UPPSALA
UNIVERSITET

UPTEC STS 25045
Examensarbete 30 hp
November 2025

# Optimizing Short-Term Electricity Price Forecasting

## Evaluating the Impact of Data-Driven Generation Shift Keys on Power Transfer Distribution Factor Estimation

Julia Malmås

Civilingenjörsprogrammet i system i teknik och samhälle

UPPSALA
UNIVERSITET

# Optimizing Short-Term Electricity Price Forecasting

Julia Malmås


## Abstract

The ongoing transformation of the Nordic power system, driven by the increasing share of renewable generation and the implementation of Flow-Based Market Coupling (FBMC) in October 2024, has introduced new challenges for electricity price forecasting. To improve short-term price models, market participants require an accurate representation of network sensitivities, which depend on the estimation of Power Transfer Distribution Factors (PTDFs). A key input in these calculations is Generation Shift Keys (GSKs), which define whether a node responds to a price change and participates in a change of a bidding zone's net position. The choice of a GSK strategy directly influences how PTDFs are represented in market models, thus impacting the capacity allocation and market results. This thesis, conducted at Vattenfall AB, investigates how data-driven GSKs based on actual generation data affect the accuracy of zonal PTDF calculations within the Nordic day-ahead electricity market. Three time-resolution levels – monthly, weekly, and daily – were evaluated and compared against a simplified flat GSK reference. Each GSK was computed using hourly production data and complemented with internal capacity documentation. The PTDFs were calculated using Vattenfall's grid model based on PyPSA which is an open-source framework for simulating and optimizing electrical networks, and performance was assessed using the Root Mean Square Error (RMSE) between modeled and official PTDFs from the Joint Allocation Office (JAO). The results show that all three data-driven approaches outperform the flat reference by improving PTDF accuracy across most bidding zones. Among them, the monthly GSKs achieve the best overall balance between accuracy and numerical stability, while the weekly and daily GSKs capture short-term variations more dynamically but at the cost of increased volatility. These findings suggest that even relatively simple, data-based GSK formulations can enhance the reliability of power flow representation in flow-based models. Improved PTDF estimation can in turn contribute to more accurate flow-based capacity calculations and, ultimately, to more reliable short-term electricity price forecasts.



**Teknisk-naturvetenskapliga fakulteten**

**Uppsala universitet, Utgivningsort Uppsala**

Handledare: Olof Nilsson  Ämnesgranskare: Göran Ericsson

Examinator: Elísabet Andrésdóttir

# Populärvetenskaplig sammanfattning

Energisektorn genomgår en omställning i takt med strävan efter att minska växthusgasutsläppen och nå klimatmål. En central del i denna omställning är det ökade beroendet av förnybara energikällor såsom vind- och solkraft. Även om dessa teknologier är avgörande för att uppnå ett hållbart och koldioxidsnålt energisystem medför de även betydande tekniska och operativa utmaningar. Till skillnad från konventionell fossilbaserad elproduktion som kan styras efter behov är elproduktion från förnybara källor i större grad väderberoende och därmed varierande. Detta skapar nya utmaningar i arbetet med att säkerställa tillförlitligheten och stabiliteten i moderna elsystem.

En av de mest kritiska tekniska förutsättningarna för stabil nätverksdrift är att upprätthålla systemfrekvensen 50 Hz. Denna frekvens speglar den realtidsmässiga balansen mellan elproduktion och elanvändning. På så sätt kan betydande avvikelser leda till skador på utrustning, systeminstabilitet eller till och med omfattande elavbrott. Historiskt sett har det varit relativt enkelt att balansera systemet tack vare den höga styrbarheten hos konventionell elproduktion men i takt med att andelen intermittenta förnybara energikällor ökar, och gränsöverskridande handel med el blir mer omfattande, har det blivit betydligt mer komplext att upprätthålla denna balans. Dessutom finns det ofta ett glapp mellan de fysiska kraftflödena och de ekonomiska marknadsutfallen, vilket ytterligare försvårar driften av elsystemet.

För att hantera dessa utmaningar infördes en ny metod för kapacitetsberäkning – *Flow-Based Market Coupling (FBMC)* – i den nordiska dagen före-marknaden den 29 oktober 2024. Syftet med FBMC är att förbättra effektiviteten på elmarknaden genom att integrera elnätets fysiska begränsningar direkt i algoritmerna för marknadsavräkning. Genom att ge en mer exakt bild av överföringskapacitet och potentiella flaskhalsar förväntas FBMC bidra till bättre samordning av gränsöverskridande kraftflöden. Införandet av denna metod innebär dock också en ökad komplexitet i prognostiseringen av elpriser, vilket är en kritisk aktivitet för marknadsaktörer som exempelvis Vattenfall. Tillförlitliga prisprognoser är nämligen avgörande för budstrategier, produktionsplanering och långsiktiga investeringsbeslut. I praktiken är det dock långt ifrån enkelt att utveckla sådana prognoser. En av de största utmaningarna är bristen på transparens kring vissa nätrelaterade data. I Sverige är till exempel information om överföringsbegränsningar inte offentligt tillgänglig på grund av säkerhetsskäl. Denna brist på data gör det svårare att modellera marknadsbeteenden och förutsäga prisrörelser.

Mot denna bakgrund undersöks i detta examensarbete hur så kallade *Generation Shift Keys (GSKs)* påverkar noggrannheten i beräkningen av *Power Transfer Distribution Factors (PTDFs)*. I FBMC används GSK:er för att översätta elsystemets detaljerade information på nodnivå till en mer övergripande nivå per elområde. Nod PTDF:erna beskriver hur kraftflöden påverkas i varje enskild punkt i nätet, medan zon PTDF:erna visar sambandet mellan hela elområden. Eftersom elhandeln i den nordiska marknaden sker på områdesnivå är det zon PTDF:erna som används i marknadsmodellen för att beräkna överföringskapacitet och elpriser. GSK:er fungerar därmed som länken mellan dessa två nivåer och avgör hur förändringar i ett elområdes export eller import påverkar kraftflödena i det fysiska nätet. Hur GSK:er definieras påverkar även vilka noder som antas reagera på prisförändringar. Prisberoende anläggningar, såsom vattenkraft eller termiska kraftverk, förväntas kunna anpassa sin produktion vid förändringar i elpriset, medan mer statiska produktionskällor, som kärnkraft eller annan väderoberoende elproduktion, ofta antas vara oförändrade. Valet av GSK-strategi har därför stor betydelse för hur realistiskt modellen kan återge samspelet mellan elmarknaden och det fysiska elsystemet.

I detta arbete har tre data-drivna metoder utvecklats baserade på faktisk produktionsdata för att beräkna GSK:er. Det innebär att varje kraftverk tilldelas en GSK-faktor som speglar dess verkliga andel av den totala produktionen i ett elområde, vilket ger en mer realistisk bild av hur el produceras och transporteras i nätet. För att utvärdera effekten av dessa metoder har de jämförts med en så kallad *flat GSK strategy*, där alla kraftverk inom ett elområde antas delta lika mycket i förändringen av nätpositionen i ett elområde. Denna strategi används här enbart som en referenspunkt för att mäta hur mycket de datadrivna metoderna förbättrar beräkningarna. I verkligheten används mer avancerade GSK-strategier som bygger på systemoperatörernas bedömningar av produktionsmönster och tekniska begränsningar.

Resultaten visar att alla de data-drivna metoderna förbättrar noggrannheten jämfört med den förenklade flat-referensen. Särskilt den månadsbaserade modellen uppvisade den bästa balansen mellan precision och stabilitet, medan de veckobaserade och dagsbaserade modellerna gav något mer varierande resultat. Det tyder på att även relativt enkla justeringar i hur GSK:er beräknas kan bidra till en mer realistisk och tillförlitlig representation av kraftflödena i det nordiska elsystemet. På längre sikt kan sådana förbättringar bidra till mer exakta beräkningar av överföringskapacitet, vilket i sin tur skulle kunna bidra till bättre prisprognoser och effektivare marknadsutfall.

# Acknowledgment

I would like to thank my supervisor at Vattenfall, Olof Nilsson, for his guidance and support throughout this master's thesis. I also wish to thank the Vattenfall team for creating an environment of collaboration and learning. To Tamás Börcsök, thank you for the many helpful discussions that guided my work. To Hampus Engsner, thank you for explaining the price-forecasting process at Vattenfall and for sharing your technical insights. Finally, I would like to acknowledge my subject reader at Uppsala University, Göran Ericsson, for his support and thoughtful guidance throughout this journey.

# Abbreviations

| | |
|---|---|
| *BZ* | Bidding Zone |
| *CGM* | Common Grid Model |
| *CNE* | Critical Network Element |
| *CNEC* | Critical Network Element with a Contingency |
| *CCR* | Capacity Calculation Region |
| *ENTSO-E* | European Network of Transmission System Operators for Electricity |
| *FBMC* | Flow-Based Market Coupling |
| *GSK* | Generation Shift Key |
| *IGM* | Individual Grid Model |
| *JAO* | Joint Allocation Office |
| *MCO* | Market Coupling Operator |
| *MCP* | Model Calculation Platform |
| *NEMO* | Nominated Electricity Market Operator |
| *NP* | Net Position |
| *NTC* | Net Transfer Capacity |
| *OSM* | OpenStreetMap |
| *PTDF* | Power Transfer Distribution Factor |
| *RAM* | Remaining Available Margin |
| *RCC* | Regional Coordination Centre |
| *RMSE* | Root Mean Square Error |
| *TSO* | Transmission System Operator |
| *UMM* | Urgent Market Message |

# Contents

# 1 Introduction

The ongoing transformation of the European energy system towards carbon neutrality is fundamentally reshaping the structure and operation of electricity markets [1]. As the share of renewable generation, such as wind and solar power continues to increase, the variability and uncertainty associated with these sources introduce new challenges for maintaining system balance and ensuring efficient market functioning. Unlike conventional fossil-based generation, renewable production is highly weather-dependent and less controllable, leading to fluctuating power flows. This increasing variability makes it more difficult to maintain the reliability and stability of modern power systems [2].

One of the most critical technical requirements for stable grid operation is maintaining the system frequency at 50 Hz. This frequency reflects the real-time balance between electricity production and consumption. Any substantial deviations can cause equipment damage, system instability, or even large-scale blackouts [3]. Historically, balancing the system was relatively straightforward thanks to the controllability of conventional power plants [1]. However, as the share of intermittent renewables increases and cross-border electricity trading becomes more extensive, maintaining this balance has become significantly more complex [3]. Furthermore, there is often a discrepancy between the physical power flows in the grid and the market outcomes determined by economic optimization, which further complicates system operation [4].

To address these challenges, the *Flow-Based Market Coupling (FBMC)* method has been introduced in several European regions, including the Nordic day-ahead market in October 2024 [5]. FBMC aims to improve market efficiency by incorporating physical transmission constraints directly into the market-clearing algorithm, providing a more realistic representation of available cross-border capacities. However, this increased level of detail also adds complexity to electricity price forecasting, as the underlying network representation must accurately describe how power flows respond to changes in generation and demand [6]. In practice, developing such forecasts remains a major challenge, partly due to limited transparency regarding network-related data. In Sweden, for instance, detailed information about critical network elements is not publicly available for security reasons. This lack of data makes it more difficult to model market behavior and predict price movements accurately [7].

This master's thesis was carried out at Vattenfall AB, within the Electricity Market department, where ongoing model development focuses on improving short-term price forecasting under the new flow-based framework. The work contributes to this development by analyzing how alternative formulations of *Generation Shift Keys (GSKs)* may enhance the estimation of *Power Transfer Distribution Factors (PTDFs)*, which in turn affect how transmission constraints are represented in market models. By improving the accuracy of PTDF calculations, the study supports Vattenfall's broader goal of producing more reliable price forecasts and market simulations in the evolving Nordic power system.

A key element in this process is the definition of GSKs, which determine how changes in a bidding zone's net position are distributed among its internal nodes. GSKs are used to transform the highly detailed, nodal PTDFs into zonal PTDFs, which are required since electricity trading in the electricity market takes place on a bidding-zone level rather than at individual nodes. The choice of a GSK strategy directly affects how accurately zonal PTDFs represent real power flow sensitivities in the grid [8].

## 1.1 Aim

The aim of this master's thesis is to evaluate how different methods for calculating GSKs influence the accuracy of zonal PTDF estimations within the Nordic electricity market. The study focuses on developing and testing data-driven GSKs based on actual generation data and compares their performance with a flat reference strategy. A particular emphasis is placed on how the time resolution used in GSK calculation—monthly, weekly, and daily—affects both the accuracy and stability of the resulting zonal PTDFs. By identifying which configuration provides the most realistic and reliable representation of zonal sensitivities, the study aims to support more accurate flow-based capacity calculations and, ultimately, contribute to improved electricity price forecasting.

## 1.2 Research Questions

Based on this aim, the following research questions are addressed in this master's thesis:

1. How do data-driven GSKs derived from actual generation data affect the accuracy of zonal PTDF estimations compared to a flat reference case?

2. How does the choice of time resolution of data-driven GSKs influence the trade-off between accuracy and stability in the resulting zonal PTDFs?

The underlying hypothesis is that GSKs derived from actual generation data provide a more accurate estimation of zonal PTDFs than those based on uniform assumptions.

## 1.3 Delimitations

Power plants with an installed capacity below 20 MW are excluded from the analysis. The study is restricted to a single day, 1 April 2024, as Vattenfall's grid model was developed for system conditions after this date. Moreover, the NO1 bidding zone is omitted due to its GSK strategy being based on load rather than generation. DK1 is excluded because it belongs to a different synchronous area, while DK2 is excluded due to difficulties in decoding power plant names. Virtual bidding zones are also excluded from the analysis. A more thorough explanation of these delimitations is provided in Section 5.2.4.

# 2 Background

This section provides the background needed to understand how the electricity market functions and how grid constraints affect price formation. It first describes the structure of the Nordic electricity markets, including trading mechanisms, and bidding zones. The section then explains the recent transition from the Net Transfer Capacity (NTC) method to the FBMC approach, which enables a more efficient use of the transmission grid. Finally, it outlines Vattenfall's price forecasting process and its connection to the new flow-based methodology.

## 2.1 The Electricity Market

The electricity market in Sweden was deregulated in 1996 and has since then been a part of the European electricity market [9]. Consequently, the power grids are interconnected, allowing electricity to flow across national borders. The purpose of the free market is to utilize the combined resources as efficiently as possible to meet consumer demand. However, unlike other free markets, the electricity market requires a constant balance between consumption and production to maintain the functionality of the physical system. As a result, the electricity market is designed to align with the technical requirements of the power grid, which shapes roles, responsibilities, and the types of products traded [10].

To maintain the critical balance between consumption and production in the electricity market, trading is organized across different time horizons and sub-markets, each serving a specific purpose to enhance system reliability and efficiency. Long-term planning takes place in the *future* or *forward markets*, where electricity is traded up to many years ahead in order to ensure financial stability and minimize risk in a longer time perspective [11]. Closer to delivery, the *day-ahead market*, also called the *spot market*, facilitates hourly electricity for the following day. When referring to the electricity market, the day-ahead market is often meant since the largest volume of electricity is traded there. Prices are determined based on supply and demand forecasts, and this market serves as the foundation for electricity trading across Europe [12]. As delivery approaches, the *intraday market* allows participants to make short-term adjustments, typically up to one hour before consumption. This market provides flexibility to address unexpected changes, such as weather forecasts, unexpected outages, or other factors that become more certain closer to physical delivery. While smaller in volume compared to the day-ahead market and involving fewer participants, the intraday market plays a crucial role in maintaining flexibility and reliability. Therefore, it is primarily used by actors to trade themselves into balance to ensure that supply and demand remain aligned [13]. Finally, the *balancing market* operates in real time and is managed by Transmission System Operators (TSOs) to ensure balance between electricity consumption and production. In Sweden, Svenska kraftnät handles this responsibility by producing ancillary services, which include actions like adjusting generation or consumption to stabilize the grid [14]. This complex market structure ensures that electricity is delivered reliably and cost-efficiently while addressing both predictable patterns and unexpected fluctuations in the system.

The Nordic electricity market is segmented into bidding zones, where each zone has a uniform price, and the local TSO determines the division of these areas within each country. This segmentation reflects physical constraints in the transmission network [15]. To address physical constraints, such as bottlenecks in the Swedish transmission grid, Sweden was divided into four bidding zones (SE1, SE2, SE3, and SE4) in 2011. This division highlights areas with transmission limitations, and within Sweden, bottlenecks are particularly evident between the northern and southern regions, where an excess of low-cost electricity production in the north contrasts with high consumption in the south. Hence, the division into bidding zones is critical for managing transmission limitations, as it provides a structured approach to addressing grid overload risks where the system cannot physically transfer the volume of electricity demanded by the market [16].

According to [17], a Capacity Calculation Region (CCR) is defined as a geographical area where TSOs coordinate the capacity calculation across bidding zone borders. Within the Nordic CCR, there are 12 real bidding zones. Additionally, HVDC interconnectors are represented as virtual bidding zones under the flow-based methodology. As a result, 19 virtual bidding zones are included, bringing the total number of bidding zones in the Nordic CCR to 31. Figure 1 presents the flow-based network topology for the Nordic CCR, illustrating both real and virtual bidding zones [18].

