UPPSALA
UNIVERSITET

UPTEC STS 26031
Examensarbete 30 hp
June 2026

Estimation of daily wind power potential in Sweden’s bidding areas

A machine learning approach for reconstructing the historical technical potential in wind power production

Olivia Lundbergh

Civilingenjörsprogrammet i system i teknik och samhälle

UPPSALA UNIVERSITET

Estimation of daily wind power potential in Sweden's bidding areas: A machine learning approach for reconstructing the historical technical potential in wind power production

Olivia Lundbergh

Abstract

As the share of Variable Renewable Energy (VRE) sources, such as wind, expands globally, its fluctuating nature introduces inherent volatility into power systems. In Sweden, wind production for 2025 reached 39 TWh, accounting for 24.3% of total national generation. Low VRE marginal costs depress electricity prices during high production, while transmission constraints across Sweden's four bidding areas cause regional price divergences. In response, operators may curtail production during non-profitable periods. Quantifying this unrealized potential is vital for understanding grid inefficiencies and optimizing system operations to reach zero-emission climate goals. This study proposes a novel machine learning framework designed to reconstruct the historical technical potential of wind power generation independent of electricity market dynamics.

Applied to the aggregated production data for Sweden's bidding areas, the methodology utilizes region-specific Histogram-based Gradient Boosting Regression (HGBR) models tuned with the Bayesian optimization framework Optuna, with meteorological features dimensionally reduced using Principal Component Analysis. The machine learning models were trained exclusively on data from profitable hours (>3 €/MWh) to prevent the algorithms from learning price driven curtailment, isolating the relationships between meteorological variables, technical availability, and production. The models were then deployed to predict the unconstrained technical production capacity across all hours, including previously unseen non-profitable hours. Subsequently, these predictions were refined using a three constraint transformation approach combining model outputs with actual production data to estimate the true technical potential.

To validate the price independence of the reconstructed technical potential, secondary validation models were trained on the reconstructed datasets and evaluated against baseline configurations, across feature sets that either included or omitted price variables. The evaluation used MAE, RMSE, and R^2 metrics to verify whether price data was required to maintain predictive accuracy across both model groups. Furthermore, the SHapley Additive eXplanations (SHAP) framework was leveraged to interpret model behaviour and how price dynamics impacted performance.

The empirical results reveal that an estimated 2.90 TWh of technically feasible energy was curtailed in Sweden during 2025. Notably, 87.13% of this volume (2.53 TWh) occurred when Day-Ahead prices fell below 3 €/MWh, confirming that market price highly influence wind asset operations and highlighting future integration challenges. Regional imbalances were apparent, SE2 alone accounted for 64.77% of total curtailment, followed by SE1 at 23.16%, while the southern areas exhibited lower curtailment volumes. This extensive deviation underscores the need for targeted grid investment in transmission capacity and smart energy storage solutions.

Teknisk-naturvetenskapliga fakulteten

Uppsala universitet, Utgivningsort Uppsala/Visby

Handledare: Jonatan Westholm Ämnesgranskare: Yuan Yao

Examinator: Elísabeth Andrésdóttir

Populärvetenskaplig sammanfattning

Världen står inför stora utmaningar kopplade till klimatmål och energiproduktion. I enlighet med Parisavtalet och internationella direktiv är målsättningen att nå nettonollutsläpp av växthusgaser. I Sverige finns det liknande mål om att ha en 100% fossilfri elproduktion år 2040. Parallellt med detta klimatarbete visar prognoser att elbehovet kommer att öka drastiskt under de kommande årtiondena. Mellan 2025 och 2045 förväntas denna ökning motsvara uppemot 140 terawattimmar (TWh), vilket nästan är en dubblering jämfört med 2025 års produktion. Denna enorma ökning kräver omfattande investeringar i elsystemet och i ny fossilfri elproduktion.

Vindkraft förväntas spela en nyckelroll i denna omställning och globalt uppskattas vindkraft nå produktionsnivåer nära 4500 TWh till år 2030. Vindkraften i Sverige stod för ungefär en fjärdedel av den nationella totala elproduktionen under 2025. Den svenska vindkraftsproduktionen är dock ojämnt fördelad över landet, där mer än hälften av produktionen under 2025 (drygt 60%) härstammade från områden i norra Sverige.

Att införa mer väderberoende elproduktion innebär dock stora utmaningar för elnätets stabilitet och drift, eftersom produktionen är helt beroende av väderförhållanden. Vidare påverkar vindkraften energimarknaderna, vars prisbildning styrs av tillgänglig elproduktion och det aktuella elbehovet. Eftersom vindkraften är en billig produktionskälla med låga driftskostnader, pressar den ner elpriserna kraftigt under perioder när hög vindkraftsproduktion sammanfaller med ett lågt elbehov. När priserna faller under vad det kostar att producera, eller till och med blir negativa, händer det att vindkraftsoperatörer stänger av sina turbiner för att undvika ekonomisk förlust, vilket kallas för marknadsreglerad nedreglering. Detta är en av anledningarna till att man ibland kan se vindkraftverk stå helt stilla, trots att det blåser för fullt! I praktiken innebär detta att vi går miste om ren energi på grund av marknadsmekanismer och begränsningar i elnätet.

I samarbete med Svenska kraftnät, den myndighet som ansvarar för driften av det svenska elnätet, har detta arbete utvecklat en metod för att uppskatta vindkraftens faktiska, historiska potential baserad enbart på väder och vindkraftverkens tekniska tillgänglighet. Teknisk tillgänglighet beskriver huruvida vindkraftverken är fungerande eller på grund av underhåll eller tekniska fel inte är i bruk. Målet har varit att rensa bort elmarknadens påverkan för att se den sanna potentialen, med andra ord vad som maximalt hade varit möjligt att producera givet rådande väder.

Detta har utförts med hjälp av maskininlärning, ett område inom Artificiell Intelligens (AI) där algoritmer används för att lära sig mönster i data för att kunna göra förutsägelser. Man skulle kunna likna AI-modellen vid en digital expert som har lärt sig exakt hur mycket el som bör produceras vid en viss vindhastighet. Metoden bygger på att en avancerad AI-modell tränats för varje elområde i Sverige baserat på data från 2025.

Det unika med metoden är att modellerna enbart fått lära sig sambandet mellan den totala elproduktionen, väder och tillgänglighet under de timmar då elpriset översteg 3 €/MWh. Genom att använda dessa prisfilter lär sig modellen hur mycket vindkraftverken producerar under lönsamma förhållanden, då utgångspunkten är att vindparksoperatörer under dessa timmar inte bromsar eller drar ner sin produktion.

Modellerna har sedan använts för att förutsäga produktionen för samtliga timmar under 2025. Eftersom modellen saknar information om elpriser, och enbart vet hur produktionen ser ut under lönsamma förhållanden, kan den räkna ut hur mycket el som *borde* ha producerats under de timmar då priserna i verkligheten varit låga. Genom att kombinera dessa förutsägelser med den faktiska produktionen har en slutgiltig rekonstruktion av den tekniska vindkraftspotentialen för respektive elområde kunnat återskapas. För att säkerställa att elprisets påverkan hade eliminerats från konstruktionerna, utvärderades metoden vidare med hjälp av nya kontrollmodeller, statistiska mätetal samt visuella analyser.

Resultatet visar att det är möjligt att använda maskininlärning för att återskapa vindkraftens dolda potential. De rekonstruerade potentialerna konstaterar att stora volymer av grön vindkraftsproduktion gick förlorad under 2025 på grund av marknadens påverkan. Totalt uppgår denna förlorade volym till hela 2.90 TWh. Om denna energi hade tagits tillvara i nätet hade det motsvarat en ökning på 7% av den svenska vindkraftsproduktionen för år 2025. Av denna el härstammade nästan 90% från de norra elområdena SE1 och SE2.

Sammanfattningsvis understryker resultatet ett tydligt behov av djupare studier och konkreta åtgärder för att nå klimatmålen och skapa ett mer effektivt elnät. För att kunna ta tillvara på denna förlorade, nedreglerade potential krävs stora investeringar i elnätet, men också smarta energilösningar som storskalig batterilagring.

Nyckelord

Vindkraft, prognoser, elmarknad, elområden, maskininlärning, optimering

Acknowledgements

I want to express my deepest gratitude to everyone who has supported and guided me throughout this project. First, I want to thank everyone at Svenska Kraftnät for generously sharing their time, valuable insights and expertise, with a special mention to the Data Analytics team.

I would like to extend a special thank you to my supervisor, Jonatan Westholm, for his invaluable guidance, encouragement and all the hours of brainstorming that made this work possible. Another thanks goes to my subject reviewer, Yuan Yao, for his time in reviewing this project.

Finally, I want to thank my friends and family for standing by me during this project and throughout my five years at Uppsala University. Without you, none of this would have been possible.

Uppsala, July 2026

Olivia Lundbergh

Table of contents

List of abbreviations

AI	Artificial Intelligence
FCR	Frequency Containment Reserve
FFR	Fast Frequency Reserve
FRR	Frequency Restoration Reserve
HGBR	Histogram-based Gradient Boosting Regression
ML	Machine Learning
MTU	Market Time Unit
NWP	Numerical Weather Prediction
PCA	Principal Component Analysis
PPA	Power Purchase Agreement
SHAP	SHapley Additive exPlanations
Svk	Svenska kraftnät
TSO	Transmission System Operator
UMM	Urgent Market Message
VRE	Variable Renewable Energy
WPF	Wind Power Forecasting

1. Introduction

The Swedish power system is monitored and operated by the state-owned enterprise Svenska kraftnät (Svk). Some of their core responsibilities include the development and maintenance of the Swedish national electrical transmission system, as well as ensuring constant balance between the production and consumption of electricity within the country (Svenska kraftnät, 2023).

Today approximately 98% of Sweden's total energy production is fossil free, and about half of it originates from renewable energy sources such as hydropower and wind (Regeringskansliet, 2026). Wind power is growing increasingly important for meeting the power consumption demands, and accounted for about 25% of the country's total energy production in 2024 (Svenska kraftnät, 2026a). It is a highly intermittent source of energy, as the production depends on external and often uncontrollable factors, such as the weather. This leads to heavy fluctuation of wind power generation.

It is however important to note that wind power is not solely a passive energy source, but is increasingly becoming more involved in the Swedish ancillary service market. There are several different types of ancillary services, and their purpose is to help maintain grid's frequency at 50 Hertz (Hz) and ensure stability through either downregulation or increasing production when necessary. By providing these services wind power production contributes to uphold the stability of the electrical system. New reports suggest that wind power will be even more prominent on this market in the future (Green Power Sweden, 2023).

Another notable shift is that wind power is becoming increasingly price sensitive, with actual production now more closely linked to and influenced by the dynamics of the electricity trading market. As a consequence, wind turbines may not necessarily operate whenever it is windy. Due to this fluctuating and complex nature it is important for Transmission System Operators (TSOs), such as Svenska kraftnät, to have accurate and reliable forecasts to efficiently and safely operate the transmission systems (Shafiullah et al., 2013).

1.1 Motivation

The motivation for this thesis is that its result will hold both scientific and societal relevance. For Svk, estimates of the technical potential of wind power constitute an important strategic input, as it provides indications on how much additional wind power could be realised by increased transmission capacity. Such estimates also help to identify the volumes of unused energy (TWh) that could be utilized through flexibility solutions, for example battery storage. Although this potential often gets overlooked because it is not directly observable, highlighting it is essential for both system operators, industry actors and other market participants.

Sweden presents as a particularly suitable case for this study. There is a high establishment and expansion of intermittent energy within the country, combined with a highly liberalized power market. The existence of multiple balancing markets to participate in combined with significant internal transmission constraints within the countries four bidding areas, increases the price elasticity of supply of wind power on the day-ahead market.

As other countries move toward similar system conditions and mechanisms, understanding how these factors interact and behave in Sweden provides insights of a wider international relevance. Studying the Swedish case therefore enables an improved understanding of how large-scale integration of intermittent energy, such as wind and also solar power, may affect usage patterns in future power systems.

The method employed within this thesis is uniquely designed to isolate and quantify the potential of wind power in the absence of market-driven influences. The framework is structured into four steps, which include training advanced machine learning models, specifically optimized Histogram-based Gradient Boosting regression models, for each bidding area exclusively on economically profitable hours to strictly capture the meteorological and production relationships. Their predictions are then, in combination with the actual production data, used for a specialized reconstruction of the historical potential. For validation of the technical potential a variety of statistical measures and visualization tools are employed in combination with new validation models.

1.1.1 Aim of thesis

As wind power production in Sweden grows and develops into a more dynamic part of the larger electricity system, it becomes harder to actually predict when and where wind power will be produced and enter the power system.

The study's purpose is to provide a deeper understanding of this complex dynamic nature. The project will revolve around developing a method for estimating the maximum technical production of wind power in the electricity areas of Sweden, based exclusively on weather data and technical availability.

1.1.2 Research questions

For this thesis the following research questions have been stated:

- To what extent does the Day-Ahead market price signal correlate with wind power curtailment in Sweden?
- How can machine learning be utilized and validated to estimate the hypothetical maximal potential of wind power production, in the absence of market-based curtailment?
- In what ways do the regional production behaviours differ across the four Swedish bidding areas, when market influences are removed?
- To what degree does the reconstructed technical potential exceed realized production, and what are the implications for the Swedish grid system?

1.2 Delimitations

Certain aspects and concepts are outside the scope of this thesis, and thus, will not be considered nor addressed in paper. These include:

- Production technologies other than wind power.
- Long-term market effects of low prices, such as decreased investments. The focus is strictly on the utilization of existing power assets.
- Production on an individual wind farm level, the work only uses the aggregated total production per bidding area.
- Wind power production behaviours outside of Sweden's bidding areas.
- A distinction between onshore and offshore wind assets.

1.3 Previous work

Extensive research has been conducted on methods for effectively and accurately predicting wind power production, both globally and in Sweden. Today, such forecasts are essential for the safe operation, control and planning of transmission system balance. An in-depth presentation of the topics related to Wind Power Forecasting (WPF) is given in Chapter 5.

However, in previous research the power market is usually omitted as a factor in how wind power production behaves, and rarely discusses the relationship between the two. Additionally, no previous work has focused on specifically trying to estimate the technical maximum possible production exclusively without market influence, which makes this project especially interesting and important.

1.4 Collaborator

This thesis was conducted in collaboration with the Data Analytics unit at Svenska Kraftnät. They are responsible for providing forecasts and analyses for operations, power system planning and the electricity market, with a growing responsibility for the strategy of the system and the enterprise's broad societal accountability.

2. The Swedish electricity system

In this following chapter a presentation and background will be provided of the Swedish electricity system, to give context of the topic which this thesis revolves around. It will include a comprehensive description of how the power grid works, is operated and which technologies constitutes the energy that is produced and consumed within Sweden.

2.1 Infrastructure and governance

A robust and efficient electricity system is a necessity of modern society, for industrial competitiveness, critical infrastructure such as healthcare, and daily life. To maintain a secure power supply, a well established infrastructure is required to transport electricity from producers to users. This infrastructure, the Swedish electricity grid, operates at three hierarchical levels: the transmission grid, the regional distribution networks, and the local distribution networks (Svenska kraftnät, 2024b).

The transmission grid serves as the backbone of the system, operating at high voltages of either 400 or 220 kV. It facilitates the long distance transportation from major producers to the regional distribution nets (Svenska kraftnät, 2024b). This grid stretches across the whole country, and as of 2021 its transmission lines amounted to approximately 16 000 km (Svenska kraftnät, 2021). From the transmission grid level the electricity is stepped down to the regional distribution networks, which further distributes it to local distribution networks for delivery to the end consumers (Svenska kraftnät, 2024b). In addition to the domestic connections, the Swedish electricity system is also highly intertwined with neighbouring Nordic countries, which entails a shared responsibility of balancing the Nordic power system to a frequency of 50 Hz, and the possibility of import and export across borders (Svenska kraftnät, 2025b).

Svenska Kraftnät (Svk) is the Swedish Transmission System Operator (TSO), a government authority tasked with managing the operation, expansion and maintenance of the transmission grid. Furthermore, Svk is responsible for the real time operational security of the power grid, ensuring the transmission capacity remains stable and the continuous balance between production and consumption at all times (Svenska kraftnät, 2021).

2.2 National energy targets and electrification

To understand how the infrastructure is organised and operated today, it is important to touch on the transformation of the Swedish energy landscape over the last couple of decades. The Swedish system has shifted from a historical dependency on oil, to a production mix which is today predominantly fossil free. Driven by political agreements and technological advancements, Sweden is continuously improving energy efficiency in alignment with climate objectives (Statens energimyndighet, 2026).

While the electricity demand within Sweden has historically been relatively constant, the consumption is projected to substantially increase in the coming decades. This increase in consumption comes as a consequence of the large scale electrification of the industrial and transport sectors, as well as the move away from fossil fuels. This development is aligned with the Paris Agreement, the European Union's (EU's) climate regulations and national objectives in Sweden. In 2018, the Swedish government established goals to increase energy efficiency by 50%, compared to 2005, and to achieve a 100% fossil free electricity production by 2040 (Regeringskansliet, 2024).

In 2025, Sweden produced a total of 160 TWh (Svenska kraftnät, 2026a), however strategic planning now estimates a projected demand of 300 TWh by 2045 to ensure the necessary conditions for a carbon neutral future are met (Regeringskansliet, 2024). To meet this surge in demand, the power grid's capacity must be expanded by approximately 8 TWh annually between 2024 and 2045. This expansion will require large-scale investments of both the domestic energy production as well as the transmission capacity of the power grid (Statens energimyndighet, 2026). As the TSO, Svenska Kraftnät plays a central role in the acceleration, and new government issued instructions establish Svk's directive to safely facilitate this expansion and electrification of Sweden (Sveriges riksdag, 2025). Current initiatives include plans to expand the transmission grid with roughly 800 km of new lines, and the refurbishment of 1700 km lines of existing infrastructure between 2021 and 2030 (Svenska kraftnät, 2021).

2.3 The Swedish production mix

Furthermore, a crucial compontent to reach the climate objectives is to increase the domestic production of fossil free energy. Of the 160 TWh produced in 2025, approximately 97% originated from fossil free sources, out of which 42.4% was sourced from hydropower, 24.3% from wind, 27.9% from nuclear and 1.9% from solar power (Svenska kraftnät, 2026a). Notably, a small portion of the production is excluded from the balance settlement due to origin in non-concession networks (Svenska kraftnät, 2025c). The power production in the Swedish power system can be broadly categorized into two groups: weather dependent production and production that can be planned in advance (Svenska kraftnät, 2025j). While the production mix historically has been dominated by hydropower, nuclear power and cogeneration, a significant shift is occurring, as a growing share of energy originates from weather dependent production (Almqvist, 2023). This category is commonly referred to as Variable Renewable Energy (VRE), which includes energy sources that fluctuate based on uncontrollable environmental factors, primarily wind and solar power (Ejuh Che et al., 2025)

While VRE are essential components in reaching net zero emissions by 2040, their intermittent and variable nature has made the power system increasingly complex and poses challenges to the grid stability, due to high production variability and fluctuations (Svenska kraftnät, 2021). This requires strategies such as sophisticated forecasting, energy storage and market policy initiatives for mitigating the inherent drawbacks of an increasingly complex and variable power mix (Ejuh Che et al., 2025), as well as a balanced power with the support of dispatchable energy sources, to ensure a flexible power system for long term operational security (Svenska kraftnät, 2025j; Svenska kraftnät, 2021).

2.4 Bidding areas and price formation

The Swedish electrical grid has since 2011 been divided into four bidding areas (Svenska kraftnät, 2025h). This restructuring was implemented following pressure from the European Commission, which argued that the previous single-zone model in Sweden violated EU competition rules. Under the old system, Sweden often managed internal transmission bottlenecks by limiting export capacity, effectively favouring domestic consumers over those in neighbouring countries (Statens energimyndighet, 2026; Svenska kraftnät, 2025e).

The current four area division: SE1, SE2, SE3 and SE4, as seen in Figure 1, was designed to reflect the physical constraints of the Swedish transmission grid, optimize both domestic and foreign trade, and provide price signals which accurately reflect local supply and demand. Additionally, the boundaries between these bidding areas, which are formally known as cuts, correspond to major structural bottlenecks in the transmission grid. These include Cut 1 (between SE1 and SE2), Cut 2 (between SE2 and SE3) and Cut 4 (between SE3 and SE4) (Svenska kraftnät, 2025e). Through segmenting the market like this, Svk can more effectively manage grid stability and maximize the socioeconomic benefits of the power system (Svenska kraftnät, 2025h).

