain, we use the separate volume method. The reason for this is that we solely want to find out which characteristics change the flow-based domain, while the outcome of comparison in market clearing prices can be dependent on the orders of that day. If we want to find out which characteristics are of influence on the flow-based domain the market orders should not be of influence.

To conclude, in order to determine whether an approximation is ‘good’ or not, we compare the market clearing prices in combination with the separate volume method. First we look at the outcome in market clearing price (and how much they are the same as the actual market clearing price) and secondly we use the separate volume method to qualify the similarity of the polytopes.

## 4.2 Testing Different Methods on (a Subset of) Hours

When trying to find the characteristics of influence on the flow-based domain, we need a method to verify whether we can say that a characteristic has influence or not. In order to solve that problem we create the so-called Training-Set (*TS*). This training-set consists of 500 hours between May 2015 till October 2015, which are randomly selected. This selection is random but representative for the data-set from May 2015 till October 2015. For the training-set we only use hours between May till October, one of the reasons for this is that the temperature of this period does not change too much, which limits seasonal influences. Also the data since the 20<sup>th</sup> of May 2015 is more accurate and complete than the data from before 20<sup>th</sup> of May 2015. We use the hours in this training-set to test the methods described in Section 4.5. The specific hours of the training-set can be found in Appendix A.1.1.

Training-Set



When verifying to what extent our approximate is a good approximate, we need a set of hours called the approximation-set (*AS*), on which we test the approximation, these hours can be found in Appendix A.1.2. For the hours in this set we need the original market clearing prices and original flow-based domain. The hours selected for the approximation-set consists of 100 hours between 1 January 2015 and 31 December 2015, because those are the only days for which we have enough information to calculate the actual market clearing prices. This selection is random but representative for the entire data-set (2013, 2014 and 2015). The approximation-set and training-set have no hours in common. The reason for this is that we want to create as much as possible independence.

Approximation-  
Set

Quite some data of the flow-based domain is available, this data can be used to find out which characteristics are of influence on the flow-based domain and to approximate the flow-based domain. Since the 20<sup>th</sup> of May 2015 the flow-based model is used in the NWE market coupling model for the CWE countries and detailed data of the flow-based domain is available since then. The parallel run, done before this date, also provides some data. For the years 2013, 2014 and 2015, the data is available for most parameters within the flow-based model. Also for these years (almost) all data of the characteristics is available. The data for the parameters within the flow-based model can be found at [11]. The data about the characteristics is provided by Eneco.

The training-set and approximation-set are representative for the used data-set. Here representative means representative in the characteristics vector. When a characteristics vector is the same for many hours, a large subset of those hours is in the approximation-set. On the other hand, when a characteristics vector only appears for one (or not so many) hour(s) in the data-set, a small (or no) subset of those hours is in the approximation-set. The consequence of this is that generally the hours selected for the approximation-set have characteristics vectors that appear for a large amount of hours in the data-set. The approximation methods use hours in the data-set with the same characteristics vector when approximating the flow-based transport capacity constraints. As a result the conclusions of this research only hold when we want to forecast an hour in the future which has a certain amount of hours with the same characteristics vector in the data-set.

### 4.3 Likely Candidates to be Characteristics of Influence on the Flow-Based Domain

In order to approximate the flow-based domain, the input of the functions  $g_{tso}$  or  $\tilde{g}_{tso}$  must be known. For that, we first have to find out how the characteristics vector **Ch** is composed. Thus we want to find out which characteristics have influence on the flow-based domain. The first step is to identify possible candidates. From Chapter 2, a list of relevant candidates are derived:

1. **Renewable energy forecast** is the forecast of solar and wind energy.
2. **Consumption forecast** is the forecast of the amount of electricity consumers will use.



3. **Congestion forecast** is the forecast of flow on the branches in the network.
4. **Known outages** are the branches and generators that are non-functional.
5. **Generation shift keys** are the forecasts of the increase/decrease of import/export mapped to the generators.
6. **Day of the week** is an indicator for demand of electricity.
7. **Season** is an indicator for the possible temperature and weather.

The forecast of renewable energy and the forecast of consumption, can be done quite accurate. To forecast the congestion, the entire design of the electrical network must be taken into account. Making a program which can simulate the entire electrical network and the change in flow in this electrical network would be time consuming and the correctness that is gained by making this very accurate, detailed program would be nullified by the errors made by the use of forecasts. The information about known outages is not very useful, because it cannot be coupled directly to a specific branch  $b(n)$  or directly to the design of the flow-based domain. The reason that this is not possible has to do with the publication of the parameters within the flow-based model by the TSOs. After the TSOs have calculated the parameters within the flow-based model they publish the parameters on the different platforms, for example at [11]. In this publication every row in the PTDF-matrix (and corresponding RAM-value) corresponds to a certain branch  $b(n)$ , every  $b(n)$  has a unique number which is published with the constraint. This number is fixed, so every time it appears it refers to the same branch  $b(n)$ , but it is unknown which physical branch  $b(n)$  corresponds to which number in the publication. This makes it impossible to couple a known outage to a specific constraint or the design of the flow-based domain. The influence of the generation shift keys on the parameters within the flow-based model is important, but a problem with this forecast is that it is only useful if a change in a generation shift key can be related to a specific constraint for branch  $b(n)$  or to the design of the flow-based domain, which is not the case. The day of the week (and hour of the day) and season can be of influence on the parameters within the flow-based model and this information is known. Concluding, we only consider the following likely candidates: renewable energy forecast, consumption forecast, day of the week, hour of the day and season.

## 4.4 Definition of Equivalence Classes

To find out if there is a relation between certain characteristics and the flow-based domain we have to define a certain concept of equality. If we have a definition of equality we can determine when a characteristic is ‘equal’ for two different data points (in our case, hours) and what the consequence is from this equality for the flow-based domain.

The natural definition of equality would be that the characteristics have the exact same value. The only problem is that our data is very detailed and our data-set is limited. As a consequence, we (almost) never have the same value for a certain characteristic, which makes it impossible to find hours with the exact same value for even one characteristic, let alone for all the characteristics. This is why we define equivalence classes. Two values of one characteristic are called ‘equal’ if they are in the same equivalence class.



Assume we have characteristic vectors  $\mathbf{Ch}^{h_1}$  and  $\mathbf{Ch}^{h_2}$  for two hours  $h_1$  and  $h_2$ . We split up the domains  $\mathcal{D}_k$  in  $l_i$  sub-domains, where  $l_i$  is specific for a certain  $Ch_i$ . For those sub-domains holds:

$$\begin{array}{ccc} \mathcal{D}_1 = \mathcal{D}_1^1 \cup \dots \cup \mathcal{D}_1^{l_1} & \text{and } \forall i, j \ (i \neq j) \ \mathcal{D}_1^i \cap \mathcal{D}_1^j = \emptyset, \\ \vdots & & \vdots \\ \mathcal{D}_k = \mathcal{D}_k^1 \cup \dots \cup \mathcal{D}_k^{l_k} & \text{and } \forall i, j \ (i \neq j) \ \mathcal{D}_k^i \cap \mathcal{D}_k^j = \emptyset, \end{array}$$

These sub-domains are used to define our equivalence class.

$$\mathbf{Ch}^{h_1} \approx \mathbf{Ch}^{h_2} \iff \text{for all } i \text{ the values of } Ch_i^{h_1} \text{ and } Ch_i^{h_2} \text{ are in the same sub-domain } \mathcal{D}_i^j.$$

If the characteristics vector of two hours ( $\mathbf{Ch}^{h_1}$  and  $\mathbf{Ch}^{h_2}$ ) are in the same equivalence class, we call them ‘equal’.

Note, that each domain can be split up into sub-domains in many different ways. In our case we use the average value of a certain characteristic  $Ch_i$  to define the middle of the entire domain  $\mathcal{D}_i$ . After that the standard deviation of characteristic  $Ch_i$  defines the size of the sub-domains. Each domain, which is not the first or last domain, has a length equal to 1 time the standard deviation. For example if we split up the domain  $\mathcal{D}_i$  of characteristic  $Ch_i$  in 5 sub-domains the split up is as in Figure 4.7.

$$\bar{x} = Ch_i^{average}$$

$\sigma$  = standard deviation of  $Ch_i$

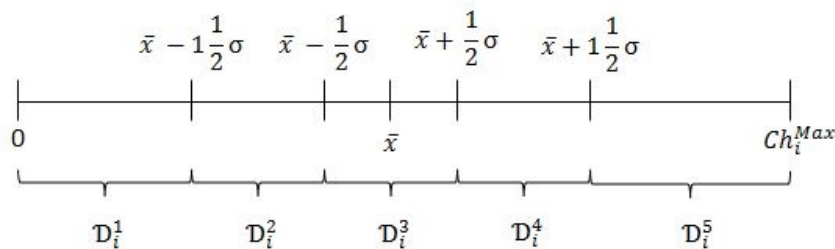


Figure 4.7: Example of the split up of domain  $\mathcal{D}_i$  of characteristic  $Ch_i$ , where  $l_i = 5$ .

#### 4.5 Analyse Methods to find the Characteristic of Influence

Based on the equivalence classes, we can define methods to find out which set of characteristics should be used as input for the methods to approximate the flow-based domain. This section describes the methods to find out which characteristics (and split up of the domains corresponding to the characteristics) best can be used as input to approximate the functions  $g_{tso}$  or  $\tilde{g}_{tso}$ . The first method is quite comprehensive, here we explore the shape and size of the flow-based domain.



### 4.5.1 (Almost) Equal Flow-Based Domains Method

The first method to find characteristics of influence is the method where one reference flow-based domain is compared to all other flow-based domains in our data-set. For this comparison we use the separate volume method and look at hours that have the most similar flow-based domain (that have the least separate volume). All domains that have a maximum score of 15% separate volume, are selected. For the hours that satisfy this threshold, we look at the characteristics and compare them with the characteristics of the reference domain. This gives us a way to say something about the influence of characteristics on the flow-based domain.

Every hour belongs for every characteristic to a certain sub-domain. It turns out that, when looking at all hours in the data-set, not an equal amount of hours belongs to each sub-domain. For example if we divide characteristic ‘Wind’ into three sub-domains, more hours belong to ‘Wind Low’ than ‘Wind High’. We can calculate the percentage of hours which belong to a certain sub-domain per characteristic, relative to the number of hours in the data-set. The results of this are given in Table 4.1. For this method every hour has three characteristics,  $Ch_1$ ,  $Ch_2$  and  $Ch_3$ , which are forecast of wind generation, forecast of solar generation and forecast of demand and are divided into three sub-domains, so  $l_1 = 3, \dots, l_3 = 3$ . The interpretation of this table is as follows: For 40.87% of the hours within our data-set the characteristic ‘Wind’ is in sub-domain ‘Low’.

Characteristic $Ch_i$	$\mathcal{D}_i^1 = \text{‘Low’}$	$\mathcal{D}_i^2 = \text{‘Medium’}$	$\mathcal{D}_i^3 = \text{‘High’}$
$Ch_1 = \text{Wind}$	40.87 %	35.37 %	23.76 %
$Ch_2 = \text{Solar}$	56.17 %	18.41 %	25.42 %
$Ch_3 = \text{Demand}$	36.90 %	26.89 %	36.21 %

Table 4.1: Assignment of hours to the different sub-domains,  $\mathbb{E}[Ch_i \in \mathcal{D}_i^j]$ .

Table 4.1 gives the expected percentage of hours which belong to each sub-domain when randomly picking a subset of hours from our data-set. If the number of hours which belong to a certain sub-domain is higher than the expected number of hours which belong to that sub-domain, we call that characteristic ‘significant’. When a characteristic is ‘significant’ for many reference hours this might indicate that the flow-based domain is influenced by this characteristic.

The procedure for this method is described in detail in Algorithm 1. Every hour of the training-set (also referred to as the ‘reference’ hour) is compared to all other hours ( $AH \setminus TS_i$ ) in the data-set, without comparing to itself. First the hours that have a maximum separate volume of 15% with the reference hour, are selected and collected in the set of good hours (GH). After that we calculate the number of times that a good hour is assigned to the same sub-domain as the reference hour, we do this for each characteristic and we calculate the percentage. When this percentage is higher than the expected percentage as given in Table 4.1, we label the characteristic  $Ch_i^{ts}$  of reference hour  $ts$  as ‘significant’. After all hours of the training-set are done, the percentage of significance can be calculated per characteristic. The result of this analysis can be found in Section 4.6.



**Data:** All hours from data-set May 2015 - October 2015, denoted by  $AH$ , with specified  $Ch_1$ ,  $Ch_2$  and  $Ch_3$  and  $l_1 = 3, \dots, l_3 = 3$  & the training-set, denoted by  $TS$ .

**Result:** Percentage of significance for  $Ch_1$ ,  $Ch_2$  and  $Ch_3$ .

```

for  $ts \leftarrow 1$  to  $|TS|$  do
  for  $ah \leftarrow 1$  to  $|AH \setminus ts|$  do
    Calculate separate volume of  $\mathcal{P}_{ts}$  and  $\mathcal{P}_{ah}$ 
    if separate volume  $\leq 15\%$  then
      | Good Hours ( $GH$ )  $\leftarrow ah$ 
    else
      |
    end
  end
  for  $i \leftarrow 1$  to  $|Ch^{ts}|$  do
    Calculate the percentage of time  $Ch_i^{gh \in GH} \in \mathcal{D}_i^j$ , when known is  $Ch_i^{ts} \in \mathcal{D}_i^j$ 
    if Percentage of time  $Ch_i^{gh \in GH} \in \mathcal{D}_i^j > \mathbb{E}[Ch_i \in \mathcal{D}_i^j]$  then
      |  $Ch_i^{ts}$  is significant.
    else
      |
    end
  end
end

```

Calculate Percentage of the significance of  $Ch_1^{ts \in TS}, \dots, Ch_{|Ch^{ts}|}^{ts \in TS}$ .

**Algorithm 1:** Finding (almost) equal flow-based domains.

#### 4.5.2 Different Variations Method

The (almost) equal flow-based domains method is a specific way to say something about the characteristics that are of influence on the flow-based domain and needs many assumptions. Another method, with less assumptions, to analyse the characteristics that are of influence on the flow-based domain is the method to compare different variations of the set of characteristics and of the split up of the domains of those characteristics. In this method not the ‘best’ variation is looked for, but we look for the best method among a few methods. There are infinite ways to split up the domains into sub-domains and many ways to choose the characteristics to approximate the flow-based domain. Based on the knowledge of Chapter 2, we select a subset of all possible variations. The selected variations are described below. They are specified with the characteristics that are used and how the domains are split up.

- **Variation 1:** *Basic Variation*, three characteristics: forecast of wind generation, the forecast of solar generation and the forecast of demand. The characteristics are divided into three sub-domains, namely ‘Low’, ‘Medium’ and ‘High’.
- **Variation 2:** *Basic Plus Variation*, six characteristics. The three extra characteristics are ‘Time’, divided into two sub-domains, ‘Peak’ and ‘Off peak’, ‘Day’, divided into three sub-domains, ‘Weekday’, ‘Weekend Day’ and ‘Monday/Friday’, ‘Season’, divided into four sub-domains, ‘Summer’, ‘Winter’, ‘Spring’ and ‘Autumn’.



Variation	$Ch_1$	$Ch_2$	$Ch_3$	$Ch_4$	$Ch_5$	$Ch_6$
1	‘Wind’ $l_1 = 3$	‘Solar’ $l_2 = 3$	‘Demand’ $l_3 = 3$	- -	- -	- -
2	‘Wind’ $l_1 = 3$	‘Solar’ $l_2 = 3$	‘Demand’ $l_3 = 3$	‘Time’ $l_4 = 2$	‘Day’ $l_5 = 3$	‘Season’ $l_6 = 4$
3	‘Wind BE’ $l_1 = 3$	‘Wind DE’ $l_2 = 3$	‘Wind FR’ $l_3 = 3$	‘Wind NL’ $l_4 = 3$	‘Solar’ $l_5 = 3$	‘Demand’ $l_6 = 3$
4	‘Wind’ $l_1 = 3$	‘Solar BE’ $l_2 = 3$	‘Solar DE’ $l_3 = 3$	‘Solar FR’ $l_4 = 3$	‘Demand’ $l_5 = 3$	- -
5	‘Wind’ $l_1 = 3$	‘Solar’ $l_2 = 3$	‘Demand BE’ $l_3 = 3$	‘Demand DE’ $l_4 = 3$	‘Demand FR’ $l_5 = 3$	‘Demand NL’ $l_6 = 3$
6	‘Wind BE’ $l_1 = 5$	‘Wind DE’ $l_2 = 5$	‘Wind FR’ $l_3 = 5$	‘Wind NL’ $l_4 = 5$	‘Solar’ $l_5 = 5$	‘Demand’ $l_6 = 5$
7	‘Wind’ $l_1 = 5$	‘Solar BE’ $l_2 = 5$	‘Solar DE’ $l_3 = 5$	‘Solar FR’ $l_4 = 5$	‘Demand’ $l_5 = 5$	- -
8	‘Wind’ $l_1 = 5$	‘Solar’ $l_2 = 5$	‘Demand BE’ $l_3 = 5$	‘Demand DE’ $l_4 = 5$	‘Demand FR’ $l_5 = 5$	‘Demand NL’ $l_6 = 5$

Table 4.2: Specifications of the different variations.

- **Variation 3:** *Wind Market Areas Variation*, in this variation the forecast of wind generation is specific per market area.
- **Variation 4:** *Solar Market Areas Variation*, in this variation the forecast of solar generation is specific per market area. The information about solar generation for NL is not available.
- **Variation 5:** *Demand Market Areas Variation*, in this variation the forecast of demand is specific per market area.
- **Variation 6:** *Wind Focus Variation*, in this variation the forecast of wind, solar and demand are divided into five intervals instead of three intervals, also wind is specific per market area.
- **Variation 7:** *Solar Focus Variation*, in this variation the forecast of wind, solar and demand are divided into five intervals instead of three intervals, also solar is specific per market area. The information about solar generation for NL is not available.
- **Variation 8:** *Demand Focus Variation*, in this variation the forecast of wind, solar and demand are divided into five intervals instead of three intervals, also demand is specific per market area.

The first variation is the basic variation. The three characteristics are chosen based on the information from the calculation process of the TSOs as in Chapter 2. This simple variation is our starting point. In order to find out whether time of the day, day of the week and time of the year have influence, the second variation is designed. If the second variation performs better than the first variation, one of the three time specifications might influence the flow-based domain. The third, fourth and fifth variation are there to find out if it is useful to look at the wind generation, solar generation or demand forecasts,



specific for each market area. The sixth, seventh and eight variation do the same, but also indicate whether it is useful to split up all the domains not in three parts, but in five parts.

The same training-set as in the previous method is used for these variations. For every variation Algorithm 2 calculates the results. The set Similar Hours (SH) denotes the set with all hours that are equal to the reference hour  $ts \in TS$  (according to the definition of equal described in Section 4.4). The Remaining Hours (RH) is the set with all hours without the similar hours (the considered hour is also no part of this set). The result of this analysis is given in Section 4.6.

This method first searches for the hours similar to the considered instance of TS (similar in sense of equal characteristics vectors), those hours are collected in the similar hour set  $SH(ts)$  of hour  $ts$ . After that, the reference flow-based domain ( $\mathcal{P}_{ts}$ ) is compared to each flow-based domain of the hours in the similar hour set ( $\mathcal{P}_{sh \in SH}$ ). Then the method also compares the domains of all hours that are not in the similar hour set ( $AH \setminus SH(ts)$ ), to the reference domain. This results in two sets of separate volume percentages. The method takes the mean of both sets, which is the output of the method.

**Data:** All hours ( $AH$ ) from data-set May 2015 - October 2015, with specified  $Ch_k$  and  $l_k$  (as specified in Table 4.2) & the training-set ( $TS$ ).

**Result:** Mean separate volume percentage of similar hours & Mean separate volume percentage of remaining hours.

```

for  $ts \leftarrow 1$  to  $|TS|$  do
    Find  $SH(ts)$ ;
    for  $sh \leftarrow 1$  to  $|SH(ts)|$  do
        Calculate separate volume between  $\mathcal{P}_{ts}$  and  $\mathcal{P}_{sh}$ ;
    end
    for  $rh \leftarrow 1$  to  $|RH(ts)|$  do
        Calculate separate volume  $\mathcal{P}_{ts}$  and  $\mathcal{P}_{rh}$ ;
    end
    Calculate mean percentage of the outcome of the separate volume calculation for the
    similar hours;
    Calculate mean percentage of the outcome of the separate volume calculation for the
    remaining hours.
end

```

**Algorithm 2:** Algorithm for generating the results of the different variations.

### 4.5.3 Currently Used Variation Method

In order to find a good approximation for the flow-based domain, it is useful to have a benchmark, because a better performance than the benchmark is already an improvement. In this thesis the benchmark is the flow-based domain Eneco now uses to forecast the market clearing prices. Eneco uses two methods which both will be used as a benchmark. The first possibility is to use the most recent flow-based domain to approximate the flow-based domain of a certain hour, this is the domain of the same hour but the day before. The second possibility is to use the flow-based domain of the most recent comparable weekday



in the past. For Tuesday, Wednesday, Thursday and Friday this is the day before. For the Monday this is the Friday before, for Saturday this is the Saturday before and for Sunday this is the Sunday before. Those possibilities give Eneco a method to use old flow-based domains to forecast the market clearing prices.

In order to test if our approximation methods work better than the currently used method, we test how well this currently used method works. This test is done by comparing the flow-based domain of the reference hour with the candidate flow-based domain: the flow-based domain of the same hour the day before or the flow-based domain of the same hour of the most recent comparable weekday in the past. For the characteristic analysis only the first possibility is used, this is because this is the one that Eneco uses most. Further on, both methods are used as a reference benchmark.

For this research MATLAB<sup>®</sup> is used to run the described algorithms (the algorithms in this chapter but also in the following chapters). For the calculation of the volume of different polytopes the Multi-Parametric Toolbox 3.0 is used, [13].

## 4.6 Results of the Characteristics Analysis

In this section we present the results achieved with the methods presented in the previous sections, combined with the conclusions that can be drawn from this results. Based on this, a decision can be made for the set of characteristics and the split up of these characteristics that is used in the methods to approximate the flow-based domain.

### 4.6.1 Results of the (Almost) Equal Flow-Based Domains Method

Algorithm 1 calculates which characteristics are called ‘significant’, which means that more than the expected number of ‘good’ hours are in the same sub-domain as the reference hour. Recall that we define a ‘good’ hour as an hour which has a flow-based domain with a maximum separate volume of 15% with the reference hour. Table 4.3 contains the results of the analysis.

Characteristic	Percentage Significant
$Ch_1 = \text{Wind}$	88.31 %
$Ch_2 = \text{Solar}$	74.24 %
$Ch_3 = \text{Demand}$	80.30 %

Table 4.3: Results of the (almost) equal flow-based domains method.

The interpretation of the results is as follows: for the characteristic wind, in 88.31% of the investigated cases, more than the expected number of good hours is in the same sub-domain as the reference hour. For the characteristic solar, in 74.24% of the investigated cases, more than the expected number of good hours is in the same sub-domain as the reference hour. For the characteristic demand, in 80.30% of the investigated cases, more



than the expected number of good hours is in the same sub-domain as the reference hour.

When there would be no dependency between a characteristic and the shape of the flow-based domain we would expect a percentage of significance of 0, because in that case every set of good hours of a reference hour are divided among the sub-domains as would be expected. Or we would expect a percentage of significance equal to the percentage of ‘not significant’, where not significant means that the the percentage is less than the expected percentage. When we assume a certain characteristic does not have any influence we would expect the number of times the percentage is less than the expected percentage equal to the number of times the percentage is higher than the expected percentage. Table 4.4 contains the number of times a characteristic is significant, not significant and neutral (where neutral means that the percentage is equal to the expected percentage). Due to sometimes the lack of hours with a separate volume percentage with a maximum of 15%, the numbers do not add up to 500.