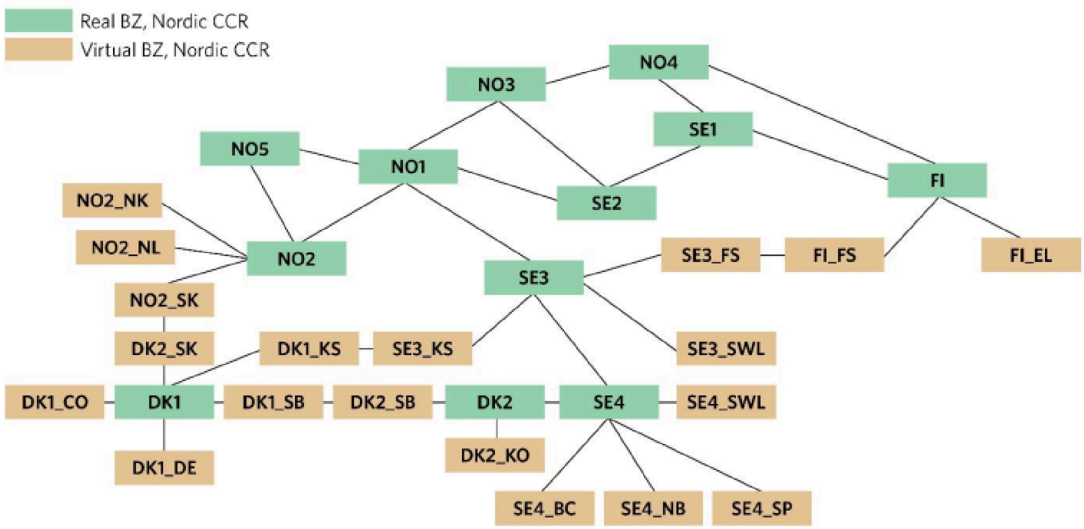


*Figure 1: Topology for the Nordic CCR including the total of 31 bidding zones, containing both real and virtual bidding zones* [18].

### 2.1.1 The Day-Ahead Market

The majority of electricity trading occurs in the day-ahead market, where producers, traders, and large energy consumers participate. This market operates within a unified European framework, and in Sweden, market participants can submit buy and sell bids for Swedish bidding zones on two electricity markets: Nord Pool and EPEX Spot. These markets serve as the Nominated Electricity Market Operators (NEMOs) in Sweden [6]. Additionally, market participants have the option to trade directly with one another through bilateral agreements [10]. All electricity markets in the EU are interconnected, enabling a common European auction where the electricity price is determined for the following 24 hours. This process, known as market coupling, ensures that buyers and sellers face the same price, regardless of which electricity market they place their bids on. Moreover, market coupling extends beyond the EU, allowing participation from electricity markets in non-EU countries[1] [12].

Before electricity trading begins, the TSOs determine the available transmission capacity between bidding zones across Europe [20]. Each TSO develops an Individual Grid Model (IGM), incorporating relevant grid data along with input from market participants. Each IGM is constructed from ”snapshots” of real-time power system operations, typically taken two days before the delivery day. To improve the accuracy of flow-based parameters, these snapshots must be adjusted to reflect the most probable conditions for the delivery day. The resulting ”updated snapshots” are merged into what is known as the base case, which can be modified based on factors such as planned outages, load forecasts, and production forecasts [21]. In addition, the base case includes net positions for each bidding zone and time interval, providing an overview of the expected injections within each zone [4]. Large-scale power systems generally operate according to the N–1 criterion, which ensures that the grid can continue functioning within security limits even if a single component fails. To assess N–1 stability, the TSO utilizes the base case alongside a list of potential contingencies and monitors specific grid elements for threshold exceedances. These monitored elements, referred to as Critical Network Elements with a Contingency (CNECs), help identify critical points in the network under contingency scenarios [21].

The IGMs are then combined into a Common Grid Model (CGM), which provides a comprehensive overview of key sections of the European power grid, including forecasts for electricity generation and consumption. For the day-ahead market, this process generates 24 hourly capacity scenarios, defining how much electricity can be transmitted across the system. The aim is to create a balanced CGM, which is obtained by retrieving the IGMs at the same point in time. Otherwise, it can cause differences in net positions and flows, resulting in an unbalanced CGM. In a balanced CGM, it is a requirement that the difference between generation and consumption, the so-called net position, within a bidding zone is exactly zero [21].

[1] Excluding Switzerland, Bosnia and Herzegovina, Serbia, Montenegro, North Macedonia, and Albania [19].

Before this capacity is allocated to the market, each TSO reviews and approves the results to ensure they align with operational security requirements. Since the TSOs oversee the stability of the power system, they are best positioned to verify the accuracy and reliability of the calculations. Transmission capacity is affected by factors such as weather conditions and ongoing maintenance, as each power line and grid component has a physical limit on how much electricity it can carry [21]. To uphold system reliability, the TSOs provide daily updates on these constraints to the market [20]. At Nord Pool and the Joint Allocation Office (JAO), transmission capacities are published at 10:00 CET, giving market participants until 12:00 CET to place their final bids. Once the bidding window closes, an auction mechanism determines electricity prices and the distribution of power across the grid [22].

After the bidding deadline, orders are submitted to the NEMOs and forwarded to the Market Coupling Operator (MCO). Here, they are matched through *EUPHEMIA*, an algorithm that is based on explicit economic welfare optimization and calculates the electricity price for each hour between 00:00 and 24:00 in every bidding zone [23]. The resulting electricity price represents an equilibrium point where supply meets demand, commonly referred to as the spot price. This price is determined based on the marginal pricing principle, meaning that the final price is set by the cost of producing the last kilowatt-hour required to meet demand. As supply and demand fluctuate throughout the day, auction prices vary by the hour. Following the auction, both buyers and sellers receive confirmation of the electricity volumes they will trade during each of the upcoming 24 hours. Importantly, the market treats all electricity equally regardless of its production source; i.e, all electricity is traded in the same manner [12].

## 2.2 A Shift in Capacity Calculation Methods

The electricity transmission grid imposes physical constraints on how much electricity can be transferred between different parts of the system. While investments in grid expansion can alleviate these limitations, unlimited expansion would be both costly and impractical. Therefore, optimizing grid utilization through efficient capacity calculation methodologies is crucial [6]. Another factor driving the need for more precise capacity calculation is the increasing integration of renewable energy sources. Unlike conventional power plants, renewable energy production is often concentrated in specific geographic areas and affected by weather patterns, resulting in high variability in production volumes. This creates a greater need for efficient electricity transmission across regions to balance supply and demand [2].

To address these challenges, a new capacity calculation methodology was introduced in the Nordic day-ahead electricity market on October 29, 2024. This new method, known as FBMC, represents a major shift in how capacity is determined [5]. While FBMC was new to the Nordic region, it had already been operational in Central Western Europe (CWE) since May 2015 [24]. The transition was necessary because the previous method, NTC, was no longer accurate enough to handle the growing complexity of the power system. The flow-based method allows for more efficient use of the grid, enabling greater electricity transfers from producers to consumers without exceeding transmission limits [5]. To understand why this shift was necessary and how the methods differ, a more detailed explanation follows.

### 2.2.1 The Previous Net Transfer Capacity Method

For many years, the Nordic electricity markets relied on the NTC method to determine the maximum allowable power transfers between bidding zones. In this approach, the TSOs established fixed capacity limits for each bidding zone border and international connection, setting separate values for import and export directions. When cross-border connections were involved, the TSOs of the respective countries coordinated their calculations and chose the lower capacity value in order to maintain operational safety [5].

A fundamental drawback of the NTC method is that the TSOs determine capacity limits without knowing how the market will behave. Since demand patterns remain uncertain at the time of calculation, it is difficult to anticipate which bidding zone borders may experience congestion. To prevent grid overloading, safety margins must be applied, ultimately reducing the transmission capacity available to the market [5]. Transmission networks are also highly interconnected, making it impractical to determine reliable NTC values between two areas without considering the influence of other connected regions. In large systems, this complexity increases, and the NTC method struggles to capture the true dynamics of the transmission system, limiting its effectiveness in capacity calculation [25].

### 2.2.2 The Shift to Flow-Based Market Coupling

The flow-based method was introduced to better reflect the physical constraints of the power system. Instead of applying fixed capacity limits between bidding zones, FBMC identifies the actual grid elements where congestion occurs and uses a more detailed representation of electricity flows. This enables a more efficient allocation of available capacity [6].

A key factor for adopting the flow-based method in the Nordic electricity market was to better manage east–west power flows through Sweden. Under the NTC approach, simultaneous exchanges between NO1, SE3, and DK1 were handled with simplified capacity allocations, which often caused internal congestion in SE3 and limited grid utilization. With the flow-based method, these interactions are represented more accurately, allowing available capacity to be allocated optimally [26].

According to [6], some of the main benefits of FBMC include:

- A better representation of physical grid constraints, leading to more efficient use of available capacity.

- More accurate price signals, reducing unnecessary market inefficiencies.
- Increased socioeconomic welfare through optimized electricity distribution.

The need for flow-based market coupling has become even more critical due to recent developments in the power system:

- The expansion of HVDC interconnections across Europe.
- The growing share of renewable energy, which increases volatility and complexity in the grid.
- The increasing interdependency between different electricity markets [6].

These factors made it difficult for NTC to efficiently allocate transmission capacity across multiple bidding zones, leading to the shift to FBMC.

## 2.3 An Introduction to Flow-Based Method

In the flow-based approach, the market coupling algorithm *EUPHEMIA* processes constraints in the form of *zone-to-slack Power Transfer Distribution Factors (PTDFs)* and *Remaining Available Margins (RAMs)* [23]. This differs from the previous NTC method, where capacity was defined at bidding zone borders. Furthermore, a bidding zone's *Net Positions (NPs)* are constrained by their impact on *Critical Network Elements (CNEs)* under various contingency scenarios [27]. RAM represents the transmission capacity available to the market, while zone-to-slack PTDFs describe the sensitivity of flows on these CNEs [6].

It is neither necessary nor practical to include all transmission lines in the FBMC model. Instead, the TSOs monitor a selected set of grid elements that are most relevant for maintaining system security and avoiding overloads. These monitored components, referred to as CNEs, may represent individual transmission lines, transformers, or aggregated cross-sections of several lines. Determining which elements to include is one of the main challenges of the approach. A conservative selection, where too many elements are monitored, greatly increases the number of parameters, making the market coupling problem computationally demanding. Conversely, if too few elements are included, the feasible domain may become overly large, increasing the operational burden on TSOs to maintain system reliability [6].

Equation 1 illustrates the so-called flow-based constraints, where *RAM* and *NP* are vectors, and $z2sPTDF$ is a matrix. As previously mentioned, the main focus of this master's thesis is on PTDFs. Consequently, RAM and NP are neither computed nor discussed in detail.

$$\mathbf{z2sPTDF} \cdot NP \leq RAM \tag{1}$$

To illustrate the differences between the flow-based and NTC approaches, a simplified three-node transmission grid is presented in Figure 2 [6].

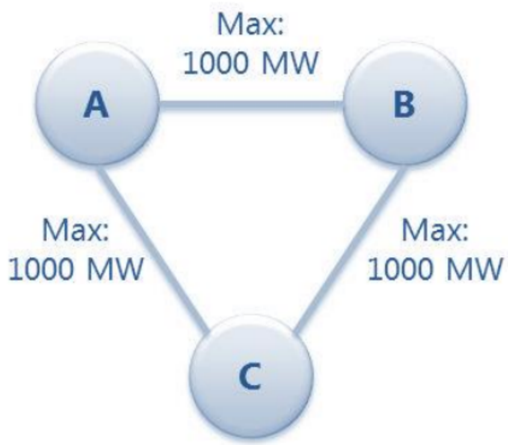


*Figure 2: An example of a transmission grid with three nodes* [6].

In Figure 2, no internal constraints are present, meaning that the only limiting factors in the grid are the transmission lines between the nodes. Each transmission line has a thermal capacity of 1000 MW, referred to as the RAM. In this example, node A and B function as generation nodes, while node C serves as a consumption node and acts as a slack node. This implies that all power injected into node C is theoretically absorbed there. Due to the grid topology, generating 1 MW in node A causes a power flow of 2/3 MW on line A-C, 1/3 MW on line A-B, and 1/3 MW on line B-C. Similarly, generation in node B results in a flow of -1/3 MW on line A-B, 1/3 MW on line A-C, and 2/3 MW on line B-C. These values represent the PTDFs, which describe how electricity flows through the grid [6]. The combined influence of nodes A, B, and C on power flows defines the PTDF matrix, as shown in Table 1 [6].

*Table 1: PTDF matrix of the grid in Figure 2* [6].

| 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 |

Consider an NTC scenario where generation in node A is 2000 MW. This results in a power flow of 1333 MW on line A-C, exceeding the line's thermal capacity and causing an overload. To prevent such violations, TSOs must limit the maximum transfer capacity. In this case, the line capacities need to be restricted to 1500 MW, which means the power flow on each line is adjusted to $2/3 \cdot 1500 = 1000$ MW. Figure 3 illustrates the solution domain, which represents the range of secure net positions for this three-node grid example. The figure also shows that if either node A or B generates 1500 MW, the other cannot produce electricity without exceeding the secure operating limits. Alternatively, both nodes A and B can generate 750 MW each, remaining within the feasible solution domain [6].

As shown in Figure 3, all NTC solutions fall within the flow-based solution domain, meaning that the flow-based method provides more solutions than the NTC approach.

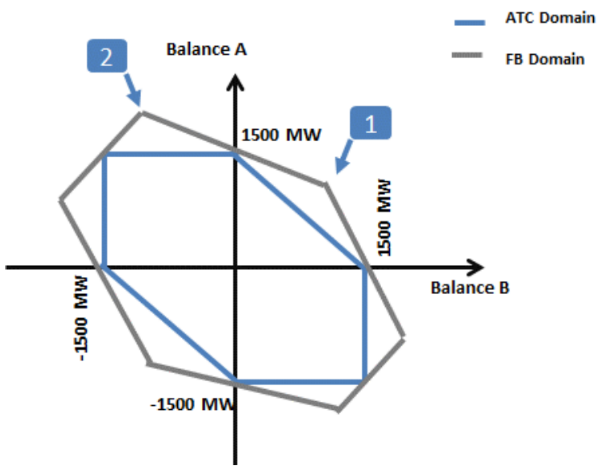


*Figure 3: Solution domains for NTC and flow-based approaches* [6].

By applying the flow-based method to the three-node grid, both node A and node B can generate 1000 MW without exceeding the capacity limits of the lines. For instance, the power flow on line A-C is determined by $1000 \cdot 2/3 + 1000 \cdot 1/3 = 1000$ MW. This solution, represented by point 1 in Figure 3, demonstrates that the flow-based method outperforms the NTC approach. The NTC method fails to enable this solution due to its lack of detailed representation of the grid's physical constraints. Consequently, the flow-based method allows for more trading opportunities while maintaining the same level of operational security [6].

However, it is important to acknowledge certain simplifications within the flow-based approach. In a pure flow-based setup, known as nodal pricing, each node would act as an independent bidding zone with its own price. However, in the European implementation of flow-based, nodes are instead aggregated into bidding zones to maintain a manageable market structure. For instance, the Nordic CCR consists of 12 real and 19 virtual bidding zones in total (see Figure 1). This zonal approach introduces a key challenge: unlike nodal pricing, it does not automatically balance supply and demand at each node. Instead, the system must estimate how changes in a zone's overall power balance affect individual nodes. Since prices cannot balance the system locally as they do under nodal pricing, additional measures are needed to keep the system stable [6].

This is addressed using GSKs. GSKs define how a change in a bidding zone's net position is allocated across individual nodes within the CGM. They play a crucial role in converting nodal PTDFs into zonal PTDFs. The way GSKs are formulated determines which nodes in a bidding zone respond to price fluctuations. Only price-responsive nodes adjust their net position in response to price changes, while price-independent nodes remain unchanged [6].

The methodology used to assign GSK factors is referred to as a GSK strategy. The choice of strategy influences PTDF calculations, which in turn affect capacity allocation and market outcomes. In the Nordic CCR, one custom and eight predefined GSK strategies have been developed by the Nordic Regional Coordination Centre (RCC), which is an independent organization owned by the TSOs of Sweden, Denmark, Finland, and Norway [8], [28]. Since the GSK approach is a linear approximation of a non-linear system, selecting an appropriate strategy is crucial. TSOs can either choose one of the eight predefined strategies for each bidding zone or define a custom GSK strategy with specific GSK factors assigned to individual generation and load units in the CGM. Different bidding zones may require different GSK strategies depending on generation technology mix, geographic distribution, and hourly variations in production and demand. Typically, flexible generation technologies, such as hydropower, coal, oil, and gas-fired plants, are assigned higher GSK factors, meaning they are expected to respond to price changes. Meanwhile, non-flexible generation, such as nuclear, wind, and solar power, is usually excluded from the GSK calculation and assigned a GSK factor of zero, meaning these units do not adjust their output in response to market changes [8]. The different GSK strategies will be described in depth in Section 4.2.2.

## 2.4 Vattenfall's Planning and Price Forecasting Process

Price forecasting is an essential part of Vattenfall's daily operations and long-term planning. Reliable forecasts help the company plan production, trading, and maintenance while reducing financial risks and supporting profitability. Forecasts are also used for investment and grid planning. The purpose of forecasting depends on the time horizon. In the short term (hours to two weeks), forecasts guide bidding strategies for the day-ahead and intraday markets. Medium-term forecasts (weeks to months) support hydropower scheduling, capacity planning, and maintenance. Long-term forecasts (years ahead) are used for strategic planning, such as new generation and grid investments [29].

Vattenfall's forecasting models use a wide range of data sources to describe both market and system conditions. Market data, such as historical prices and traded volumes, help identify price patterns and market behavior. System data, including weather forecasts, demand, and production, show changes in supply and demand. Grid data is especially important in the FBMC framework, where transmission constraints strongly affect price formation and power flows. In addition, external information such as regulatory updates and Urgent Market Messages (UMMs) are included to account for operational events that may impact grid or generation capacity. For example, UMMs include unexpected generator outages, planned or unplanned grid maintenance, or major capacity reductions. However, UMMs are still published according to the NTC domain, meaning they refer to single borders rather than the flow-based domain. This makes it difficult to use them directly in FBMC-based forecasting models [29], [30].