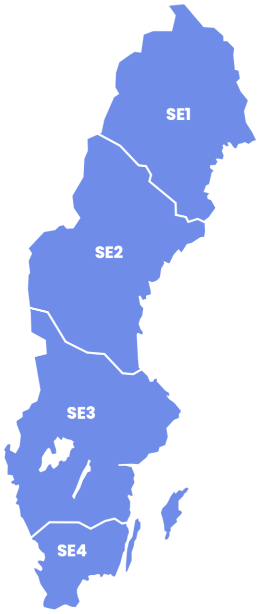


Figure 1: Division of electrical areas in Sweden, image taken from (Svenska kraftnät, 2025h, p. 27)

A fundamental challenge within the Swedish power system is the geographical imbalance between production and consumption. In the northernmost areas, SE1 and SE2, there is consistent generation surplus, characterized by vast hydropower and wind power resources. In contrast, the southern bidding areas, SE3 and SE4, face high consumption and production deficits. This necessitates a high-capacity transmission system to transport electricity from northern Sweden to the south (Statens energimyndighet, 2026; Svenska kraftnät, 2025h). Regional, long term imbalances within a bidding area with low energy self-sustainability introduce vulnerabilities during system disturbances. Addressing this requires substantial investments to enhance internal transmission capacity and improve conditions for import from neighbouring countries (Statens energimyndighet, 2026). Through large scale investment plans, Svk is working toward reducing these bottlenecks over the coming decades, thereby improving overall transmission capacity within the grid (Svenska kraftnät, 2025h).

Electricity price formation is determined by available generation and demand, which both fluctuate across different years, seasons and times of day. These shifting patterns in supply and demand cause high price volatility. For instance the prices are typically substantially lower during the summer months, than in the winter when high demand drives up the prices. Additionally the availability of hydropower and yearly weather patterns impact the prices. Beyond domestic factors, prices are heavily influenced by developments in the broader European market. For instance, the shutdown of German nuclear plants in 2022 increased external demand, which subsequently drove up electricity prices in Sweden (Sandberg, 2024). Consequently, when there is a clear deficit in local generation relative to market demand, prices rise, and vice versa (Svenska kraftnät, 2021).

The pricing mechanism relies on a marginal pricing model, meaning that the equilibrium price is determined by the final and most expensive production unit required to meet demand. When the transmission grid has insufficient capacity to transport low-cost electricity from the northern areas, to the southern consumption centres, the bottleneck limitations cause the market to split, resulting in arising price differences between bidding areas (Energimarknadsinspektionen, 2025; Svenska kraftnät, 2024a). As an increasing volume of VRE enters the power system, the pricing formation and the merit order are significantly impacted. This is due to the extremely low marginal operating costs of VRE, which allows them to supply electricity at low prices. This in consequence means VRE assets are positioned at the base of the merit order curve and are dispatched first. This shifts the supply curve and drives down the equilibrium price, also known as the merit order effect. In turn, this can diminish the competitiveness and utilization of more expensive technologies, such as nuclear and coal, which carry significantly higher operational costs, thus struggling to compete during periods of high VRE output (Zeier, 2026).

In recent years, the electricity market in Sweden has generally maintained some of the lowest prices average prices in the EU, albeit some deviations with significant price spikes and large regional disparities. For instance, while prices fluctuated dramatically during 2022-2023, they stabilized in the following years to reach the lowest levels since 2022 (Statens energimyndighet, 2026). Price deviations between bidding areas have become more pronounced since 2020, characterised by the, at times, very high prices in SE3 and SE4, and significantly lower prices in the northernmost areas (Statens energimyndighet, 2026).

3. The electricity markets of Sweden

To analyse how electricity prices influences the wind power production in Sweden, it is essential to first establish an understanding of how the Swedish electricity markets function. Historically, the Swedish electricity market was price-regulated. This changed with the electricity market reform and the introduction of the new Electricity act, which was passed in 1994 and entered into force in early 1996. This transition introduced a new era, where the transmission grid became open to all users, and electricity began to be traded in a competitive and non regulated environment. The reform ensured free trade and increased the possibilities for both import and export of electricity. Furthermore, this reform meant that no one single actor could control pricing, that the markets remain competitive and that the concession owner would have to allow share of transmission capacity in exchange for a fee (Heden, 2012).

Since the deregulation, the Swedish electricity markets have operated under free competition.

Today, the trade of energy and capacity occurs across four primary markets, each serving distinct purposes and with operational timeframes, as illustrated in Figure 2. Various actors operate within these markets, including producers, electricity retailers and larger consumers. Beyond these organized markets, participants can also decide to conduct trade by utilizing bilateral agreements outside of them (Svenska kraftnät, 2025i)

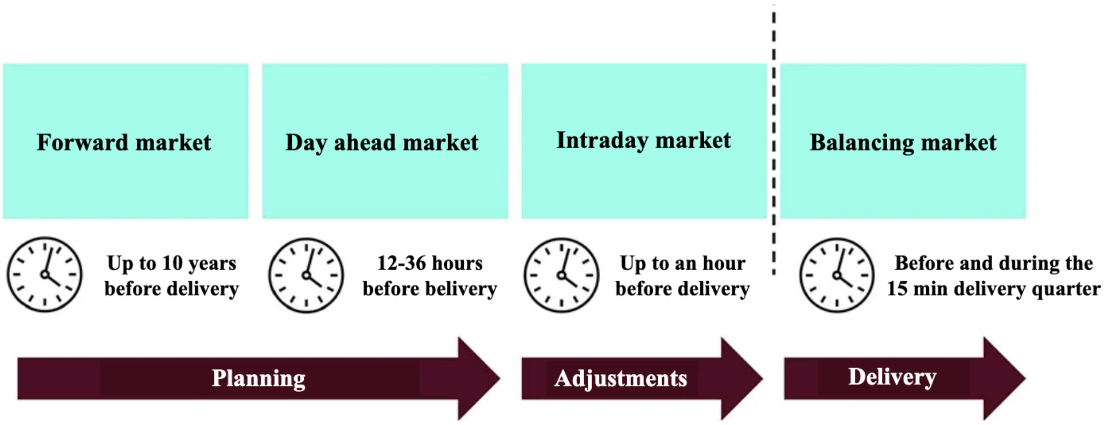


Figure 2: Electricity markets, taken and translated from Svenska kraftnät (2025i)

3.1 Forward market

The market segment with the longest time horizon is the forward market, which serves as the primary hedging mechanism for participants seeking to mitigate risks associated with future price volatility and uncertainty. These contracts, which can be established years, months or week prior to the actual delivery date, provide financial stability for both consumers and producers, ensuring predictable revenue streams for long term investments (European Court of Auditors, 2023).

In the Nordic synchronous area, these financial contracts, or futures, were historically traded on Nasdaq Commodities. However due to recent market reconstruction, as of March 2026, the Nord Pool market is the primary platform for conducting these transactions. The most common futures include system price futures, which hedge against the Nordic reference price Electricity Price Area Differential's (EPAD's). EPAD's are designed to hedge against internal price differences between electricity areas (Svenska kraftnät, 2025f).

3.2 Day-Ahead market

The Day-Ahead market (DA), commonly referred to as the "spot market", is where the majority of electricity is bought (Svenska kraftnät, 2025d). It is a physical market, where prices and volumes are established for the following 24 hour delivery period, to ensure optimal balance between production and consumption (Brolin, Hamon, and Nyström, 2021).

The DA operates based on a marginal pricing, also known as a pay-as-clear model (Brolin, Hamon, and Nyström, 2021). Market participants submit buy and sell orders for each quarter of the upcoming day (Svenska kraftnät, 2025d), which are aggregated into supply and demand curves for each bidding area. The intersection of these curves determines the market clearing price and the cleared volume. Under this principle all participants buy or sell at this uniform market clearing price, regardless of their initial bid price (Brolin, Hamon, and Nyström, 2021).

In Sweden there are two power exhanges which facilitate the DA trading, Nord Pool and EPEX Spot. Despite having multiple operators, the markets are integrated through a European framework known as market coupling. This ensures that prices are determined through a shared auction process across Europe (Svenska kraftnät, 2025d).

The daily trading cycle is structured as following (Nord Pool, 2026):

1. 10:00 Available transmission capacities for the grid are published from the Transmission System Operators (TSO's)

2. 12:00 Deadline for participants to submit their bids, specifying volume, price and what bidding area they are located in, for all quarters for the following day.

3. 12:45 (at the earliest) The spot prices are announced, followed by distribution of individual results to all participants

It is Svenska Kraftnät, Sweden's TSO, which allocates capacity in the Swedish transmission grid, which is what is used for trading. It is calculated both internally and between bidding areas and neighbouring countries (Svenska kraftnät, 2025d). In the Nordic synchronous area a newly introduced flowbased capacity method is utilized to estimate the available capacity in the grid. It is implemented on the DA market, and in comparison to its predecessor it is more efficient, safe and has improved opportunities for the complex power system (Svenska kraftnät, 2026b).

3.3 Intra-Day market

Following the DA transactions, participants have the opportunity to adjust their previous bids on the Intra-Day (ID) market. This market serves as a vital complement to the DA market, due to its mechanism for maintaining system balance. Trading in the ID market is conducted both through a continuous process and three specific auctions. The Market Time Unit (MTU) is set at 15 minutes (Svenska kraftnät, 2025g). Recently, a transition was made from the previous 1 hour MTU to this current MTU. This change was implemented to more accurately reflect the behaviour of a power system with a high penetration of VRE, thereby enhancing operational safety and market efficiency (Svenska kraftnät, 2021).

In Sweden, trade for each specific hour closes 60 minutes prior to delivery. The three ID auctions are strategically scheduled, where the two initial auctions (IDA1 and IDA2) opens at 15:00 and 22:00 the day before delivery (D-1), and the final auction (IDA3) at 10:00 on the actual day of delivery. Continuous trade opens immediately following the initial auction. The principle of market coupling also applies to the intra-day markets, as long as there is transmission capacity between bidding areas (Svenska kraftnät, 2025g).

While the DA capacity calculations currently utilize the new flowbased model (Svenska kraftnät, 2025d), the ID market still operates using the predecessor method, Net Transfer Capacity (NTC). Once the DA market results are finalized, the remaining available capacity on the grid is calculated using NTC (Svenska kraftnät, 2025g). The decision to not implement the flowbased method on the intra-day markets is primarily due to its high complexity, where current algorithms do not yet support such calculations in a market characterized by continuous trading (Brolin, Hamon, and Nyström, 2021).

This ID market is crucial for producers and Balance Responsible Parties (BRP's) as a complement to the DA market. On a systematic level, it has been identified as a factor in decreasing the volume of activated ancillary services. For VRE producers, production is inherently tied to forecast uncertainty. As production forecasts become increasingly reliable closer to the MTU of delivery, due to more accurate weather data, the importance of the intra-day markets grows. It provides the necessary flexibility to adjusting production plans, by either increasing or decreasing allocated volumes (Brolin, Hamon, and Nyström, 2021)

Other market actors may face unplanned technical outages, introducing a need to rebalance them in accordance to market requirements. Furthermore, fluctuations in demand, changes in allocated capacity, or variations in available production capacity serve as strong incentives to participate in the ID markets to minimize imbalances (Brolin, Hamon, and Nyström, 2021).

3.4 Balancing market

The balancing market represents the final stage of the trading process, as seen per Figure 2. Its primary objective is to ensure the safe operability of the power grid by maintaining the system frequency within the Nordic synchronous area at 50 hertz (Hz) (Svenska kraftnät, 2025b). Following the outcomes of the previous markets, the balancing markets are used to handle real-time physical deviations. To maintain this balance at every 15 minute MTU, actors are activated for either up-regulation (increasing production or decreasing consumption), or down-regulation (decreasing production or increasing consumption) (Svenska kraftnät, 2024a).

These imbalances often stem from the inherent uncertainty and variability of production plans, particularly as the share of Variable Renewable Energy (VRE) increases. To manage these deviations and disturbances, Svk utilizes ancillary services and remedial actions. These services are traded on specific balancing markets, where control room operators at Svk are responsible for activating the resources and services to uphold grid stability across both Sweden and the wider Nordic System (Svenska kraftnät, 2025b).

All participants on the Swedish electricity markets are personally responsible for their own balance. They are incentivized to strive for balance, in accordance with existing terms and conditions, and when a participant causes an imbalance on the grid they are penalized given the polluter pays principle (Svenska kraftnät, 2024a). This is managed through an imbalance settlement, performed following every delivery quarter. This process quantifies the magnitude of the imbalance, identifies responsible parties, and calculates the associated costs which are then distributed among the relevant actors who caused it (Svenska kraftnät, 2025a).

There are a variety of different ancillary services, which are procured by Svenska kraftnät on separate markets, each with distinct purposes, participants, time horizons and technical requirements. This division of services and markets ensures efficient resource allocation, while simultaneously enabling a strong, competitive and transparent markets. These ancillary services are typically categorized into three main levels of response: Fast Frequency Reserve (FFR), Frequency Containment Reserve (FCR) and Frequency Restoration Reserve (FRR) (Svenska kraftnät, 2024a).

Amongst the three categories, FFR, has the fastest response time, at almost 1 second, and is used to stabilize disturbances at low levels of system inertia. FCR, on the other hand, is designed for frequency containment when there are frequency anomalies in the Nordic system. It is constituted of two products: FCR-D (Disturbance), which responds automatically if the frequency deviates from the normal frequency interval (49.9 - 50.1 Hz) and is almost as efficient as FFR, and FCR-N (Normal), which is a symmetric service allowing continuous upward and downward regulation within the normal band. Finally, FRR is constituted by services, which are designed to restore the frequency to 50 Hz, after a disturbance has occurred. FRR services include both automatic (aFRR) and manual (mFRR) activation. Notably, mFRR is further divided into an Energy Activation Market (EAM) and a Capacity Market (CM), to ensure both system stability and long term flexibility (Svenska kraftnät, 2024a).

Participation in these balancing markets is voluntary and based on competitive bidding. However, to ensure operational reliability, any resource must undergo a pre-qualification process to prove its technical capability before it can participate (Svenska kraftnät, 2025k).

Due to its controllability and applicative qualities, it is hydropower that has dominated the ancillary services in Sweden historically. While hydropower still is vital to the balancing and still is a large contributor to deliver these services, an increasing share of activated ancillary services from other production types have begun to enter these markets, especially wind power, as the technology improves and market needs evolve (Almqvist, 2023).

4. Wind Power

The human utilization of wind is a tradition spanning millennia, with historical applications ranging from usage at sea to propel our ships, to water pumping and grain milling. The first documented electricity generating wind turbine was demonstrated at the 1883 Vienna International Electrical Exhibition. Following this initial development, large scale usage was delayed until the 1970s. During this period significant technical breakthroughs allowed wind power to emerge as a viable competitor to the conventional technologies, such as coal and early hydropower (Möllerström, Gipe, and Ottermo, 2025).

By the late 20th century, wind energy had established itself as one of the most prominent sustainable energy sources (Ackermann, 2012). In following decades the technology improved further, with rapid advancements such as larger rotor diameters, increased capacity factors and the deployment of active pitch control (Möllerström, Gipe, and Ottermo, 2025). This scaling trend is Figure illustrated in 3, providing insight to the fast size evolution of turbines.

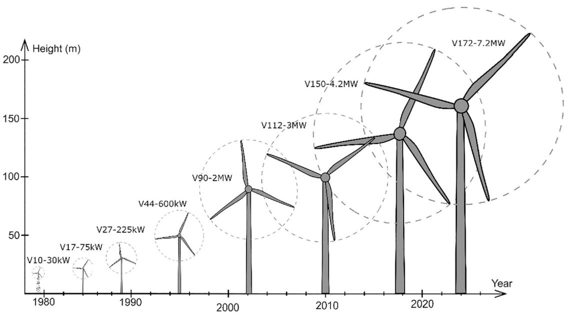


Figure 3: Evolution of wind turbines, note some turbines are even larger (Möllerström, Gipe, and Ottermo, 2025, p. 507)

Today, the absolute majority of wind power turbines are constructed as Horizontal Axis Wind Turbines (HAWTs), and features almost exclusively a three-bladed rotor configuration. These systems have a propeller like appearance, and consists of a tower with a nacelle mounted at its top. The nacelle houses the electrical generator, the gearbox, and the rotor. Modern systems integrate mechanisms for automated yaw control, which actively adjusts the orientation of the nacelle relative to the direction of the wind, for optimizing the aerodynamic efficiency as well as ensuring the structural safety (Ackermann, 2012)

These turbines operate primarily on the principle of aerodynamic lift. As airflow passes over the profile of the blades, it generates a so called lift force, perpendicular to the wind velocity. This lift force is magnitudes greater than the opposing drag force, thus providing the necessary aerodynamic torque required to rotate the blades and drive the internal generator (Ackermann, 2012).

4.1 Physical Principles and wind power extraction

The total kinetic power [Watt] available in an unobstructed, ideal wind stream is governed by fluid dynamics and is expressed as Equation 1:

$$\mathrm{P} = \frac{1}{2}\rho A V^3 \tag{1}$$

where V = wind speed [m/s], A = the rotor swept area, and ρ represents the air density, which varies dynamically with atmospheric pressure and temperature (Ackermann, 2012).

However, the power defined in Equation 1 represents a theoretical maximum. In practical applications, it is impossible for a wind turbine to extract all of this kinetic energy existent in the wind (Ackermann, 2012). The actual mechanical power captured by the rotor, P_T, is instead quantified using the Power coefficient, C_p, which represents the ratio of extracted power to the total available power (Alex, 2017). This coefficient is a based on the pitch angle of the blades β, as well as the tip speed ratio, λ (Bilendo et al., 2023). Consequently, the actual power captured by the turbine is formulated in Equation 2.

$$P_{\mathrm{T}} = \frac{1}{2}\rho A V^3 * C_p(\lambda, \beta) \tag{2}$$

According to the Betz limit, the maximum theoretical efficiency for turbines is $C_p = 0.59$ or (59%). Therefore, no conventional turbine can convert more than 59% of the available kinetic energy of wind into mechanical energy (Alex, 2017). An additional takeaway from Equation 2 is the cubic relationship between wind speed and power output (Ackermann, 2012). Furthermore this relationship implies that minor changes or variations in wind speed yields a large increase in the available kinetic energy, for instance a doubling of wind speed results in an eight fold increase in the available energy (Alex, 2017).

To evaluate operational performance, the power output of a turbine is characterized by its power curve, which presents the relationship between wind speed at hub height and the electrical output (Bilendo et al., 2023). A typical, ideal, power curve is visualized in Figure 4.

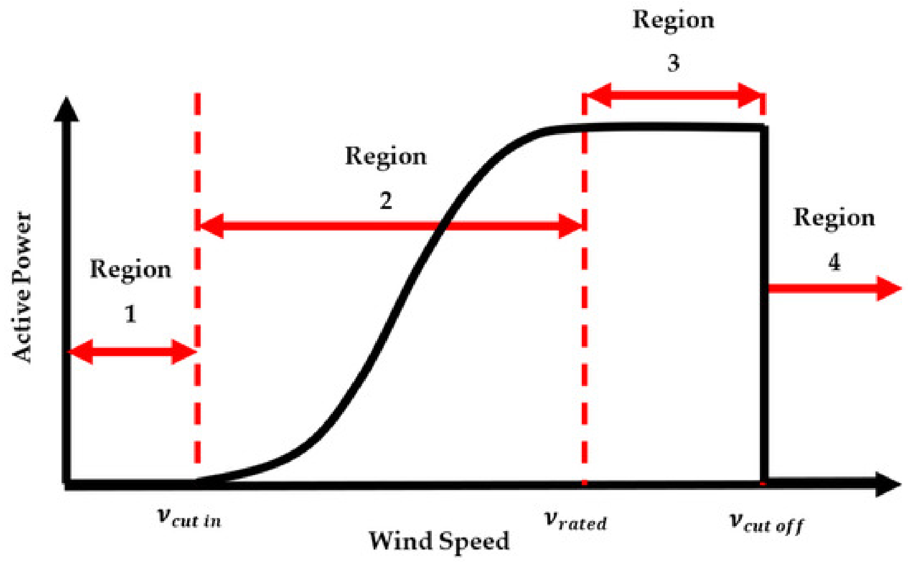


Figure 4: Ideal wind turbine power curve (Bilendo et al., 2023, p.3)

As illustrated in Figure 4, the operational profile is categorized into four distinct regions, bounded by three critical thresholds: the cut-in speed ($v_{\text{cut-in}}$), the rated speed (v_{rated}), and the cut-out speed ($v_{\text{cut-out}}$) (Bilendo et al., 2023).

In Region 1, the wind speed is insufficient for the wind turbine to generate power, keeping the production at zero until the wind reaches the cut-in speed, $v_{\text{cut-in}}$. Within Region 2, spanning from $v_{\text{cut-in}}$ to v_{rated}, the electrical power output increases cubically with the wind speed (Bilendo et al., 2023). When the rated wind speed is reached, v_{rated}, the turbine transitions into Region 3, where it generates power at a constant, maximum rated power output, even as the wind speed continues to increase (Ackermann, 2012). With the help of modern pitch regulation technology, operators can reach the rated power at lower wind speed and thus faster reaching their stable maximum output level (Möllerström, Gipe, and Ottermo, 2025).