Characteristic	Significant	Not significant	Neutral
$Ch_1 = \text{Wind}$	408	34	20
$Ch_2 = \text{Solar}$	343	75	44
$Ch_3 = \text{Demand}$	371	65	26

Table 4.4: Number of instances that are significant, not significant and neural per characteristic.

When the realised percentage of significance (as in Table 4.3) is (much) higher for a specific characteristic, this means that when two domains are ‘comparable’ (comparable in sense that the separate volume is at maximum 15%), it is likely that they belong to the same sub-domain of the considered characteristic. The other way around also holds. But still it is hard to draw a conclusion from the results of this method, because of the limited variations used in the characteristics. To conclude, this method gives an indication that wind energy generation might have some more influence than solar energy generation and demand. Also demand has some higher influence than solar.

#### 4.6.2 Results of the Different Variations Method

Algorithm 2 calculates for each variation what the mean separate volume is for similar hours within the equivalence class of reference hour  $h$  and what the mean separate volume is of hours outside the equivalence class of the reference hour  $h$  (this means the hours in other equivalence classes than the equivalence class of hour  $h$ ). The results of the different variations are given in Table 4.5.

The most relevant factor is the difference between inside and outside the equivalence class, because this indicates whether the given set of characteristics has influence on the separate volume between the reference domain and the candidate domain or not. Variation 6 has the most impact among the given variations. The second best variation is Variation 2.

This method also indicates that the characteristic ‘forecast of wind generation’ has an



Variation	Inside Equivalence Class	Outside Equivalence Class	Difference Inside and Outside	Standard Deviation Inside	Standard Deviation Outside
1	48.51%	51.33%	2.82%	5.75%	4.85%
2	40.40%	51.23%	10.83%	9.34%	4.86%
3	43.63%	51.19%	7.57%	10.25%	4.86%
4	47.90%	51.30%	3.40%	6.39%	4.83%
5	44.53%	51.28%	6.75%	8.77%	4.87%
6	38.77%	51.18%	12.41%	14.64%	4.86%
7	45.68%	51.25%	5.57%	10.13%	4.85%
8	42.10%	51.20%	9.10%	11.21%	4.87%

Table 4.5: Results of the variations method.

Variation	Minimum Inside	Minimum Outside	Maximum Inside	Maximum Outside
1	36.13%	43.34%	65.46%	69.58%
2	3.77%	43.03%	64.87%	69.37%
3	4.21%	43.19%	66.91%	69.38%
4	24.76%	43.24%	67.17%	69.37%
5	6.58%	43.11%	64.46%	69.46%
6	3.97%	43.13%	75.46%	69.37%
7	3.23%	43.26%	70.70%	69.39%
8	7.17%	43.09%	64.75%	69.44%

Table 4.6: Table 4.5 continued.

impact on the flow-based domain when divided in 5 sub-domains. Also it seems that ‘time’, ‘season’ and ‘day’ have some impact. For ‘forecast of solar generation’ this has hardly any added value for the difference between inside and outside the equivalence class. ‘Forecast of demand’ has also quite some impact but not so much as the ‘forecast of wind generation’. The last conclusion which can be drawn from Table 4.6 is that the basic scenario is very limited and generates hardly any difference between inside and outside the equivalence class.

### 4.6.3 Results of the Currently Used Variation Method

In order to use the currently used variation as a benchmark, also the separate volume is calculated for the hours of the training-set compared with the hour of the day before, using the most recent flow-based domain. The results of this are given below.

If we compare the outcome of the difference between the day before calculation and the other hours calculation with the outcome of the different variations, we see that the currently used variation works better than the result of the best variation. The difference in the best variation is 12.41% and the outcome of the difference of the currently used method is 18.98%. The conclusion that can be drawn from Table 4.7 in combination with



Currently Used Variation	
Mean Separate Volume Hour Day Before	32.17%
Mean Separate Volume Other Hours	51.15%
Difference Day Before and Other Hours	18.98%
Standard Deviation Hour Day Before	12.23%
Standard Deviation Other Hours	4.86%
Minimum Separate Volume Hour Day Before	6.16%
Minimum Separate Volume Other Hours	42.91%
Maximum Separate Volume Hour Day Before	68.78%
Maximum Separate Volume Other Hours	69.36%

Table 4.7: Outcome of the Currently Used Variation.

the results in Table 4.5 is that the currently used method works relatively well compared to the best variation. This means that the similarity (similar in sense of separate volume method) between the flow-based domains of the reference day and the day before is quite good.

## 4.7 Number of Hours per Equivalence Class

Approximating the flow-based transport capacity constraints based on data from the past requires a selection procedure to select flow-based domains or transport capacity constraints from this data. First we select a set of characteristics which have relatively the most influence on the flow-based domain. When we want to make an approximation for hour  $h$  in the future for which we know the characteristics vector  $\mathbf{Ch}^h$  (with the value for each characteristic in the set of characteristics), we select the so-called similar hours  $SH(h)$  from the data-set, those are the hours that are equal to hour  $h$  in sense of the equivalence relation described in Section 4.4. The methods described in Chapter 5, use those similar hours for a procedure to make an approximation. A problem with the methods described in Section 4.5 is that the result might indicate a certain use and split up of characteristics, but that in practice there are not enough similar hours  $SH(h)$  to use for the methods which make an approximation. Before we make a decision about the set and split up of the characteristics, we should take into account the number of similar hours the set and split up create.

The variations described in Section 4.5.2 have all a specific set and split up of characteristics, this results in a number of equivalence classes. For example for the basic variation (variation 1) one of them is ‘wind=low, solar=high and demand=low’, in this way 27 equivalence classes are specified for variation 1. Table 4.8 contains the number of equivalence classes per variation.

When we know the characteristics vector of hour  $\mathbf{Ch}^h$ , this determines in which equivalence class this hour is. Also our hours from the past can be divided into the equivalence classes specified per variation and we can calculate the number of hours that is in each equivalence class. The reason for doing this, is that this number also indicates the number



Variation	Number of Equivalence Classes
1	27
2	648
3	729
4	243
5	729
6	15625
7	3125
8	15625

Table 4.8: Number of equivalence classes per variation.

of similar hours  $SH(h)$  which are available to use for the method to make an approximation of the flow-based domain of  $h$ . If this number is too low, this causes problems for the methods to approximate the flow-based domain. This is why we should take the mean number of hours per equivalence class into account when deciding which variation is used as set and split up of characteristics.

Table 4.9 contains the mean number of hours in the different equivalence classes. This mean number gives an indication of the spread of the hours among the equivalence classes. Also the minimum and maximum are given. Starting from variation 2 there are also equivalence classes without any hours in it. The consequence of this is that when we want to make an approximation for a specific hour  $h$ , it can be the case that there are no hours with the same characteristics vector  $\mathbf{Ch}^h$  in the data-set, which can be used in the approximation methods. So for this hour we cannot make an approximation. Table 4.9 also contains the mean number of hours per equivalence class without taking into account the empty equivalence classes. This gives an indication of the number of hours in an equivalence class, when we are able to make an approximation. For example an empty equivalence class can be when wind generation is very high and solar generation is also very high, because those events rarely happen together.

If we look at the outcome of the mean number of hours per equivalence class for each variation, we see that variation 1 has a high number of hours per equivalence class, this might indicate that variation 1 is too coarse. The same kind of reasoning holds for variation 4. Variation 7 has a special kind of result. Special about this result is that apparently in the equivalence classes of variation 7, the hours are collected in just a few equivalence classes, which results in a low mean number of hours per equivalence class, but a very high mean number of hours when the empty equivalence classes are ignored. When we say that there is a reasonable amount of hours per equivalence class for a specific variation, we mean that when using this variation to select hours  $SH(h)$  in our data-set to make an approximation for hour  $h$  in the future, there are a reasonable amount of similar hours to use in the approximation methods. Variations 2 and 3 have a reasonable amount of hours per equivalence class and the difference between the mean number of hours with or without empty equivalence classes is not very large. The mean number of hours (with



Variation	Mean number of hours per equivalence class	Minimum	Maximum	Mean number of hours (without empty equivalence classes)	Number of equivalence classes without hours
1	801.78	107	2541	801.78	0
2	33.41	0	468	38.32	83
3	29.70	0	912	38.25	163
4	89.09	0	2465	155.74	104
5	29.7	0	1584	53.19	322
6	1.39	0	655	16.99	14351
7	6.93	0	5824	176	3002
8	1.39	0	1120	24.63	14746

Table 4.9: Mean number of hours per equivalence class for each variation.

empty equivalence classes) is very low for variation 6 and 8, but without empty equivalence classes both variations have still a reasonable result.

## 4.8 Definitive Set of Characteristics

In order to decide which set of characteristics and which split up should be used to approximate the flow-based domain, the results of the different variations are leading. This is because this is the most broad and best comparable analysis. Although it is hard to draw a conclusion from the (almost) equal flow-based method, this method indicates that wind generation forecast might have a higher influence on the flow-based domain than solar generation or demand. The currently used method is only used as benchmark to compare the designed approximation methods with.

As already stated in Section 4.6.2, variation 6 and variation 2 behave best among the outcome of the variations. This is why those variation will play a role in the approximation of the flow-based domain. First only variation 6 is used as set and split up of characteristics. In a second step also the combination of variation 6 and variation 2 is used as input to approximate the flow-based domain. If we only use variation 6 as set and split up of characteristics in order to approximate the flow-based domain, every equivalence class in our data-set has on average almost 16 hours (see Table 4.9) to use for the approximation methods. The combination of variation 6 and variation 2 results in a mean number of 3 hours per equivalence class, which is not so much. The reason we still use the combination of variation 6 and 2 is that this combination gives a somewhat more precise selection of similar hours in the data-set, which might result in a better performance.



## Chapter 5

# Approximation Methods

The characteristic analysis done in Chapter 4 provides us with a set of characteristics which can be used as input to approximate the flow-based domain. Based on this, we now treat the problem of interest: approximating the flow-based domain. To tackle this problem we designed different approximation methods, which are described in this chapter.

### 5.1 Using Similar Hours to Approximate the Flow-Based Domain

The first method to approximation the flow-based domain for a specific hour  $h$  in the near future, is the so-called ‘Similar Hour’ method. This method looks for a set  $SH(h)$  of similar hours in the past. Similar, in this sense, means that the characteristic vectors of the hours in  $SH(h)$  are equal to the characteristic vector  $\mathbf{Ch}^h$  of hour  $h$  (as defined in Section 4.4), i.e. are in the same equivalence class as hour  $h$ . After the similar hours of hour  $h$  are found, the market clearing prices are calculated when using the transport capacity constraints of the hours  $sh \in SH(h)$ .

The similar hour method does not approximate the functions  $g_{tso}$  or  $\tilde{g}_{tso}$ , but uses a workaround. This means that it does not approximate the functions but uses the transport capacity constraints of the similar hours to directly forecast the market clearing prices. The only problem that remains is, to specify how we use the found market clearing prices per market area for the hours in  $SH(h)$  to forecast the market clearing prices for the different market areas for hour  $h$ . In the following paragraphs possible solutions for this problem are presented.

The similar hour method results in  $|SH(h)|$  potential market clearing prices per market area. A straight forward method to transform the many market clearing prices, is by taking the mean of the market clearing prices of a certain market area to forecast the market clearing price for that market area.

Another method to find a forecast of the market clearing price per market area is taking

Mean Mar-  
ket Clearing  
Prices



the median. The motivation for this is that, when the variation between the market clearing prices calculated with the transport capacity constraints of the similar hours becomes very large, the median is a more robust forecast. A more detailed description of this method can be found in Algorithm 3.

Median Mar-  
ket Clearing  
Price

**Data:** All hours  $AH$  from data-set May 2015 - October 2015, with specified  $Ch_k$  and  $l_k$  (as described in 4.8) & the approximation-set ( $AS$ ).

**Result:** Forecast of the market clearing price per market area calculated with the similar hour method.

```

for  $as \leftarrow 1$  to  $|AS|$  do
    Find  $SH(as)$  of  $as$ ;
    for  $sh \leftarrow 1$  to  $|SH(as)|$  do
        Calculate market clearing prices when using the transport capacity constraints of
        hour  $sh$ .
    end
    Calculate mean market clearing price per market area;
    calculate median of the market clearing prices per market area.
end

```

**Algorithm 3:** Algorithm corresponding to the similar hour method to forecast the market clearing price per market area.

The essence of the similar hour method is not to approximate the flow-based domain but to forecast the market clearing prices directly. In fact, the flow-based domain of each similar hour of hour  $h$  is used to forecast the market clearing price per market area for hour  $h$ . Using only one or a small subset of similar hour(s) to approximate the flow-based domain (and consequently the market clearing prices) has the risk to be quite unreliable and might lead to a high standard deviation in correctness. This is why all similar hours are used to forecast the market clearing prices and to transform these market clearing prices to one market clearing price per market area.

## 5.2 Using Probabilities to Approximate the Flow-Based Domain

The similar hour method is a simple and quite direct way to forecast the market clearing prices. A somewhat more extensive way to approximate the flow-based domain is adding a probabilistic component to the approximation. The method in which this is done, is called the probabilistic method and focuses somewhat more on the appearance of constraints  $C_{b(n)}$  corresponding to a branch  $b(n)$ .

The probabilistic method also starts with finding the similar hours of a certain reference hour  $h$  (the hour for which we want to make the approximation). The similar hours of hour  $h$  are collected in the set  $SH(h)$ . Every similar hour  $sh \in SH(h)$ , has a set of flow-based transport capacity constraints, denoted by  $\mathbf{C}^{sh,fb} = \{C_1^{sh}, C_2^{sh}, \dots, C_j^{sh}\}$ . Every constraint in this set corresponds to a certain branch  $b(n)$ . Using these constraints, the method generates sample sets which are used as approximation of the flow-based domain



of hour  $h$ . For the generation of a sample set, we use the set  $\mathcal{C}_{coll}^{h,fb} = \bigcup_{sh \in SH(h)} \mathbf{C}^{sh,fb}$ , in which a certain constraint  $C_{b(n)}$  can appear multiple times in different realisations (this means with different parameters). More precisely, branch  $b(n)$  is with probability

$$\mathbb{P}[h, b(n)] = \frac{|H(h, b(n))|}{|SH(h)|}, \quad (5.1)$$

selected for the sample set, where  $H(h, b(n)) = \{\tilde{h} | C_{b(n)} \in \mathbf{C}^{\tilde{h},fb} \text{ \& } \tilde{h} \in SH(h)\}$ . Note, that the cardinality of this set gives the number of constraints which represent  $b(n)$  in  $\mathcal{C}_{coll}^{h,fb}$ . The expected number of branches in the sample set is given by:

$$\mathbb{E}[z(h)] = \sum_{b(n) \in B^{total}} \frac{|H(h, b(n))|}{|SH(h)|},$$

where  $z(h)$  is the number of constraints in the sample set generated to approximate the flow-based domain of hour  $h$ .

After this, it is clear which branches  $b(n)$  are represented in the sample set, but this branch  $b(n)$  can have different realisations within  $\mathcal{C}_{coll}^{h,fb}$ . What is left, is to select a specific  $C_{b(n)}$  for each branch  $b(n)$  selected for the sample set among the different realisations within  $\mathcal{C}_{coll}^{h,fb}$ . This constraint is drawn randomly from the different realisations of constraints corresponding to branch  $b(n)$  in  $\mathcal{C}_{coll}^{h,fb}$ .

The resulting sample set now gives the constraints which form a flow-based domain, denoted by  $\mathbf{C}^{h,fb,for}$ . In this way,  $w$  different sample sets are created, which results in  $w$  market clearing prices. In our case we have chosen  $w = 100$ . The mean or median of the market clearing prices generated by the sample sets is our forecast.

In order to test whether the approximated flow-based domain is a good approximation the separate volume of the reference domain and the candidate domain is calculated. The following algorithm gives a more detailed description of this method.



**Data:** All hours  $AH$  (from the 2013, 2014 and 2015 data-set), with specified  $Ch_k$  and  $l_k$  & the approximation-set ( $AS$ ).

**Result:** Mean and median market clearing price calculated with the probabilistic method & Mean separate volume of every instance of the approximation-set.

```

for  $as \leftarrow 1$  to  $|AS|$  do
    Find  $SH(as)$ ;
    Find  $C_{coll}^{as,fb}$ ;
    Calculate  $\mathbb{P}[as, b(n)] \quad \forall b(n) \in B^{total}$  according to Equation (5.1);
    for  $i \leftarrow 1$  to  $w$  do
        for  $b(n) \leftarrow 1$  to  $|B^{total}|$  do
             $b(n)$  is selected with probability  $\mathbb{P}[as, b(n)]$ ;
            if  $b(n)$  is selected then
                select a  $C_{b(n)}$  among the  $C_{b(n)} \in C_{coll}^{as,fb}$  (every realisation of  $C_{b(n)}$  has equal probability to be selected);
                 $C^{as,fb,for} \leftarrow C_{b(n)}$ ;
            else
                end
            end
        end
        Calculate the market clearing price per market area with  $C^{as,fb,for}$ ;
        Calculate separate volume of  $\mathcal{P}_{as}$  and  $\mathcal{P}_{for}$ ;
    end
    Calculate mean separate volume;
    calculate mean market clearing price per market area, this results in our approximation by using the mean;
    calculate the median of the market clearing prices per market area, this results in our approximation by using the median.
end

```

**Algorithm 4:** Algorithm corresponding to the probabilistic method.

In order to clarify this method, an example is given. Assume we want to approximate the flow-based domain of hour  $h$ . For this we know the characteristics-vector of hour  $h$ , namely  $Ch^h$ . Corresponding to this characteristics-vector we find 3 similar hours in our data-set, hour  $sh_1, sh_2$  and  $sh_3$ , so  $SH(h) = \{sh_1, sh_2, sh_3\}$ . Table 5.1, Table 5.2 and Table 5.3 give the constraint sets of the similar hours. Those constraint sets form together the set

$b(n)$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$RAM_{id}$
BN-1	-1	0	0	0	2486
BN-2	0.27087	-0.02267	0.14195	-0.17573	1168
BN-5	-0.05442	0.20454	0.0969	0.23226	629

Table 5.1: Constraints of similar hour  $sh_1$ .

$C_{coll}^{h,fb}$ , which is given in table 5.4.

To calculate the probabilities, we need the set  $H(h, b(n))$  for each  $b(n)$ . For  $b(n) = \text{BN-1}$ ,  $C_{b(n)}$  is in all three similar hours, which results in:  $H(h, \text{BN-1}) = \{sh_1, sh_2, sh_3\}$ . For



$b(n)$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$RAM_{id}$
BN-1	-1	0	0	0	2486
BN-3	-0.2168	0.01784	-0.11255	0.16684	815

Table 5.2: Constraints of similar hour  $sh_2$ .

$b(n)$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$RAM_{id}$
BN-1	-1	0	0	0	3291
BN-4	0.30142	0.23585	0.23344	0.20325	369
BN-5	-0.05443	0.20454	0.09892	0.23144	632

Table 5.3: Constraints of similar hour  $sh_3$ .

$b(n)$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$PTDF_{id,m_1}$	$RAM_{id}$
BN-1	-1	0	0	0	2486
BN-1	-1	0	0	0	2486
BN-1	-1	0	0	0	3291
BN-2	0.27087	-0.02267	0.14195	-0.17573	1168
BN-3	-0.2168	0.01784	-0.11255	0.16684	815
BN-4	0.30142	0.23585	0.23344	0.20325	369
BN-5	-0.05443	0.20454	0.09892	0.23144	632
BN-5	-0.05442	0.20454	0.0969	0.23226	629

Table 5.4: Constraints of similar hour  $h_3$ .

$b(n) = \text{BN-2}$ ,  $C_{b(n)}$  is only in  $\mathbf{C}^{sh_1}$ , which results in  $H(h, \text{BN-2}) = \{sh_1\}$ . This leads to the following probabilities.

$$\begin{aligned}\mathbb{P}[h, \text{BN-1}] &= \frac{|H(h, \text{BN-1})|}{|SH(h)|} = \frac{3}{3}, \\ \mathbb{P}[h, \text{BN-2}] &= \frac{|H(h, \text{BN-2})|}{|SH(h)|} = \frac{1}{3}, \\ \mathbb{P}[h, \text{BN-3}] &= \frac{|H(h, \text{BN-3})|}{|SH(h)|} = \frac{1}{3}, \\ \mathbb{P}[h, \text{BN-4}] &= \frac{|H(h, \text{BN-4})|}{|SH(h)|} = \frac{1}{3}, \\ \mathbb{P}[h, \text{BN-5}] &= \frac{|H(h, \text{BN-5})|}{|SH(h)|} = \frac{2}{3},\end{aligned}$$

These values denote the probability that a certain branch  $b(n)$  is selected for the sample sets. The expected number of constraints in a sample set is  $\mathbb{E}[z(h)] = (1 \cdot 1 + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3}) = 2\frac{2}{3}$ . After the construction of the sample set, the method calculates the market clearing price per market area with the sample set as flow-based transport capacity constraints. After  $w$  sample set calculations, the mean or median is the forecast for the market clearing price per market area.



### 5.3 Probabilistic Method Including Correlation

When using the probabilistic method we assume that the constraints  $C_{b(n)}$  are statistically independent. This means that the probability that  $C_{b_1(n)}$  is critical for hour  $h$  on day  $d$  is independent of the probability that  $C_{b_2(n)}$  is critical for the same hour  $h$  on day  $d$ . From the documentation about the calculation of the transport capacity constraints by the TSOs we suspect that this is not the case. Also the outcome of the probabilistic method suggest that this independence is not necessarily true. This is why we design a method where we take correlation between transport capacity constraints into account.

We suspect a relationship between the criticality of different branches. In order to calculate if there exists a relationship between two random variables the *correlation coefficient* can be used. The correlation coefficient for (0/1)-variables (in our case a branch  $b(n)$  is critical ( $=1$ ) or not ( $=0$ )) is defined as follows:

Correlation  
Coefficient

$$\rho(b_1(n), b_2(n)) = \frac{\mathbb{P}[b_1(n) \wedge b_2(n)] - \mathbb{P}[b_1(n)] \cdot \mathbb{P}[b_2(n)]}{\sqrt{\mathbb{P}[b_1(n)] \cdot (1 - \mathbb{P}[b_2(n)]) \cdot \mathbb{P}[b_2(n)] \cdot (1 - \mathbb{P}[b_1(n)])}}.$$

$\rho(b_1(n), b_2(n))$  denotes the correlation coefficient between branch  $b_1(n)$  and branch  $b_2(n)$ , which specifies whether there is a positive, negative or neutral relationship between the occurrence of branch  $b_1(n)$  and branch  $b_2(n)$  in the constraint set. It holds that  $\rho(b_1(n), b_2(n)) \in [-1, 1]$  and a correlation coefficient near +1 means that branch  $b_1(n)$  and branch  $b_2(n)$  are typically critical together or both are not critical and a correlation coefficient near -1 means that when branch  $b_1(n)$  is critical, branch  $b_2(n)$  typically is not critical and vice versa.

