FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Second-Level Wind Energy Forecasting

Sérgio Rafael Dias Alves

FINAL VERSION

Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Ricardo Bessa

July 15, 2024

Resumo

A previsão da produção renovável para as próximos horas e dias atingiu um nível de maturidade a nível comercial e várias empresas comercializam o serviço para operadores do sistema elétrico e mercado de eletricidade. A natureza dinâmica das previsões meteorológicas numéricas e das observações de energia em tempo real, resultam em múltiplas previsões ao longo do dia, criando problemas de decisão interessantes. Decisões críticas, como nos mercados de energia, dependem da previsão de flutuações de incerteza em previsões subsequentes - conhecidas como previsões de segundo nível. Uma previsão de segundo nível refere-se à previsão das flutuações destas incertezas em previsões subsequentes da produção de energia eólica. Uma vez que as previsões meteorológicas numéricas (PNT) e as observações de energia em tempo real são atualizadas ao longo do dia, a previsão de segundo nível centra-se na antecipação das flutuações na previsão de produção de energia eólica prevista para horizontes temporais futuros, normalmente de 6 em 6 horas.

Nesta dissertação, é utilizado um método de previsão de energia eólica de segundo nível para quantificar estas flutuações nas previsões de incerteza para a produção de energia eólica e para estudar de que forma esta vai impactar a tomada de decisão nas diferente áreas de atuação do sector da energia. Por exemplo, para aferir se o operador deve ativar toda a flexibilidade disponível ou não, o conhecimento da variação da incerteza das previsões futuras permite aos operadores de mercado tomar uma decisão mais informada sobre se devem aguardar pela próxima previsão de produção e adiar a decisão.

Esta dissertação tem como objetivo desenvolver um algoritmo de aprendizagem automática para previsões de energia eólica de segundo nível, com foco em atualizações de previsões que ocorrem a cada 6 horas. Utilizando como recurso dados de parques eólicos na Noruega e PNT do modelo da Europa Central, espera-se desenvolver uma metodologia baseada em gradient boosting trees (GBT) para prever flutuações em diferentes atualizações de produção eólica. A investigação visa contribuir com conhecimentos e ferramentas valiosos para os decisores no sector das energias renováveis, oferecendo capacidades melhoradas para navegar nos meandros dos problemas de decisão em tempo real e otimizar as estratégias operacionais em resposta à evolução das informações de previsão.

Palavras Chave: Previsão de segundo nível, energia eólica, aprendizagem automática, previsão de incertezas, flutuações das previsões

Abstract

Renewable energy production forecasting has matured to a commercial level, with services offered to electricity system operators and the market. The dynamic nature of numerical weather forecasts and real-time power observations results in multiple forecasts throughout the day, posing various decision challenges. Critical decisions, such as in energy markets, hinge on the forecast of fluctuations of uncertainty in subsequent predictions—known as second-level forecasts. A second-level forecast refers to the prediction of fluctuations of these uncertainties in subsequent wind energy production forecasts. As numerical weather predictions (NWP) and real-time power observations are updated throughout the day, the second-level forecast focuses on anticipating variations in the forecasted wind energy production for future time horizons, typically every 6 hours.

For this dissertation, a second-level wind energy forecasting method is employed to quantify these fluctuations in uncertainty forecasts for wind power output and to study its contribution to decision-makers in the energy sector. For example in economic dispatches knowing how future forecast uncertainty will vary, allows market operators to make a more informed decision on whether should he dispatch all available production now or wait for the next forecast and postpone the decision.

This dissertation aims to develop and assess various machine learning algorithms for second-level wind energy forecasts, with a focus on forecast updates occurring every 6 hours. Leveraging data from wind farms in Norway and NWP from the Central European model, a methodology based on gradient boosting trees (GBT) to forecast fluctuations in different wind production updates is expected to be developed. The research aims to contribute valuable insights and tools for decision-makers in the renewable energy sector, offering enhanced capabilities to navigate the intricacies of real-time decision problems and optimize operational strategies in response to evolving forecast information.

Keywords: Second-level forecasting, wind energy, machine learning, uncertainty forecasting, forecast fluctuations

Sustainable Development Goals (SDGs) of the United Nations

The United Nations Sustainable Development Goals (SDGs) provide a comprehensive framework for tackling global challenges, including energy sustainability and climate action.

This section explores the performance indicators related to specific SDGs and establishes metrics to measure the impact of second-level wind energy forecasting in contributing to these goals. The SDGs of focus are presented in the following table 1:

Table 1: United Nations Sustainable Development Goals

SDG	Target	Contribution	Performance indicators and metrics
7	2	By 2030, increase substantially the share of renewable energy in the global energy mix	Measure the proportion of energy supplied by wind power in the regional or national energy mix before and after implementing advanced wind energy forecasting models.
	b	By 2030, expand infrastructure and upgrade technology for supplying modern and sustainable energy services for all in developing countries, in particular least developed countries, small island developing States, and land-locked developing countries, in accordance with their respective programmes of support	Track the number and scale of infrastructure projects (e.g., new wind farms, grid integration projects) facilitated by better forecasting and measuring the increase in the number of people in developing regions with access to reliable wind energy due to enhanced forecasting
13	2	Integrate climate change measures into national policies, strategies and planning	Assess the extent to which improved forecasting data is integrated into national climate and energy policies and estimate the reduction in carbon emissions achieved by optimizing wind energy usage via better wind energy forecasting accuracy

The rationale behind the metrics chosen for the each SDG is as follows:

SDG 7.2

- **Proportion of Energy Supplied by Wind Power:** This metric directly measures the impact of improved forecasting on the adoption of wind energy, reflecting progress toward increasing the share of renewables in the energy mix.

- **Accuracy of Wind Energy Forecasts:** By tracking forecasting accuracy, stakeholders can evaluate the effectiveness of the second-level forecasting models and their role in enhancing energy reliability and stability.

SDG 7.b

- **Number and Scale of Infrastructure Projects:** This indicator captures the tangible outcomes of improved forecasting in terms of infrastructure development, demonstrating the practical benefits of enhanced forecasting capabilities.
- **Access to Reliable Wind Energy:** Expanding access to sustainable energy is a core goal of SDG 7, and this metric measures the real-world impact of forecasting on energy accessibility in developing regions.

SDG 13.2

- **Integration into National Policies:** The inclusion of forecasting data in national policies indicates a strategic commitment to leveraging technology for climate action, aligning with SDG 13's objectives.
- **Reduction in Carbon Emissions:** This metric quantifies the environmental benefits of optimized wind energy usage, providing a clear link between forecasting improvements and climate action outcomes.

By employing these performance indicators and metrics, the dissertation can effectively evaluate and demonstrate the contributions of second-level wind energy forecasting to the relevant SDGs, offering a comprehensive assessment of its impact on sustainable development goals.

Agradecimentos

Gostaria de expressar o meu agradecimento a todas as pessoas que de alguma forma me ajudaram, não só durante a realização do presente trabalho, como também no meu desenvolvimento académico e pessoal.

Em primeiro lugar e acima de tudo, estendo o meu mais profundo agradecimento aos meus pais, cujo apoio inabalável e encorajamento têm sido a base do meu sucesso. Estou eternamente grato por todas as oportunidades que me proporcionaram.

Ao meu orientador, Ricardo Bessa pela qualidade da sua orientação, pelos conhecimentos partilhados e pelas sugestões que enriqueceram o trabalho desenvolvido. Quero também expressar a minha gratidão ao João Paulo Viana, investigador do Centro de Sistemas de Energia do INESC-TEC. A sua orientação, apoio e experiência foram fundamentais para a realização deste trabalho. Estou genuinamente grato pela paciência e disponibilidade para me ajudar a ultrapassar desafios técnicos.

À Francisca Valente e à Joana Dias, vocês são o expoente máximo do espírito de equipa e de colaboração. Foi um prazer aprender ao vosso lado e tornar-me vosso amigo.

Aos meus camaradas de trabalhos e amigos Tiago, Rui, Nuno, Fábio e Renato pelo companheirismo, as jantaradas e os bons momentos que foram fundamentais para tornar toda esta jornada de 5 anos uma experiência que vai deixar saudades.

"Nothing would be more tiresome than eating and drinking if God had not made them a pleasure as well as a necessity"

Voltaire

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

PNT	Previsões Meteorológicas Numéricas
ECMWF	European Centre for Medium-Range Weather Forecasts
ED-ANN	Encoder-Decoder Artificial Neural Network
EnKTF	Ensemble Kalman Transor Filter
LS	Least Squares
LSTM	Long Short-Term Memory
MAC	Mean Absolute Change
MAE	Mean Absolute Error
ML	Maximum Likelihood
MSE	Mean Squared Error
GBM	Gradient Boosting Machines
GBT	Gradient Boosting Trees
NWP	Numerical Weather Preditcions
PDF	Probability Density Function
RES	Renewable Energy Sources
RMSC	Root Mean Squared Change
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
SO	System Operator
SGD	Sustainable Development Goal

Chapter 1

Introduction

As the level of integration of renewable energies increases, both power system operators and energy market operators face new challenges caused by the variability and uncertainty of power output, which is strongly dependent on local meteorological conditions. As such Wind power forecasting has been utilized to predict the uncertainty associated with wind power output in order to contradict the challenges imposed by the variability of the power output. For example for a distribution system operator, when integrating renewable energy sources into the grid, the variability in their power output can cause over/under-voltages in a given network node. Being able to forecast power generation uncertainty ahead of time gives the system operator insight on whether should he take preventive measures now based on a forecast with higher uncertainty or wait for a later and more accurate forecast and decide later but risk having fewer options available to solve grid congestion problems. Throughout the day, various forecasts are released due to updates of NWP.

Presently renewable energy forecasting has already achieved a certain level of maturity in its commercial use [1], which has traditionally taken form in single point forecasts even though these may not provide the most reliable information regarding future forecasts. As such using advanced forecasting techniques, such as uncertainty forecasts can provide additional information for these cases and aid forecast users in the decision-making process [2].

A second-level forecasting model approach is employed to forecast the uncertainty (represented by a set of quantiles) and fluctuations of subsequent forecasts. Having more knowledge on how the uncertainty will vary in future weather forecasts allows forecast users to decide whether should they act now on the basis of the current forecast or to wait for the next forecast in the hopes of it being more accurate and postpone the decision [3].

Context

Datasets containing numerical weather predictions are typically generated every 6 hours on a certain day. As such the target value can be forecasted for a certain time horizon, using datasets generated on the day D, at 00h, 06h, 12h or 18h, as illustrated below.

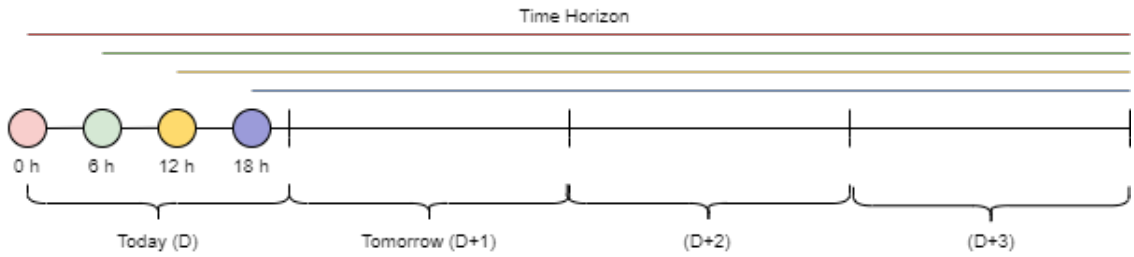

Figure 1.1: Second-Level concept ilustration

It is worth mentioning that, even though, forecasts that are associated with a shorter lead time are characterized by being more accurate [4], it is also incurred by the risk of losing options that were available before if action was taken earlier on the basis of a less accurate forecast with a longer lead time. These decision problems can be complex to assess and can carry a significant cognitive load on the user.

As such, a Second-Level wind energy forecasting method is proposed, where the uncertainty surrounding wind energy is quantified for a time horizon to be defined. Although the focus of this dissertation is on wind energy forecasting, the second-level forecasting concept can also apply to solar energy forecasting.

The term Second-Level refers to the requirement that, additionally, the algorithm needs to be able to forecast how the uncertainty fluctuates in future forecasts [3]. For instance, a forecast for the day D+1 is performed using NWP generated at 00h00 in day D, it is expected to obtain probabilistic forecasts the time horizon mentioned but also how the uncertainty of the same target values, varies if the forecast was conducted using datasets of NWP generated at 06h00, 12h00 or 18h00 on the day D. The implemented algorithm is based on the non-parametric GBT algorithm using as inputs raw NWP data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and new variables created via feature engineering.

Objectives

The goal is to establish a methodology using machine learning to forecast how probabilistic wind forecasts change with each update of the Numerical Weather Prediction (NWP) that occurs every 6 hours.

The developed models will enable the forecast user to anticipate the wind production forecasts released at different times of the day, essentially creating a forecast of the forecasts. This introduces the concept of second-level forecasting, where forecasts generated at the first level are used as input data. To achieve this, the following objectives have been defined:

- A study of the available data generated by the ECMWF, analyzing correlations between the different NWP variables and the production history for each wind farm, and assessing which variables prove to have predictive capacity for forecasting wind production;

- Obtain first-level probabilistic forecasting models for 3 wind farms in Norway for a 48-hour time horizon. Generate new forecasts every 6 hours until delivery time;
- Obtain second-level forecasting models using the first-level forecast results as input data to estimate the forecast fluctuations during each NWP update;
- Quantitative assessment of forecast errors and qualitative analysis of model effectiveness in decision support for electricity system operators and market participants by comparing the accuracy of the second-level forecast to a relevant benchmark model;

Document Structure

This document is structured as follows.

Chapter 1 presents the introduction, context, and rationale behind the concept of second-level forecasting.

Chapter 2 serves as the foundation for this dissertation, providing a comprehensive exploration of the existing knowledge landscape related to wind energy uncertainty forecasting. Through an extensive literature review, this chapter will dissect the historical evolution of forecasting techniques, highlighting key advancements, challenges, and emerging trends in the field.

Chapter 3 articulates the problem at hand and outlines the proposed solutions. It elucidates the intricacies of the forecasting challenge and delineates the methodologies that will be employed to address it. This chapter provides a roadmap for readers, detailing the step-by-step approach to be undertaken in solving the forecasting problem.

Chapter 4 centers around the result analysis of the first-level and second-level forecasts performed on real data.

Chapter 5 wraps the dissertation up by offering final remarks about the contributions and implications of the research.

Chapter 2

Background and State of the Art

2.1 Definition of Forecast Uncertainty

To ensure effective communication and utilization of wind power forecast uncertainty data, it is necessary to establish a common understanding of key terms. This includes differentiating between forecast error and forecast uncertainty.

Forecast Error is the deviation between a forecast value and a measured value at a specific time, either in the past or present. It can also refer to an average error measured by metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE).

Forecast uncertainty, in contrast to error, refers to the potential range of future errors or true values. It describes the spread within which the actual outcome might fall, highlighting the inherent limitations of forecasts. [5]

2.1.1 Forecast Interval

The most common way to represent forecast uncertainty is through a forecast interval [6], also known as a marginal forecast interval. Mathematically, it is expressed as:

$$Prob\{P_{t+k}^{\tau^L} \leq P_{t+k} \leq P_{t+k}^{\tau^H}\} = \tau^H - \tau^L = \alpha \tag{2.1}$$

Where:

- τ^H and τ^L are the quantile nominal proportions of the interval limits H and L; [5]
- P_{t+k} is the forecasted value at lead time $t+k$; [5]
- α is the coverage rate that corresponds to the probability of having the observed value P_{t+k} inside the forecast interval at hour $t+k$; [5]

2.1.2 Quantiles and Cumulative Distribution Function

In wind power forecasting, quantiles represent values that a future wind power measurement has a certain probability of exceeding. These quantiles are derived from the cumulative distribution

function (CDF). When creating prediction intervals, the median is typically used as the center of the interval.

2.2 Time Horizons

The time horizon is a key characteristic of a wind power forecasting system, referring to the future period for which wind generation is predicted. Similar to load forecasting, wind power forecasting can be categorized based on its time horizon: very short-term, short-term, medium-term, and long-term [7].