However this constant maximum production has an upper limit, and when the wind reaches the cut-out speed, $v_{\text{cut out}}$, the turbine enters Region 4 and shuts down. Typically, the cut out speed lies somewhere in the interval of 20 and 25 m/s (Ackermann, 2012). The shut down is typically done as a safety measure to prevent mechanical or structural failures, as turbines are sensitive to extremely high wind speeds (Bilendo et al., 2023). Additionally, these power curves profiles, are highly sensitive to a variety of factors, such as geographical conditions, rotor aerodynamics and overall turbine size (Ackermann, 2012).

Looking ahead, the global expansion of wind energy usage is accelerating. By 2030, the global wind energy capacity is forecasted to reach nearly 2000 GW, while total annual generation is projected to exceed 4500 TWh. Additionally, driven by its affordability, wind power, together with solar PV, is expected to account for up to 96% of all new power installations worldwide (International Energy Agency, 2025).

4.2 Wind power in Sweden

In 2025, Sweden's total electricity production amounted to 160.5 TWh, of which nearly 39 TWh originated from wind power, representing approximately 24.3% of the country's total production. Onshore wind assets contributed to the absolute majority of this volume, in total 38.7 TWh, whereas offshore installations, which are almost exclusively located in SE4, accounted for 0.25 TWh (eSett Oy, 2026)

The geographical distribution of this production among the bidding areas are detailed in Table 1. The data highlights a significant regional imbalance, where the vast majority of wind power production originates from SE2, accounting for roughly 41.4% of the national total wind power generation, followed by SE3, while SE1 and SE4 exhibit the lowest production volumes (eSett Oy, 2026).

Table 1: Aggregated wind power production [TWh] per bidding area in 2025 (eSett Oy, 2026)

SE1	**SE2**	**SE3**	**SE4**
6.54	16.14	10.78	5.51

Evaluating the historical trajectory of installed capacity provides insights into the dynamics of deployment and capacity expansion in Sweden. Table 2 illustrates these regional development patterns from 2023 to 2025. Notably, SE2 has the highest installed capacity at $\approx$ 7.77 GW (7765 MW) in 2025, following a substantial regional increase of over 1000 MW within two years. While all bidding areas experienced net capacity growth, SE4 observed the most constrained growth, indicating lower investment rates (Energimyndigheten, 2026).

Table 2: Installed wind power capacity [GW] in each bidding area, from 2023-2025 (Energimyndigheten, 2026)

	2023	**2024**	**2025**
SE1	3.00	3.07	3.29
SE2	6.82	7.08	7.77
SE3	3.97	4.23	4.49
SE4	2.43	2.44	2.58

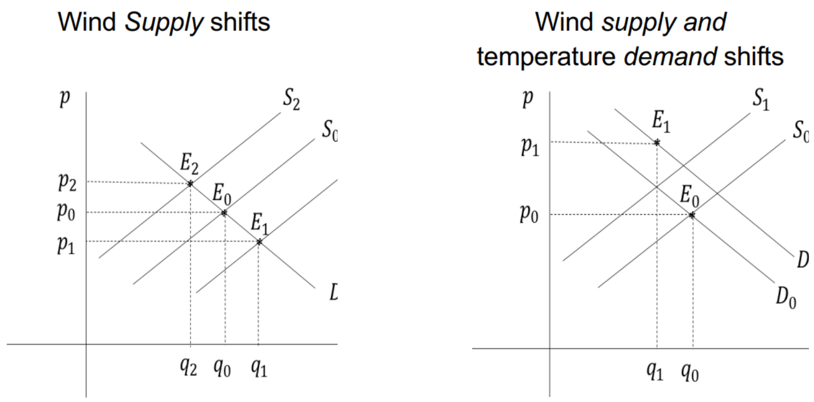
This capacity is closely related with the number of turbines present in the each area, as shown in Table 3. Interestingly, SE4, features more turbines than SE1, despite SE1 both having higher capacity and total production volume in 2025. This discrepancy could potentially be analytically explained by the generational and technological variations within the regions, where the average capacity per turbine is greater in SE1 than in SE4. A potential explanation could be that the SE1 wind power infrastructure is constituted by larger turbines, with higher capacities and that the meteorological conditions in the northern areas may be more beneficial for wind power generation than those in the southern regions.

Table 3: Number of wind power turbines in each bidding area, from 2023-2025 (Energimyndigheten, 2026)

	2023	**2024**	**2025**
SE1	814	823	872
SE2	1902	1941	2038
SE3	1647	1658	1684
SE4	1134	1125	1126

4.3 Wind power influence on the electricity markets

It is important to investigate how wind production influences prices on the electricity markets. As previously established by the merit order effect, generation sources with lower marginal costs are cleared first, and the market clearing price is determined by the most expensive bid required to meet demand. A more in depth conceptualization of how electricity prices are influenced by the supply of wind power production and temperature is visualized in Figure 5.



(a) Illustration of how how prices may change due to supply

(b) Illustration of how how prices may change due to shift in supply and demand

Figure 5: Supply and Demand curves based on wind supply and temperature (Sandberg, 2024, p. 11)

Figure 5a demonstrates how the market equilibrium shifts based on wind availability. The intersection point E_0 denotes the market equilibrium during a day with average wind conditions, with a quantity q_0 and a price of p_0. When wind speeds increase, the corresponding surge in electricity production shifts the supply curve from S_0 to S_1. Given a constant demand, this shifts the equilibrium to E_1, reducing the price to p_1 while increasing the quantity to q_1. Since $p_1 < p_2$ the increase in wind supply results in a decrease of the market price. Conversely, on days with lower wind conditions the opposite is true, meaning the supply curve shifts inward, resulting in an equilibrium which features a higher market price (Sandberg, 2024).

The dynamic becomes more complex when a thermal dimension is introduced, integrating temperature variations into the supply and demand curves, as illustrated by Figure 5b. In this scenario E_0 is the equilibrium during an average windy and warm day. During periods where low temperatures coincide with low wind speeds, which is a common occurrence during winter months, power demand rises sharply due to needs for heating, while the wind power generation drops. Consequently, the supply curve shifts inward from S_0 to S_1, while the demand curve shifts outward from D_0 to D_1. This creates a new market equilibrium E_1, with a cleared quantity of q_1 and a price of p_1. As seen in the figure, p_1>p_0, meaning the price increases, however, as q_1<q_0, the supply decreases. This visualization represents a highly constrained market scenario, with a lower available volume at a significant higher cost. Furthermore, it is important to nuance this relationship, as the correlation between wind and temperature varies seasonally and has different impacts depending on local characteristics of the bidding areas, and that additional factors also have impacts on the supply and demand curves (Sandberg, 2024).

As a consequence of the price dynamics presented in Figure 5, market driven curtailment emerges as a prominent operational strategy. When the regional supply peaks coincide with internal transmission bottlenecks, electricity prices may plunge very low, or even under zero, which could fall beneath operational thresholds, forcing producers to intentionally downregulate production to avoid financial penalties.

4.3.1 Operational optimization and ancillary services

When wind power operators participate in Day Ahead markets, as well as the ancillary markets, they must utilize forecasting tools to estimate their real time production. Bidding based on inaccurate forecasts lead to over- or underestimation of the volume and capacity delivered in real time. These deviations cause energy imbalances, forcing the BRP to pay financial penalties through imbalance pricing. Interestingly, these imbalances can sometimes be leveraged for operational optimization, for instance if the Day Ahead price is negative and an operator produces less than their scheduled production plan, it can technically mitigate losses or generate a positive surplus, depending on the specific imbalance price at that hour (Hamon and Persson, 2023).

Furthermore, the participation of wind power on the ancillary market has grown in both in scale and systematic importance as the installed capacity increases. The segment where wind assets are most critically anticipated to contribute are the down regulation markets. During periods of high wind and low demand, operators can help balance through curtailing production by using active pitch control and braking technologies (Almqvist, 2023).

Historically, wind power participation has been limited on the balancing markets. However, its increasing involvement is today vital for enhancing grid flexibility, safe operability, and overall socioeconomic value (Svenska kraftnät, 2021). Today, wind power is gradually entering and participating on almost all balancing markets (SWECO, 2023), emerging as a trending technology on both the FCR-D down service, as well as on the mFRR up and down services. Its presence is the most pronounced on the mFRR down regulation market, where over 900 MW was prequalified in the fourth quarter of 2024. Their involvement will, besides the assistance in balancing the grid, further improve the liquidity and enhance competition within these markets, which in turn will lower prices (Svenska kraftnät, 2024a).

Despite these advancements, certain structural obstacles continue to limit widespread market participation. A primary barrier is the lack of knowledge about how wind assets can technically contribute and how the complex markets operates. Furthermore bidding carries inherent financial risks, as weather dependent forecasts remain volatile. The age of a wind asset also plays a decisive role, as older wind farms often fail to meet the technical requirements to participate, while operators of younger farms frequently exhibit uncertainty regarding market dynamic and financial risks. Finally ownership structures and Power Purchase Agreements (PPAs) can heavily complicate strategic decision making. In certain cases a wind farm can have multiple stakeholders and owners, which can prove to complicate aligning on a balancing market strategy. Similarly, standard commercial PPAs pose a barrier, as actors bound by such a contract are often obligated and required to deliver maximum output to their corporate buyer, leaving them with no uncommitted capacity to allocate to the ancillary markets (Almqvist, 2023).

5. Wind Power Forecasting (WPF)

Accurate and high-quality Wind Power Forecasting (WPF) is crucial for the safe and reliable operation of modern power systems. For instance high-quality WPF enable stable grid utilization and minimize the need for costly backup power (Z. Liu et al., 2025). When combined with load forecasts, WPF contributes to more efficient power generation planning, which further strengthens the balance and stability of the grid (Saxena et al., 2024).

The inherent intermittency and variability of wind power increase the demand for precision in forecasts, as fluctuations in production can threaten the balance and complicate system planning (Z. Liu et al., 2025). Furthermore, WPF is vital for reducing the uncertainty in wind power production, which facilitates for higher penetration in the power system (Hanifi, X. Liu, et al., 2020). This is complicated by that the field of WPF is shaped by volatile weather patterns, maintenance requirements and unpredictable equipment failures, as well as regulatory constraints imposed by the electricity grid (Suárez-Cetrulo et al., 2022).

5.1 WPF Classification

Due to its societal importance, wind power forecasting is a highly established research field with various methodologies and definitions. Yang et al. (2025) categorizes WPF based on four dimensions: time-horizon, spatial range, input data, and chosen method of modelling. The reasoning behind this is to distinguish between different approaches based on specific application areas.

In contrast, Hanifi, X. Liu, et al. (2020) simplifies this classification by focusing on the two most dominant categories: time-horizons and applied methodology. In this simplified view, the method determines the required input data and resulting level of model complexity.

5.1.1 Time Horizons

The time-horizon is a fundamental classification in WPF, though definitions vary across the literature. Commonly four subgroups are used: very/ultra-short, short-term, medium-term and long-term horizons. Typical ranges are summarized in Table 4.

Table 4: Examples of time horizon definitions across papers

Authors	Very/Ultra short	Short term	Medium term	Long term
Soman et al., 2010	< 30 min	30 min - 6 hours	6 hours - 24 hours	1 day - 7 days
Z. Liu et al., 2025	< 30 min	30 min - 1 hour+	Days - months	1 year+
Yang et al., 2025	15 min - 4 hours	4 hours - 3 days	3 days+	3 days+
Hanifi, X. Liu, et al., 2020	< 30 min	30 min - 6 hours	6 h - 24 hours	1 day - 1 month

The application of a forecast depends heavily on its horizon. For instance, forecasts created within the very-short term horizon are commonly applied for grid stability and real-time scheduling and operation of wind power farms. Short term forecasts can also be applicative for this operational purpose, but are often also utilized for Day-Ahead (DA) transactions on electricity markets (Z. Liu et al., 2025).

5.1.2 Spatial Resolution and input data

The spatial dimensions refers to the "level" at which the forecasts is performed. Commonly, forecasts are conducted at a unit/turbine level, on a singular wind farm level or performed across a larger regional cluster area, which could entail creating forecasts for multiple wind farms or potentially even larger regions (Yang et al., 2025).

The choice of input data is highly intertwined with what method of modelling is chosen, as different methods typically require different type of data (Yang et al., 2025). Suárez-Cetrulo et al. (2022) distinguishes between traditional models and Artificial Intelligence (AI) models. Z. Liu et al. (2025) further divides AI approaches into two distinct groups: Machine Learning (ML) models and ensemble learning.

5.2 The Wind Power Prediction Process

The literature suggests at a standardized process in WPF, involving three overarching key stages (Koukaras et al., 2024; Nair, Vanitha, and Mohan, 2017):

1. Data preprocessing.
2. Model training, tuning and optimization.
3. Model testing and evaluation.

Below details of each stage, and their respective approaches and techniques, are presented in depth.

5.2.1 Data preprocessing

Data preprocessing involves manipulating data to optimize model performance. Nair, Vanitha, and Mohan (2017) include data acquisition in their WPF process as a preliminary step to preprocessing. Once collected one can apply methods which enable the selection of relevant features to make the data more appropriate for the model. Some methods which can be used for reducing complexity of advanced data are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and t-distributed Stochastic Neighbor Embedding (t-SNE). They also facilitate extraction of meaningful feature patterns and effectively reduces dimensions of data, helpful for redundant datasets (Dubey et al., 2026). One can also apply Exploratory Data Analysis (EDA) for this purpose, in which the data is examined through examination of distributions and trends in data, aiding in feature selection (T.A. and Columbus, 2025).

In addition to, this common steps taken during data preprocessing include resampling to a desired time frequency, adding "dummy" variables to the dataset, and handling missing observations in the dataset (Koukaras et al., 2024). It can also include additional data cleaning and normalization of values (Nair, Vanitha, and Mohan, 2017). Data cleaning is critical as it ensures duplicate records as well as outliers are removed, which help reduce model bias (T.A. and Columbus, 2025).

Finally the data is split into model training and testing sets. Common allocations for this include 80/20 or 70/30 splits to enhance model robustness. When time series data is utilized the split the split is created with seasonality in mind (not randomized data splits) (Koukaras et al., 2024; T.A. and Columbus, 2025)

5.2.2 Model training and optimization

The second stage in the streamlined WPF process includes constructing the model architecture, training the model as well as hyperparameter tuning. Model performance, its ability generalization ability and robustness is largely impacted by the values of the hyperparameters. Specifically the hyperparameters governs the models architecture and they fit to the data. Thus their optimization is of high importance to the performance of the trained model (Hanifi, Cammarono, and Zare-Behtash, 2024).

While grid search is a traditional approach for finding optimal parameters (Koukaras et al., 2024; Aslan et al., 2025), it is computationally expensive and unreliable as results from previous trials are not taken into consideration. This downside of inefficiency is further amplified by the risk of it finding non-optimal solutions. To mitigate these risks, more sophisticated Bayesian optimization based methods, such as Sckit-opt, Optuna and Hyperopt can be utilized. These are more efficient and ensure higher accuracy, as they use statistics and consider results from past trials when they choose the next combinations of parameter values. This then leading to better and highly performing models while being computationally efficient (Hanifi, Cammarono, and Zare-Behtash, 2024).

To further mitigate risks of overfitting and enhancing model generalizability techniques like cross-validation are frequently employed during the tuning process. One approach is to incorporate five-fold cross-validation, which show good results in improving robustness (T.A. and Columbus, 2025).

5.2.3 Model testing and evaluation

Finally the WPF forecasting methodology is finished with testing and interpretation of results. Model performance is typically evaluated using statistical metrics. There are a variety of different metrics used across literature and studies. Examples of metrics commonly applied for WPF include Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and R^2 (Dubey et al., 2026; T.A. and Columbus, 2025; Z. Liu et al., 2025). Additionally normalized metrics, such as Normalized Root Mean Square Error (NRMSE), are also often applied for measuring a models relative accuracy across different scales (Koukaras et al., 2024).

Furthermore, beyond statistical measures the inference time is a key metric for real-time applications (Koukaras et al., 2024). Visual comparisons between model predictions and the actual production are also valuable as they provide insights into model performance and behaviour (T.A. and Columbus, 2025).

To address the "black box" nature of complex ML models one can apply advanced analysis tools for evaluating model behaviours and which assists in interpreting and clarifying which factors influence the models predictions. This also increases the reliability of results, as transparency increases confidence in the model's performance. Examples of such of interpretability tools are SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). Both are typically used for better understanding model predictions, where LIME provides local explanations for model predictions, whereas SHAP visualizes how each input feature affects the models predictions (Aslan et al., 2025).

5.3 Forecasting methodologies

As previously mentioned there exists a variety of methodologies which can be implemented when one is predicting wind power production.

5.3.1 Persistence model

The simplest approach to WPF is the persistence method which assumes that the wind speed at time $t + x$ will be identical to time t. In other words the method performs under the assumption that the past and future wind speed is highly correlated (Wu and Hong, 2007). This, often times called "naïve" predictor, performs surprisingly well for very/ultra-short term horizons and frequently outperforms other types of models. It is typically used as a benchmark against more complex models (Soman et al., 2010). However the accuracy of the persistence model quickly deteriorates with time as the time-horizon x increases (Wu and Hong, 2007).

5.3.2 Traditional models

Traditional models can be divided into physical and statistical approaches. Physical models estimate power output using Numerical Weather Prediction (NWP) technology, which take a variety of factors into account. By simulating the impact of meteorological conditions the power output is calculated. These models are however both computationally heavy and complex, and can be very sensitive to initial conditions (Z. Liu et al., 2025).

These types of models do not require historical data for training, but are instead dependent on complex equations and mathematical models. The variables used are both meteorological data and geographical details of the wind turbines/farms area to determine wind speed. The windspeed prediction is then used combined with the wind turbine power curve to forecast wind power output (Hanifi, X. Liu, et al., 2020). Since the characteristics significantly affect the distribution of local wind speeds and therefore also generated wind power, traditional physical models which take these factors into account usually achieve a better and more accurate performance in the long term horizon compared to other models (Suárez-Cetrulo et al., 2022; Z. Liu et al., 2025).

On the contrary statistical models establish mathematical and statistical relationships between input and output data. Common examples include Autoregressive Moving Average (ARMA) models and Support Vector Machines (SVM). Unlike the physical models, statistical approaches do not require details about the physical factors at play, but instead rely on the historical wind speed and output data at the site (Z. Liu et al., 2025). A big advantage is that they are easy to formulate and are computationally efficient (Hanifi, X. Liu, et al., 2020), however they can sometime struggle with understanding the non-linear and erratic nature of wind data (Z. Liu et al., 2025).

5.3.3 Machine learning and Ensemble models

Machine Learning (ML) models offer significant advantages over the traditional physical and statistical methods, by effectively capturing non-linear and dynamic relationships without requiring explicit physical features (T.A. and Columbus, 2025). Within WPF, Artificial Neural Networks (ANNs) are frequently employed due to their good performance and predictive accuracy on the complex input and output mappings. Whilst there exists numerous configurations of these Nerual Networks, techniques such as Long-Short term memory (LSTM) and Convolutional Neural Networks (CNN), have been proven efficient for ultra-short and short term WPF (Hanifi, X. Liu, et al., 2020; Z. Liu et al., 2025).

To further improve robustness and reduce bias ensemble models have proven as among the top performers in WPF (Z. Liu et al., 2025). The ensemble approach is built upon combining multiple predictors to achieve superior performance compared to single based learners (Ren, Suganthan, and Srikanth, 2015). T.A. and Columbus (2025) found that ensemble models such as Gradient Boosting Decision Trees (GBDT) and XGBoost consistently outperformed stand-alone Ml models.

The tree-based structure that some ensemble models, such as XGBoost, have imply a higher time during model hyperparameter optimization and tuning, due to its iterative structure. However, the inference time is more efficient than deep learning models in an actual prediction scenario (T.A. and Columbus, 2025).

Some ensemble models also employ boosting, a technique that by training predictors sequentially, each subsequent predictor is aiming to reduce and correct the error of its predecessor, turning multiple weaker learners into stronger ones (Koukaras et al., 2024). The GDBT model is in some aspects inefficient during training, leading to computational inefficiency. To address these limitations newer implementations reduce complexity by utilizing binning. LightGBM, which was presented in 2017 by Ke et al. (2017), is a novel implementation of GBDT which by grouping continuous features into discrete histograms reduces model complexity and computational inference while maintaining almost the same accuracy (Ke et al., 2017).

5.4 Summary WPF

Numerous WPF technologies exist today, ranging from traditional physical models to highly advanced and complex ML models, such as ensemble boosting ones. Each has their own distinct advantages and application areas. Several significant improvements within energy systems have been enabled through integrating machine learning for power forecasts, which strengthens the power grid (Dubey et al., 2026).