The forecasting process combines automated and manual work. Data from Nord Pool, JAO, and other sources is imported into Vattenfall's internal models, which generate forecasts for prices, generation, and grid constraints. Most of this process is automated, but some information, especially incomplete or qualitative UMMs requires manual checking. Forecasts are updated continuously when new data becomes available, such as updated weather forecasts or grid changes. This ensures that the forecasts stay aligned with current market and system conditions [29].

The shift from NTC to FBMC has made price forecasting more complex. Under NTC, Vattenfall forecasted transmission capacity for each border. In FBMC, the company must instead forecast the entire flow-based domain. According to [30], this is challenging because the data needed to accurately reproduce the flow-based domain is not publicly available. As a result, forecasting errors can occur when grid constraints change or when reduced capacities are not fully reflected in published data. The limited transparency increases uncertainty in short-term forecasts and makes it more difficult to include network constraints in price models [30].

Despite these challenges, Vattenfall continues to improve its forecasting tools. The company highlights the need for more transparency and cooperation between TSOs, the Nordic RCC, and market participants to better understand how network constraints influence prices. Better data sharing and clearer methods would not only improve forecast accuracy but also strengthen overall market efficiency and competition [29], [30].

# 3 Previous Research

Recent research in electricity market forecasting and network modeling have increasingly emphasized the importance of open and transparent data for analyzing price formation in the Nordic and European electricity systems. With the introduction of FBMC, the need for models that accurately capture the influence of network constraints on market outcomes has become even more apparent. However, this remains particularly challenging in the Nordic context, where transparency and data availability lag behind practices in many other European regions.

A foundational step toward more open modeling was the development of the Nordic490 model (N490) [15], an extension of the open-source PyPSA-Eur framework tailored to the Nordic synchronous area. By combining geographic information, demographic-based load allocation, and updated transmission parameters, the N490 model showed that open nodal representations can replicate Nordic power flows with reasonable accuracy. While the model was limited by incomplete generator data and approximate line parameters, it demonstrated that meaningful system analysis is possible even when full data transparency is lacking.

Building on this work, the study in [7] addresses one of the major limitations of the Nordic FBMC implementation: the lack of publicly available static grid models and the anonymization of CNECs, especially in Sweden. The authors present a method for reconstructing network topology using only open data, including reverse-engineering the structure of the grid, deanonymizing CNECs, and dynamically calculating GSKs. The model, based on a refined version of N490, includes over 500 buses, 600 lines, and 1000 power units, with parameter values estimated using geographical data. This approach enables accurate prediction of grid constraints and improves market transparency, two key factors in achieving better short-term price forecasts and more efficient trading strategies.

At the European level, similar efforts have focused on building transparent and scalable infrastructure models. In [31], the authors present a high-resolution model of the European transmission system (covering 220–750 kV AC and all DC interconnectors) based on OpenStreetMap (OSM) data. Their methodology combines OSM with geographic filtering, engineering-based parameter estimation, and validation against European Network of Transmission System Operators for Electricity statistics (ENTSO-E). The resulting dataset, released through PyPSA-Eur and PyPSA-Earth, offers improved topological accuracy and geographic resolution compared to existing representations. These qualities make the model especially valuable for reproducible energy system studies, scenario analyses, and optimization applications across the interconnected European grid.

Despite these advances, the Nordic electricity market continues to present unique modeling challenges. Unlike CWE, where FBMC was implemented earlier with greater data transparency, the Nordic CCR does not publish static grid models. Moreover, nodal PTDF matrices and GSKs are not available to the public, only zonal PTDFs for each CNEC are published. This restricts the ability of market participants to understand and forecast market behavior, since the market clearing algorithm relies on internal representations that cannot be externally reproduced. While a parallel run of FBMC alongside the previous NTC-based method was conducted prior to implementation, transparency concerns raised during that period persist today [7].

To overcome these issues, the method proposed in [7] uses the reconstructed grid model to simulate outage scenarios and dynamically adjust GSKs, enabling the prediction of how constraints evolve under changing network conditions. Validation against publicly available data from the JAO confirms the model's ability to replicate constraint behavior with high accuracy. This demonstrates not only the feasibility but also the practical relevance of building detailed power system models using entirely open-source data. A similar structural approach is adopted in [31], although it is applied at the European scale, where the focus lies more on infrastructure representation than market-specific forecasting. Nevertheless, both studies highlight the value of open, high-quality datasets and transparent modeling workflows that support ongoing development by the broader energy system research community.

All together, these efforts reflect a growing trend toward open, reproducible, and technically robust modeling frameworks that support both academic research and market-oriented applications. In the Nordic context, they provide a foundation for understanding the implications of FBMC and for improving the quality of short-term forecasting models. At the European level, they enable consistent cross-border analysis and long-term energy system planning based on publicly available datasets.

# 4 Theoretical Background

This section provides the theoretical background necessary to understand how PTDFs and GSKs are applied in the flow-based method. It first explains how PTDFs are derived from the AC power flow equations and simplified through the DC load flow model to represent network behavior. The section then introduces the concept of GSKs and how they are used to convert node-to-slack PTDFs into zone-to-slack PTDFs. It also describes the different GSK strategies used by the TSOs. Finally, it discusses how the choice of a GSK strategy influences the accuracy and robustness of capacity calculation.

## 4.1 Power Transfer Distribution Factors

This subsection provides an introduction to the underlying AC power flow equations and outlines the derivation of the DC load flow model that serves as a simplified representation of the grid in the flow-based method.

### 4.1.1 Power Flow Equations

In steady-state conditions, the active and reactive power balances at each node in an AC network are defined as

$$P_i = V_i \sum_{k=1}^{n} V_k (G_{ik} \cos(\delta_i - \delta_k) + B_{ik} \sin(\delta_i - \delta_k)) \tag{2}$$

$$Q_i = V_i \sum_{k=1}^{n} V_k (G_{ik} \sin(\delta_i - \delta_k) - B_{ik} \cos(\delta_i - \delta_k)) \tag{3}$$

where:

| | | |
|---|---|---|
| $P_i$ | = | Active power balance in node $i$ [pu MW] |
| $Q_i$ | = | Reactive power balance in node $i$ [pu Mvar] |
| $i,k$ | = | Node numbers |
| $n$ | = | Number of nodes |
| $V_i$ | = | Voltage magnitude in node $i$ |
| $\delta_i$ | = | Voltage angle of node $i$ |
| $\delta_k$ | = | Voltage angle of node $k$ |
| $G_{ik}$ | = | Conductance between node $i$ and $k$ with negative sign |
| $G_{ii}$ | = | Sum of all conductances connected to node $i$ |
| $B_{ik}$ | = | Susceptance between node $i$ and $k$ with negative sign |
| $B_{ii}$ | = | Sum of all susceptances connected to node $i$ |

Equation 2 and 3 describe the balance of each node, summing power flows on transmission lines and shunt elements. The objective is to determine voltage magnitudes and angles at all buses, which in turn allows for calculating power flows, losses and currents. However, the PTDF calculation relies on a DC approximation to simplify these equations. This involves the following assumptions:

1. Node voltages are constant at 1 pu.
2. Transmission line resistance is neglected.
3. Voltage angle differences are small.

Applying these simplifications, the power flow equations reduce to:

$$P_i = \sum_{k=1}^{n} B_{ik}(\delta_i - \delta_k) \tag{4}$$

$$Q_i = \sum_{k=1}^{n} -B_{ik} \tag{5}$$

Since Equation 5 shows that $Q_i$ becomes a constant, it does not influence the network's power flow. Therefore, only the active power equation (Equation 4) is used to calculate voltage angles for given generation and consumption values [21].

For an example with three nodes, the voltage angles can be calculated by Equation 6 by adding "+1" to one of the diagonal elements. For example, if node 1 is used as a slack node, the voltage angles can be computed as

$$[\delta] = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} = \begin{bmatrix} 1 + B_{12} + B_{13} & -B_{12} & -B_{13} \\ -B_{21} & B_{21} + B_{23} & -B_{23} \\ -B_{31} & -B_{32} & B_{31} + B_{32} \end{bmatrix}^{-1} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = [Zbus][P] \tag{6}$$

where the inverted Y-bus matrix is referred to as the *Zbus* [21].

### 4.1.2 Computing the Power Transfer Distribution Factors

The PTDFs determine a certain node's flow participation on a specific line. The PTDF matrix is thereby defined by each node's PTDF for all lines in the system [21].

In a generic form, the node-to-slack PTDF can be expressed as:

$$n2sPTDF_j^{\alpha} = B_j(Zbus_{i,\alpha} - Zbus_{k,\alpha}) \tag{7}$$

In Equation 7, the parameter $n2sPTDF_j^{\alpha}$ represents the sensitivity of the power flow on transmission element $j$, which connects nodes $i$ and $k$, to a power injection at node $\alpha$ with a corresponding consumption at the slack node. The term $B_j$ denotes the susceptance of line $j$ between nodes $i$ and $k$, while $Zbus_{i,\alpha}$ and $Zbus_{k,\alpha}$ are elements of the bus impedance matrix corresponding to nodes $i$, $k$, and $\alpha$ [32]. The PTDF matrix represents how the net position of the nodes influences the power transfers on the lines [21].

## 4.2 Generation Shift Keys

This section introduces the concept of GSKs and it begins by explaining how node-to-slack PTDFs are converted into zone-to-slack PTDFs using GSKs. The following part describes the different predefined GSK strategies used by the Nordic TSOs and how they differ in terms of assumptions and practical application. The section concludes with a discussion on how the choice of strategy involves balancing accuracy and robustness in the capacity calculation process.

### 4.2.1 Converting Node-to-Slack PTDFs into Zone-to-Slack PTDFs

As mentioned earlier, GSKs determine how a change in net position within a bidding zone is distributed across the nodes in that zone. Equation 8 illustrates how the flow-based method converts node-to-slack PTDFs into zone-to-slack PTDFs, ensuring that the sum of GSK factors within each bidding zone equals 1 [32].

$$z2sPTDF_j^A = \sum_{\alpha} GSK^{\alpha} \cdot n2sPTDF_j^{\alpha}, \qquad \text{and} \qquad \sum_{\alpha} GSK^{\alpha} = 1 \tag{8}$$

$z2sPTDF_j^A$ = Sensitivity of CNE $j$ to injection of 1 MW in bidding zone $A$

$n2sPTDF_j^{\alpha}$ = Sensitivity of CNE $j$ to injection of 1 MW in node $\alpha$

$GSK^{\alpha}$ = Weight of node $\alpha$ on the PTDFs of bidding zone $A$

To illustrate the concept of GSKs, consider a simplified power system with three bidding zones (A, B, and C) each consisting of four nodes. The bidding zones are interconnected through four CNEs. When assessing how a power injection in zone A affects CNE 1, the four nodes in zone A must be aggregated. In this simplified example, GSK factors are applied to assign a weighting factor to each of the four nodes in zone A [21]. The aggregated zone-to-slack PTDF between zone A and a given CNE can then be expressed as Equation 9:

$$z2sPTDF_A^{CNE1} = GSK^{A1} \cdot n2sPTDF_{A1}^{CNE1} + ... + GSK^{A4} \cdot n2sPTDF_{A4}^{CNE1} \tag{9}$$

Repeating this calculation for all CNEs and bidding zones results in an aggregated zone-to-slack PTDF matrix, as shown in Equation 10. This matrix describes how a power injection in each bidding zone affects the loading of network elements when transferring power to the slack node and serves as the PTDF matrix used in the market coupling process [21].

$$z2sPTDF = \begin{array}{cccc} & ZoneA & ZoneB & ZoneC \\ CNE1 & z2sPTDF_A^{CNE1} & z2sPTDF_B^{CNE1} & z2sPTDF_C^{CNE1} \\ CNE2 & z2sPTDF_A^{CNE2} & z2sPTDF_B^{CNE2} & z2sPTDF_C^{CNE2} \\ CNE3 & z2sPTDF_A^{CNE3} & z2sPTDF_B^{CNE3} & z2sPTDF_C^{CNE3} \\ CNE4 & z2sPTDF_A^{CNE4} & z2sPTDF_B^{CNE4} & z2sPTDF_C^{CNE4} \end{array} \quad (10)$$

In the Nordic CCR, the PTDF matrix consists of 31 columns, representing both real and virtual bidding zones, and hundreds of rows corresponding to the CNEs for each hour of the day. The zone-to-slack PTDF matrix thus captures the aggregated sensitivities of the transmission network and forms the foundation for flow-based capacity calculation [21].

### 4.2.2 Generation Shift Key Strategies

To enhance capacity calculation in the Nordic day-ahead market, eight predefined and one custom GSK strategy have been established. These strategies define how different generators and loads contribute to changes in net position. Notably, price-insensitive nodes are assigned a GSK factor of zero, meaning they do not adjust their output based on market fluctuations [21].

The first GSK strategy, GSK 0, is a custom strategy and it is designed to more accurately reflect on a bidding zone's generation and/or load in response to changes in its net position, particularly when such dynamics cannot be effectively captured by the other eight strategies [8].

In GSK 1, GSK factors are only assigned to generators with available downward regulation capacity when adjusting a bidding zone's net position. A generator's participation is determined by its available downward margin, which is calculated as the difference between current generation and minimum generation levels, $\max(P_g - P_{min}, 0)$. Consequently, generators with greater downward flexibility receive higher GSK factors, meaning they play a more significant role in balancing the system. This strategy is particularly useful in bidding zones where downward regulation is limited and where reducing generation is necessary to maintain system stability [8].

A similar approach is used in GSK 2, where only generators capable of increasing their output are assigned GSK factors. A generator's participation is determined by its available upward margin, calculated as the difference between maximum generation capacity and current generation, $\max(P_{max} - P_g, 0)$. Generators with a higher capacity to ramp up production are assigned larger GSK factors. This strategy is beneficial when an increase in generation is needed to balance supply and demand [8].

In GSK 3, GSK factors are assigned proportionally to the installed maximum capacity of each generator, without considering their real-time output or flexibility constraints, $P_{max}$. This means that larger power plants contribute more to net position adjustments, regardless of their current generation levels. This strategy provides a simplified allocation method, assuming that installed capacity determines a generator's ability to participate [8].

In GSK 4, GSK factors are evenly distributed among all participating generators, regardless of their size, capacity, or flexibility. For example, in a bidding zone with four price-responsive generators, each would be assigned a GSK factor of 0.25, meaning that each generator contributes equally to the adjustment. This strategy assumes that all generators respond identically to market fluctuations and does not differentiate based on operational characteristics [8].

Instead of using installed capacity or equal distribution, GSK 5 assigns GSK factors proportionally to real-time generation levels, $P_g$. This means that generators currently producing more power will contribute more to the net position change, while those generating less will have a smaller impact. This method ensures that net position adjustments follow the current generation mix rather than theoretical capacity limits [8].

While the previous strategies focus exclusively on generation, GSK 6 is the only strategy that incorporates both generation and load in the distribution of net position changes. In this case, generators participate based on their current production levels, while load units contribute based on their current consumption. This strategy ensures a balanced allocation of adjustments across both supply and demand, making it particularly useful in bidding zones where both generation flexibility and demand response play important roles [8].

Another approach, GSK 7, assigns GSK factors exclusively to load units when adjusting net positions. GSK factors are determined in proportion to each load's current consumption, $P_l$. This means that larger electricity consumers contribute more to the adjustment than smaller consumers. This approach is particularly relevant in regions where demand-side flexibility plays a key role in grid balancing [8].

Finally, GSK 8 distributes GSK factors equally among all loads, similar to how GSK 4 distributes them among generators. In this case, all price-sensitive loads receive the same GSK factor, regardless of their actual electricity consumption. This strategy assumes that all loads respond equally to price fluctuations, providing a simplified and uniform allocation [8].

To summarize, the strategies have different approaches to determine the GSK values for a specific bidding zone. Generally, GSK 1-5 are based on generation, GSK 7 and GSK 8 are based on load, and GSK 6 is determined by both load and generation [8]. The chosen GSK strategies by each TSO are listed below in Table 2.

*Table 2: Chosen GSK strategies by each TSO for GSK calculations* [33].

| TSO | GSK strategy |
|---|---|
| Svenska kraftnät (SE) | GSK 6 |
| Fingrid (FI) | GSK 6 |
| Energinet (DK) | GSK 0 |
| Statnett (NO) | NO1: GSK 7, NO2-NO5: GSK 6 |

As previously mentioned, each TSO considers different power sources as price dependent or price independent. Consequently, certain power sources are excluded from participating in price changes within a bidding zone. In Sweden, Svenska kraftnät includes hydro power and thermal power plants in the price-dependent category, while nuclear and wind power are excluded. In Finland, Fingrid includes hydro power, fossil gas, fossil hard coal, fossil oil, biomass, and fossil peat, whereas nuclear and wind power are excluded. In Denmark, Energinet considers offshore wind farms, biomass, and fossil hard coal, fossil oil, and fossil gas as included power sources and excludes onshore wind power. Finally, in Norway, Statnett includes hydro power but excludes wind power from participating in price changes [33].

### 4.2.3 Finding the Optimal Generation Shift Key Strategy

The choice of GSK strategy has a direct impact on the calculation of PTDFs, which in turn influences both capacity allocation and market outcomes. As mentioned in [21], “there should be harmonized rules guiding how GSKs are defined in order to avoid potential obscure incentives and to ensure transparency in the capacity calculation process.”