Very short-term forecasts have a time horizon of between a few seconds and an hour. They are mostly used to ensure the quality and stability of the network [8].

Short-term forecasts typically range from very short-term up to 48 or 72 hours. While some studies consider horizons up to 72 hours, most focus on 48 hours, and occasionally even 36 hours. This timeframe is particularly relevant for participating in day-ahead markets. For example, the Iberian Electricity Market (MIBEL) requires bids for the next day to be submitted by 10:00 AM. Therefore, a 38-hour forecast effectively covers the entire following day. It's important to note that submission deadlines for offers can vary across countries, impacting the optimal forecast horizon. Furthermore, short-term forecasts can be used for scheduling maintenance, especially when considering a 72-hour horizon [7].

Medium-term forecasting encompasses forecasts with a time horizon ranging from a few weeks to a few months. These forecasts play a crucial role in predicting future electricity availability and informing decisions around system planning.

Long-term wind power forecasting covers a timeframe of several months to a year. These forecasts are primarily used to anticipate the growth of electricity networks, enabling well-informed planning for production, transportation, and distribution infrastructure. Additionally, long-term wind power forecasting contributes to effective organization of electricity distribution, considering both energy needs and operating guarantees [9].

2.3 Uncertainty Forecasting Methods

The production of electricity from wind resources relies on natural atmospheric processes. Essentially, winds are created due to variances in atmospheric pressure. When variations in atmospheric pressure occur, air shifts from areas of higher pressure to those with lower pressure, giving rise to winds. Numerous factors, such as terrain, temperature, and friction, can influence the generation of wind resources. Consequently, the dynamic nature of wind behavior is intricate, leading to considerable variability in wind power generation [10].

The advantage of models that employ probabilistic forecasting methods over deterministic models is that their output, in addition to providing the average value, also represents the uncertainty associated with the forecast. This uncertainty can be represented in different ways, such as a confidence interval, quantiles, or a distribution of probabilities. Uncertainty can be derived

from either the variability of the weather or from deriving wind power from weather variables. Weather forecasts typically provide data on variables like wind speed and direction. Translating these variables into wind power output requires models that consider factors like the efficiency of wind turbines, the layout of wind farms, and local geographical features. Each of these factors can introduce errors and uncertainty. For example, small errors in predicting wind speed can lead to larger errors in predicting wind power because the relationship between wind speed and power output is nonlinear. Probabilistic methods help account for these uncertainties by generating a range of possible power outputs, again giving a more comprehensive picture of the potential outcomes.

2.3.1 Weather Uncertainty Forecasting

Uncertainty forecasts refer to predictions or estimates that not only provide a point estimate of a future outcome but also quantify the associated uncertainty or variability in that estimate. Bessa et al. [5] makes the distinction between forecast errors and forecast uncertainty, where forecast error is defined as the deviation between the forecast value and a measured value at a certain point in time.

Forecast uncertainty gives the range of forecasting errors (or true values) that are associated with a certain forecast. As such it is acknowledged that there are inherent uncertainties in various predictions, and these types of forecasts seek to provide a clearer picture of the potential range of outcomes of different scenarios. Reliable weather forecasting accuracy hinges on addressing various sources of uncertainty such as initial condition uncertainty, physical approximation uncertainties, and boundary condition uncertainties. Some of these uncertainties can be quantified via the following methods:

- **Singular vector methods** that are used to assess and quantify uncertainty in weather forecasts. These methods help forecasters understand how small perturbations in the initial conditions of weather models can lead to significant differences in the forecast outcomes. They identify the directions in which the forecast is most sensitive to uncertainty. This information can be used to generate ensemble forecasts, which are a collection of forecasts that are generated by perturbing the initial conditions of the forecast model in the directions of the singular vectors. The spread of the ensemble forecasts provides a measure of the uncertainty in the forecast. The larger the spread, the more uncertain the forecast is.

- **Stochastic Physics** adds a layer of randomness to the model's physical tendencies, aiming to improve the representation of uncertainty arising from the simplifications made in the NWP models. The stochastically Perturbed Parameterization Tendencies method is used to introduce perturbations aligned with the unperturbed tendencies where the multivariate distribution of the original scheme is replaced by an univariate distribution in order to achieve perturbations that are more consistent with the model physics.

- **Sequential and Local Ensemble Kalman Transor Filter (EnKTF)** is a data assimilation method that generates an ensemble of model states, each of which is perturbed from the prior state estimate. The ensemble is then propagated forward in time. Observations are then used to update the ensemble states, taking into account the model's uncertainty. For weather forecasting, for example, real-time observational data can be assimilated into the model as it becomes available to improve the accuracy of the ongoing forecast.

- **Multi-Model Approach** involves using multiple numerical weather predictions to generate a range of forecasts, with the assumption that the combination of different model systems is a good representation of model error and thus creates a good representation of uncertainty for a certain forecast.

2.3.2 Wind Power Uncertainty Forecasting

Probabilistic predictions stand as the prevailing method for expressing uncertainty in wind power forecasting. In the realm of wind power output, these forecasts typically adopt the structure of a probability density function.

Diverging from commonly employed point forecasts that yield a single expected value for wind power output, probabilistic forecasts offer supplementary quantitative insights into the associated uncertainty. The construction of predictive distributions primarily relies on two key methodologies: parametric and non-parametric approaches [10].

2.3.2.1 Parametric Probabilistic Forecasts

The parametric approach operates on the assumption that the predictive distribution adheres to a predefined shape, which can be delineated through an analytical expression [10].

The Gaussian, normal, and beta distributions are frequently employed in parametric forecasting models, although various other forms find application. Here is a generic representation of the utilized distributions:

$$\hat{d} = f(p, \mu, \sigma^2) \tag{2.2}$$

Where:

- $\hat{d}$ represents the predicted distribution as a function of p;
- p represents the central point;
- μ corresponds to the form factor;
- σ is the scale factor;

In the field of parametric forecast approaches, four aspects are subjects of much research, including:

- shape assumption of predictive distribution;

- estimator of location parameter;
- estimator of scale parameter;
- parameter evaluation theory;

As previously discussed, parametric methods are expressed through a probability distribution function. These distributions are conveniently characterized by shape and scale parameters. Selecting a shape that appropriately accommodates the renewable energy source production within the bounds of 0 to the maximum is crucial. Among the most prevalent distributions are the normal, beta, and Gaussian distributions.

For probabilistic wind power forecasts using parametric models, parameters are typically derived through the Maximum Likelihood (ML) or Least Squares (LS) methodology.

A Beta Probability Density Function (PDF) is employed to characterize the wind power forecast in the persistence model outlined by Bludszuweit et al [11]. Initially, the forecast outcomes are categorized into distinct power classes or bins to represent the power distribution using the Beta PDF. The aggregate forecast error PDF is derived by combining the calculated error PDFs for each bin. The outcomes underscore the efficacy of the suggested forecast error PDF in determining the optimal storage size.

The Cauchy distribution is recommended as the model distribution to characterize the forecast error in the persistence model discussed in [12]. Notably, the configuration of the distribution is observed to undergo substantial alterations depending on the forecasting horizon. The impact of the introduced wind power forecast uncertainty is demonstrated.

2.3.2.2 Non-Parametric Probabilistic Forecasts

When it comes to the derivation of wind power from weather variables, this can be achieved by constructing non-parametric distributions that describe wind power forecasts in a probabilistic form (in the form of quantiles) without relying on any assumption about the distribution. Quantile Regression is a non-parametric method for modeling predictive quantiles. Unlike ordinary least-squares regression, quantile regression is asymmetric and doesn't have a closed-form solution to perform parameter estimation. Decision-tree-based methods are an effective machine learning technique capable of performing quantile regression in wind power forecasting producing well-calibrated forecasts.

Bessa et al [5], mention that the top two performers in the 2014 Global Energy Forecasting Competition wind power segment adopted tree-based methodologies, incorporating a substantial array of input parameters from deterministic NWP models. For the mentioned competition, the top performer used a tree-based method, Gradient Boosting Machines (GBM), which was adapted for quantile regression tasks to learn the non-linear relationship between wind power production and the wind kinetic energy, turbine capacity factor, wind direction, cut-out speed and cut-in speed to effectively produce probabilistic wind power forecasts.

GBM is an ensemble learning technique that combines the predictions of multiple weak learners (decision trees) to create a strong predictive model. It employed the primary technique of Freund and Schapire's (1996) AdaBoost for fitting weak learners consecutively where the algorithm builds trees sequentially, and each tree corrects errors made by the previous ones. In the context of wind energy forecasting, each tree contributes to predicting the energy output, considering the information from previous trees. GBMs have hyperparameters that control the model's behavior, such as the learning rate, the number of trees, and the depth of each tree. These hyperparameters are tuned during training to find the optimal configuration for the specific forecasting task.

Identifying which meteorological variables (or features) need to be included in the model is imperative for a successful implementation of statistical learning. As such, features should be selected by a cross-validation procedure to understand which meteorological variables have the most significant impact on the model's predictions.

It is also worth noting that in gradient-boosting machines, the statistical model may become over-specified in relation to the training set, resulting in overfitting. In order to increase the accuracy and reliability of the predictions, k-fold cross-validation can be employed to ensure that the model becomes more generalized. GBMs can provide not only point predictions but also probabilistic predictions. This is crucial in wind energy forecasting, where uncertainty is inherent. The algorithm can output not just a single predicted value but a distribution, including quantiles, capturing the range of possible outcomes with respect to wind energy generation.

2.4 Error measurement and forecast performance

Inaccurate forecasts can result in substantial economic losses and power system reliability issues. As such to extract the full potential value of the wind forecasts, choosing evaluation parameters and benchmarks suitable for each model is essential to assess the performance of the chosen approaches adequately [13].

Assessing the effectiveness of forecasting methods involves directly comparing predicted wind power values at a specific time with the actual observations. Therefore, evaluating a forecasting method's quality revolves around analyzing the differences between the forecasted values and the real data. This process entails examining a substantial series of predictions to gather sufficient data for statistical performance evaluation.

Error measurement metrics can be used to highlight issues that are inherent to the conceived models used to predict wind power production. They enable comparisons of the performance between the different models, providing insight into which one performs its function better.

It is essential to ensure a fair comparison between different models by using the same forecasting problems and application conditions.

When evaluating probabilistic forecasts, which are inherently more complex, assessing multiple facets of the models becomes necessary. Some researchers categorize these facets into reliability, sharpness, and resolution [14], and the objective of a probabilistic forecast is to maximize

forecast sharpness, subject to calibration [15]. Reliability refers to the statistical consistency of a probabilistic forecast. Sharpness evaluates how closely grouped the possible outcomes are in a forecast, without thinking about whether those predictions are accurate or not. And lastly, forecast resolution is the ability of the method to produce case-dependent forecasts.

Error measurements can be classified be either being of deterministic or probabilistic nature depending on the type of the prediction model being used.

2.4.1 Deterministic Error Measurement

The error between the predicted production values and the real measured values can be obtained via the following calculation:

$$e_{t+k|t} = y_{pred} - y_{meas} \tag{2.3}$$

Where:

- e_n is the measured error between the predicted value corresponding to time $t+k$ made at time t and the real measured value;
- y_{pred} corresponds to the predicted value;
- y_{meas} corresponds to the actual measured value;

The prediction error obtained using the formula mentioned above can be normalized by the total capacity of the respective wind power farm:

$$e_{t+k|t} = \frac{e_{t+k|t}}{P_{Inst}} \tag{2.4}$$

Where

- P_{Inst} is the total capacity installed in a given wind power farm;

As mentioned before, error measurements are typically used to assess the performance of different prediction models. For instance, the MSE can be used to identify the contribution of positive and negative error values throughout the test dataset, which can be given by the following formula [7].

$$MSE_k = \frac{1}{N}\sum_{n=1}^{N} e_{t+k|k}^2 \tag{2.5}$$

In addition to the MSE, two other fundamental criteria are commonly used to demonstrate a model's performance: the MAE and the RMSE. The MAE can be defined as:

$$MAE_k = \frac{1}{N}\sum_{n=1}^{N} |e_{t+k|k}^2| \tag{2.6}$$

The RMSE corresponds to the square root of the MSE, and is defined as such:

$$RMSE_k = \frac{1}{N}\sqrt{\sum_{n=1}^{N} e_{t+k|k}^2} \tag{2.7}$$

Similar to the MAE, the RMSE criterion is impacted by both systematic and random errors. However, RMSE is advantageous as it provides information in the same units as the predicted variable, making it more straightforward to understand compared to MSE. When normalized by the installed capacity or average production of the wind farm, they become NMAE (Normalized Mean Absolute Error) and NRMSE (Normalized Root Mean Square Error).

The decision to use MAE or RMSE as the primary evaluation metric for wind power forecasts depends on how sensitive end-users are to errors, as represented by a loss function. Opting for RMSE implies a quadratic loss function is appropriate, while MAE corresponds to a linear function. If a method was trained to minimize MSE forecasts, RMSE is the suitable choice. However, if the loss function representing user sensitivity isn't clearly defined, it's advisable to report both MAE and RMSE [7].

2.4.2 Probabilistic Error Measurement

Reliability and sharpness play crucial roles in evaluating probabilistic forecasts, but their interpretation can be subjective, especially when comparing different methods. To address this issue, skill scores or scoring rules are introduced to ensure objectivity, requiring that the best forecast receives the highest score [16].

Some measures commonly used for this purpose are the continuous ranked probability score (CRPS) and the pinball loss function.

The CRPS is particularly robust as it evaluates both reliability and sharpness. It has the advantage of simplifying to the absolute error for deterministic forecasts, facilitating comparisons between probabilistic and point forecasts. The CRPS is formulated as follows: [17]

$$CRPS(F,x) = \int_{-\infty}^{\infty} (F(y) - \Gamma(y_{Meas} \leq y_{Pred}))^2 dy \tag{2.8}$$

Where

- $F(y)$ is the predicted probability distribution function;
- Γ represents the Heaviside step function;
- y_{Meas} corresponds to the actual observed values;
- y_{Pred} is the predicted value;

Since the CRPS has a direct relation to the absolute error, a low CRPS indicates an accurate probabilistic forecast.

Like the CRPS, the pinball loss function also accounts for both reliability and sharpness, but it is tailored specifically for quantile forecasts [18]. It is defined as follows [19]:

$$Pinball(y_{Pred},x) = \begin{cases} (1-q)|y_{Pred} - y_{Meas}| & \text{if} \quad y_{Pred} \leq y_{Meas} \\ q|y_{Pred} - y_{Meas}| & \text{if} \quad y_{Pred} > y_{Meas} \end{cases} \tag{2.9}$$

Where:

- q is the target quantile;

After calculating for each quantile, these can be summed over the forecast horizon to attain the pinball loss [18]. It also worth noting that a low pinball loss is indicative of an accurate probabilistic forecast model [18].

2.5 Second-level forecasting methods

2.5.1 Gradient Boosting Trees and Encoder-Decoder Artificial Neural Network

Second-level forecasting takes probabilistic forecasts, in the form of quantiles, from multiple forecasting algorithms and combines them to produce a distribution model of the uncertainty that is associated with the forecasts from the first level.

In the article [3], Gradient Boosting Trees model with feature engineering is used with truncated generalized Pareto distribution for the distribution's tails. The GBT used was extended to include as inputs a set of probabilistic forecasted values obtained from the forecast for the day ahead (D+1) with NWP data generated at 0h00 of the day D.

Another model employed in the mentioned article was an Encoder-Decoder Artificial Neural Network (ED-ANN). The Encoder component comprised two layers, specifically Long Short-Term Memory (LSTM) network layers, designed to handle temporal or sequential information with feedback connections, while the Decoder, consisted of three internal layers: one LSTM and two fully connected layers utilizing a linear activation function. The output is produced iteratively by the Decoder, and at each forecast step, the hidden state generates the subsequent forecast. To train the Decoder, first-level forecast values were used as input.

The performance of both these models was also assessed by evaluating the mean squared error that is obtained by the difference between the forecasted quantile and the second-level forecast of the same quantile. Using a naive model as a benchmark, where both the forecasted quantile and the second-level forecast quantile are the same, the MSE improvement over the benchmark model was computed for both GBT and ED-ANN models. Results demonstrated that both models improve over the naive model, resulting in increased forecasting skill when employing any of these methods thus illustrating the potential of second-level forecasts.

