Forecasting of Zonal Power Transfer Distribution Factors in Flow-Based Market Coupling

Mika Andersson

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.
Helsinki 20.11.2023

Supervisor

Prof. Ahti Salo

Advisor

M.Sc. (Tech.) Jouni Mäenpää

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Mika Andersson

Title Forecasting of Zonal Power Transfer Distribution Factors in Flow-Based Market Coupling

Degree programme Engineering Physics

Major Advanced Energy Technologies **Code of major** SCI

Supervisor Prof. Ahti Salo

Advisor M.Sc. (Tech.) Jouni Mäenpää

Date 20.11.2023 **Number of pages** 50 **Language** English

Abstract

This thesis focuses on forecasting the Zonal Power Transfer Distribution Factor (PTDF) in the context of Flow-Based Market Coupling (FBMC). The growing share of renewable energy in European energy sector has brought challenges to electricity markets and accurate forecasting of price. The FBMC methodology was introduced to address the challenges of renewable dominated energy system, but little research has been conducted on predicting Flow-Based (FB) parameters.

The thesis develops a model based on the physical characteristics of nodal PTDF and Generation Shift Keys (GSKs) to forecast zonal PTDF values. The nodal PTDF represents the sensitivity of power flows in critical network elements (CNEs) during a power transfer between different nodes, while GSKs capture the relationship between zonal power transfers and nodal power injections. The model is built on the perspective of a market participant and therefore, only utilizes publicly available data. The model's effectiveness is tested using historical data from the External Parallel Run (EPR) of the capacity calculation by the Nordic Regional Coordination Centre (CCR). The results are reasonably accurate, with generally small errors in the predicted PTDF values. Suggestions for improving the model focus on the functioning of the metric used for the forecasted values. Additionally, the thesis adds to the current research on the theory and behaviour of the FBMC parameters in changing electrical systems.

Keywords Flow-Based Market Coupling, Zonal Power Transfer Distribution Factor, Nodal Power Transfer Distribution Factor, Generation Shift Key, Capacity Calculation Methodology, Flow-Based method, energy markets, market participant, electricity price, Nordics, forecast, linear model

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Diplomityön tiivistelmä

Tekijä Mika Andersson

Työn nimi Alueellisen voimansiirron jakautumiskertoimen ennustaminen Flow-Based markkinakytkennässä

Koulutusohjelma Engineering Physics

Pääaine Advanced Energy Technologies **Pääaineen koodi** SCI

Työn valvoja Prof. Ahti Salo

Työn ohjaaja DI Jouni Mäenpää

Päivämäärä 20.11.2023 **Sivumäärä** 50 **Kieli** Englanti

Tiivistelmä

Tämä diplomityö syventyy alueellisen voimansiirron jakautumiskertoimen (PTDF) arvon ennustamiseen Flow-Based markkinakytkennän (FBMC:n) yhteydessä. Euroopan energiasektorilla uusiutuvan energian kiihtyvä kasvu on tuonut haasteita sähkömarkkinoille ja hintojen tarkalle ennustamiselle. FBMC metodologia luotiin vastaamaan näihin uusiin haasteisiin, joita uusiutuvalla energialla toimiva energiajärjestelmä tuottaa, mutta tutkimusta Flow-Based (FB) parametrien ennustamisesta on toistaiseksi tehty niukasti.

Tässä diplomityössä kehitetään malli, joka perustuu nodaalisen PTDF:n ja generaation muutosvakion (GSK:n) fyysisiin ominaisuuksiin ennustaakseen alueellisen PTDF:n arvojen muutosta. Nodaalinen PTDF kuvastaa voimansiirtolinjojen herkkyyttä eri noodien välisten voimansiirtojen seurauksena, kun taas GSK:t ilmaisevat suhteen alueellisten voimansiirtojen ja noodien tuotannon välillä. Malli rakennetaan markkinatoimijan näkökulmasta ja käyttää siten vain julkisesti saatavilla olevaa tietoa. Mallin tehokkuutta testataan historiadatalla Nordic Regional Coordination Centre (CCR) kapasiteetinlaskennan testiajosta (EPR:stä) Pohjoismaissa. Tulokset ovat kohtuullisen tarkkoja ja virheet ennustetuissa PTDF arvoissa ovat pääosin pieniä. Mallin tarkkuuden parantamiseksi todetaan arvojen ennustamisessa käytetyn metriikan toiminnan olevan oleellisin tekijä. Lisäksi diplomityö avartaa nykyistä käsitystä FBMC parametrien teoriasta muuttuvassa sähköisessä systeemissä.

Avainsanat Flow-Based markkinakytkentä, alueellinen voimansiirron jakautumiskerroin, nodaalinen voimansiirron jakautumiskerroin, generaation muutosvakio, kapasiteetinlaskentametodologia, Flow-Based laskentamenetelmä, energiamarkkinat, markkinatoimija, sähkön hinta, Pohjoismaat, ennuste, lineaarinen malli

Preface

Studying Engineering Physics at Aalto University has undoubtedly been the most challenging and rewarding experience of my life thus far. Having the opportunity to conclude this journey with such a great team and friends I have made along the way means the most to me.

As I write these words, the reality that this experience is finally coming to an end still feels quite surreal. I want to give special thanks to Jouni, my advisor and colleague, who consistently demonstrated professionalism and guided me through every obstacle encountered during the research for this thesis. I would also like to extend my gratitude to my other colleagues in trading for their lighthearted humor and encouraging words throughout the writing process. Most importantly, I want to express my profound appreciation to my closest friends and family for their unwavering support during the best and most challenging days. It would have been considerably more difficult without all of you.

I will strive to preserve the knowledge and teachings I have received during my time as a student and continue to embrace the journey of lifelong learning. I hope that the collective experiences I have gained will contribute to creating something positive and good for the world.

Helsinki, 20.11.2023

Mika Andersson

Contents

Abstract 3

Abstract (in Finnish) 4

Preface 5

Contents 6

Symbols and Abbreviations 8

1 Introduction 10
1.1 Electricity Markets . . . 10
1.2 Capacity Allocation and Congestion Management Regulation . . . 11
1.3 Research Question and Motivation . . . 12

2 Transition to Flow-Based Market Coupling 14
2.1 Electricity Markets and the Price Calculation . . . 14
2.2 The Net Transfer Capacity Method . . . 16
2.3 Flow-Based Method . . . 18

3 Physical Nature of Flow-Based Parameters 21
3.1 Zonal Power Transfer Distribution Factor . . . 21
3.2 Nodal Power Transfer Distribution Factor . . . 22
3.2.1 Outages and Load Flow Distribution . . . 22
3.2.2 Time Dependency and Simulation . . . 23
3.3 Generation Shift Key . . . 25
3.3.1 Static GSK Strategies . . . 26
3.3.2 Dynamic GSK Strategies . . . 27

4 A Linear Model of the Zonal PTDF 30
4.1 Theoretical Derivation . . . 30
4.2 Application to Market Data . . . 31

5 Forecasting the Zonal Power Transfer Distribution Factors 33
5.1 Setting the Forecast . . . 33
5.1.1 Zonal PTDF Data . . . 33
5.1.2 Application of the Data . . . 34
5.1.3 Outage Report Data . . . 34
5.1.4 Selection Metric for the Outage . . . 36
5.2 Results of the Forecast . . . 37
5.2.1 Individual Elements . . . 37
5.2.2 The Zonal PTDF Matrix . . . 40
5.2.3 Further Analysis of the Results . . . 43

6 Conclusions 45

References **47**

Symbols and Abbreviations

Symbols

$GSK_{n,z}$	generation shift key of grid node n and bidding zone z
$\boldsymbol{GSK}$	generation shift key matrix
$\Delta \boldsymbol{L}$	load flow distribution matrix
$\Delta \boldsymbol{L^Z}$	zonal load flow distribution matrix
$PTDF^N_{l,z}$	nodal power transfer distribution factor of critical transmission element l and bidding zone z
$\boldsymbol{PTDF^N}$	nodal power transfer distribution factor matrix
$PTDF^Z_{l,z}$	zonal power transfer distribution factor of critical transmission element l and bidding zone z
$\boldsymbol{PTDF^Z}$	zonal power transfer distribution factor matrix

Operators

$\boldsymbol{A} \times \boldsymbol{B}$	matrix product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$
$\boldsymbol{A}^T$	transpose of matrix $\boldsymbol{A}$
$\sum_n$	sum over index n

Abbreviations

ANN	artificial neural network
ATC	available transfer capacity
BZ	bidding zone
CACM	capacity allocation and congestion management
CCM	capacity calculation method
CWE	central Western Europe
CNE	critical network element
CNEC	critical network element and contingency
D-1	one day prior (equivalent to day-ahead)
D-2	two days prior
DA	day-ahead
EPR	external parallel run
EU	European Union
FB	flow-based
FBMC	flow-based market coupling
GSK	generation shift key
IEA	International Energy Agency
JAO	Joint Allocation Office
MAE	mean absolute error
NTC	net transfer capacity
PTDF	power transfer distribution factor
QQ	Quantile-Quantile
RMSE	root mean squared error
SW	Shapiro-Wilk test
TRM	transmission reliability margin
TSO	transmission system operator

1 Introduction

1.1 Electricity Markets

Since the first major steps in climate policies in Europe in the 1990s [1], the share of renewable energy in the European energy sector has grown significantly [2, 3]. According to Eurostat, the statistical office of European Union (EU), the share of renewable energy in the gross final energy consumption increased from 8.3 % in 2004 to 20.8% in 2020. Moreover, the total installed renewable capacity in Europe reached 536 GW in the same year, up from 5.2 GW in 1990, with majority of the capacity coming from wind power [3].

A similar trajectory has continued in recent years. The adoption of renewable technologies in energy production has accelerated due to stricter climate targets and policies set by the European Commission in the European Green Deal in 2019 [4]. Additionally, the events in Ukraine, which caused a major energy crisis in Europe, have further expedited the pace of building new power plants independent of fossil fuels [5]. According to the International Energy Agency's (IEAs) forecast, the cumulative renewable electricity capacity in Europe is expected to increase by nearly 2400 GW (equivalent to the entire installed power capacity of China today) between 2022 and 2027, representing an 85% acceleration compared to the previous five-year period. This is the largest upward revision ever made by the IEA.

These changes in the source of the electrical energy production have also imposed changes in the energy system, which includes the electrical grid used in the distribution and transmission of electricity [6]. The main challenge in adopting a vast amount of renewable energy in the energy system is the unpredictable nature of the renewable energy sources and the variability of the generation [7, 8]. The most commonly used variable generation sources are solar and wind which both depend on weather conditions. Consequently, due to fluctuating weather changes, there will also be fluctuations in the energy output of the system. To integrate this variable energy into the grid requires new investments in the grid's flexibility and resilience [6, 9]. Although investments have been made in smart grid and storage technologies, more is needed to accommodate future demands [7, 9].

The problem of unpredictability and flexibility in energy production extends to the economics of the electricity market in Europe. Trading in the electricity markets relies on the ability to forecast the future prices as accurately as possible. As the electricity prices in the market are determined by the intersection of supply and demand, large fluctuations in electricity generation, i.e., the supply, directly impact the price [10]. These fluctuations pose a major challenge for accurate forecasting. The difficulty of predicting such price fluctuations affects all market participants, such as energy companies, industries and grid operators, and ultimately alters the social welfare of Europe [11]. When supply and demand are out of balance, prices may converge at either the low or high end of the demand-supply curve [10]. In some cases, this imbalance causes problems for consumers outside the immediate electricity market. For instance, the recent energy crisis resulted in soaring prices due to the lack of gas production in Europe after severing trade with Russia [12]. As a result,

some governments even distributed funds to the general public to alleviate the high energy bills [13, 14]. Similar issues arise when significant amounts of renewables are integrated into the grid without sufficient mechanisms to control the variable generation. Extreme weather conditions, both short and long term, cause distinctive imbalances in the supply and demand of energy [15]. Therefore, the importance of flexibility and stability in both the energy system and the energy markets is now more essential than ever.

1.2 Capacity Allocation and Congestion Management Regulation

In 2015, a new methodology called Capacity Allocation and Congestion Management (CACM) Regulation was implemented in Central Western Europe (CWE) with the purpose of addressing the challenges posed by increasing renewables in the grid. These challenges include ensuring security of energy supply, enhancing competitiveness, and allowing all consumers to purchase energy at affordable prices [16]. The methodology introduced a new procedure for calculating available capacity in the energy markets. In general, the capacity calculation determines the available capacity for trading in the market for each given time step. Therefore, the amount of trade that can be conducted by the market participants is restricted by the calculation procedure defined by this regulation. The method employed by the CACM regulation is called as the Flow-Based (FB) capacity calculation method (CCM), also known as the Flow-Based Market Coupling (FBMC) [17].

Before this implementation, electricity in the markets was primarily supplied by easily controllable energy sources like hydro power, gas and coal. The price calculation was performed using a different method known as the Net Transfer Capacity (NTC) method. The NTC method relied on the predictability of these energy sources and employed a simplified approach to calculate available capacity in the energy system. The method assumes that electricity flow follows the actual trade and only flows bilaterally through a single connection. However, physical reality is different. According to Kirchhoff's laws, electricity flows through all available paths between two points and is distributed across multiple connections in the grid. This simplification led to inaccuracies as renewables were increasingly integrated into the grid, resulting in increased congestion and price differences between zones in the electricity markets. A congestion is an event where the capacity limits of grid connections are fully used and no more electricity can be transferred between two zones. Price differences will emerge due to imbalance of electricity supply and demand between the zones. The main objective of the FBMC is to address these inaccuracies in determining the available capacity and to readjust to the new energy system with renewable production [17].

Under the implemented FBMC, the physical reality in the flow of electricity is integrated in the capacity calculation. It introduces two types of parameters; the power transfer distribution factors (PTDFs) and generation shift keys (GSKs). PTDFs describe how the flow of a transfer is distributed among other connections in the grid, indicating the fraction of flow through a particular connection. GSKs

describe the generation units involved in the transfer and their contributions. The transmission system operator (TSO) of the market area defines these parameters. The FBMC, therefore, maximizes the available information in the capacity calculation with the aim of making the utilization of the grids capacity more efficient [17].

Several reports have been conducted to assess the feasibility and benefits of the FBMC [18, 19, 20, 21, 22]. There is a general consensus that when properly implemented, the FBMC enhances grid reliability, stability, and market performance. Since the method takes into account factors such as the network topology and the interaction between different parts of the grid, it allows for a more accurate and dynamic calculation of the available capacity. This reduces the risk of overloading, grid failures, and bottlenecks [18]. The reports have also shown increases in the domain of feasible trades with the FBMC in comparison to the domain achieved in the NTC approach [23]. This increase in the domain is accompanied by improved price convergence and welfare gains.