GSK strategies are built on assumptions about which generating units respond to market changes and how their production or load levels adjust relative to the total generation and consumption in a bidding zone. The combination of these assumptions determines how accurately a GSK strategy reflects actual system behavior. A *flat GSK strategy* assumes that all price-sensitive nodes contribute equally to changes in generation. While this approach is not fully realistic, it is considered robust because it does not depend on detailed assumptions about local dispatch patterns or unit-specific responses. In contrast, *marginal or production-based GSK strategies* incorporate stronger assumptions about which generators are more likely to adjust their output, aiming to better reflect short-term operational behavior [21].

The key difference between these approaches lies in their trade-off between robustness and accuracy. When the market outcome closely resembles the base case used in the capacity calculation, a marginal GSK strategy typically provides more accurate PTDF estimates, as it captures the expected generation distribution more precisely. However, when market conditions deviate from the base case, such strategies may misrepresent the true system response. In such cases, the flat GSK approach tends to perform better, as its uniform weighting makes it less sensitive to errors in predicting individual generator behavior [21].

In summary, selecting an optimal GSK strategy involves balancing accuracy and robustness. Marginal or production-based approaches can improve precision when forecasts are reliable, while the flat strategy offers greater stability under uncertain or changing market conditions. Therefore, the most suitable choice depends on how well the assumed base case aligns with actual market operations and how sensitive the system is to deviations from that reference state [21].

# 5 Methodology

This section outlines the methodology used to calculate GSKs for the Nordic CCR and the approach taken to evaluate their performance compared to a flat GSK strategy. The goal was to investigate whether a more detailed, data-driven GSK calculation improved the accuracy of zonal PTDFs estimates compared to the official reference PTDFs from the network model dated 2024-04-01 7:00 CET to 2024-04-02 00:00 CET.

Data were primarily sourced from Vattenfall's Model Calculation Platform (MCP) (retrieved from ENTSO-E's Transparency Platform), which provided hourly generation data per energy source and power plant, complemented with internal documentation on installed capacities. Since not all power plants reported hourly production data, missing generation was estimated proportionally based on installed capacity within each bidding zone and energy source.

The calculated GSKs represented the average share of each power plant's generation relative to the total generation in its bidding zone over selected time intervals: *monthly (1-30 April), weekly (1-7 April), and daily (1 April).* These GSKs were then implemented in Vattenfall's grid model based on PyPSA and grid data from Energinet to simulate PTDFs. To benchmark the effectiveness of the data-driven GSKs, they were compared against a flat GSK strategy that assumed equal distribution of generation across all nodes in a bidding zone. The evaluation metric was the Root Mean Square Error (RMSE) between the PTDFs generated by each GSK model and the official PTDFs provided by the JAO.

Some simplifying assumptions were made to address data gaps and modeling constraints. For example, power plants below 20 MW capacity were excluded to focus on significant generation units. For bidding zones with comparability limitations (NO1, DK1, DK2, and virtual bidding zones), a flat GSK was assigned for modeling completeness, but these bidding zones were excluded from the PTDF error comparison.

The following subsections detail the specific tools, datasets, and calculation steps used to implement the methodology described above.

## 5.1 Programming Language and Software Libraries

All analyses and data processing in this master's thesis were performed using Python. The language was selected for its readability and wide array of libraries, particularly for geographic data and numerical computations.

The following libraries were essential:

- `pandas`: for time-series handling, especially to manage and process the generation datasets on a per-hour basis.

- `geopandas`: to process and analyze geographic data, especially for assigning coordinates to generators.
- `geopy`: to compute geodesic distances between geographic points.
- `fuzzywuzzy`: for fuzzy matching of generator names across datasets.

With the software environment established, the next step was to calculate GSK values using the available generation and capacity data.

## 5.2 Generation Shift Key Calculation Framework

This subsection explains how the GSKs were calculated. It describes the main steps, which were collecting and processing data, calculating GSK values, and linking them to the grid model. Both internal and open data sources were used to ensure reliable results. The subsection also describes the main assumptions and delimitations of the methodology, and includes a flat GSK as a reference for comparison.

### 5.2.1 Data Collection

The primary source of data used was Vattenfall's internal platform, MCP. MCP provides hourly time-series data of the total generation per energy source (such as hydro, wind, biomass, etc.) across different bidding zones. Additionally, it includes individual generation data for each power plant that is listed on ENTSO-E's Transparency Platform [34]. However, since not all power plants are included on ENTSO-E, the MCP data is incomplete.

To address this gap, internal documentation from Vattenfall was used to complement the MCP data. These documents included each power plant's installed capacity, power source, and its bidding zone. This combination of MCP data and internal documentation allowed for a more complete and accurate representation of generation in the Nordic CCR.

### 5.2.2 Generation Shift Key Calculation Methodology

The method used to calculate GSKs for each country in the Nordic CCR followed the same standardized process, described below (with the exception and exclusions noted in Section 5.2.4). The approach is a hybrid of GSK 3 and GSK 5, and explicitly *production-based* because load data were limited.

**Step 1: Download data from MCP**
The first step was to download the data from the MCP platform. As previously mentioned, the dataset included:

- Hourly total production in MW per energy source and bidding zone.
- Hourly production data in MW per individual power plant.

**Step 2: Structuring of power plant data**

In order to calculate GSKs accurately, production data per power plant was organized by bidding zone and energy source. Where multiple generators existed within a plant, they were summed (e.g., Harsprånget units 1-5 combined).

**Step 3: Estimation of missing production data**

To address the gap between the total production reported per bidding zone and the sum of the available individual power plant data, the remaining production was calculated. This gap exists because not all power plants report hourly generation at the unit level, leading to incomplete datasets. To ensure accurate estimation of GSKs, it is important to approximate this missing generation.

The remaining production for each energy source within a specific bidding zone was computed using Equation 11:

$$P_{\text{rest},t} = P_{\text{tot},t} - P_{\text{plant},t}, \qquad \text{where} \qquad t = \text{hour} \tag{11}$$

- $P_{\text{tot},t}$ is the total reported production for a specific energy source and bidding zone at hour $t$,
- $P_{\text{plant},t}$ is the summed generation of all individual power plants, for a specific energy source and bidding zone, with available data at hour $t$,
- $P_{\text{rest},t}$ is the calculated remaining production for a specific energy source and bidding zone at hour $t$.

To illustrate how this calculation was applied in practice, Figure 4 shows an example for hydro power in NO5 over a 72-hour period. The blue solid line represents the total production, based on bidding zone-level data. The green solid line indicates the aggregated generation from known individual power plants. The difference between these two is shown as the orange dashed line, with a shaded area highlighting the remaining production. This represents the portion of generation not accounted for by the available plant-level data.

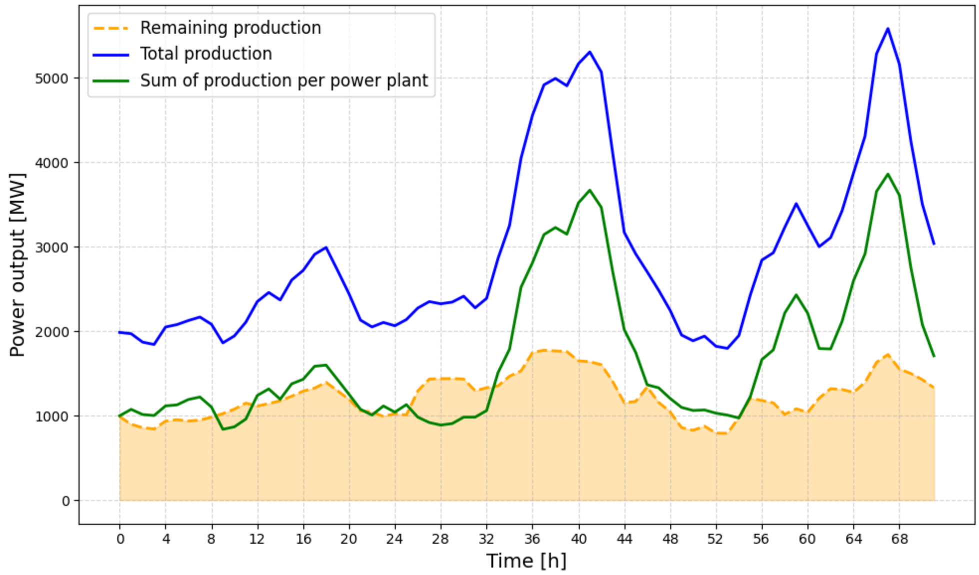


*Figure 4: Hydro power production in NO5 over a 72-hour period. The graph shows the total production, the sum of known plant-level production, and the remaining production calculated using Equation 11.*

The remaining production is later distributed proportionally across unidentified power plants using their installed capacities, allowing for the reconstruction of a more accurate generation profile per bidding zone.

**Step 4: Allocation of remaining production**

The remaining production calculated in Step 3 was distributed proportionally among the power plants for which detailed generation data was not available. This allocation was based on internal documents including installed capacities of these plants, assuming that their contribution to the missing generation is proportional to their capacity within the bidding zone and energy source.

To perform this allocation, internal documents listing the names and installed capacities of power plants were used. These power plants were sorted by bidding zone and energy source. For each power plant, the share of the remaining production at each hour was calculated according to Equation 12:

$$P_{i,t} = \frac{P_{\text{installed},i}}{\sum_j P_{\text{installed},j}} \cdot P_{\text{rest},t} \tag{12}$$

where:

- $P_{i,t}$ is the estimated generation of power plant $i$ at hour $t$,
- $P_{\text{installed},i}$ is the installed capacity of power plant $i$,
- $\sum_j P_{\text{installed},j}$ is the total installed capacity of all unreported power plants within the same bidding zone and energy source,

- $P_{\text{rest},t}$ is the remaining production to be allocated at hour $t$ (from Equation 11).

This proportional allocation enables reconstruction of a more complete generation profile at the plant level, which is essential for the calculation of GSKs.

**Step 5: Calculation of GSK values**

Once complete hourly generation profiles were established for the power plants, both those with reported data and those with estimated values, GSKs were calculated. For each bidding zone, and for each time interval (month, week, day), the mean generation of each plant was divided by the total mean generation across all plants within the same bidding zone (aggregated over relevant energy sources). This resulted in a proportional value between 0 and 1 for each power plant, representing its average share of the total generation within the bidding zone. The GSK for each plant was calculated using Equation 13:

$$GSK_i = \frac{\overline{P}_i}{\overline{P}_{\text{total}}} \tag{13}$$

where:

- $GSK_i$ is the GSK for power plant $i$,
- $\overline{P}_i$ is the mean generation of plant $i$ over the chosen interval (monthly, weekly or daily),
- $\overline{P}_{\text{total}}$ is the total mean generation of all plants in the bidding zone, defined as

$$\overline{P}_{\text{total}} = \sum_{N=j} (\overline{P}_{\text{MCPplants}} + \overline{P}_{\text{estimated}}), \quad j = \text{energy sources (hydro, thermal, etc).}$$

GSK values were assigned to all power plant names that directly matched the corresponding entries in Vattenfall's grid model.

### 5.2.3 Assigning Coordinates to Power Plants Using Open Data

For the remaining power plants without a direct match, a coordinate assignment procedure was applied to ensure their inclusion in the grid model for PTDF computation. A key challenge was that many nodes in the grid model were labeled only with numerical identifiers (e.g., `"5672"`) rather than descriptive power plant names. This made it difficult to directly link the generation data to the network model using names alone. Therefore, a coordinate-based matching procedure was required to ensure that each generator could be connected to the correct node in the grid model.

The primary reference dataset for coordinates was the `Spotnic` dataset, which contains power plant names (e.g., `"5672"`) and their corresponding coordinates. The coordinates in the dataset were then linked with OSM to retrieve the actual name of the power plant. OSM is a collaborative, open-source platform where contributors worldwide map infrastructure such as roads, buildings, land use, and energy facilities. A notable strength of OSM is its global coverage and rich tagging system, allowing attributes like `power=plant` and `name=Harsprånget` to be associated with mapped features [35].

To extract relevant power plant data from OSM, the query tool Overpass Turbo was utilized [36]. An Overpass QL query was constructed to retrieve all features tagged with `power=plant` within Sweden, Norway, and Finland, as illustrated in Figure 5. The extracted data were exported in GeoJSON format for further processing in Python.

```
[out:json][timeout:180];

area["ISO3166-1"="SE"]->.se;
area["ISO3166-1"="NO"]->.no;
area["ISO3166-1"="FI"]->.fi;

// All power plants in each area
(
  way["power"="plant"](area.se);
  way["power"="plant"](area.no);
  way["power"="plant"](area.fi);
);

// Include all geometry
(._; >;);
out body geom;

```

*Figure 5: Overpass Turbo query used to extract power plants from OSM.*

For GSK names without direct matches in the grid model, fuzzy string matching was applied with the `fuzzywuzzy` library to link each unmatched GSK name to a corresponding power plant name in the OSM dataset. To increase match rates, several common suffixes (e.g., `"kraftstation"`, `"power plant"`) were appended to GSK names during matching.

Many OSM power plant features are represented as polygons rather than points, which complicates coordinate extraction. To address this, the `representative_point()` method from the `GeoPandas` library was used. Unlike centroids, which can lie outside irregular polygons, the representative point is guaranteed to lie within the polygon's boundaries, ensuring a valid location within the mapped plant area. Once a representative point was obtained for a matched OSM feature, it was used to find the geographically nearest power plant within the Spotnic dataset, restricted to the same bidding zone, by calculating geodesic distances using the `geopy` library. Only matches whose nearest Spotnic coordinate fell within a **5 km** radius of the OSM-derived representative point and within the same bidding zone were accepted and assigned a GSK. Nodes that did not satisfy this 5 km criterion did not receive a direct match-based GSK. Instead, the remaining GSK mass in the zone was allocated uniformly to the unmatched nodes according to Equation 14:

$$GSK_{\text{unmatched}} = \frac{1 - \sum_{i \in M} GSK_i}{N}, \tag{14}$$

where $M$ is the set of matched nodes and $N$ is the set of unmatched nodes in the bidding zone.

### 5.2.4 Assumptions and Delimitations

The methodology presented above relies on a set of assumptions and delimitations that ensure a consistent GSK calculation across all bidding zones in the Nordic CCR. First, in line with Table 2, Statnett applies GSK 7 in NO1, which is a load-based approach. Since this master's thesis focuses on generation, the NO1 approach could not be replicated. To maintain model completeness, all generators in NO1 were therefore assigned equal factors (a flat strategy). For the PTDF error comparison, however, NO1 is excluded.

A plant filter based on installed capacity was also applied to manage the size of the dataset. All power plants with installed capacity below 20 MW were excluded from the analysis and hence did not receive GSKs.

For comparability with the official reference PTDFs, DK1 and DK2 are likewise excluded from the error analysis. DK1 belongs to a different synchronous area, which complicates direct comparison, while DK2 presented challenges in decoding power plant names. Virtual bidding zones are also excluded. Flat GSKs are still assigned to NO1, DK1, and DK2 inside the grid model so that the network remains solvable, but these bidding zones do not enter the error metrics.

### 5.2.5 Flat Generation Shift Key Strategy

To evaluate whether the GSK models provided a more accurate estimation of the PTDFs, they were compared against a benchmark approach: the flat GSK, referred to in this master's thesis as GSK 4. This method assumes an equal distribution of generation across all available nodes within each bidding zone. In practice, this meant that each node was assigned a GSK factor of $\frac{1}{N}$, where $N$ was the total number of nodes in the zone. This assumption-based strategy served as a reference point to assess whether more data-driven GSK methods offered improved PTDF estimations.

## 5.3 Power Transfer Distribution Factor Calculation Framework

This subsection describes how PTDFs were calculated and evaluated using Vattenfall's grid model. It outlines the model structure and how the results were compared with official reference PTDFs from the JAO. The evaluation is based on the RMSE metric, which quantifies how accurately different GSK approaches reproduce observed power flow patterns across the Nordic CCR.

### 5.3.1 Vattenfall's Grid Model

The zonal PTDF calculations in this master's thesis are based on Vattenfall's grid model which is derived from PyPSA and grid data from Energinet. PyPSA is an open-source framework for simulating and optimizing electrical networks [37]. The model follows a DC load-flow approximation, and the power system is represented by buses, transmission lines, and transformers, with line flows governed by their susceptances. Where line parameters were not publicly available, standardized assumptions based on typical equipment characteristics were applied.

All PTDF comparisons refer to the official reference day 2024-04-01, and both network topology and parameters are aligned with that date. PyPSA's generator coordinates are approximate, which complicated one-to-one matching between OSM power plants and PyPSA nodes and therefore motivated the conservative 5 km acceptance rule described in Section 5.2.2. The network model and evaluation are also developed for a single day, and extending the analysis to multiple days would enable the identification of temporal trends and support broader generalizations.

In total, the grid model contains 11 real bidding zones (DK1 is excluded) and 14 virtual zones, and 502 GSK values as inputs to the model.

After using the grid model to compute PTDFs for both the flat and the production-based GSK approaches, a quantitative comparison was carried out to assess how well each method reproduces observed flow patterns. The following subsection details the evaluation procedure.

### 5.3.2 Evaluation of Generation Shift Key Performance

To evaluate how accurately the different GSK models reproduce the physical flow patterns, a comparison was carried out between the modeled PTDFs and the official reference PTDFs provided by the JAO. The accuracy of each GSK model was assessed using the RMSE metric, which quantifies the average deviation between two sets of values. In this context, a lower RMSE indicates a closer agreement between the PTDFs calculated from the model and those from the official reference [38].

The RMSE was calculated for each GSK model according to Equation 15, where the difference between each modeled PTDF value and the corresponding JAO reference value was squared, averaged across all evaluated elements, and then square-rooted to obtain the overall error magnitude.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(PTDF_i^{\text{model}} - PTDF_i^{\text{JAO}})^2} \tag{15}$$

- $PTDF_i^{\text{model}}$ is the PTDF value for a CNE and bidding zone pair as computed using the modeled GSKs,

- $PTDF_i^{\text{JAO}}$ is the corresponding reference PTDF value from the JAO, and
- $N$ is the total number of compared PTDFs.