2.5.2 Multivariate Model

Kolkmann et al [20], proposes a multivariate model of wind forecasting trajectories to support power trading strategies under continuous trading with repeated updates of wind forecasts considering the correlations of forecast changes for subsequent delivery periods. A copula-based approach is employed to capture the interdependencies between forecast updates for subsequent future periods. This multivariate probabilistic forecast update modeling, in contrast to single-point forecasts, offers more comprehensive and valuable insights for optimizing wind power bidding strategies and risk management in the continuous intraday electricity market.

Copulas have been shown to enhance wind power forecasting by considering random variables as a joint distribution rather than individual entities. The key concept lies in first characterizing marginal distributions for the uncertain factors and then employing a copula to describe the multivariate distributions based on the marginal cumulative distribution functions. Pinson et al [21] employed an application with basics of the modeling with copulas for wind power forecasting. The method proposed in [21] models forecast updates as stochastic processes, capturing the dynamic nature of renewable infeed taking into account the continuous evolution of information through forecast updates.

This approach centers on multivariate modeling of wind forecasting paths to aid in power trading strategies within the context of continuous trading and recurring wind forecast updates. To achieve this, the authors model the short-term dependencies between forecast updates for multiple subsequent wind infeed hours. Theoretical properties are derived that should be satisfied by such forecast updates and a copula approach is employed to derive an empirically sound model. This methodology is applied to a case study of hourly wind power forecast updates and evaluated against alternative formulations. In contrast to other multivariate probabilistic forecasts, the authors concentrated on interrelations between forecast updates released at the same time step but for multiple future periods. As a result, multivariate forecast paths or update processes are derived, serving as a crucial component and driver, particularly for intraday electricity trading [20].

The proposed approach is capable of simultaneously replicating autocorrelation and cross-correlation of sequences of forecasts and when using a t-copula, provides a balanced performance across three distinct metrics measuring overall forecast performance as well as the ability to replicate dependencies along and between forecast trajectories. The author also highlights the impact of appropriate copula choices on the model's performance, recommending exploiting additional data sources and to investigate methods for better calibration of the copula parameters [20].

2.6 Use cases for second-level forecasting

The approach outlined by Bessa et al [3] proposes using a predictive algorithm to procure grid flexibility by the system operator. The aim is to alleviate the cognitive load on the human operator when confronted with the analysis of various flexibility options and trajectories concerning the forecasted load. This method incorporates the time-to-decide strategy, where the operator must decide to procure flexibility either immediately when decision costs are lower or wait for the subsequent forecast update when forecast accuracy might be higher albeit with higher decision costs.

The time-to-decide problem is formulated based on the concept of second-level forecasting, aiming to predict how uncertainties in Renewable Energy resources and load forecasts evolve over time with the availability of new information [3]. The findings of this study indicate that the methodology grounded in uncertainty forecasts results in cost savings related to managing technical grid issues. Additionally, strategically reserving flexibility at the optimal moment, guided by an understanding of how forecasts' uncertainty evolves, leads to cost savings when compared

to a strategy that exclusively reserves flexibility when costs are lower. This underscores the significance of having enhanced knowledge about forecast changes over time for informed decision-making. In the context of uncertainty forecasts, GBT can be used to generate predictions and estimate associated uncertainties.

Wanke and Greenbaum [22] addressed the air traffic congestion management issue with a novel sequential decision-making approach, wherein the uncertainty in traffic and weather predictions is quantified and explicitly considered in developing efficient congestion resolution strategies. The proposed method utilizes Monte Carlo simulation to generate traffic and weather outcomes based on specific forecasts. Various sequential decision strategies are then assessed against the spectrum of potential outcomes to identify the optimal course of action.

Decisions are guided by a quantitative evaluation of the expected delay cost distribution, with a focus on targeting resolution actions at specific flights rather than entire flows. To validate the approach, a simulation was conducted to explore a weather-induced airspace congestion scenario, considering three distinct levels of weather forecast uncertainty. For the first scenario, perfect weather was assumed with no uncertainty associated with the predictions.

Results for this scenario show that decision paths where action to prevent congestion is taken closer to the last moment possible are more viable with fewer hours of delay and fewer aircraft affected than compared to decision paths with a more aggressive strategy where preventive action is taken earlier. However, when taking weather uncertainty into account, the results were different. As the severity of weather uncertainty increased, results showed that for these scenarios, the previous decision path for the first scenario was no longer viable as decision paths with actions taken earlier resulted in less number of aircraft affected and fewer delays in flights as well.

As such, this study illustrates the effects of uncertainty in weather forecasts in decision-making and how quantifying these uncertainties can lead to different solutions for air traffic congestion management than when compared to a scenario where uncertainty isn't taken into account.

Jewson et al [23] discuss the need for users of meteorological forecasts to decide whether to act on current forecasts or wait for potentially more accurate future forecasts and analyze this question as an extension of the time-dependent cost–loss model discussed by Murphy and Ye [24]. This decision-making process is informed by understanding the probabilities of forecast accuracy and potential changes in weather predictions. The authors conclude that to maximize the value of probabilistic forecasts for decision-making, forecast providers should consider including supplementary information regarding forecast changes alongside their forecasts.

The use case of second-level forecasts in the mentioned article can be seen in the decision-making scenarios where more accurate subsequent forecasts are anticipated. For instance, when an event organizer needs to decide whether to cancel an event based on current weather forecasts or wait for the next update, the second-level forecasts can provide crucial information. These forecasts would predict how the initial forecast might change, allowing the organizer to make a more informed decision on whether to cancel early (with lower costs) or wait (and risk higher costs or losses).

Chapter 3

Forecasting Methodology

3.1 Problem Formulation

The forecasting of wind energy production involves the use of input data that characterizes the atmospheric conditions influencing wind behavior. Different data sources are employed in advanced modeling techniques, where the complex interactions influencing wind energy production is captured, and as thus providing reliable estimates of uncertainty. The goal is to deliver probabilistic forecasts that convey the range of possible outcomes given the inherent variability in atmospheric conditions.

A second-level forecast, denoted by the equation below, predicts the conditional quantile (α) for a future time interval (t + k) based on information available at the present time (t). This forecast, designed for a specific proportion (α between 0 and 1), is updated every z time interval. The update itself is essentially the expected value of the next forecast's quantile (α) for the same future time interval (t + k), generated z time intervals later (t + z). [3]

To achieve this, a forecasting model (f) specific to each quantile level (α) is employed. This model utilizes two types of inputs:

- Quantile forecasts ($\widehat{q}^{\alpha}_{t+k|t}$): These are the initial forecasts for the future time interval (t + k) generated at the current time (t);
- Combined features ($X^{\alpha}_{t+k|t}$): This combines l explanatory variables that capture the uncertainty level in the initial power forecast at the current time (t) and exogenous variables like weather data (NWP);

$$\hat{q}^{\alpha}_{t+k|t} = \mathbb{E}[\hat{q}^{\alpha}_{t+k|t+z}|\hat{q}^{\alpha}_{t+k|t}, X^{\alpha}_{t+k|t}] = f\left(\hat{q}^{\alpha}_{t+k|t}, \hat{x}^{1}_{t+k|t}, \ldots, \hat{x}^{l}_{t+k|t}\right) + e_{t+k|t} \tag{3.1}$$

Second-level forecasting is employed using different machine learning approaches which are used to make a prediction based on input datasets containing NWP that characterizes atmospheric conditions influencing wind behavior. The algorithm describes the behavior of the dataset, model

input features for the expected output, and forecast output features with respect to its historical records. If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this "evaluate and optimize" process, updating weights autonomously until a threshold of accuracy has been met.

Initially the provided data is split to divide the dataset into training and testing sets. The training set is used to train the GBT model, while the testing set is reserved for evaluating its performance. Then relevant features are identified by assessing the correlation between the provided input data and wind production. The model is constructed using the training dataset, incorporating the validation of decisions and/or rules pertaining to the structure of the model [7]. It should be noted that the risk of overfitting should be accounted for. Overfitting occurs when the network becomes too adapted to the training set, losing its ability to generalize. For the implementation of the GBT algorithm, different python frameworks can be used such as XGBoost [25] or scikit-learn [26]. Using the testing dataset, the evaluation of the model is assessed by verifying how well the model generalizes to new, unseen data. Appropriate metrics for regression problems can be used to compare the deviation between the estimated values with the measurements, such as mean absolute error or root mean squared error [7]. Forecast change and forecast skill variation between different hours can be quantified by using the root-mean-square size of forecast changes (RMSC) or mean absolute change (MAC), allowing the user to understand the sizes of forecast changes [27].

GBT is adapted for quantile regression to estimate different quantiles of the uncertainty distribution [28]. Also, the GBT needs to be extended to accommodate additional variables for second-level forecasting, namely the quantiles from the first-level forecast [3]. The model should be continuously refined by adjusting hyperparameters based on validation results and additionally be iteratively improved by adding new features that can improve its performance and forecasting skill. Once the model's performance meets standards, it is deployed for operational use.

3.2 Machine Learning Approaches

3.2.1 Gradient Boosting Trees

Andrade and Bessa [29], outline a forecasting framework designed to leverage insights from a grid of numerical weather predictions (NWP) for wind and solar energy applications. The approach integrates the gradient boosting trees algorithm with feature engineering techniques, aiming to extract maximum information from the NWP grid.

Gradient Boosting Trees is an ensemble machine learning technique applicable to classification and regression challenges. This algorithm employs numerical optimization through steepest-descent minimization in function space. It iteratively combines base learners on adapted data derived from previous iterations. GBT is a particular case of the gradient boosting algorithm where the base learners are regression trees.[29] The mentioned article, used an initial model,

where the average of the target value is calculated by $F(x) = \sum_i^N y_i/N$ which is then followed by the following iterative process for the training phase:

- For each training sample i, calculate the negative gradients: $-g(x) = -\partial L(y_i, F(x_i))/\partial(F(x_i)) = y_i - F(x_i)$, where L is the loss function;
- Fit a regression tree h (x; a) using the negative gradients $-g(x_i)$ as target variable; a is the set of hyperparameters of the regression tree;
- Calculate $F_m(x) = F_{m-1}(x) - \tau \times h(x,a)$, where τ is the set of hyperparameters of the regression tree;

The described algorithm presents three key benefits for addressing the renewable energy sources forecasting challenge:

- Its non-parametric nature allows avoiding strict assumptions about the underlying data distribution. This flexibility allows the algorithm to adapt to diverse and complex patterns present in renewable energy sources (RES) data.
- It accommodates various loss functions, providing the flexibility to generate distinct types of forecasts. For instance, it can produce point forecasts using absolute or square loss functions and, alternatively, generate probabilistic forecasts with the quantile loss function. This versatility ensures that the algorithm can tailor its predictions to different forecasting requirements.
- The algorithm exhibits scalability, making it suitable for handling a substantial number of explanatory variables. This scalability is particularly advantageous for practical applications in industrial settings where forecasting models may need to consider numerous factors influencing renewable energy production.

Hyperparameters in Gradient Boosting Trees are configuration settings that guide the training process and overall behavior of the model. These parameters are set before the training phase and are learned from the data. The hyperparameters are categorized those specific to regression trees and those about the boosting process. The tree-specific hyperparameters are as follows:

- **Maximum Depth:** Controls the maximum depth of each tree in the ensemble and regulates the complexity of individual trees, preventing overfitting; [29]
- **Minimum Samples to Split:** Higher values prevent overfitting by avoiding the selection of specific samples for a tree. Ranges between 150 and 350; [29]
- **Minimum Samples at Leaf Node:** Regulates overfitting, tuned by maximum depth and minimum samples split. Set between 20 and 80; [29]
- **Number of Features for Best Split:** Determines the number of features considered when looking for the best split. Choosing fewer features reduces variance and increases bias; [29]

The boosting-specific hyperparameters are as follows:

- **Learning Rate:** Controls the contribution of each tree in the additive training process. A lower value shrinks contributions, enhancing model generalization. Set between 0.01 and 0.05; [29]
- **Number of Boosting Iterations:** Defines the base learners in the final model. Considered alongside the learning rate to avoid overfitting. Ranges between 500 and 800; [29]
- **Fraction of Samples for Base Learners:** Slightly below 100% enhances model robustness by reducing variance. Typically around 80%; [29]

The mentioned hyperparameters can be estimated by Grid search [30], Random Search [31] or by the the Bayesian Optimization algorithm [29].

For this work, the Bayesian Optimization algorithm was used to estimate the hyperparameters of the GBT. This involved a 10-fold cross-validation and starting with 25 initial random combinations of hyperparameters for the parametrization. For the first-level forecasts, the GBT was fitted the features described in subsection 3.2.3 while for the second-level forecasts, the GBT model was fitted with additional features that pertains to variables that were obtained from the first level forecasts as described in subsection 3.2.5.

The authors in [29] concluded that the methodology combining GBT with feature engineering demonstrated superior forecasting performance, both in terms of point forecasts and probabilistic forecasts, when compared to a simpler model that considers only a single NWP point for a specific location. The improvement is quantified as 13.11% for solar power and 12.06% for wind power in respect to probabilistic forecasts. This suggests that the new approach not only improves the accuracy of point forecasts but also enhances the model's ability to provide probabilistic forecasts, which capture the uncertainty associated with predictions.

3.2.2 Bayesian Optimization For Hyper-Parameters Tunning

Hyperparameter tuning involves finding the best settings for a machine learning model, but the performance metric we're optimizing for can be complex and difficult to directly evaluate. This makes it challenging for traditional optimization methods like Newton's or gradient descent, which rely on readily available derivatives. Bayesian optimization offers a powerful approach to navigating these "black-box" objective functions. It's a powerful approach for finding the extrema of computationally expensive functions, which is particularly useful for optimizing functions that lack a closed-form expression, are costly to evaluate, have difficult-to-calculate derivatives, or are non-convex. The objective in this context is to identify the maximum value of an unknown function f at a sampling point, denoted as:

$$x^{+} = \arg\max_{x \in A} f(x) \tag{3.2}$$

where A represents the search space of x.

Bayesian optimization is rooted in Bayes' theorem, which states that the posterior probability $P(M|E)$ of a model M, given evidence E, is proportional to the likelihood $P(E|M)$ of observing E given M, multiplied by the prior probability $P(M)$:

$$P(M|E) \propto P(E|M)P(M)$$

This principle forms the foundation of Bayesian optimization. It combines the prior distribution of the function $f(x)$ with sample information (evidence) to derive the posterior distribution of the function. This posterior information is then used to determine where the function $f(x)$ is maximized according to a specific criterion, represented by a utility or acquisition function u. The acquisition function u helps identify the next sampling point to maximize expected utility, balancing exploration (sampling in high-uncertainty areas) and exploitation (sampling in high-value areas). This balance helps minimize the number of samples needed and enhances performance, even in functions with multiple local maxima.

Bayesian optimization also relies on the prior distribution of the function f, which is crucial for statistical inference of the function's posterior distribution. This prior distribution can be based on subjective beliefs and does not need to be entirely objective. Generally, Gaussian processes are well-suited as the prior distribution in Bayesian optimization due to their flexibility and ease of handling. Gaussian processes are used to fit the data and update the posterior distribution effectively in Bayesian optimization. [32]

3.2.3 First Level Forecast Features

The features tested for the various forecasting models were categorized using a methodology similar to that used by Andrade and Bessa [29], which is as follows:

- Group A: Meteorological variables namely wind speed and direction at 10 and 100 meters altitude, instantaneous wind gust speed at 10 meters altitude and u and v components of wind speed at 10 and 100 meters altitude. Wind speed (ws) and wind direction (wd) at 10 and 100 meters was determined via the following formulas. Here, u and v represent the provided wind components, and d represents density, which was assumed to be a constant value of 1.0:

$$ws = \sqrt{u^2 + v^2} \tag{3.3}$$

$$wd = \frac{180}{\pi} \times \arctan(u, v) \tag{3.4}$$

- Group B: Transformations of meteorological variables such as the respective mean, temporal variance and standard average deviation of the various instances. The temporal variance (σ^2_{time}) can be calculated for each weather research and forecasting (WRF) variable time series x(t + k). These metrics consider a moving window of Nh hours (where Nh can be 3,

7, or 11) centered on a specific forecast time horizon (t + k).

$$\sigma^2_{time}(t+k) = \frac{\sum_{t+k-Nh/2}^{t+k+Nh/2}(x(i)-\bar{x})^2}{Nh-1} \tag{3.5}$$

- Group C: The influence of a numerical weather prediction (NWP) variable on a regression model can be captured by incorporating its past values (lags) and future values (leads) relative to a specific forecast time horizon. Lags are included as preceding time steps (t + k - z, where z represents the lag period), and leads are included as following time steps (t + k + z).