The correlation coefficient gives a relation between two variables, but when we have to decide whether a certain branch is added to the sample set we have to take into account all branches that are already in the sample set. Let us assume that  $\mathbf{C}^{h,fb,sample} = \{C_{b_1(n)}, \dots, C_{b_\beta(n)}\}$  are the constraints that are already in the sample set,  $\mathbf{C}^{h,fb,sample} = \emptyset$  is also possible. For the simple case that  $\mathbf{C}^{h,fb,sample} = \{C_{b_1(n)}\}$ , we can calculate the correlation between branch  $b_1(n)$  and any other branch  $b_2(n)$  and decide based on this correlation to add branch  $b_2(n)$  to the sample set or not. For example, we can decide to add branch  $b_2(n)$  when the correlation between the two branches is strong positive. The probability that branch  $b_2(n)$  is critical given that branch  $b_1(n)$  is critical, is the formal probability for the decision to add branch  $b_2(n)$  to the sample set or not. When the sample set already consists of more than one constraint we aim to calculate the probability that  $b(n)$  is critical given that  $b_1(n), \dots, b_\beta(n)$  are already critical. The probability for branch



$b_1(n)$  to be critical given that  $\{b_2(n), \dots, b_\beta(n)\}$  are already critical is given by:

$$\begin{aligned}
 \mathbb{P}[h, b_1(n)] &= \mathbb{P}[b_1(n) = 1 | b_2(n) = 1 \wedge \dots \wedge b_\beta(n) = 1], \\
 &= \frac{\mathbb{P}[b_1(n) = 1 \wedge b_2(n) = 1 \wedge \dots \wedge b_\beta(n) = 1]}{\mathbb{P}[b_2(n) = 1 \wedge \dots \wedge b_\beta(n) = 1]}, \\
 &= \frac{|H(h, b_1(n) \wedge b_2(n) \wedge \dots \wedge b_\beta(n))|}{|SH(h)|} \cdot \frac{|SH(h)|}{|H(h, b_1(n))|}, \\
 &= \frac{|H(h, b_1(n) \wedge b_2(n) \wedge \dots \wedge b_\beta(n))|}{|H(h, b_1(n))|},
 \end{aligned} \tag{5.2}$$

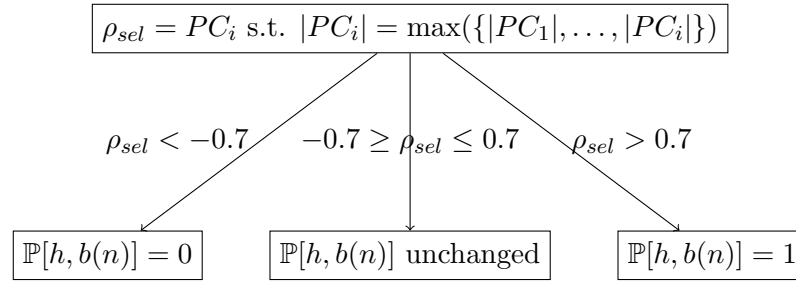
where  $H(h, b_1(n) \wedge b_2(n) \wedge \dots \wedge b_\beta(n)) = \{\tilde{h} | (C_{b_1(n)} \wedge C_{b_2(n)} \wedge \dots \wedge C_{b_\beta(n)}) \in \mathbf{C}^{\tilde{h}, fb} \text{ \& } \tilde{h} \in SH(h)\}$ .

If we want to use the conditional probability of Equation (5.2), we have to calculate the probability for each branch for each sample set. Recall that  $t$  is the number of different branches that appear in the data-set and  $\gamma$  denotes the number of different hours in the data-set. Per sample set we have to decide  $t$ -times whether we denote a specific branch  $b(n)$  as critical or not. For each decision  $\mathbb{P}[h, b(n)]$  is calculated. When calculating  $\mathbb{P}[h, b(n)]$  we have to search for branch  $b(n)$  within the hours  $sh \in SH(h)$  but also for the branches that are already denoted critical. The part of the time that branch  $b(n)$  appears in  $sh \in SH(h)$  with the branches that are already denoted critical is the conditional probability. The set  $SH(h)$  is at most of size  $\gamma - 1$ , per similar hour we search at maximum for  $t$  branches. Searching for a constraint representing a specific branch can be done in  $\mathcal{O}(\log(t))$ , where the a set of constraints of a similar hour has length at most  $t$ . The running time of generating only one sample set is of order  $\mathcal{O}(\gamma \cdot t^2 \cdot \log(t))$ , when doing this for  $w$  sample sets for each instance of the approximation-set this is  $\mathcal{O}(|AS| \cdot w \cdot \gamma \cdot t^2 \cdot \log(t))$ . Based on the relatively long running-time, we restrict our-self to calculate only correlations between pairs of branches and use those correlations to adjust the probability  $\mathbb{P}[h, b(n)]$ .

If we only use correlations to adjust the probability  $\mathbb{P}[h, b(n)]$ , we can calculate the pairwise correlation for the considered branch  $b_1(n)$  and the branches which are already critical. When more than one branch is already denoted critical we have to compare the pairwise correlations in order to adjust the probability  $\mathbb{P}[h, b_1(n)]$ . Assume we consider to add a constraint of branch  $b_1(n)$  to the sample set  $\mathbf{C}^{h, fb, sample}$  which already consists of  $\beta$  constraints,  $\{C_{b_2(n)}, \dots, C_{b_\beta(n)}\}$ . We base our decision on the pairwise correlations between  $b_1(n)$  and  $b_2(n), \dots, b_\beta(n)$ , which are collected in the vector  $PC(b_1(n), \mathbf{C}^{h, fb, sample}) = \{\rho(b_1(n), b_2(n)), \dots, \rho(b_1(n), b_\beta(n))\}$ . Using this vector, we consider the correlation with maximal absolute value. If this is unique and is larger than 0.7, we set the probability that  $b_1(n)$  is selected 1, i.e.  $\mathbb{P}[b_1(n)] = 1$ . If this is unique and it is smaller than  $-0.7$ , we set the probability that  $b_1(n)$  is selected to 0, i.e.  $\mathbb{P}[b_1(n)] = 0$ . In all other cases  $\mathbb{P}[b_1(n)]$  is unchanged. Figure 5.1 illustrates the steps, here  $\rho_{sel}$  is the correlation with maximal absolute value.

Note, that this method is to some extend the same as the probabilistic method. Only the probability that branch  $b(n)$  is selected for the sample set depends on the correlation between  $b(n)$  and the branches already in the sample set. Consequently the algorithm cor-



Figure 5.1: Decision tree for the adjustment of  $\mathbb{P}[h, b(n)]$ 

responding to this method is also quite similar to the algorithm used for the probabilistic method. For this we add the corrections of  $\mathbb{P}[as, b(n)]$  based on the pairwise correlations, a detailed description of the algorithm corresponding with the correlation method is given in Algorithm 5.

**Data:** All hours  $AH$  (from the 2013, 2014 and 2015 data-set), with specified  $Ch_k$  and  $l_k$  & the approximation-set ( $AS$ ).

**Result:** Mean and median market clearing price calculated with the correlation method & Mean separate volume of every instance of the approximation-set.

```

for  $as \leftarrow 1$  to  $|AS|$  do
  Find  $SH(as)$ ;
  Find  $C_{coll}^{as,fb}$ ;
  Calculate  $\mathbb{P}[as, b(n)] \quad \forall b(n) \in B^{total}$  according to Equation (5.1) in combination with
  decision tree 5.1;
  for  $i \leftarrow 1$  to  $w$  do
    for  $b(n) \leftarrow 1$  to  $|B^{total}|$  do
       $b(n)$  is selected with probability  $\mathbb{P}[as, b(n)]$ ;
      if  $b(n)$  is selected then
        select a  $C_{b(n)}$  among the  $C_{b(n)} \in C_{coll}^{as,fb}$  (every realisation of  $C_{b(n)}$  has equal
        probability to be selected);
         $C^{as,fb,for} \leftarrow C_{b(n)}$ ;
      else
      end
    end
    Calculate the market clearing price per market area with  $C^{as,fb,for}$ ;
    Calculate separate volume of  $\mathcal{P}(C^{as,fb})$  and  $\mathcal{P}(C^{as,fb,for})$ ;
  end
  Calculate mean separate volume.
  Calculate mean market clearing price per market area, this results in our
  approximation by using the mean.
  Calculate the median of the market clearing prices per market area, this results in
  our approximation by using the median.
end
  
```

**Algorithm 5:** Algorithm corresponding to the Correlation Method.



# Chapter 6

## Results

For generating results, it is important to notice that an approximation is considered a ‘good’ approximation of the flow-based transport capacity constraints when the approximation generates a market clearing price which is close to the actual market clearing price. For this we need a definition of the error between the actual market clearing price and the forecast market clearing price. The method for this error calculation is explained in the first section of this chapter. The following section presents a visualisation and interpretation of the results of the different methods. A more detailed presentation of this visualisation and interpretation of the results can be found in Appendix A.3. After that the different methods are statistically compared. The last section contains some validation and verification that is done during the generation of the results.

### 6.1 Different Ways to Calculate an Approximation Error

Approximation  
Error

There are different ways to measure the difference between an approximation and a real value, this difference is called the ‘*approximation error*’. A few commonly used error calculations are:

- mean absolute error;
- mean absolute percentage error;
- mean squared error;
- root mean squared error.

#### Mean Absolute Error (MAE)

This method of calculating the approximation error between a vector  $\mathbf{y} \in \mathbb{R}^u$  of approximations and a real value  $\bar{y} \in \mathbb{R}$  is the most common:

$$MAE = \frac{1}{u} \sum_{i=1}^u |\bar{y} - y_i|,$$



In our case the real value  $\bar{y}$  is the actual market clearing price and the vector  $\mathbf{y}$  is the vector of the forecasts of the market clearing price.

### Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error method is quite similar to the mean absolute error, only this method is scaled to the size of  $\bar{y}$ :

$$MAPE = \frac{1}{u} \sum_{i=1}^u \left| \frac{\bar{y} - y_i}{\bar{y}} \right|.$$

### Mean Squared Error (MSE)

A very popular calculation method is the mean squared error:

$$MSE = \frac{1}{u} \sum_{i=1}^u (\bar{y} - y_i)^2,$$

### Root Mean Squared Error (RMSE)

A way to adjust the mean squared error to a more realistic error is by taking the root. This method is called the root mean squared error.

$$RMSE = \sqrt{\frac{1}{u} \sum_{i=1}^u (\bar{y} - y_i)^2}.$$

The above described ways to calculate an approximation error are all (slightly) different, but the mean absolute error and the mean absolute percentage error are quite similar (similar when we want to rank methods according to the outcome of the error), also the mean squared error and the root mean squared error are quite similar. In order to keep the results clear, we only use two types of error calculation. For Eneco the absolute error is most important, because that error says something about the amount of money the forecast of the market clearing prices generated with the approximation of the flow-based domain, is off. In order to base the conclusions not only on one type of error calculation method, we also use the root mean squared error as this method gives a relatively high weight to large errors and large errors are undesirable for Eneco because large errors lead to large losses.

## 6.2 Visualisation and Interpretation of the Results

This section contains the visualisation and interpretation of the results generated with the currently used method, the similar hour method, the probabilistic method and the correlation method. The similar hour method, the probabilistic method and the correlation



method have two variations, one using the mean market clearing price as forecast and one using the median of the market clearing prices as forecast. The currently used method also has two variants, one where the most recent flow-based domain is used as approximation and a second where the most recent comparable flow-based domain is used as approximation. All methods are tested with the approximation-set of variation 6 and with the approximation-set of variation 2 & 6. Also we use two different error measuring methods, the mean absolute error (MAE) and the root mean squared error (RMSE). The outcome of the results differ for each country of the CWE region, which are the Netherlands (NL), Belgium (BE), Germany (DE) and France (FR).

The first set of figures contain the MAE for the different methods in combination with both approximation-sets. The corresponding results for NL and BE are given in Figure 6.1 and Figure 6.2. The results are sorted in ascending order, starting with the lowest mean absolute error. The results of DE and FR are given in Appendix A.4, respectively in Figure A.2 and Figure A.3. The explanation of the labels of the bars in the figures in this chapter is given in the Table 6.1.

Label	Explanation
B	Currently used method, most recent day.
BS	Currently used method, most recent comparable day.
S	Similar hour method.
P	Probabilistic method.
C	Correlation method.
d	Median used as forecast.
m	Mean used as forecast.
6	Approximation-set of variation 6.
26	Approximation-set of variation 2 & 6

Table 6.1: Explanation of the labels used in the following figures.

When we look at the results of the MAE for NL, BE, DE and FR (Figure 6.1, Figure 6.2, Figure A.2 and Figure A.3), it is noticeable that the second half of the figures often contains the results of the similar hour method and the correlation method. For NL there is even a noticeable increase between the first half and the second half. For BE this difference is less clear, but there the results of the correlation methods are relatively worse than the results of the other methods. For DE there is not really a clear distinction between the two halves, but for FR and DE the similar hour methods are noticeable worse than the other methods. Also for FR a slight distinction between the first and second half can be seen.

The most visible of the results of the MAE is that the market clearing price of Belgium seems relatively hard to forecast compared to the market clearing prices of Germany and France. The currently used methods, but also the probabilistic methods seem to be able to forecast the market clearing prices within a range of €1,- to €2,- in mean absolute error, while forecasting the market clearing price of Belgium starts around an mean absolute error of €4,-. For the Netherlands this range is around €3,- to €5,-. This difference in



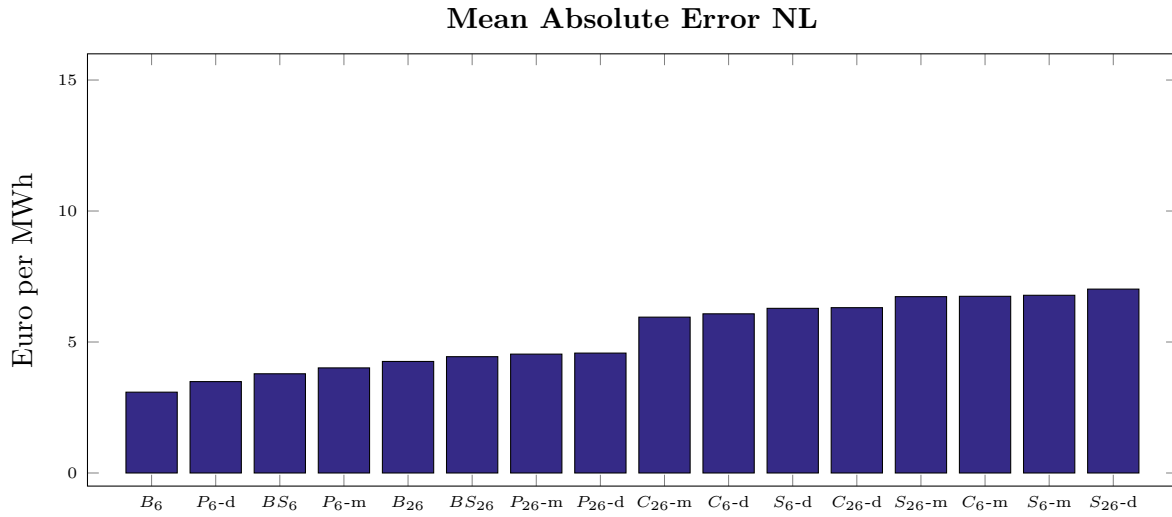


Figure 6.1: Mean absolute error results for the Netherlands.

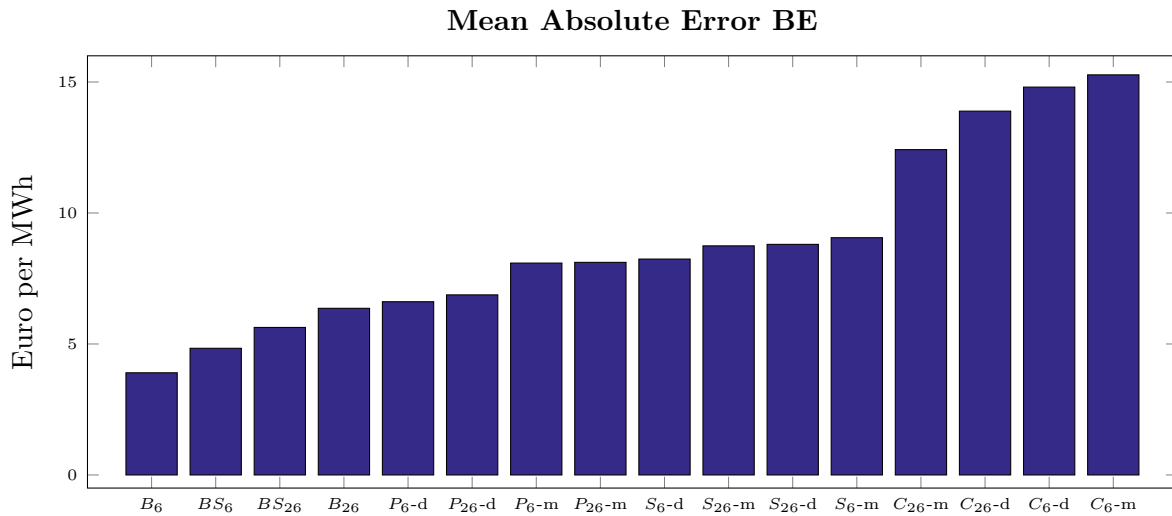


Figure 6.2: Mean absolute error results for Belgium.

forecasts has probably to do with the fact that Belgium is a relatively small country which is sensitive to changes in import and export. When the approximation of the flow-based domain is relatively bad, the price in Belgium tends to peak quickly. For Germany and France this is not the case, this mainly has to do with the large volume and with the block orders that are traded there. As a consequence in Germany and France the market clearing prices seem to be relatively stable.

The results of the error measure RMSE for NL, BE, DE and FR can be found in the Appendix A.4 in Figure A.4, Figure A.5, Figure A.6 and Figure A.7. Those results are similar (similar in sense of what is described in the paragraph above) to the results generated with the MAE. However, the results of BE are an exception, here the correlation method is still relatively bad, but the similar hour methods behave quite good relative to the probabilistic methods, but the difference is small. In NL the distinction between the



first and second half is less clear than the MAE results. The MAE results for FR and DE are quite similar to the RMSE results.

Up to now we only compared the results based on mean absolute error and root mean squared error, but we did not take the standard deviation of those errors into account. In Figure 6.3 the standard deviation of the MAE NL is given too. The bottom part of the bars is still the MAE but the upper part is the standard deviation of the error, in this way we visualise the standard deviation in combination with the mean absolute error. The ordering of the methods is the same as in the previous MAE results. The results for BE, DE and FR are given in the appendix in Figure A.8, Figure A.9 and Figure A.10. For NL we see that the probabilistic methods have a somewhat smaller standard devi-

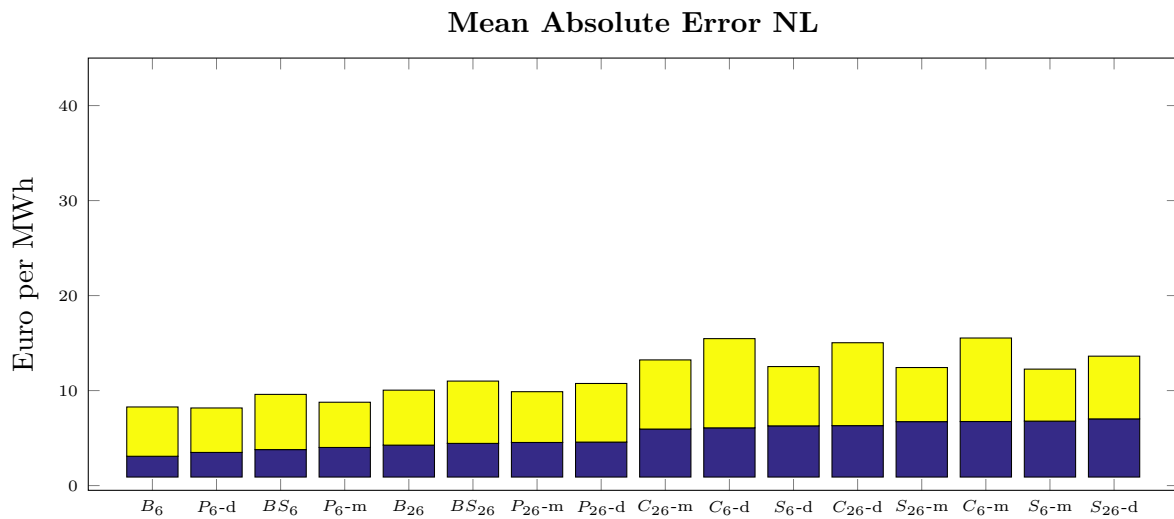


Figure 6.3: MAE with added standard deviation for NL.

ation than the currently used methods, also the similar hour methods have a relatively smaller standard deviation than the correlation methods. In the results of BE it can be seen that the similar hour methods have a relatively smaller standard deviation, which brings these methods closer to the probabilistic methods. The correlation methods have also a quite large standard deviation for BE. The results for DE and FR are similar to the results described above. The results of the standard deviation for the RMSE can be found in the appendix in Figure A.11, Figure A.12, Figure A.13 and Figure A.14. For NL the results are similar to the RMSE, noticeable is that the currently used method with the most comparable day seems to have a slightly larger standard deviation. For BE the similar hour methods seem to have a smaller standard deviation than the probabilistic methods. The results for DE and FR are similar to the results of the MAE and the RMSE.

The results of the standard deviation already give an idea about how much the results are spread. Another way to gain more insight on the dispersion of the outcome is with a boxplot. A boxplot represents the minimum of the results, the maximum of the results and divides the results in four quartiles. Outliers are plotted as individual points. The middle of the boxplots is the median ( $Q_2$ ), the edges of the boxes are the first ( $Q_1$ ) and third quartile ( $Q_3$ ), which are the data numbers that respectively denote the lower 25%



and higher 75% of the data. A data point is considered an outlier when it is smaller than  $Q_1 - \zeta \cdot (Q_3 - Q_1)$  or larger than  $Q_3 + \zeta \cdot (Q_3 - Q_1)$ , where in our case  $\zeta = 1.5$ .

In Figure 6.4 the boxplots are plotted for the MAE in NL, the dotted line represents where the data is cut (above this line only outliers are plotted). The results are now ordered from ‘simple’ to ‘complex’, starting with the currently used method, most recent day with the approximation-set of variation 6, following by the same method with the approximation-set of variation 2 & 6. After that the currently used method with most recent comparable day follows, also with both variation 6 and variation 2 & 6, following by the similar hour methods, probability methods and finally the correlation methods. From the boxplots of the MAE for NL we can see that in general the similar hour methods and correlation methods have a relatively large dispersion. The boxplots for BE, DE and FR can be found in the appendix in respectively Figure A.15, Figure A.16 and Figure A.17. For BE it is also the case that the similar hour methods and correlation methods have a relatively larger dispersion, with some exceptions for the currently used methods for the approximation-set of variation 2 & 6. For DE the most noticeable is that all methods seem to result in a quite small dispersion except for the similar hour methods. The results of FR with respect to dispersion are similar to the results of DE, only the correlation methods also have a somewhat larger dispersion. If we do the same for the MSE calculation method we get similar results. The figures corresponding to the boxplots of the MSE calculation can be found respectively in Figure A.18, Figure A.19, Figure A.20 and Figure A.21.