First, the RMSE was computed for the flat GSK strategy, which served as a benchmark for evaluating the relative improvements achieved by the data-driven configurations. Next, RMSE values were calculated for each of the production-based GSK models. These results were then compared directly with the flat case to assess whether and to what extent the data-driven approaches improved zonal PTDF accuracy.

To further examine variability, the RMSE was also computed individually for each bidding zone. This zonal evaluation highlights how the performance of different GSK models varies across the Nordic CCR.

# 6 Results

This section presents the outcomes from implementing the GSK calculation framework and evaluating its influence on zonal PTDF estimation accuracy. The objective is to determine how different GSK models affect alignment with the official reference PTDFs published by the JAO. Throughout this section, performance is assessed using the RMSE between modeled and reference PTDF matrices. The results are structured to first present the calculated GSK values, followed by the evaluation of the resulting PTDFs. The analysis focuses on comparing the flat GSK benchmark with the data-driven, production-based approach. The comparison explores how different time resolutions affect model accuracy, stability, and zonal variation.

## 6.1 Generation Shift Key Values

The calculated GSK values represent the average contribution of each generation unit to the total production within its respective bidding zone. Two main approaches were compared:

1. A flat GSK strategy, assuming equal participation of all generators within each bidding zone (see Appendix A).
2. A data-driven GSK strategy, were GSKs were calculated based on actual generation data and installed capacities over different time-frames.

To investigate how the level of time resolution affects the results, the data-driven GSKs were calculated for three different time frames:

- Monthly GSKs: based on the monthly average generation share per plant (see Appendix B).
- Weekly GSKs: based on the weekly average generation shares per plant (see Appendix C).
- Daily GSKs: based on the daily average generation share per plant (see Appendix D).

The Appendix tables list the final GSK factors and constitute the input for the zonal PTDF calculations described in Section 6.2. Regarding matching, the share of generation units successfully matched either through direct identification or coordinate matching was as follows:

- In Norway, 31% in NO2 (17 of 55), 24% in NO3 (6 of 25), 65% in NO4 (13 of 20), and 47% in NO5 (16 of 34).
- In Finland, 17% (10 of 59).
- In Sweden, 50% in SE1 (9 of 18), 47% in SE2 (30 of 64), 28% in SE3 (19 of 68), and 6% in SE4 (1 of 17).

The generation units that were not matched, received a GSK factor according Equation 14.

## 6.2 Evaluation of Power Transfer Distribution Factor Calculations

Based on the GSK values derived in Section 6.1, PTDFs were calculated to evaluate how different GSK approaches affect the estimation accuracy. As described above, three data-driven time-frames were compared against both the flat GSK strategy and the official reference PTDFs provided by the JAO.

The RMSE between the modeled and reference PTDF matrices was used as the primary evaluation metric. The analysis included the overall RMSE for the complete PTDF matrix, RMSE values per bidding zone, and the percentage change in RMSE for each bidding zone compared with the flat GSK case.

### 6.2.1 Flat Generation Shift Key Strategy

The flat GSK strategy, in which all generation units within each bidding zone are assigned equal factors, serves as a reference case for evaluating the performance of the modeled approaches. This configuration represents a simplified assumption that does not consider the actual generation or the energy mix of production within in the bidding zones. The resulting PTDFs provide a baseline against which improvements from the more detailed and modeled GSKs can be measured. The RMSE between the PTDFs obtained by the flat GSK strategy and the reference PTDFs from the JAO was **0.02279**, which is used as the benchmark value throughout the subsequent analyses. The RMSE distribution for the PTDF matrix for this case is shown in Figure 6.

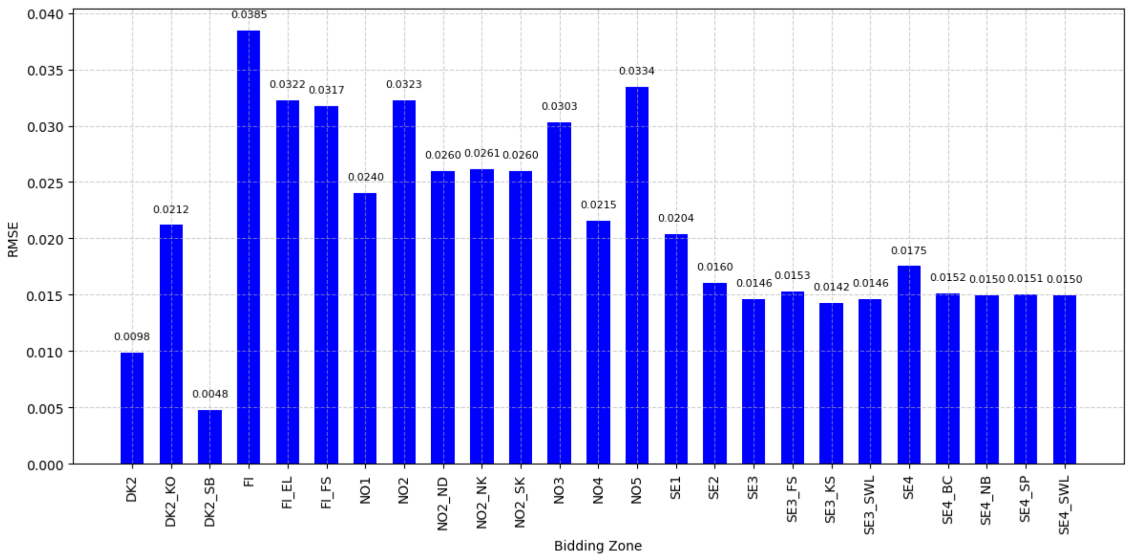


*Figure 6: RMSE distribution for the flat GSK reference case across the Nordic CCR. The figure shows the deviation between the flat strategy PTDFs and the official reference PTDFs from the JAO.*

Figure 6 illustrates how the deviation between the flat GSK strategy and the reference PTDFs varies across the bidding zones within in the Nordic CCR. As previously mentioned, virtual bidding zones, and NO1 and DK2, will not be included in the analysis.

### 6.2.2 Monthly Data-Driven Generation Shift Keys

Replacing the flat GSK reference case with GSKs based on the mean monthly generation of each generation unit resulted in a slightly improved alignment with the reference PTDFs. The overall RMSE for the complete PTDF matrix decreased to **0.02251**, corresponding to an improvement of approximately **1.25%** compared to the flat GSK benchmark. Figure 7 presents the RMSE for both the flat reference case and the monthly GSK model, separated by bidding zone.

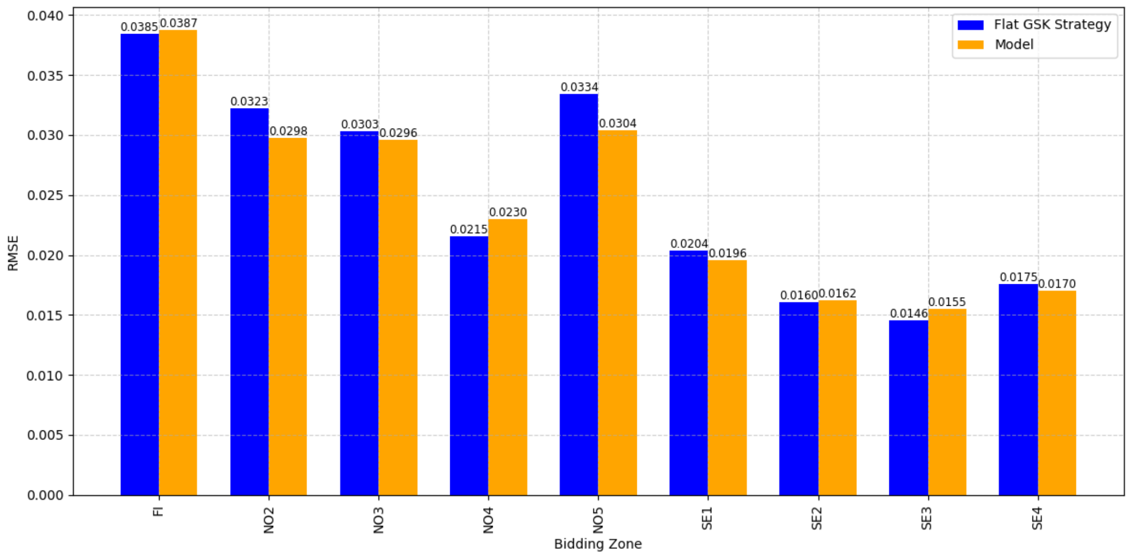


*Figure 7: RMSE per bidding zone for the flat reference and monthly data-driven GSKs.*

To examine the performance difference more clearly, Figure 8 displays the RMSE difference between the monthly GSKs and the flat reference for each bidding zone, where negative values correspond to improved accuracy and positive values correspond to worsened accuracy when applying the monthly GSKs. The majority of the bidding zones show negative differences, confirming that the data-driven approach enhances PTDF estimation.

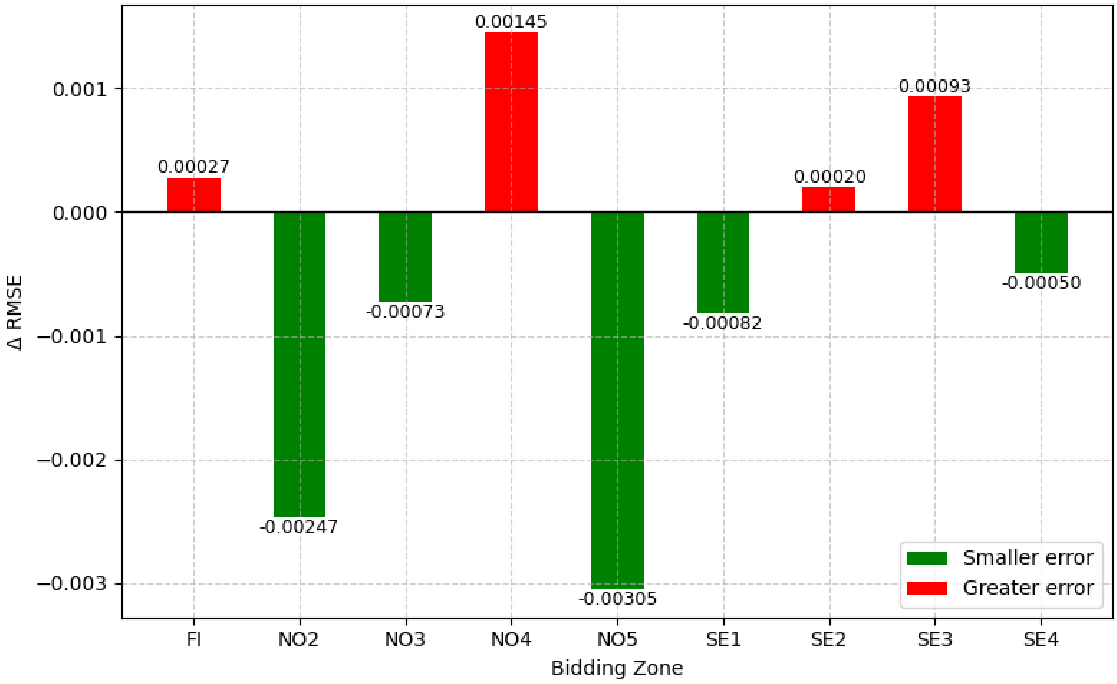


*Figure 8*

Additionally, Figure 9 shows the percentage change in RMSE in each bidding zone when comparing the monthly data-driven GSKs to the flat reference case. It also shows that the largest reductions in RMSE occur in NO5 (-9.14%), NO2 (-7.67%) and SE1 (-4.01%), and that the GSK model does not outperform the flat reference case in FI, NO4, SE2 and SE3.

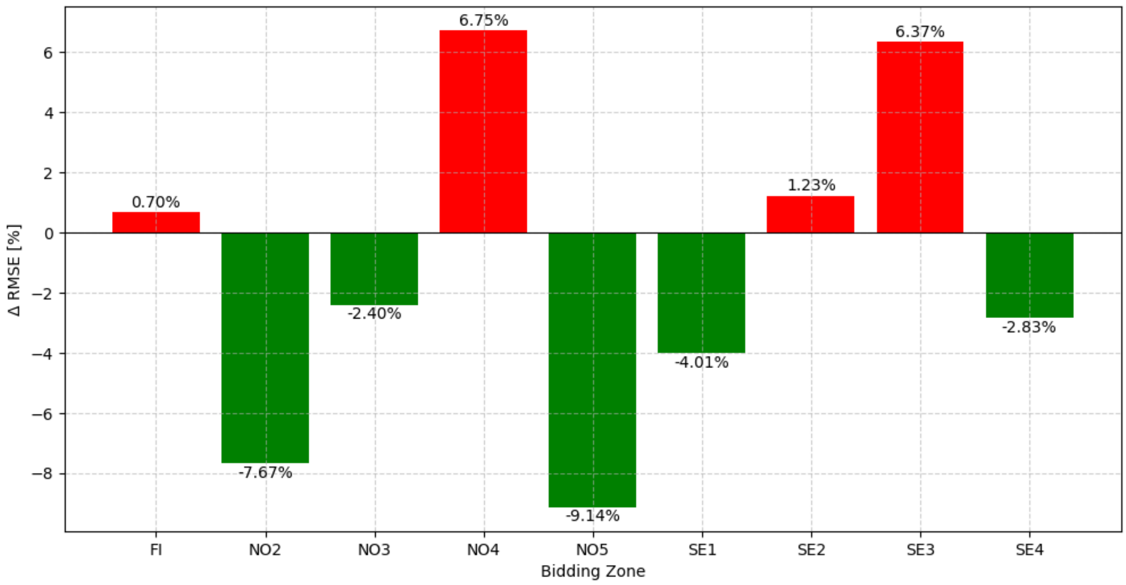


*Figure 9: Percentage change in RMSE between the flat reference and monthly GSKs per bidding zone.*

In summary, the monthly data-driven GSKs provide a modest improvement compared to the flat reference case. The results indicate that incorporating average monthly generation data better reflects long-term production patterns and seasonal variations, leading to slightly more accurate PTDF estimations. However, since the improvement remains small, the benefits of using monthly GSKs are limited to capturing gradual shifts in the generation mix rather than short-term operational changes. This suggests that while monthly GSKs enhance model realism, their impact on overall accuracy is relatively minor.

### 6.2.3 Weekly Data-Driven Generation Shift Keys

To evaluate how a shorter time resolution affects the result, weekly average generation data were used to calculate GSKs for each bidding zone. The overall RMSE for the weekly GSKs was **0.02252**, compared with 0.02279 for the flat reference. This corresponds to an average improvement of approximately **1.17%**, indicating that the weekly resolution slightly improves the overall PTDF estimation accuracy. This approach captures changes in production that happen on a weekly basis. Compared to the monthly setup, the weekly GSKs are expected to follow the actual generation mix more closely, but they may also create larger variations from week to week. Figure 10 shows the RMSE for each bidding zone when comparing the flat reference case with the weekly GSKs.

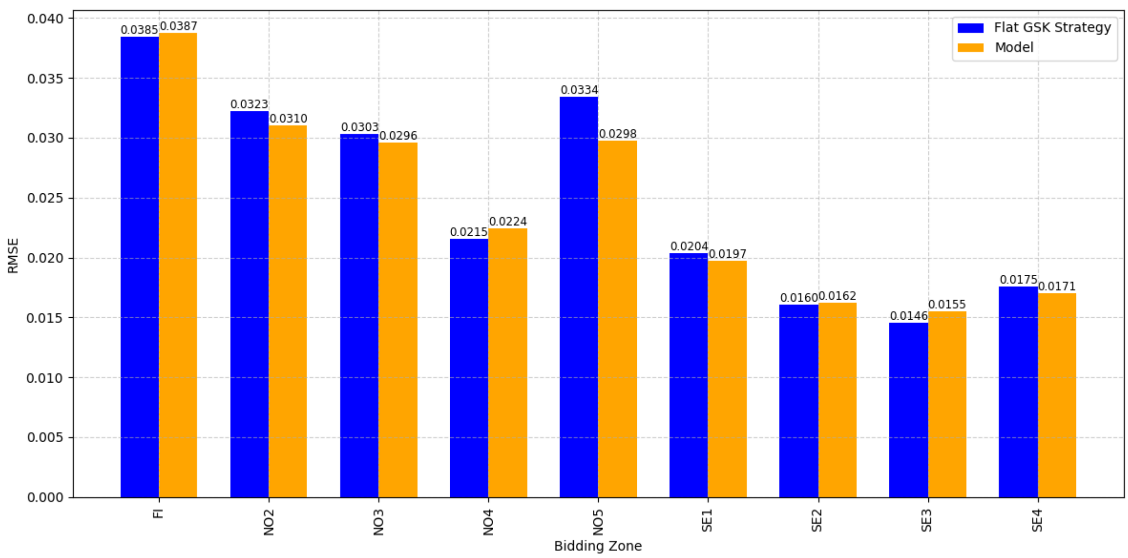


*Figure 10: RMSE per bidding zone for the flat reference and weekly data-driven GSKs.*

The differences between the two approaches are shown more clearly in Figure 11, which presents the change in RMSE for each bidding zone. Negative values mean that the weekly GSKs perform better, while positive values show where the flat GSKs give lower errors.