However, a gap remains in the current literature. While most studies focus strictly on features such as meteorological or geographical features, the vast majority of research omits and disregards the complex relationship between wind power production and the electricity market dynamics. This is of course the focus of this paper.

6. Methodology

The methodology for this project has been carefully selected to meet its goals and provide answers to the research questions. The first phase of the project revolved around conducting research into relevant topics to obtain a thorough understanding of Sweden's electrical system, wind farms operation and functionality, and how machine learning can be applied to develop a technique for comprehending the maximal technical potential of wind power production. This was conducted through a vigorous literature study, where related papers and previous work was evaluated, to establish key topics and workflows possible to integrate into this project.

6.1 Machine learning model selection

As previously discussed in Chapter 5, Wind Power Forecasting (WPF) generally uses three primary modelling techniques: physical, statistical, and machine learning (ML). While physical models can be computationally intensive and statistical models often struggle to capture the non-linear and intermittent nature of wind data, ML models have emerged as the superior alternative. These ML models generally offer higher accuracy, greater robustness, and stronger ability to capture the complex relationships between meteorological features and actual wind power production.

Previous research indicates that ensemble based gradient boosting models, including eXtreme Gradient Boost (XGBoost) and Light Gradient Boosting Machine (LightGBM), frequently outperform other methods and are identified as some of the most suitable models for WPF. For instance, in a WPF study conducted by T.A. and Columbus (2025), XGBoost was found to marginally outperform other ML models, such as Random Forest, where it was both effective and adequately capable of capturing wind power generation trends in regards to the non-linear relationships. Similar results were found by Suárez-Cetrulo et al. (2022), who identified that XGBoost was the best performing WPF model, noting its excellent predictive accuracy and efficient inference time.

Furthermore, LightGBM was found to be among the top three model performers by Koukaras et al. (2024). Their findings show that LightGBM was significantly faster than its competitors, while maintaining high statistical accuracy, particularly at a one hour resolution. Similarly, Aslan et al. (2025) implemented LightGBM due to its superior ability to maintain high computational efficiency even with large, complex datasets, without compromising on robustness prediction accuracy.

With the requirements of model robustness, efficiency and ability to capture the dynamic and complex relationships present in the data, scikit-learn's Histogram Gradient Boosting Regressor (HGBR) model was selected. HGBR is an implementation inspired by the LightGBM algorithm. Where both LightGBM and XGBoost are highly effective variations of the Gradient Boosting Decision Tree (GBDT).

GBDT models offer high accuracy, by utilizing an ensemble of decision trees, trained in sequence to minimize a loss function by fitting negative gradients. While both simple and accurate, traditional GBDT can be inefficient during training and demanding in terms of energy and memory consumption. LightGBM addresses these inefficiencies by introducing two novel concepts: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These techniques allow the model to maintain a similar accuracy while simultaneously significantly reducing computational overhead, by over 20 times (Ke et al., 2017).

The primary inefficiency in traditional GBDT stems from the computationally expensive processing of scanning all data points to find optimal split points in the decision trees. LightGBM addresses this by focusing on instances with larger gradients, as those with smaller gradients already have a small training error, suggesting the model is already well trained on those. For a feature *j* at a split point *d*, the variance gain $V_j(d)$ is estimated by prioritizing and keeping all high gradient instances (set *A*), and performing random sampling of instances with small gradients (set *B*). This efficiently reduces the search space. The estimation of variance gain is mathematically represented by the following equation (Ke et al., 2017):

$$V_j = \frac{1}{n}\left(\frac{(\sum_{x_i \in A_l} g_i + \frac{1-a}{b}\sum_{x_i \in B_l} g_i)^2}{n_l^j(d)} + \frac{(\sum_{x_i \in A_r} g_i + \frac{1-a}{b}\sum_{x_i \in B_r} g_i)^2}{n_r^j(d)}\right) \tag{3}$$

In Equation 3, A_l and A_r represents the the sets of instances with high gradients, on the left and right side of the split point *d* respectively. B_l and B_r similarly represents the sets of sampled small gradient instances. In mathematical terms this is expressed as $A_l = \{x_i \in A : x_{ij} \leq d\}, A_r = \{x_i \in A : x_{ij} > d\}, B_l = \{x_i \in B : x_{ij} \leq d\}, B_r = \{x_i \in B : x_{ij} > d\}$. To compensate for the altered distribution which is caused by the low gradient sampling, the coefficient $\frac{1-a}{b}$ is introduced for normalization. This coefficient normalizes the sum of gradients in set *B* to reflect the total population of low gradient instances, thus ensuring low bias within the model despite the data volume reduction (Ke et al., 2017).

HGBR, as scikit-learn's implementation of LightGBM, has shown the ability to perform magnitudes faster on large datasets, compared to standard gradient boosting methods. Through the utilization of discrete histograms, the algorithm effectively reduces the number of evaluated split points during tree construction. Another feature of HGBR is its native support for missing values in data, which increases usability and simplifies the preprocessing phase (scikit-learn, 2026). HGBR is also a prominent top performer, for example in one study it, together with LightGBM, was among the top three performers in regards to both statistical metrics and computational time for short-term time horizon forecasts (Koukaras et al., 2024).

6.2 Framework

The central objective of this project is to reconstruct the historical technical potential of wind power across Sweden's four bidding areas (SE1, SE2, SE3 and SE4), for the year 2025. Historical generation data acts as a confounded time series, reflecting a combination of weather conditions, technical availability and market driven operational behaviour. To isolate the true technical maximum capacity of wind power production, based solely on weather and availability, a specialized novel modelling framework is proposed in four stages.

1. Data preprocessing, model optimization and training
   For each of the four bidding areas, data will be preprocessed and then used to train an Optuna optimized HGBR model. The training set will exclusively contain weather features and technical information data, from hours where the corresponding Day Ahead spot price exceeds 3 €/MWh, filtering out historical intervals prone to economic driven curtailment.

2. Model prediction and evaluation
   Each optimized and trained model will be deployed to predict hourly wind power generation over the entirety of 2025. This data set will include both the high priced periods used during training, and the lower priced hours which were omitted during the training.

3. Reconstruction of the historical maximum theoretical potential
   Model predictions and actual production data will be combined, under predefined structural assumptions, to create the final historical reconstruction of the technical maximum production.

4. Potential verification and analysis
   The reconstructed potential will be validated and examined through a variety of tests and visualizations, to verify that market price signals have been eliminated. The results from this will provide insights into the behaviours and potential of wind power generation across the bidding areas.

By training the HGBR models exclusively on highly profitable hours, the algorithms will learn the relationship between meteorological variables and power production, under the assumption that wind farms operate at unconstrained capacity during high price conditions. This training logic is therefore uniquely designed to isolate and remove the market signals from the models, thus removing the influence of market driven curtailment and allowing the models to predict what the models could have produced during low priced hours based only on present weather conditions.

In the following sections the stages of the process will be presented in depth. The initial steps of this proposed methodology closely aligns with the traditional WPF process, containing preprocessing, model optimization, training and prediction. Finally, the potential creation and evaluation steps will be presented.

6.2.1 Preprocessing

During the preprocessing phase, the required data was firstly acquired. According to Yan et al. (2022) WPF requires a certain training set data size, to sufficiently capture and include the cyclic seasonal patterns behaviour a region may experience. One suggestion is to utilize a one-year data set (Yan et al., 2022), which was utilized in this work.

Given the large geographical area for each bidding area, the meteorological data contained a high volume of measurement points. Thus Principal Component Analysis (PCA) was employed to mitigate the "curse of dimensionality" and effectively reduce the dimensions within the feature space. For each meteorological feature, the high dimension data was projected onto five principal components. This projection summarizes the original data by focusing on the axes which represent the highest variance. This allows for a significant reduction in computational overhead, while still retaining critical information of the data (Leskovec, Rajaraman, and Ullman, 2014).

Feature engineering was further used for improved model performance. This included adding ± 1 hour time lags to the wind speed variables at all used altitudes and transforming the cyclic variables into Cartesian coordinates (sine and cosine). This was performed for wind direction, and makes the circular nature of the variable more apparent.

Finally the most crucial step during the preprocessing was filtering the training data, so that hours where the price was below a manually defined threshold of 3€/MW was excluded all together. By omitting these hours we prevent the model from learning drops in production from economic decisions and optimization, rather than weather conditions.

6.2.2 Model optimization and training

To ensure optimal performance, this step involved optimizing the models through tuning their hyperparameters. This was performed through the utilization of the Optuna framework, an automated Bayesian optimization method.

Optuna allows for a dynamically constructed parameter search space χ, where the search aims to find the optimal configuration of hyperparameters x^* within χ by maximizing the objective function $f(x)$, as seen in Equation 4 (Akiba et al., 2019).

$$\chi^* = \arg\max_{x \in \chi} f(x) \tag{4}$$

The Optuna search direction was set as to maximize the coefficient of determination, R^2, as it provides a normalized measure of how well the model explains the variance in wind production, given by Equation 5. In this equation, n is the number of total observations, y_i is the actual observed value for index i, $\hat{y_i}$ is the model's predicted outcome for index i, and where the mean of the observed values is represented as $\bar{y}$ (T.A. and Columbus, 2025).

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y_i})^2}{\sum_{i=0}^{n}(y_i - \bar{y})} \tag{5}$$

Furthermore the choice of parameter search space was inspired by Cai et al. (2024), which utilized Optuna during their tuning of a LightGBM model, with some minor deviations. Additionally, an L2 regularization term, which was not included in their search space, was included to penalize complexity and prevent overfitting. The concluding parameter search spaces are depicted in Table 5.

Table 5: Hyperparameter search spaces used for Optuna optimization

Hyper parameter	Lowest value	Highest value	Parameter type	Log scale
Learning rate	0.01	0.3	float	True
L2 regularization	0.01	0.1	float	True
Max depth	5	10	integer	False
Max iterations	400	1000	integer	False
Max leaf nodes	15	256	integer	False
Min samples leaf	5	100	integer	False

The tuning was performed through 100 trials, which means that for each model 100 different configurations of hyperparameter values were tested. To ensure reliability a 5-fold cross validation was applied during the tuning, as well as a fixed value of 42 as a random seed for reproducibility.

Following this optimization, the models will be trained on the data using their respective found optimal hyperparameters.

6.2.3 Model prediction and evaluation

Following the previous steps, the trained models will be employed to predict production for all hours of 2025. Thus, it will predict on the already seen hours, where the price was above 3 €/MWh, but also the unseen data with hours which features lower prices. To use the models in the coming potential reconstruction, the results will first be evaluated and tested using both metrics and visual representation to investigate their behaviour. This will be done based on price, to see how the models behave for low and high prices respectively. The metrics included for evaluation are presented below.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{6}$$

The MAE metric is given by Equation 6. It is given by the absolute sum of all prediction errors, where x_i and y_i are the actual value and model prediction for index i and n is the total number of observations (Suárez-Cetrulo et al., 2022). MAE is an applicable metric for WPF model evaluation since it provides a useful general average error magnitude, without being too highly impacted by occasional large deviation (Z. Liu et al., 2025).

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{7}$$

In contrary to MAE, which is not sensitive to larger errors, RMSE (seen in Equation 7) takes the squared error between the actual true value x and the model prediction y, and not the magnitude of the error in each prediction, which makes larger errors weigh heavier than lower errors (Suárez-Cetrulo et al., 2022)

R^2 was, as previously presented used for the Optuna optimization, and is given in Equation 5. It is also commonly referred to as the coefficient of determination, as it provides an explanation on how well the model prediction reflect the variability present in the actual data. The maximum possible value is 1, indicating a stronger generalization ability and high predictive ability (Z. Liu et al., 2025)

A non statistical test that was also utilized for evaluation was scatter diagrams. Using these diagrams allows for a visual examination of the models behaviours. On these figures each dot represents one hour, and the colour signals the price during that hour. During hours where the price is below the threshold of 3 €/MWh, the dots are pink, where lower the prices are reflected in darker. For hours where the price is larger than the set threshold the colour is instead green. The x-axes of the plots are the models prediction, and the y-axes show the actual production, both in megawatts [MW]. The diagonal in each plot represents the optimal prediction, where the prediction equals the true actual production. When the dots are below this line this means that the model has predicted a higher production than the actual realised production during said hours, and vice versa if the dots lie above the diagonal.

6.2.4 Creation of the technical potential

To systematically eliminate the historical curtailment from the actual production data, a reconstructed technical maximum potential, P, is established by combining empirical observations with the model predictions. Let E represent the actual, aggregated wind power production reported by eSett for a given bidding area. Furthermore, let Y denote the HGBR model's prediction, and let C represent the total regional installed capacity within that same bidding area. The final, reconstructed technical maximum potential, P, will be generated using the following requirements.

- During very high profitable hours, when the price is greater than 10 €/MWh, we assume that wind assets are operating that their maximum possible capacity. This is motivated by the strong economic incentives for owners to maximize utilization for economic gain. Thus, during these hours we let the potential be equal to the actual production.

$$P = E \tag{8}$$

- During hours where the prices fall below 10 €/MWh, and approaches operational thresholds, curtailment behaviour is likely to be observed. To capture this lost potential, while respecting periods where the model might underpredict the real production, the technical potential will adopt the maximum value between the actual production and the model prediction.

$$P = max(E, Y) \tag{9}$$

- To ensure physical plausibility, the reconstructed maximum potential can never exceed the total installed capacity.

$$P = min(P, C) \tag{10}$$

By adhering to these requirements we ensure that the maximum potential P, will **always** be larger than, or equal to, the true actual production E, such that $P \geq E$. This upholding constraint guarantee that the observed generation is never underestimated.

The resulting historical maximum potential, for each bidding area, is finally compiled into a continuos time series with an hourly frequency.

6.2.5 Potential reconstruction testing

To verify whether the price signals and curtailment factors have been efficiently removed from our constructed theoretical maximum potential, the following experimental strategy was implemented. Four HGBR ML models were trained for each bidding area, and evaluated and compared based on their feature inputs. All four of these models were trained on 10 months worths of data from 2025, and tested on the months of June and July.

- Technical potential models (P_1 and P_2)
  Both models use the historical reconstructed technical potential, P, as their target variable. Model P_1 is trained strictly on weather features and UMM features, whereas P_2 in addition to these variables will also be trained on electricity price data.

- Actual production baseline models (E_1 and E_2)
  Both models will use the actual production data, E, as their target variable. Similar to the technical potential models, E_1 will solely be trained on weather and technical availability data, while E_2 also includes the electricity price data in its feature set.

The objective of this strategy is to analyse the models sensitivities to price. If the proposed methodology for reconstructing the historical technical potential successfully isolates the pure technical potential from market driven curtailment, the performance metrics of the technical potential models P_1 and P_2 should be very similar, such that $P_1 \approx P_2$. This would indicate that adding price features yields no predictive advantage, as the target variable itself has become price invariant.

Conversely the baseline models E_1 and E_2, trained on the actual wind power generation data, are expected to display larger differences in performance. Because the actual production data is affected by pricing, the E_2 model is expected to have an advantage over E_1. These baseline models are created for verification and validation of the results of P_1 and P_2.

The metrics used to evaluate these models performance were the same ones which were presented in Chapter 6.2.3: MAE (Equation 6), RMSE (Equation 7) and R^2 (Equation 5). Additionally, to further explore these relationships, feature importance and interaction effects are examined using Shapley Additive ExPlanations (SHAP), which are discussed in detail in Chapter 6.2.7.

6.2.6 Analysis of the created historical potential

It stands to reason that the lower the price, the larger the fraction of curtailed power. Therefore, the absolute volume of electricity lost to market driven curtailment or economic optimization, according to the recreated potential, can be quantified as a residual vector, calculated as $P - E$. This residual represents a meteorological and availability based estimation of how much additional generation was physically feasible, and therefore lost due to operational curtailment.

To evaluate how this curtailment behaviour scales in relation to market condition and price, a residual framework is structured by categorizing the data based on price intervals. These intervals have been defined into the following bins: (-∞, -5], (-5, -3], (-3, 0], (0, 3], (3, 5] and (5, 10] in Euro per MWh. Within each of these price intervals the following metrics will be calculated: the average hourly residual, the total accumulated residual volume, the percentage residual relative to the total technical potential as well as the sample count. This structural breakdown enables a precise quantitative assessment of curtailment trends during varying price levels. This will be presented both in statistical data, as well as visualized using histograms for higher interpretability.

Complementing these statistical metrics, scatter diagrams were once again employed to provide a visual presentation of model behaviour. It follows the same logistics as previously presented, with one difference: the x-axis represents the technical reconstructed potential P and not the original model prediction. This serves as a visual indicator of uncaptured potential.

6.2.7 SHapley Additive exPlanations (SHAP)

Using SHAP provides a tool for investigating if, and to what extent profitability factors influenced the ML models. As previously established in Chapter 5.2.3, SHAP serves as an advanced visualization tool designed to break down how each feature influences the final predictions.

Originating from collaborative game theory, SHAP calculates each feature's contribution to a model's output. A key property of this method is the efficiency property which that states that for every prediction *f(i)*, the following criteria must be satisfied:

$$f(i) = \phi_0 + \sum_{\text{features}} \phi_{\text{i, j}} \tag{11}$$

where ϕ_0 is the models average prediction and $\sum_{\text{features}} \phi_{\text{i, j}}$ is the sum of all individual SHAP values for instance *i*. This ensures that the total credit for a prediction is distributed fairly among all available features (Ponce-Bobadilla et al., 2024).

By utilizing the open-source SHAP library on the P_2 and E_2 models (which were trained on price), the resulting SHAP values will indicate whether they rely heavily on price features for their predictions. Thus, the structure of the SHAP analysis will allow for comparison between the price features impact on the technical potential model P_2 and actual production model E_2.

6.3 Software and development environment

To take advantage of its extensive libraries the Python programming language was chosen for this project. This provided the opportunity to use existing, specialized machine learning and data analysis tools. Specifically the project utilized libraries such as scikit-learn for model construction, SciPy for performing advanced statistical computations. Data manipulation and exploratory analysis were performed with pandas and NumPy, while plotly was used for visualization. Furthermore the SHAP package was integrated into the workflow, for interpretation of feature impact on model predictions.

Additionally internal software packages from Svk were implemented to improve performance and ensuring compliance to organizational standards and workings.

7. Data

A primary focus was identifying available and suitable data for the project. As data directly impacts the predictive performance of ML models, poor data quality can lead to suboptimal performance, misleading results and instability (Gong et al., 2023). To mitigate these critical risks data quality dimensions, completeness, accuracy, timeliness, consistency and relevance (Wang et al., 2024), were implemented as a decision support during data selection process.

Once a certain level of data quality is reached the uncertainty of WPF is reduced (Yan et al., 2022). However as this study is conducted as a retrospective analysis, rather than a real time forecasting application, the timeliness dimension is of secondary importance.

7.1 Data sources

Multiple datasets and data sources have been used in the process of this work, to be able to answer the proposed research questions. The used data can be divided three subgroups, which are: wind power production data, meteorological data and electricity market data.

7.1.1 Wind power production data

Two potential and available wind power data sources were evaluated for utilization in this project. One available source was SCADA data, which includes production information at a wind farm level in Sweden. While this has high resolution and includes highly detailed information, which enables the option to map farms to their geographical location, it struggles with missing data and concerns with regarding completeness, accuracy and consistency. While the SCADA data source has good potential to perform an in-depth analysis, it was not used in this work due to its weaknesses in the data quality dimensions and for confidentiality reasons.

Another available data source was eSett, which was chosen for this work due to its high data quality. The eSett wind production data consists of the aggregated total production of wind power [MW] in each Swedish bidding zone at an hourly frequency. This data is concluded and finalized two weeks after delivery, thus, the timeliness quality is somewhat limited as the data is not available in real-time. However, as previously mentioned, this is a retrospective analysis, and it is therefore acceptable as long as we are confident that the available data is correct. The reason behind the two week delay is to ensure an accurate settlement of the actual production is delivered. Consequently, this data exhibits a very high level of accuracy, completeness and consistency, necessary for this work.

7.1.2 Meteorological data

The majority of input features into the models consists of meteorological data. These were sourced and fetched from three different sources to ensure high accuracy and make the model prediction more robust. Since wind conditions have a major impact on wind power production, it is crucial that the meteorological data is of high quality.

The sources used were Arome provided by the Swedish institute, SMHI (SMHI, 2026), data from The European Centre for Medium-Range Weather Forecasts, (ECMWF, 2026) and data from the Global Forecasting System (GFS, 2026). These sources have different spatial resolutions, with Arome providing the highest resolution of them all all.

The implemented meteorological features include temperature, wind speed, and wind direction at all available altitudes (10, 100 and 150 m), gathered from all sources.

7.1.3 Market data

Data for prices and Urgent Market Messages (UMM) were sourced from Volue. The price data, comprising of the Day-Ahead spot prices for each bidding area, was resampled to an hourly basis, following the market restructuring from a 1 hour to a 15 minute Market Time Unit (MTU) in late 2025 (Nord Pool, 2025).