While the results of the FBMC are promising for most market participants, there have been criticisms regarding the transparency in the calculation and implementation of the method. Participants outside of the TSOs, which provide the parameters, find it challenging to access and interpret the FB market results [22, 24]. The complexity of the method and the involvement of multiple stakeholders make it difficult for market participants and external observers to fully understand and assess the market outcomes and the allocation of available transfer capacity [24]. Additionally, TSOs have some flexibility in applying models and setting parameters to ensure security of supply, making it impossible for parties other than TSOs to precisely model the FBMC method. As a result, market participants find it challenging to forecast the capacities that TSOs will make available for future trading intervals [22].

1.3 Research Question and Motivation

Research on forecasting the FBMC parameters for future trading intervals and their effectiveness is limited. Van Stiphout (2016) compares multiple methods for forecasting capacities in the Day-Ahead (DA) market coupling, relying on past FB capacity data. One approach is to search for similar hours from historical data and using that information for future capacities. Other approaches utilize probabilities and correlations found in the historical data. Wallin (2016) addresses similar topics but focuses on past data of electricity flows resulting from trading directly. The study compares a multiple linear regression model with various Tobit models, with a specific emphasis on NTC-based transfers between selected European countries. Abdel-Khalek et al. (2019) employ an Artificial Neural Networks (ANN) model fitted to existing FB capacity data, comparing it to a simple carry-on approach. According to the knowledge of these authors, no other models addressing predictions of the FBMC domain exist, indicating that the current research on the subject is limited [25, 26, 27].

Therefore, there is a need for further research on forecasting FBMC parameters and exploring different methods for doing so. The aforementioned attempts primarily focused on short-term trading intervals ranging from a few hours to days. Addition-

ally, little emphasis was given to the actual physical characteristics of the FBMC parameters, and the reports preferred a black-box type of approach.

In this thesis, an approach which forms the basis of the model in the physical characteristics of the individual FBMC parameters is proposed. Thus, the thesis focuses on the theoretical behaviour of the nodal PTDF values and the GSKs, in order to forecast their effect on the zonal PTDF. Based on the theoretical review and existing literature on these individual parameters (the nodal PTDF and the GSK), approximations which are in line with their characteristics are made for the predictive model. The forecasted time period extends to months and is primarily limited by the availability of data. Differing from the previous research, the model mainly focuses on planned/unplanned outages, which are known to market by publications of TSOs. In addition, only publicly available data is used for the model. The outlook of the thesis is on the Nordic region of the electricity markets, which is undergoing a change in the capacity calculation, from NTC method to FBMC. The motivation behind this research is to provide a framework for forecasting FBMC parameters for all market participants and contribute to the current understanding of the topic.

2 Transition to Flow-Based Market Coupling

The capacity calculation plays a crucial role in determining the electricity price within electricity markets. In this section, the general principles of an electricity market and its functioning is explained with a focus on the capacity calculation itself. This is followed by a detailed description of the transition from the NTC method to the FB method, and its ability to better describe the physical characteristics of the grid.

2.1 Electricity Markets and the Price Calculation

The electricity markets are comprised of market participants and bidding zones. A bidding zone is a defined geographical area within a power system (commonly set by the borders of countries) where the participants can submit their offers and bids to buy or sell electricity. Generation units offer their production where as consumption units bid their expected loads for predefined time intervals in the market. All the offers are collected in a pool for each bidding zone. A supply curve is formed for the sell offers (generation) and a demand curve from the buy offers (consumption). The supply bids are set in a merit order, which ranks the sources of production in an ascending order of price. Here the price is defined from the short-run marginal costs of producing energy. The spot price of the electricity is found at the intersection of these curves, defining the price for the whole bidding zone [10]. The price formation is illustrated in Figure 1.

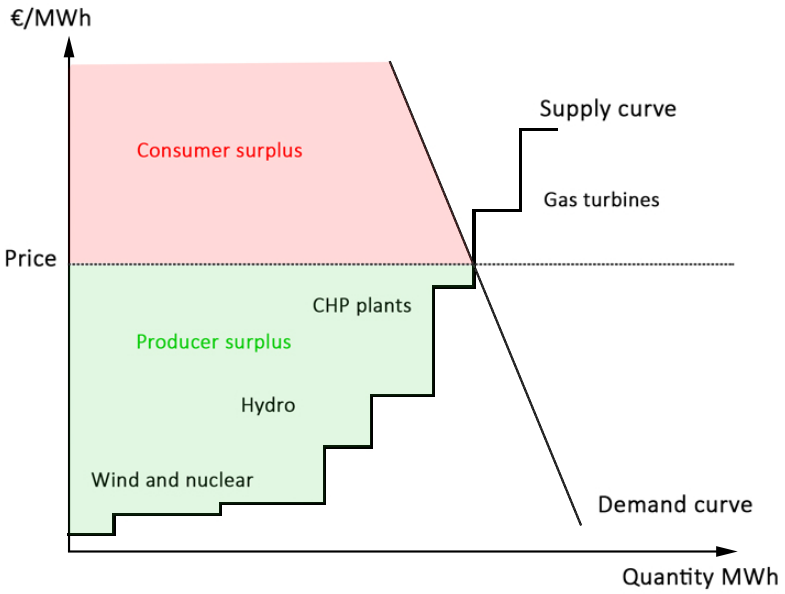

Figure 1: The clearing price determined by the demand and supply curves in a bidding zone. The production offers of different energy production units are shown in the supply curve. The demand curve is simplified as a straight line. Consumer surplus can be seen as the red area, whereas the producer surplus as green.

During this process, the net positions for all of the bidding zones in the market are defined, which represents the balance between generation and consumption of the zone. A positive net position indicates that the total generation within the bidding zone exceeds the total consumption. Such a zone is a net exporter of electricity. The opposite is true for a negative net position. Transfers between zones with different net positions are optimized within the capacity limits of the network. If the capacity between bidding zones is too low to conduct a transfer, a congestion occurs, where price difference between zones emerge. By optimizing the net positions of the bidding zones, the social welfare of the market can be maximized. The exact definition of the welfare is the sum of the consumer surplus, producer surplus, and the congestion income.

In Figure 1, the consumer surplus is illustrated as the area between the supply curve and the price, left of the intersection. Similarly, the producer surplus is the area between the demand curve and the price. Congestion income on the other hand, is a revenue that arises in the electricity market in the case of congestion between zones. In the price calculation in general, this sum is maximized to ensure the best outcomes for the market [17, 23].

The calculation of the spot price of electricity, as described in the previous paragraphs, only occurs in the day-ahead (DA) market. However, electricity markets operate on various time intervals to ensure a better match between supply and demand at the time of delivery, which is crucial for the electrical grid. The structure of these markets is depicted in Figure 2. Trading of electricity begins in the long-term markets through long-term bilateral contracts, forward markets, and future markets. Long-term markets can commence years in advance and continue until two days before the delivery date (D-2). Following this, the DA market takes place one day before the delivery date (D-1), where the market clearing sets the price for electricity for each hour of the next day. After the market clearing, there is still room for trade,

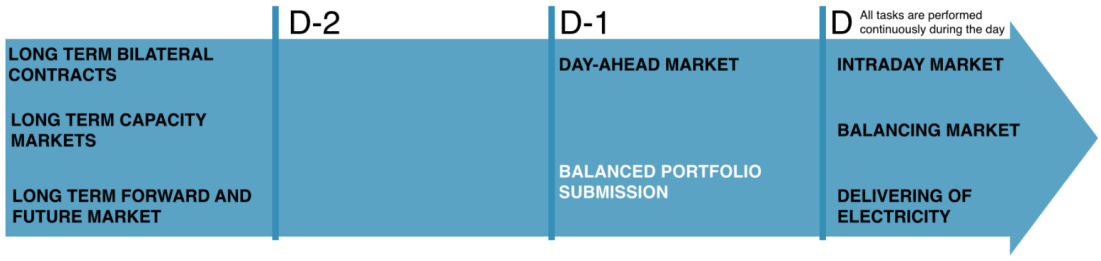

Figure 2: The different timelines where the electricity is traded before the delivery date [28].

which occurs on the delivery day. Most of this trading takes place in the intraday market, where the supply and demand of electricity are further balanced. The initial price in this market is defined by the spot price for each hour, but it can deviate once trading begins. The volumes traded in the intraday market are typically relatively small. Although the intraday trading largely balances out the demand and supply of electricity, there are often some final adjustments to be made. Any remaining

imbalances are addressed in the balancing market, where electricity is sold to the TSO of the bidding zone [10, 17].

The timeline which is affected by the CAMC regulation mentioned in the introduction, is the DA market. As was described in the previous paragraph, the DA price sets the starting point for the intraday market and also defines whether a long term trade was profitable. Therefore, the capacity calculation which comes in to the picture when different bidding zones are combined, has central place in setting the prices of the market.

2.2 The Net Transfer Capacity Method

At the moment, the Nordic electricity markets are operated with the Net Transfer Capacity method. In the NTC method, the TSO allocates the available capacity for trading based on a basecase scenario developed two days (D-2) before the moment of delivery. The basecase scenario is formed on historical data, forecasted demand, generation availability and network topology. As the market participants submit their offers in the DA market for every hour of the upcoming day, these offers are coupled by an algorithm which utilizes these capacity constraints given by the TSOs. The capacity constraints deduced from the basecase scenario contain the capacities for all the border connections between bidding zones. The domain formed by these capacity limits describes the possibilities inside which the algorithm needs to find the optimum [17, 23, 29].

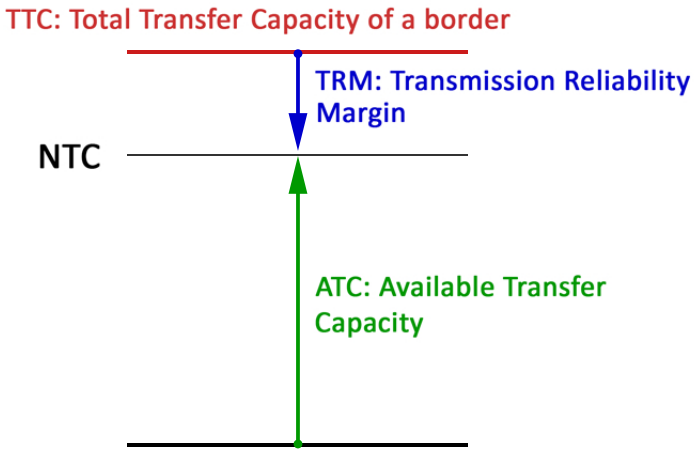

Figure 3: The capacity allocation in the NTC method. The Total Transfer Capacity (TTC) is the maximum capacity of a border connection, which also takes into account all the grid related issues required to maintain a security of supply defined by the TSO. The TTC is divided to Transmission Reliability Margin (TRM) and Available Transfer Capacity (ATC) as depicted in blue and green, respectively.

As shown in Figure 3, NTC limits are composed of two components; the Trans-

mission Reliability Margin (TRM) and the Available Transfer Capacity (ATC). The ATC describes the actual capacity that is available for trade in the market, when all the factors in the basecase scenario have been taken into account. This is the value which is used in the optimization of the net positions in the price calculation. The TRM describes all the capacity which is left unused in order to ensure a reliable and secure operation of the power system. It functions as a buffer for stability in unexpected changes in generation, demand or network conditions [18, 23].

Since the NTC constraints only describe the border connections, an unexpected result in market conditions during the price calculation may lead to increased congestion. As electricity is transferred from bidding zone to another, it takes a path along all the connections in the grid due to Kirchhoff's law. An incorrect result in the forecast of market conditions will provide capacities for a market situation that does not align with physical reality. In the NTC method, there is no mechanism to directly address to this error, since the algorithm only has the information of the border capacities predicted in D-2 and the information of the actual paths of electricity is missing. This increases the risk of stability issues in the grid and necessitates additional safety measures in the provided NTC values, such as the TRM. Since the risk of such a forecast error has increased among the adding of vast amounts of renewables in the grid, the capacity provided by the NTC is more limited and leaves capacity unused by the market [17, 18, 23].

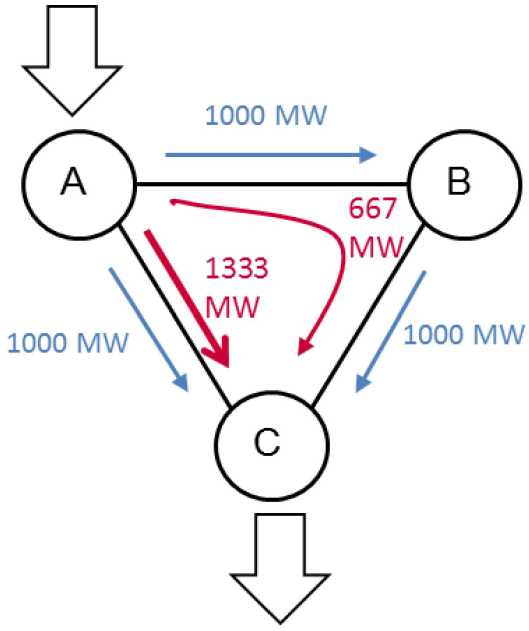

Figure 4: Physical flows resulting from the trade between zone A and zone C. Due to the physical properties of the grid, the transfer of 2000 MW is divided between the connections and in this case, overloading the connection between A and C [30].

The difference between the NTC flows and the physical flows can be seen in Figure 4. Here the line capacities allow 1000 MW transfers per line and all lines are expected to have the same impedance (electrical resistance). A market transfer of

2000 MW is made from zone A to zone C, which in the assumptions of the NTC method should be possible, since all the 1000 MW connections can participate to fulfill the transfer. This is depicted in blue in the picture. In reality however, the flows in such a transfer follows the laws of physics and the individual flows are divided between all the connections in the grid based on their individual physical properties and topology. Using Kirchhoff's law, the flow is divided into 1333 MW flow between A and C, and 667 MW from A to B to C, depicted in red. Due to the physical properties of the grid, this trade leads to a situation where the line from A to C is overloaded. In this situation the TSO would need to limit the capacity to 1500 MW a trade, in order to avoid overloading of the 1000 MW lines [17].

2.3 Flow-Based Method

The flow-based method derives its parameters from the nodal description of the grid. In the nodal description, the grid is accounted fully; all the nodes and lines in the grid are considered, and therefore, during market clearing the results of the price calculation are exact. Looking at the example of Figure 4, the nodal FB parameters for the system can be directly derived by using Kirchhoff's first and second law. The calculation is a basic direct current circuit exercise, where the fractions of power flows on the lines are solved assuming the impedance of the circuit lines equal. The impedance of lines could be any other value as well, but the equality is assumed for simplicity of the demonstration. The resulting nodal power transfer distribution factors (PTDFs) which describe the flow influence of each bidding zone to each connecting line, are shown in Table 1. Here, the node C was used as a reference for the values. This means that all power injected in bidding zones A and B is (mathematically) absorbed in bidding zone C. The same holds true for bidding zone C itself, all power injected in bidding zone C is absorbed in C (hence, the zero values) [17].

Line	A	B	C
$A \Rightarrow B$	1/3	−1/3	0
$A \Rightarrow C$	2/3	1/3	0
$B \Rightarrow C$	1/3	2/3	0

Table 1: The nodal PTDFs describing the power flow in the three node system, when C is chosen as the reference node.