- Group D: Calendar variables such as month and time so that seasonality information associated with the wind resource is provided. It is important to note that these are cyclical variables and as such, can be normalized and then transformed into periodic sine and cosine functions.

$$Hour_{Normalized,Cos} = cos\left(\frac{Hour}{24} \times 2\pi\right) \tag{3.6}$$

$$Hour_{Normalized,Sin} = Sin\left(\frac{Hour}{24} \times 2\pi\right) \tag{3.7}$$

3.2.4 Second-Level Model Architecture

Unlike first-level forecasting, which directly predicts the wind power output, second-level forecasting delves into how the uncertainty of these predictions fluctuates over time during the various NWP predictions updated that occur during the day.

The same gradient boosting regressor that was employed for the first-level forecasts is again used for second-level forecasting, which was extended to incorporate input variables specifically engineered for second-level forecasting. These additional variables were derived from the quantiles predicted in the first-level forecast.

The target variables for the second-level model are the quantiles predicted by the first-level forecast that occur with every NWP update on a given day. The model is expected to output predictions for a range of quantiles from 5% to 95%, with increments of 5%. This range provides a comprehensive view of the potential variability in wind power production for each NWP update.

Second-level forecasting aims to anticipate the changes in forecast uncertainty that occur with the updates of NWP, which are typically issued every six hours at 00h00, 06h00, 12h00, and 18h00.

The figure 3.1 illustrates the layout of a second-level forecast for a lead time of 24h generated at 00h00, using the quantile predictions from the first-level forecasts ($Q(t+25)$). In this example, second-level forecasting aims to predict how the uncertainty of wind power production for the respective quantile will vary by the next NWP update at 06h00, using data available at 00h00 ($Q(t+25|t+06)$).

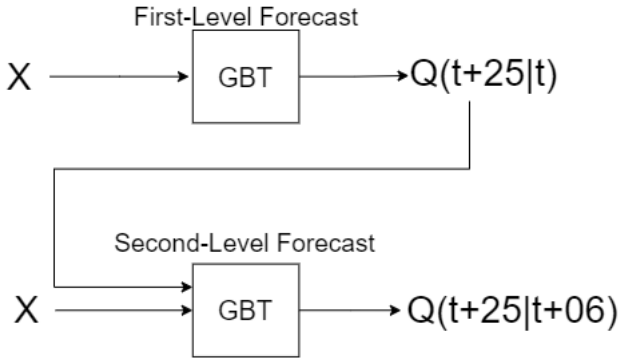

Figure 3.1: Modelization of a second level forecast

This forecasting approach provides a deeper insight into the reliability and potential errors of wind power forecasts, enhancing the decision-making process for grid operators and energy planners.

In the second-level prediction, the prediction model makes a deterministic prediction that corresponds to minimizing the mean absolute error (corresponding to the 50% quantile).

The hyperparameters were adjusted for each quantile to be predicted, as each quantile requires its own model and therefore the hyperparameters used are also different.

The model's performance is evaluated using a separate test set to ensure its generalizability. Key performance metrics include the mean absolute error and root mean squared error for the predicted quantiles.

3.2.5 Second-Level Forecast Features

The features tested for the different forecasting models were grouped as follows:

- Group 1 (G1): Meteorological variables such as wind speed at 10 and 100 meters altitude, instantaneous wind gust speed at 10 meters altitude;
- Group 2 (G2): Calendar variables such as month and time so that seasonality information associated with the wind resource is provided. It is important to note that these are cyclical variables and as such, can be normalized and then transformed into periodic sine and cosine functions;
- Group 3 (G3): Corresponds to variables extracted from the first-level forecasts such as the quantiles 5% to 95% and the target quantile. Usually including the quantiles that are neighboring the target quantile two positions before and two positions after, yielded the best results;
- Group 4 (G4): It corresponds to transformations of the set of variables in group 3, namely the mean, average standard deviations and difference between neighboring quantiles of the

target quantile with a distance of 2 positions ahead and 2 positions behind and the interquartile range (difference between the 75% and 25% quantiles), the mean. It was also extracted the temporal mean ($\bar{x}_{time}(t+k)$) and temporal variance ($\sigma^2_{time}(t+k)$) considering a $N_h \in$ 1,2,3 hours moving window centered in a specific lead-time of interest (t + k), which are given by the following formulas respectively:

$$\bar{x}_{time}(t+k) = \frac{\sum_{t+k-N_h/2}^{t+k+N_h/2} x(i)}{N_h - 1} \tag{3.8}$$

$$\sigma^2_{time}(t+k) = \frac{\sum_{t+k-N_h/2}^{t+k+N_h/2} (x(i) - \bar{x})^2}{N_h - 1} \tag{3.9}$$

- Group 5 (G5): lags and leads ($t+k\pm z$) of NWP variables;
- Group 6 (G6): Transformations of NWP variables such as the respective mean, temporal variance and standard average deviation of the various instances;

Chapter 4

Numerical Results

In this chapter, the feature selection process for the prediction models utilized in first-level forecasting will be outlined. This will be followed by a demonstration of how the previously designed forecasting models were expanded to meet the requirements for making second-level forecasts, in order to predict fluctuations in production forecasts between the updates of the intraday weather forecasts, which occur every 6 hours. Lastly, a result analysis of the forecasts is conducted. All the forecasting models were designed using the GBT machine learning model available in the scikit-learn framework with a 48-hour time horizon.

4.1 Data Description

The available data was provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) running the numerical weather prediction model HRES. HRES is a single forecast (horizontal resolution around 9 km) that provides a description of one possible evolution of the weather out to 10 days ahead and assimilates vast amounts of observational data from satellites, ground-based stations, and other sources to provide accurate and reliable weather forecasts. The data used for this dissertation characterizes the wind behavior in three different Wind Farms located in Norway.

It is also worth noting that, meteorological data is forecasted and updated every 6 hours.

4.2 Description of the Data

The wind farms selected for wind production forecasting in this dissertation have the following characteristics: [33]:

Table 4.1: Wind Farm's characteristics

Wind Farm	Capacity (MW)	Nº of Turbines	Start Date	Latitude(º)	Longitude (º)
Fakken	60	18	01/06/2012	70.1023	20.1076
Lista	71.3	31	27/09/2012	58.1557	6.6818
Bessakerfjellet	57.5	25	10/09/2007	64.2248	10.3746

The process of selecting features for the forecasting models commenced with a basic model incorporating only one pertinent variable for predicting wind production. In the case of the first-level forecasts, the correlation between actual production values and the diverse variables predicted using the HRES model was examined, ultimately identifying a strong correlation for all wind farms between wind production and wind speed at 100 meters of altitude.

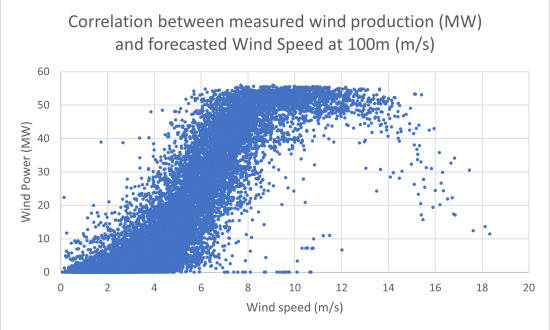

Figure 4.1: Correlation between wind power production and wind speed at 100 meters of altitude

After generating a forecast, the residuals between the forecast values and the actual production values were computed. These residuals will be utilized to evaluate the potential of new variables to refine the initial model by examining the correlation between the respective residuals and the candidate variable to address any deviations observed in the base model.

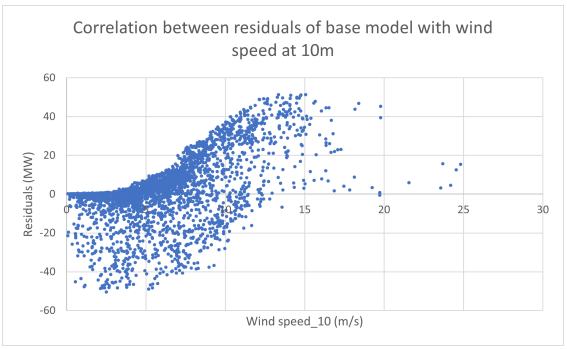

Figure 4.2: Correlation between base model residuals and wind speed at 10 meters of altitude

The graph above in 4.2 demonstrates that the candidate variable (specifically the wind speed at 10 meters above sea level) has the capability to rectify the deviations observed in the base model. The varying wind speeds correspond to different residual values, suggesting that including the wind speed at 10 meters in the next model could help mitigate the errors identified in the base model. This process is repeated by exhaustively testing the other variables until the RMSE, MAE, and pinball losses are minimized.

In the case of second-level forecasts, it was observed that the quantile generated in the first-level forecast at the time the second-level forecast was launched displayed a high correlation with the target quantile.

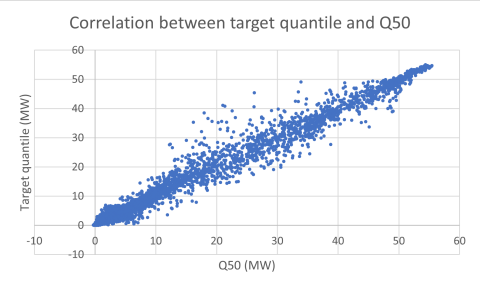

Figure 4.3: Correlation between target quantile generated at 06h and quantile Q50 generated at 00h

The analysis of the relationship between the residuals of the second-level forecasts and the different candidate variables proved to be unproductive, as the second-level forecasts exhibited low errors, making it challenging to visually assess the impact of the features on correcting the discrepancies in the forecast models. In this scenario, the selection of variables for the second-level prediction models was conducted through a trial-and-error approach. A new variable was introduced if it showed some correlation with the target quantile, such as neighboring quantiles within a 2-quantile range and their respective means. After incorporating the new candidate variable into the model, a second-level forecast is generated and to evaluate the mean absolute error and root mean squared error. If the inclusion of the variable led to a reduction in these metrics, it is retained in the model. Otherwise, the variable is excluded from the forecasting model. This process is repeated until the MAE and RMSE are minimized.

The variables used for the first-level forecasts are presented in the following table:

Table 4.2: Variables used for first-level forecasting

Variable Group	Variable	Description
A	WindSpeed100_t_z	Wind speed at 100 meters altitude generated at z hours with lead time t (m/s)
A	WindSpeed10_t_z	Wind speed at 10 meters altitude generated at z hours with lead time t(m/s)
A	WindDirection100_t_z	Wind direction at 100 meters altitude generated at z hours with lead time t (º)
A	WindDirection10_t_z	Wind direction at 10 meters altitude generated at z hours with lead time t (º)
A	i10fg_t_z	Instantaneous wind gust speed at 10 meters of altitude generated at z hours with lead time t (m/s)
A	u100_t_z	U wind component speed at 100 meters of altitude generated at z hours with lead time t (m/s)
A	v100_t_z	V wind component speed at 100 meters of altitude generated at z hours with lead time t (m/s)

Table 4.2 continued from previous page

Variable Group	Variable	Description
A	u10_t_z	U wind component speed at 10 meters of altitude generated at z hours with lead time t (m/s)
A	v10_t_z	V wind component speed at 10 meters of altitude generated at z hours with lead time t (m/s)
B	AVG_WindSpeed100	Average of all wind speeds at 100 meters of altitude (m/s)
B	AVG_WindSpeed10	Average of all wind speeds at 10 meters of altitude (m/s)
B	AVG_WindDirection100	Average of all wind directions at 100 meters of altitude (º)
B	AVG_WindDirection10	Average of all wind directions at 10 meters of altitude (º)
B	AVG_i10fg	Average of all instantaneous wind gust speeds at 10 meters of altitude (m/s)
B	VARWindSpeed100_t_z	Temporal variance of wind speed at 100 meters of altitude generated at z hours with lead time t (m/s)
B	VARWindSpeed10_t_z	Temporal variance of wind speed at 10 meters of altitude generated at z hours with lead time t (m/s)
C	LAG3WindSpeed100_t_z	3 hour lagged Wind speed at 100 meters altitude generated at z hours with lead time t (m/s)
C	LAG3WindSpeed10_t_z	3 hour lagged Wind speed at 10 meters altitude generated at z hours with lead time t (m/s)
C	LAG3WindDirection100_t_z	3 hour lagged Wind direction at 100 meters altitude generated at z hours with lead time t (º)
C	LAG3WindDirection10_t_z	3 hour lagged Wind direction at 10 meters altitude generated at z hours with lead time t (º)
D	Cos(Hour)	Normalized hour via cosine transformation
D	Sin(Hour)	Normalized hour via sine transformation

The variables used in the second-level forecasting models that successfully reduced the MAE and RMSE are outlined in table 4.3. It's important to note that the features related to the G5 group of variables (lags and leads of NWPs) did not lead to a reduction in deviations between the forecasted and real values and, consequently, were not included in any of the second-level forecast models.

Table 4.3: Variables used for second-level forecasting

Variable Group	Variable	Description
G1	WindSpeed100_t_z	Wind speed at 100 meters altitude generated at z hours with lead time t (m/s)
G1	WindSpeed10_t_z	Wind speed at 10 meters altitude generated at z hours with lead time t (m/s)
G1	WindDirection100_t_z	Wind direction at 100 meters altitude generated at z hours with lead time t (m/s)
G1	WindDirection10_t_z	Wind direction at 10 meters altitude generated at z hours with lead time t (m/s)
G1	i10fg_t_z	Instantaneous wind gust speed at 10 meters of altitude generated at z hours with lead time t (m/s)
G2	Month	Month in which wind resource was predicted
G2	Hour	Hour in which wind resource was predicted
G3	Q_N	Quantile target extracted from the first-level forecast
G3	Q_N+1	Neighbouring quantile with a neighbourhood of 1 quantile ahead of the target quantile
G3	Q_N+2	Neighbouring quantile with a neighbourhood of 2 quantiles ahead of the target quantile
G3	Q_N-1	Neighbouring quantile with a neighbourhood of 1 quantile behind of the target quantile
G3	Q_N-2	Neighbouring quantile with a neighbourhood of 2 quantiles behind of the target quantile
G4	AVG{QN-2_QN+2}	Average of the quantiles between neighbouring quantiles with a neighbourhood of up to 2 quantiles before and after
G4	AVG_QN	Temporal mean of the target quantile QN with a 3 hour moving window
G6	AVG_WindSpeed100	Average of all wind speeds at 100 meters of altitude (m/s)
G6	AVG_WindSpeed10	Average of all wind speeds at 10 meters of altitude (m/s)

4.3 First Level Forecasts

The forecast models refer to the Bessakerfjellet, Fakken, and Lista wind farms with a capacity of 57.5, 60, and 71.3 MW, respectively. The historical data used covers a time period from June 4th, 2019, to November 30th, 2020, with no missing values in all of its span.

The time horizon is 48 hours where a new wind production forecast is released every 6 hours to take into account the NWP updates that occur during the day.

4.3.1 Training and Test Dataset

Due to the high variability of the wind resource, it is essential to ensure that the training group is large enough for sudden changes in the local wind dynamics and larger-scale wind-flow patterns to be captured by the GBT model. It was therefore decided to dedicate a percentage of 80% to the training group, i.e. the months of June 2019 to May 2020 were used for the training phase and the months of June to August for the test phase. Since the first-level forecasts will later be the forecast target for the second level, the test period was extended to include September, October, and November 2020, using the same training period as previously used for the aforementioned months. This resulted in 6 months of data to be used later in the 2nd level forecasts, allowing the models designed to be measured more accurately.