The UMM data quantifies the known unavailable production volumes, due to planned maintenance or failures in each bidding area. This data is also aggregated on an hourly basis. The UMM data was specifically focused on losses from wind assets, which are based on operator reported outages in capacity [MW].

7.2 Available data

Finally, the data used in this work was taken from the whole year of 2025, which exhibited 100% availability and contained no missing data.

Following the preprocessing step, where prices below the threshold of 3 €/MWh were filtered out, the final training and test set sizes for each bidding area are outlined in Table 6. While there is an imbalance in these dataset sizes, this variation directly reflects the underlying price dynamics in each respective area, specifically higher frequency of low priced hours in the northern regions.

Table 6: Training and test sets sizes

Bidding Area	**Training [hours]**	**Share of Year [%]**	**Test [hours]**	**Share of Year [%]**
SE1	6612	75.5%	8760	100%
SE2	6134	70.0%	8760	100%
SE3	7777	88.8%	8760	100%
SE4	7840	89.5%	8760	100%

8. Results

In this section the empirical results are presented. First, an initial summary highlighting the key findings across all bidding areas is provided. Subsequently, the detailed results are presented in accordance with the order of the framework structure established in Chapter 6.2.

8.1 Initial summary of results from all the bidding areas

Given the created reconstructed maximum theoretical potential for each bidding area, the total estimated curtailment across Sweden during 2025 amounts to approximately 2.90 TWh, as seen in Table 7. Distributing this figure evenly across the year amounts to an hourly curtailment of 82.679 MW, which increases to 161.472 MW during hours when prices fall below 10 €/MWh. This substantial figure is the theoretical volume which could have been generated, based on available weather conditions. Interestingly, 87.31% of this total curtailment occurred during hours when the price dropped below 3 €/MWh.

Table 7: Summary statistics over all bid zones

	All prices	**Price =< 10**	**Price < 3**
Total curtailment [TWh]	2.90	2.90	2.53
Average hourly curtailment [MW]	82.679	161.472	302.055

8.2 Model optimization

Following the preprocessing of data, Optuna was employed for optimizing the models' performances. This process resulted in a variety of different optimal hyperparameter values, each specific for each bidding area. These are visualized in Table 8.

Table 8: Results from Optuna tuning

Hyperparameter	**SE1**	**SE2**	**SE3**	**SE4**
Learning rate	0.015	0.0475	0.030	0.012
Max depth	6	5	9	8
Max iterations	503	548	528	898
L2 regularization	0.019	0.038	0.035	0.059
Max Leaf nodes	129	256	141	16
Min samples leaf	68	50	76	34
Time	5 min 43 s	4 min 56 s	7 min 16 s	8 min 22 s
Best R^2 score	0.761	0.858	0.925	0.906

As seen in Table 8 all of the optimal trials performed with very good results, resulting in R^2 scores ranging between 0.761 and 0.925, where the perfect score is 1. Additionally we can observe that the many of the found optimal hyperparameters were similar across areas.

Overall the whole optimizing process took a little under 30 minutes. For all models it was found that the value of learning rate had the highest parameter importance, with importances of 71.03%, 86.56%, 58.27% and 86.58% for bidding areas SE1-4 respectively.

8.3 Model prediction and evaluation

The models predictive ability across all bidding areas during high priced hours were remarkably high. This outcome was anticipated, given that the models were trained specifically in that data. For instance, all models achieved an R^2 score above 0.97, demonstrating a robust capability for making predictions during hours. Important, and interestingly the models behaved significantly worse on the unseen, low priced, data. Some regional discrepancies were observed, where SE2 for instance exhibited the poorest performance for these hours, with an R^2 score of 0.39 and an MAE of 696.98 MW. In contrast, the magnitude of error was considerably lower for the remaining areas, especially for the southern areas.

A common trend across all models was a pronounced tendency to overpredict production during low priced hours, while they for the more profitable hours presented with a more standard distribution of predictions.

Using scatterplots provided invaluable insights into these dynamics, as they visually reveal whether a model systematically mispredicts in a specific direction or if the residuals are normally distributed.

8.3.1 SE1

In Table 9 the performance metrics for the SE1 model are presented. The data highlights a substantial performance difference between the two price categories. Specifically the MAE score increases by 213% during the low priced hours, and the R^2 score drops from 0.97 to 0.67.

Table 9: Metrics over model performance for SE1

	MAE	RMSE	R^2
All hours	129.89	201.50	0.90
Low priced hours	266.95	351.88	0.67
High priced hours	85.37	116.50	0.97

A deeper intuition and understanding of this behavioural shift between the price groups is given by the scatterplot in Figure 6. The plot illustrates that the model systematically overestimates production during the pink, low priced hours. More specifically, the model overestimated production in 1661 of the hours compared to 487 hours of underprediction when the price was below 3 €/MWh. For the more profitable hours above this threshold, the distribution of predictions was more evenly balanced, with 3738 overpredictions and 2875 underpredictions. Notably, this number does not tell any information about the magnitude of error.

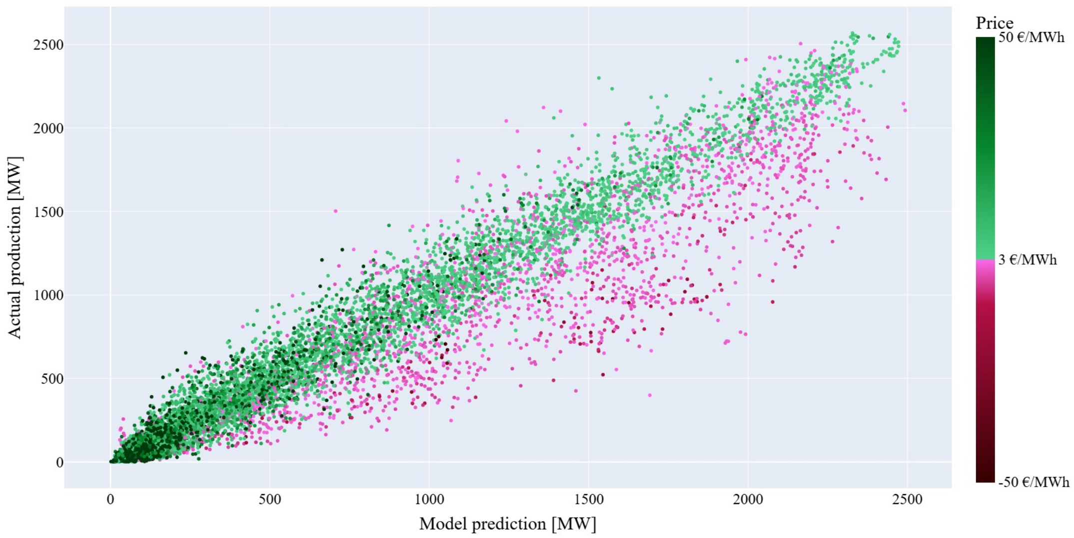


Figure 6: Scatterplot of SE1, with the model prediction against the actual production [MW]

8.3.2 SE2

The SE2 model presented with the most severe performance degradation between high versus low priced hours. It performed excellent on the previously seen, high priced, data, with a near perfect R^2 score and with both low MAE and RMSE values, as seen in Table 10. In comparison, the model performed significantly worse for the lower priced hours. This indicates that while the model could predict production during profitable hours, external factors influenced the other hours, such that it could not effectively predict, despite using the same features and hyper parameters.

Table 10: Metrics over model performance for SE2

	MAE	**RMSE**	R^2
All hours	269.15	491.91	0.86
Low priced hours	696.98	880.55	0.39
High priced hours	86.02	116.92	0.99

This finding is further supported through the visual analysis using the corresponding scatterplot in Figure 7. Here, with a few exceptions, the data forms into two distinct clusters, where the low priced hours deviate systematically from the optimal prediction diagonal. Within this low priced subset, the model overpredicted production approximately 85% of the time, resulting in 2223 overestimations against 393 underestimations. During the high priced hours the distributions were in comparison significantly more evenly distributed, with 3186 overpredictions and 2949 underpredictions, such that it overpredicted around 4% more.

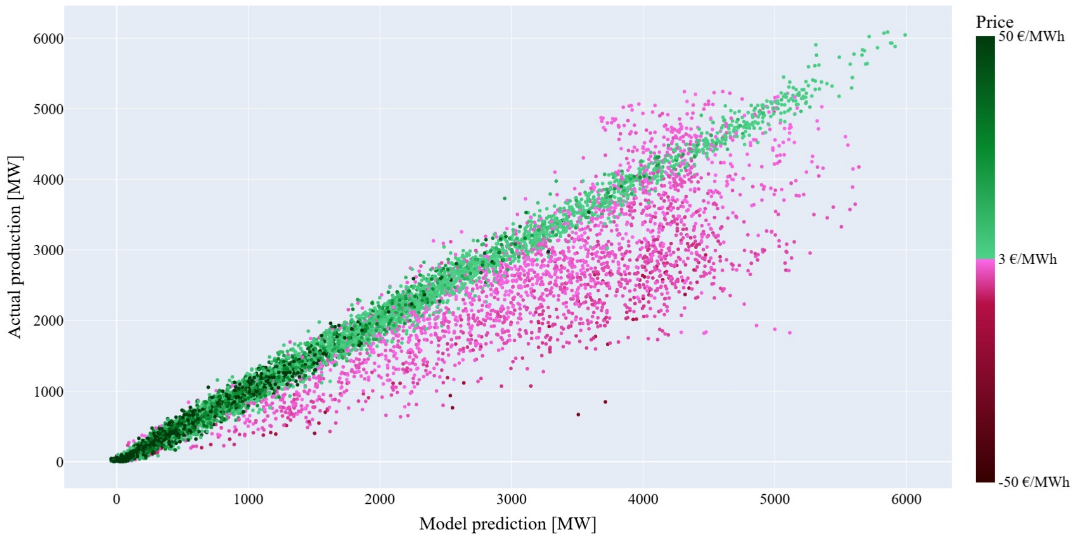


Figure 7: Scatterplot of SE2, with the model prediction against the actual production [MW]

8.3.3 SE3

The observed model performance for SE3 is visualized in Table 11. In comparison to the northern SE1 and SE2 models, this model shows somewhat smaller sensitivity in regards to pricing, maintaining a relatively stable R^2 score across both groups. However, there is a significant difference in other error metrics across both pricing groups. The model generated a much higher MAE for the low priced hours (230.54 MW), than the higher priced ones (41.68 MW), which represents an increase of error totalling 188.86 MW.

Table 11: Metrics over model performance for SE3

	MAE	RMSE	R^2
All hours	62.87	114.74	0.98
Low priced hours	230.54	303.07	0.89
High priced hours	41.68	56.74	1.00

In the scatterplot in Figure 8 we observe a more concise and tighter pattern than previously seen, where the predictions are placed closer to the optimal prediction diagonal. Still, a disproportionate number of low priced hours are positioned below this diagonal, confirming a bias to overestimating production. A pattern which is confirmed by a ratio of 736 overpredictions to 247 underpredictions for these hours. During the remaining higher priced hours the distribution stabilized at 3982 overpredictions and 3796 underpredictions.

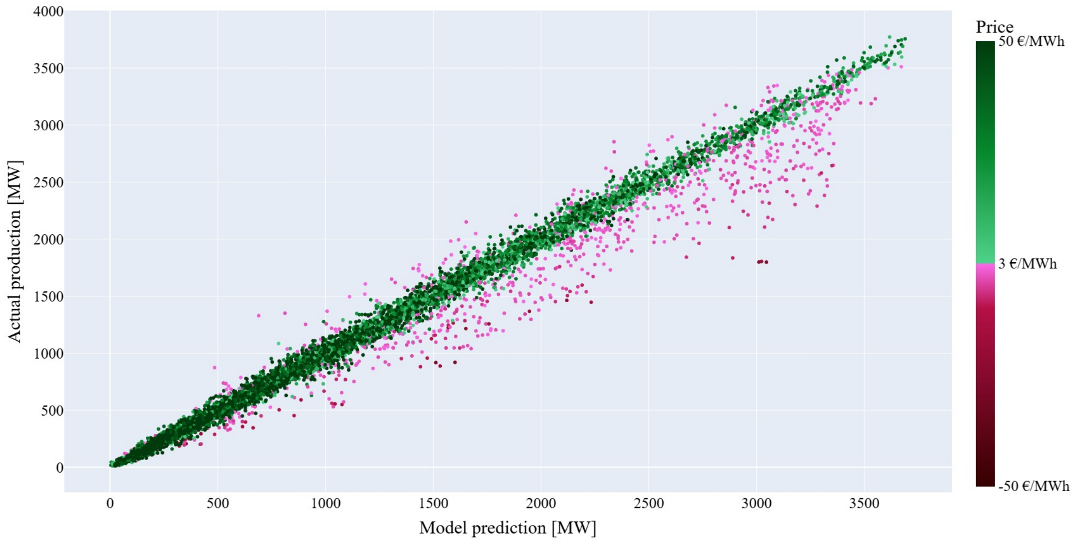


Figure 8: Scatterplot of SE3, with the model prediction against the actual production [MW]

8.3.4 SE4

Finally, the SE4 bidding area model demonstrates a behaviour which closely resembles that of the SE3 model. It obtained high R^2 between the price categories, yet there was an increase in MAE for the low priced hours, rising from 54.10 MW to 140.56 MW, as presented in Table 12.

Table 12: Metrics over model performance for SE4

	MAE	RMSE	R^2
All hours	63.18	89.90	0.97
Low priced hours	140.56	185.35	0.84
High priced hours	54.10	70.71	0.98

The corresponding scatterplot in Figure 9 presents itself similar to the one observed in the SE3 analysis. The overestimation patterns are less visually apparent, than for those in the northern bidding areas. However, the overestimating pattern still remains present. The model overestimated production in 702 instances compared to 218 underpredictions for low priced hours. The residual distribution for the higher prices proved more balanced, where the model overpredicted 3998 times, and underpredicted 3843 times.

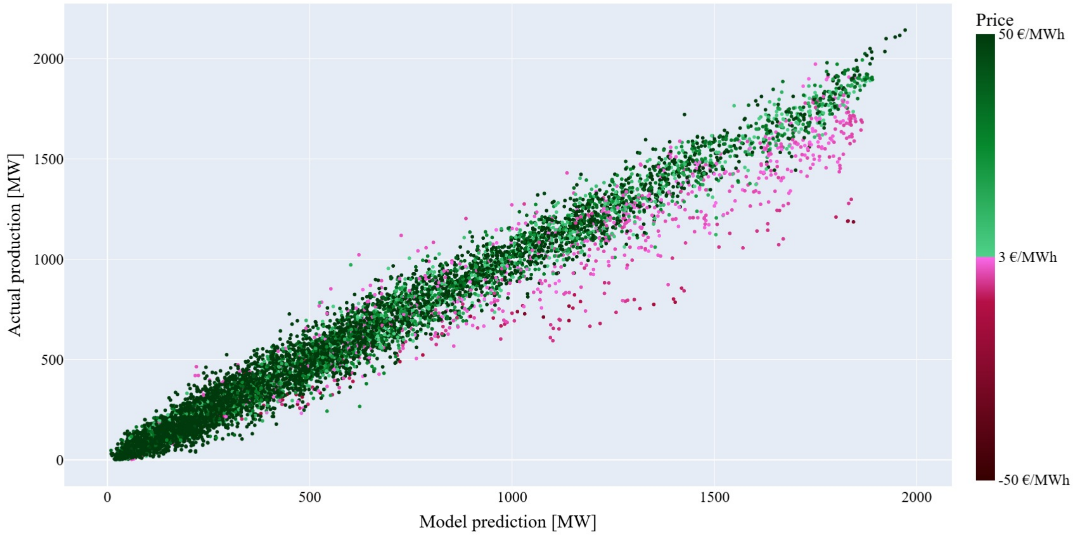


Figure 9: Scatterplot of SE4, with the model prediction against the actual production [MW]

8.4 Reconstruction of technical potential

Here the reconstruction of technical potential is presented. A general takeaway, based on the requirement presented in Equation 10, is that none of the model predictions ever exceeded the installed capacity in their respective bidding area.

The respective installed capacities in each bid zone for 2025 was previously presented in in Table 2. The maximum estimation of the different bid zone potential reconstructions in comparison to the installed capacities is shown in Table 13.

However, worth noticing is that due to the choice of utilizing HGBR ML models, the models predictive ceilings are naturally bounded by the physical capacity limits observed within the training distribution, thus making it infeasible for any model to produce a prediction larger than the maximum observed data.

Table 13: Summary statistics over all bid zones

	Total capacity 2025 [GW]	Max estimated potential [GW]	Share of installed capacity [%]
SE1	3.29	2.57	78.09%
SE2	7.77	6.09	78.38%
SE3	4.49	3.77	83.96%
SE4	2.58	2.14	83.15%

8.4.1 SE1

In Figure 10 the scatterplot for the reconstructed technical potential of SE1 is presented. Here we can observe the limit bounds that the requirements of the strategy induce, that the potential P, is either equal to or larger than, the actual production E. Thus all dots, representing the hours, lie either perfectly on the diagonal, or under it. The dots positioned under the diagonal represents that the recreated technical potential estimates that a higher production would be possible given the technical availability and meteorological features.

It is observable, as it was in Figure 6, that there is a distinct difference between the two categories where the pink dots are more prevalent far from the diagonal.

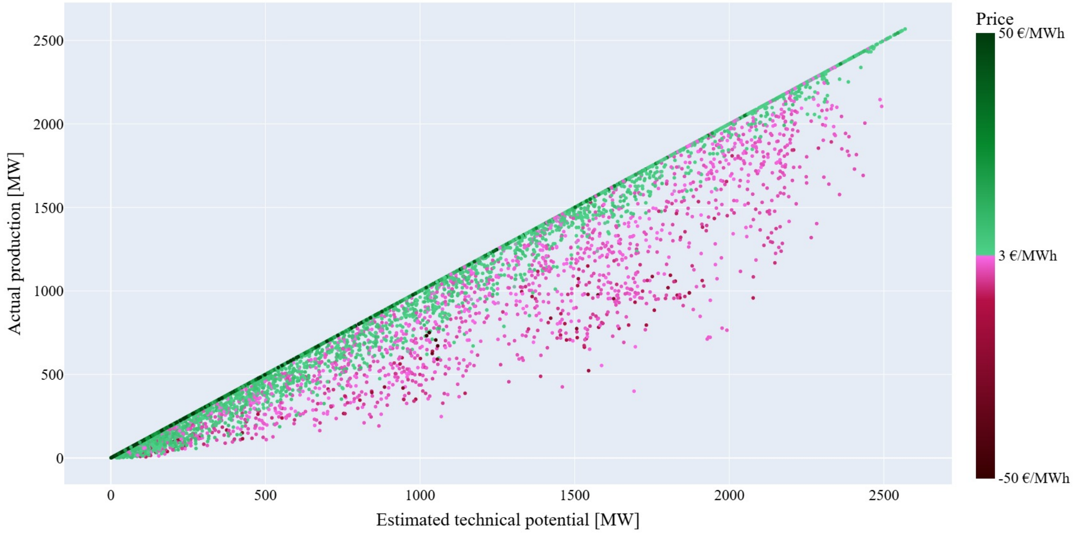


Figure 10: Scatterplot of SE1, with the created recreated technical potential against the actual production [MW]

8.4.2 SE2

The clustering behaviour between the two price groups is even more prevailing in Figure 11, which illustrates the historical reconstruction of technical potential in SE2. The Figure visualizes the distinct differences between the two groups, where the green, high priced hours are closely positioned near the diagonal, while the pink, lower priced hours often fall far below this optimal diagonal.

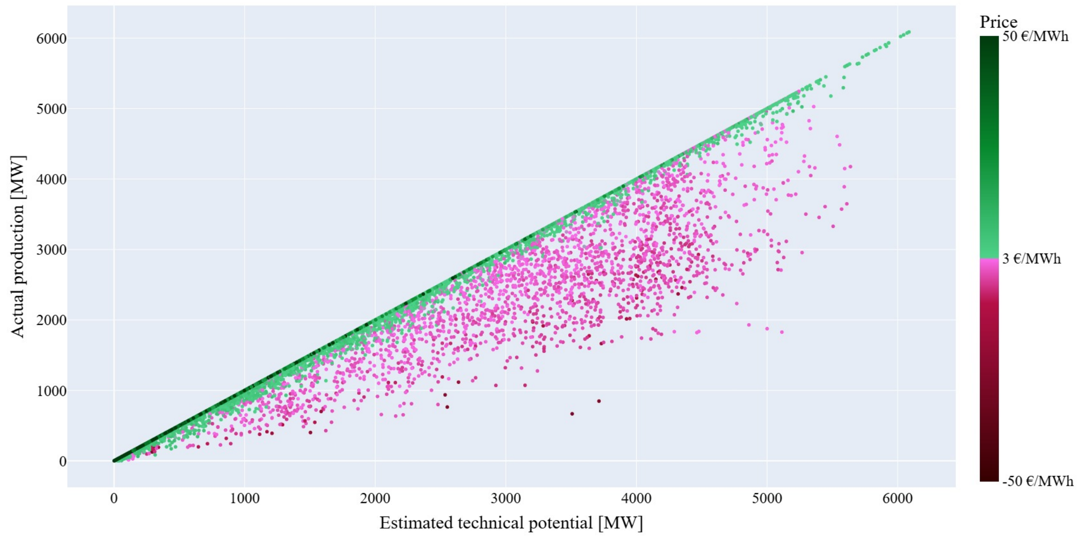


Figure 11: Scatterplot of SE2, with the created recreated technical potential against the actual production [MW]

8.4.3 SE3

For SE3, the technical potential is less widespread, and more closely placed in relation to the diagonal, shown in Figure 12. Here, the green, high priced hours, are distributed extremely close to the diagonal, with no significant over-representation below this line. While the overestimation of low priced hours seem less extreme than those for SE1 and SE2, the behaviour is still observable.