The nodal PTDFs are a crucial part of the information added in the flow-based capacity calculation. However, the method suggested by the CAMC regulation has some added simplifications due to the existence of bidding zones. As described in the previous subsection, the capacity calculation in the NTC allowed for the markets to function on the zonal level. The nodal description above requires nodal pricing. Therefore, to keep the zonal interpretation, the nodal PTDF needs to be reduced by an approximate parameter. This parameter is the generation shift key (GSK), which translates the nodal system included in the nodal PTDFs to a zonal system required by the markets. The GSKs describe the contribution of the individual

power plants to the bidding zones total production. In Figure 5, the effect of the GSK approximation is illustrated. The nodes of the countries are shifted to act as a one zone and a three node system emerges instead of the original nine node system. This coupling of the bidding zones by using FB parameters is defined as the FBMC [17, 18, 23].

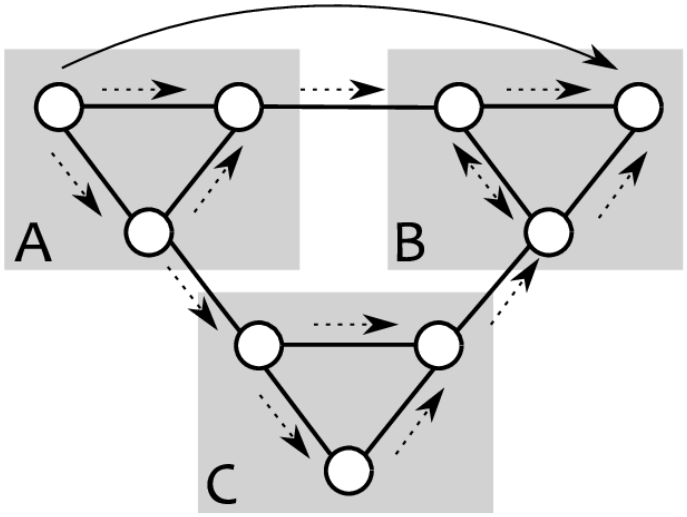

Figure 5: A zonal approximation of the nodal system. The long arrow indicates a transfer between the two outer nodes and the bilateral path of the electricity between the bidding zones depicted as gray areas. The physical reality of the actual flow of electricity is depicted in smaller arrows [18].

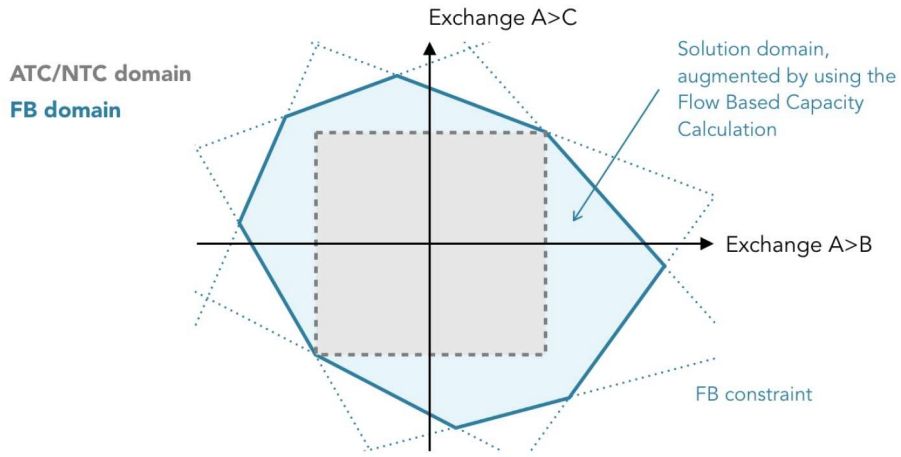

Figure 6: A representation of the solution domain for both NTC and FB method in the allocation of the transfer capacity between different zones. The solution domain of the FB method depicted in blue includes the solution domain of the NTC method depicted in gray [31].

In FBMC, similar to the NTC calculation, both parameters (nodal PTDFs and GSKs) are predetermined in the base case D-2 prior to delivery, introducing a certain

level of uncertainty in the capacity calculation. However, the availability of actual flow information during transfers between bidding zones enables the method to adjust to the real-time market outcome. Once the market result of the day-ahead (DA) market is known, the optimization of bidding zone net positions incorporates the impact of trade on all grid connections. Consequently, FBMC is expected to provide a more accurate assessment of capacities and allocate them more efficiently in the market. It can be shown that all solutions within the NTC domain are included in the FB domain, but the reverse is not true. This is depicted in Figure 6, where the gray area represents the solution domain based on NTC constraints, while the blue area represents results within the FB domain. The blue area allows for additional possibilities for zone-to-zone electricity exchanges [17, 18, 23].

The FB parameters, therefore, play a crucial role in determining the outcomes of the price calculation and are influenced by the grid topology and its physical characteristics. To understand the zonal PTDF values employed in the FBMC and their evolution in time, understanding the individual behaviour of its constituents is essential. This will be the focus of the next section.

3 Physical Nature of Flow-Based Parameters

In this section, a theoretical review is conducted on the properties of the parameters that determine the values of the zonal PTDFs. The analytical behavior of each parameter, namely the nodal PTDF and GSK, is derived based on their definitions, and the impact of the electrical system on these parameters is modeled. This will enhance the understanding of the behavior of critical line flows in an evolving electrical system and function as a baseline for the predictive model of the zonal PTDF.

3.1 Zonal Power Transfer Distribution Factor

In the flow-based market coupling the zonal power transfer distribution factors, $PTDF^Z_{l,z}$, are defined as the weighted sum of the nodal power transfer distribution factors, $PTDF^N_{l,n}$, and generation shift keys, $GSK_{n,z}$,

$$PTDF^Z_{l,z} = \sum_n PTDF^N_{l,n} \cdot GSK_{n,z} \ , \tag{1}$$

where n is the node, l is the critical transmission element (cross-border line, internal transmission line or transformer) and z is the bidding zone [18]. Critical transmission elements are defined by the TSOs in the basecase and are chosen on the basis of how the flows are affecting each element in the grid. Commonly only elements which contribute at least 5% to the flows are taken into account [17]. The relation in Equation 1 can also be expressed in matrix form, which will be the preferred notation

$$\boldsymbol{PTDF}^Z = \boldsymbol{PTDF}^N \times \boldsymbol{GSK} \tag{2}$$

and in opened up form

$$\begin{bmatrix} PTDF^Z_{1,1} & \cdots & PTDF^Z_{1,z} \\ PTDF^Z_{2,1} & \cdots & PTDF^Z_{2,z} \\ \vdots & \ddots & \vdots \\ PTDF^Z_{l,1} & \cdots & PTDF^Z_{l,z} \end{bmatrix} = \begin{bmatrix} PTDF^N_{1,1} & \cdots & PTDF^N_{1,n} \\ PTDF^N_{2,1} & \cdots & PTDF^N_{2,n} \\ \vdots & \ddots & \vdots \\ PTDF^N_{l,1} & \cdots & PTDF^N_{l,n} \end{bmatrix} \times \begin{bmatrix} GSK_{1,1} & \cdots & GSK_{1,z} \\ GSK_{2,1} & \cdots & GSK_{2,z} \\ \vdots & \ddots & \vdots \\ GSK_{n,1} & \cdots & GSK_{n,z} \end{bmatrix} . \tag{3}$$

Here the dimensions of the zonal PTDF are defined by the number of zones, z, in the capacity calculation and the number of critical transmission elements, l, which are also referred as critical network elements and contingencies (CNECs).

The parameters that constitute the zonal PTDF matrix, $\boldsymbol{PTDF}^Z$, in equation (2) were briefly introduced in Section 2.3. The nodal PTDF matrix, $\boldsymbol{PTDF}^N$, and the GSK matrix, $\boldsymbol{GSK}$, serve as the foundation for the behavior of the zonal PTDFs. Changes in these parameters have a combined effect that propagates to the zonal PTDF. Therefore, understanding the nature of each parameter is crucial for predictive analysis. In general, the nodal PTDFs are derived from the physical characteristics of the grid and are sensitive to any real-world changes in the grid. On the other hand, the GSKs are defined by the TSO of the respective zone and can vary based on the TSO's preferences. The GSKs are virtual parameters and theoretically independent of physical changes [18, 32]. Since the zonal PTDF is the product of these parameters, their effects on the zonal PTDF are examined individually, while one of the parameters remains constant.

3.2 Nodal Power Transfer Distribution Factor

The analysis on this subsection is largely based on electrical networks and circuit analysis, which can be further explored in reference [33]. In essence, the properties of networks are defined by the number of nodes and lines (topology) and the characteristics of its components. Networks with similar topology exhibit similar properties regardless of other factors.

3.2.1 Outages and Load Flow Distribution

Assume a circuit depicted in Figure 7a represents a physical network of a bidding zone. It consists of four generation nodes shown in blue and five transmission lines illustrated in gray. The nodal PTDF for the network can be obtained as

$$\boldsymbol{PTDF}_a^N = \begin{bmatrix} 0.25 & -0.375 & 0 & 0.125 \\ 0.25 & 0.625 & 0 & 0.125 \\ 0.25 & 0.125 & 0 & 0.625 \\ 0.25 & 0.125 & 0 & -0.375 \\ 0.5 & 0.25 & 0 & 0.25 \end{bmatrix}, \tag{4}$$

assuming equal line impedances. The footnote, a, in, $\boldsymbol{PTDF}_a^N$ refers here to the network of Figure 7a (a similar reference will be done for the network in Figure 7b). Node C is chosen as the reference node, indicated by the red arrow in Figure 7, for which Kirchhoff's laws are applied in order to derive the values in the matrix (calculated similarly as in Section 2.3). The matrix is structured such that its rows represent the individual transmission lines in numerical order and its columns represent the generation nodes in alphabetical order as in Figure 7a, resulting in a five-by-four matrix.

The fraction of flow through each line in cases of changes in the net position of a node in equation (4) was primarily determined by the network topology. To highlight this point, the same parameter, the nodal PTDF, is calculated for a different network with the same impedance of transmission lines. Assume a change happens in the topology of the network shown in Figure 7a. Specifically, assume that one of the lines, denoted as l_5, is taken out of service for maintenance. As a result, the network transforms into the configuration shown in Figure 7b, where the line l_5 is effectively removed.

The nodal PTDFs corresponding to this new network configuration are then calculated as

$$\boldsymbol{PTDF}_b^N = \begin{bmatrix} 0.5 & -0.25 & 0 & 0.25 \\ 0.5 & 0.75 & 0 & 0.25 \\ 0.5 & 0.25 & 0 & 0.75 \\ 0.5 & 0.25 & 0 & -0.25 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5}$$

The remaining non-zero matrix elements in equation (5) have higher values compared to equation (4) on average. In essence the flow on the line l_5 is now redistributed to the remaining transmission lines, thus increasing the fraction of flow through these

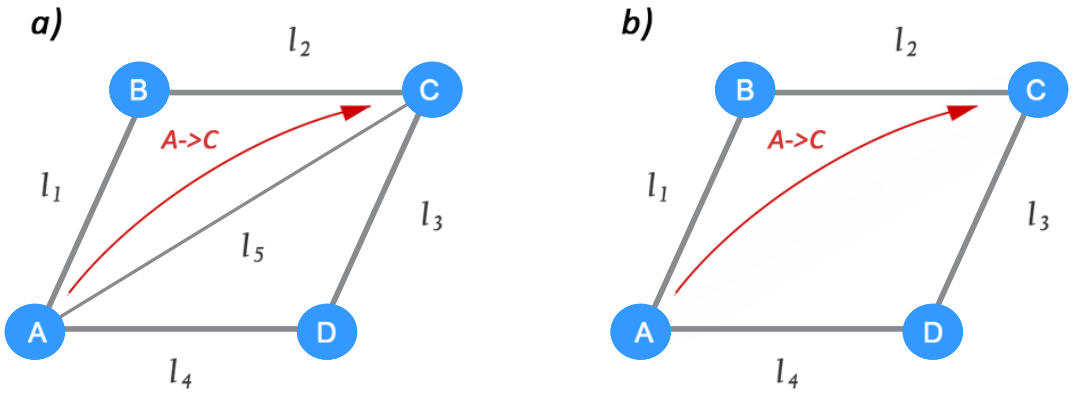

Figure 7: Grid topology of two different grids, with one having a diagonal line, l_5, connecting the zones A and C, and one without the connection. The lines are illustrated in gray and the nodes in blue. The red arrow shows the reference node in the PTDF calculation.

lines in the network. By computing the difference, $\Delta \boldsymbol{L} = \boldsymbol{PTDF}_b^N - \boldsymbol{PTDF}_a^N$, the redistribution of flows resulting from the outage of line l_5 can be determined as

$$\Delta \boldsymbol{L} = \begin{bmatrix} 0.25 & 0.125 & 0 & 0.125 \\ 0.25 & 0.125 & 0 & 0.125 \\ 0.25 & 0.125 & 0 & 0.125 \\ 0.25 & 0.125 & 0 & 0.125 \\ -0.5 & -0.25 & 0 & -0.25 \end{bmatrix}. \tag{6}$$

The load flow distribution, $\Delta \boldsymbol{L}$, in equation (6), thus describes the change in nodal PTDF due to an outage [32]. If line l_5 were to be replaced with a new one after the maintenance, the nodal PTDFs would return to their previous values as in (4).

The example highlights how a physical change in the grid leads to a change in the nodal PTDF values. Common event such as a transmission line outage, will push the grid to another temporary static state, and after the outage has been taken care of, the grid returns to its original state. When there is no physical change, the nodal PTDF matrix remains constant and can be determined directly based on the known topology and relative impedance of the lines. Consequently, a static grid corresponds to static nodal PTDF values. The nodal PTDF is therefore, independent of any market conditions which is essential in the modelling of the values.

3.2.2 Time Dependency and Simulation

Using the description above, the time evolution of the nodal PTDF of individual network elements can be represented as a step function. Assuming the GSKs as

constant, the effect of outages to the zonal PTDF is evaluated. With the definition of the load flow distribution, $\Delta \boldsymbol{L}$, the change in the zonal PTDF through time, t, due to changes in the nodal PTDF is derived from equation (2) as

$$\boldsymbol{PTDF}^Z(t) = (\boldsymbol{PTDF}^N + \Sigma_j \Delta \boldsymbol{L}_j(t)) \times \boldsymbol{GSK} \ . \tag{7}$$

The load flow distribution step function, $\Delta \boldsymbol{L}(t)$, for each outage, j, is defined as

$$\Delta \boldsymbol{L}(t) = \left\{ \begin{array}{ll} \Delta \boldsymbol{L}, & t_1 \leq t \leq t_2 \\ 0, & \text{otherwise} \end{array} \right\} \ , \tag{8}$$

where $[t_1, t_2]$ are the start and end date of the outage. $\boldsymbol{PTDF}^N$ is the static value of the nodal PTDF when no outages are present and $\Sigma_j \Delta \boldsymbol{L}_j(t)$ the linear sum of the effect of outages on the load flows.