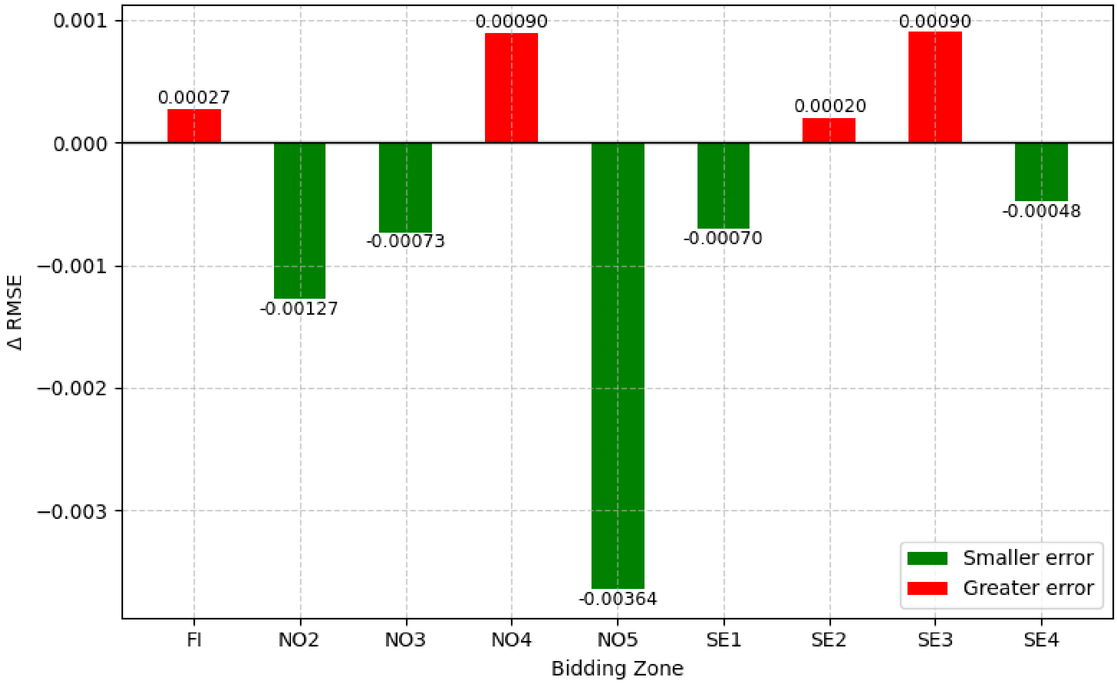


*Figure 11: Difference in RMSE between the flat reference and weekly data-driven GSKs.*

Figure 12 shows the percentage change in RMSE in each bidding zone when comparing the weekly data-driven GSKs to the flat reference case. It also shows that the largest reductions in RMSE occur in NO5 (-10.90%), NO2 (-3.94%) and SE1 (-3.43%), and that the GSK model does not outperform the flat reference case in FI, NO4, SE2 and SE3. In contrast to the monthly GSK model, the error increases in NO2, SE1, SE2, and SE4, while in this time frame it decreases in NO4, NO5, and SE3.

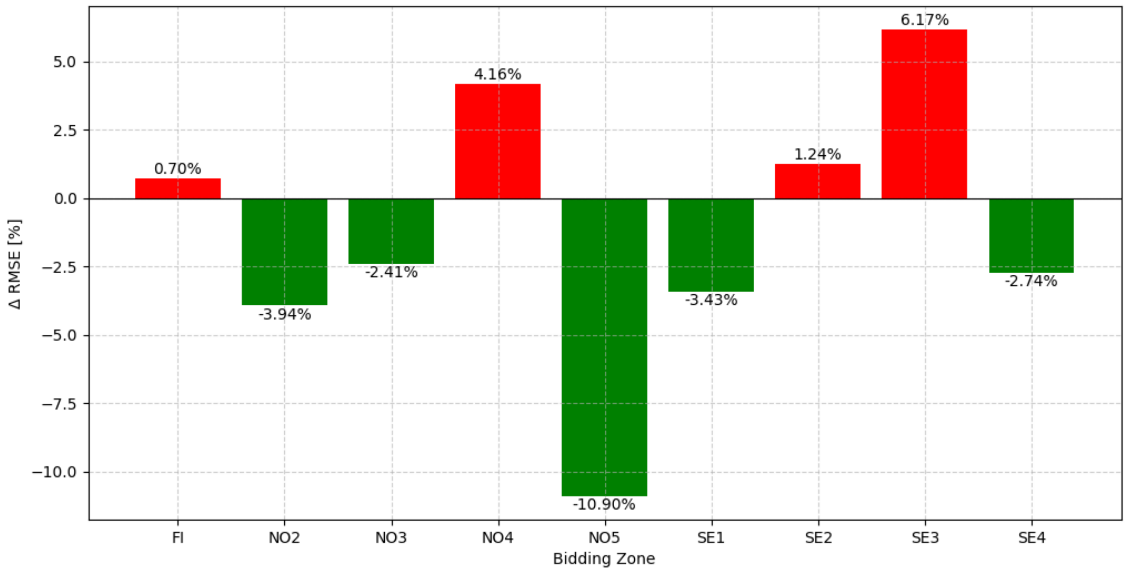


*Figure 12: Percentage change in RMSE between the flat reference and weekly GSKs.*

In summary, the weekly GSKs provide a more detailed description of how generation changes over time, but the results vary between bidding zones. The method would hypothetically work best in areas with stable production, such as the northern hydro-dominated zones, but less well in regions where production varies rapidly. Although weekly data make the model more responsive to actual operational changes, they also reduce the overall stability of the PTDF calculations.

### 6.2.4 Daily Data-Driven Generation Shift Keys

To test how a very short time resolution affects the results, daily average generation data were used to calculate GSKs for each bidding zone. While this approach gives a more detailed description of how the system operates on a daily basis, it can also introduce instability, since production patterns can change significantly within short time periods. The overall RMSE for the daily GSKs was **0.02267**, compared with 0.02279 for the flat case. This corresponds to an improvement of only about **0.54%**, meaning that daily GSKs lead to slightly lower overall error but do not significantly enhance model accuracy. Instead, they increase variability between days and across bidding zones. Figure 13 shows the RMSE per bidding zone for both the flat reference and the daily GSKs.

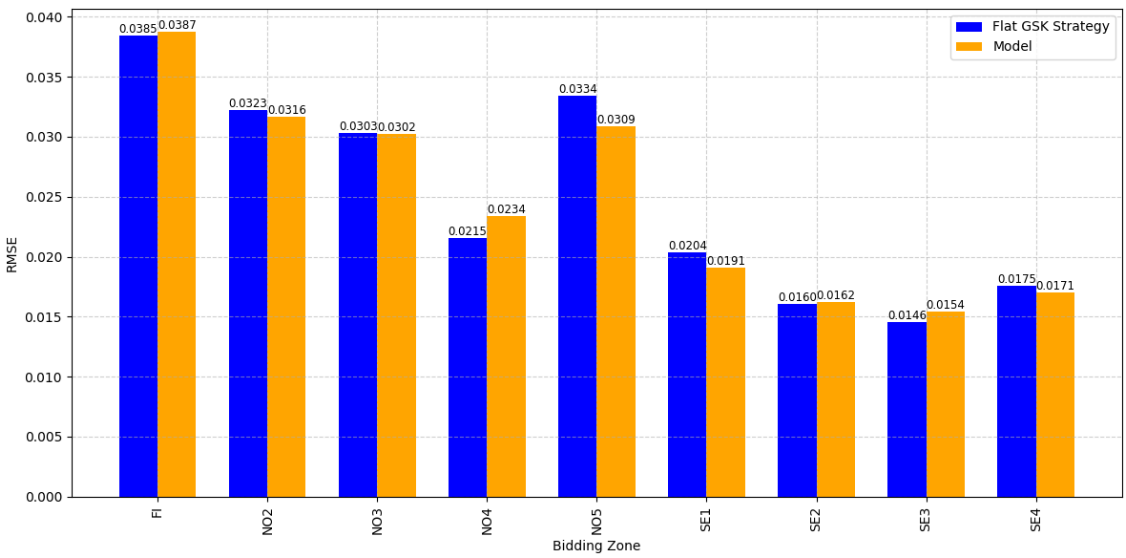


*Figure 13: RMSE per bidding zone for the flat reference and daily data-driven GSKs*

The differences between the two methods are presented more clearly in Figure 14, which shows the absolute change in RMSE for each bidding zone. Negative values represent an improvement when using daily GSKs, while positive values indicate that the flat reference performs better.

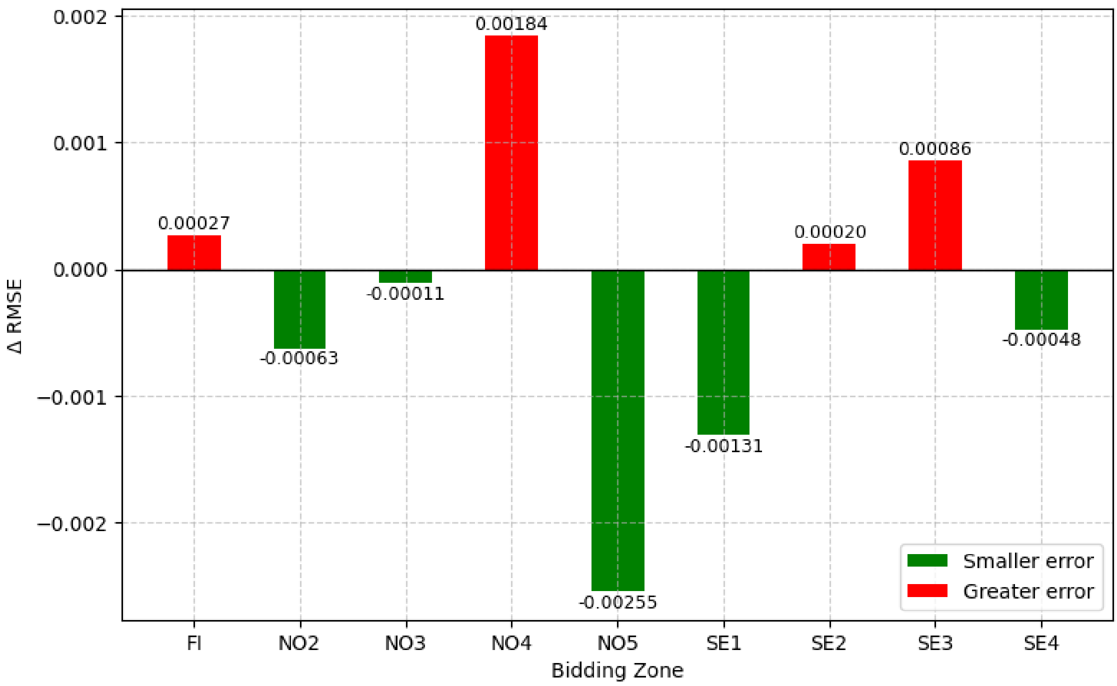


*Figure 14: Difference in RMSE between the flat reference and daily data-driven GSKs*

Figure 15 shows the same results as percentage changes compared to the flat case. It also shows that the largest reductions in RMSE occur in NO5 (-7.62%), SE1 (-6.43%) and SE4 (-2.74%), and that the GSK model does not outperform the flat reference case in FI, NO4, SE2 and SE3.

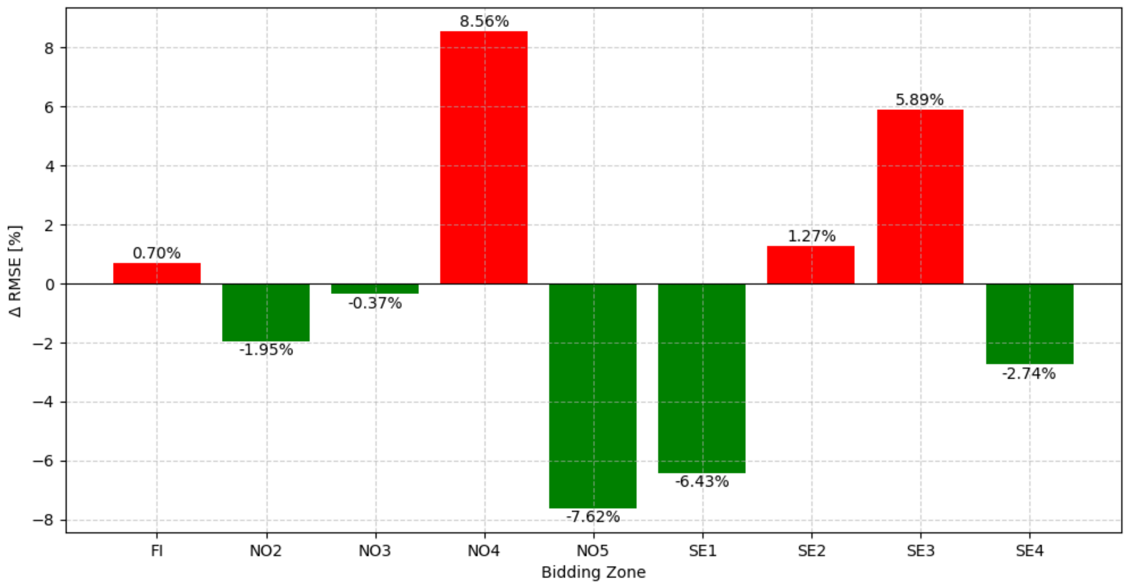


*Figure 15: Percentage change in RMSE between the flat reference and daily GSKs*

In summary, the daily GSK approach captures the most detailed generation variations and can reflect short-term system conditions with high precision for specific days. However, this high level of detail also makes the results less stable over time, as daily production changes are often driven by temporary weather conditions.

## 6.3 Summary of Result

This section summarizes the results from the evaluation of the different GSK approaches. The purpose is to understand how the level of time resolution influences the accuracy and stability of the PTDFs. Both the overall system performance and the zonal differences are presented below.

Table 3 shows the overall RMSE values for each GSK strategy compared to the flat reference. All data-driven GSKs perform better than the flat case, although the level of improvement varies. The monthly GSKs achieved the lowest RMSE value, **0.02251**, corresponding to an improvement of about **1.25%** compared with the flat reference. The weekly GSKs followed closely with a value of **0.02252** and an improvement of **1.17%**, while the daily GSKs resulted in a value of **0.02267** which a smaller improvement of **0.54%**. These results show that including generation data generally enhances PTDF estimation, but a higher time resolution does not necessarily lead to better overall accuracy. The monthly GSKs therefore provide the most balanced and reliable configuration, combining accuracy with numerical stability.

*Table 3: Overall RMSE and relative improvement for each GSK strategy compared to the flat reference.*

| GSK model | RMSE | Improvement |
|---|---|---|
| Flat reference | 0.02279 | - |
| Monthly | 0.02251 | 1.25% |
| Weekly | 0.02252 | 1.17% |
| Daily | 0.02267 | 0.54% |

Table 4 provides a zonal comparison, showing the relative change in RMSE (%) for each bidding zone when comparing the monthly, weekly, and daily GSKs with the flat case. Negative values indicate an improvement (a lower RMSE) compared to the flat GSK, while positive values represent an increase in error. To visualize these results, a continuous color scale is applied. Green shades indicate an improvement relative to the flat reference, with darker green representing a larger improvement. Moreover, red shades indicate an increased error compared to the flat case, where darker red corresponds to a larger increase in RMSE.

*Table 4: Percentage change in RMSE per bidding zone for monthly, weekly, and daily GSKs compared to the flat reference. Negative values indicate an improvement. The color scale represents the magnitude and direction of change relative to the flat case: green shades indicate an improvement (darker green = larger improvement), while red shades indicate a deterioration (darker red = larger error).*

| BZ | Monthly | Weekly | Daily | Best Model |
|---|---|---|---|---|
| FI | 0.70 | 0.70 | 0.70 | Equal |
| NO2 | -7.67 | -3.94 | -1.95 | Monthly |
| NO3 | -2.40 | -2.41 | -0.37 | Weekly |
| NO4 | 6.75 | 4.16 | 8.56 | Weekly |
| NO5 | -9.14 | -10.90 | -7.62 | Weekly |
| SE1 | -4.01 | -3.43 | -6.43 | Daily |
| SE2 | 1.23 | 1.24 | 1.27 | Monthly |
| SE3 | 6.37 | 6.17 | 5.89 | Daily |
| SE4 | -2.83 | -2.74 | -2.74 | Monthly |

The table shows that the optimal time resolution varies between zones. Several bidding zones, including NO2, NO3, NO5, SE1 and SE4, show lower RMSE values (green shades) for all time resolutions. However, NO4, SE2 and SE3 display higher RMSE values (red shades) across all configurations. In Finland (FI), all models produce nearly identical results, with minimal differences between time resolutions.

When grouping the bidding zones by country, the results show that the monthly and weekly models perform best overall in Norway, where most zones have their lowest RMSE for the weekly GSKs. In Sweden, the best-performing model varies between zones: SE1 and SE3 achieve the lowest RMSE with the daily GSKs, while SE2 and SE4 perform best with the monthly GSKs. In Finland, the RMSE values remain consistent across all models, indicating no significant differences between the time resolutions.

In summary, all data-driven GSK models show overall improvements in RMSE compared to the flat reference. However, the magnitude of the improvement differs between time resolutions and across bidding zones. On a system level, the monthly GSK achieves the lowest overall RMSE, followed by the weekly and daily models, and at a zonal level, the optimal time resolution varies.

# 7 Discussion

This section discusses the main findings of this master's thesis in relation to the research questions, with a focus on how the choice of a GSK model and its time resolution influence the accuracy of zonal PTDF estimations. The discussion also considers methodological aspects such as data limitations, coordinate matching, and differences between the implemented GSK models and the operational GSK strategies. Finally, implications for future flow-based market modeling in the Nordic CCR are outlined.

## 7.1 Evaluation of the GSK Results

The results show that including real generation data improves the estimation of PTDFs compared to the flat reference case. All three data-driven configurations—monthly, weekly, and daily—produced lower RMSE values than the flat GSK strategy, confirming that generation-based participation better represents the physical response of the power system. The monthly GSKs achieved the lowest overall RMSE, followed by the weekly and daily configurations. Although the differences between them are small, the pattern reveals how the level of time resolution affects both accuracy and stability.