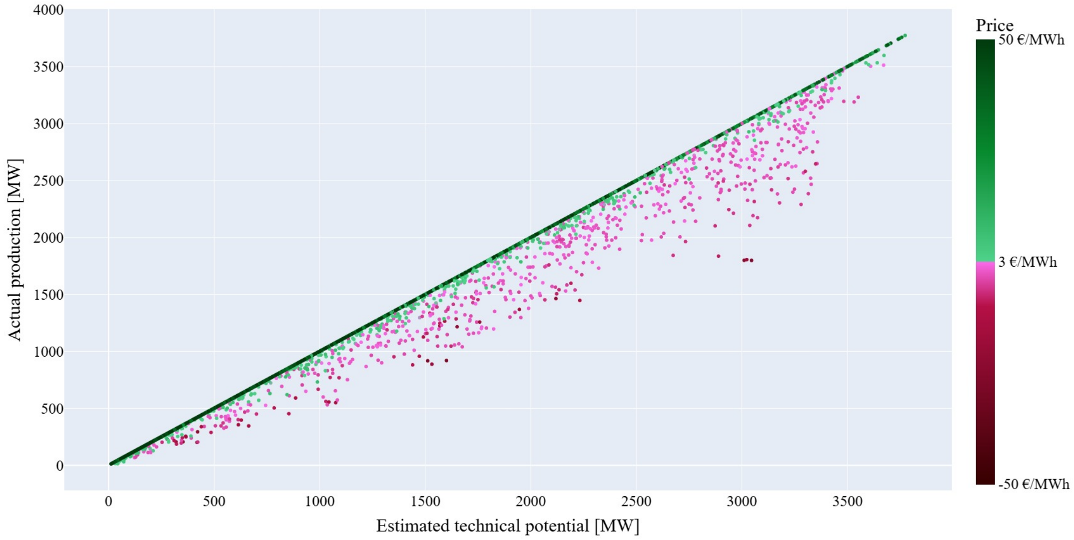


Figure 12: Scatterplot of SE3, with the created recreated technical potential against the actual production [MW]

8.4.4 SE4

The reconstructed technical potential for SE4 seems to behave similar to that of SE3. In its corresponding scatterplot, presented in Figure 13, it is once again observable that the green, high priced hours almost form a straight line, while the low priced hours show overestimation.

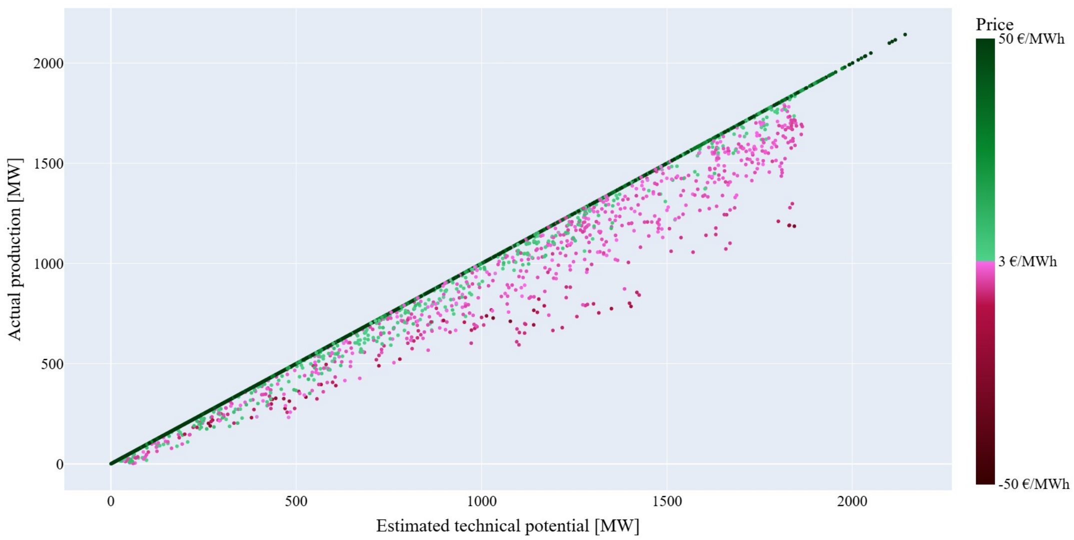


Figure 13: Scatterplot of SE4, with the created recreated technical potential against the actual production [MW]

8.5 Evaluation and analysis of reconstructed potential

The evaluation of the reconstructed technical potential reveals distinct differences between the theoretical potential models and the actual baseline models, across all four bidding areas. All the models with the technical potential as target input, were found to be more invariant to prices. In contrast the baseline models proved more price sensitive, as emerged from the metric testing as well as the SHAP analysis. This was particularly pronounced for northern bidding areas, of which SE2 proved to have the highest price impact on the actual production model.

Using the recreated technical potential to estimate total curtailment, a key finding across all regions, was the distribution of average volume curtailed during different price intervals. The lower price corresponded to the highest average estimated loss of energy, further describing market driven curtailment patterns.

8.5.1 SE1

The performance metrics for the evaluation of the reconstructed technical potential for SE1, featuring the two corresponding technical potential models, P_1 and P_2, and the baseline models, E_1 and E_2, are presented in Table 14. Notably, while the technical potential models performed similarly, larger deviations were observed between the two baseline models.

Introducing a price feature into the potential model P_2 results in a negligible change in accuracy, slightly worsening the MAE score by 2.59%, from 84.58 MW to 86.77 MW, while the other metrics remained relatively constant. This indicates that the target variable, the reconstructed technical potential already was largely invariant to price, and primarily required wind condition data and technical availability to effectively predict production.

In contrast, the baseline models trained on actual production exhibit larger differences. Incorporating price data in E_2 improves MAE by 11.41%, compared to price oblivious E_1. More notably was that the R^2 increased from a moderate 0.68 to a stronger 0.76 through introducing the price variable. This significant improvement demonstrates that the actual behaviour in SE1 is shaped by the dynamics of price, which the price aware model is able to capture.

Table 14: Results model comparison SE1

Model	**MAE**	**RMSE**	R^2
P_1 (Potential - Weather Only)	84.58	118.41	0.88
P_2 (Potential - Weather and Price)	86.77	117.60	0.88
E_1 (Esett - Weather Only)	116.17	161.31	0.68
E_2 (Esett - Weather and Price)	102.91	141.13	0.76

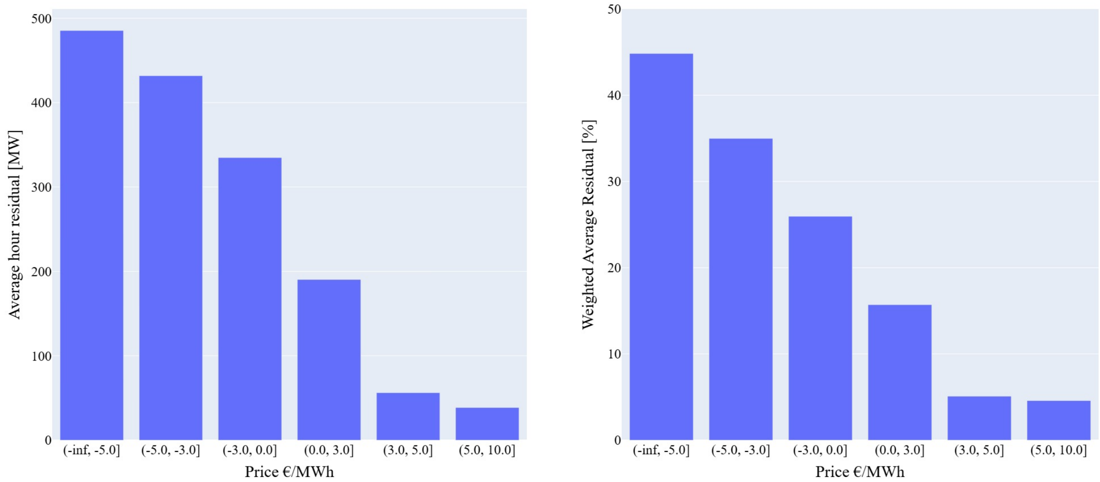
The SHAP analysis further highlights this divergence in model logistics and allows for further understanding of exactly how Day-Ahead electricity price in the region influenced the model predictions. For the potential model, P_2, price exhibits a minor global mean absolute SHAP impact of 21.58 MW and a negative correlation coefficient of -0.73. This negative relationship reflects the relationships, previously presented in Chapter 4.3, where periods of high wind production coincides with lower prices.

In comparison, the actual production E_2 model obtained a slightly higher global mean SHAP impact value of 37.85 MW. This model had a price correlation value of 0.47. This suggests that that during hours with exceptionally low prices, the feature's SHAP impact on the model was highly negative, and thus the model actively predicts a drop in output during low priced hours. For instance during an hour in early June the price dropped to -1.1 €/MWh which prompted to model to apply a massive downward output adjustment by -292.96 MW.

Following the evaluation of the reconstructed technical potential, the data was analysed to estimate the volume of curtailment throughout 2025 according to P, as summarized in Table 15. The table details the average curtailment in megawatts, the average curtailment percentage, calculated as the ratio of total residuals to the total maximum potential in Gigawatt-hour, and the total sum of curtailment across specific price intervals. It is important to note that the hours are not evenly distributed, and as such the sum of curtailment must be interpreted as such. During hours with highly negative prices, below -5 €/MWh, the average curtailment, $(P - E)$, reached 485.31 MW, which represents a large contrast to hours with prices above 5 €/MWh, where the corresponding curtailment only reached 38.56 MW. These numbers are also visualized in Figure 14, which illustrates the average residual volumes and percentages as histograms in subplots 14a and 14b respectively.

Table 15: Results from maximum technical potential estimation from bidding area SE1

Price bins	Average residual [MW]	Average residual [%]	Sum of residuals [GWh]	Count
(-inf, -5.0]	485.31	44.83	39.31	81
(-5.0, -3.0]	431.71	34.99	23.74	55
(-3.0, 0.0]	334.72	25.96	136.23	407
(0.0, 3.0]	190.31	15.70	307.73	1617
(3.0, 5.0]	56.09	5.08	85.32	1521
(5.0, 10.0]	38.56	4.58	78.58	2038



(a) Average curtailment volume [MW] (b) Relative curtailment [%]

Figure 14: Distribution of hourly average curtailment and relative percentage across price intervals in SE1

In summary, the total estimated curtailment of actual production due to market behaviour in the SE1 bidding area amounted to 670.91 GWh, or approximately 0.67 TWh during 2025. Over the entire year, the average hourly curtailment was 76.59 MW. During hours when prices were below 10 €/MWh, this average increased to 117.31 MW. Out of the total curtailment of 0.67 TWh, 75.47% occurred during hours with prices below 3 €/MWh, illustrating how lower prices suppress actual wind power utilization. These statistics are compiled in Table 16.

Table 16: Statistics from bidding area SE1

	Number of hours	**Total curtailment [GWh]**	**Average curtailment [MW]**
All prices	8760	670.91	76.59
Price =< 10	5719	670.91	117.31
Price < 3	2148	506.37	235.74

8.5.2 SE2

The results from the SE2 bidding area stands out against the other regions, exhibiting more pronounced differences in model performance and display a much higher volume of estimated curtailment. As shown in Table 17, the technical potential model remains robust using weather data alone, and adding a price feature to P_2 results in a virtually identical MAE change of 1.51 MW.

A more notable shift occurs in the E_1 and E_2 models, which target the actual production data. For these models, predictive accuracy significantly improves with inclusion of the dimension of price. The MAE decreases by nearly 20%, the RMSE by nearly 25% and the R^2 score improves from 0.77 to 0.87. This represents a significant enhancement in the model's capacity to capture the actual production behaviour. Given that these discrepancies between the technical potential models and the actual production models are so great, it could be argued that price information is largely redundant for the potential models, which remain robust using weather data alone.

Table 17: Results model comparison SE2

Model	**MAE**	**RMSE**	R^2
$\boldsymbol{P_1}$ (Potential - Weather Only)	202.84	267.15	0.93
$\boldsymbol{P_2}$ (Potential - Weather and Price)	204.35	272.68	0.93
$\boldsymbol{E_1}$ (Esett - Weather Only)	265.08	374.23	0.77
$\boldsymbol{E_2}$ (Esett - Weather and Price)	213.43	282.12	0.87

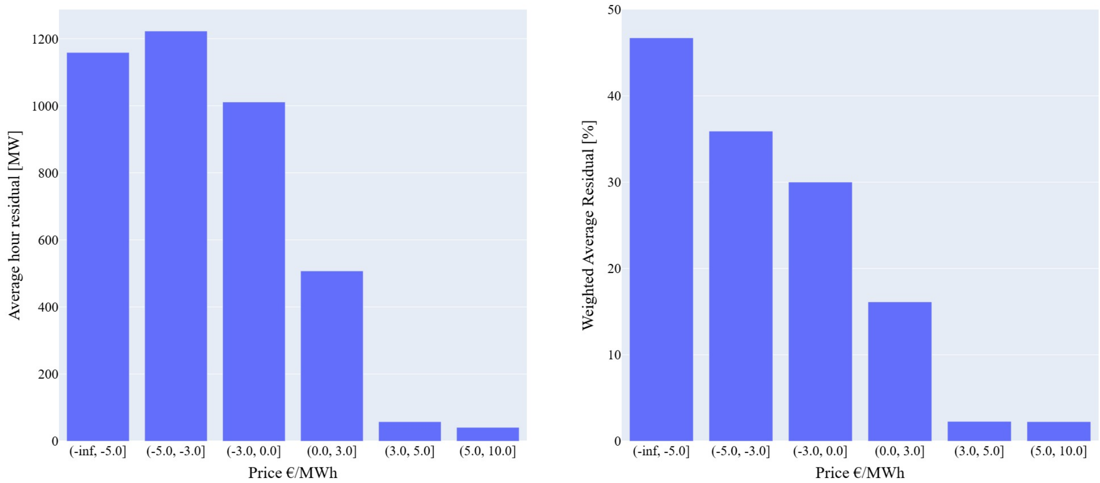
Illustrated by the further SHAP analysis is the absolute contrast in price influence over the models. The global mean absolute SHAP value of the price feature was almost five times higher for the actual production model E_2, for which it obtained a substantial value of 170.01 MW, than for the potential model P_2 for which it was only 34.21 MW. Furthermore, the directional correlation between price and its SHAP impact also diverges. While P_2 captures the natural dynamic of high wind supply depressing price (a correlation of -0.51), E_2 instead relies on operational behaviour, predicting production drops during hours with low prices (with a correlation of 0.44). This is exemplified during a late June hour, where price had dropped to -24.0 €/MWh, for which price had a substantial SHAP impact of -772.78 MW on the E_2 prediction. Thus demonstrating negative price signals prompting massive downward adjustment in model output.

Subsequently, the reconstructed technical potential was analysed across the predefined price intervals, as presented in Table 18. The distribution that appears to clearly underscore a pattern of market driven curtailment, and overall large volumes of unutilized clean energy. All price intervals below 0 €/MWh exhibit average curtailment volumes exceeding 1010.97 MW, peaking at 46.71% of total maximum technical potential during the lowest pricing bin. As the average curtailment during the highest pricing intervals only amounted to an average curtailment close to 50 MW, this indicates larger operational curtailment during the extremely unprofitable hours.

Table 18: Results from maximum technical potential estimation from bidding area SE2

Price bins	Average residual [MW]	Average residual [%]	Sum of residuals [GWh]	Count
(-inf, -5.0]	1158.57	46.71	57.93	50
(-5.0, -3.0]	1222.48	35.91	59.90	49
(-3.0, 0.0]	1010.97	30.00	631.86	625
(0.0, 3.0]	507.07	16.12	969.52	1912
(3.0, 5.0]	57.39	2.27	94.52	1647
(5.0, 10.0]	40.44	2.24	62.76	1552

The statistics for the average residuals, both in terms of absolute volume and relative to the total maximum technical potential, are visualized in the histogram in Figures 15a and 15b. These subplots visually illustrate the pronounced curtailment during the low price hours, as evidenced by the sharp contrast in bar heights between the intervals.



(a) Average curtailment volume [MW]

(b) Relative curtailment [%]

Figure 15: Distribution of hourly average curtailment and relative percentage across price intervals in SE2

Based on these findings, the following conclusions regarding the scale of total curtailment are presented in Table 19. According to the maximal technical potential estimation, a total volume of 1876.49 GWh was curtailed in SE2 during 2025. This significant volume, equivalent to approximately 1.88 TWh of energy, estimated purely based on weather conditions, such as wind speed, underscores a massive volume of unutilized generation.

Out of this total volume, nearly 1.72 TWh, or 91.59%, was curtailed during low priced hours when prices dropped below 3 €/MWh, forcing the average hourly curtailment to surge from an annual baseline of 214.21 MW up to 654.21 MW. Thus, emphasizing the intense impact of market dynamics on wind power production in SE2.

Table 19: Statistics from bidding area SE2

	Number of hours	Total curtailment [GWh]	Average curtailment [MW]
All prices	8760	1876.49	214.21
Price =< 10	5835	1876.49	321.59
Price < 3	2626	1717.71	654.12

8.5.3 SE3

Moving to the southern bidding areas, the evaluation metrics for the SE3 technical potential reconstruction generated a more uniform performance. As outlined in Table 20, both model categories perform almost identically, regardless of whether the price variable was included. The most notable variation is a 5.03% RMSE improvement for the price aware baseline model E_2 compared to E_1, reducing the error from 174.78 MW to 165.98 MW.

Table 20: Results model comparison SE3

Model	MAE	RMSE	R^2
$\boldsymbol{P_1}$ (Potential - Weather Only)	114.68	151.36	0.96
$\boldsymbol{P_2}$ (Potential - Weather and Price)	112.68	150.83	0.96
$\boldsymbol{E_1}$ (Esett - Weather Only)	126.70	174.78	0.94
$\boldsymbol{E_2}$ (Esett - Weather and Price)	123.42	165.98	0.94

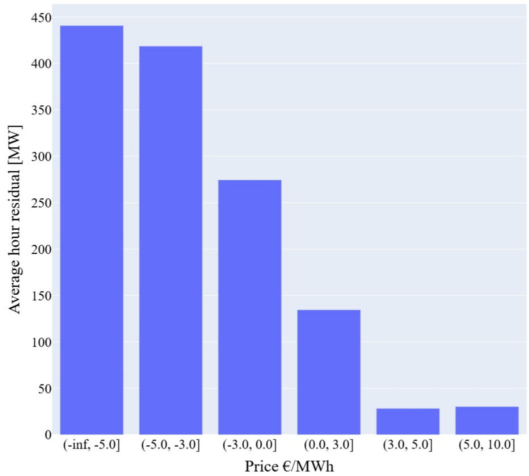
Smaller magnitude of price SHAP impacts were observed for SE3, in comparison to those observed in the northern areas. The global price impacts for P_2 and E_2 was 8.51 MW and 20.40 MW respectively. Despite lower magnitudes of impact on model outputs, the underlying structural logics as previously observed for SE1 and SE2, persist for SE3. P_2 maintains a negative price correlation coefficient of -0.78, while E_2 upholds a positive coefficient for E_2 (0.37). This operational sensitivity to negative pricing is still visible for extremely price negative hours, where an hour with an electricity price of -15.08 €/MWh caused the E_2 model to lower its model prediction by 332.21 MW. The peak impact for P_2, on the contrary, was a minor reduction of 40.94 MW during an 102.31 €/MW hour.

An examination of the historical reconstructed potential shows an consistent overestimation of production during low priced hours, in comparison to those more profitable. This is presented more in depth in Table 21, which shows that curtailment was highly pronounced during the 74 hours where prices fell below -3 €/MWh, with an average hourly loss exceeding 418.75 MW. For higher price hours, the relative curtailment drops to roughly 2%, indicating large differences in price response actions.