It is important to note that the combined impact of multiple outages is not equivalent to the sum of their individual effects. However, the linearity here is assumed as an approximation in the perspective of a market participant that lacks the exact knowledge of the grid topology and characteristics. The linearity assumption is the most reasonable when the deviations from the operating grid conditions due to the outage are small.

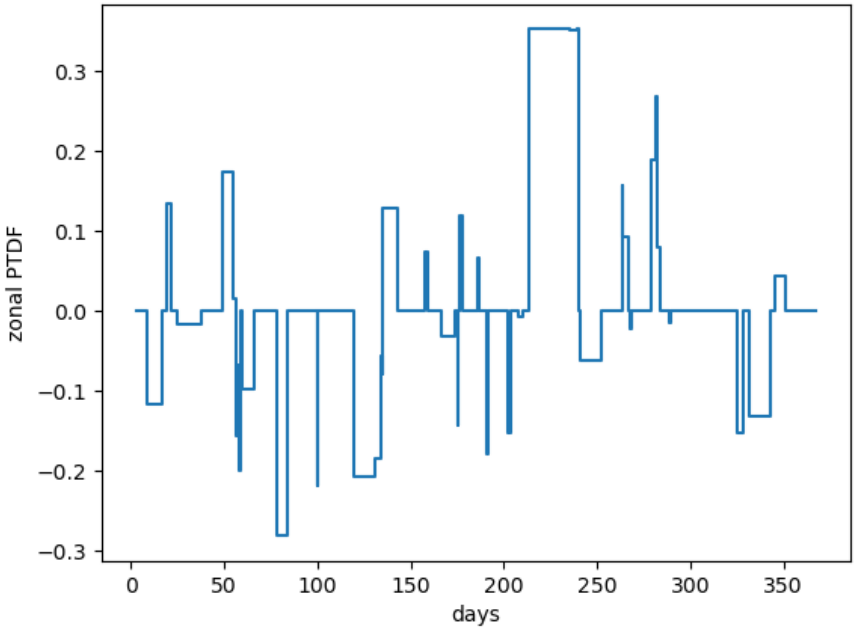

Figure 8: A simulation of an effect of outages on the PTDF values of a single critical transmission element, where a step function is used as an assumption for the effect of each outage.

A simulation is done to illustrate this time evolution in Figure 8 for a single critical transmission element in a four node system. In the simulation, the step size, $\Delta \boldsymbol{L}$, and the step duration, $\Delta t = t_2 - t_1$, are chosen by modified normal distributions.

By definition, the zonal PTDF matrix can experience values in the range of $[-1,\ 1]$. Since the sum of GSK values per bidding zone column always equal 1, the maximum change in a nodal PTDF matrix element is the difference between its static value to both extremes of the [-1, 1] range. The simplest example of this behavior can be illustrated when the static value is assumed as zero, $\boldsymbol{PTDF}_l^N = [0,\ 0,\ 0,\ 0]$ (i.e. a negligible flow is assumed on the critical transmission line when no outages are present). Here, the footnote l refers to a single critical transmission element, which values are a single row from the full nodal PTDF matrix. Therefore, for the step size, a normal distribution with a mean of $\mu = 0$ and standard deviation of $\sigma = 0.3$ is used, and only the values between $[-1, 1]$ are chosen. Each load flow distribution matrix element is chosen by this normal distribution, thus making the electrical characteristics (the number of transmission lines and their impedances) and the outages of the simulated system fairly random.

Based on the reported outages in the Nordpool's Urgent Market Messaging (UMM) platform [34], the duration of outages can last from hours to months and occur frequently throughout the year. Therefore, for the step duration, Δt, the positive values of a normal distribution with a mean of $\mu = 2$ and a standard deviation of $\sigma = 5$ are chosen with the unit being days. For the occurrence dates, t_1, random integer values are taken between $[0, 365]$.

To simulate the tendency for the zonal PTDF value to stay in its static state, roughly 65% of the step sizes are also set to zero by random (65 out of 98 data points in the illustrated case). The GSK value is set at a constant $\boldsymbol{GSK} = [0.25, 0.25, 0.25, 0.25]^T$, which reflects a flat GSK strategy for a four node system (see Section 3.3.1). The result of the simulation is shown in Figure 8. If the actual zonal PTDF values are solely influenced by similar changes, a pattern illustrated in the simulation would be expected.

In practise, the grid can undergo additional changes including the outages, which can cause deviation from the simplified example shown above. The previous assumption of the grid returning to its original state after a topological change slightly falls off. Such changes are mediated by the TSO's, both in short term and long term. Short term changes include line switching, transformer tap adjustments and load shedding. Long term changes are caused by network reconfiguration and grid development [35]. The effect of these actions are TSO and grid specific. Therefore, there are some factors that change the nodal PTDF values which are not forecastable and that are left out in these derivations.

3.3 Generation Shift Key

Another parameter affecting the change of the zonal PTDF values is the GSK. GSKs attempt to predict, which generating units participate in net position changes of the zone they are located in. The evolution of the parameter is determined by the TSO of the bidding zone and can thus, have different effect on the zonal PTDF values depending on the zone [17]. Examples of different strategies for the GSK value are shown in Table 2.

Table 2: An example of GSK strategies currently set by the TSOs of the Nordics. Denmark (Energinet) can be seen to have a static strategy, while others have dynamic ones [36].

TSO	**Strategy number with description/comments**
Energinet	Custom GSK strategy with equal factors assigned to thermal powerplants and offshore windfarms
Fingrid	Generators and loads participate relative to their current power injection. Nuclear, wind power and non-conform loads are excluded from participation (participation factor set to 0)
Statnett	**For areas NO1-NO4:** Generators and loads participate relative to their current power injection. Wind power excluded from participation (participation factor set to 0) **For areas NO5:** Generators participate relative to their current power injection. Wind power excluded from participation (participation factor set to 0)
Svenska kraftnät	Generators and loads participate relative to their current power injection. Nuclear and wind power excluded from participation (participation factor set to 0)

3.3.1 Static GSK Strategies

The simplest of the strategies assumes the contribution of each power plant node to the total net position change to be equal (a flat GSK strategy). An equal contribution for each node leads to GSK values that are unchanging in time. Another strategy with constant values include generators participating to their maximum (installed) capacity [17]. These type of strategies are referred to here as static strategies. In such strategies, the changes in nodal PTDF values determine the timing and duration of change in the zonal PTDF, since the GSK values are set constant. The magnitude of change in these strategies, however, is dependent on both parameters.

An example of the calculation of the GSK values in a static strategy can be carried out for the network introduced in the previous subsection in Figure 7 a. Here an equal participation of each generation node (A, B, C and D) i.e, a flat strategy, produces a GSK matrix

$$\boldsymbol{GSK} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} , \tag{9}$$

where the whole network is considered as a single zone.

3.3.2 Dynamic GSK Strategies

A more complicated picture emerges with dynamic GSK strategies, where the change of the GSK values is dependent on time. Here, the GSKs are usually tied to the size and output of the relevant generation units. The change can then be expected to be induced by events such as power plant outages and the regulation of generation during market hours. However, due to the virtual nature of the parameter, the GSK values do not necessary react to such changes directly, since the parameter is based on assumptions of generation made by the TSO. This decreases the predictability of the actual GSK values by participants outside TSOs, since information of the TSO's own forecasts of the grid and market is not part of the required publicly published data [17].

A simulation is made which focuses only on the unpredictable participation of generators in the market and leaves out the possible influence of TSOs. Therefore, a dynamic strategy where the generation nodes participate to power injections relative to their current output is chosen for the demonstration. The nodal PTDF is kept as a constant value and the zonal PTDF value of a single critical transmission element is simulated. Two systems are compared; one with four generation nodes and one with eight. The fraction of change to the output of a generation node, n, is defined by a normal distribution with a mean value of $\mu = 0$ and a standard deviation of $\sigma = 0.3$. The values are then multiplied by the assumed mean output of the generation node, to normalize the change to the size of the generation node. This change is denoted as $\Delta GSK_{n,z}$. The assumed mean output, $GSK_{n,z}$, is dependent on the number of generation nodes. Derived from (1), the change in the zonal PTDF value with a constant nodal PTDF and a dynamic GSK is obtained as

$$PTDF^{Z}_{l,z} = \sum_{n} PTDF^{N}_{l,n} \cdot (GSK_{n,z} + \Delta GSK_{n,z}(t)) \ , \tag{10}$$

where l is the critical transmission element, z the number of zones and n the number of nodes. Since in the simulation, the change, $\Delta GSK_{n,z}$, is obtained from the normal distribution there is also a need for normalization for the resulting GSK matrix by the definition of the GSK parameter. Therefore, the results of the GSK values for each node, n, and zone, z, combination are normalized, so that the sum of the values equal 1 per bidding zone.

With four nodes, $n = 4$, the nodal PTDF is assumed to be as in equation (4), where the nodal PTDF of the transmission line l_2, $\boldsymbol{PTDF}^{N}_{l_2} = [0.25,\ 0.625,\ 0,\ 0.125]$ is chosen for illustration, the footnote l_2 referring to the second row of the matrix. The choice is here arbitrary, and transmission line l_2 is chosen for illustrative purposes, since there are clear differences between the individual nodal PTDF values for the line. The system is considered as a single zone and $z = 1$. The generation nodes are expected to participate equally most of time (which may not always be the case in the physical world). Therefore, with four nodes the assumed mean participation is then $GSK_{n,z} = 0.25$ for each node in the bidding zone.

For the case of eight nodes, the simulation is made similarly with the mean $GSK_{n,z} = 0.125$ for all, n, and the system corresponding to a single zone $z = 1$. An

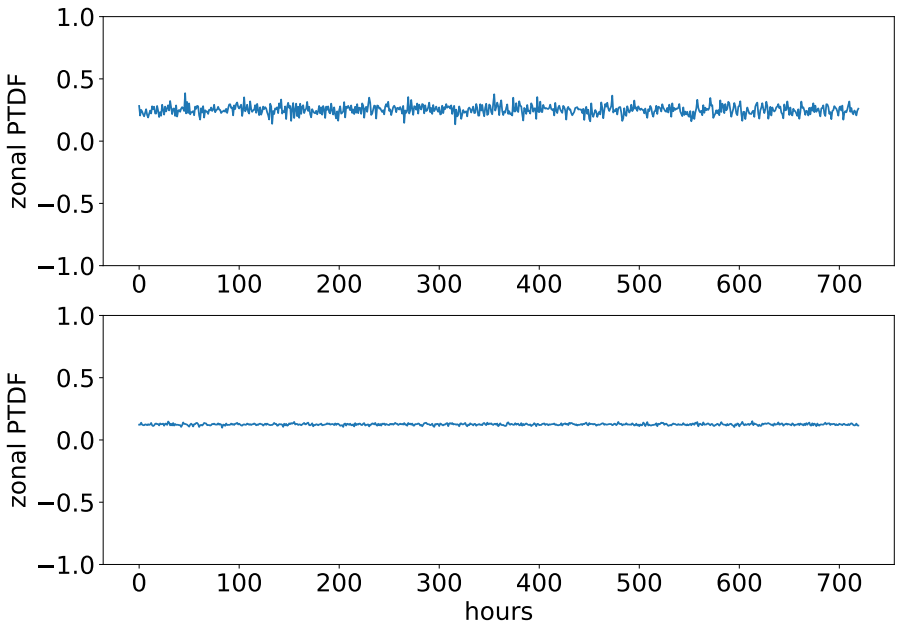

Figure 9: A simulation of the change in zonal PTDF value with a constant nodal PTDF and a dynamic GSK strategy. Above, the number of generator units is four, where a clear change in the zonal PTDF value is observed. Below, the unit number is eight and the effect of change in the GSK value is less significant.

arbitrary nodal PTDF row

$$\boldsymbol{PTDF}^{N}_{l_8} = [0.1,\ 0.2,\ 0.05,\ 0.125,\ 0.125,\ 0.25,\ 0.1,\ 0.05] \tag{11}$$

is designed for illustration, where the footnote l_8 similarly refers to row values. The exact values of such a nodal PTDF row in an eight node system would depend on its actual topology and electrical characteristics. Here, the values are set to be close to the compared nodal PTDF row, $\boldsymbol{PTDF}^{N}_{l_2}$, and the sum of the row values is set close to 1 (which can differ depending on the system) as an approximation for a nodal PTDF row in an arbitrary eight node system.

The results of the simulation are in Figure 9. It can be seen that the dynamic GSK strategy has less effect the more generation nodes are taken into account. The number of generation nodes for one bidding zone in practise, can be much higher than 8 (for example, Finland has around 400 power plants distributed to nodes throughout the country [37]), thus the effect of the change in GSK on the zonal PTDF can be observed to be small. Moreover, for an observation time more than a month (>720 hours), the changes in the zonal PTDF value due to changes in GSK become less significant if the relative contribution of each generation node is expected to stay around a mean value.

Research on the effect of different GSK strategies on the FBMC domain also indicate a minor influence. According to [32], on an aggregated level, existing

literature suggests that GSK strategies often have a small effect on the FMBC results when comparing the influence of different strategies globally. In addition, when comparing the effect of different strategies per bidding zone, the difference in performance remain less than 4% for most bidding zones, with few exceptions at 8 %. For some bidding zones, such as NO1, the differences are considered negligible [38]. This suggests that a static GSK strategy has a total effect on the zonal PTDF values close to that of a dynamic one. Moreover, a flat strategy is often related with an average performance. However, power flows on individual critical network elements (CNEs) have been shown to vary substantially depending on the GSK strategy. Therefore, when approximating the effect of GSKs on the zonal PTDFs the total effect of a GSK strategy on the zonal PTDF matrix is less significant, although errors can be found on the level of individual zonal PTDF values of CNEs [32, 38, 39].

4 A Linear Model of the Zonal PTDF

A model for forecasting the zonal PTDF values is presented. The model is first derived by using the theoretical framework presented in previous section and then applied to example data.

4.1 Theoretical Derivation

As described in the previous sections, the zonal PTDF is dependent on the nature of its constituents. The change in the grid topology due to planned outages will affect nodal PTDF for the duration of the outages. It will push the zonal PTDF to another static state and when the outage is taken care of, the zonal PTDF value will then return to its original state. In addition, depending on the GSK strategy chosen by the TSO, the change in the zonal PTDF can either be completely defined by this change in the nodal PTDF in a static GSK strategy, or it can be the combination of the change created by a dynamic GSK strategy and the nodal PTDF.