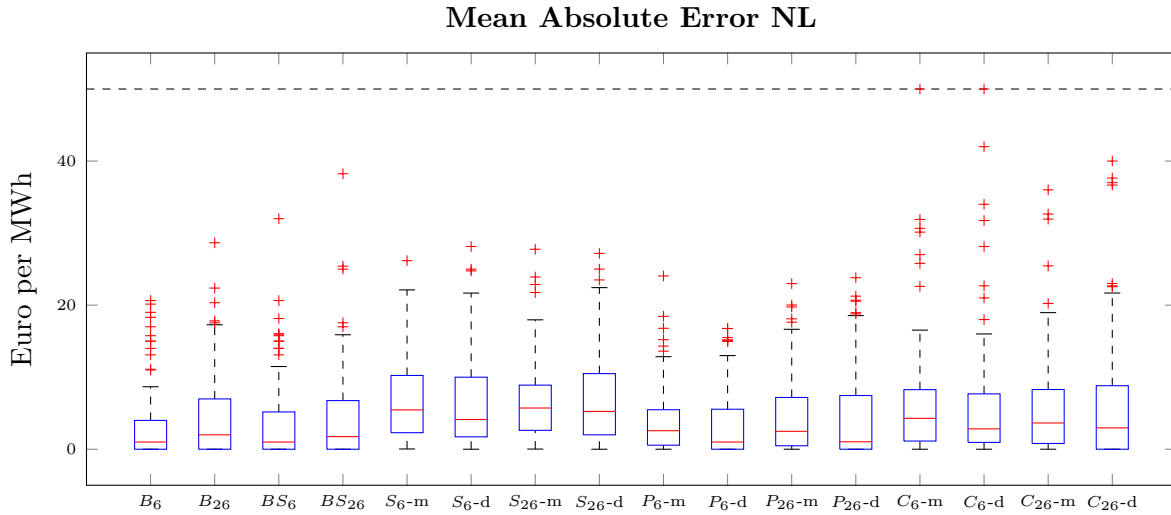


Figure 6.4: Boxplots of the MAE results in NL.

One thing the results above have in common is that for each country, for each error measuring method and for each approximation-set, it holds that the correlation methods (almost) always behaves worst. When ranking the methods the correlation methods always belong to the worst half, or even are last. Also in dispersion the correlation methods stand out in negative sense. The likely reason for this is that in the correlation method, when adjusting the probability that a certain constraint is added to the sample set, this probability is only adjusted positively, which results in (many) more constraints. More constraints can make



the feasible region smaller, which makes the approximated flow-based domain tighter and can make the flow-based domain less realistic. In that case the market clearing prices become harder to forecast. Consequently the correlation method is not a good way to approximate the flow-based transport capacity constraints. This is why we do not consider this method anymore from now on. Some more explanation about the difference between the correlation methods and the probabilistic methods is given in Section 6.2.2.

The similar hour methods behave also quite bad but sometimes it seems to have a quite small standard deviation and for BE the similar hour methods behave relatively good. This is why we cannot draw a conclusive conclusion at this point and we still consider the results of the similar hour methods. In the results presented above we see that the MAE gives similar results as the RMSE. From now on we continue to look only at the results of the MAE, this is because these results are most interesting for Eneco and from the previous results we can conclude that the results seem to be not really different when using the MAE in comparison with the RMSE.

The currently used methods behave pretty good as approximation method for the market clearing prices in all market areas of the CWE region. Also the probabilistic methods can provide a relatively good approximation. The problem with the currently used method is that this method works good for forecast horizons which are in the very near future (day ahead), but when forecasting a day further in the future this might not be the case. For the behaviour of the currently used methods in a forecasting horizon which is further than one day ahead, more research is necessary. An advantage of the probabilistic methods is that it is not limited to a certain forecast horizon, with the input of the characteristics vector of a certain hour in the future, the probabilistic method is able to make a relatively good forecast.

It is not always the case that the approximation-set of variation 6 gives the same results as the approximation-set of variation 2 and 6. Figure 6.5 shows this for NL, where the left bar is variation 6 and the right bar variation 2 & 6. In NL for the currently used methods, the probabilistic methods and the similar hour method when using the median, variation 6 gives relatively better results than variation 2 & 6. For the similar hour method when using the mean, the results for both approximation-sets are very close to each other. For BE for the currently used methods variation 6 works relatively better. For the remaining methods the difference is inconclusive. The results for DE and FR are quite the same, namely inconclusive results for the currently used methods and probabilistic methods and for the similar hour methods, variation 2 & 6 works relatively better. The figures which show the results when comparing the approximation-set of variation 6 with the approximation-set of variation 2 & 6 for BE, DE and FR can be found in the appendix, respectively in Figure A.22, Figure A.23 and Figure A.24. Summarising, whether the approximation-set of variation 6 or variation 2 & 6 gives relatively better forecast is at this point yet unknown.

The last comparison we do, is to compare using the mean v.s. using the median. This comparison can only be made for the similar hour methods and the probabilistic methods, because for the currently used methods this is not relevant. The result of this for NL



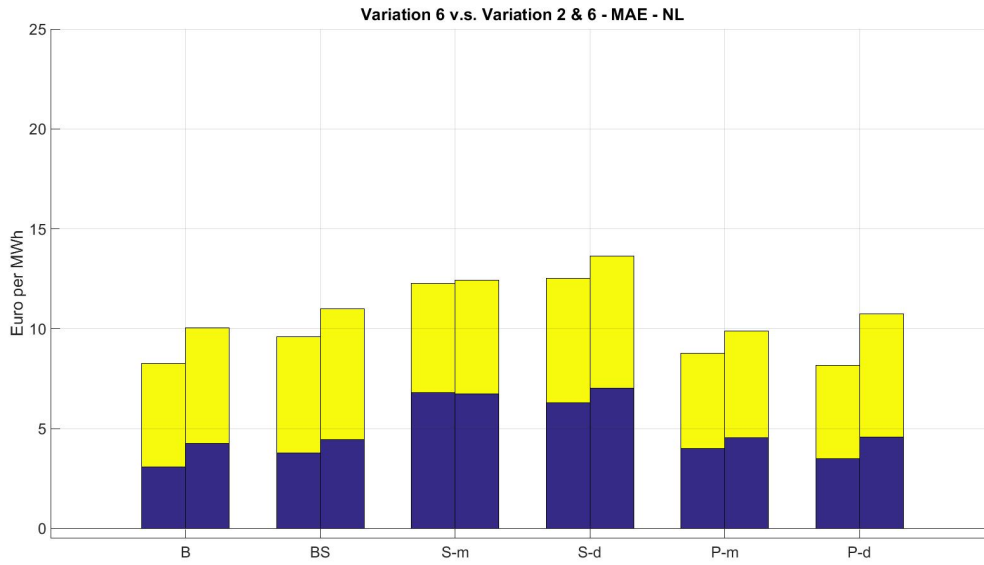


Figure 6.5: The MAE outcome of using the approximation-set of variation 6 compared with using the approximation-set of variation 2 & 6 plotted in one figure for NL, the left bar is variation 6 and the right bar variation 2 & 6.

are given in Figure 6.6. When looking at the figure we see hardly any difference between using the mean v.s. using the median. For BE the results seem also very close to each other but the median seems to have a little advantage. Also for FR and DE the difference seems hardly noticeable. The figures for BE, DE and FR can be found in the appendix, respectively in Figure A.25, Figure A.26 and Figure A.27.

In the above, different comparisons are made in order to be able to draw a general conclusion about the behaviour of the different methods. This section ends with an overview of the conclusions and methods in order to clearly summarise the main results. Table 6.2 contains the main results for all countries considered with all different methods. Table 6.3 contains the results when we compare using the approximation-set of variation 6 v.s. variation 2 & 6. Finally Table 6.4 contains the results when we compare using the mean v.s. the median. Note that all results are relative to the other methods, where the currently used methods are our benchmark. Also -- stands for relatively very negative results, - for relatively negative results, 0 for neutral results which means inconclusive, + for relatively positive results, ++ for relatively very positive results.

A conclusion that can be drawn from Table 6.2 is that in general the benchmark methods generate a relatively good forecast, but that also the probabilistic methods stay not far behind. The correlation methods do not even come close to the benchmark methods. The similar hour methods usually generate a relatively bad forecast, but most of the time still better than the correlation method. For all countries this seems to be the case, except for Belgium this is not entirely true. For Belgium the probabilistic methods do not rank as high as for the other countries, also the similar hour methods rank a bit higher than with the other countries.



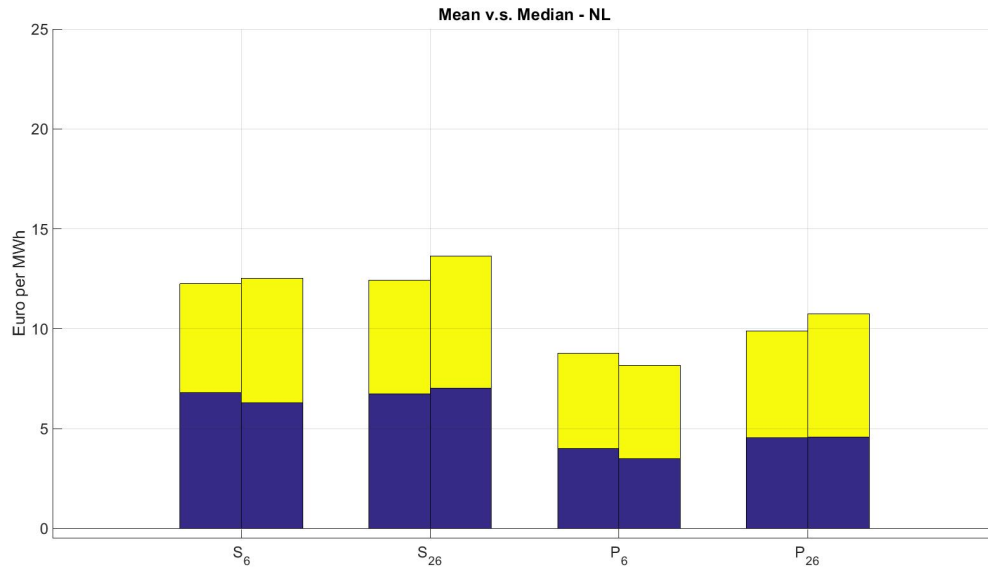


Figure 6.6: The MAE outcome of using the mean compared with using the median plotted in one figure for NL, the left bar is the mean and the right bar is the median.

	NL				BE				DE				FR			
	B	S	P	C	B	S	P	C	B	S	P	C	B	S	P	C
MAE	++	--	+	-	++	-	+	--	++	--	+	-	++	--	+	-
RMSE	+	-	++	--	++	+	-	--	++	--	+	-	++	--	+	-
Stacked MAE	+	-	++	--	++	+	-	--	+	--	+	0	++	--	+	-
Stacked RMSE	+	-	++	--	++	+	-	--	+	--	+	0	++	--	+	-
Boxplot MAE	++	--	+	0	+	-	0	--	+	--	+	0	++	--	+	-
Boxplot RMSE	+	--	+	-	++	-	0	-	+	--	+	0	++	--	+	-

Table 6.2: Summary of the main results.

	NL		BE		DE		FR	
	Var 6	Var 2 & 6	Var 6	Var 2 & 6	Var 6	Var 2 & 6	Var 6	Var 2 & 6
B	++	--	++	--	0	0	0	0
BS	+	-	0	0	-	+	-	+
S-m	0	0	-	+	-	+	-	+
S-d	+	-	0	0	-	+	-	+
P-m	+	-	+	-	0	0	-	+
P-d	++	--	+	-	0	0	-	+

Table 6.3: Results of comparing the use of the approximation-set of variation 6 to the use of the approximation-set of variation 2 & 6.

Comparing the approximation-set of variation 6 with the approximation-set of variation 2 & 6 is quite hard. There is not a significant distinction between the results generated



with both approximation-sets. For DE and FR variation 2 & 6 seems somewhat better, but for NL and BE this seems to be the other way around.

	NL		BE		DE		FR	
	m	d	m	d	m	d	m	d
$S_6$	0	0	-	+	+	-	+	-
$S_{26}$	+	-	+	-	+	-	+	-
$P_6$	-	+	---	++	0	0	0	0
$P_{26}$	+	-	---	++	0	0	0	0

Table 6.4: Results of comparing the use of the mean to the use of the median.

Likewise comparing the mean and the median used as forecast is also quite hard. There is not a significant distinction between the results generated with both options. For DE and FR the mean seems to have a little advantage, but for BE this is the other way around. In NL there is not really a distinction. However, in all cases (except for the probabilistic methods in BE) the difference is not significantly large between the two options.

### 6.2.1 Results of Comparing Polytopes

The results presented in the previous section only look at the outcome of the market clearing prices in order to rank the designed methods. Another method to rank the designed methods is looking at the polytope generated by the approximation. Especially when the outcome of the net positions is not restraint by the flow-based domain this polytope can still say something about the accurateness of the approximation. For this, we compared the polytopes of the approximations with the actual flow-based domain using the separate volume method. This comparison method is significantly different than the comparison methods used in previous section, so we consider the results of all approximation methods.

For the currently used methods the flow-based domain of the (most comparable) previous day is used to forecast the market clearing prices. Per instance of the approximation-set one flow-based domain is used as approximation of the flow-based domain, this approximated flow-based domain has a certain separate volume with the actual flow-based domain of the hour for which we want to make a forecast. In order to rank the different methods, we take the mean of the outcomes in separate volume of the different instances of the approximation-set. For the currently used methods this leads to the results presented in Table 6.5. For the similar hour methods, probabilistic methods and correlation methods this works a little different. For these methods we usually find more than one flow-based domain per instance of the approximation-set. For example, in the similar hour method we find  $|SH(h)|$  similar hours from which we use the flow-based domain as approximation in order to forecast the market clearing prices. The flow-based domain of each similar hour is compared to the actual flow-based domain which results in  $|SH(h)|$  separate volumes. From these separate volumes we take the mean (or median), which gives us an indication on the average separate volume between the actual flow-based domain and the flow-based domain of the similar hours. The number presented in Table 6.5 is the mean of these outcomes, so the mean of the mean (or median) of the results of the instances



of the approximation-set. Also in this table the standard deviation (S.D.) of the mean (or median) of the outcome for the separate volume method is given. Figure 6.7 gives some insight in the results, here the results are ordered in ascending order (only taken into account the mean).

Method	Mean	S. D.
$B_6$	0.28	0.10
$B_{26}$	0.30	0.11
$BS_6$	0.30	0.11
$BS_{26}$	0.30	0.11
$S_6 - m$	0.50	0.07
$S_6 - d$	0.51	0.09
$S_{26} - m$	0.49	0.09
$S_{26} - d$	0.51	0.10
$P_6 - m$	0.55	0.06
$P_6 - d$	0.55	0.06
$P_{26} - m$	0.51	0.08
$P_{26} - d$	0.51	0.08
$C_6 - m$	0.65	0.09
$C_6 - d$	0.66	0.10
$C_{26} - m$	0.53	0.10
$C_{26} - d$	0.54	0.11

Table 6.5: Results of the separate volume comparison.

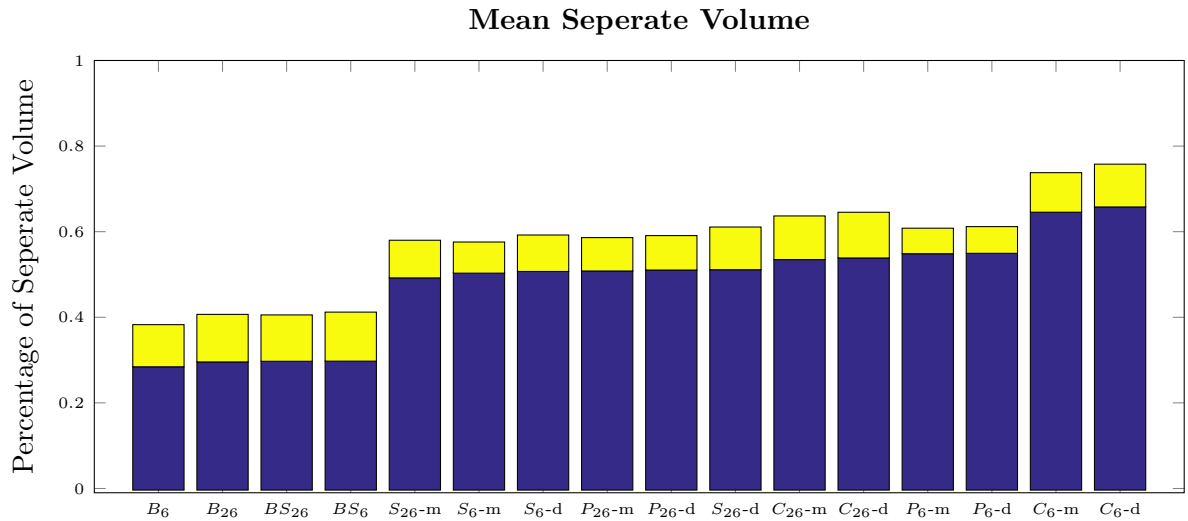


Figure 6.7: Results of the separate volume comparison.

Noticeable is that the currently used methods have a significantly smaller mean, which means that in terms of the separate volume method those flow-based domains are quite comparable. However, the standard deviations are relatively large when we compare them with the standard deviation of the probabilistic methods. This probably has to do with



the fact that the currently used method only has one flow-based domain as approximation, while the other methods usually have multiple flow-based domains which are used as approximations to forecast the market clearing prices. Still the difference in mean is much larger than the difference in standard deviation. The other methods have quite the same mean, with an exception for the correlation methods when using the approximation-set of variation 6, those methods have a somewhat higher mean.

### 6.2.2 Increase in the Expected Number of Constraints for the Correlation Method

The results presented up to this section are conclusive on one thing: it seems that the approximation methods which takes correlation into account do not have good results in the forecast of market clearing prices, relative to the probabilistic methods. The only difference between those methods is that the correlation methods have an adjustment of the probability that a certain branch  $b(n)$  is selected for the sample set, as given in Figure 5.1. Those adjustments can be positive (when the maximal absolute correlation is strongly positive) or negative (when the maximal absolute correlation is strongly negative). In order to find an explanation for the large difference in rank of those methods, we try to find out what the impact of the adjustment of the probabilities is. For this we calculate the expected number of constraints when adjusting the probabilities and when we do not adjust the probabilities. Table 6.6 gives those results, where the second column gives the mean of the difference between the expected number of constraints with and without adjustment of the probabilities, also the minimum difference, maximum difference and standard deviation of the difference is given. A positive difference means that the expected number of constraints with adjustment is higher than without adjustment.

Method	Mean	Minimum	Maximum	Standard Deviation
Correlation, Variation 6	34.75	3.75	146.98	21.27
Correlation, Variation 2 & 6	60.94	9.09	184.19	35.39

Table 6.6: The difference between the expected number of constraints with and without adjustment of the probabilities.

Table 6.6 shows that the impact of the adjustment of the probabilities is only positive, which means that the expected number of constraints in the correlation method is higher than in the expected number of constraints in the probabilistic method. Consequently the realised number of constraints is higher for the correlation method. More constraints can indicate a tighter flow-based domain, which can result less transport of electricity between countries. This can result in less realistic prices.

## 6.3 Statistically Comparing Different Methods

In the visualisation of the results, a difference can be seen between the different approximation methods, but sometimes the mean and standard deviation of those methods are



very close. In order to make a distinction between the mean error of two methods, a statistical test is applied. This test decides with a certain confidence level ( $1 - \alpha$ ), whether two methods significantly differ in mean or not. Often used confidence levels are 95% or 99%, in our case we use a confidence level of 95%, which results in  $\alpha = 0.05$ . Our basic hypothesis is that the different methods do not significantly differ in mean, so when the outcome of the test is smaller or equal to 0.05, we reject the basic hypothesis and say that the tested two methods do significantly differ in mean.

There are many different statistical tests to test whether two methods significantly differ in mean or not. A few of them are:

- **Wilcoxon signed-rank test** ([16]): The Wilcoxon signed-rank test is a statistical test for a non-parametric hypothesis. The hypothesis compares two paired samples, from which is not assumed that they are normally distributed.
- **Welch's t-test** ([27]): The Welch's t-test is a statistical test for the hypothesis that two populations have equal mean. The samples do not have to be paired. Welch's t-test assumes that the populations are normally distributed.
- **Paired t-test** ([18]): The paired t-test is similar to the Welch's t-test, only the samples must be paired.
- **Paired z-test** ([14]): The paired z-test assumes normal distribution of the samples difference and assumes that the standard deviation of the population is known.

In our case we do not have a good reason to assume that the results are normally distributed. However, we have paired observations. This is why we use the Wilcoxon signed-rank test, to test whether the methods differ significantly in mean or not. The test has been applied per market area. In Table 6.7 the results of the methods using the approximation-set of variation 6 are given. In Table A.14 in the appendix the results of the same test but when using the results of the approximation-set of variation 2 & 6 are given.

In Tables 6.7 and A.14 an entry '1' means that the result is below the confidence level of 5%, which means that our basic hypothesis is rejected. As in table 6.7 most entries are a '1', statistically not many methods are said to be equal in mean. Only the similar hour methods when using the mean and median seem to be equal in mean. Also for all market areas except BE, the probabilistic method using the median seems to be equal in mean with the currently used method when using the most recent comparable day. For the approximation-set of variation 2 & 6 the results are somewhat more positive. Here we see that the currently used methods seem to be equal in mean. Also the probabilistic methods seem to be equal in mean with the currently used methods (except for DE). Besides that, the similar hour methods seem to be equal in mean with each other (when testing using the mean v.s. using the median), the same holds for the currently used methods, the probabilistic methods and the correlation methods. So for the approximation-set of variation 2 & 6 it does not significantly matter whether we use the mean or median to approximate the flow-based transport capacity constraints.



	<i>B</i>				<i>BS</i>				<i>S - m</i>				<i>S - d</i>			
	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR
<i>B</i>	-	-	-	-												
<i>BS</i>	0	1	1	1	-	-	-	-								
<i>S - m</i>	1	1	1	1	1	1	1	1	-	-	-	-				
<i>S - d</i>	1	1	1	1	1	1	1	1	0	0	0	0	-	-	-	-
<i>P - m</i>	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1
<i>P - d</i>	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1
<i>C - m</i>	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1
<i>C - d</i>	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1

	<i>P - m</i>				<i>P - d</i>				<i>C - m</i>				<i>C - d</i>			
	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR
<i>B</i>																
<i>BS</i>																
<i>S - m</i>																
<i>S - d</i>																
<i>P - m</i>	-	-	-	-												
<i>P - d</i>	0	1	0	1	-	-	-	-								
<i>C - m</i>	1	1	1	1	1	1	1	1	-	-	-	-				
<i>C - d</i>	1	1	1	1	1	1	1	1	0	0	0	0	-	-	-	-

Table 6.7: Results of the Wilcoxon signed-rank test when testing the different methods with the results when using approximation-set of variation 6.

## 6.4 Validation and Verification

In order to assure that the methods meet the requirements, verification and validation is necessary. For the validation part we constantly observed the methods and observed how they behaved after implementing them. Here we noticed that the similar hour method and probabilistic method seemed not to outperform the currently used method. During the validation phase of those methods, a new method was designed. We identified that a possible problem with the probabilistic method could be that we assume that the constraints are independent. The correlation method takes the correlation between branches into account and emanated from the validation of the probabilistic method.

Besides validating the designed methods we also had to verify that the algorithms that are used to perform the methods behave according to that method. While implementing the methods, tests were done to check this. Also we checked whether the number of constraints is in agreement with the expected number of constraints. The results of this can be found in Table 6.8. First we calculate the expected number of constraints for the probabilistic method and compare the expected number of constraints with the realised number of constraints. From this difference we calculated the mean. For the correlation method we did the same, only the expected number of constraints differs in the correlation method per random drawing because the expected number of constraints is dependent on the order in which the constraints are selected. After the process of selecting constraints



for the sample set, the expected number of constraints was calculated. This expected number was compared to the realised number of constraints. For completeness the minimum, maximum and standard deviation of the difference between the expected number of constraints and the realised number of constraints are also given.