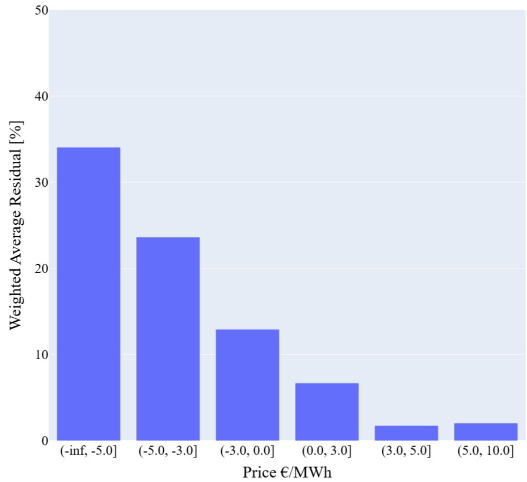
Table 21: Results from maximum technical potential estimation from bidding area SE3

Price bins	Average residual [MW]	Average residual [%]	Sum of residuals [GWh]	Count
(-inf, -5.0]	440.92	34.07	18.96	43
(-5.0, -3.0]	418.75	23.62	12.98	31
(-3.0, 0.0]	274.55	12.94	79.89	291
(0.0, 3.0]	134.51	6.69	83.53	621
(3.0, 5.0]	28.24	1.74	9.86	349
(5.0, 10.0]	30.21	2.03	12.05	399

To inspect the relationship and magnitude variations across the different price intervals, the corresponding distributions are visualized in Figure 16.



(a) Average curtailment volume [MW]



(b) Relative curtailment [%]

Figure 16: Distribution of hourly average curtailment and relative percentage across price intervals in SE3

In total, market driven curtailment in SE3 during 2025 is estimated to a volume of 217.28 GWh (roughly 0.22 TWh), as presented in Table 22. The predominant share of this loss of capacity was concentrated within hours falling below the price threshold of 3 €/MWh. During these specific hours, the average hourly curtailment was 198.59 MWh, accumulating a total volume of nearly 0.20 TWh over the entire year.

Table 22: Statistics from bidding area SE3

	Number of hours	Total curtailment [GWh]	Average curtailment [MW]
All prices	8760	217.28	24.80
Price =< 10	1734	217.28	125.30
Price < 3	983	195.22	198.59

8.5.4 SE4

Finally, the results for the southernmost bidding area, SE4, are presented. Table 23 details the metrics for the region's corresponding technical potential reconstruction models, P_1 and P_2, alongside the actual production models E_1 and E_2. The results demonstrate highly uniform performance between the potential models, with performance variations often remaining negligible, showing less than a 2% degradation upon including price information in P_2. The actual production model however exhibit a benefit from price awareness, reducing by 7.21% improvement (a 5.76 MW improvement), and improving its RMSE by 9.86% in comparison to the weather only baseline.

Table 23: Results model comparison SE4

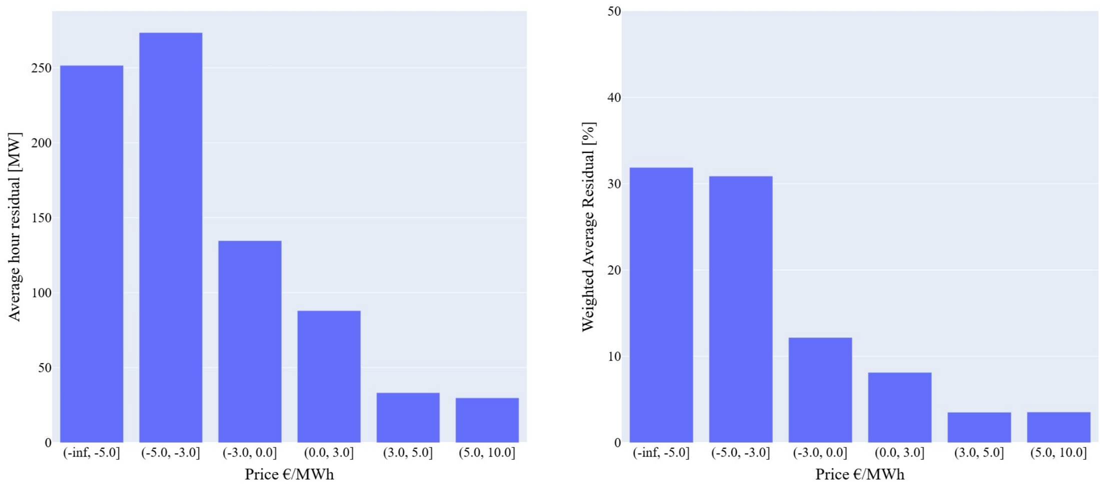
Model	MAE	RMSE	R^2
P_1 (Potential - Weather Only)	70.11	90.01	0.96
P_2 (Potential - Weather and Price)	70.44	90.80	0.96
E_1 (Esett - Weather Only)	79.97	107.20	0.95
E_2 (Esett - Weather and Price)	74.21	96.63	0.96

The global mean absolute SHAP impact reinforces the minimal role plays in establishing theoretical potential in this zone, averaging just 5.68 MW for P_2 compared to a slightly higher 14.05 MW for E_2. The price correlation coefficient was -0.85 for P_2, and 0.38 for E_2, showing the models' internal structures remain consistent with previous regions. Operational adjustments during price shocks are still distinct for the actual production model, E_2, with a considerable downward shift in prediction of 228.97 MW when prices dropped to -5.27 €/MWh.

In terms of total curtailment, SE4, recorded the lowest overall curtailment among the four areas. However, the breakdown across different pricing intervals, compiled in Table 24, still illustrates a clear price dependent distribution trend, with higher volumes concentrated during low priced hours, than for the more profitable periods. This pattern is further illustrated by the histograms in Table 17. During the lowest priced interval (prices below -5 €/MWh), an estimated total volume of 12.83 GWh was curtailed, which equates to an hourly average curtailment of 251.52 MW.

Table 24: Results from maximum technical potential estimation from bidding area SE4

Price bins	Average residual [MW]	Average residual [%]	Sum of residuals [GWh]	Count
(-inf, -5.0]	251.52	31.87	12.83	51
(-5.0, -3.0]	273.43	30.87	10.12	37
(-3.0, 0.0]	134.64	12.17	40.26	299
(0.0, 3.0]	87.92	8.12	47.22	537
(3.0, 5.0]	33.27	3.51	11.65	350
(5.0, 10.0]	29.80	3.54	10.34	347



(a) Average curtailment volume [MW]

(b) Relative curtailment [%]

Figure 17: Distribution of hourly average curtailment and relative percentage across price intervals in SE4

Throughout 2025, the total estimated curtailment amounted to 132.40 GWh (0.13 TWh), as seen in Table 25. True to the national trend an overwhelming majority of this energy loss, about 83.22% (110.19 GWh), took place during hours with prices below 3 €/MWh. Thus confirming that even in southern areas, low prices serves as a cause for production down regulation.

Table 25: Statistics from bidding area SE4

	Number of hours	Total curtailment [GWh]	Average curtailment [MW]
All prices	8760	132.40	15.11
Price =< 10	1621	132.40	81.68
Price < 3	920	110.19	119.77

9. Discussion

While traditional Wind Power Forecasting (WPF) literature often overlooks market influences on wind power generation, this work has specifically focused on the role of market driven curtailment. The following chapter interprets the empirical results, structured around the research questions defined in the introduction. Furthermore, an evaluation of the study's limitations will be performed, as well as suggestions for future work on the subject.

9.1 Market driven curtailment behaviour in Sweden

Research question 1: *To what extent does the Day-Ahead market price signal correlate with wind power curtailment in Sweden?*

As visualized through the histograms and descriptive statistics for each of the bidding areas recreated potential models, there is clear evidence of significant impact of Day-Ahead price signal on estimated curtailment. As the largest volumes of curtailed energy occur during the absolute lowest prices, exhibiting curtailment volumes that are orders of magnitude larger than those observed during profitable hours, this signals that wind power operators generally act according to rational strategies driven by market dynamics. In SE2 the difference between the average curtailment behaviour in prices below -5 €/MWh and +5 €/MWh was 1118.13 MW, or an 93.25% larger volume of average hourly curtailment during low priced hours. Even more striking is the histograms visualizing the differences in average hourly curtailment for each bidding area, for instance Figure 15 detailing pricing curtailment in SE2. During negative prices, or prices close to the operational economic thresholds of production, it is inferred that wind asset operators execute downregulation strategies to minimize negative revenue exposure, even though physical conditions for higher production are present. This is in line with the information presented by Hamon and Persson (2023), that during low negative hours wind power operators can optimize their financial position or mitigate imbalance, which is crucial for avoiding imbalance penalties.

From the analysis it is evident that meteorological and technical availability alone is insufficient to capture operational reality of Swedish wind power production. The improvement in accuracy for the actual production models by the inclusion of Day-Ahead prices as a feature confirms that price signals act as a significant hidden variable in production forecasting.

Through the SHAP analysis, comparing the technical potential models P_2 and the actual production models E_2 the impact of price also became clear. Using the direction of correlation between the price feature of the two models, we see fundamental differences in model characteristics, where the actual wind power production models E_2 have a positive price correlation, while for P_2 this correlation is negative across all four bidding areas. The positive correlation could be a consequence of down-regulation during negative prices, and higher production during periods of higher prices.

Thus, it is clear that the general wind power production in Sweden is not solely a weather dependent generation source, but instead use operation strategies to not only maximize physical possible output, but also to minimize exposure to imbalance costs and negative revenue periods.

Furthermore, apart from the clear improvement in the actual production model performance with the integration of price knowledge, this is further observed by the vast differences in the potential model, as previously discussed.

9.2 Implementation of machine learning in estimating historical wind production potential

Research question 2: *How can machine learning be utilized and validated to estimate the hypothetical maximal potential of wind power production, in the absence of market-based curtailment?*

This study employed a novel framework to address the second research question. This framework utilized Histogram-Based Gradient Boosting Regression (HGBR) models, which were, for each bidding area, trained on historical data from hours with profitable market conditions, during the year 2025. Subsequently, these models were employed to predict the production output for all hours of the same year. To construct the finalized reconstructed technical potential models, the predictions from the HGBR models were combined with actual production data, subject to three specific requirements.

A potential limitation of this methodological setup is that the models were trained on and tested on data from the same calendar year, which stems from the nature of historical reconstruction.

By contrast, a strength of this approach is the utilization of HGBR models for production prediction, which allows the framework to effectively capture the complexity and non-linearity present within the data. This is achieved through the employment of a boosting ensemble of decision trees (Ke et al., 2017), which has proved highly efficient in multiple studies including one by Koukaras et al. (2024). Furthermore, integrating the Optuna framework with a five-fold cross validation implementation to optimize the models hyper parameters enhances both their accuracy and reliability, as supported by Hanifi, X. Liu, et al. (2020) and T.A. and Columbus (2025).

The robustness of this approach is reflected in the evaluation metrics of the HGBR models during profitable hours, where the models consistently obtained high performance. The Mean Absolute Errors (MAE's) ranged between 86.02 MW to 41.68 MW, and the R^2 scores exceeded 0.97 for all models during these profitable periods, which represents an excellent standard in WPF. However, it is critical to note that models had already observed these specific data points during training, thus the excellent performance and high metrics during profitable periods raise concerns regarding overfitting, rather than generalized high performance. Nevertheless, the results from the five-fold cross validation mitigate this concern. The optimal performance across all tuned models obtained R^2 scores exceeding 0.761 on unseen data, which highlights the HGBR models capability to generalize and successfully capture the non linear relationship between meteorological conditions and regional aggregated power output, when production remains profitable. This resilience against overfitting can further be attributed to the found optimal low learning rates during the tuning process, as detailed in Table 8.

One striking finding from the analysis is the predictive inability of the HGBR models to reproduce similarly high accuracy during hours where prices fall below the threshold of 3 €/MWh. Across all four regions the MAE increased by an average of 403.82% during low priced hours. This deterioration was particularly pronounced in SE2, where the MAE escalated from 85.37 MW during profitable periods, to 692.98 MW during low price hours, representing a 810.25% decrease in performance. For the other areas, SE1, SE3 and SE4, the MAE increased with 213%, 553.12% and 38.9% respectively. This systematic degradation in predictive capability signals that regardless of high performance accuracy during profitable hours, the models fail to maintain precision and the same level of accuracy during the low price hours due to external influences which were absent during training. Since the primary distinction between these data subsets is the corresponding Day-Ahead price, this economic variable is presumed to be the underlying factor causing model deterioration. Rather than indicating poor model design, this drop in accuracy supports the core hypothesis of this thesis, that wind assets undergo a significant downregulation or curtailment during low price hours. This creates a discrepancy between actual production and meteorological potential that the training data could not account for, due to its structure.

The resulting technical potential, constructed by combining the HGBR models predictions with the actual production in accordance with the established criteria, show significant results. The underlying assumptions and criteria for the technical potential themselves can be discussed. Particularly, the decision to make the potential equal to the true production when electricity prices exceeded 10 €/MWh. During these hours with high prices it was assumed reasonable to presume that operators maximise their output due to elevated profitability.

Furthermore, the decision to apply a retraining process as an evaluation method, using the technical potential as the target variable to compare price inclusive and price exclusive models against baseline models, requires further discussion.

In relation to the traditional WPF process, standard practices suggest utilizing a training dataset of one year to allow machine learning models to capture the cyclic and seasonal patterns, as recommended by Yan et al. (2022). However, because the historical reconstruction of the technical potential was limited to the year 2025, and a separate subset was required for the purpose of validation, a ten month training set was utilized. This configuration resulted in an 83.33% to 16.67% training, testing split, closely adhering to the recommendation of a 80/20 ratio as presented by Koukaras et al. (2024) and T.A. and Columbus (2025). The selection of the summer season, June and July, for testing and validation was motivated by the historically lower average electricity prices induced by drops in seasonal demand, as noted by Sandberg (2024). Consequently, this period provides an optimal window to detect and analyse the presence of any market induced curtailment.

When comparing the performance of these models, P_1, P_2, E_1 and E_2, even though their target variables differ, important insights emerge. The results indicate that the potential models do not experience a significant increase in accuracy from the inclusion of price. This then signalling that the engineered target variable, is largely invariant to price fluctuations. In comparison the baseline models exhibited a measurable improvement in predictive accuracy when they were allowed to attain information of price. Therefore, this method of evaluation validates the proposed approach to remove market influences from the technical wind potential.

This interpretation is further substantiated by applying the SHAP framework, which, as demonstrated by Aslan et al. (2025), allows for deeper interpretation of model logistics. The SHAP analysis revealed that while the actual production models are highly impacted and sensitive to the value of price, particularly within the northern areas, it does not obtain the same dependency within the technical potential trained models. For instance, the technical potential models all displayed a negative correlation with price, which aligns with the standard market where high wind penetration depresses electricity prices, as presented in Figure 5. For the actual production price aware models this correlation was instead positive, reflecting low power output prediction during low priced hours. While the SHAP framework provides valuable insight, it is worth noting that when highlighting the hours with the highest SHAP price in Chapter 8 it might introduce a bias if those specific hours contain data outliers. Even so, the underlaying systematic logic remains robust, the price aware baseline models adjusted their output downwards during low hours, whereas the technical potential price aware models heightened their output during profitable hours based on meteorological conditions.

9.3 Regional characteristics in production behaviour

Research question 3: *In what ways do the regional production behaviours differ across the four Swedish bidding areas, when market influences are removed?*

Based on the results procured by the reconstructed technical maximal potential, we can draw some critical conclusions to address the third research question. The calculation of potential power generation lost in each bidding area reveal substantial regional variations and disparate operational responses to market price signals. The most prominent indication of large scale curtailment was present in bidding area SE2, which accounted for 64.77% (1.72 TWh) of the total curtailed volume of 2.9 TWh. Bidding area SE1 also exhibited notable curtailment, comprising 23.16% (0.67 TWh) of the total volume. The southernmost bidding areas demonstrated significantly less volumes of down regulation, representing only 7.50% (0.22 TWh) and 4.57% (0.13 TWh) of the total curtailment, in SE3 and SE4 respectively.

A potential explanation for this large difference between the northern and southern areas, could possibly be found in the fundamental characteristic of the Swedish power grid. These regional disparities are intrinsically linked to the structural structure of the Swedish transmission grid, which is characterized by bidding area capacity bottlenecks as explained by Svenska kraftnät (2025e), meaning the grid at times have insufficient capacity for transportation of power. Additionally there is, as previously presented, often an abundance of available power in the northern areas and a deficit in production in combination with high consumption needs in the southern areas, which may influence the differences.

As observed in Table 6, there are less hours with prices below 3 €/MWh present in the yearly data for SE3 and SE4, than for the northern regions. For instance the largest difference is between SE2 which had the largest number of these sub threshold hours, 2626 hours in total, while SE4, which experienced the least hours under these conditions only amounted to 920. This could mean, that since the frequency of low priced hours is higher within SE2, that they have to adhere more to the market dynamics to mitigate penalties during the lowest priced hours, than in SE4 where the phenomenon is less widespread. The same conclusion could possibly be drawn for SE3, which experienced 983 hours below the 3 €/MWh limit.

During an in depth examination the distribution of occurrences below 3 €/MWh, it is further clear that SE1 and SE2 experienced larger hours with negative prices than the southern regions. In total SE1 and SE2 experienced 543 and 724 hours where the prices fell below 0 €/MWh respectively, while the same condition took place 365 and 387 times for SE3 respective SE4. This could further explain the regional differences market curtailment practices. Alternative explanatory variables include regional variations in asset configurations, such as turbine capacities, turbine sizes and differences in ownership structures.

9.4 Curtailment volumes of 2025, and system implications

Research question 4: *To what degree does the reconstructed technical potential exceed realized production, and what are the implications for the Swedish grid system?*

As presented earlier, the total estimated loss of generation from wind power during 2025 from all four bidding areas, due to market influence, amounted to a significant volume of approximately 2.90 TWh. In total the production from all energy sources in 2025 equaled 190 TWh, of which 39.0 TWh originated from wind power assets. Thus a volume of 2.90 TWh is a significant number, and if it was realized it would have increased the total production by $\approx 1.53\%$ and the wind power sourced production by $\approx 7.43\%$.

This, in turn, means that the inclusion of this curtailed volume into the realized potential would have meant a large amount of possible energy to utilize. Enabling the acquisition of this hidden potential is imperative for reaching global and national electrification targets and accelerating the electrification. One of which is to be able to meet a projected demand of 300 TWh for 2045, as formulated by Regeringskansliet (2024).

In accordance with Ejuh Che et al. (2025), a usable strategy with VRE, such as wind power, is to use efficient energy storage solutions, which in relation to the material presented in this work could mean an increase of production when wind speeds exceeds the technical cut in thresholds, $v_{\text{cut\_in}}$, under unprofitable periods.

Other ways to improve the reality of this potential is the reinforcement of the Swedish transmission grid. This would enable larger quantities of volumes to be transformed across areas with an abundance of production, to where it is needed, thereby alleviating localized low pricing and. This could possibly impact the SE2 bidding area, which in 2025 experienced high levels of curtailment, which then could possibly have been sold and exported outside the region.

9.5 Limitations within the study

The content of this study can be argued to have the following limitations, which should be kept in mind in relationship to the conclusions that have been drawn.

Since the study only focuses on production and prices for 2025, information about previous years is lost. This could be a potential weakness, as there is no way to compare results between years, as operational behaviours, wind conditions and pricing outlooks may transform with time. The choice to perform a one year analysis was taken in regards to time constraints of the project. Other reasons lie within lower data availability for years preceding 2025.

Another potential limitation of the topic of temporal usage is the exclusion of all hours below the 3 €/MWh threshold in the training datasets. Thus there is a potential risk that the models are not trained on some meteorological patterns. In accordance with Sandberg (2024), high supply of wind conditions often coincide with lower prices, as the available production of wind power can meet demand thus saturating the market. This means that eliminating these hours could mean hours with higher wind speeds are lost. Furthermore, the fact that the analysis is done at a 1 hour frequency could have several implications, as the nature of VRE, and wind power in particular, is characterized by quick fluctuations. As derived from Equation 1, the relationship between wind speed and power output is cubic, thus a fast change in wind speed during an hour has a large impact on the realised production. Furthermore this temporal resolution, where the market prices have been aggregated into the 1 hour frequency, could as a consequence average out the fast changes of price where the prices averages out prices. This could mean that an hour experiencing negative prices and fast goes up could be included in the training data if the average is above the threshold. As a consequence this smoothing may introduce a systematic underestimation of the true curtailment frequency and magnitude present within individual hours.