From this description a general mathematical model for forecasting the zonal PTDF may be derived. Let us assume we can define the load flows, $\Delta \boldsymbol{L}$, after a topological change, i.e., how the load of a network element is distributed among the remaining network elements after an outage. The effect of multiple outages is described as the linear sum of the effect of individual load flows, $\Sigma_j \Delta \boldsymbol{L}_j(t)$. Moreover, the change in the GSK values will be described with the parameter $\Delta \boldsymbol{GSK}(t)$. Then, the equation for the zonal PTDF at any given time unit, $\boldsymbol{PTDF}^Z_{fc}(t)$ (footnote, fc, refers to the forecasted value), with the effect of any GSK strategy and the nodal PTDF outages can be defined from the equations (2), (7) and (10) as

$$\boldsymbol{PTDF}^Z_{fc}(t) = (\boldsymbol{PTDF}^N + \Sigma_j \Delta \boldsymbol{L}_j(t)) \times (\boldsymbol{GSK} + \Delta \boldsymbol{GSK}(t)) \ , \tag{12}$$

where $\boldsymbol{PTDF}^N$ is the nodal PTDF of the network before any topological changes and $\boldsymbol{GSK}$ the mean of the GSKs in the dynamic strategy. Depending on the dynamic GSK strategy, $\boldsymbol{GSK}$ can correspond to the relative power outputs of the power plants and is equivalent to the $\boldsymbol{GSK}$ matrix of a static GSK strategy.

Based on the analysis of Section 3.3, the effect of different GSK strategies on the net position of a bidding zone is relatively small. Moreover, the change mediated by a dynamic strategy is less significant on individual zonal PTDF values of CNEs when higher number of power plants are present. Therefore, in this case the model will assume a static GSK strategy in the calculation of the zonal PTDF values, and the parameter $\Delta \boldsymbol{GSK}(t)$ is approximated as zero, leaving behind a constant $\boldsymbol{GSK}$ value. The equation then becomes

$$\begin{aligned} \boldsymbol{PTDF}^Z_{fc}(t) &= (\boldsymbol{PTDF}^N + \Sigma_j \Delta \boldsymbol{L}_j(t)) \times \boldsymbol{GSK} \\ &= \boldsymbol{PTDF}^N \times \boldsymbol{GSK} + \Sigma_j \Delta \boldsymbol{L}_j(t) \times \boldsymbol{GSK} \\ &= \boldsymbol{PTDF}^Z + \Sigma_j \Delta \boldsymbol{L}^Z_j(t) \ , \end{aligned} \tag{13}$$

where $\boldsymbol{PTDF}^N$ and $\Sigma_j \Delta \boldsymbol{L}^Z_j(t)$ resemble the zonal values of the respective parameters.

In practice, the parameter $\boldsymbol{PTDF}^Z$ describes the static state of the zonal PTDF value to which $\boldsymbol{PTDF}^Z_{fc}(t)$ returns in the absence of outages. $\Sigma_j \Delta \boldsymbol{L}^Z_j(t)$ on the other hand, describes the load flows caused by the outages after multiplied by the static GSK factors. Note that the formulation of the equation resembles exactly that of the nodal PTDF, a linear sum of a constant value and step functions.

4.2 Application to Market Data

For a market participant outside the TSOs the exact conditions of the grid are unknown. Therefore, the parameters $\boldsymbol{GSK}$ and $\boldsymbol{PTDF}^N$ are not available for the calculation of the zonal PTDF as depicted in the equation (13). In this situation, the transformation of the matrix products, $\boldsymbol{PTDF}^N \times \boldsymbol{GSK}$ and $\Sigma_j \Delta \boldsymbol{L}_j \times \boldsymbol{GSK}$, in (13) to the respective zonal components, $\boldsymbol{PTDF}^Z$ and $\Sigma_j \Delta \boldsymbol{L}^Z_j$, serves a purpose. The market participants are given only the zonal data for the analysis of the markets and thus, the components $\boldsymbol{PTDF}^Z$ and $\Delta \boldsymbol{L}^Z$ need to be defined from the zonal data directly.

A simplified example is given to demonstrate the application of the model in the case of real market data. Assume the historical zonal PTDF data of an critical network element is the following shown in Figure 10. The CNE has experienced two outages in the past, at $hour = 200$ for a duration of 35 hours and at $hour = 324$ for a duration of 12 hours. These outages are simulated as step functions as explained in Section 3.2.2, where base value of the zonal PTDF is produced as in the example of Section 3.3.2 equation 10 with 8 generators. A planned outage is set for the CNE for

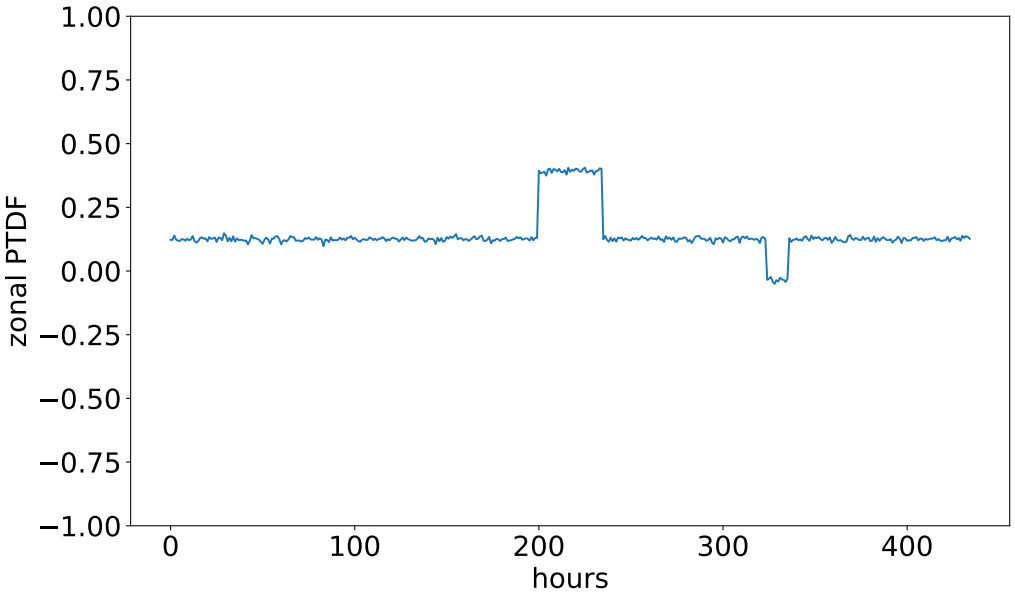

Figure 10: Simulated market data.

an upcoming hour at date $hour = 623$, which is expected to affect the grid similarly

as the historical outage seen in the data at *hour* = 200. No other outages are planned for the network element in the near future.

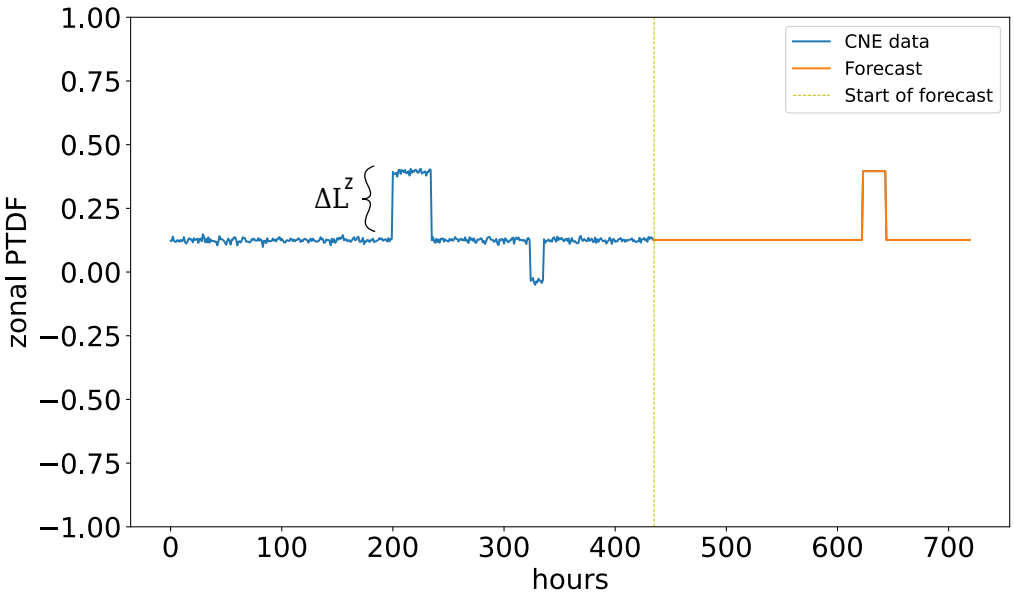

Figure 11: A forecast for the simulated market data.

For this CNE, a forecast can be made based on the information. The static component of the forecast $\boldsymbol{PTDF}^Z$ can be obtained from the median of the critical network element historical data. Next, the effect of the planned outage, $\Delta \boldsymbol{L}^Z$, on the zonal PTDF is approximated from the historical data as the effect of the known outage at *hour* = 200. This is obtained by computing the difference of the zonal PTDF value at the occurrence of this outage as shown in Figure 11. This effect is then added to the median as a step function as demonstrated by the equation (13) of the model. This way the forecast of the zonal PTDF value of the CNE is obtained as in Figure 11.

In this example, an additional assumption was made on the functioning of the grid which simplified the approach considerably. Historical outages are not reliable predictions of future planned outages, since the grid can experience changes in its topology and electrical properties over time (some of which were described in Section 3.2). For estimating the effect of outages as a market participant, reports for planned outages have been given which give a detailed effect of the future outage [40]. These reports would be utilized when calculating the load flow distributions. However, for the current state of the FB implementation, this reporting is yet to be developed for the FBMC [41]. Therefore in this analysis, the method for approximating the effect of future outages is based on historical outage data and the current outage reports in NTC methodology.

5 Forecasting the Zonal Power Transfer Distribution Factors

The theoretical model developed in the previous section for forecasting zonal PTDF values is now applied to real market data. The initialization of the model for the forecast is first presented, with detailed explanations on the required data sets and metrics. The results of the forecast are then computed, and the effectiveness of the model is evaluated.

5.1 Setting the Forecast

There are two data sets that are needed to make the forecast; the zonal PTDF data of the market area in question and the outage report data of the same market, which gives details on the effect of outages. In addition, there needs to be a metric which links the zonal PTDF data with future and past outages. All these three inputs and their application are presented in this subsection.

5.1.1 Zonal PTDF Data

The zonal PTDF data used for the model comes from the External Parallel Run (EPR) of the Nordic FB methodology, which relates to its implementation for the day-ahead and intraday market time frame in the Nordic electricity markets. It is run by Nordic TSOs and the Nordic SRC in respond to the CAMC regulation, starting in 8th of March in 2022. During the EPR, the FB parameters will be calculated parallel to the real market and published daily on the JAO publication tool website [42]. The EPR defines whether the methodology is ready for go-live, and the transition from NTC method to FBMC can be done.

For the testing of the model, the data from the EPR is used from 8th of March in 2022 to 6th of January in 2023 (the starting date of the thesis). The data is obtained from the JAO publication tool [42], which in essence includes the zonal PTDF values for each CNEC in the Nordic grids chosen by the TSOs with respect to each bidding zone in the Nordics. However, only data from the TSOs of Norway, Denmark and

	BZ1	**BZ2**	$\cdots$	**BZn**
CNEC1	0.02	0.005	$\cdots$	0.09
CNEC2	0.20	0.03	$\cdots$	0.10
$\vdots$			$\ddots$	$\vdots$
CNECn	0.01	0.07	$\cdots$	0.45

Table 3: The zonal data in matrix format, where CNECs are as rows and bidding zones (BZ) as columns. A similar matrix is for each time unit.

Finland are non-anonymous, leaving out the data of Sweden. Due to the quality of outage report data (explained in detail in Section 5.1.3) the analysis limits the zonal PTDF data to CNECs under the TSO of Norway. In addition, only CNECs which

have data for the whole time period of the EPR data set are chosen, thus 50 out of the 163 CNECs under the TSO of Norway are included in the data set. These are the network elements which most often restrict the FB domain. This data from JAO is then processed into a matrix format, where the bidding zones of the Nordic market area act as columns and the CNECs of the TSO of Norway as rows, shown in Table 3.

5.1.2 Application of the Data

Before introducing the other data set and the metric, a brief view is done to obtaining the necessary values, $\boldsymbol{PTDF}^Z$ and $\Delta\boldsymbol{L}^Z$, from the zonal PTDF data set as was demonstrated in the example of Section 4.2.

The exact value of the static parameter $\boldsymbol{PTDF}^Z$ is unknown. However, due to the nature of the nodal PTDF, the zonal PTDF should always return to the base value, $\boldsymbol{PTDF}^Z$, when no topological changes are present and when the effect of the GSK is approximated as constant. Due to this property the parameter $\boldsymbol{PTDF}^Z$ can be approximated by a median value of the zonal PTDF time series. The median emphasizes the value in which the zonal PTDF of a CNEC spends the most time on and leaves out the deviations due to the outages, which commonly last some days. The median is calculated for each CNEC and bidding zone pair in the zonal PTDF data.

The zonal load flow distribution for each outage, $\Delta\boldsymbol{L}^Z$, on the other hand, can be approximated from the zonal PTDF data when the hourly event date of an outage is known, denoted here as T (this date value is obtained from the outage reports, see Section 5.1.3). This value for the load flows of the outage is then calculated from the difference of the zonal PTDF value before the recorded starting period of the outage, $t_1 = T - x$, to the value of the zonal PTDF during the outage $t_2 = T + x$. If $\boldsymbol{PTDF}^Z_{jao}(t)$ refers to the JAO publicaton tool data at a time unit, t, [42], the value is then defined as

$$\Delta\boldsymbol{L}^Z = \boldsymbol{PTDF}^Z_{jao}(t_2) - \boldsymbol{PTDF}^Z_{jao}(t_1)\ . \tag{14}$$

In this analysis the constant, x, is set to the value of 3 hours. Once the outage is set to end, the value is assumed to change back a similar amount, $-\Delta\boldsymbol{L}^Z$. This describes a step function as explained in Section 3.2.1. Whether there is a change in the value of the static parameter, $\boldsymbol{PTDF}^Z$, during the outage, for example in a monthly forecast in which the static parameter is also updated monthly, the difference of the preceding and the new static value is reduced from the load flow distribution, $-\Delta\boldsymbol{L}^Z$, thus also updating the step function. The total effect of the outages on the whole timeline of the forecast is the linear sum of the step functions defined by the individual load flows of outages.

5.1.3 Outage Report Data

The other needed data set is collected from outage reports, which are provided by TSO's. In order to inform the market the timing and effects of outages, TSO's are

required to publish data of planned/unplanned unavailability in transmission to transparency platforms such as ENTSO-E [43]. The reports carry information of the available capacity of the line in question and the date, duration, and (approximate) location of the outage event. The reports also include other border connections which are unavailable at the same time period (either affected by the outage of the grid component or otherwise unavailable) and their individual NTCs [44]. An example of the report is given in Figure 12. These reports are made with the assumptions of the NTC method (the implementation for a flow-based reporting on outages is set to go live year after the FB method itself [41]), which provides an extra challenge on determining the effect of outages on the FB parameters. However, in the absence of more developed data, the metric for choosing the corresponding historical outage (see Section 5.1.4) is done by utilizing these outage event reports, which currently give the best approximation on the characteristics of an outage.