The monthly GSKs provided the most stable and reliable results. By averaging production across longer periods, they filter out short-term fluctuations and capture structural generation behavior. The best performance of the monthly GSKs was found in zones such as NO2, NO5, SE1, and SE4, which are characterized by stable hydropower and thermal generation. These zones displayed generation patterns consistent with their monthly averages, leading to PTDFs that aligned closely with the official reference.

The weekly GSKs performed nearly as well as the monthly ones, with only a marginal increase in RMSE. Weekly aggregation reflects system conditions over a shorter horizon, allowing the model to capture adjustments in flexible hydropower generation without introducing excessive noise. The Norwegian zones NO3-NO5 showed the clearest improvement using weekly GSKs, consistent with their high share of flexible hydro capacity. Weekly GSKs therefore provide a good balance between detail and stability, especially in systems where generation changes often but follows regular patterns.

The daily GSKs were based on the actual generation on the reference day, giving the highest level of detail in time. However, they did not produce the most accurate results. Even though they matched the timing of the JAO reference PTDFs perfectly, the improvement in RMSE was small (only 0.54% compared to the flat case). This shows that more detailed data do not always lead to better accuracy. In practice, daily GSKs can become too sensitive to short-term changes, capturing random fluctuations instead of typical system behavior.

Overall, the results show that the best time resolution depends on how the model will be used. Monthly GSKs are the most stable and reliable choice for long-term or market-level analysis. Weekly GSKs work better for short-term operational studies where generation changes more often. Daily GSKs provide more detail but are also more sensitive to small fluctuations that may not reflect real physical changes. For flow-based market coupling, monthly GSKs therefore offer the best balance between accuracy and stability.

The comparison with the flat GSK strategy highlights the balance between stability and accuracy. The flat GSK, which gives all generators in a bidding zone the same weight, is simple and very stable. It does not depend on assumptions about which power plants are marginal or how their production changes with prices. Because of this, the flat GSK works well when the power system behaves differently from the base case used in the flow-based model. In such situations, spreading the changes evenly across all generators avoids relying too much on uncertain or rapidly changing data. However, this stability comes at the cost of lower accuracy. The flat GSK ignores that different generation types behave differently, for example, hydro and thermal plants often adjust production with price, while nuclear and wind plants do not. When the system is close to the base case and generation patterns are predictable, the flat GSK can miss these differences and underestimate real changes in power flow. In contrast, the data-driven GSKs capture these variations more realistically by weighting each generator based on its actual share of production.

The results confirm this trade-off. The flat GSK had higher RMSE values overall but produced consistent results across zones, while the data-driven GSKs achieved lower RMSEs but showed more variation between bidding zones. This suggests that both approaches are useful in different situations. When data are limited or uncertainty is high, the flat GSK is a stable and reliable option. When detailed and accurate generation data are available, data-driven GSKs give a more realistic picture of how the grid behaves.

## 7.2 Methodological Reflections

The method developed in this master's thesis was designed to balance data availability, computational efficiency, and accuracy. A hybrid approach was used, combining parts of GSK 3 and 5. This made it possible to represent both the size and mix of generation within each bidding zone in a consistent way.

A key difference between this study and current TSO practice is that all Nordic TSOs, except Statnett for NO1, use GSK 6, which includes both load and generation. In this master's thesis, only generation-based GSKs were used, since detailed load data were not available. As a result, demand-side flexibility, which can have a significant effect in some bidding zones, was not included. In regions with large industrial consumers or flexible loads, this likely caused the model to overestimate the impact of generators and underestimate the role of demand. Including both load and generation in future work would make the results more consistent with how GSK 6 is used in practice and could reduce these differences.

The completeness and accuracy of the data also affected the results. Not all power plants reported hourly generation and missing data were estimated based on installed capacity within each bidding zone and energy source. This ensured that total zonal generation matched reported values but assumed that plants of the same type behave similarly. While this assumption works reasonably well for thermal power, it is less accurate for hydropower, where local water conditions strongly affect production. This simplification may have caused some power plants to receive slightly overestimated GSKs while others were underestimated, which might have influenced the zonal PTDFs.

Matching the generation data to the grid model was an important step that could also influence the overall accuracy of the results. Power plants were connected to PyPSA nodes using both name matching and geographic coordinates from the Spotnic and OSM datasets. Since many nodes in the grid model only had numerical identifiers (for example, "5672"), coordinate matching was necessary to correctly place each generator. When a direct name match was not possible, fuzzy matching and a conservative 5 km distance limit were applied. Plants that could not be matched within this range were assigned uniform GSK values within their bidding zone to ensure that the model remained solvable. This approach reduced the risk of incorrect matches but also introduced some uncertainty in how generation was distributed across nodes.

The varying success rate of coordinate matching across bidding zones may have affected how generation was weighted in the model. In bidding zones with lower match rates, many power plants did not receive a direct coordinate match and were therefore assigned uniform GSKs. This means that some generators could have been given either too high or too low participation weights compared to their actual role in the system. For example, non-price-sensitive units such as nuclear or wind power may have been included even though they do not normally adjust their output in response to price changes, while important flexible generators without coordinates may have been underrepresented. These discrepancies could have influenced the GSKs and, as a result, slightly affected the accuracy of the calculated zonal PTDFs.

Finally, units smaller than 20 MW were excluded to simplify the dataset. Small plants contribute little to the total generation. while this somewhat reduces representativeness, it likely makes the model more robust by reducing noise and uncertainty.

## 7.3 Implications for Flow-Based Market Modeling

The results of this study are directly relevant for flow-based market modeling in the Nordic electricity market. Accurate PTDFs are essential for calculating available transmission capacities and for understanding how changes in generation and demand affect prices. Even small inaccuracies in PTDFs can lead to differences in price outcomes, capacity allocation, and congestion patterns.

For market participants such as Vattenfall, using data-driven GSKs would most likely improve the accuracy of internal models and short-term price forecasts. By basing the GSKs on real generation data, the PTDFs better reflect actual system behavior and network constraints. Even relatively simple improvements such as updating GSKs monthly using actual production data could noticeably enhance model precision and predictability.

However, several challenges remain. The Nordic system is large and complex, and simplified DC models cannot fully capture all network dynamics or nonlinear interactions between zones. In addition, anonymized network data, particularly in Sweden, make it difficult to verify or link generators precisely to their corresponding nodes in the grid. This lack of transparency limits how accurately power flow models can be validated and calibrated using open data sources.

## 7.4 Future Research

Future research could continue developing the data-driven GSK framework in several ways to improve both accuracy and usability.

One important step would be to include demand-side flexibility in the model. In this master's thesis, the GSKs only represent how generation changes, but in reality, demand also varies and affects the power balance in each bidding zone. By including flexible demand, especially from large industries or controllable loads, the model would better capture how both generation and consumption respond to market changes. This would also make the approach more consistent with the operational GSK 6 strategy currently applied in most bidding zones in the Nordic day-ahead market.

Another interesting direction would be to test several different GSK strategies, rather than focusing only on GSK 6. Since GSK 6 is the standard approach for most Nordic countries, it would be valuable to examine whether it actually provides the best performance for all bidding zones or if alternative strategies could be better suited in certain areas. Comparing different strategies could show whether specific zones benefit more from generation-based, load-based, or mixed approaches.

A third improvement would be to increase the precision of how power plants and nodes are located in the grid model. In this study, some nodes were identified only through approximate coordinate matching, which can make it difficult to assign the correct GSK value to the right grid node. By improving geographic data and linking each generator more accurately to its corresponding node, future models could assign weights with higher confidence and reduce the risk of mismatches. This would make the calculated PTDFs more representative and further improve the overall reliability of the results.

# 8 Conclusions

This master's thesis evaluates how different GSK models affect the estimation of zonal PTDFs in the Nordic day-ahead electricity market. The analysis focused on two main research questions:

1. How do data-driven GSKs derived from actual generation data affect the accuracy of zonal PTDFs compared to a flat reference strategy?
2. How does the choice of time resolution of data-driven GSKs influence the trade-off between accuracy and stability in the resulting zonal PTDFs?

Regarding the first research question, the results show that data-driven GSKs improve the accuracy of PTDF estimation compared to the flat reference strategy. All data-driven configurations achieved lower RMSE values than the flat case, confirming that incorporating real generation data leads to a more realistic representation of how power flows respond to zonal net position changes. This improvement demonstrates that generation patterns carry essential information about grid sensitivities that a uniform distribution cannot capture.

The second research question explored how the time resolution of GSKs affects accuracy and stability. The results revealed that while higher time resolution allows GSKs to follow short-term variations in generation, it does not necessarily improve overall accuracy. In fact, the daily GSKs introduced more noise and instability. The monthly GSKs, on the other hand, provided the most robust and consistent performance, offering the lowest RMSE values and best reflecting average system behavior. The weekly GSKs served as a useful middle ground, capturing operational dynamics without excessive sensitivity.

The study also highlighted several methodological factors that influence the reliability of PTDF estimation. Incomplete generation data, uncertainties in coordinate matching, and simplifications in plant classification can affect the allocation of GSK values. In some cases, non-participating generators such as nuclear or wind power plants may have been incorrectly assigned a GSK, which could slightly distort zonal sensitivities. Addressing these limitations through better data validation, improved coordinate accuracy, and the inclusion of demand-side participation would further enhance the realism of data-driven GSKs.

In conclusion, data-driven GSKs represent a step toward more accurate and transparent flow-based market modeling in the Nordic region. By basing GSKs on real generation data and selecting an appropriate time resolution, market participants such as Vattenfall can improve internal models for price forecasting and capacity calculation. Future work should continue to refine the GSK framework with integrating load participation, testing additional GSK strategies, and improving coordinate precision to ensure that flow-based modeling reflects the physical realities of the Nordic power system as closely as possible.

# References


[1] I. M. Ydersbond, “Aiming to be Environmental Leaders, but Struggling to go Forward: Sweden and Norway on Energy System Transformation”, *Energy Procedia*, vol. 58, pp. 16–23, 2014. DOI: https://doi.org/10.1016/j.egypro.2014.10.403.

[2] P. Malinina and N.-H. M. Von Der Fehr, “Flow-Based Market Coupling and Renewable Energy Integration in Nordic”, in *2025 21st International Conference on the European Energy Market (EEM)*, 2025, pp. 1–6. DOI: https://doi.org/10.1109/EEM64765.2025.11050288.

[3] Svenska kraftnät. “Balansering av kraftsystemet”. (n.d), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-systemansvaret/balansering-av-kraftsystemet/. [Accessed: 2025-11-11].

[4] D. Schönheit, M. Kenis, L. Lorenz, D. Möst, E. Delarue, and K. Bruninx, “Toward a fundamental understanding of flow-based market coupling for cross-border electricity trading”, *Advances in Applied Energy*, vol. 2, 2021. DOI: https://doi.org/10.1016/j.adapen.2021.100027.

[5] Svenska kraftnät. “Kapacitetsberäkning”. (2024), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/kapacitetsberakning/. [Accessed: 2025-01-27].

[6] Nordic Regional Coordination Centre, *Supporting document for the second amendment of 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*, n.d. [Online]. Available: https://nordic-rsc.net/wp-content/uploads/2018/10/Stakeholder-consultation-document-and-Impact-Assessment-for-the-Capacity-Calculation-Methodology-Proposal-for-the-Nordic-CCR.pdf, [Accessed: 2025-01-27].

[7] A. Seleznev and M. Dolmatova, “Modeling and Predicting Constraints in Nordic Flow-Based Market Using Open Data”, in *2024 20th International Conference on the European Energy Market (EEM)*, (Istanbul, Turkey), 2024, pp. 1–6. DOI: https://doi.org/10.1109/EEM60825.2024.10608916.

[8] Nordic Regional Coordination Centre, *GSK Methodology Elaboration*, 2024. [Online]. Available: https://nordic-rcc.net/wp-content/uploads/2024/12/GSK-Methodology-Elaboration.pdf, [Accessed: 2025-02-03].

[9] P. Högselius and A. Kaijser, “The politics of electricity deregulation in Sweden: The art of acting on multiple arenas”, *Energy Policy*, vol. 38, no. 5, pp. 2245–2254, 2010. DOI: https://doi.org/10.1016/j.enpol.2009.12.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421509009501.

[10] Svenska kraftnät. “Om elmarknaden”. (2024), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/. [Accessed: 2025-01-23].

[11] P. Falbo, D. Felletti, and S. Stefani, “Electricity Futures”, in *The World Scientific Handbook of Futures Markets*. World Scientific Publishing Company, 2015, ch. 18, pp. 545–565. DOI: 10.1142/9789814566926_0018.

[12] Svenska kraftnät. “Dagen före-marknaden - fysisk handel med el”. (2023), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/dagen-fore-marknaden--fysisk-handel-med-el/. [Accessed: 2025-01-24].

[13] Svenska kraftnät. “Intradagsmarknaden - justering av dagen före-handel”. (2024), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/intradagsmarknaden--justering-av-dagen-fore-handel/. [Accessed: 2025-01-24].

[14] Svenska kraftnät. “Balansmarknaden”. (2022), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/balansmarknaden/. [Accessed: 2025-01-24].

[15] A. S. Kumar, I. Kouveliotis-Lysikatos, E. Nycander, *et al.*, “Open Nodal Power Flow Model of the Nordic Power System”, in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6. DOI: https://doi.org/10.1109/PowerTech46648.2021.9494886.

[16] Svenska kraftnät. “Elområden”. (2022), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/. [Accessed: 2025-01-23].

[17] ENTSO-E. “Capacity Calculation Regions”. (n.d), [Online]. Available: https://www.entsoe.eu/network_codes/ccr-regions/. [Accessed: 2025-10-15].

[18] Nordic Regional Coordination Centre, *External Parallell Run*, 2023. [Online]. Available: https://nordic-rcc.net/wp-content/uploads/2023/03/CCM-EPR-handbook_V3.pdf, [Accessed: 2025-02-12].

[19] ENTSO-E. “Single Day-ahead Coupling”. (n.d), [Online]. Available: https://www.entsoe.eu/network_codes/cacm/implementation/sdac/#How%20Does%20the%20Single%20Day-Ahead%20Coupling%20Work%20in%20Detail?. [Accessed: 2025-11-11].

[20] Svenska kraftnät. “Flaskhalsinkomster”. (2024), [Online]. Available: https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/om-flaskhalsinkomster/. [Accessed: 2025-01-27].

[21] Nordic Regional Coordination Centre, *Methodology and concepts for the Nordic Flow-Based Market Coupling Approach*, n.d. [Online]. Available: https://www.fingrid.fi/globalassets/dokumentit/fi/tiedotteet/sahkomarkkinat/2015/methodology-and-concepts-for-the-nordic-flow-based-market-coupling-approach.pdf, [Accessed: 2025-02-06].

[22] Nord Pool. “Day-ahead market”. (n.d), [Online]. Available: https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/. [Accessed: 2025-01-27].

[23] All NEMO Committee, *EUPHEMIA Public Description: Single Price Coupling Algorithm*, 2024. [Online]. Available: https://www.nemo-committee.eu/assets/files/euphemia-public-description-2024.pdf, [Accessed: 2025-11-10].

[24] M. Ovaere, M. Kenis, K. Van den Bergh, K. Bruninx, and E. Delarue, “The effect of flow-based market coupling on cross-border exchange volumes and price convergence in Central Western European electricity markets”, *Energy Economics*, vol. 118, 2023. DOI: https://doi.org/10.1016/j.eneco.2023.106519.

[25] ENTSO-E, *Net Transfer Capacities (NTC) and Available Transfer Capacities (ATC) in the Internal Market of Elevtricity in Europe (IEM)*, 2020. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/pre2015/ntc/entsoe_NTCusersInformation.pdf, [Accessed: 2025-02-20].

[26] Nordic Regional Coordination Centre, *External parallel run evaluation report - For assessment by the NRAs of the Nordic CCR, as required by the Nordic DA/ID CCM*, 2024. [Online]. Available: https://nordic-rcc.net/wp-content/uploads/2023/06/Parallel-run-report_final_public.pdf, [Accessed: 2025-11-10].

[27] L. Wyrwoll, M. Nobis, S. Raths, and A. Moser, “Evolution of Fundamental Price Determination within Electricity Market Simulations”, *Energies*, vol. 14, no. 17, pp. 1–17, 2021. DOI: https://doi.org/10.3390/en14175454.

[28] Nordic Regional Coordination Centre. “Nordic Regional Coordination Centre”. (n.d), [Online]. Available: https://nordic-rcc.net/. [Accessed: 2025-02-01].

[29] H. Engsner, *Interview on Vattenfall’s planning and price forecasting process*, 2025.

[30] O. Nilsson and R. B. Bernes, *Challenges of FBMC implementation from hydro-scheduling perspective*, 2024. [Online]. Available: https://nordic-rcc.net/wp-content/uploads/2024/09/06-Vattenfall_FBMC_Stakeholder_Meeting.pdf.

[31] B. Xiong, D. Fioriti, F. Neumann, I. Riepin, and T. Brown, “Modelling the high-voltage grid using open data for Europe and beyond”, *Scientific data*, vol. 12, no. 1, pp. 277–, 2025. DOI: https://doi.org/10.1038/s41597-025-04550-7.

[32] Nordic Regional Coordination Centre, *Proposal for a Capacity Calculation Methodology for the balancing timeframe for the Nordic Capacity Calculation Region in accordance with Article 37(3) of Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing*, 2023. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/nc-tasks/230816_Art%2037(3)_EB%20Regulation_Nordic%20CCR_BT%20CCM%20-%20amended%20legal%20proposal_Vf.pdf, [Accessed: 2025-10-23].