4.3.2 Model

Wind power production depends on several key factors, including the wind's available kinetic energy, turbine capacity, wind direction, and the speeds at which turbines turn on and off (cut-in and cut-out speeds). While an ideal approach might involve determining the prevailing wind patterns and operational limits for each zone to build a model based on these physical characteristics, forecast uncertainties make this challenging. Therefore, a machine learning approach that can capture the complex interactions between these factors is more desirable.

For this dissertation, the GBT model available in the scikit-learn library was implemented using a quantile loss function to generate probabilistic forecasts. Because of its non-parametric nature, making assumptions about the data distribution is avoided, allowing the algorithm to adapt to diverse and complex patterns present in renewable energy sources (RES) data. Each probabilistic forecast is represented by a set of quantiles between 5% and 95%, with a 5% increment.

Table 4.4: Hyperparameter sets for windfarm Bessakerfjellet

Parameter	**Quantiles 05% - 10%**	**Quantiles 15% - 95%**
Number of Boosting Iterations	1000	190
Maximum Depth	3	4
Minimum Samples to Split	212	80
Minimum Samples at Leaf Node	200	108
Learning Rate	0.0343	0.0495
Subsample	0.146	0.503
Max Features	6	6

The hyperparameters were adjusted using the Bayesian optimization algorithm described in section 3.2.2, where a fold of 10 cross-validation and 25 random hyperparameter combinations

Table 4.5: Hyperparameter sets for windfarm Fakken

Parameter	**Quantiles 05% - 10%**	**Quantiles 15% - 95%**
Number of Boosting Iterations	130	192
Maximum Depth	4	9
Minimum Samples to Split	2	162
Minimum Samples at Leaf Node	200	147
Learning Rate	0.0838	0.0466
Subsample	0.369	0.805
Max Features	3	2

Table 4.6: Hyperparameter sets for windfarm Lista

Parameter	**Quantiles 05% - 10%**	**Quantiles 15% - 95%**
Number of Boosting Iterations	100	685
Maximum Depth	10	10
Minimum Samples to Split	300	159
Minimum Samples at Leaf Node	1	192
Learning Rate	0.0145	0.0466
Subsample	0.342	1.00
Max Features	6	6

were used to start. The predictions for the 5% and 10% quantiles required a different set of hyperparameters from the other quantiles in order to obtain more accurate results for these quantiles in question.

The different hyperparameters used for each wind farm are presented in tables 4.4, 4.5 and 4.6

4.3.3 Metrics

To gauge the quality of the first-level forecasts, the 50% quantile forecasts (i.e. Q50) were compared with the actual values using error metrics such as MAE and RMSE. It is also worth noting that all errors were normalized by the rated power of it's respective windfarm.

The probabilistic forecasts are represented by a set of quantiles between 5% and 95% with increments of 5% whose forecast quality is evaluated using the pinball loss function in which a low value is indicative of an accurate and sharp probabilistic forecast. The hyperparameters were adjusted using the objective function of minimizing pinball losses.

The forecasting skill of each first-level forecast model is assessed by comparing it to a simple model that includes only group A variables, namely the wind speed and direction at 10 and 100 meters of altitude at various times.

4.3.4 First-Level Forecasts Results

This section presents the results of the models developed for probabilistic wind energy forecasting. The main objective of this analysis is to identify which combinations of input variables produce the most accurate forecasts, minimizing the error when compared to the actual observed values.

Various combinations of the four groups of input variables listed in table 4.2 were tested and selected based on their relevance to wind power forecasting. Each group of variables represents different aspects and influences on the behavior of wind power generation, providing a comprehensive view of the factors affecting energy production.

The final model's selection involved an exhaustive analysis of the combinations of variables to identify the one with the lowest forecast error. This process allowed for an in-depth understanding of the influence of each group of variables on the accuracy of the forecasts, providing insight for future implementations and improvements of wind energy forecasting models.

Below are the detailed analysis results for June, July, August, and September 2020, including the model performance metrics and the tested variable combinations. The final model employed for forecasting for each lead time is bolded in tables 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12. The figures 4.6, 4.11, and 4.19 depict the hourly MAEs of the final models utilized for the different first-level forecasts at each lead time.

4.3.4.1 Result analysis for Bessakerfjellet

The graphs below display the initial forecast results, illustrating how different sets of variables impact the MAE and RMSE of the forecast models over the forecast horizon. First-level forecasts are launched every 6 hours. Model M1 comprises group A variables, while model M2 includes a combination of group A and group B variables. If the metrics show improvement of M2 over M1, the subsequent model M3 will consist of group A, B, and C variables. However, if including group B has a negative impact, the next model M3 will only include group A and C variables, and so on to the final model M4. It is important to highlight that forecast models should exclusively incorporate variables that are accessible at the time the forecast is initiated. For instance, if a forecast is implemented with a lead time of 48 hours, the corresponding forecast model should only encompass NWP variables generated with a minimum lead time of 48 hours.

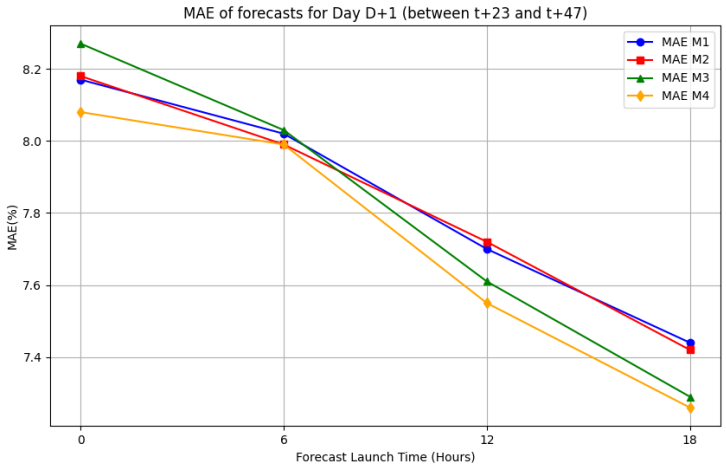

Figure 4.4: MAE for the different models tested for Bessakerfjellet

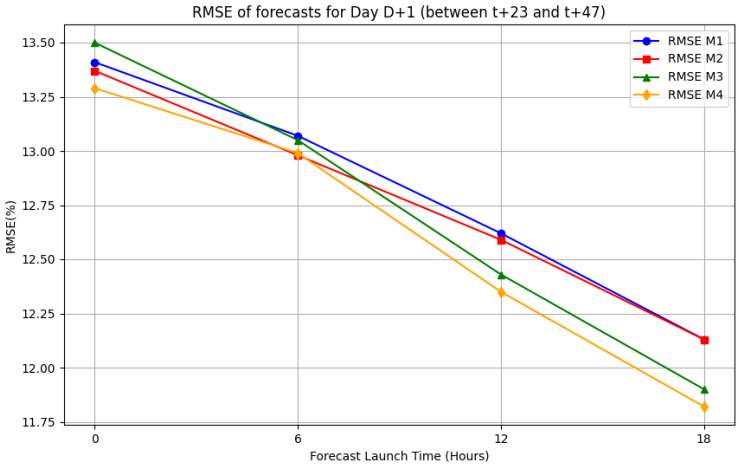

Figure 4.5: RMSE for the different models tested for Bessakerfjellet

Table 4.7: Bessakerfjellet metrics for lead time t+48|t

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+48|t	Group A	8.17	13.41	1.78
	Group A, Group B	8.18	13.37	1.77
	Group A, Group B, Group C	8.27	13.50	1.78
	Group A, Group B, Group D	**8.08**	**13.29**	**1.76**
t+42|t	Group A	8.02	13.07	1.74
	Group A, Group B	7.99	12.98	1.73
	Group A, Group B, Group C	8.03	13.05	1.73
	Group A, Group B, Group D	**7.99**	**12.99**	**1.72**
t+36|t	Group A	7.70	12.62	1.66
	Group A, Group B	7.72	12.59	1.65
	Group A, Group B, Group C	7.61	12.43	1.64
	Group A, Group B, Group C, Group D	**7.55**	**12.35**	**1.63**
t+30|t	Group A	7.44	12.13	1.59
	Group A, Group B	7.42	12.13	1.59
	Group A, Group B, Group C	7.29	11.90	1.56
	Group A, Group B, Group C, Group D	**7.26**	**11.82**	**1.55**

The results illustrated in figures 4.4, 4.5 and table 4.7 show that as the lead time decreases, the deviations of the forecasts from the actual values also decrease. This is due to the use of weather forecast updates that occur during the day, which will consequently improve the quality and accuracy of wind production forecasts. It is also possible to conclude that the models differ between forecasts released at different times within the same day. For example, the use of group 3 variables, which concern the lags and leads of meteorological variables, did not lead to an improvement in forecasts for lead times t+48|t and t+42|t, but for forecasts with lead times t+36|t and t+30|t, their inclusion led to a reduction in all measured errors. It can also be seen that the forecasts generated at 18h with 48 hours in advance show a 12.92% improvement in pinball losses on the forecasts generated at 00h on the same day.

For all lead times, with the exception of forecasts with lead time t+24|t, presented in table 4.8, where the inclusion of transformed meteorological variables (means, standard deviations,

Table 4.8: Bessakerfjellet metrics for lead time t+24|t

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+24|t	Group A	6.97	11.44	1.50
	Group A, Group B	7.03	11.49	1.51
	Group A, Group C	**6.90**	**11.32**	**1.49**
	Group A, Group C, Group D	6.92	11.32	1.48
t+18|t	Group A	6.62	10.96	1.43
	Group A, Group B	6.62	10.94	1.42
	Group A, Group B, Group C	**6.47**	**10.71**	**1.40**
	Group A, Group B, Group C, Group D	6.52	10.73	1.40
t+12|t	Group A	6.61	10.92	1.43
	Group A, Group B	6.57	10.85	1.43
	Group A, Group B, Group C	6.57	10.76	1.41
	Group A, Group B, Group C, Group D	**6.55**	**10.71**	**1.40**
t+6|t	Group A	6.32	10.25	1.36
	Group A, Group B	6.30	10.19	1.36
	Group A, Group B, Group C	6.28	10.15	1.35
	Group A, Group B, Group C, Group D	**6.23**	**10.03**	**1.34**

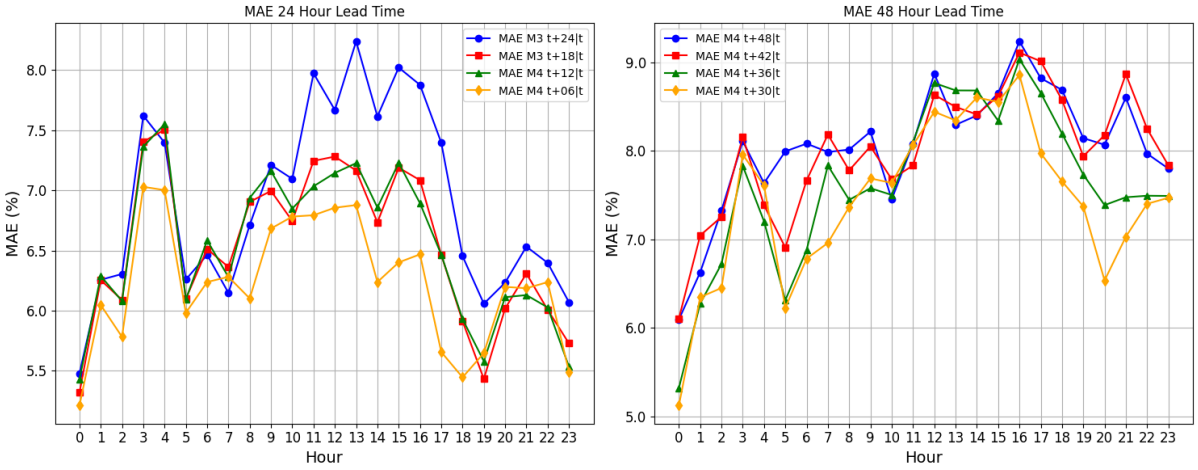

Figure 4.6: Hourly MAE for the final models employed for each lead time

variances) did not yield any improvement, adding more groups generally improves the forecast performance (i.e., lower MAE, RMSE, and pinball loss).

The performance metrics generally improve as the lead time decreases (from t+24|t to t+6|t), indicating that shorter-term forecasts are more accurate, which corroborates what was mentioned earlier above that the inclusion of more current NWPs leads to more accurate and reliable forecasts.

The differences in performance metrics between the variable group combinations become more pronounced at longer lead times (e.g., t+24|t), highlighting the importance of incorporating more relevant variable groups for longer-term forecasts.

Adding calendar variables (Group D) to the combination of meteorological variables and their lags and leads further improves the performance, though the improvement is marginal compared to adding lags and leads.

Figure 4.6 shows the hourly MAE of the final model used for each forecast at different lead times. At certain hours, particularly between 11h and 17h, the error appears to be higher. This

might indicate that these times correspond to periods of greater climatic instability and, therefore, more variable wind patterns, resulting in lower accuracy in wind production forecasts. Additionally, it can also be seen that the accuracy of the forecasts improves as the lead time decreases.

The figures 4.7 and 4.8 depict a 2-week graphical representation of the probabilistic forecast for the first-level forecasts at lead times t+48|t and t+06|t. They showcase the 5/95% confidence intervals, the 50% quantile utilized as the point forecast, and the actual production values derived from the production history of the specified wind farm.

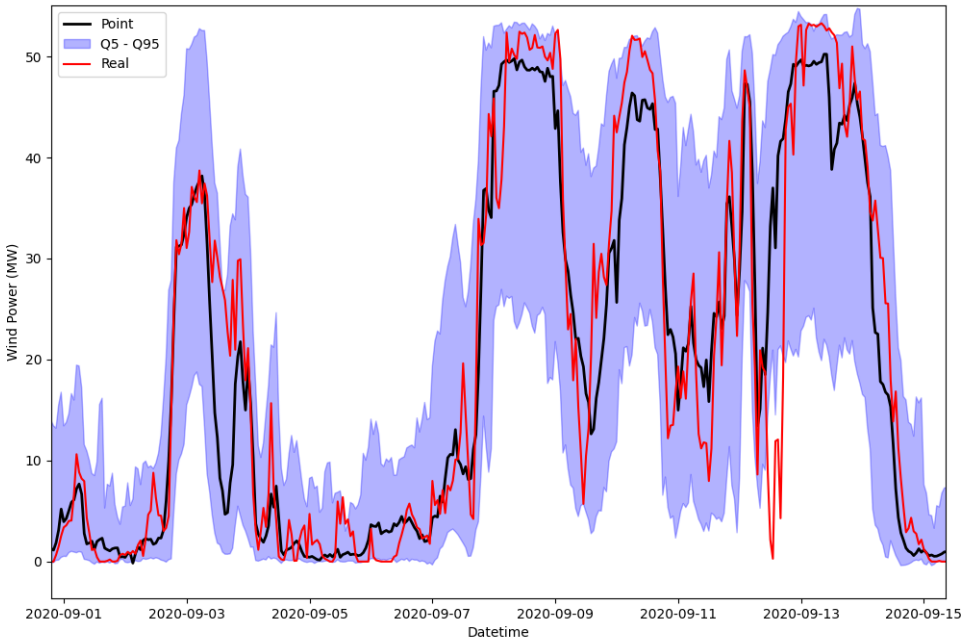

Figure 4.7: Forecasted wind power production for lead time t+48|t for wind farm Bessakerfjellet

Upon reviewing the probabilistic forecasts presented in the figures 4.7 and 4.8, it is apparent that there is an overall enhancement in the sharpness of the forecasts (i.e., size of the 5%-95% interval) as the lead time decreases. This improvement also correlates with a reduction in pinball losses, decreasing from 1.78 for forecasts with lead time t+48|t to 1.34 for forecasts with lead time t+06|t. The interval forecast consists of a lower bound (Q05) and an upper bound (Q95), which represent two forecasted quantiles. The midpoint of the interval is commonly assumed to be the point forecast, representing the production forecasts for the 50th quantile. As the lead time decreases, the differences between the actual values and the Q50 forecasts also decrease, as indicated by the analysis of the MAE and RMSE, which demonstrated a reduction in these metrics.