	A	B	C	D	E	F	G
1	**Unavailability in Transmission Grid**						
2	Planned Unavailability in the Transmission Grid [10.1.A]						
3	Changes in Actual Availability in the Transmission Grid [10.1.B]						
4	**Event start - Event Stop (Time Zone)**	**Publication Date/Time (Time Zone)**	**Publisher**	**Event Status**	**Type of Unavailability**	**Version**	**Reason Code**
5	24.10.2023 09:00 - 24.10.2023 15:00 (UTC)	19.10.2023 12:37 (UTC)	Statnett	Active	Planned	3	Foreseen Maintenance
6	**Remarks**						
7							
8	**Reason for The Unavailability**						
9	Impedance measurement.						
10	**Affected Assets or Unit**	**Affected Assets or Unit EIC Code**		**Market Participant**		**Market Participant Code**	
11	M 420 Nea-Klabu	50TL00000001192F		Statnett		10X1001A1001A38Y	
12	**Message ID**						**Unit of measurement**
13	NU_g3EY2lhQkKlPzOpQHc0iPA_003						MW
14	**Interval start - Interval Stop (UTC)**	**Biding Zone**	**Installed Capacity [MW]**	**Available Capacity [MW]**		**Unavailavle Capacity [MW]**	
15	24.10.2023 09:00 - 24.10.2023 10:00	SE2 > NO4	300	250		50	
16	24.10.2023 10:00 - 24.10.2023 11:00	SE2 > NO4	300	250		50	
17	24.10.2023 11:00 - 24.10.2023 12:00	SE2 > NO4	300	250		50	
18	24.10.2023 12:00 - 24.10.2023 13:00	SE2 > NO4	300	250		50	
19	24.10.2023 13:00 - 24.10.2023 14:00	SE2 > NO4	300	250		50	
20	24.10.2023 14:00 - 24.10.2023 15:00	SE2 > NO4	300	250		50	
21	24.10.2023 09:00 - 24.10.2023 10:00	SE2 > NO3	1000	0		1000	
22	24.10.2023 10:00 - 24.10.2023 11:00	SE2 > NO3	1000	0		1000	
23	24.10.2023 11:00 - 24.10.2023 12:00	SE2 > NO3	1000	0		1000	
24	24.10.2023 12:00 - 24.10.2023 13:00	SE2 > NO3	1000	0		1000	
25	24.10.2023 13:00 - 24.10.2023 14:00	SE2 > NO3	1000	0		1000	
26	24.10.2023 14:00 - 24.10.2023 15:00	SE2 > NO3	1000	0		1000	
27	24.10.2023 09:00 - 24.10.2023 10:00	NO4 > SE2	250	150		100	
28	24.10.2023 10:00 - 24.10.2023 11:00	NO4 > SE2	250	150		100	
29	24.10.2023 11:00 - 24.10.2023 12:00	NO4 > SE2	250	150		100	
30	24.10.2023 12:00 - 24.10.2023 13:00	NO4 > SE2	250	150		100	
31	24.10.2023 13:00 - 24.10.2023 14:00	NO4 > SE2	250	150		100	
32	24.10.2023 14:00 - 24.10.2023 15:00	NO4 > SE2	250	150		100	
33	24.10.2023 09:00 - 24.10.2023 10:00	NO3 > SE2	600	0		600	

Figure 12: A view of the outage report excel file by one of the sites for transparency; Nordic Unavailability Collection System (NUCS) [44]. The reports in NUCS platform are the source of outage report data used in the analysis.

Since there are multiple outages happening in different zones and different connections throughout the year, it is difficult to asses the performance of the model without limiting the data. Therefore, a particular border connection between the bidding zones NO1 and SE3 is investigated and only outage reports in this location are analyzed. The connection is chosen due to its high activity in terms of outages during the whole EPR zonal PTDF data set. In general, the period of the available zonal PTDF data set is short; around 10 months. Finding a connection which provides enough occurrences of outages for this time interval is crucial. For the said border connection, 26 transmission line outages was found in the time period from 8th of March in 2022 to 6th of January in 2023. These 26 outage reports will comprise the outage report data set in this analysis.

The downside of limiting the outage data is the noise in the zonal PTDF values due to outages in other bidding zones and border connections outside the selection. Nevertheless, most connections have too few occurrences of outages in the timeline

of the EPR zonal PTDF data to yield reliable results to be included in the analysis. This limitation can affect the accuracy of the model's forecast depending on the amount of outages in the network of the market area. However, majority of the effect caused by the outage is seen in bidding zones (physically) close to the location of the connection affected by the outage. To reduce the effect of noise on the results, this analysis will focus only on zones close by, which will further limit the data. As the chosen connection is between a Norwegian and a Swedish bidding zone, only the zonal PTDF data from the TSO of Norway will be used in the analysis (data from Swedish bidding zones would also be preferred, however, the data is anonymous and therefore, unusable).

5.1.4 Selection Metric for the Outage

As introduced previously in this subsection, there are two data sets which provide all the available information for the analysis of the model. The data of the outage reports on the chosen border connection, NO1-SE3, now needs to be combined with the zonal PTDF data from the affected bidding zones.

Once the effect of the outages have been defined as in Section 5.1.2, a metric is made to combine this information with the outage report data and determine the effect of future outages on the forecast of the zonal PTDF data. The goal of the metric is to compare historical and future outage reports and determine which outages can be approximated to have the same effect. Consequently, a future outage event will have the exact effect on the zonal PTDF values as a similar outage in the past. To define the similarity between outages, the metric focuses on comparing the similarity of constants in the outage event reports. The similarity is defined by the number of affected border connections, n, the reported NTC value, c, and the location, x, of the outage. A crude assumption is made so that a future outage with values n_f, c_f and x_f in the outage report (footnote, f, referring to a future value), corresponds to a historical outage with the same n_h, c_h and x_h values (footnote, h, referring to a historical value). Moreover, whether such an outage is not found in the historical data, the metric approximates the effect of the outage to be similar to an outage closest to these values.

The constant, x, is kept at an absolute and the metric only compares outages within the same location. For the other constants, the closeness is defined by calculating the distance between the values in comparison. Assume the constant in comparison is denoted as c_f for the future outage. This value is then compared to the historical outage data set, where each outage has a similar value denoted as c_h^j, where j is the outage number. Then the distance is calculated by the difference

$$d_j = |c_f - c_h^j| \tag{15}$$

for each outage j. The minimum distance is then defined as

$$min\{d_1, d_2, ..., d_j\} \ . \tag{16}$$

This procedure is done to the constants n and c.

When comparing outages and an exact match is not found ($d_j \neq 0$), the metric prioritises first the location, x, then the number of affected border connections, n, and lastly the NTC value, c. Only outages that are in the same location x are compared. The number of border connections, n, are rounded down to the closest, i.e. if the future outage affects 3 border connections, and there are only outages that affect 2 or 4 connections in the historical data (thus making the distance variable equal for two different outages), the outage with $n = 2$ is chosen. Lastly, the metric compares the NTC values. If no exact match is found for the NTC value, the metric will only compare outages with greater NTC values (the greater the available NTC, the less impact the outage has), and similarly chooses the one with least distance. The metric therefore, prefers an outage which parameters indicate a lesser impact on the zonal PTDF values, when an exact match is not found.

Roughly half of the outages (14) are set as historic data for the metric, for which the effect of forecasted outages in the rest of the data set will be based on. This will also define the historic data for the zonal PTDF data set. The median for each element will be taken from the period 8th of March to 7th of August, which comprises the historic outages, from the zonal PTDF data set. By using the two data sets and the metric, a monthly forecast is produced for the remaining time interval 8th of August to 6th of January, where the historic data (both the median data and the outage report data) is updated after each month.

5.2 Results of the Forecast

As described in previous subsection, the model applies two data sets and a metric to forecast the zonal PTDF values. In the upcoming sections, the performance of the model is analyzed on the zonal PTDF values of this selected zonal PTDF data set. First, the results for individual CNECs are discussed, followed by an analysis of the entire matrix comprised of these network elements.

5.2.1 Individual Elements

In total, the zonal PTDF values of 50 network elements are modelled for the forecasted time period. The actual data from the time period is plotted in comparison for the predicted values. Figures 13, 14, 15 show samples of the results for the individual elements with differing success in the forecasted values. Here, the blue line represents the real data and the red one the forecasted values. A vertical dotted line is plotted as the start of the forecasted time-period, thus also illustrating the amount of historical data used in the beginning as the training data set. Four columns from the Norwegian bidding zones are chosen for illustration purposes for each of the individual elements.

In Figure 13, an example where the model has successfully predicted multiple of the outages is shown. The actual element data in the occurrence of an outage closely mimics the theoretically expected behavior explained in Section 3.2.2 on the effect of the outage to the zonal PTDF value. The model reacts similarly to this behaviour, expecting an abrupt jump in the zonal PTDF value before returning back to the base value after the outage ends. The forecast can be observed to overlap with the

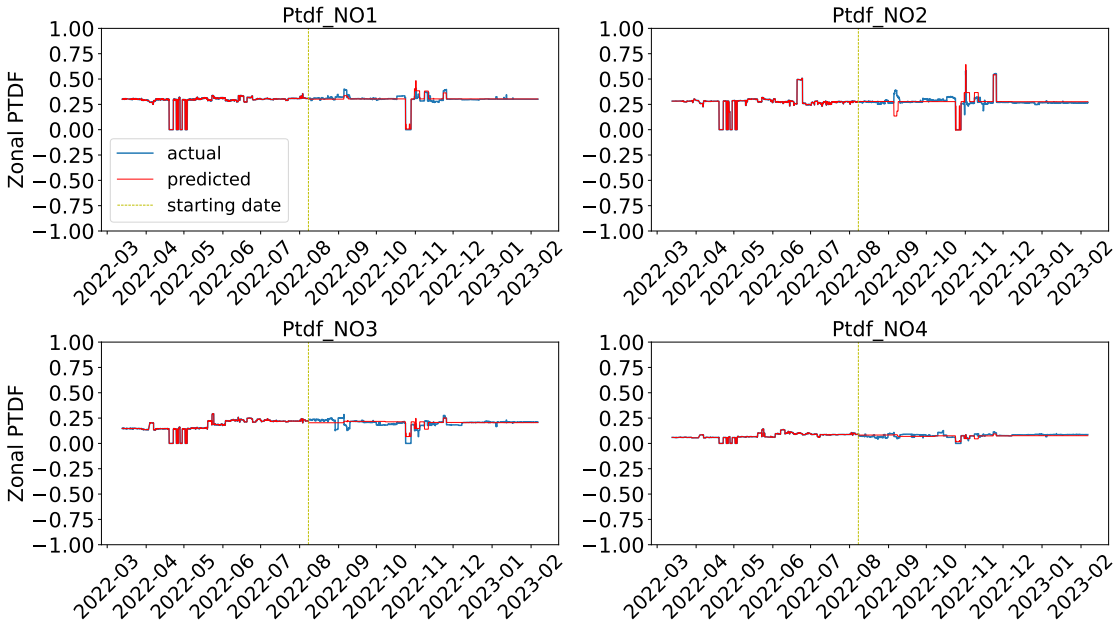

Figure 13: The result of the forecast for the individual values of a single network element for four bidding zones. Here the forecast the overlaps with the real value, depicted in blue, and the predicted value, depicted in red, for multiple outages. The dotted vertical line in yellow, illustrate the start of the forecasted period, before which the real historical data is shown.

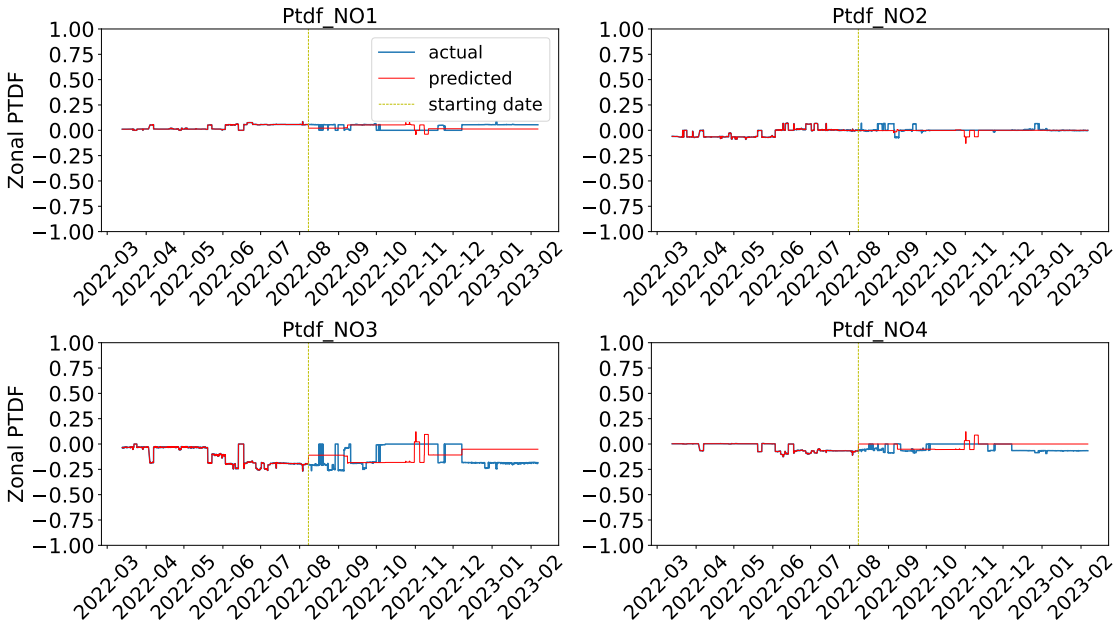

Figure 14: The result of the forecast for the individual values of a single network element for four bidding zones. Here the forecast the fails to predict the real values, thus having less overlap between the graphs.

actual data on outages after the date 2022-10. In addition, the median part of the forecast stays relatively same during the whole time period and close to the base value into which the actual element data returns to after the end of an outage.

For the second example, an element where the forecast misses the effect of the outages is shown in Figure 14. In addition to the noise created by unaccounted outages, it can be seen that the outages predicted from the history data are not in line with the real data. For example, the forecasted effect of outages around 2022-10 show a major effect to the zonal PTDF, whereas the real data show little to no effect. In addition, due to the long time period of some of the outages, the median part of the forecast fails to catch the base value of the network element in some of the bidding zones, mainly NO3.

The last example illustrates an element which experiences little to no change in its zonal PTDF value due to outages. Here, the median of the forecast predicts the real value sufficiently for three of the bidding zones (NO1, NO3 and NO4). The forecast also predicts the effect of outages correctly, i.e. with little to no effect on these bidding zones.