Method	Mean	Min.	Max.	S. D.	Mean $\mathbb{E}$ [Number of Constraints]	Mean Number of Constraints
Probabilistic, Variation 6	3.15	0	16	2.40	22.05	22.04
Probabilistic, Variation 2 & 6	3.02	0	18.87	2.34	22.34	22.35
Correlation, Variation 6	3.03	0	17.93	2.29	82.99	83.02
Correlation, Variation 2 & 6	2.74	0	15	2.08	57.09	57.12

Table 6.8: Outcome when validating the difference between the expected number of constraints and the realised randomly drawn number of constraints.

The mean difference for all methods is around 3 constraints, with a minimum of 0, maximum around 17 and a standard deviation around 2. When we compare the mean of the expected number of constraints and the mean of the realised number of constraints we see that those two are very close to each other, which indicates that the random generator works properly. The mean difference around 3 constraints is relatively low compared to the mean number of realised constraints.



## Chapter 7

# Conclusions, Recommendations and Discussion

The conclusions are described in this chapter. From the conclusions some further research opportunities arises. This chapter also describes the limitations of this research which should also be taken into account when interpreting the conclusions.

### 7.1 Main Conclusions of the Results

In this research, four different types of methods are compared which approximate the flow-based transport capacity constraints which are used to forecast the market clearing prices for four market areas, which are the Netherlands, Belgium, Germany and France. The compared types of methods are: the currently used methods, the similar hour methods, the probabilistic methods and the correlation methods. For these methods we use two test-sets, the so-called approximation-sets, which consist of different hours. For each instance of the approximation-sets each method is used to make an approximation of the flow-based domain. Those approximation-sets are respectively based on the set of characteristics which is focused on generation of wind energy (so-called variation 6) and on the set of characteristics which also is focused on the time and season (so-called variation 2 & 6). The currently used methods consist of two methods, one for which the most recent day is used to approximate the flow-based transport capacity constraints and the second where the most recent comparable day is used as approximation. The similar hour methods, probabilistic methods and correlation methods also consist of two methods each, namely the method where the mean market clearing price is used as forecast and the method where the median market clearing price is used as forecast. For the results of these methods (which can be found in the previous chapter) two types of error measurement methods are used to specify the approximation error, namely the mean absolute error and the root mean squared error. When comparing these error measurement methods both methods overall lead to similar results, in sense of boxplots, mean and standard deviation of the error measurement methods. For this reason we only continued to use the mean absolute error.



The main conclusion of this research is that, among the tested methods, the currently used methods can best be used to approximate the flow-based capacity constraints, when forecasting one day in the future. When forecasting a day further in the future the currently used methods can also be used, but the probabilistic method will give a forecast as good as the currently used methods, especially when using the set of characteristics of variation 2 & 6 (even when forecasting only one day ahead this methods work as good as the currently used methods). The similar hour methods do not lead to good approximations for the Netherlands, Germany and France, but lead to relatively good approximations for Belgium. The correlation method always is ranked last, so it is not wisely to use this approximation method to forecast the market clearing prices. When we compare the results of the two approximation-sets we cannot draw a general conclusion of which one works better. If we compare the results between the methods when using the mean v.s. the median, a general conclusion cannot be drawn. These results remain inconclusive. The results of the comparison in separate volume indicate that the currently used methods give relatively good results, the similar hour methods and the probabilistic methods behave somewhat worse than the currently used methods and again the correlation method is ranked last.

## 7.2 Recommendations for Further Research

As the time for this research is limited, not all research around this subject has been done. This section contains some suggestions for further research.

Based on the conclusions and some superficially investigation of the data, we suspect that the data of the flow-based transport capacity constraints generated in the time that the project was not yet live, differs from the data since the moment that the CWE market areas actually started using the flow-based transport capacity constraints for real. The flow-based domains from the parallel run are possibly different and the amount of critical branches might be lower. This might have an influence on the outcome of this research and might explain why the currently used methods work better. In order to find out whether this presumption is true or not, a more broad investigation of the data is necessary. Further research can be investigating the data and based on this investigation implement a time-dependent method, which favours hours in the close past.

One of the conclusions of this research is that the benchmark method (the currently used method) works pretty well, better than or similar to the proposed methods. In order to find a method which beats the benchmark, it can be useful to find out when the benchmark fails to work properly. This can be done from two angles. The first angle is to find out in which equivalence class the benchmark does not work properly. Another angle is based on the presumption that the data of the parallel run is different than the real life data. In this case a possible method can be to take into account the distance in time of the hour that is used to approximate the flow-based domain. For example this method favours hours that are quite recent compared to hours far in the past.

This research focuses mainly on the critical branches, while the flow-based domain is



only represented by a subset of those critical branches. Consequently the critical branches which are no bounds for the flow-based domain are of less interest and never directly constrain the outcome of the net positions. Another way to investigate the problem can be to focus only on the critical branches that delimit the feasible flow-based domain.

A conclusion from the results is that the correlation method does not have real added value. In general the probabilistic methods, the similar hour methods and also the currently used methods seem to give better results. The reason for this might be that, the correlation methods add too many constraints which only have a high correlation based on one or a few appearances. This might increase the number of constraints substantially, which can make the flow-based domain too small. Consequently, less trades between countries are possible. This results in diverging prices. In further research a more sophisticated method can be designed to use correlation in the probabilistic methods, which can result in a better approximation. Another option is to change the decisions which adjust the probability that a certain constraint is added to the sample set.

The part of the NWE region which now uses the flow-based model to calculate the transport capacity constraints only consists of the CWE countries. When the flow-based model turns out to indicate more available transport capacity for the market, the flow-based model might be implemented in the entire NWE region. In further research it should be investigated what influence this has on the methods that approximate the flow-based domain.

### 7.3 Discussion and Remarks on this Research

In this research we simplified some aspects to be able to achieve an approximation within a reasonable computation time. The main adjustments are done to the market coupling method. Originally this method calculates the market clearing price and accepted orders for all countries in the NWE region. However, we do not take into account the countries which are not in the CWE region. The reason for this is that the flow-based method is only implemented for the CWE countries. The non-CWE countries still work with the ATC-model. The supply and demand orders of non-CWE countries in the NWE region can have an influence on the prices in the CWE region. In our case this problem is solved by inserting a buy- or sell-order in the set of real orders for respectively import- and export to non-CWE countries, with respectively the maximum allowable price and minimum allowable price. This implies that they are always accepted. By this, the influence of countries outside of the CWE is still taken into account in the calculation process of the market coupling method. Furthermore, when adjusting the market coupling method to a method for only the CWE countries we also deleted certain types of (complicated) orders.

Another adjustment that is made to the market coupling method is that it is simplified to an integer linear program, while the actual algorithm is a quadratic linear program. The orders that cause this, are the interpolated orders. Those are orders that start being accepted at a certain price  $p_0$  and are fully accepted at another price  $p_1$ . In our program we do not allow interpolated orders, and by this the model result in an integer linear program.



The approximation-set is representative of the entire data-set, but when calculating the number of similar hours for each instance of the approximation-set, the minimum of this number is still 56 for variation set 6 and 23 for variation set 2 & 6. This means that the results only hold when using the approximation methods for an hour  $h$  which has at least 23 similar hours in the data-set.

Intuitiveness

Another aspect of the market coupling method that is ignored, is the so-called *intuitiveness*. An intuitive market outcome means that each market area with price  $p_A$  is not exporting to market areas with a price  $p_B < p_A$  and not importing from a market area with a price  $p_B > p_A$ . In the market coupling method used in this research, a counter intuitive market outcome is possible but happens seldom. In order to prevent counter intuitive market outcomes extra constraints have to be added. In the actual market coupling method these constraints are present, in our simplified market coupling method they are not present. The reason why we left these constraints out is because the method becomes more complex and needs more calculation time. Also a counter intuitive market outcome almost never occurs. In the case that it occurs the prices are in the simplified market coupling method slightly higher for some market areas and slightly lower for other market areas. This adjustment does not bias this research, because the methods used to approximate the flow-based transport capacity constraints are mutually compared. This means that when ranking the different methods, a counter intuitive outcome does not influence the ranks.



# Symbols and Abbreviations

## Symbols

Definition	Symbol	Set	Value	Unity	Page
Approximation-Set	$as$	$AS$	-	-	38
Set of All Hours	$h$	$AH$	-	-	41
ATC-bound	$ATC$	$\mathbb{R}^+$	-	MegaWatt (MW)	9
Branch	$b$	$B$	-	-	13
Branch combined with scenario $n$	$b(n)$	$B^{crit}, B_n, B^{total}$	-	-	24
Constraints	-	$C_{transport}, C_{balance}, C^{h,fb}, C^{h,fb,for}$	-	-	13
Collected flow-based constraints of $SH(h)$	-	$C_{coll}^{h,fb}$	-	-	53
Corner point	$cp$	$CP$	-	-	33
Characteristic	$Ch$	$Ch^h$	-	-	29
Day	$d$	$D$	-	-	25
Distance Function	$Dis(\mathcal{P})$	-	-	-	33
Domain	$\mathcal{D}$	-	-	-	29
Flow of electricity	$f$	$\mathbb{R}^+$	-	MegaWatt (MW)	11
Reference flow	$f^{ref}$	$\mathbb{R}^+$	-	MegaWatt (MW)	26
Maximum allowable flow	$f^{Max}$	$\mathbb{R}^+$	-	MegaWatt (MW)	86
Mapping function	$g$	-	-	-	13
TSO function	$g_{tso}$ or $\tilde{g}_{tso}$	-	-	-	13
Good Hours	$gh$	$GH$	-	-	41
Hour	$h$	$\{1, 2, \dots, 24\}$	-	-	14
Set of representatives of $b(n)$ in $SH(h)$	-	$H(h, b(n))$	-	-	53
Maximum current	$I^{Max}$	$\mathbb{R}^+$	-	Ampere (A)	86
Number of Characteristics	$k$	$\mathbb{N}$	-	-	29

Table 7.1: Explanation of the symbols and where to find them.



Definition	Symbol	Set	Value	Unity	Page
Number of domain splits	$l$	$\mathbb{N}$	-	-	40
Market area	$m$	$M$	-	-	12
Outage Scenario	$n$	$N$	-	-	23
Net position	$NEP$	$\mathbb{R}$	-	MegaWatt (MW)	13
Order	$o$	$O$	-	-	12
Set with Correlations	$pc$	$PC$	-	-	57
Polytope	$\mathcal{P}$	-	-	-	30
Price	$p$	$\mathbb{R}$	-	Euro per MegaWatt hour (€/MWh)	11
PTDF-value	$PTDF$	$\mathbb{R}$	-	-	15
Quantity	$q$	$\mathbb{R}^+$	-	MegaWatt hour (MWh)	12
25 % Quartile	$Q_1$	-	-	-	63
75 % Quartile	$Q_3$	-	-	-	63
Cross-border branch parameter	$r$	$\{-1, 0, 1\}$	-	-	16
RAM-value	$RAM$	$\mathbb{R}$	-	-	17
Remaining Hours	$rh$	$RH$	-	-	44
Set of constraints	-	$S_j$	-	-	30
Sensitivity Function	$Sen(b(n))$	-	-	-	25
Similar Hours of hour $h$	$sh$	$SH(h)$	-	-	44
Number of different branches $b(n)$ in the data-set	$t$	-	-	-	30
Training-Set	$ts$	$TS$	-	-	37
Number of approximations	$u$	-	-	-	59
Nodes	$v$	$V$	-	-	15
Number of sample sets	$w$	$\mathbb{N}$	100	-	53
Reference voltage	$W$	$\mathbb{R}^+$	-	KiloVolt (kV)	86
Level of acceptance	$x$	$(0, 1)$	-	-	12
Real market clearing price	$\bar{y}$	-	-	-	59
Approximated market clearing price	$y_i$	-	-	-	59
Number of constraints in hour $h$	$z(h)$	$\mathbb{N}$	-	-	53
Confidence level	$\alpha$	-	0.05	-	71
Number of constraints already in the sample set	$\beta$	-	-	-	56

Table 7.2: Table 7.1 continued.



Definition	Symbol	Set	Value	Unity	Page
Number of hours in the data-set	$\gamma$	-	-	-	30
Boxplot coefficient	$\zeta$	-	1.5	-	64
Correlation coefficient	$\rho$	-	-	-	56
Correlation with maximal absolute value	$\rho_{sel}$	-	-	-	57

Table 7.3: Table 7.1 continued.

## Abbreviations

Abbreviation	Explanation	Page
APX	Amsterdam Power Exchange	4
ATC	Available Transfer Capacity	9
B	Currently used method, most recent day	44
BE	Belgium	6
BS	Currently used method, most recent comparable day	44
C	Correlation method	56
CWE	Central Western Europe	6
d	Median used as forecast	52
DE	Germany	6
EC	External Constraint	27
FAV	Final Adjustment Value	86
FR	France	6
FRM	Flow Reliability Margin	87
GSK	Generation Shift Key	27
ILP	Integer Linear Program	12
LU	Luxembourg	6
m	Mean used as forecast	51
MAE	Mean Absolute Error	59
MAPE	Mean Absolute Percentage Error	60
MCP	Market Clearing Price	5
MSE	Mean Squared Error	60
NEP	Net Position	9
NL	the Netherlands	6
NWE	North Western Europe	8
P	Probabilistic method	52
PTDF	Power Transfer Distribution Factor	15
RA	Remedial Action	27
RAM	Remaining Available Margin	17
RMSE	Root Mean Squared Error	60
S	Similar hour method	51

Table 7.4: Explanations of the abbreviations and where to find them.



Abbreviation	Explanation	Page
S. D.	Standard Deviation	<a href="#">69</a>
TSO	Transmission System Operator	<a href="#">6</a>
6	Approximation-set of variation 6	<a href="#">42</a>
26	Approximation-set of variation 2 & 6	<a href="#">42</a>

Table 7.5: Table [7.4](#) continued.



# Appendix A

## Appendix

### A.1 The Different Test-Sets

#### A.1.1 Training-Set

The training-set described in Section 4.5.2 contains the following hours:

Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
21-May-15	8	20-Jun-15	22	24-Jul-15	24	29-Aug-15	14	02-Oct-15	21
21-May-15	11	21-Jun-15	16	25-Jul-15	1	29-Aug-15	17	03-Oct-15	6
21-May-15	12	21-Jun-15	22	25-Jul-15	2	30-Aug-15	22	03-Oct-15	8
22-May-15	1	22-Jun-15	8	25-Jul-15	11	31-Aug-15	8	03-Oct-15	9
22-May-15	5	22-Jun-15	11	25-Jul-15	23	31-Aug-15	12	03-Oct-15	21
22-May-15	13	22-Jun-15	16	26-Jul-15	1	31-Aug-15	14	04-Oct-15	2
22-May-15	19	22-Jun-15	20	26-Jul-15	4	31-Aug-15	24	04-Oct-15	4
22-May-15	24	23-Jun-15	1	26-Jul-15	19	01-Sep-15	4	04-Oct-15	9
23-May-15	6	23-Jun-15	16	27-Jul-15	12	01-Sep-15	6	04-Oct-15	13
23-May-15	11	23-Jun-15	19	27-Jul-15	13	01-Sep-15	9	04-Oct-15	24
23-May-15	12	23-Jun-15	24	27-Jul-15	15	01-Sep-15	16	05-Oct-15	7
23-May-15	13	24-Jun-15	24	27-Jul-15	17	01-Sep-15	24	05-Oct-15	9
23-May-15	18	25-Jun-15	5	28-Jul-15	2	02-Sep-15	15	05-Oct-15	23
24-May-15	14	25-Jun-15	20	28-Jul-15	8	03-Sep-15	13	05-Oct-15	24
24-May-15	16	25-Jun-15	22	28-Jul-15	9	03-Sep-15	18	06-Oct-15	12
25-May-15	2	26-Jun-15	1	28-Jul-15	10	03-Sep-15	22	06-Oct-15	21
25-May-15	8	26-Jun-15	3	29-Jul-15	5	04-Sep-15	7	06-Oct-15	22
25-May-15	15	26-Jun-15	4	29-Jul-15	21	04-Sep-15	9	07-Oct-15	4
26-May-15	2	26-Jun-15	15	30-Jul-15	11	04-Sep-15	16	08-Oct-15	2
26-May-15	7	26-Jun-15	21	30-Jul-15	12	04-Sep-15	19	08-Oct-15	4
26-May-15	10	27-Jun-15	5	30-Jul-15	16	04-Sep-15	21	09-Oct-15	12
27-May-15	2	27-Jun-15	8	31-Jul-15	5	05-Sep-15	5	09-Oct-15	15
27-May-15	3	27-Jun-15	19	31-Jul-15	13	05-Sep-15	11	09-Oct-15	22

Table A.1: Dates and Hours of the training-set.



Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
27-May-15	10	28-Jun-15	7	01-Aug-15	7	06-Sep-15	4	10-Oct-15	2
27-May-15	11	28-Jun-15	13	01-Aug-15	9	06-Sep-15	15	10-Oct-15	8
27-May-15	20	29-Jun-15	23	01-Aug-15	10	07-Sep-15	6	11-Oct-15	5
27-May-15	23	30-Jun-15	15	02-Aug-15	10	07-Sep-15	11	11-Oct-15	8
28-May-15	13	30-Jun-15	23	02-Aug-15	13	07-Sep-15	19	12-Oct-15	3
29-May-15	23	01-Jul-15	3	02-Aug-15	18	08-Sep-15	2	12-Oct-15	7
30-May-15	3	01-Jul-15	12	03-Aug-15	7	08-Sep-15	7	12-Oct-15	15
30-May-15	6	01-Jul-15	21	03-Aug-15	12	08-Sep-15	11	12-Oct-15	16
30-May-15	22	01-Jul-15	23	03-Aug-15	16	08-Sep-15	20	12-Oct-15	17
31-May-15	10	02-Jul-15	9	04-Aug-15	3	08-Sep-15	21	13-Oct-15	3
31-May-15	15	02-Jul-15	21	04-Aug-15	8	08-Sep-15	22	13-Oct-15	12
31-May-15	24	04-Jul-15	6	05-Aug-15	9	08-Sep-15	24	13-Oct-15	22
01-Jun-15	2	04-Jul-15	13	05-Aug-15	11	09-Sep-15	5	14-Oct-15	12
01-Jun-15	4	05-Jul-15	12	05-Aug-15	16	09-Sep-15	7	14-Oct-15	15
01-Jun-15	5	05-Jul-15	23	05-Aug-15	18	09-Sep-15	14	14-Oct-15	18
01-Jun-15	19	06-Jul-15	8	06-Aug-15	5	09-Sep-15	16	14-Oct-15	21
01-Jun-15	21	06-Jul-15	10	06-Aug-15	10	09-Sep-15	19	14-Oct-15	22
01-Jun-15	23	06-Jul-15	11	06-Aug-15	11	09-Sep-15	21	15-Oct-15	2
02-Jun-15	1	06-Jul-15	15	06-Aug-15	13	10-Sep-15	16	15-Oct-15	3
02-Jun-15	10	06-Jul-15	18	07-Aug-15	10	11-Sep-15	2	15-Oct-15	7
04-Jun-15	1	07-Jul-15	3	07-Aug-15	12	11-Sep-15	4	15-Oct-15	8
04-Jun-15	4	07-Jul-15	5	08-Aug-15	18	11-Sep-15	8	16-Oct-15	7
04-Jun-15	9	07-Jul-15	7	08-Aug-15	21	11-Sep-15	9	16-Oct-15	8
04-Jun-15	13	07-Jul-15	9	09-Aug-15	11	11-Sep-15	21	16-Oct-15	9
04-Jun-15	14	07-Jul-15	15	09-Aug-15	24	11-Sep-15	22	16-Oct-15	11
05-Jun-15	2	07-Jul-15	20	10-Aug-15	3	12-Sep-15	4	16-Oct-15	24
05-Jun-15	4	07-Jul-15	22	10-Aug-15	8	13-Sep-15	2	17-Oct-15	2
05-Jun-15	21	07-Jul-15	23	10-Aug-15	15	13-Sep-15	3	17-Oct-15	10
06-Jun-15	2	08-Jul-15	20	11-Aug-15	7	13-Sep-15	5	17-Oct-15	11
06-Jun-15	3	09-Jul-15	9	11-Aug-15	10	13-Sep-15	9	17-Oct-15	24
06-Jun-15	14	09-Jul-15	13	11-Aug-15	18	13-Sep-15	19	18-Oct-15	17
06-Jun-15	16	09-Jul-15	22	11-Aug-15	24	14-Sep-15	4	18-Oct-15	18
06-Jun-15	20	10-Jul-15	8	12-Aug-15	20	14-Sep-15	16	18-Oct-15	21
07-Jun-15	2	10-Jul-15	19	13-Aug-15	9	15-Sep-15	2	18-Oct-15	24
07-Jun-15	3	11-Jul-15	4	13-Aug-15	14	15-Sep-15	3	19-Oct-15	3
07-Jun-15	13	11-Jul-15	7	14-Aug-15	15	15-Sep-15	17	19-Oct-15	9
07-Jun-15	17	11-Jul-15	9	14-Aug-15	19	15-Sep-15	21	19-Oct-15	12
08-Jun-15	10	11-Jul-15	21	15-Aug-15	10	15-Sep-15	24	19-Oct-15	13
08-Jun-15	18	11-Jul-15	24	15-Aug-15	23	16-Sep-15	4	19-Oct-15	24
08-Jun-15	23	12-Jul-15	6	16-Aug-15	1	16-Sep-15	11	20-Oct-15	20
09-Jun-15	13	13-Jul-15	14	16-Aug-15	4	16-Sep-15	14	20-Oct-15	21
09-Jun-15	18	13-Jul-15	22	16-Aug-15	6	17-Sep-15	17	20-Oct-15	23
09-Jun-15	22	14-Jul-15	14	16-Aug-15	17	17-Sep-15	18	21-Oct-15	7
09-Jun-15	24	15-Jul-15	2	17-Aug-15	6	17-Sep-15	24	21-Oct-15	13

Table A.2: Table A.1 continued.



Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
10-Jun-15	7	15-Jul-15	3	17-Aug-15	12	18-Sep-15	3	21-Oct-15	18
10-Jun-15	9	15-Jul-15	7	17-Aug-15	15	18-Sep-15	21	23-Oct-15	6
10-Jun-15	11	15-Jul-15	9	17-Aug-15	21	20-Sep-15	2	24-Oct-15	2
10-Jun-15	18	15-Jul-15	11	17-Aug-15	24	21-Sep-15	2	24-Oct-15	5
10-Jun-15	22	15-Jul-15	17	18-Aug-15	3	21-Sep-15	18	24-Oct-15	12
10-Jun-15	23	15-Jul-15	19	18-Aug-15	14	22-Sep-15	10	24-Oct-15	13
11-Jun-15	14	16-Jul-15	3	19-Aug-15	15	22-Sep-15	22	24-Oct-15	21
11-Jun-15	24	16-Jul-15	17	19-Aug-15	23	23-Sep-15	13	25-Oct-15	1
12-Jun-15	1	16-Jul-15	18	20-Aug-15	9	23-Sep-15	17	25-Oct-15	4
12-Jun-15	2	16-Jul-15	22	20-Aug-15	14	24-Sep-15	14	25-Oct-15	5
12-Jun-15	9	17-Jul-15	14	20-Aug-15	23	24-Sep-15	16	25-Oct-15	20
12-Jun-15	11	18-Jul-15	10	22-Aug-15	9	24-Sep-15	20	26-Oct-15	4
12-Jun-15	22	18-Jul-15	20	22-Aug-15	13	25-Sep-15	4	26-Oct-15	7
13-Jun-15	3	19-Jul-15	4	23-Aug-15	10	25-Sep-15	7	26-Oct-15	11
13-Jun-15	6	19-Jul-15	9	23-Aug-15	19	25-Sep-15	16	26-Oct-15	14
13-Jun-15	16	19-Jul-15	11	24-Aug-15	3	25-Sep-15	18	26-Oct-15	16
14-Jun-15	10	19-Jul-15	15	24-Aug-15	4	26-Sep-15	22	27-Oct-15	1
14-Jun-15	18	20-Jul-15	3	24-Aug-15	18	27-Sep-15	8	27-Oct-15	5
15-Jun-15	1	20-Jul-15	24	24-Aug-15	23	27-Sep-15	17	27-Oct-15	14
15-Jun-15	2	21-Jul-15	8	25-Aug-15	8	27-Sep-15	19	27-Oct-15	19
15-Jun-15	7	21-Jul-15	11	25-Aug-15	9	28-Sep-15	9	27-Oct-15	23
15-Jun-15	10	22-Jul-15	5	25-Aug-15	17	29-Sep-15	8	28-Oct-15	6
16-Jun-15	7	22-Jul-15	11	25-Aug-15	22	29-Sep-15	9	28-Oct-15	8
16-Jun-15	13	22-Jul-15	12	25-Aug-15	24	30-Sep-15	8	28-Oct-15	11
17-Jun-15	12	23-Jul-15	5	26-Aug-15	6	30-Sep-15	17	29-Oct-15	1
18-Jun-15	18	23-Jul-15	17	26-Aug-15	9	30-Sep-15	19	29-Oct-15	3
18-Jun-15	23	23-Jul-15	19	27-Aug-15	2	30-Sep-15	20	29-Oct-15	5
19-Jun-15	18	23-Jul-15	23	27-Aug-15	11	01-Oct-15	12	29-Oct-15	9
19-Jun-15	22	24-Jul-15	1	27-Aug-15	13	01-Oct-15	20	29-Oct-15	12
20-Jun-15	3	24-Jul-15	5	27-Aug-15	15	02-Oct-15	2	29-Oct-15	18
20-Jun-15	8	24-Jul-15	6	28-Aug-15	5	02-Oct-15	4	29-Oct-15	23
20-Jun-15	18	24-Jul-15	8	28-Aug-15	14	02-Oct-15	11	29-Oct-15	24
20-Jun-15	20	24-Jul-15	20	28-Aug-15	24	02-Oct-15	12	30-Oct-15	11

Table A.3: Table A.1 continued.

### A.1.2 Approximation-Set Variation 6

The approximation-set described in Section 4.2 contains the following hours:

Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
05-Jan-15	5	03-Apr-15	19	08-May-15	5	10-Jul-15	22	03-Sep-15	17
09-Jan-15	7	04-Apr-15	10	15-May-15	5	15-Jul-15	15	08-Sep-15	6
10-Jan-15	3	07-Apr-15	5	22-May-15	4	16-Jul-15	2	09-Sep-15	18

Table A.4: Dates and Hours of the approximation-set Variation 6.



Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
16-Jan-15	22	08-Apr-15	6	22-May-15	16	16-Jul-15	6	10-Sep-15	20
19-Jan-15	19	11-Apr-15	5	28-May-15	1	16-Jul-15	23	11-Sep-15	11
22-Jan-15	11	11-Apr-15	9	29-May-15	19	28-Jul-15	1	19-Sep-15	23
31-Jan-15	8	16-Apr-15	2	03-Jun-15	23	31-Jul-15	11	25-Sep-15	15
01-Feb-15	2	20-Apr-15	20	06-Jun-15	21	02-Aug-15	19	03-Oct-15	10
01-Feb-15	6	21-Apr-15	19	07-Jun-15	21	02-Aug-15	22	08-Oct-15	15
10-Feb-15	21	22-Apr-15	18	12-Jun-15	5	04-Aug-15	22	09-Oct-15	2
15-Feb-15	4	23-Apr-15	6	13-Jun-15	1	06-Aug-15	4	14-Oct-15	16
15-Feb-15	14	23-Apr-15	9	15-Jun-15	13	10-Aug-15	23	16-Oct-15	18
17-Feb-15	19	26-Apr-15	4	22-Jun-15	22	17-Aug-15	14	18-Oct-15	9
05-Mar-15	3	26-Apr-15	18	25-Jun-15	11	20-Aug-15	5	18-Oct-15	15
08-Mar-15	1	30-Apr-15	8	25-Jun-15	17	21-Aug-15	12	20-Oct-15	14
09-Mar-15	13	02-May-15	22	25-Jun-15	24	21-Aug-15	22	24-Oct-15	18
13-Mar-15	23	05-May-15	6	28-Jun-15	10	22-Aug-15	4	31-Oct-15	16
18-Mar-15	2	05-May-15	24	29-Jun-15	21	28-Aug-15	8	02-Nov-15	3
24-Mar-15	5	07-May-15	16	30-Jun-15	12	28-Aug-15	18	05-Nov-15	1
24-Mar-15	15	07-May-15	20	04-Jul-15	21	03-Sep-15	5	06-Nov-15	9

Table A.5: Table A.4 continued.

### A.1.3 Approximation-Set Variation 6 & 2

The approximation-set described in Section 4.2 contains the following hours:

Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
06-Jan-15	9	26-Apr-15	16	30-Jun-15	12	10-Aug-15	14	10-Sep-15	1
14-Jan-15	3	02-May-15	2	05-Jul-15	2	11-Aug-15	12	12-Sep-15	3
10-Feb-15	16	07-May-15	21	05-Jul-15	7	20-Aug-15	6	13-Sep-15	11
11-Feb-15	4	07-May-15	24	05-Jul-15	11	20-Aug-15	17	23-Sep-15	22
17-Feb-15	2	21-May-15	22	07-Jul-15	6	20-Aug-15	19	01-Oct-15	23
25-Feb-15	5	27-May-15	1	15-Jul-15	16	20-Aug-15	20	03-Oct-15	23
01-Mar-15	9	04-Jun-15	18	15-Jul-15	21	20-Aug-15	24	08-Oct-15	13
07-Mar-15	1	04-Jun-15	24	16-Jul-15	1	21-Aug-15	1	08-Oct-15	19
11-Mar-15	1	11-Jun-15	9	16-Jul-15	10	21-Aug-15	13	08-Oct-15	21
17-Mar-15	9	14-Jun-15	11	21-Jul-15	23	21-Aug-15	18	08-Oct-15	24
21-Mar-15	4	16-Jun-15	5	23-Jul-15	12	21-Aug-15	19	16-Oct-15	15
23-Mar-15	4	17-Jun-15	5	01-Aug-15	4	22-Aug-15	19	16-Oct-15	23
23-Mar-15	22	18-Jun-15	5	03-Aug-15	15	29-Aug-15	4	17-Oct-15	4
28-Mar-15	4	21-Jun-15	23	03-Aug-15	23	29-Aug-15	5	17-Oct-15	17
04-Apr-15	1	25-Jun-15	15	05-Aug-15	17	29-Aug-15	12	18-Oct-15	10
04-Apr-15	3	26-Jun-15	5	06-Aug-15	9	29-Aug-15	15	20-Oct-15	13
08-Apr-15	14	29-Jun-15	8	06-Aug-15	15	29-Aug-15	24	21-Oct-15	5
09-Apr-15	20	29-Jun-15	12	06-Aug-15	24	30-Aug-15	21	31-Oct-15	24
20-Apr-15	10	30-Jun-15	1	07-Aug-15	3	30-Aug-15	23	01-Nov-15	14

Table A.6: Dates and Hours of the approximation-set Variation 6 &amp; 2.



Date	Hour	Date	Hour	Date	Hour	Date	Hour	Date	Hour
21-Apr-15	4	30-Jun-15	2	07-Aug-15	17	07-Sep-15	24	06-Nov-15	9

Table A.7: Table A.6 continued.

## A.2 Detailed Calculation of the RAM-value

The RAM-value is an important factor in the transport capacity constraints. The RAM-value is dependent of certain factors. In order to be able to calculate the RAM-value, the TSOs need the factors described below.

### Maximum Allowable Flow ( $f_{b(n)}^{Max}$ )

The most important factor for the calculation of the RAM-value is the maximum allowable flow. The maximum allowable flow is the physical capacity of a branch multiplied by certain factors to calculate the actual maximum allowable flow. The relation between the maximum allowable flow ( $f_{b(n)}^{Max}$ ) and these factors is described in the following equation:

$$f_{b(n)}^{Max} = \frac{\sqrt{3} \cdot I_{b(n)}^{Max} \cdot W \cdot \cos(\phi)}{1000},$$

where  $I_{b(n)}^{Max}$  is the maximum allowable current (in Ampere)(explained below),  $W$  is fixed to a reference voltage (e.g. 225 kV or 400 kV) given in kV and  $\cos(\phi)$  is 1. The maximum allowable flow is given in MW.

### Maximum Allowable Current ( $I_{b(n)}^{Max}$ )

The maximum current on a critical branch is necessary to calculate the maximum flow ( $f_{b(n)}^{max}$ ) on  $b(n)$ . The  $I_{b(n)}^{Max}$  is the physical capacity of  $b(n)$ . This value is given in Ampère and can be dependent on the temperature. The TSOs fix the  $I_{b(n)}^{Max}$  per season but can review the value if the announced temperature is much higher or lower than foreseen.

### Final Adjustment Value ( $FAV_{b(n)}$ )

Another import factor for the calculation of the RAM-value is the Final Adjustment Value ( $FAV_{b(n)}$ ). The final adjustment value is a way to model experience and complex operational skills that cannot be expressed in a ‘normal’ transport constraint. The final adjustment value increases or reduces the RAM-value to maintain the security of the electrical network. The calculation of the final adjustment value is not very precise documented and differs for the different TSOs.



### Flow Reliability Margin ( $FRM_{b(n)}$ )

The last factor which influences the RAM-value is the Flow Reliability Margin ( $FRM_{b(n)}$ ), which is a fixed parameter and should be updated at least once a year. This flow reliability margin models the uncertainties that arise in the flow-based parameter calculation process. The flow reliability margin decreases the RAM-value to ensure the security of the electrical network. This is done by comparing the difference in the realised flow and the estimated flow over a certain period of time. The calculation process is summarised in Figure A.1.

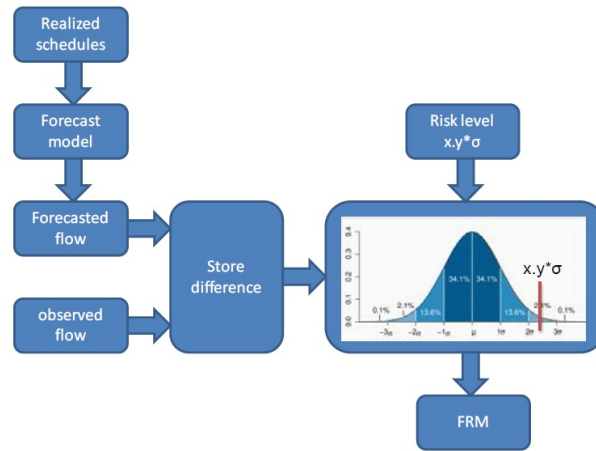


Figure A.1: Summary of the calculation of the FRM-value.

Source: [3].

The purpose of the FRM-value aims to cover the following effects:

- Uncertainty in the renewable energy forecast;
- uncertainty in the demand forecast;
- uncertainty in the generation forecast;
- topological changes;
- internal trades conducted within the market area;
- trades outside the CWE region;
- unintended flow deviations.

The flow reliability margin is different from the external constraints in the sense that flow reliability margin is based on the past and can only prevent situations that happened in the past. It cannot prevent situations that are so extreme that they never happened before. The FRM-value is a fixed value for each branch  $b(n)$ .



## A.3 Results

This appendix contains the results of the different methods researched in this thesis.

### A.3.1 Results of the Similar Hours Method

In the similar hour method,  $|SH(h)|$  market clearing prices per market area are calculated, those market clearing prices are used to calculate the mean and median, which results in the forecast of the market clearing prices. In order to generate the results, the forecast of the market clearing prices is compared to the actual market clearing price, with two different error calculation methods. Table A.8 contains the results when testing this method on the approximation-set. The first column of the table indicates whether the approximation-set of variation 6 or the approximation-set of variation 2 and 6 is used, also here is denoted whether the mean or median is used as approximation and which error calculation is used. The error calculation methods that are used are the mean absolute error (MAE) and the (root) mean squared error ((R)(MSE)). The standard deviation (S.D.) of the different error calculation methods is also given, which indicates how much spread there is between the different instances of the approximation set. The numbers in the table are in euros.

Var. 6	NL	BE	DE	FR
Mean MAE	€6.78	€9.06	€5.18	€5.45
S.D. Mean MAE	€5.48	€9.24	€4.10	€4.54
Median MAE	€6.29	€8.24	€5.12	€5.24
S.D. Median MAE	€6.24	€9.26	€4.80	€5.17
Mean RMSE	€8.70	€12.91	€6.60	€7.08
S.D. Mean MSE	€109.97	€364.96	€61.93	€75.75
Median RMSE	€8.84	€12.36	€7.01	€7.35
S.D. Median MSE	€138.58	€365.29	€88.61	€102.99

Var. 2 & 6	NL	BE	DE	FR
Mean MAE	€6.73	€8.75	€4.22	€4.94
S.D. Mean MAE	€5.69	€8.28	€3.70	€4.63
Median MAE	€7.02	€8.80	€ 4.16	€5.00
S.D. Median MAE	€6.60	€9.06	€4.42	€4.84
Mean RMSE	€8.80	€12.01	€5.60	€6.75
S.D. Mean MSE	€128.83	€288.90	€51.57	€100.79
Median RMSE	€9.61	€12.60	€6.06	€6.94
S.D. Median MSE	€150.36	€335.08	€69.49	€87.00

Table A.8: Results of the similar hour method.



### A.3.2 Results of the Probabilistic Method

The presentation of the results of the probabilistic method is similar to the presentation of the results of the similar hour method.

Var. 6	NL	BE	DE	FR
Mean MAE	€4.01	€6.61	€0.97	€1.93
S.D. Mean MAE	€4.76	€11.10	€1.12	€3.19
Median MAE	€3.49	€8.09	€1.04	€2.08
S.D. Median MAE	€4.68	€10.63	€1.31	€3.38
Mean RMSE	€6.21	€13.69	€1.52	€3.79
S.D. Mean MSE	€83.88	€685.06	€4.75	€58.32
Median RMSE	€5.82	€12.48	€1.63	€3.88
S.D. Median MSE	€66.94	€519.85	€6.29	€60.01

Var. 2 & 6	NL	BE	DE	FR
Mean MAE	€4.54	€8.12	€0.92	€1.85
S.D. Mean MAE	€5.35	€13.41	€1.29	€2.68
Median MAE	€4.58	€6.87	€0.88	€1.70
S.D. Median MAE	€6.17	€12.22	€1.45	€2.86
Mean RMSE	€6.99	€15.62	€1.58	€3.25
S.D. Mean MSE	€96.70	€902.75	€8.17	€42.14
Median RMSE	€7.66	€13.97	€1.69	€3.31
S.D. Median MSE	€115.93	€642.66	€9.70	€46.11

Table A.9: Results of the probabilistic method.

### A.3.3 Results of the Correlation Method

The presentation of the results of the correlation method is similar to the presentation of the results of the similar hour method and the probabilistic method.

### A.3.4 Results of the Currently Used Method

The currently used method exists of two variations. The first variation is using the ‘most recent’ flow-based domain as approximation and the second variation is using the ‘most recent comparable’ flow-based domain as approximation.



Var. 6	NL	BE	DE	FR
Mean MAE	€6.74	€15.27	€1.53	€3.14
S.D. Mean MAE	€8.79	€22.13	€1.75	€4.46
Median MAE	€6.08	€14.81	€1.50	€3.30
S.D. Median MAE	€9.39	€28.49	€1.99	€4.72
Mean RMSE	€11.04	€26.80	€2.31	€5.43
S.D. Mean MSE	€393.47	€2365.97	€11.41	€107.72
Median RMSE	€11.15	€31.98	€2.49	€5.74
S.D. Median MSE	€431.20	€4114.50	€14.49	€107.77

Var. 2 & 6	NL	BE	DE	FR
Mean MAE	€5.95	€12.42	€1.05	€2.47
S.D. Mean MAE	€7.28	€18.97	€1.61	€3.21
Median MAE	€6.31	€13.89	€1.07	€2.53
S.D. Median MAE	€8.73	€27.34	€1.91	€3.49
Mean RMSE	€9.37	€22.59	€1.91	€4.04
S.D. Mean MSE	€210.85	€1668.08	€13.41	€54.33
Median RMSE	€10.73	€30.54	€2.18	€4.29
S.D. Median MSE	€293.89	€3853.06	€18.39	€59.28

Table A.10: Results of the correlation method calculated.

Var. 6	NL	BE	DE	FR	Var. 2 & 6	NL	BE	DE	FR
MAE	€3.09	€3.90	€0.72	€1.20	MAE	€4.26	€6.36	€0.68	€1.50
S.D. MAE	€5.19	€8.01	€1.15	€2.06	S.D. MAE	€5.79	€9.73	€0.98	€2.05
RMSE	€6.02	€8.88	€1.35	€2.38	RMSE	€7.16	€11.58	€1.18	€2.53
S.D. MSE	€90.91	€281.05	€5.02	€20.68	S.D. MSE	€118.94	€356.49	€3.45	€17.36

Table A.11: Results of the most recent currently used method.

Var. 6	NL	BE	DE	FR
MAE	€3.79	€4.84	€0.97	€1.63
S.D. MAE	€5.81	€8.51	€1.43	€2.66
RMSE	€6.91	€9.75	€1.72	€3.11
S.D. MSE	€127.32	€284.90	€8.25	€30.51

Var. 2 & 6	NL	BE	DE	FR
MAE	€4.44	€5.63	€0.74	€1.58
S.D. MAE	€6.56	€7.89	€1.20	€2.21
RMSE	€7.89	€9.66	€1.40	€2.70
S.D. RMSE	€178.32	€204.78	€5.11	€16.49

Table A.12: Results of the most recent comparable currently used method.



## A.4 Visualisation of the Results continued

Label	Explanation
B	Currently used method, most recent day.
BS	Currently used method, most recent comparable day.
S	Similar hour method.
P	Probabilistic method.
C	Correlation method.
d	Median used as forecast.
m	Mean used as forecast.
6	Approximation-set of variation 6.
26	Approximation-set of variation 2 & 6

Table A.13: Explanation of the labels used in the following figures.

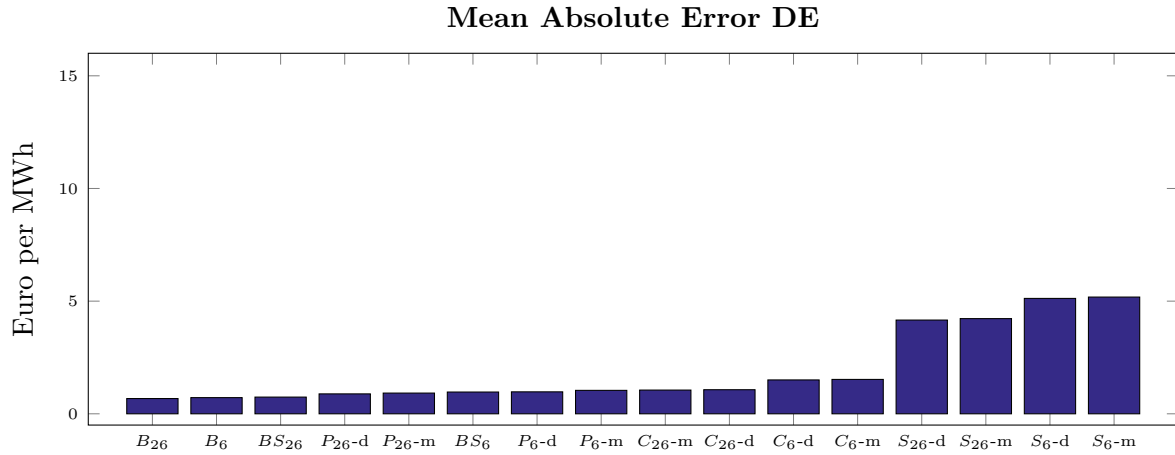


Figure A.2: Mean Absolute Error results for Germany.

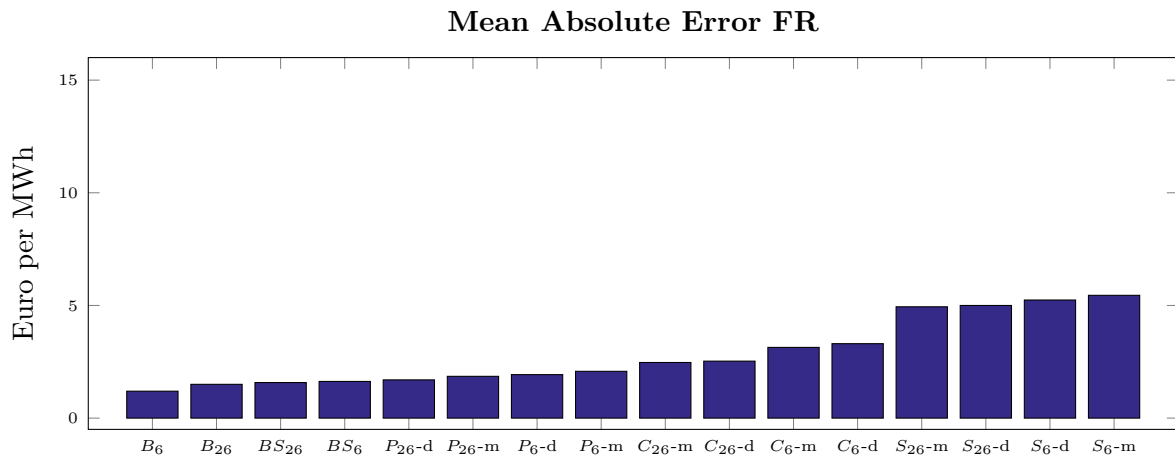


Figure A.3: Mean Absolute Error results for France



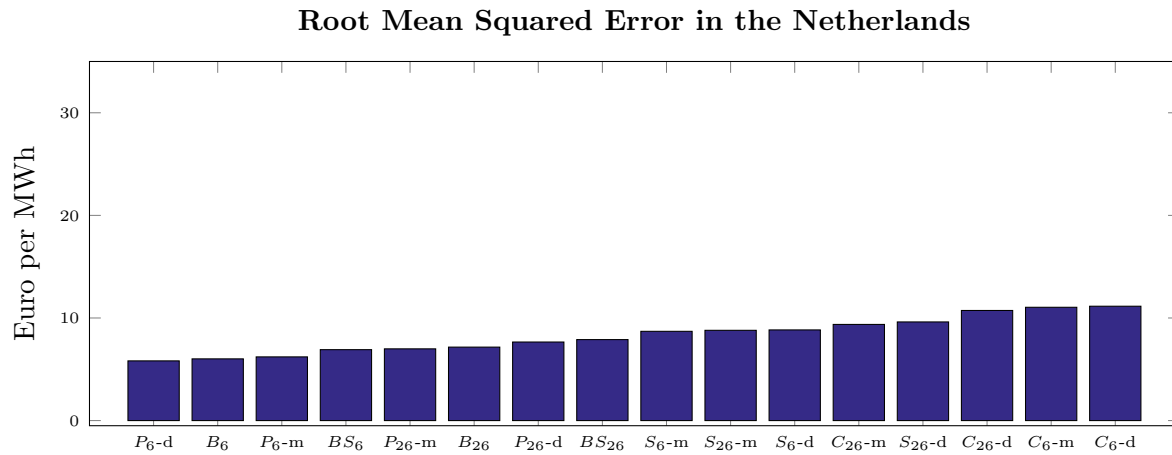


Figure A.4: Root Mean Squared Error results for the Netherlands.