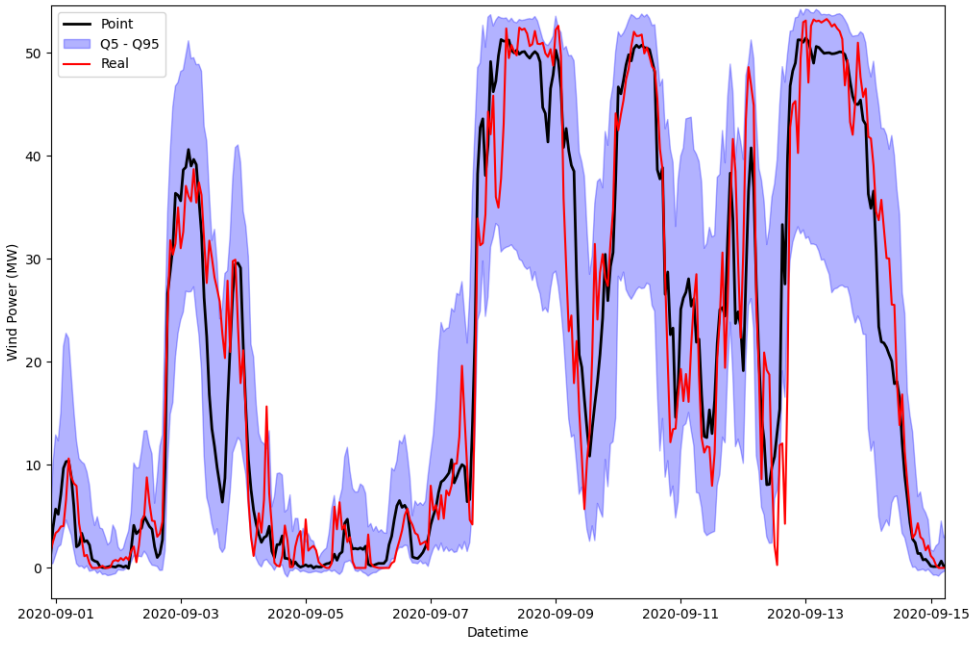

Figure 4.8: Forecasted wind power production for lead time t+06|t for wind farm Bessakerfjellet

4.3.4.2 Result analysis for Fakken

The results of the various models tested for the Fakken wind farm are presented in figures 4.9, 4.10 and tables 4.9 and 4.10, demonstrating the influence of different variable groups on wind forecasts across the time horizon.

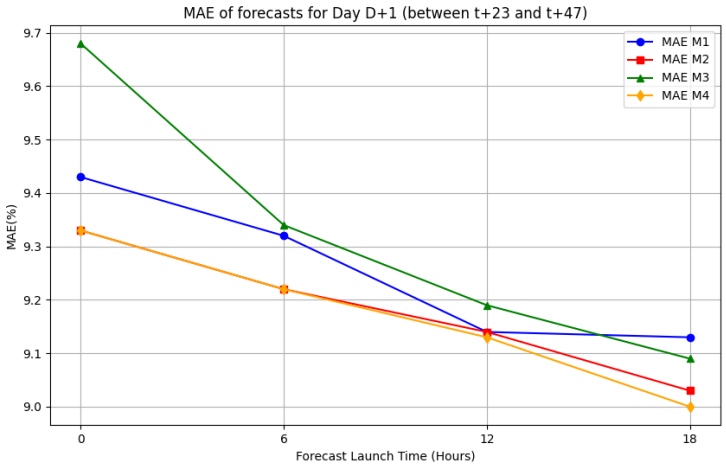

Figure 4.9: MAE for the different models tested for Fakken

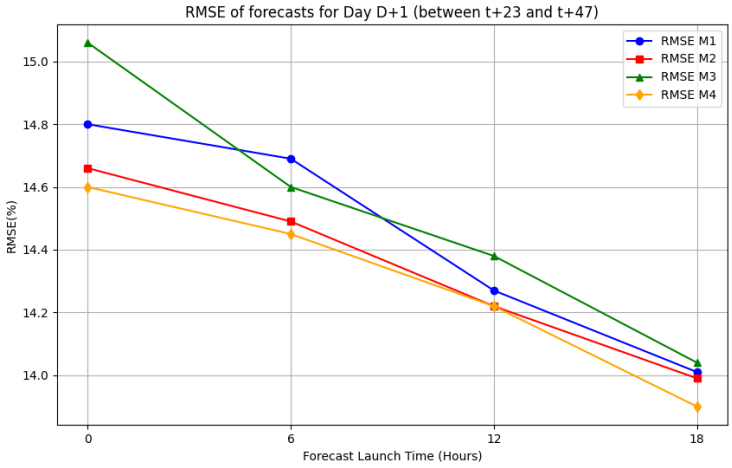

Figure 4.10: RMSE for the different models tested for Fakken

Table 4.9: Fakken metrics for lead time t+48|t

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+48|t	Group A	9.43	14.80	2.15
	Group A, Group B	9.33	14.66	2.14
	Group A, Group B, Group C	9.68	15.06	2.17
	Group A, Group B, Group D	**9.33**	**14.60**	**2.14**
t+42|t	Group A	9.32	14.69	2.14
	Group A, Group B	9.22	14.49	2.10
	Group A, Group B, Group C	9.34	14.60	2.12
	Group A, Group B, Group D	**9.22**	**14.45**	**2.10**
t+36|t	Group A	9.14	14.27	2.07
	Group A, Group B	9.14	14.22	2.06
	Group A, Group B, Group C	9.19	14.38	2.08
	Group A, Group B, Group D	**9.13**	**14.22**	**2.05**
t+30|t	Group A	9.13	14.01	2.04
	Group A, Group B	9.03	13.99	2.03
	Group A, Group B, Group C	9.09	14.04	2.05
	Group A, Group B, Group D	**9.00**	**13.90**	**2.03**

The results shown in table 4.9 show that the combination of Group A, Group B, and Group D (meteorological variables + transformed meteorological variables + calendar variables) consistently shows the best performance across all metrics and lead times.

Adding Group D which refers to calendar variables, namely the cosine and sine transformations of the hour of the day, tends to provide the best improvements when combined with Group A and Group B, showing lower MAE, RMSE, and pinball loss compared to the other combinations.

Adding lags and leads of meteorological variables does not consistently improve the performance and sometimes increases the forecast errors. For longer lead times, the inclusion of more relevant variable groups becomes increasingly important for maintaining forecast accuracy.

Table 4.10: Fakken metrics for lead time t+24|t

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+24|t	**Group A**	**8.81**	**13.83**	**1.99**
	Group A, Group B	8.84	13.83	1.99
	Group A, Group C	8.85	13.78	2.01
	Group A, Group D	8.90	13.95	2.00
t+18|t	Group A	8.71	13.68	1.98
	Group A, Group B	**8.70**	**13.67**	**1.96**
	Group A, Group B, Group C	8.88	13.85	1.98
	Group A, Group B, Group D	8.74	13.76	1.97
t+12|t	Group A	8.51	13.48	1.91
	Group A, Group B	**8.42**	**13.27**	**1.90**
	Group A, Group B, Group C	8.48	13.41	1.91
	Group A, Group B, Group D	8.44	13.30	1.91
t+6|t	Group A	8.32	13.14	1.88
	Group A, Group B	8.27	13.07	1.86
	Group A, Group B, Group C	8.30	13.08	1.86
	Group A, Group B, Group D	**8.21**	**13.02**	**1.86**

For lead time t+24|t presented in table 4.10, the study results indicate that including transformed meteorological variables (Group B) generally enhances the performance of the gradient-boosting regressor for wind energy forecasts. Furthermore, incorporating calendar variables (Group D) also leads to additional improvements, especially noticeable in lead time t+6|t. The combination of variable groups A, B and D shows the best the best performance in this regard.

However, the addition of lags and leads of meteorological variables (Group C) does not consistently improve forecast accuracy and may, in some cases, even degrade it. Therefore the optimal combination for the most accurate forecasts involve meteorological variables and their respective transformations, with calender variables also making a positive contribution, especially for shorter lead times.

It is also worth noting that performance metrics generally improve as the lead time decreases (from t+24|t to t+6|t), indicating that shorter-term forecasts are more accurate. The forecasts with a lead time of t+6|t show a 6.53% improvement over those with a lead time of t+24|t.

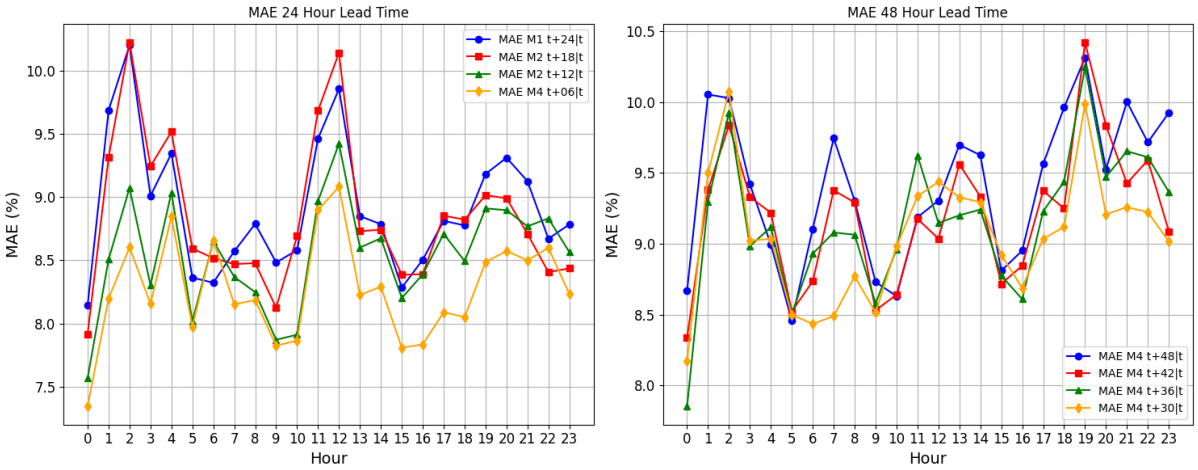

Figure 4.11: Hourly MAE for the final models employed for each lead time

From the results in figure 4.11, it can be seen that the range of variation in MAEs between forecasts launched with different lead times is low for almost all hours of the day, which indicates that the accuracy of wind forecasts for the Fakken wind farm is not very dependent on the forecast horizon.

The wind forecasts for the Fakken wind farm exhibited greater uncertainty compared to those for the Bessakerfjellet wind farm, resulting in less precise forecasts. This discrepancy can be explained by the historical production distribution of the Fakken wind farm. The dataset revealed low production during the training period, posing challenges for the model to adjust forecasts during production peaks, thereby leading to higher uncertainty. This suggests what was already mentioned above, that the quality of the forecasts for the Fakken wind farm is less sensitive to the time horizon. This behavior can be observed when examining the graphs in 4.12 and 4.13. It is apparent that the confidence interval is narrower during low production periods and wider during high production periods, signifying greater uncertainty during the latter. Furthermore, both RMSE and MAE errors were higher during production peaks, indicating that the models were more adept at adjusting forecasts to real values during periods of low production.

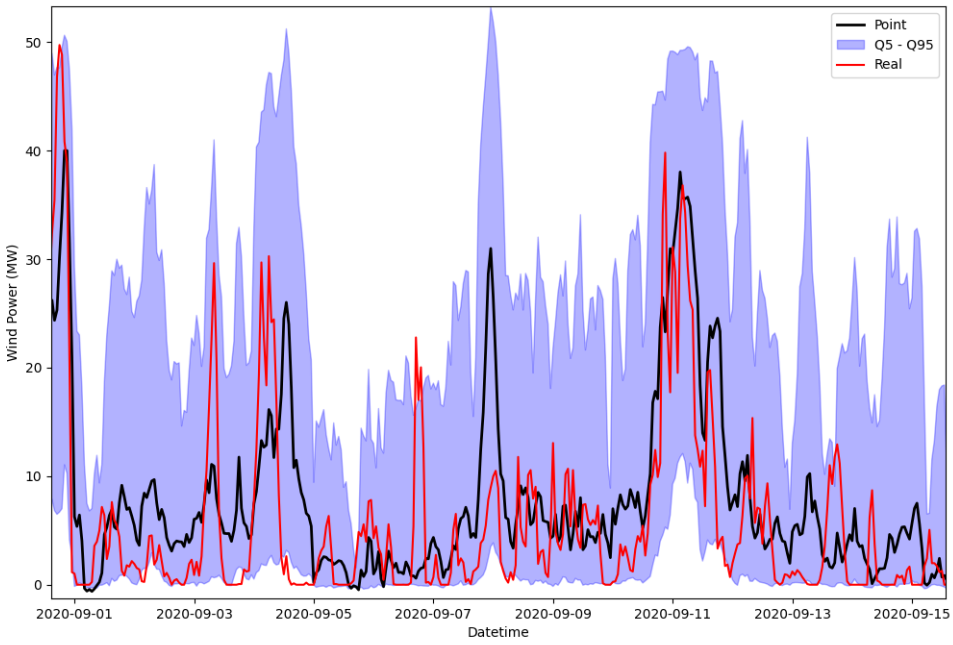

Figure 4.12: Forecasted wind power production for lead time t+48|t for wind farm Fakken

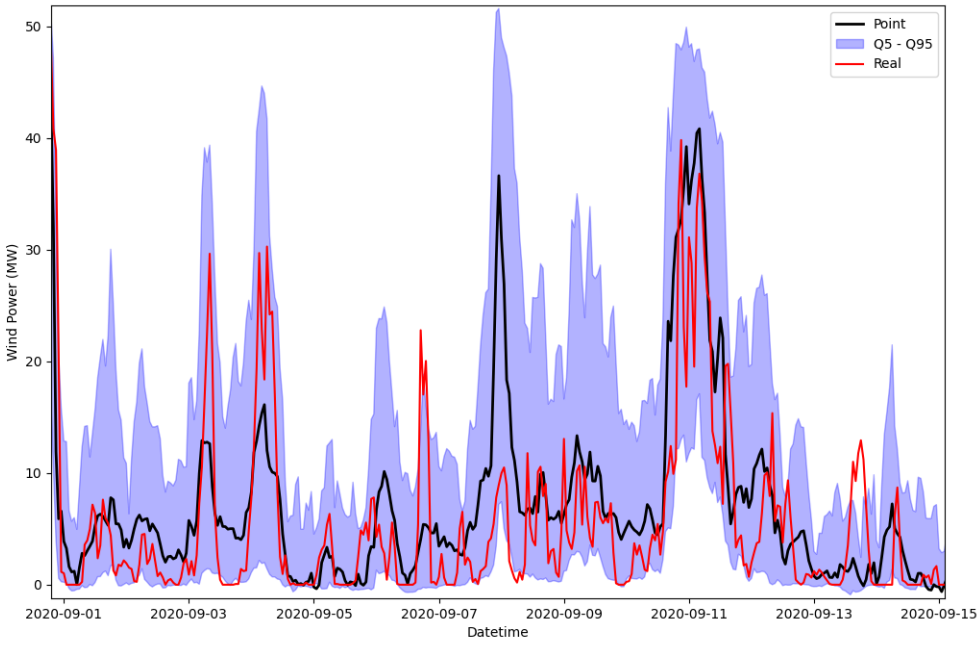

Figure 4.13: Forecasted wind power production for lead time t+06|t for wind farm Fakken

4.3.4.3 Result analysis for Lista

Applying the same rationale as in the previous subsection, the results for the different models of the Lista wind farm were as follows:

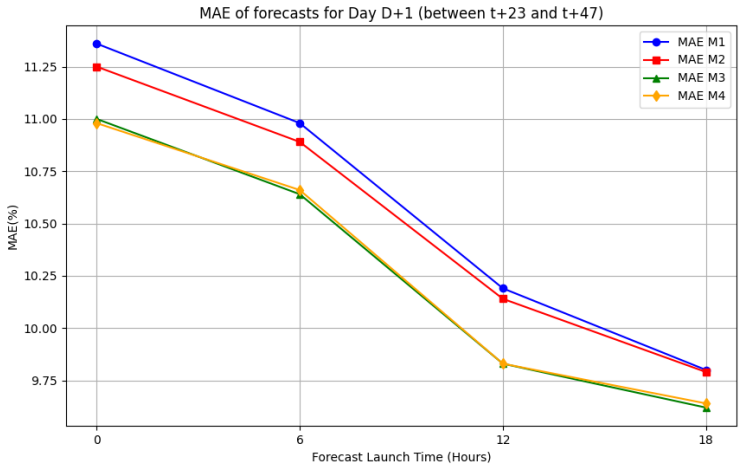

Figure 4.14: MAE for the different models tested for Lista

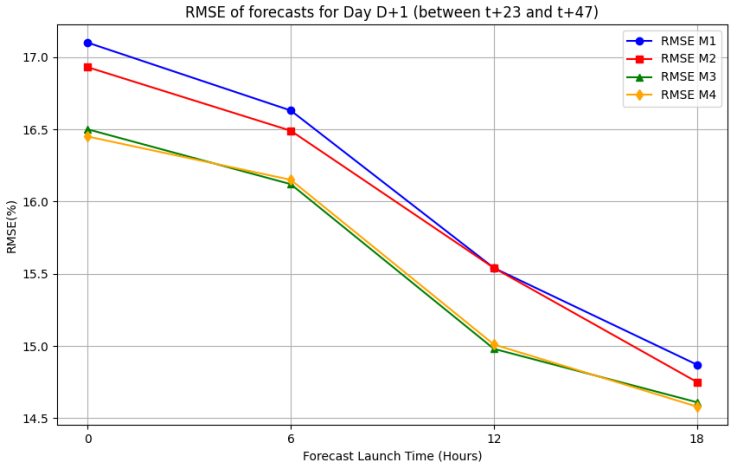

Figure 4.15: RMSE for the different models tested for Lista

Table 4.11: Lista metrics for lead time t+48lt

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+48lt	Group A	11.36	17.10	3.04
	Group A, Group B	11.25	16.93	3.03
	Group A, Group B, Group C	11.00	16.50	2.96
	Group A, Group B, Group C, Group D	**10.98**	**16.45**	**2.95**
t+42lt	Group A	10.98	16.63	2.94
	Group A, Group B	10.89	16.49	2.91
	Group A, Group B, Group C	**10.64**	**16.12**	**2.87**
	Group A, Group B, Group C, Group D	10.66	16.15	2.87
t+36lt	Group A	10.19	15.54	2.72
	Group A, Group B	10.14	15.54	2.71
	Group A, Group B, Group C	9.83	14.98	2.63
	Group A, Group B, Group C, Group D	**9.83**	**15.01**	**2.62**
t+30lt	Group A	9.80	14.87	2.62
	Group A, Group B	9.79	14.75	2.62
	Group A, Group B, Group C	**9.62**	**14.61**	**2.57**
	Group A, Group B, Group C, Group D	9.64	14.58	2.57

Analyzing figures 4.4 and 4.5 and table 4.11, it is evident that incorporating transformed meteorological variables (Group B) and lags and leads of meteorological variables (Group C) significantly improves the performance of the GBT for wind energy forecasts. Adding calendar variables (Group D) also contributes positively, particularly for longer lead times.

The higher values for MAE, RMSE, and pinball losses in the forecasts for the above wind farm compared to others can be attributed to a combination of factors, such as the wind patterns that might be highly variable or influenced by local meteorological phenomena, which can increase the difficulty of accurate predictions. This phenomena can be assessed by analysing the power distribution of each wind farm.

The wind farm's power histogram in 4.16 shows a wider distribution of power outputs compared to the other wind farms illustrated in 4.17 and 4.18. This spread in power generation indi-

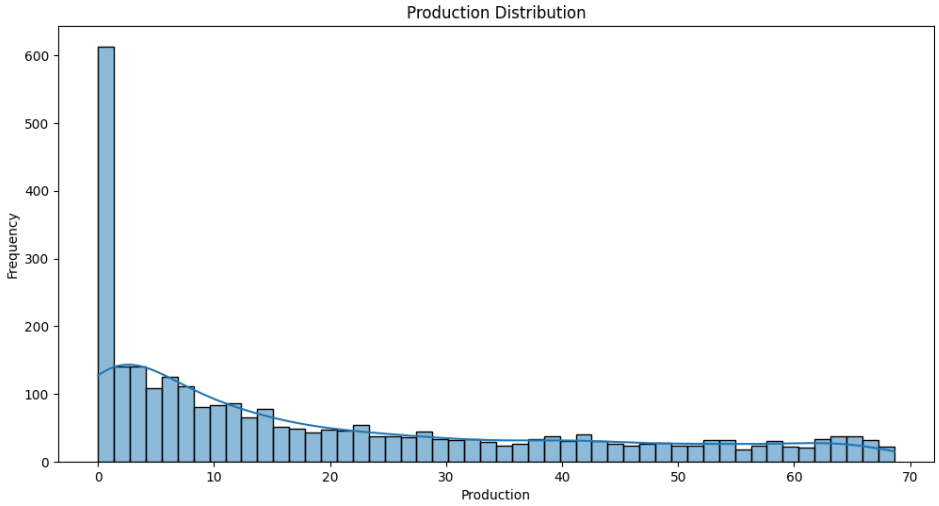

Figure 4.16: Histogram for Wind Farm Lista

cates greater uncertainty in wind conditions, making it more challenging to forecast future production and, therefore, potentially leading to increased MAE, RMSE, and pinball losses.

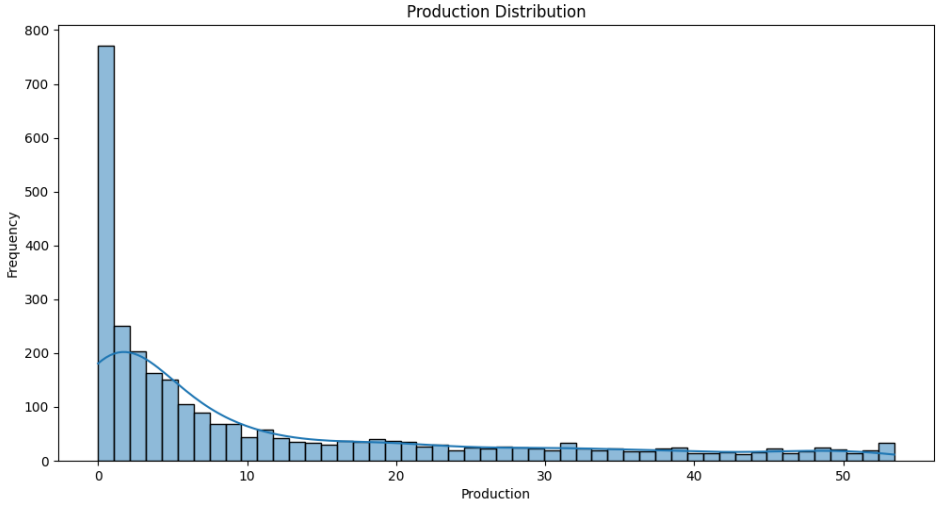

Figure 4.17: Histogram for Wind Farm Bessakerfjellet

When analyzing the historical wind production for the Fakken and Bessakerfjellet wind parks, there is a smaller distribution of production, which is more concentrated in the lower values, thus indicating less uncertainty and, therefore, easier to forecast, which is reflected in smaller deviations from the actual values. Consequently, a less dispersed power distribution will contribute to lower values for MAE, RMSE, and pinball losses.

The results presented in table 4.12, indicate that including more variable groups in the model generally improves the forecasting accuracy, and shorter lead times yield more accurate predic-

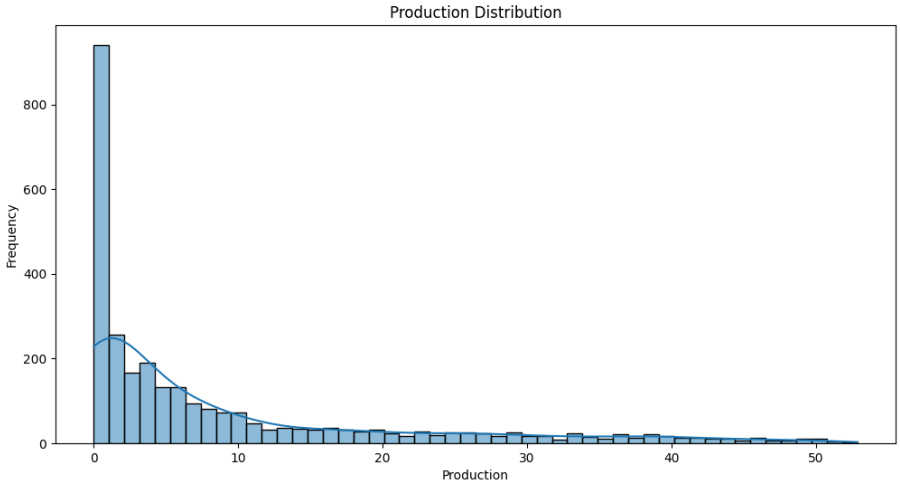

Figure 4.18: Histogram for Wind Farm Fakken

Table 4.12: Lista metrics for lead time t+24|t

Lead Time	Variable Groups	MAE(%)	RMSE(%)	Pinball Loss
t+24|t	Group 1	9.24	14.07	2.49
	Group A, Group B	9.20	14.10	2.48
	Group A, Group B, Group C	9.08	13.87	2.44
	Group A, Group B, Group C, Group D	**9.05**	**13.79**	**2.44**
t+18|t	Group A	8.55	12.75	2.30
	Group A, Group B	8.54	12.64	2.29
	Group A, Group B, Group C	8.34	12.52	2.24
	Group A, Group B, Group C, Group D	**8.30**	**12.44**	**2.24**
t+12|t	Group A	7.77	11.49	2.09
	Group A, Group B	7.76	11.51	2.08
	Group A, Group B, Group C	7.60	11.32	2.05
	Group A, Group B, Group C, Group D	**7.58**	**11.27**	**2.04**
t+6|t	Group A	7.32	10.79	1.97
	Group A, Group B	7.35	10.86	1.98
	Group A, Group C	7.22	10.71	1.95
	Group A, Group C, Group D	**7.19**	**10.66**	**1.94**

tions. For the best predictive performance, it is beneficial to use a comprehensive set of variables and focus on shorter lead times. The most significant improvements are seen when moving from using only Group A to including Group B, and further improvements are observed when Groups C and D are added. For shorter lead times, the predictions are more accurate, which is reflected in lower MAE, RMSE, and pinball loss values.

The best performance is observed for the shortest lead time (t+6|t) with the least complex forecasting model incorporating almost all but group B of variables (Group A, Group C, Group D): MAE: 7.19%, RMSE: 10.66%, pinball loss: 1.94. The worst performance is observed for the longest lead time (t+24|t) with the least comprehensive variable set (Group A): MAE: 9.24%, RMSE: 14.07%, Pinball Loss: 2.49.

Based on the findings presented in figure 4.19, it is apparent that there is significant variation in

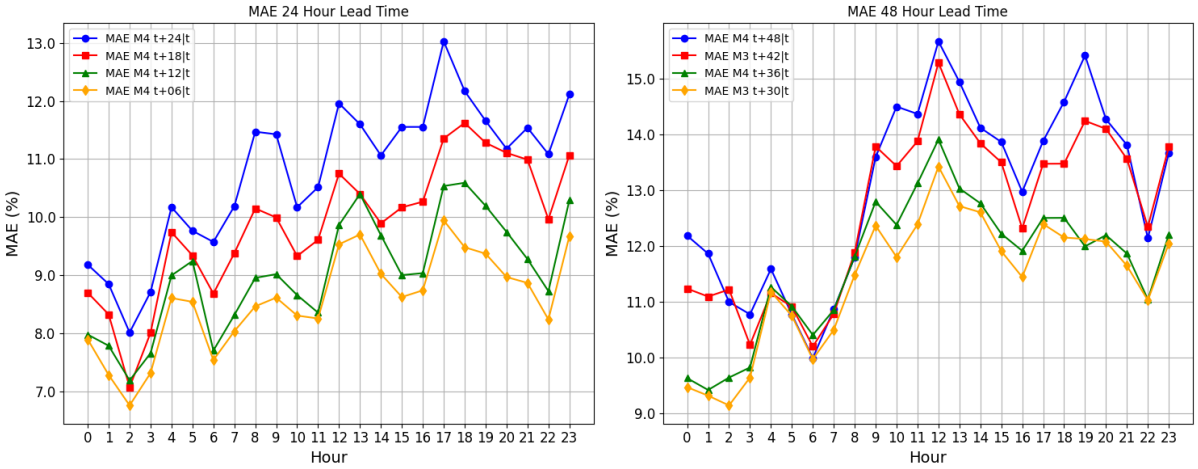

Figure 4.19: Hourly MAE for the final models employed for each lead time

MAEs across forecasts with different lead times throughout the day. This clearly indicates that the accuracy of wind forecasts for the Lista wind farm is heavily influenced by the forecast horizon. As a result, it may be more advantageous to only produce forecasts for this wind farm with a 24-hour forecast horizon, as NWP generated within this timeframe exhibits greater predictive capability compared to those generated with a lead time exceeding 24 hours.

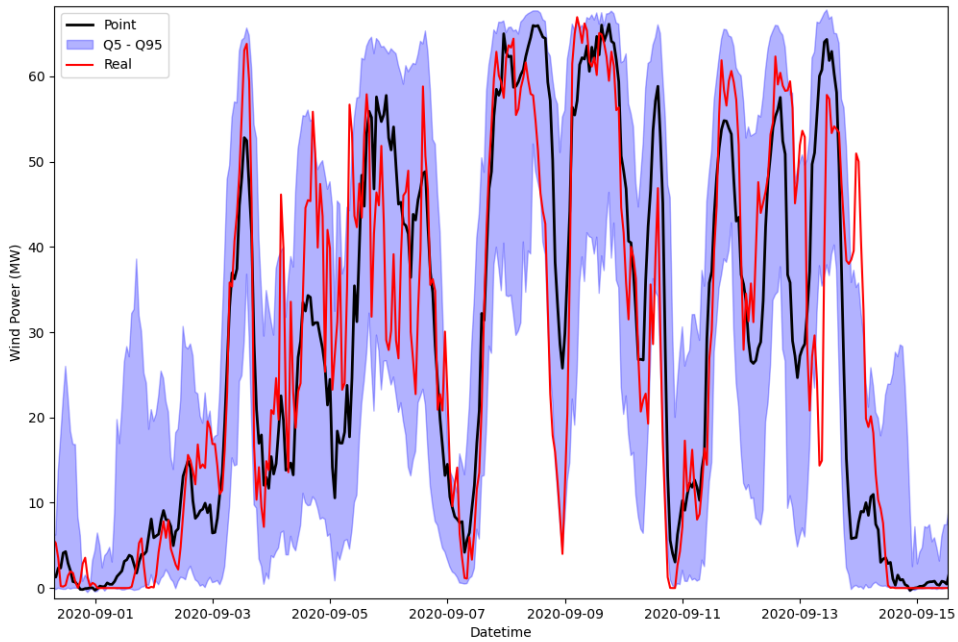

Figure 4.20: Forecasted wind power production for lead time t+48|t for wind farm Lista

In the case of the Lista wind farm, it was noted in graphs 4.20 and 4.21 that the accuracy

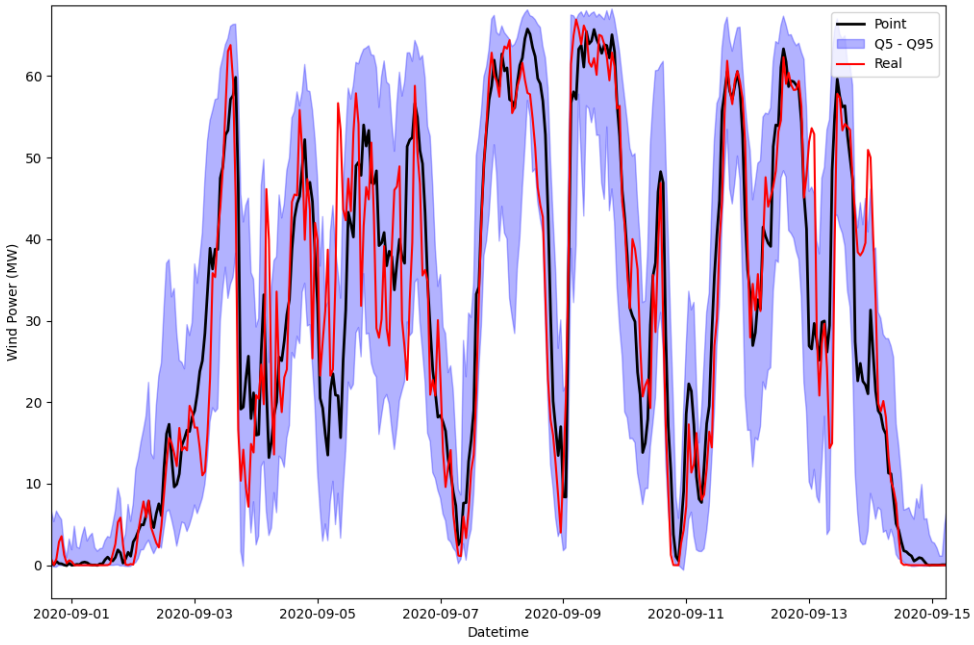

Figure 4.21: Forecasted wind power production for lead time t+06|t for wind farm Lista

of wind forecasts is closely tied to the forecast horizon. Forecasts made 48 hours in advance exhibited high uncertainty, with pinball losses reaching 2.95, reflecting wide confidence intervals. In contrast, forecasts made 6 hours in advance showed improved reliability and sharpness, with pinball losses of 1.94, representing a 34.24% enhancement over the 48-hour forecasts. As a result, it can be concluded that the Lista wind farm demonstrated the most significant improvement in forecast accuracy compared to other wind farms. Consequently, it is advisable to opt for 24-hour lead time forecasts for this wind farm, as the 48-hour forecasts are associated with a notably high degree of uncertainty.