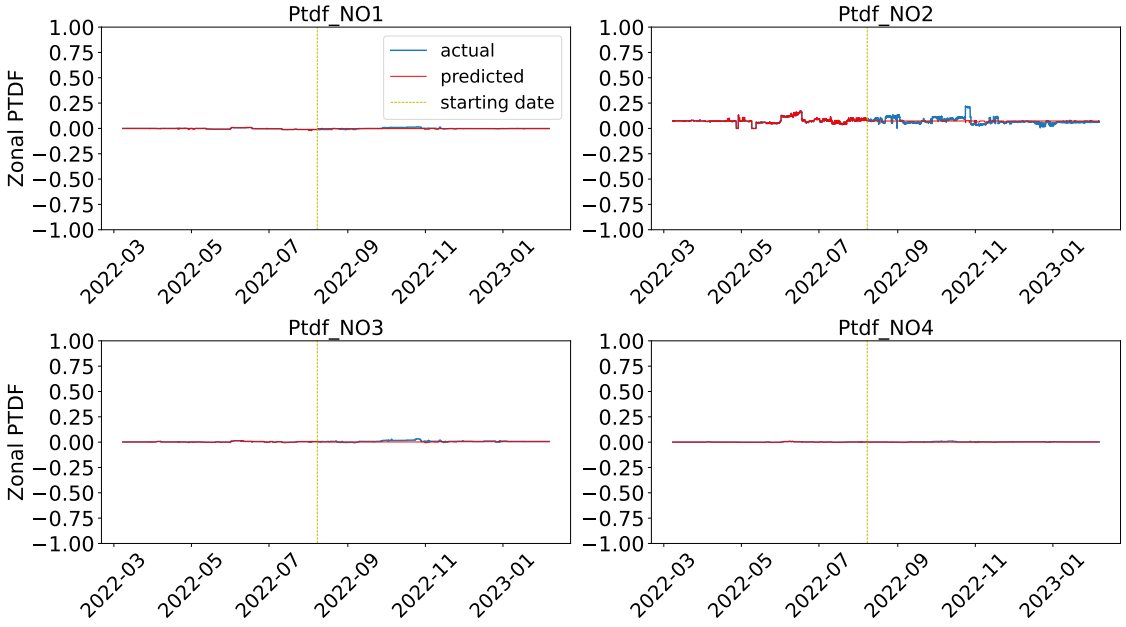

Figure 15: The result of the forecast for the individual values of a single network element for four bidding zones. Here the forecast is applied to an element with little to no change in its zonal PTDF value. The forecast overlaps with the real data of the element for three of the bidding zones: NO1, NO3 and NO4.

In general, significant error is seen on the level of individual zonal PTDF values of CNECs. As described in Section 3.3.2, different GSK strategies have a direct impact on the zonal PTDF values of individual CNECs. By using a static GSK strategy in the model as approximation, the individual zonal PTDF values can have similar error when the assumed static strategy differs from the strategy used by the TSO in the EPR. This is indeed the case, as the Stattnett has a dynamic GSK strategy

as defined in Table 4. In addition as described in Section 4.2, historical outages are not fully comparable with future outages. This attributes to the error in some of the predicted outages. Moreover, the NTC method behind the outage reports limits the effectiveness of the metric. Other sources of error are the noise due to outages outside the selected reports, and possible TSO specific changes to the nodal PTDF or GSK.

5.2.2 The Zonal PTDF Matrix

The data from all of the 50 elements forecasted by the model is combined, thus creating the prediction of the zonal PTDF matrix for the time period. In order to analyze the effectiveness of the model in depth, a residual plot is made for the forecasted values. The residuals are measured from the hourly starting date, t_{start}, of the outage, $t_{residual} = t_{start} + 3\ hours$, where the change to the zonal PTDF value is expected to already been happened. The difference is calculated as $residual = \hat{y}_i - y_i$, of the actual zonal PTDF value y_i to the predicted zonal PTDF value, $\hat{y}_i$, at the said time unit, $t_{residual}$. Therefore, there exists one data point for each forecasted outage of every element for each bidding zone of the element

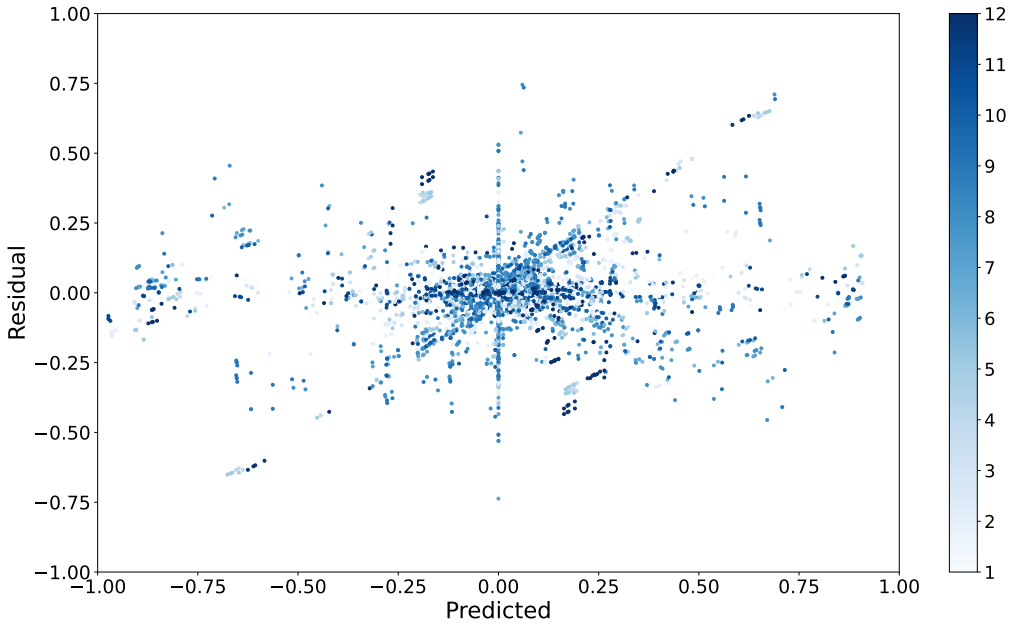

Figure 16: The residual scatter plot of the forecasted data set. Each shade of blue corresponds to a different outage, with darkest shade for the most recent one.

($rows \times columns \times outages = 50 \times 31 \times 12 = 18000$ data points in total). The residual analysis, therefore, compares whether the value of the zonal PTDF after an outage is correctly predicted by the model.

The residuals of the data set are shown in the scatter plot in Figure 16. Each outage is plotted with a different shade of blue, where the most recent outages have a darker tone. The color bar is therefore set from 1 to 12 for the number of outages.

For example, the residual of the effect of the first outage of the forecasted period (color bar code 1) on all of the network elements is plotted with the lightest shade of blue.

A histogram of the residuals is shown in Figure 17. Lastly, in order to highlight the performance of the model on the vast data, summary measures of the residuals across each outage is plotted in Figure 18. The plot shows the Mean, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Median of the residuals over time as the training data is updated after each month.

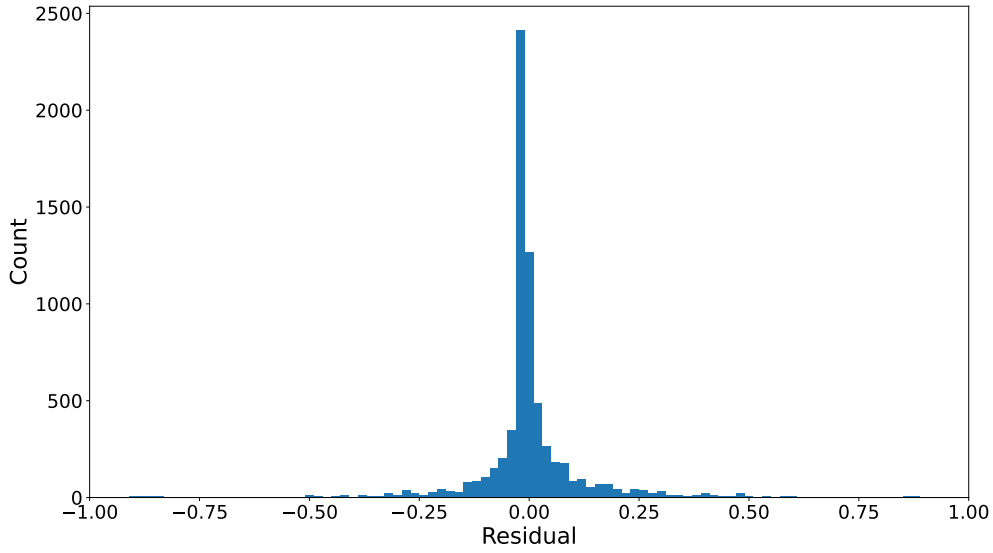

Figure 17: Bar plot of the residuals of the forecasted data set.

These graphs show the model produces forecasts that take into account most of the available information, but not quite. In the residual scatter plot, Figure 16, some patterns are observable through the data; an incline line and a vertical line through the origin. This suggests that there are predictable parameters missing from the model. However, these consistencies are for minor part of the data and good residual patterns, i.e., symmetrical distribution and clustering to lower digits, are found for the remaining data. For example in Figure 17, highest count is found for the value of zero and the count cascades towards the ends of the x-axis to higher values indicating clustering towards lower digits. In addition, the values in the histogram are evenly distributed around the mean (highest count), suggesting a symmetrical distribution for the residuals.

Comparing the summary measures in Figure 18, similar results are obtained for the median and the mean in each outage. The mean of the residuals provides an estimate of the average deviation of the model's predictions from the actual observed values. If the mean of the residuals is close to zero, it suggests that, on average, the model is unbiased. The values observed in Figure 18 are slightly to the negative but extremely small. A slight bias could be possible. Lastly, the median measures

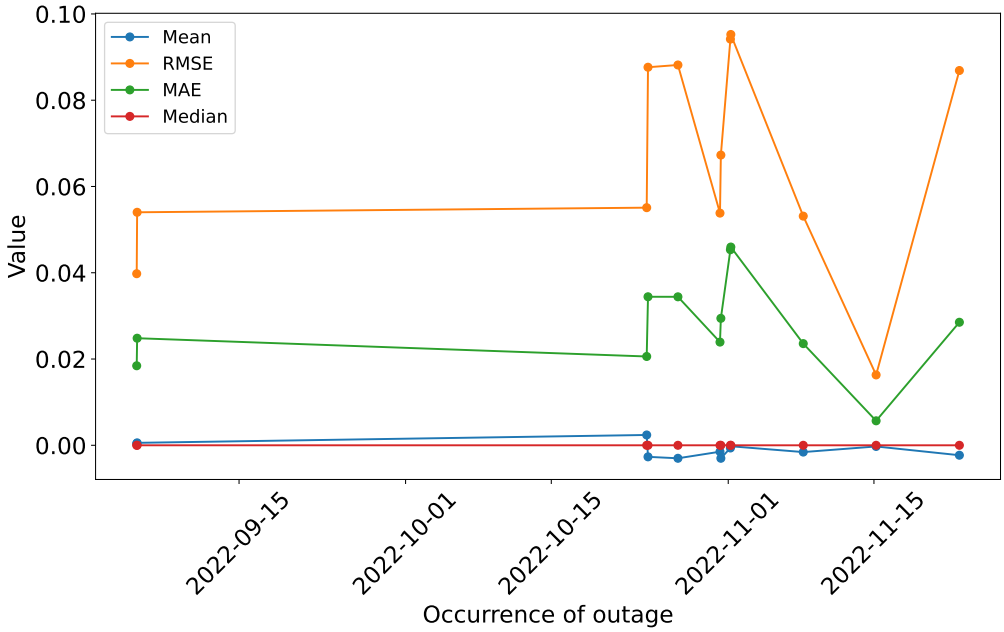

Figure 18: The summary measures of the residuals of the forecasted data set plotted for each outage. The Mean, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Median of the residuals are shown over time as the training data is updated after each month

the central tendency and can be less sensitive to outliers compared to the mean. It provides an estimate of the typical magnitude of the errors. This is found to be zero.

The summary measures of RMSE and MAE in Figure 18 also indicate a relatively good accuracy. The RMSE of the residuals provides a measure of the average magnitude of the errors. Lower RMSE values indicate better performance. When scaled to the measurement data, the maximum RMSE of the residuals is 0.048, indicating a $\sim 5\%$ error. MAE provides a measure of the average absolute deviation between the predicted and observed values, without considering the direction of the errors. As with the RMSEs, these values are relatively low and below $\sim 3\%$.

Part of the accuracy can be attributed to the nature of the zonal PTDF data. When an outage occurs on a network element, the effect is the highest in the immediate vicinity of the element. This means that the change in the zonal PTDF values diminishes for zones which are physically further away. For example, in Figure 15 the effect of the outage in bidding zones NO1, NO3 and NO4 has already diminished compared to bidding zone NO2, and the zonal PTDF values are almost constant. This type of effect is relatively trivial for the model to predict, since the change in the zonal PTDF values is low and the model uses a median value as part of the forecast. In the EPR data, much of the values that are attributed to bidding zone columns physically far away from the occurrence of the outage has similar values as in the aforementioned zones, NO1, NO3 and NO4. This attributes to the low error for the zonal PTDF matrix in the model. In addition, as described in Section 3.2.2, the effect of different GSK strategies on the whole zonal PTDF matrix (which defines

the net positions) is less significant than on individual elements. As is observed, on the scale of the whole zonal PTDF matrix the approximated flat strategy fares better than on the scale of individual CNECs.

5.2.3 Further Analysis of the Results

A closer examination of the symmetry and randomness of the distribution of the residuals is done with a statistical inference hypothesis testing. A p-value is used to decide whether the difference is large enough to reject the null hypothesis: If the p-value of the test is larger than 0.05, the assumption of normal distribution is not rejected. If the p-value of the test is smaller than 0.05, the assumption of a normal distribution is rejected. For this analysis, the Shapiro-Wilk (SW) test is chosen due to the large size of the data set. The results are shown in Table 4. The

Test	Statistic	p-value
SW	0.681	~ 0.00

Table 4: The results of the statistical inference tests.

statistic is defined as the maximum value of the difference between tested data set and normal cumulative distribution functions. Here, the p-value is smaller than 0.05 and therefore, the error in the forecasted values does not follow a normal distribution.

This suggests that the model does not account for all predictable phenomena in the forecasting of the zonal PTDFs. With large samples, however, even minor departures from normality may lead to a rejection of the null hypothesis. Therefore to support the result, a visual inspection is done with a Quantile-Quantile (QQ) plot. As illustrated in Figure 19, the quantiles of the sample data plotted against the quantiles of the theoretical distribution, depicted in blue, exhibit a noticeable deviation from the ideal linear trend represented by the red dotted line. This departure from normality, supports the SW test result.

Another essential remark deduced from Figure 18 is the distribution of the occurrences of the outages. Most of the outages are clustered in between few months of the zonal PTDF data set. The months October and November include all but one outage. This implies that the historical outage data is only updated for November and December, rather than gradually improving each month. This weakens the accuracy of the prediction for the outages in the month of October as only one outage has been added to the historical data before. Moreover, the rest of the outages occurring in November are predicted with the same historical outage report data set, which is non-optimal for similar reasons.