A final limitation in relation to the data usage, is related to the management of the meteorological features used. The bidding areas in Sweden are, as seen in Figure 1 geographically vast , but using all available data measurement points available is not computationally feasible within the constraints of this project. Thus, Principal Component Analysis, PCA, was employed to create five vectors, representing the values of each weather variable within each bidding area for every hour, which the model then could train on. While this reduces computational overhead and the feature set size, it also means local information is omitted. While this could be a problem for models of individual farms, the model performance of the aggregated regional production during the five-fold cross validation still validates that these five representative vectors contain enough information for the model to create a prediction for the area as a whole.

It is further necessary to acknowledge that both the initial HGBR models and the final recreated potential models created using these model predictions, remain susceptible and can still be sensitive to external factors not captured in this thesis. For instance, although the models incorporate the technical availability of regional wind power assets, through the utilization of Urgent Market Messages (UMMs) data, this dataset does not guarantee a comprehensive reporting of all outages, introducing potential data limitations. Additionally, dynamics from other markets, such as the intra day and ancillary markets, may include operational impacts that fell out of the scope of this research framework. As presented in Chapter 4.3.1, an increasing volume of wind power is participating on the ancillary markets, something that is not taken regard to in this work, but could have a large impact on the results.

Moreover, the thesis treats the bidding areas as isolated regions, only considering the regional Day-Ahead price and not the conditions in related areas. This structure of analyis simplifies the dynamics present between the regions and within the Nordic synchrounus system, since the trade between areas is frequent. In addition to this, another relation that is overlooked in this work is the status of the whole Nordic synchronous system and within the EU. Thus treating the bidding areas as individual models could impact model performance and accuracy of results.

Thus, the conclusions and volumetric estimations produced by this historical reconstruction must be interpreted with caution, as attributing curtailment solely to Day-Ahead prices simplifies a highly complex electrical system. However, given the substantial statistical deviations observed during the negative and low price intervals, it remains highly arguable that the Day-Ahead price exerts a dominant influence on regional wind power regulation.

9.6 Future work

In relation to the limitations previously presented, as well as the literature that was studied in the context of this project, several suggestions of possible future work will now be presented.

The structure of examination within this thesis is to investigate the aggregated production on a regional bidding area level. Future work could benefit in investigating curtailment behaviour on a wind farm level. Applying this framework at a the individual farm level would provide asset specific behavioural operational patterns, which are lost at the regional aggregation level. Performing the modelling on a farm level would also make it possible to use the specific local meteorological conditions, possibly further improving the model accuracy. This could further be performed as an comparative study, investigating farms of different sizes, ownership and geographical position.

A further suggestion, to investigate individual farm behaviour and economic optimization is to conduct a qualitative study as a complement to the technical aspects present in this work. By introducing interviews and correspondence with actors which are involved with the operation of wind power production assets or the power grid, such as owners, BRP's and TSO representatives, a deeper understanding and stronger conclusions could possibly be gained.

Conversely, future research may also perform the analysis on an even larger scale, considering the whole Nordic synchronous system and the import and export between regions and countries. This could make the model more representative of the power system as it is more complex than it, due to constraints, was modelled in this work. An examination between differences in countries, could further be very interesting, as different power mixes present in each country could as a consequence affect the operational strategies of wind power operators. This could draw conclusions on optimal strategies and system structures.

As discussed in the limitations section in Chapter 9.5, this study was performed at an hourly frequency. Additional studies may employ and utilize the current MTU of 15 minutes, to better represent the in time variations which are now lost when only a one hour frequency is used.

Other examples of what could be done within this subject, could be to elaborate the usage of the SHAP framework. One experimental idea which may prove useful is to conduct a more in-depth analysis using the SHAP impact of price, to evaluate how specific marginal shifts in price features mathematically drive the models predictions. Another possible comparative study, frequent in the WPF literature, is the comparison of a variety of machine learning models to see which obtains the best performance. As mentioned there are multiple different types of models which are often used for wind power production, such as different configurations of Artificial Neural Networks (ANN's) and XGBoost.

Further work should possibly, regardless of being a historical reconstruction of power production, hold out data for validation during the first model, to make sure of its ability to generalize to unseen data under profitable conditions (as this study trains and tests on the same data for the high price data subset).

10. Conclusions

This thesis has demonstrated the viability and possibility of using a machine learning framework, specifically the Histogram-Based Gradient Boosting Regression (HGBR) ML model, to reconstruct the historical technical potential of wind power production in Sweden's four bidding areas, in the absence of market influences.

The final technical potential reconstruction further reveal that market driven curtailment accounts for an estimated total loss of approximately 2.90 TWh across the bidding areas during the year of 2025. Interestingly, the absolute majority of this volume, 2.53 TWh or 87.31% of the total technical potential loss, occured during hours when the Day-Ahead prices fell below the threshold of 3 €/MWh.

From a regional perspective the curtailment of volume is heavily concentrated in the northern regions. In bidding areas SE1 and SE2, the volume of curtailed energy represents an operational loss of nearly 10% of each regions respective total wind power generation for 2025. Out of all bidding areas it was SE2 that exhibited the absolute highest levels of downregulated capacity, amounting to 132.40 GWh during the whole year, of which 110.19 GWh were curtailed specifically during hours below the 3 €/MWh threshold. These insights underscore the large impact of capacity transmission bottlenecks and market dynamics on the wind power production. Ultimately as Sweden works toward their national targets of providing more energy and reaching net zero carbon emissions in the coming decades, mitigating these market induced inefficiencies through grid enforcements and localized energy storage solutions will be imperative to fully achieving the nations technical wind potential.

References

Ackermann, T., (ed.) (2012). *Wind Power in Power Systems*. 2nd. ProQuest Ebook Central. John Wiley & Sons, Incorporated. https://ebookcentral.proquest.com/lib/uu/detail.action?docID=902783.

Akiba, T. et al. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. https://api.semanticscholar.org/CorpusID:196194314.

Alex, K. (Dec. 2017). Wind Power Fundamentals. In: *Wind Energy Engineering: A Handbook for Onshore and Offshore Wind Turbines*. Academic Press, pp. 17–24. ISBN: 9780128094518. DOI: 10.1016/B978-0-12-809451-8.00002-3.

Almqvist, E. (2023). *Vindkraftens möjligheter att delta på stödtjänstmarknaden*. https://www.svk.se/siteassets/om-oss/organisation/vara-rad/driftradet/2023/bilaga-5-vindkraftens-mojligheter-att-delta-pa-stodtjanstmarknaden-2023-11-29.pdf.

Aslan, E. et al. (2025). A Hybrid Machine Learning Approach for Predicting Power Transformer Failures Using Internet of Things-Based Monitoring and Explainable Artificial Intelligence. In: *IEEE Access* 13, pp. 113618–113633. https://api.semanticscholar.org/CorpusID:279652458.

Bilendo, F. et al. (2023). Applications and Modeling Techniques of Wind Turbine Power Curve for Wind Farms—A Review. In: *Energies* 16.(1), p. 180. ISSN: 1996-1073. DOI: 10.3390/en16010180. https://www.mdpi.com/1996-1073/16/1/180.

Brolin, M., Hamon, C., and Nyström, S. (2021). *Intradagmarknaden : En generell beskrivning av intradagmarknadens funktion*. Report 2021:100. Energiforsk. https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-56725.

Cai, Y. et al. (2024). The Optuna–LightGBM–XGBoost Model: A Novel Approach for Estimating Carbon Emissions Based on the Electricity–Carbon Nexus. In: *Applied Sciences* 14.(11), p. 4632. ISSN: 2076-3417. DOI: 10.3390/app14114632. https://www.mdpi.com/2076-3417/14/11/4632.

Dubey, A. K. et al. (2026). Machine learning and hybrid intelligence for wind energy optimization: A comprehensive state-of-the-art review. In: *Expert Systems with Applications* 296, p. 128926. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2025.128926. https://www.sciencedirect.com/science/article/pii/S0957417425025436.

Ejuh Che, E. et al. (2025). The Impact of Integrating Variable Renewable Energy Sources into Grid-Connected Power Systems: Challenges, Mitigation Strategies, and Prospects. In: *Energies*

18.(3), p. 689. ISSN: 1996-1073. DOI: 10.3390/en18030689. https://www.mdpi.com/1996-1073/18/3/689.

Energimarknadsinspektionen (2025). *Elmarknaden*. https://ei.se/konsument/el/elmarknaden [Accessed: 05/14/2026].

Energimyndigheten (2026). *Databas för energirelaterad statistik: Antal verk, installerad effekt och vindkraftproduktion per elområde, 2003-*. https://pxexternal.energimyndigheten.se/pxweb/sv/Energimyndighetens_statistikdatabas/Energimyndighetens_statistikdatabas__Officiell_energistatistik__Vindkraftsstatistik/EN0105_2.px/?rxid=500b004e-7586-4511-91a7-c5f07d179919 [Accessed: 05/15/2026].

eSett Oy (2026). *eSett Open Data: Production*. Open Data Service. https://opendata.esett.com/production [Accessed: 05/16/2026].

European Court of Auditors (2023). *Integrationen av den inre marknaden för el: En komplex rättslig struktur, förseningar, bristfällig styrning och ofullständig marknadskontroll gör att det ambitiösa målet inte uppnås fullt ut*. Tech. rep. Special Report 03/2023. European Court of Auditors. https://www.eca.europa.eu.

Gong, Y. et al. (2023). A survey on dataset quality in machine learning. In: *Information and Software Technology* 162, p. 107268. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2023.107268. https://www.sciencedirect.com/science/article/pii/S0950584923001222.

Green Power Sweden (2023). *Vindkraften stöttar redan idag stabilitet i elnätet – kan och vill bidra än mer*. https://greenpowersweden.se/vindkraften-stottar-redan-idag-stabilitet-i-elnatet-kan-och-vill-bidra-an-mer/ [Accessed: 02/09/2026].

Hamon, C. and Persson, M. (2023). *Wind power participation in frequency regulation: a profitability assessment for Sweden*. Tech. rep. 2022:42. RISE Research Institutes of Sweden, Measurement Technology.

Hanifi, S., Cammarono, A., and Zare-Behtash, H. (2024). Advanced hyperparameter optimization of deep learning models for wind power prediction. In: *Renewable Energy* 221, p. 119700. ISSN: 0960-1481. DOI: 10.1016/j.renene.2023.119700. https://www.sciencedirect.com/science/article/pii/S0960148123016154.

Hanifi, S., Liu, X., et al. (2020). A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. In: *Energies* 13.(15), p. 3764. ISSN: 1996-1073. DOI: 10.3390/en13153764. https://www.mdpi.com/1996-1073/13/15/3764.

Heden, H. (2012). *Energimarknadsinspektionen - en sekellång historia*. Energimarknadsinspektionen.

International Energy Agency (2025). *Renewables 2025*. Tech. rep. Paris: IEA. https://www.iea.org/reports/renewables-2025.

Ke, G. et al. (2017). "LightGBM: a highly efficient gradient boosting decision tree". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., pp. 3149–3157. ISBN: 9781510860964.

Koukaras, P. et al. (2024). Optimizing Building Short-Term Load Forecasting: A Comparative Analysis of Machine Learning Models. In: *Energies* 17.(6), p. 1450. ISSN: 1996-1073. DOI: 10.3390/en17061450. https://www.mdpi.com/1996-1073/17/6/1450.

Leskovec, J., Rajaraman, A., and Ullman, J. D. (2014). *Mining of Massive Datasets*. Second. Cambridge University Press. http://mmds.org.

Liu, Z. et al. (2025). A Comprehensive Review of Wind Power Prediction Based on Machine Learning: Models, Applications, and Challenges. In: *Energies* 18.(2), p. 350. ISSN: 1996-1073. DOI: 10.3390/en18020350. https://www.mdpi.com/1996-1073/18/2/350.

Möllerström, E., Gipe, P., and Ottermo, F. (2025). Wind power development: A historical review. In: *Wind Engineering* 49.(2), pp. 499–512. DOI: 10.1177/0309524X241260061. https://doi.org/10.1177/0309524X241260061.

Nair, K., Vanitha, V., and Mohan, J. (July 2017). "Forecasting of wind speed using ANN, ARIMA and Hybrid models". In: *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, pp. 170–175. DOI: 10.1109/ICICICT1.2017.8342555.

National Centers for Environmental Information (2026). *Global Forecast System*. https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast [Accessed: 02/09/2026].

Nord Pool (2025). *Market Coupling Steering Committee confirms go-live of 15-Minute MTU in SDAC on trading day 30 September 2025 for delivery day 1 October 2025*. Nord Pool Newsroom. https://www.nordpoolgroup.com/en/message-center-container/newsroom/exchange-message-list/2025/q3/market-coupling-steering-committee-confirms-go-live-of--15-minute-mtu-in-sdac-on-trading-day-30-september-2025-for-delivery-day-1-october-2025/ [Accessed: 05/29/2026].

Nord Pool (2026). *Day-ahead market*. https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/ [Accessed: 05/09/2026].

Ponce-Bobadilla, A. V. et al. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. In: *Clinical and Translational Science* 17.(11), e70056. DOI: 10.1111/cts.70056. https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1111/cts.70056.

Regeringskansliet (2024). *Regeringens proposition 2023/24:105, Energipolitikens långsiktiga inriktning*. https://www.regeringen.se/contentassets/2fd0739890d8484b8129d3c0e678f24d/energipolitikens-langsiktiga-inriktning-prop.-202324105.pdf [Accessed: 06/05/2026].

Regeringskansliet (2026). *Elproduktion i Sverige*. https://www.regeringen.se/regeringens-politik/sveriges-elforsorjning/elproduktion-i-sverige/ [Accessed: 02/09/2026].

Ren, Y., Suganthan, P., and Srikanth, N. (2015). Ensemble methods for wind and solar power forecasting—A state-of-the-art review. In: *Renewable and Sustainable Energy Reviews* 50, pp. 82–91. ISSN: 1364-0321. DOI: 10.1016/j.rser.2015.04.081. https://www.sciencedirect.com/science/article/pii/S1364032115003512.

Sandberg, R. (2024). *Wind and Electricity Prices in Sweden – a Statistical Analysis*. Tech. rep. Energiforsk. https://energiforsk.se/media/33100/2024-975-wind-and-electricity-prices-in-sweden-a-statistical-analysis.pdf.

Saxena, A. et al. (2024). Intelligent Load Forecasting and Renewable Energy Integration for Enhanced Grid Reliability. In: *IEEE Transactions on Industry Applications* 60.(6), pp. 8403–8417. DOI: 10.1109/TIA.2024.3436471.

scikit-learn (2026). *1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking*. https://scikit-learn.org/stable/modules/ensemble.html#histogram-based-gradient-boosting [Accessed: 04/20/2026].

Shafiullah, G. et al. (2013). Potential challenges of integrating large-scale wind energy into the power grid–A review. In: *Renewable and Sustainable Energy Reviews* 20, pp. 306–321. ISSN: 1364-0321. DOI: 10.1016/j.rser.2012.11.057. https://www.sciencedirect.com/science/article/pii/S1364032112006715.

Soman, S. S. et al. (2010). "A review of wind power and wind speed forecasting methods with different time horizons". In: *North American Power Symposium 2010*, pp. 1–8. DOI: 10.1109/NAPS.2010.5619586.

Statens energimyndighet (2026). *En samlad analys och bedömning av energiomställningen*. Tech. rep. Statens energimyndighet.

Suárez-Cetrulo, A. L. et al. (2022). Wind power forecasting using ensemble learning for day-ahead energy trading. In: *Renewable Energy* 191, pp. 685–698. ISSN: 0960-1481. DOI: 10.1016/j.renene.2022.04.032. https://www.sciencedirect.com/science/article/pii/S0960148122004931.

Svenska kraftnät (2021). *Systemutvecklingsplan 2022-2031: Vägen mot en dubblerad elanvändning*. Tech. rep. Svenska kraftnät.

Svenska kraftnät (2024a). *Balancing market outlook 2030*. Tech. rep. Svenska kraftnät. https://www.svk.se/48d672/siteassets/om-oss/rapporter/2024/svenska-kraftnat-balancing-market-outlook-2030.pdf.

Svenska kraftnät (2024b). *Sveriges elnät*. https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/sveriges-elnat/ [Accessed: 05/05/2026].

Svenska kraftnät (2025a). *Avräkning*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/avrakning/ [Accessed: 05/11/2026].

Svenska kraftnät (2025b). *Balansering av kraftsystemet*. https://www.svk.se/om-kraftsystemet/om-systemansvaret/balansering-av-kraftsystemet/ [Accessed: 05/05/2026].

Svenska kraftnät (2025c). *Beskrivning av statistiken*. https://www.svk.se/om-kraftsystemet/kraftsystemdata/elstatistik/beskrivning-av-statistiken/ [Accessed: 05/06/2026].

Svenska kraftnät (2025d). *Dagen före-marknaden – fysisk handel med el*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/dagen-fore-marknaden--fysisk-handel-med-el/ [Accessed: 05/09/2026].

Svenska kraftnät (2025e). *Elområden*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/ [Accessed: 05/14/2026].

Svenska kraftnät (2025f). *Förhandsmarknaden - prissäkringsmarknad*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/forhandsmarknaden--prissakringsmarknad/ [Accessed: 05/10/2026].

Svenska kraftnät (2025g). *Intradagsmarknaden - justering av dagen före handel*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/intradagsmarknaden--justering-av-dagen-fore-handel/ [Accessed: 05/09/2026].

Svenska kraftnät (2025h). *Network development plan 2026-2035*. Report. Svenska kraftnät. https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-56725.

Svenska kraftnät (2025i). *Om elmarknaden*. https://www.svk.se/om-kraftsystemet/om-elmarknaden/ [Accessed: 05/07/2026].

Svenska kraftnät (2025j). *Så påverkar olika kraftslag systemet*. https://www.svk.se/om-kraftsystemet/sa-paverkar-olika-kraftslag-systemet/ [Accessed: 05/07/2026].

Svenska kraftnät (2025k). *Stödtjänster och avhjälpande åtgärder*. https://www.svk.se/om-kraftsystemet/om-systemansvaret/verktyg-for-systemdrift/stodtjanster-och-avhjalpande-atgarder/ [Accessed: 05/11/2026].

Svenska kraftnät (2026a). *Elstatistik*. https://www.svk.se/om-kraftsystemet/kraftsystemdata/elstatistik/ [Accessed: 02/09/2026].

Svenska kraftnät (2026b). *Flödesbaserad kapacitetsberäkningsmetod*. https://www.svk.se/utveckling-av-kraftsystemet/systemansvar--elmarknad/flodesbaserad-kapacitetsberakningsmetod/ [Accessed: 05/09/2026].

Sveriges meteorologiska och hydrologiska institut (2026). *Meteorologiska modeller*. https://www.smhi.se/kunskapsbanken/meteorologi/meteorologiska-modeller [Accessed: 02/09/2026].

Sveriges riksdag (2025). *Förordning (2025:782) med instruktion för Affärsverket Svenska kraftnät*. Svensk författningssamling. SFS 2025:782. Issued June 26, 2025.

SWECO (2023). *De svenska marknaderna för stödtjänster och vindkraftens möjlighet till deltagande*. Report from SWECO as assignment from Swedish Wind Energy Association. https://greenpowersweden.se/vindkraften-stottar-redan-idag-stabilitet-i-elnatet-kan-och-vill-bidra-an-mer/.

T.A., R. and Columbus, C. (July 2025). Enhanced wind power forecasting using machine learning, deep learning models and ensemble integration. In: *Scientific Reports* 15. doi: 10.1038/s41598-025-05250-3.

The European Centre for Medium-Range Weather Forecasts (2026). *Forecast charts and data*. https://www.ecmwf.int/en/forecasts [Accessed: 02/09/2026].

Wang, J. et al. (Mar. 2024). Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. In: *Journal of the Knowledge Economy* 15.(1), pp. 1159–1178. doi: 10.1007/s13132-022-01096-6. https://ideas.repec.org/a/spr/jknowl/v15y2024i1d10.1007_s13132-022-01096-6.html.

Wu, Y.-K. and Hong, J.-S. (2007). “A literature review of wind forecasting technology in the world”. In: *2007 IEEE Lausanne Power Tech*, pp. 504–509. doi: 10.1109/PCT.2007.4538368.

Yan, J. et al. (2022). Uncovering wind power forecasting uncertainty sources and their propagation through the whole modelling chain. In: *Renewable and Sustainable Energy Reviews* 165, p. 112519. ISSN: 1364-0321. DOI: 10.1016/j.rser.2022.112519. https://www.sciencedirect.com/science/article/pii/S1364032122004221.

Yang, M. et al. (2025). Review of several key processes in wind power forecasting: Mathematical formulations, scientific problems, and logical relations. In: *Applied Energy* 377, p. 124631. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2024.124631. https://www.sciencedirect.com/science/article/pii/S0306261924020142.

Zeier, I. (2026). *How is the electricity price calculated?* https://www.nordpoolgroup.com/en/the-power-market/how-is-the-electricity-price-calculated/ [Accessed: 05/11/2026].