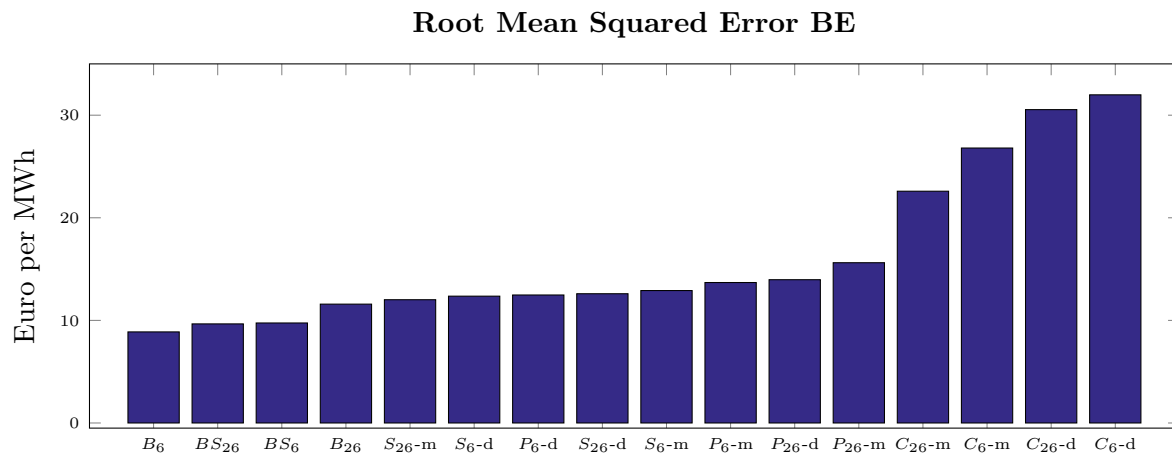


Figure A.5: Root Mean Squared Error results for Belgium.

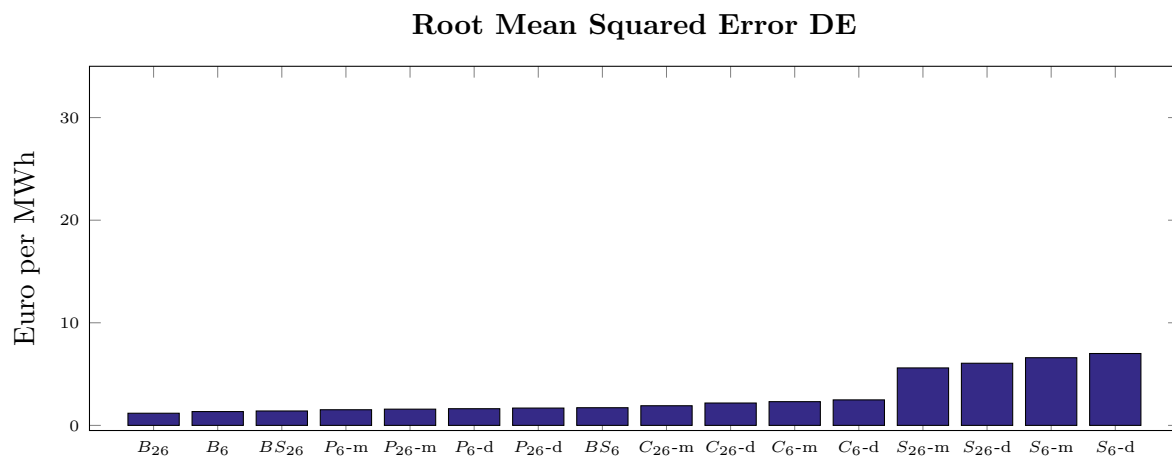


Figure A.6: Root Mean Squared Error results for Germany.



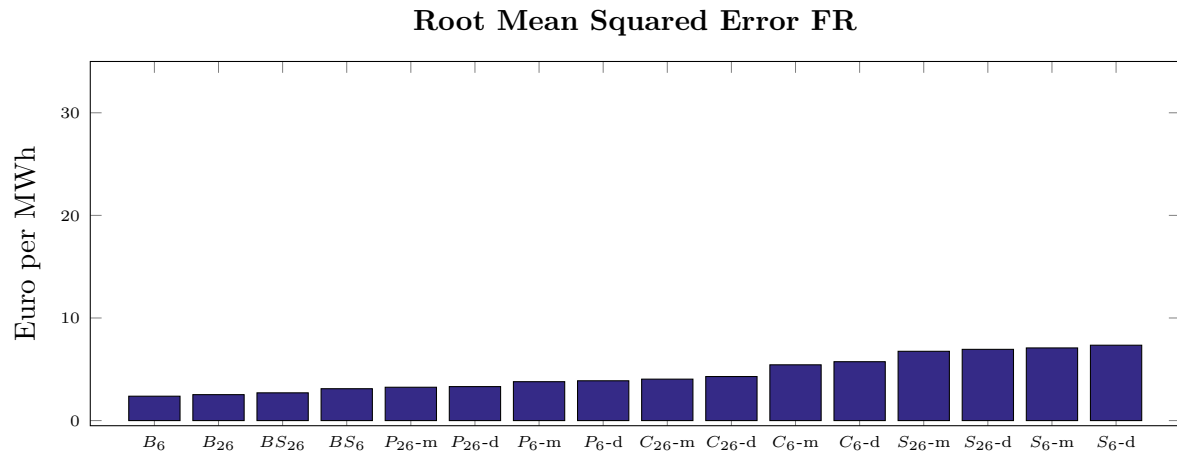


Figure A.7: Root Mean Squared Error results for France.

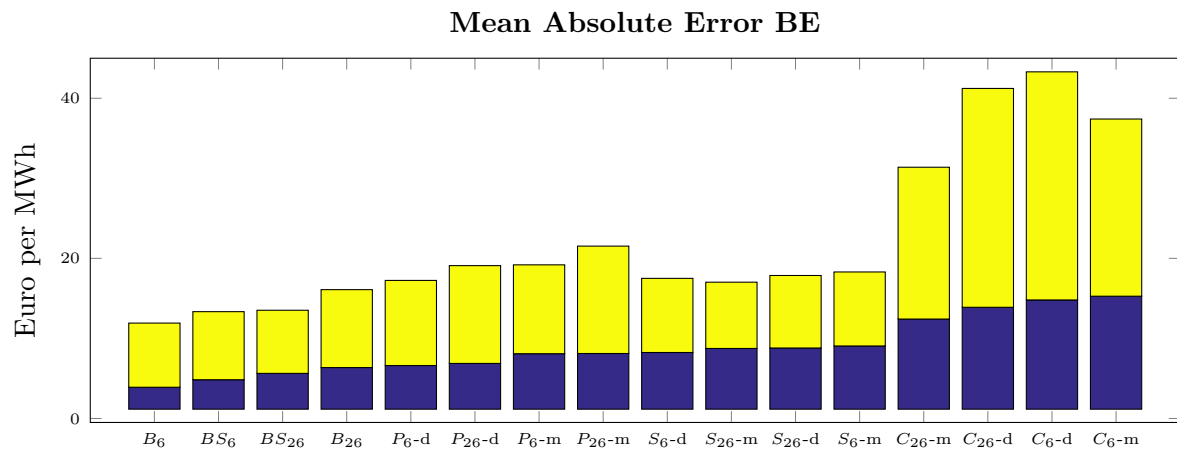


Figure A.8: MAE with added standard deviation for BE.

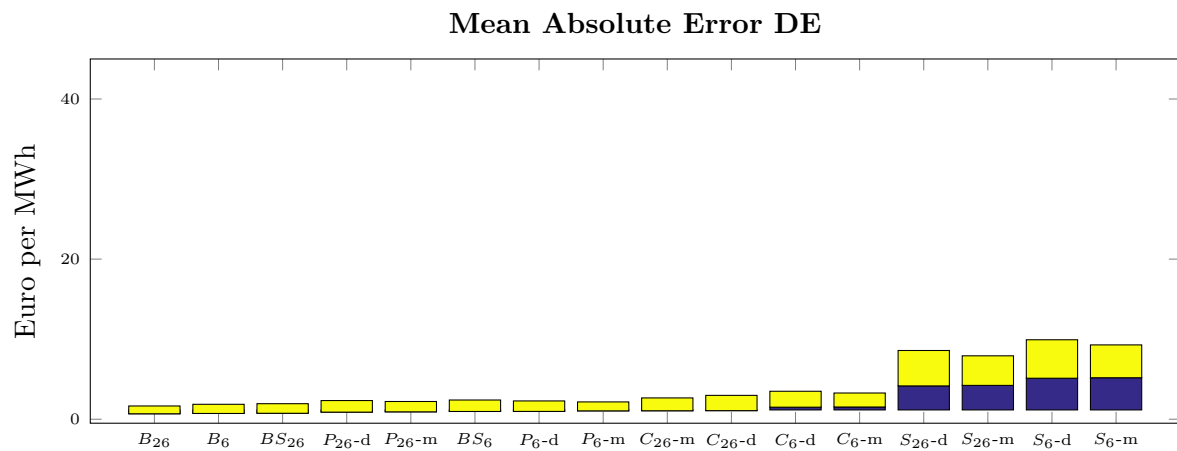


Figure A.9: MAE with added standard deviation for DE.



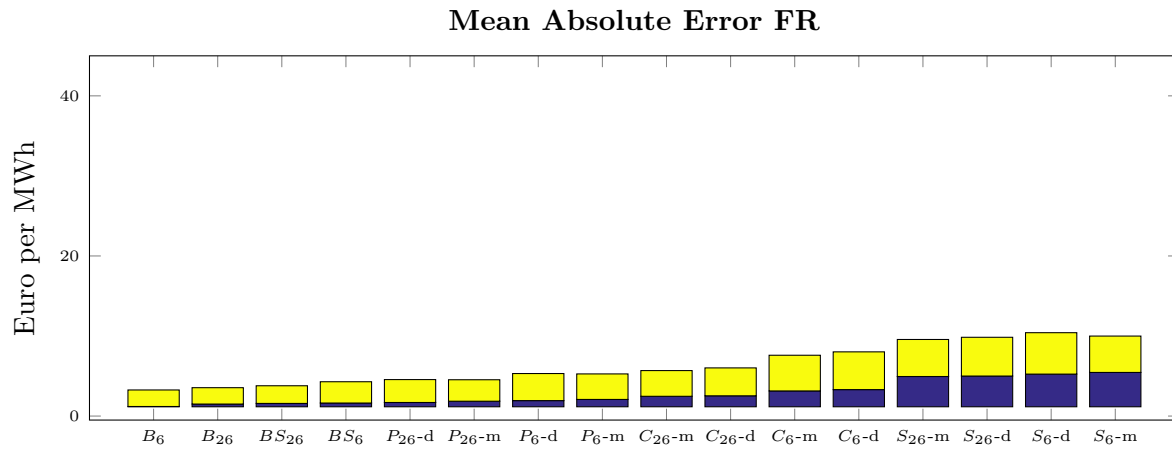


Figure A.10: MAE with added standard deviation for FR.

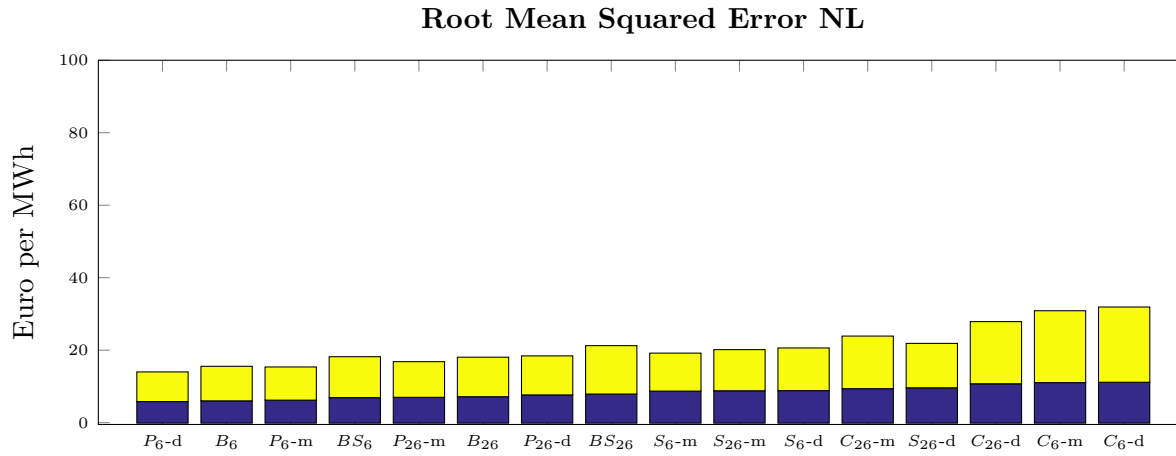


Figure A.11: RMSE with added standard deviation for NL.

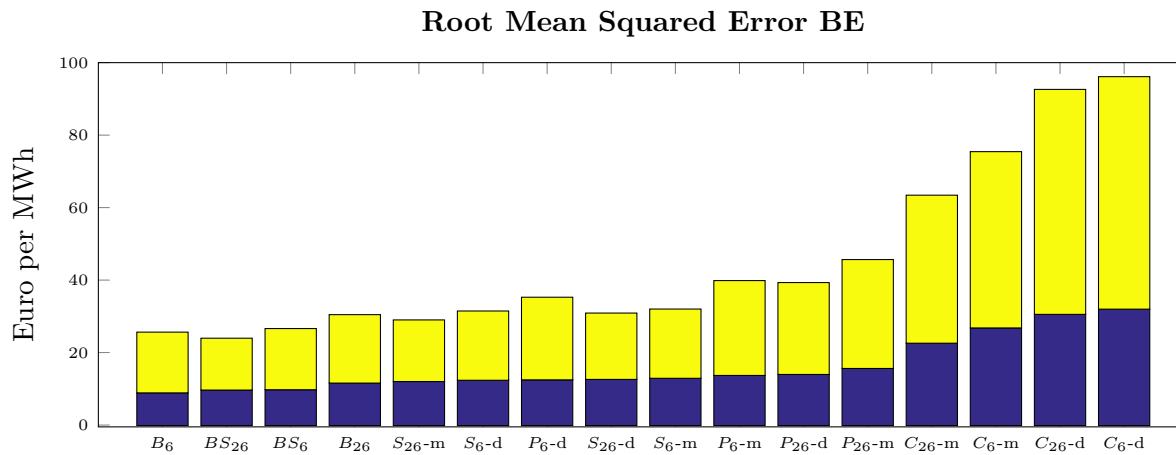


Figure A.12: RMSE with added standard deviation for BE.



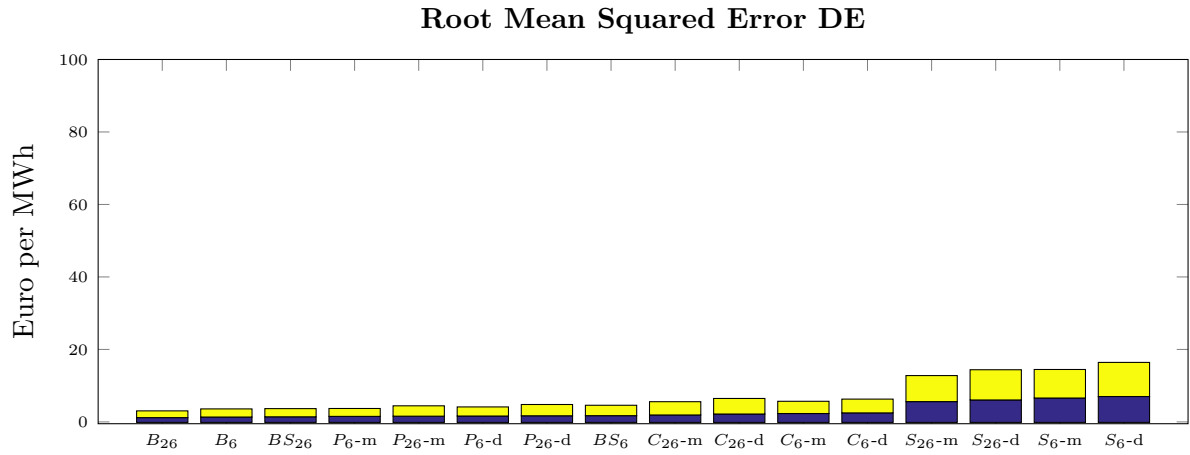


Figure A.13: RMSE with added standard deviation for DE.

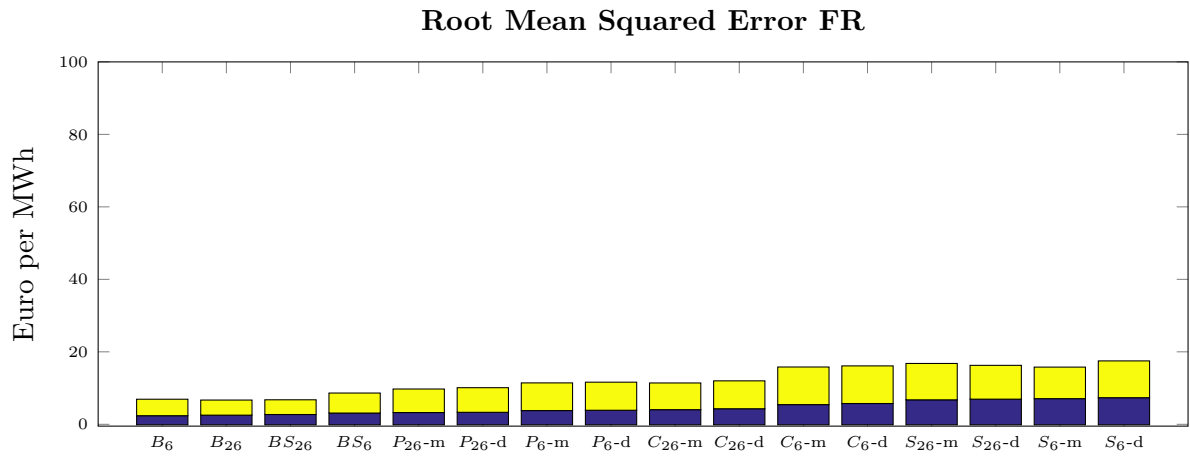


Figure A.14: RMSE with added standard deviation for FR.

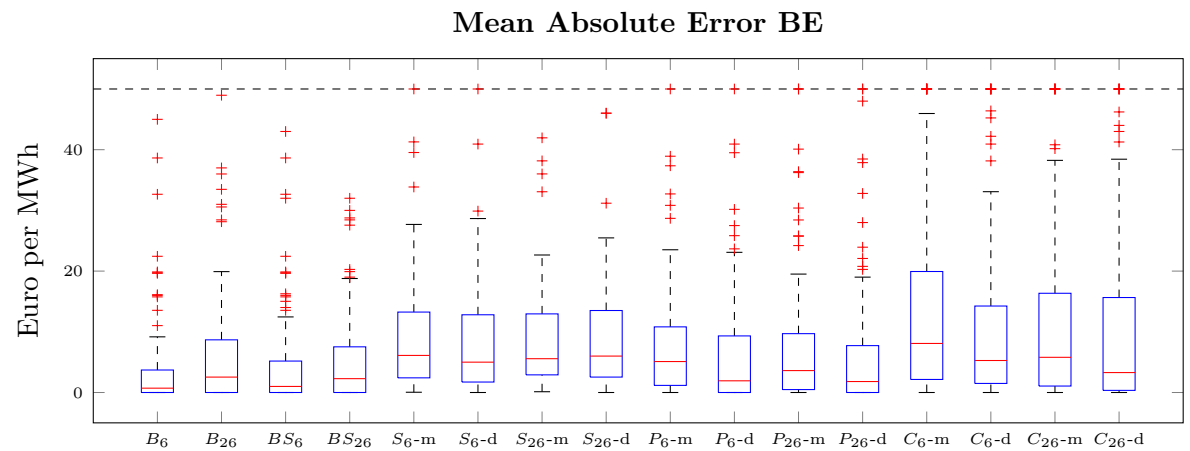


Figure A.15: Boxplots of the results calculated with the MAE in BE.



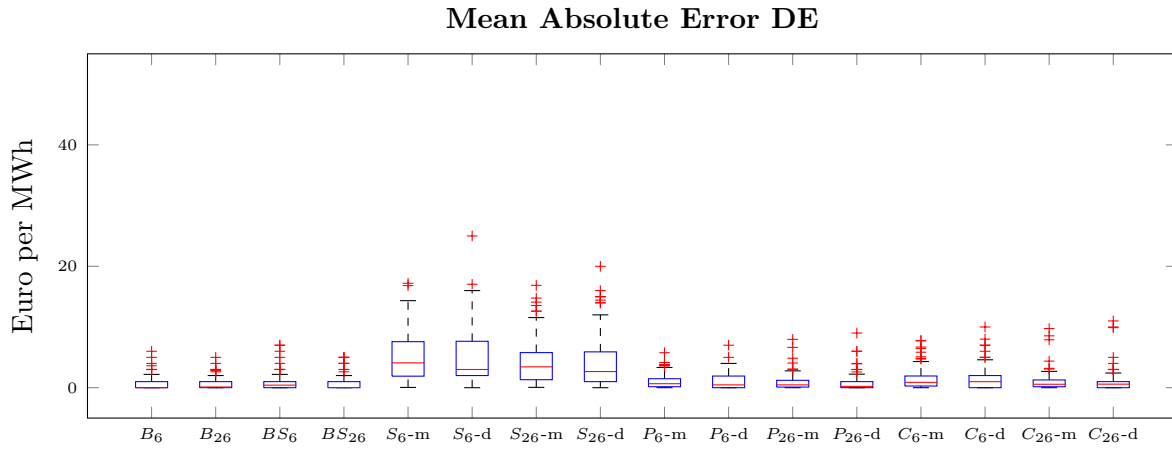


Figure A.16: Boxplots of the results calculated with the MAE in DE.

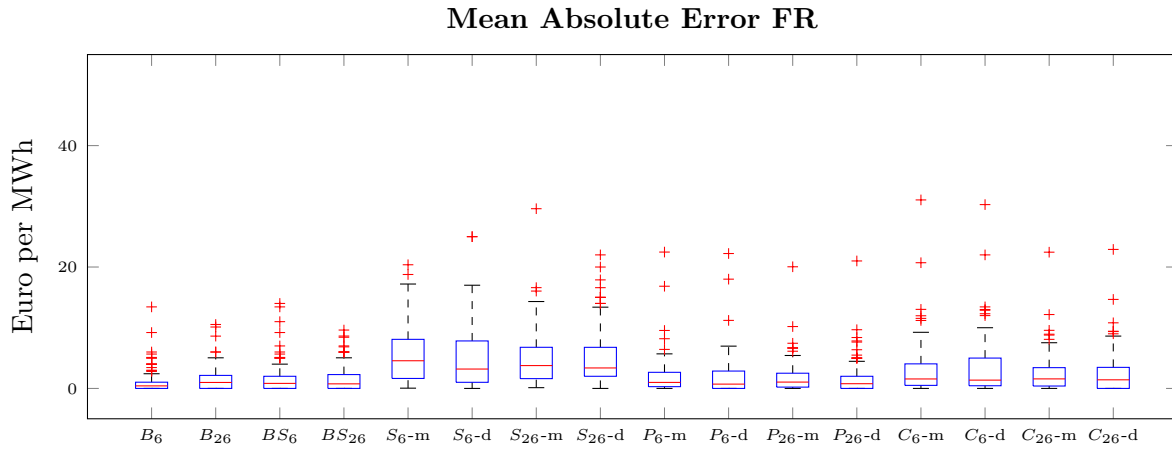


Figure A.17: Boxplots of the results calculated with the MAE in FR.