Despite the increase in outage data after the October, there does not seem to be a clear improvement in the performance of the model between the rest of the outages, as indicated by the RMSE and MAE. The RMSE and MAE values do reach the minimum in November, but bounce back to a higher value for the last outage. The model's lack of consistent improvement after each month follows from the simplistic approach used in the prediction metric. The poor distribution of outages, on the other hand, results from the limited outage data. Both of these issues stem from

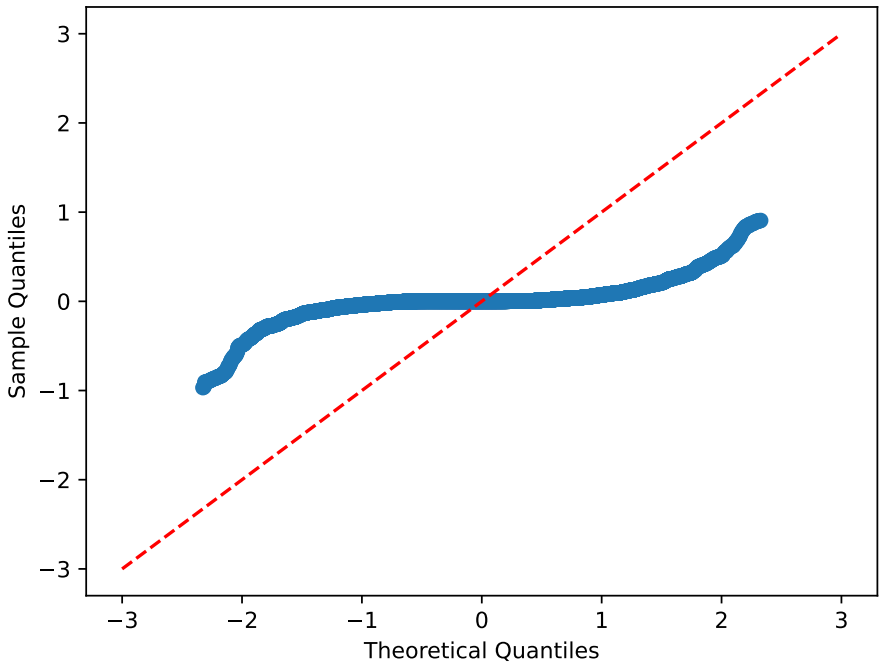

Figure 19: The Quantile-Quantile (QQ) plot reveals deviations in the sample data distribution from the expected normal distribution. As depicted, the data points (in blue) exhibit a distinct curvature away from the ideal linear trend (shown as red dotted line), indicating a significant departure from normality.

the recent implementation of FBMC in the Nordics, which is still a work in progress. With an extended training data set and improved reporting of outages with the FB methodology (as announced in [41]), these issues could be addressed, if not solved.

Currently, an improvement to the model could be done by improving the approach used in the metric. As described in Section 3.2, the topology of the grid is one of the major factors determining the nodal PTDF values, which in this model play the key role in the predicted values. Similar grid topologies exhibit similar properties. If the exact change in the topology of the grid can be pointed out accurately, for example which network element was actually affected by the outage, the following change in the PTDF values can be recorded for the said network element. Therefore, by coupling the historical outage to the forecasted outage precisely (elementwise for example), a major increase in accuracy can be achieved. A straightforward approach to increase the accuracy of the metric would be to use an Artificial Intelligence (AI) model to deduce a better logic between the relationship of future and historical outages from the current outage reports. In general, research on the dependencies of historical and future outages, and the similarities of their effect is needed for the improvement of the metric.

6 Conclusions

The development of the energy system for a more sustainable future has brought challenges to the energy markets and the ability to forecast capacities of the grid and prices of electricity. Increasing amounts of renewable energy sources in the grid with variable production couples the state of the energy system and its price to weather conditions. The main objective to alleviate the uncertainties due to the difficulty of reliably forecast the weather for longer time periods has been the development of grid flexibility and security of supply through new investments and regulations in Europe. One of the major developments is the transfer from the NTC capacity calculation to the FB capacity calculation through the CAMC regulation. Starting from CWE, the regulation is now set to launch also in the Nordic region in the near future. The FB method brings its own challenges to the market participants, who model the energy markets for trading in different time intervals. New components are added to the capacity calculation, and an accurate price forecast is dependent on the reliable forecast of the new parameters introduced by the method. So far only little research has been conducted on the topic of predicting the FB parameters, and especially on the prediction of the zonal PTDF.

In this thesis, a model for forecasting the zonal PTDF in the FBMC was built and its effectiveness on real data was tested. The model forms its basis on the physical theory of the individual components of the zonal PTDF, the nodal PTDF and GSK. The nodal PTDF describes the distribution of flows on the electrical connections and determines how much power flows through each connection when an electrical transfer is made between two nodes. The GSK, on the other hand, simplifies the nodal PTDF to a zonal format by providing the fraction of contribution of each generation node in the zone to the transfer. The contribution of each of these parameters to the changes in the zonal PTDF was simulated and the main contributor was determined to be the changes in the nodal PTDF. Changes in the nodal PTDF are mediated by changes in the physical reality of the grid where a connection in the grid is partly or completely disabled. The model assumed the effect of these changes from the historical data, and forecasted the future outages based on the information. A step function was used as a standard for the effect of an outage. In addition, the static nature of nodal PTDF was accounted in the model by a median value of the historical zonal PTDF values.

The results of the model were obtained for the elements of the Norwegian bidding zones, and the outages of a NO1-SE3 border connection were forecasted. A residual analysis revealed that on average the model performs relatively well and predicts most of the zonal PTDF values with small error. However, it could also be seen that there was room for improvement based on the residual plots and the additional statistical interference testing. In addition, major inaccuracies were discovered on the level of zonal PTDF values of individual CNECs. The difference in accuracy between the individual CNECs and the full PTDF matrix was in line with the current research on GSK strategies, which confirms part of the error. However, the rest of the inaccuracy in the model was deduced to be largely dependent on the metric, which defined the effect of the outages. Due to the limited availability of data, multiple

assumptions were needed on the functioning of the electric grid which simplified the approach used for the metric. The metric assumed the equivalence of historical and future outages, which is not fundamentally true. A possible improvement could be done with a well defined metric, either through extended data, updated outage reports or applying a AI model/different logic for the metric.

In general, this thesis introduced a new method for forecasting zonal PTDFs with reasonable accuracy for long time intervals. One key advantage of this method is that it relies solely on publicly available data, making it accessible to market participants and researchers. Additionally, the thesis adds to the research of the theory of the role of the FB parameters in the capacity calculation, and provides an in depth look to the exact factors affecting these parameters in the energy system. The methods and observations introduced in this thesis have practical implications for decision-making and planning for various participants in the energy markets and in the energy sector in general. It is important to acknowledge that the field of forecasting FBMC components is still in its early stages. Further research is necessary to ensure the effectiveness of current and new models in the CCM of EU and the ability to reliably forecast the changes to markets in the face of the ongoing renewable energy revolution.

References

[1] Delreux, T., & Ohler, F. (2019). Climate policy in European union politics. Oxford Research Encyclopedia of Politics. doi:10.1093/acrefore/9780190228637.013.1097

[2] Eurostat. (2023). Renewable energy statistics. Retrieved April 5, 2023, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable_energy_statistics

[3] Eurostat. (2017). Archive:Energy from renewable sources - Statistics Explained. Retrieved October 7, 2023, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive%3AEnergy_from_renewable_sources

[4] European Commission. (n.d.). Delivering the European Green Deal. Retrieved April 17, 2023, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en

[5] IEA (2022). Renewables 2022. IEA, Paris. Retrieved from https://www.iea.org/reports/renewables-2022, License: CC BY 4.0

[6] Smith, O., Cattell, O., Farcot, E., O'Dea, R. D., & Hopcraft, K. I. (2022). The effect of renewable energy incorporation on Power Grid Stability and resilience. Science Advances, 8(9). https://doi.org/10.1126/sciadv.abj6734

[7] Kondziella, H., & Bruckner, T. (2016). Flexibility requirements of renewable energy based electricity systems – a review of research results and methodologies. Renewable and Sustainable Energy Reviews, 53, 10–22. https://doi.org/10.1016/j.rser.2015.07.199

[8] Mehigan, L., Al Kez, D., Collins, S., Foley, A., Ó'Gallachóir, B., & Deane, P. (2020). Renewables in the European power system and the impact on system rotational inertia. Energy, 203, 117776. https://doi.org/10.1016/j.energy.2020.117776

[9] Brunner, C., Deac, G., Braun, S., & Zöphel, C. (2020). The future need for flexibility and the impact of fluctuating renewable power generation. Renewable Energy, 149, 1314–1324. doi:10.1016/j.renene.2019.10.128

[10] Biggar, D. R., & Hesamzadeh, M. (2014). The Economics of Electricity Markets (1st edition). Chichester West Sussex United Kingdom: Wiley.

[11] Goodarzi, S., Perera, H. N., & Bunn, D. (2019). The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. Energy Policy, 134, 110827. doi:10.1016/j.enpol.2019.06.035

[12] Korosteleva, J. (2022). The implications of Russia's invasion of Ukraine for the EU Energy Market and businesses. British Journal of Management, 33(4), 1678–1682. doi:10.1111/1467-8551.12654

[13] Fonte, G. (2023). Retrieved from https://www.reuters.com/article/italy-economy-idUSKBN2VU133

[14] Finnish Government. (2023). Retrieved from https://valtioneuvosto.fi/en/-/1410877/acts-on-retroactive-reimbursement-for-electricity-costs-and-extended-payment-periods-of-electricity-bills-enter-into-force-next-week

[15] van der Wiel, K., Stoop, L. P., van Zuijlen, B. R. H., Blackport, R., van den Broek, M. A., & Selten, F. M. (2019). Meteorological conditions leading to extreme low variable renewable energy production and Extreme High Energy Shortfall. Renewable and Sustainable Energy Reviews, 111, 261–275. doi:10.1016/j.rser.2019.04.065

[16] Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management. EUROPA. (2015). Retrieved April 17, 2023, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015R1222

[17] Fingrid, Energinet, Svenska kraftnät, & Statnett. (2018). (rep.). Supporting document for the Nordic Capacity Calculation Region's proposal for capacity calculation methodology in accordance with Article 20(2) of Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management. Copenhagen: Nordic RCC.

[18] Van den Bergh, K., Boury, J., & Delarue, E. (2016). The flow-based market coupling in Central Western Europe: Concepts and definitions. The Electricity Journal, 29(1), 24–29. https://doi.org/10.1016/j.tej.2015.12.004

[19] Agency for the Cooperation of Energy Regulators (ACER) (2017). Annual Report on the Results of Monitoring the Internal Electricity and Gas Markets in 2016.

[20] Voswinkel, S., Felten, B., Felling, T., & Weber, C. (2019). Flow-based market coupling – what drives welfare in Europe's electricity market design? SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3424708

[21] Lang, L. M., Dallinger, B., & Lettner, G. (2020). The meaning of flow-based market coupling on redispatch measures in Austria. Energy Policy, 136, 111061. https://doi.org/10.1016/j.enpol.2019.111061

[22] Kristiansen, T. (2020). The flow based market coupling arrangement in Europe: Implications for traders. Energy Strategy Reviews, 27, 100444. https://doi.org/10.1016/j.esr.2019.100444

[23] Schönheit, D., Kenis, M., Lorenz, L., Möst, D., Delarue, E., & Bruninx, K. (2021). Toward a fundamental understanding of flow-based market coupling for cross-border electricity trading. Advances in Applied Energy, 2, 100027. https://doi.org/10.1016/j.adapen.2021.100027

[24] Holzmann, A., Zwieb, L., Knaus, K., Harrucksteiner, A., & Sahin, A. (2020). Increasing Transparency for Flow-Based Market Coupling in European Electricity Trading. Vienna: Austrian Energy Agency GmbH.

[25] van Stiphout, F. (2016). Approximating the Flow-Based Transport Capacity Constraints for the Day-Ahead Power Market. Master's thesis, University of Twente.

[26] Wallin, P. (2016). Estimation of cross-border flow in electricity markets using a Markovian-Tobit approach. Master's thesis, KTH Swedish Royal Institute of technology.

[27] Abdel-Khalek, H., Schäfer, M., Vásquez, R., Unnewehr, J. F., & Weidlich, A. (2019). Forecasting cross-border power transmission capacities in central Western Europe using artificial neural networks. Energy Informatics, 2(S1). https://doi.org/10.1186/s42162-019-0094-y

[28] Boury, J. (2015). Methods for the determination of flow-based capacity parameters: description, evaluation and improvements. 10.13140/RG.2.2.23785.70248.

[29] NEMO Committee (2020) EUPHEMIA Public Description. rep. Nordpool.

[30] Fingrid, Energinet, Svenska kraftnät, & Statnett. (2015). (rep.). Principle Approach for Assessing Nordic Welfare under Flow-based methodology. Copenhagen: Nordic RCC.

[31] Schönheit, D., & Sikora, R. (2018). A Statistical Approach to Generation Shift Keys. 2018 15th International Conference on the European Energy Market (EEM), 1-6.

[32] Schönheit, D., Weinhold, R., & Dierstein, C. (2020). The impact of different strategies for Generation Shift Keys (gsks) on the flow-based market coupling domain: A model-based analysis of Central Western Europe. Applied Energy, 258, 114067. https://doi.org/10.1016/j.apenergy.2019.114067

[33] Young, H. D., Freedman, R. A. & Ford, A.L. (2004) Direct-Current Circuits. University Physics with Modern Physics, International Edition. Vol. 11th edition, International edition. (pp. 980–1018). Pearson.

[34] Nord Pool. REMIT UMM. Available at: https://umm.nordpoolgroup.com/#/messages?publicationDate=all&eventDate=all

[35] Thakurta, P. G. (2016). Flexible Operation of Electric Power Transmission Grids. In Sustainable Energy-Technological Issues, Applications and Case Studies. IntechOpen.

[36] Joint Allocation Office (JAO). (2023). Publication handbook. Available at: https://test-publicationtool.jao.eu/nordic/home

[37] International Atomic Energy Agency (IAEA) (2022) Country Nuclear Power Profiles, Finland 2022. Available at: https://cnpp.iaea.org/countryprofiles/Finland/Finland.htm (Accessed: 27 June 2023).

[38] Svarstad VB. (2016). Flow based market clearing - GSK strategies, Master's thesis, NTNU. URL https://brage.bibsys.no/xmlui/handle/11250/2413290

[39] Jegleim B. (2015). Flow based market coupling, Master's thesis, NTNU. URL https://brage.bibsys.no/xmlui/handle/11250/2368078

[40] ENTSO-E. (2023). Retrieved from https://transparency.entsoe.eu/outage-domain/r2/unavailabilityInTransmissionGrid/show

[41] Fingrid, Statnett, Svenska kraftnät, & Energinet. (2022). LT CC and new NUCS processes. In CCM Stakeholder Meeting. Nordic RSC.

[42] Joint Allocation Office. (n.d.). Puto Nordic CCR Test. Retrieved from https://test-publicationtool.jao.eu/nordic/flowbasedDomain

[43] ENTSO-E. (2023). Retrieved from https://transparency.entsoe.eu/dashboard/show

[44] Nordic Unavailability Collection System (NUCS). (2023). Retrieved from https://www.nucs.net/