[33] Joint Allocation Office, *Publication handbook*, 2024. [Online]. Available: https://publicationtool.jao.eu/PublicationHandbook/Nordic_PublicationTool_Handbook_v1.2.pdf, [Accessed: 2025-02-13].

[34] ENTSO-E. "ENTSOE-E Transparency Platform". (n.d), [Online]. Available: https://m-transparency.entsoe.eu/?site_preference=mobile. [Accessed: 2025-08-29].

[35] OpenStreetMap. "About". (n.d), [Online]. Available: https://www.openstreetmap.org/about. [Accessed: 2025-09-03].

[36] Overpass Turbo. "Overpass Turbo". (n.d), [Online]. Available: https://overpass-turbo.eu/. [Accessed: 2025-09-03].

[37] PyPSA. "PyPSA Website". (n.d), [Online]. Available: https://pypsa.org/. [Accessed: 2025-09-10].

[38] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not", *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022. DOI: https://doi.org/10.5194/gmd-15-5481-2022.

# A GSK Values for Flat Strategy

*Table 5: GSK values for the flat reference case. Each value represent the value liked to the nodes in each bidding zone. For example, in FI, all generation units received a value of 0.0169.*

| BZ | GSK value |
|---|---|
| DK2 | 0.0101 |
| DK2_KO | 1.0000 |
| DK2_SB | 1.0000 |
| FI | 0.0169 |
| FI_EL | 0.5000 |
| FI_FS | 1.0000 |
| NO1 | 0.0357 |
| NO2 | 0.0182 |
| NO2_ND | 1.0000 |
| NO2_NK | 1.0000 |
| NO2_SK | 1.0000 |
| NO3 | 0.0400 |
| NO4 | 0.0526 |
| NO5 | 0.0294 |
| SE1 | 0.0500 |
| SE2 | 0.0161 |
| SE3 | 0.0147 |
| SE3_FS | 1.0000 |
| SE3_KS | 1.0000 |
| SE3_SWL | 1.0000 |
| SE4 | 0.0588 |
| SE4_BC | 1.0000 |
| SE4_NB | 1.0000 |
| SE4_SP | 1.0000 |
| SE4_SWL | 1.0000 |

# B GSK Values for the Monthly Model

*Table 6: GSK values for the monthly GSK model.*

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.015374 | 0.035243 | 0.060300 | 0.031681 | 0.016864 | 0.056279 | 0.014299 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.060300 | 0.031681 | 0.016864 | 0.056279 | 0.032190 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.060300 | 0.031681 | 0.016864 | 0.056279 | 0.005112 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.060300 | 0.031681 | 0.016864 | 0.056279 | 0.013530 | 0.002946 | 0.000351 |
| 0.015374 | 0.035243 | 0.018645 | 0.031681 | 0.016864 | 0.056279 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.000381 | 0.031681 | 0.016864 | 0.056279 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.060300 | 0.031681 | 0.016864 | 0.058720 | 0.007737 | 0.012965 | 0.000351 |
| 0.015374 | 0.053422 | 0.060300 | 0.031681 | 0.016864 | 0.016545 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.011310 | 0.031681 | 0.016864 | 0.056279 | 0.013530 | 0.018280 | 0.000351 |
| 0.015374 | 0.049536 | 0.135846 | 0.031681 | 0.016864 | 0.056279 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.011310 | 0.031681 | 0.016864 | 0.073531 | 0.013530 | 0.014185 | 0.000351 |
| 0.015374 | 0.035243 | 0.051564 | 0.026799 | 0.016864 | 0.101346 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.021537 | 0.011310 | 0.031681 | 0.005103 | 0.018204 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.051564 | 0.001305 | 0.016864 | 0.172077 | 0.010445 | 0.012965 | 0.994383 |
| 0.015374 | 0.035243 | 0.029879 | 0.031681 | 0.016864 | 0.038038 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.075381 | 0.032113 | 0.060943 | 0.016864 | 0.056279 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.034885 | 0.011506 | 0.016864 | 0.007752 | 0.013530 | 0.012965 | 0.000351 |
| 0.015374 | 0.035243 | 0.059758 | 0.007049 | 0.016864 | 0.007274 | 0.013530 | 0.012965 | |
| 0.015374 | 0.035243 | 0.189635 | 0.000000 | 0.016864 | | 0.008289 | 0.012965 | |
| 0.015374 | 0.062459 | | 0.002883 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | 0.062459 | | 0.031681 | 0.085079 | | 0.013530 | 0.012965 | |
| 0.015374 | 0.040838 | | 0.029372 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.008630 | 0.035243 | | 0.023498 | 0.016864 | | 0.013530 | 0.006765 | |
| 0.015374 | 0.035243 | | 0.006104 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.008121 | 0.035243 | | 0.006577 | 0.016864 | | 0.022105 | 0.008183 | |
| 0.003894 | | | 0.010689 | 0.000000 | | 0.013530 | 0.012965 | |
| 0.015374 | | | 0.100711 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.004352 | | | 0.031681 | 0.016864 | | 0.015833 | 0.007747 | |
| 0.015374 | | | 0.031681 | 0.002089 | | 0.013530 | 0.012965 | |
| 0.015374 | | | 0.031681 | 0.019088 | | 0.013530 | 0.012965 | |
| 0.007804 | | | 0.031681 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | 0.010622 | 0.016864 | | 0.002901 | 0.012965 | |
| 0.082080 | | | 0.114056 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015857 | | | 0.017623 | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.033909 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.001782 | | 0.006286 | 0.012965 | |

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.009910 | | | | 0.006063 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.010914 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.000000 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.013530 | 0.012965 | |
| 0.015374 | | | | 0.000000 | | 0.011812 | 0.012965 | |
| 0.027537 | | | | 0.020885 | | 0.009947 | 0.012965 | |
| 0.013828 | | | | 0.020885 | | 0.021760 | 0.012965 | |
| 0.021641 | | | | 0.016864 | | 0.009146 | 0.012965 | |
| 0.021803 | | | | 0.016864 | | 0.020585 | 0.012965 | |
| 0.015374 | | | | 0.018225 | | 0.013816 | 0.012965 | |
| 0.009411 | | | | 0.018225 | | 0.016579 | 0.012965 | |
| 0.043183 | | | | 0.002134 | | 0.057390 | 0.012187 | |
| 0.095808 | | | | 0.016864 | | 0.034318 | 0.012965 | |
| 0.015374 | | | | 0.016864 | | 0.013530 | 0.012965 | |
| 0.016962 | | | | 0.048602 | | 0.013387 | 0.012965 | |
| 0.024970 | | | | 0.002004 | | 0.025973 | 0.012965 | |
| | | | | 0.016864 | | 0.021829 | 0.009493 | |
| | | | | 0.016864 | | 0.005526 | 0.012965 | |
| | | | | 0.024518 | | 0.023625 | 0.003302 | |
| | | | | 0.016864 | | 0.008980 | 0.070130 | |
| | | | | | | 0.020392 | 0.000000 | |
| | | | | | | 0.018237 | 0.070905 | |
| | | | | | | 0.054424 | 0.011457 | |
| | | | | | | 0.016150 | 0.000000 | |
| | | | | | | 0.013530 | 0.000000 | |
| | | | | | | | 0.012188 | |
| | | | | | | | 0.021108 | |
| | | | | | | | 0.072013 | |
| | | | | | | | 0.023820 | |

# C GSK Values for the Weekly Model

*Table 7: GSK values for the weekly GSK model.*

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.016303 | 0.035536 | 0.056830 | 0.031089 | 0.016864 | 0.053741 | 0.014355 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.056830 | 0.031089 | 0.016864 | 0.053741 | 0.032315 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.056830 | 0.031089 | 0.016864 | 0.053741 | 0.005132 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.056830 | 0.031089 | 0.016864 | 0.053741 | 0.013342 | 0.003231 | 0.000000 |
| 0.016303 | 0.035536 | 0.000354 | 0.031089 | 0.016864 | 0.053741 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.000354 | 0.031089 | 0.016864 | 0.053741 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.056830 | 0.031089 | 0.016864 | 0.043001 | 0.007767 | 0.013195 | 0.000000 |
| 0.016303 | 0.052777 | 0.056830 | 0.031089 | 0.016864 | 0.031798 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.009507 | 0.031089 | 0.016864 | 0.053741 | 0.013342 | 0.018267 | 0.000000 |
| 0.016303 | 0.048938 | 0.144979 | 0.031089 | 0.016864 | 0.053741 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.009507 | 0.031089 | 0.016864 | 0.053847 | 0.013342 | 0.015555 | 0.000000 |
| 0.016303 | 0.035536 | 0.048597 | 0.025343 | 0.016864 | 0.117432 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.021277 | 0.009507 | 0.031089 | 0.005103 | 0.020068 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.048597 | 0.004818 | 0.016864 | 0.194989 | 0.010485 | 0.013195 | 1.000000 |
| 0.016303 | 0.035536 | 0.028160 | 0.031089 | 0.016864 | 0.046096 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.074470 | 0.031763 | 0.050429 | 0.016864 | 0.053741 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.073438 | 0.012005 | 0.016864 | 0.004480 | 0.013342 | 0.013195 | 0.000000 |
| 0.016303 | 0.035536 | 0.056319 | 0.007355 | 0.016864 | 0.004623 | 0.013342 | 0.013195 | |
| 0.016303 | 0.035536 | 0.197940 | 0.000000 | 0.016864 | | 0.008321 | 0.013195 | |
| 0.016303 | 0.061704 | | 0.004513 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | 0.061704 | | 0.031089 | 0.085079 | | 0.013342 | 0.013195 | |
| 0.016303 | 0.039477 | | 0.030645 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.000000 | 0.035536 | | 0.024516 | 0.016864 | | 0.013342 | 0.007419 | |
| 0.016303 | 0.035536 | | 0.006369 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.010289 | 0.035536 | | 0.006862 | 0.016864 | | 0.022191 | 0.008974 | |
| 0.004218 | | | 0.011152 | 0.000000 | | 0.013342 | 0.013195 | |
| 0.016303 | | | 0.115304 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.004647 | | | 0.031089 | 0.016864 | | 0.015894 | 0.008495 | |
| 0.016303 | | | 0.031089 | 0.002089 | | 0.013342 | 0.013195 | |
| 0.016303 | | | 0.031089 | 0.019088 | | 0.013342 | 0.013195 | |
| 0.006192 | | | 0.031089 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | 0.027360 | 0.016864 | | 0.002913 | 0.013195 | |
| 0.043688 | | | 0.095335 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016931 | | | 0.018387 | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.033909 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.001782 | | 0.006310 | 0.013195 | |

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.010582 | | | | 0.006063 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.010957 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.000000 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.013342 | 0.013195 | |
| 0.016303 | | | | 0.000000 | | 0.011858 | 0.013195 | |
| 0.024589 | | | | 0.020885 | | 0.009986 | 0.013195 | |
| 0.014775 | | | | 0.020885 | | 0.021844 | 0.013195 | |
| 0.025539 | | | | 0.016864 | | 0.009181 | 0.013195 | |
| 0.023281 | | | | 0.016864 | | 0.020665 | 0.013195 | |
| 0.016303 | | | | 0.018225 | | 0.013869 | 0.013195 | |
| 0.010397 | | | | 0.018225 | | 0.016643 | 0.013195 | |
| 0.039229 | | | | 0.002134 | | 0.057612 | 0.012178 | |
| 0.097097 | | | | 0.016864 | | 0.034451 | 0.013195 | |
| 0.016303 | | | | 0.016864 | | 0.013342 | 0.013195 | |
| 0.026019 | | | | 0.048602 | | 0.013439 | 0.013195 | |
| 0.023002 | | | | 0.002004 | | 0.026074 | 0.013195 | |
| | | | | 0.016864 | | 0.021913 | 0.010410 | |
| | | | | 0.016864 | | 0.005548 | 0.013195 | |
| | | | | 0.024518 | | 0.023716 | 0.003278 | |
| | | | | 0.016864 | | 0.009015 | 0.070080 | |
| | | | | | | 0.020471 | 0.000000 | |
| | | | | | | 0.018307 | 0.070855 | |
| | | | | | | 0.058928 | 0.012564 | |
| | | | | | | 0.016213 | 0.000000 | |
| | | | | | | 0.013342 | 0.000000 | |
| | | | | | | | 0.000000 | |
| | | | | | | | 0.016355 | |
| | | | | | | | 0.071962 | |
| | | | | | | | 0.023803 | |

# D GSK Values for the Daily Model

*Table 8: GSK values for the daily GSK model.*

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.015546 | 0.036741 | 0.058513 | 0.031668 | 0.016864 | 0.062353 | 0.014408 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.058513 | 0.031668 | 0.016864 | 0.062353 | 0.032435 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.058513 | 0.031668 | 0.016864 | 0.062353 | 0.005151 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.058513 | 0.031668 | 0.016864 | 0.062353 | 0.013341 | 0.003170 | 0.000000 |
| 0.015546 | 0.036741 | 0.000389 | 0.031668 | 0.016864 | 0.062353 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.000389 | 0.031668 | 0.016864 | 0.062353 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.058513 | 0.031668 | 0.016864 | 0.102437 | 0.007796 | 0.013001 | 0.000000 |
| 0.015546 | 0.055703 | 0.058513 | 0.031668 | 0.016864 | 0.000000 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.008936 | 0.031668 | 0.016864 | 0.062353 | 0.013341 | 0.018117 | 0.000000 |
| 0.015546 | 0.051651 | 0.148395 | 0.031668 | 0.016864 | 0.062353 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.008936 | 0.031668 | 0.016864 | 0.128274 | 0.013341 | 0.015262 | 0.000000 |
| 0.015546 | 0.036741 | 0.050036 | 0.019360 | 0.016864 | 0.081492 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.022457 | 0.008936 | 0.031668 | 0.005103 | 0.022800 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.050036 | 0.035164 | 0.016864 | 0.070628 | 0.010524 | 0.013001 | 1.000000 |
| 0.015546 | 0.036741 | 0.028994 | 0.031668 | 0.016864 | 0.022244 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.078599 | 0.035313 | 0.035040 | 0.016864 | 0.062353 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.025675 | 0.014683 | 0.016864 | 0.004659 | 0.013341 | 0.013001 | 0.000000 |
| 0.015546 | 0.036741 | 0.057988 | 0.008996 | 0.016864 | 0.006288 | 0.013341 | 0.013001 | |
| 0.015546 | 0.036741 | 0.224900 | 0.000000 | 0.016864 | | 0.008352 | 0.013001 | |
| 0.015546 | 0.065125 | | 0.000000 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | 0.065125 | | 0.031668 | 0.085079 | | 0.013341 | 0.013001 | |
| 0.015546 | 0.000000 | | 0.037482 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.000000 | 0.036741 | | 0.029986 | 0.016864 | | 0.013341 | 0.007279 | |
| 0.015546 | 0.036741 | | 0.007790 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.008709 | 0.036741 | | 0.008393 | 0.016864 | | 0.022273 | 0.008805 | |
| 0.000247 | | | 0.013640 | 0.000000 | | 0.013341 | 0.013001 | |
| 0.015546 | | | 0.111040 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.004553 | | | 0.031668 | 0.016864 | | 0.015953 | 0.008335 | |
| 0.015546 | | | 0.031668 | 0.002089 | | 0.013341 | 0.013001 | |
| 0.015546 | | | 0.031668 | 0.019088 | | 0.013341 | 0.013001 | |
| 0.018345 | | | 0.031668 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | 0.038240 | 0.016864 | | 0.002923 | 0.013001 | |
| 0.038390 | | | 0.047680 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.016588 | | | 0.022489 | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.033909 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.001782 | | 0.006334 | 0.013001 | |

| NO2 | NO3 | NO4 | NO5 | FI | SE1 | SE2 | SE3 | SE4 |
|---|---|---|---|---|---|---|---|---|
| 0.010368 | | | | 0.006063 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.010997 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.000000 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.013341 | 0.013001 | |
| 0.015546 | | | | 0.000000 | | 0.011902 | 0.013001 | |
| 0.004184 | | | | 0.020885 | | 0.010023 | 0.013001 | |
| 0.008843 | | | | 0.020885 | | 0.021925 | 0.013001 | |
| 0.032650 | | | | 0.016864 | | 0.009215 | 0.013001 | |
| 0.022809 | | | | 0.016864 | | 0.020742 | 0.013001 | |
| 0.015546 | | | | 0.018225 | | 0.013921 | 0.013001 | |
| 0.009008 | | | | 0.018225 | | 0.016705 | 0.013001 | |
| 0.055416 | | | | 0.002134 | | 0.057826 | 0.012078 | |
| 0.121478 | | | | 0.016864 | | 0.034579 | 0.013001 | |
| 0.015546 | | | | 0.016864 | | 0.013341 | 0.013001 | |
| 0.033989 | | | | 0.048602 | | 0.013489 | 0.013001 | |
| 0.023683 | | | | 0.002004 | | 0.026171 | 0.013001 | |
| | | | | 0.016864 | | 0.021995 | 0.010214 | |
| | | | | 0.016864 | | 0.005568 | 0.013001 | |
| | | | | 0.024518 | | 0.023804 | 0.003195 | |
| | | | | 0.016864 | | 0.009048 | 0.069502 | |
| | | | | | | 0.020547 | 0.000000 | |
| | | | | | | 0.018375 | 0.070271 | |
| | | | | | | 0.057156 | 0.012327 | |
| | | | | | | 0.016273 | 0.000000 | |
| | | | | | | 0.013341 | 0.000000 | |
| | | | | | | | 0.000000 | |
| | | | | | | | 0.029435 | |
| | | | | | | | 0.071368 | |
| | | | | | | | 0.023607 | |