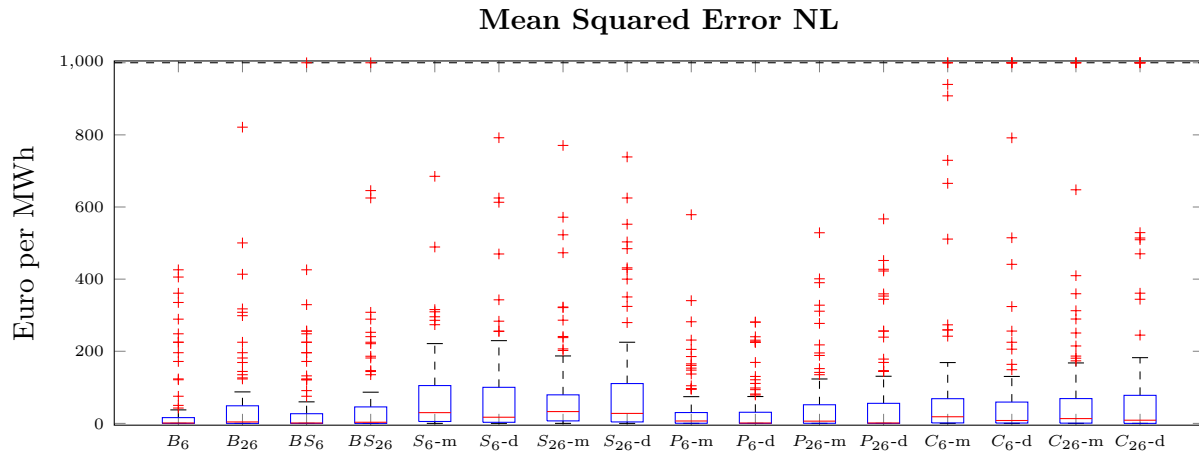


Figure A.18: Boxplots of the results calculated with the MSE in NL.



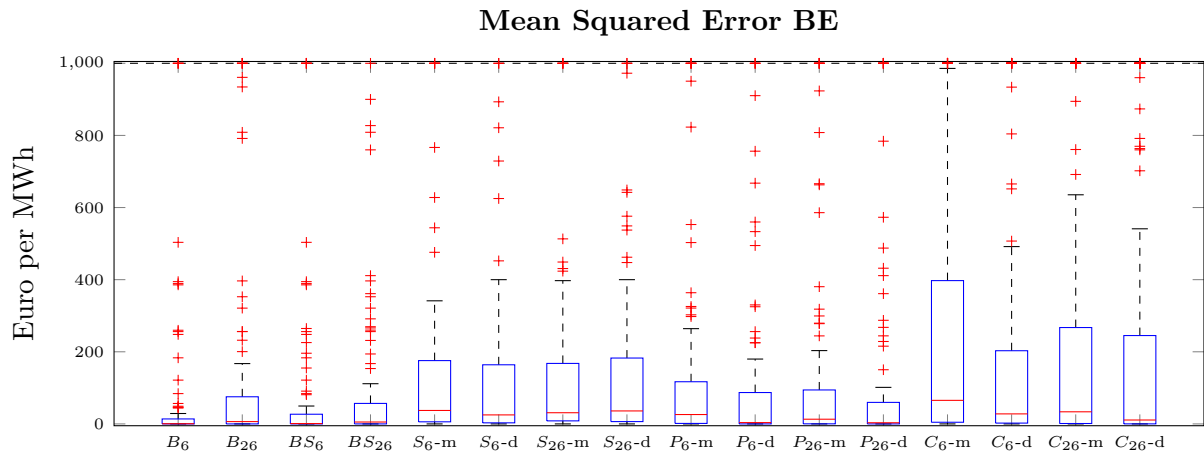


Figure A.19: Boxplots of the results calculated with the MSE in BE.

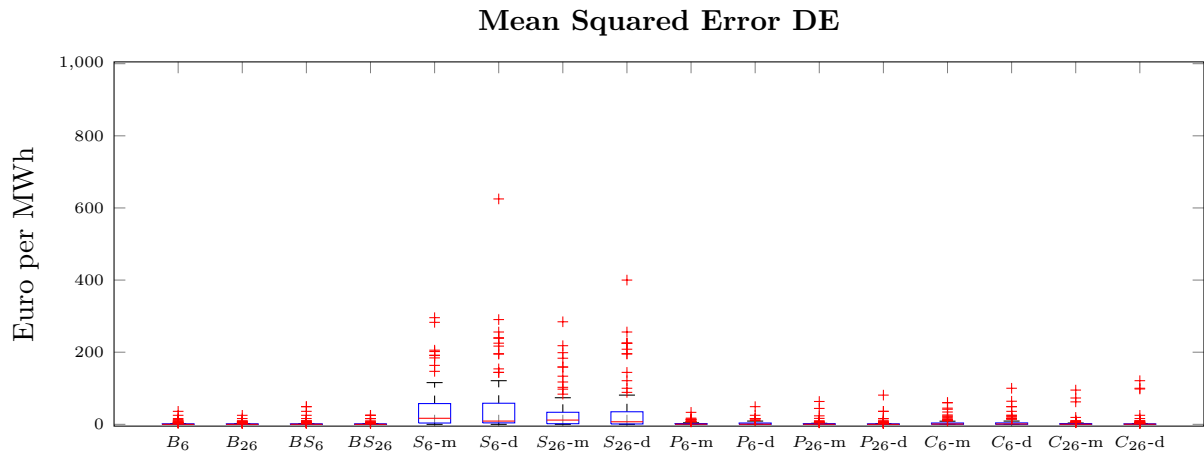


Figure A.20: Boxplots of the results calculated with the MSE in DE.

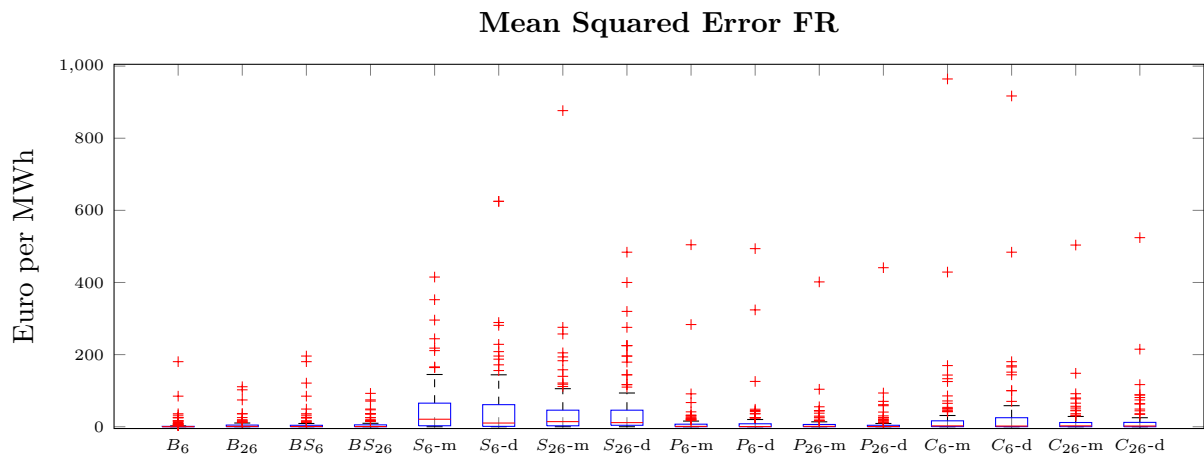


Figure A.21: Boxplots of the results calculated with the MSE in FR.



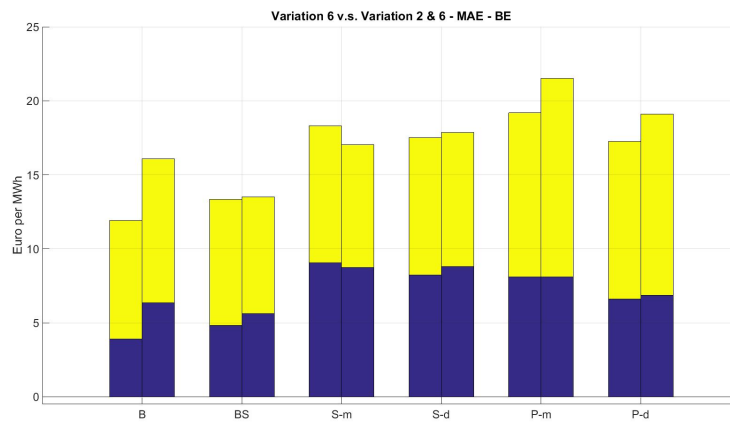


Figure A.22: Using variation 6 or variation 2 & 6 for BE. Left is variation 6, Right is 2 & 6.

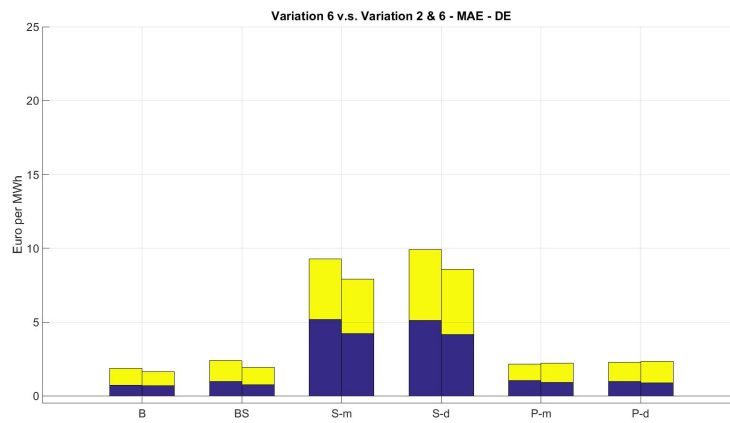


Figure A.23: Using variation 6 or variation 2 & 6 for DE. Left is variation 6, Right is 2 & 6.

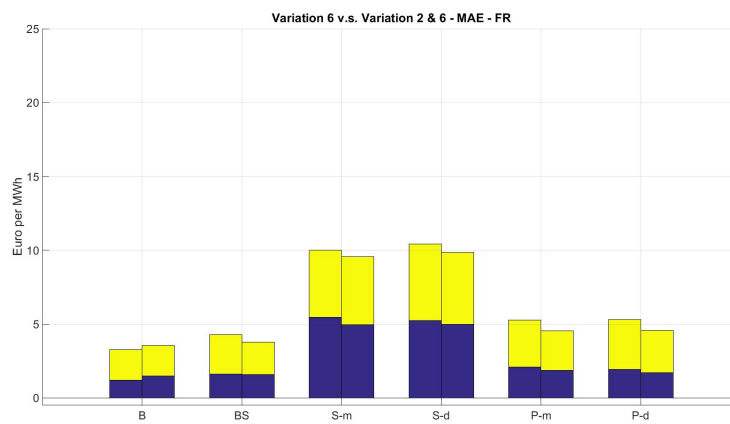


Figure A.24: Using variation 6 or variation 2 & 6 for FR. Left is variation 6, Right is 2 & 6.



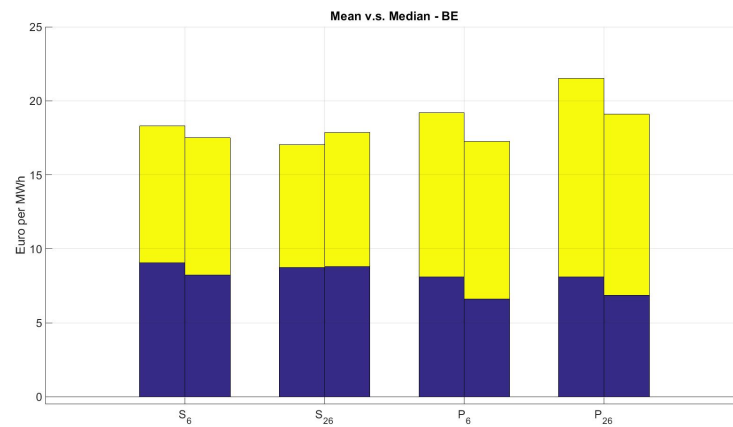


Figure A.25: Using mean or median for BE. Left is mean, right is median.

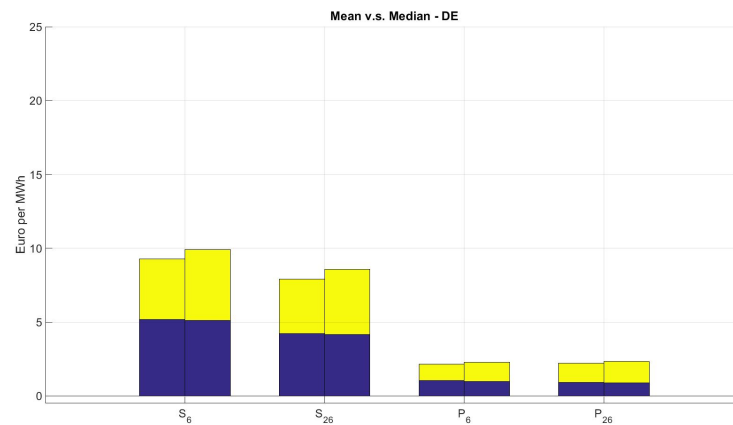


Figure A.26: Using mean or median for DE. Left is mean, right is median.

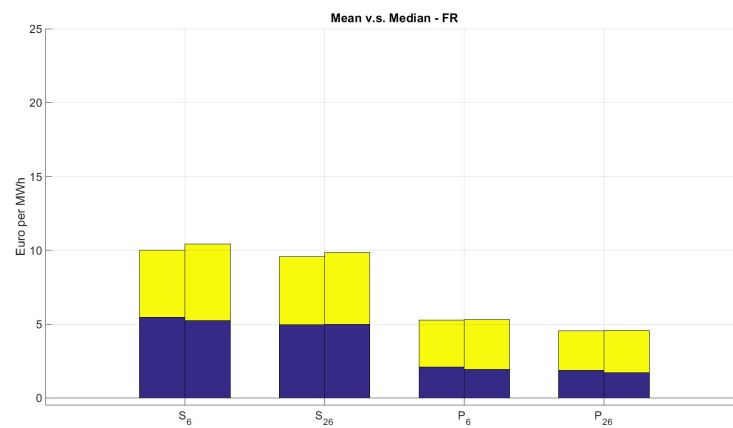


Figure A.27: Using mean or median for FR. Left is mean, right is median.



## A.5 Results Statistical Tests

	$B$				$BS$				$S - m$				$S - d$			
	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR
$B$	-	-	-	-												
$BS$	0	0	0	0	-	-	-	-								
$S - m$	1	1	1	1	1	1	1	1	-	-	-	-				
$S - d$	1	1	1	1	1	1	1	1	0	0	0	0	-	-	-	-
$P - m$	0	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1
$P - d$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
$C - m$	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1
$C - d$	0	1	0	1	0	1	0	1	0	0	1	1	0	0	1	1
	$P - m$				$P - d$				$C - m$				$C - d$			
	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR	NL	BE	DE	FR
$B$																
$BS$																
$S - m$																
$S - d$																
$P - m$	-	-	-	-												
$P - d$	0	0	0	0	-	-	-	-								
$C - m$	1	1	1	1	1	1	1	1	-	-	-	-				
$C - d$	1	1	0	1	1	1	0	1	0	0	0	0	-	-	-	-

Table A.14: Results of the Wilcoxon signed-rank test when testing the different methods with the results when using approximation-set of variation 2 & 6.



# Bibliography

- [1] M. Adamec, M. Indrakova, and P. Pavlatka. Market coupling and price coordination between power exchanges. *10th IAAEE European Conference, Vienna Austria*, 7 September 2009. [http://www.aaee.at/2009-IAEE/uploads/fullpaper\\_iaee09/P\\_521\\_Pavlatka\\_Pavel\\_7-Sep-2009,%208:59.pdf](http://www.aaee.at/2009-IAEE/uploads/fullpaper_iaee09/P_521_Pavlatka_Pavel_7-Sep-2009,%208:59.pdf).
- [2] M. Aguado, R. Bourgeois, J.Y. Bourmaud, J. van Casteren, M.A. Ceratto, M. Jäkel, B. Malfliet, C. Mestdag, P. Noury, M. Pool, W. van den Reek, M. Rohleder, P.H. Schavemaker, S. Scolari, O. Weis, and J. Wolpert. Flow-based market coupling in the Central Western European region - on the eve of implementation -. *Cigre*, pages C5-204, 2012. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.7550&rep=rep1&type=pdf>.
- [3] Amprion, APX, Belpex, Creos, Elia, EPEXSPOT, RTE, TenneT, and TRANSNET BW. Documentation of the CWE FB MC solution. *As basis for the formal approval-request*, 1 August 2014. <http://www.casc.eu/media/140801%20CWE%20FB%20MC%20Approval%20document.pdf>.
- [4] Amprion, APX, Belpex, Creos, Elia, EPEXSPOT, RTE, TenneT, and TRANSNET BW. CWE flow-based MC launch on 20 may 2015 following NRAs approval. *Press release*, 24 April 2015. [https://www.apxgroup.com/wp-content/uploads/Confirmation-Go-live\\_20May.pdf](https://www.apxgroup.com/wp-content/uploads/Confirmation-Go-live_20May.pdf).
- [5] Amprion, APX-Endex, Belpex, Creos, Elia, EPEXSPOT, RTE, TenneT, and EnBW. Questions & Answers., 12 December 2011. [http://www.apxgroup.com/wp-content/uploads/FAQ\\_-\\_update\\_121211.pdf](http://www.apxgroup.com/wp-content/uploads/FAQ_-_update_121211.pdf).
- [6] APX, Belpex, EPEX SPOT, Nord Pool Spot, 50Hertz, Amprion, Creos, Elia, Energinet.dk, Fingrid, National Grid, RTE, Statnett, Svenska Kraftnät, Tennet B.V. (Netherlands), Tennet GmbH (Germany), and TransnetBW. NWE Day-Ahead Price Coupling: Questions & Answers. [http://www.belpex.be/wp-content/uploads/NWE\\_QA.pdf](http://www.belpex.be/wp-content/uploads/NWE_QA.pdf).
- [7] APX, Belpex, EPEX SPOT, Nord Pool Spot, 50Hertz, Amprion, Creos, Elia, Energinet.dk, Fingrid, National Grid, RTE, Statnett, Svenska Kraftnät, Tennet B.V. (Netherlands), Tennet GmbH (Germany), and TransnetBW. NWE price coupling to launch 4 february 2014. *Press release*, 18 December 2013. [https://www.apxgroup.com/wp-content/uploads/20131218\\_NWE-confirmation-with-SWE-announcement.pdf](https://www.apxgroup.com/wp-content/uploads/20131218_NWE-confirmation-with-SWE-announcement.pdf).



- [8] J. Boury. Methods for the determination of flow-based capacity parameters: description, evaluation and improvements. *Thesis*, 2014 - 2015. KU Leuven. [http://www.scriptiebank.be/sites/default/files/webform/scriptie/Boury\\_Jonas\\_0.pdf](http://www.scriptiebank.be/sites/default/files/webform/scriptie/Boury_Jonas_0.pdf).
- [9] M. Burger, B. Graeber, and G. Schindlmayr. *Managing Energy Risk: An Integrated view on power and other energy markets*. John Wiley & Sons, Ltd, 2007.
- [10] European Commission. A framework strategy for a resilient energy union with a forward-looking climate change policy., 25 February 2015. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:80:FIN>.
- [11] Creos, Elia, TransnetBW GmbH, TenneT TSO GmbH, TenneT TSO B.V, RTE, Amprion, Austrian Power Grid AG, Elektro Slovenija, Independent Power Transmission Operator S.A., Swissgrid, Terna, Energinet, and Statnett. <http://www.casc.eu>.
- [12] Creos, Elia, TransnetBW GmbH, TenneT TSO GmbH, TenneT TSO B.V, RTE, Amprion, Austrian Power Grid AG, Elektro Slovenija, Independent Power Transmission Operator S.A., Swissgrid, Terna, Energinet.dk, and Statnett. NWE Day-Ahead price coupling, algorithm. *NWE Forum Presentation*, 26 September 2012. <https://www.epexspot.com/document/22759/NWE%20Stakeholder%20forum%20-%20presentation%20algorithm%2C%2026%20Sept%202012>.
- [13] M. Herceg, M. Kvasnica, C.N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17–19 2013. <http://control.ee.ethz.ch/~mpt>.
- [14] P.G. Hoel. *Introduction to Mathematical Statistics*. Wiley, 1984.
- [15] R.J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2013. <https://www.otexts.org/book/fpp>.
- [16] R. Lowry. *Concepts and Applications of Inferential Statistics*. Web Resource, 2014. <http://doer.col.org/handle/123456789/4853>.
- [17] A. Marien, P. Luickx, A. Tirez, and D. Woitrin. Importance of design parameters on flow-based market coupling implementation. *10th International Conference on the European Energy Market*, pages 1–8, 2013.
- [18] J.H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, 2014. <http://www.biostathandbook.com/>.
- [19] A.H. Murphy. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather & Forecasting*, pages 281–293, 20 January 1993. <http://journals.ametsoc.org/doi/pdf/10.1175/1520-0434%281993%29008%3C0281%3AWIAGFA%3E2.0.CO%3B2>.
- [20] K.T.S. Oldham. The doctrine of description: Gustav Kirchhoff, classical physics, and the purpose of all science in 19th-century germany. *University of California, Berkeley*, 2008.



- [21] M. Shahidehpour, H. Yamin, and Z. Li. *Market operations in electric power systems: forecasting, scheduling, and risk management*. Wiley, 2002. <http://ieeexplore.ieee.org/book/0471463949.pdf>.
- [22] EPEX Spot, APX, Belpex, Nord Pool Spot, OMIE, Mercatoelettrico (GME), and OTE. EUPHEMIA: public description. *PCR Market Coupling Algorithm*, 2 October 2013. [http://www.nordpoolspot.com/globalassets/download-center/pcr/euphemia-public-description\\_nov2013.pdf](http://www.nordpoolspot.com/globalassets/download-center/pcr/euphemia-public-description_nov2013.pdf).
- [23] EPEX Spot, APX, Belpex, Nord Pool Spot, OMIE, Mercatoelettrico (GME), OTE, and Europex. EUPHEMIA: description and functioning. *Price Coupling of Regions*, May 2014. <http://nordpoolspot.com/globalassets/download-center/pcr/euphemia-public-documentation.pdf>.
- [24] K. Van den Bergh, J. Boury, and E. Delarue. The Flow-Based Market Coupling in Central Western Europe: concepts and definitions. *Working Paper*, 2015. [https://www.mech.kuleuven.be/en/tme/research/energy\\_environment/Pdf/wpen2015-13.pdf](https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wpen2015-13.pdf).
- [25] C. Wan, Z. Xy, Y. Wang, and Z.Y. and Wong K.P. Dong. A hybrid approach for probabilistic forecasting of electricity price. *IEEE Transactions on Smart Grids*, pages 463–470, January 2014.
- [26] D. Waniek, C. Rehtanz, and E. Handschin. Flow-based evaluation of congestions in the electric power transmission system. *7th International Conference on the European Energy Market*, pages 1–6, 2010.
- [27] B.L. Welch. The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, pages 28–35, January 1947.
- [28] R. Weron. *Modeling and forecasting electricity loads and prices: A statistical approach*. Wiley, 29 July 2013. <ftp://ftp.repec.org/opt/ReDIF/RePEc/wuu/hsbook/HSBook0601.pdf>.
- [29] R. Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, pages 1030–1081, 1 October 2014. <http://www.sciencedirect.com/science/article/pii/S0169207014001083>.