4.4 Second-Level Forecasts

As wind power integrates into the electrical grid, accurate forecasting becomes crucial for grid management. Traditional wind energy forecasting predicts output based on meteorological inputs, but understanding uncertainty in these forecasts is increasingly important.

In this context, second-level wind energy forecasting provides a means to predict the variability and uncertainty inherent in wind power forecasts. This method leverages the forecasts obtained from the first level, particularly the quantiles predicted, to model and predict the uncertainty of wind power production. As such the historical data available for second-level forecasting corresponds to the data used during the testing phase in the first-level forecasts.

The forecast time horizon is 48 hours, and after a first-level forecast is launched, a second-level forecast is also generated to forecast wind production in the next NWP updates for each wind farm. In this dissertation, five second-level forecasts were conducted for each wind farm. Initially, two second-level forecasts were generated at 00:00 with a 48-hour lead time to predict wind forecast variations in the next NWP updates from 00h00 to 06h00 (t+49|t+06) and from 00h00 to 18h00 (t+49|t+18). Subsequently, two more second-level forecasts at 00h00 were produced, but with a 24-hour lead time to predict variations from 00h00 to 06h00 (t+25|t+06) and from 00:00 to 18:00 (t+25|t+18). Finally, a forecast was made to predict wind production variations from 06h00 to 12h00, also with a 24-hour lead time (t+19|t+12). The more detailed results of each second-level forecast are presented in the appendix 5.

4.4.1 Training and Test Dataset

The second-level forecasting concept builds upon a set of probabilistic forecasts generated in the first-level. Consequently, the data available for second-level forecasts corresponds to the training period used for the first-level forecasts.

The first-level forecasts were created in two separate testing phases. The first testing phase covered a three-month period from June to August 2020, using data from a training phase that lasted from June 2019 to May 2020. To increase the size of the training dataset for the second-level forecasts, a sliding window approach was used for a three-month testing period, forecasting the months of September, October, and November using the same training data as the first testing phase. This resulted in a total dataset spanning six months. As such, data from June, July, August, September, and October 2020 was used for training, reserving the data from November 2020 for the test phase.

4.4.2 Benchmark model

In order to evaluate the effectiveness and added value of the second-level wind energy forecasting model, a benchmark model was established. This benchmark assumes a scenario where there is no variation in the quantile forecasts throughout the day. Specifically, the forecast made at 00h00 is assumed to remain constant for the entire day, regardless of subsequent updates in the Numerical Weather Predictions at 06h00, 12h00, and 18h00.

This static benchmark model serves as a baseline to assess whether the additional complexity and computational effort involved in predicting the fluctuations of quantile forecasts with NWP updates offer any significant improvement. By comparing the performance of the second-level forecasting model, which dynamically updates forecasts based on the evolving uncertainty captured through NWP updates, against this static benchmark, the practical value of incorporating the second-level forecasting approach can be determined.

The benchmark model operates under the assumption that the initial forecast made at 00h00 provides an adequate estimate of wind power production for the entire day. If the second-level forecasting model, which leverages gradient boosting regression and engineered input variables,

shows a substantial improvement in prediction accuracy and reliability over this static benchmark, it justifies the extra effort and resources invested in the second-level forecasting process. Conversely, if the performance gains are minimal, it may suggest that simpler forecasting approaches could suffice for operational purposes.

An improvement metric is used to quantify the improvement offered by the second-level forecasts over the benchmark model. This metric is calculated as follows:

$$Improvement = \left(1 - \frac{MAE_{Forecast}}{MAE_{Benchmark}}\right) \times 100 \tag{4.1}$$

Since the benchmark is essentially the quantile predicted at 00h00 and functions like a deterministic forecast, the improvement is measured by comparing the Mean Absolute Errors of the benchmark quantile and the 50% quantile (median) of the second-level prediction. This comparison focuses on how well the second-level model captures the variations and uncertainties in the wind power production forecasts throughout the day.

The analysis of the results will focus on key metrics such as Mean Absolute Error and Root Mean Square Error across different quantile ranges from 5% to 95% with 25% increments.

4.4.3 Results Analysis For Windfarm Bessakerfjellet

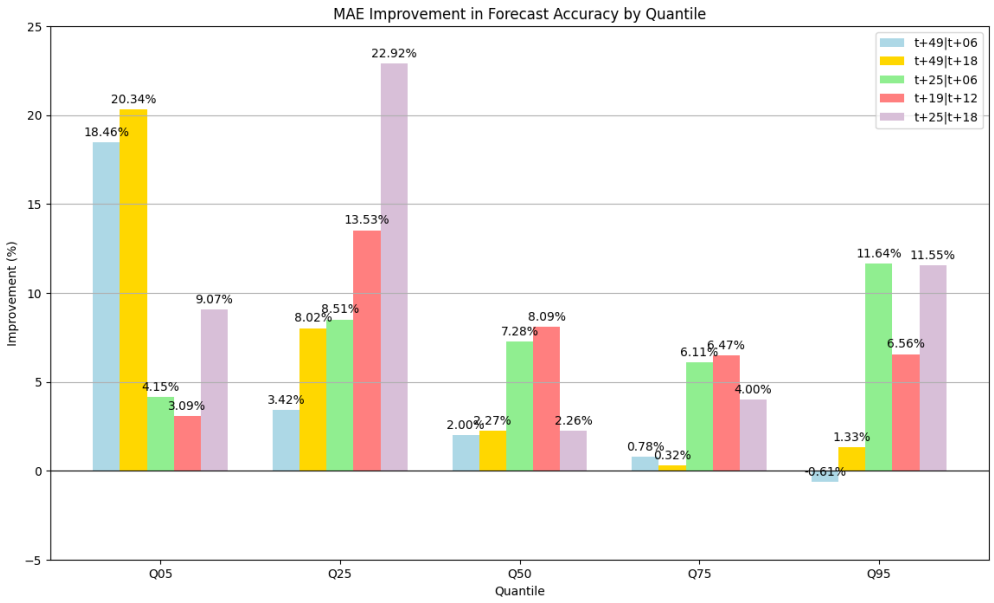

Figure 4.22: MAE improvements for windfarm Bessakerfjellet

The analysis revealed a significant improvement at the 5% quantile for lead times t+49|t+06 and t+49|t+18. However, as the quantile level increases, the benefits of the second-level forecasting model diminish progressively.

The overall improvement achieved by the second-level forecasting model for lead time t+49|t+18 was approximately 4.81%. This represents a modest enhancement, suggesting that the additional effort to predict fluctuations of the 06h forecasts using those generated at 00h with a 48h lead time may not be justified.

To better understand the impact of variability in wind production forecasts between different NWP updates on second-level forecast performance, tests were conducted to examine the effect of extending the forecast period between NWP updates. Specifically, a second-level forecast with a 48-hour lead time was generated at 00h00 to predict forecast fluctuations at 18h00. There was an improvement of 6.46% compared to the same benchmark model used for forecasts with a lead time of t+49|t+06. This suggests that assuming wind forecasts remain constant over 18 hours is not an accurate prediction of forecast conditions for that instant. It is, therefore, preferable to make the extra effort and carry out a second-level forecast to obtain a forecast that better characterizes the fluctuations in the wind.

Notably, for lead time t+49|t+06, there was a negative improvement for the 95% quantile. This might be attributed to the use of NWP variables generated with a long lead time, which tend to be less accurate and reliable compared to those with shorter lead times. Additionally, NWP variables inherently carry a forecast error that can be propagated when used in second-level forecasts, potentially introducing noise into the production forecasts.

Also, because the performance of each quantile varies, a different model is required to forecast each quantile accurately. This explains why the improvement results compared to the benchmark differ for each quantile. The information extracted by the gradient boosting regressor from each model is different, resulting in varying forecasting capabilities for each quantile.

Overall improvements of 7.54%, 7.55%, and 9.96% were observed for forecasts with lead times of t+25|t+6, t+19|t+12, and t+25|t+18, respectively. This further confirms the earlier observation that better improvements are obtained for shorter lead times, particularly for forecasts made 24 hours in advance. This is attributed to using more reliable and accurate data, which enhances the model's ability to predict the actual values.

The results presented in figures 4.23 and 4.24 pertain to the second-level forecasts of the Q25 quantile with lead times t+49|t+06 and t+49|t+18.

In figure 4.23, the real values corresponds to the forecasted values for Q25 at 06h obtained with the first-level forecasts with 42 hours of lead time and the figure 4.24 the real values are the forecasted values for Q25 at 18h obtained with the first-level forecasts with 30 hours of lead time.

Upon analyzing the data, it becomes clear that as the forecast interval between NWP updates widens, the results show a noticeable loss of forecasting accuracy. It is evident that the deviations between the real values and the forecasted values obtained with second-level forecasting, are more pronounced in figure 4.24.

After analyzing the results of the Q25 quantile forecast with a lead time of t+25|t+06, depicted in Figure 4.25, the prediction model obtained MAE and RMSE of approximately 2.58% and 4.42% respectively. This represents an enhancement over the forecasts with a lead time of t+49|t+06, which showed MAE and RMSE of 3.39%, 5.01%. Despite the temporal distance between the

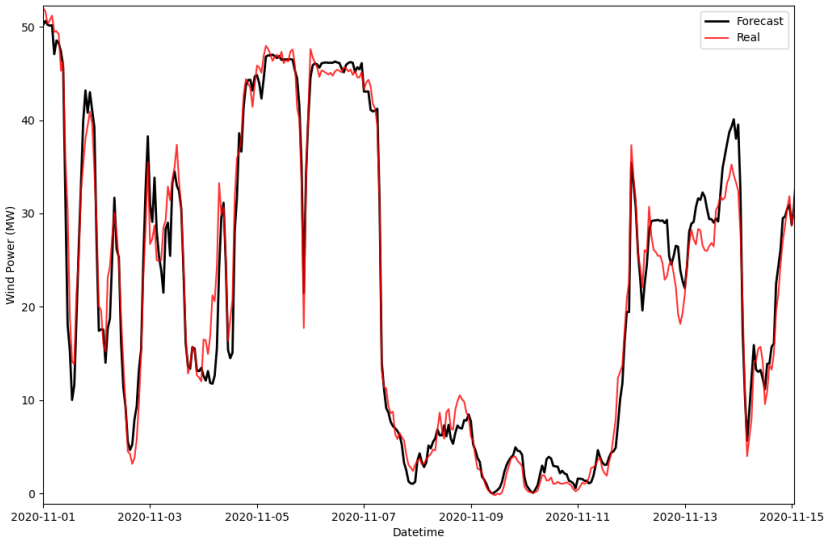

Figure 4.23: Second-level forecasts of Q25 for lead time t+49|t+06

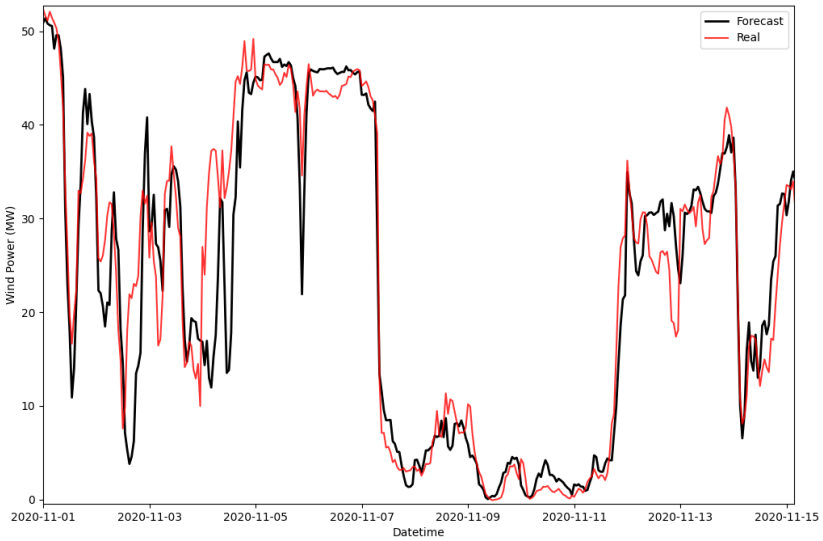

Figure 4.24: Second-level forecasts of Q25 for lead time t+49|t+18

NWP updates being the same for both forecasts, the use of data with a shorter lead time in the t+25|t+06 forecasts may account for their improved accuracy, as NWP variables forecasted with shorter lead times are often characterized by being more reliable and accurate and thus possess a greater predictive capability. This trend was consistently observed across most of the remaining wind farms, generally resulting in superior outcomes over a 24-hour time horizon.

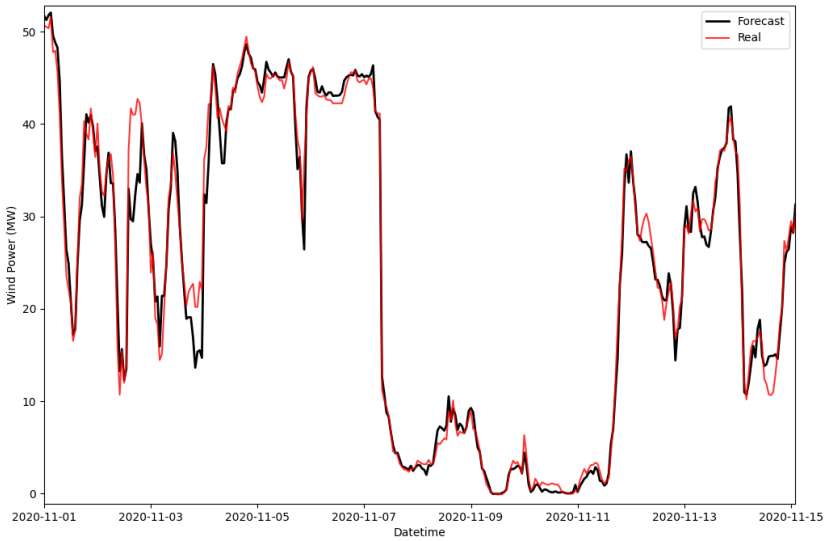

Figure 4.25: Second-level forecasts of Q25 for lead time t+25|t+06

Next, the average hourly MAE improvements compared to the benchmark were analyzed to understand the performance of certain second-level forecasting models throughout the day.

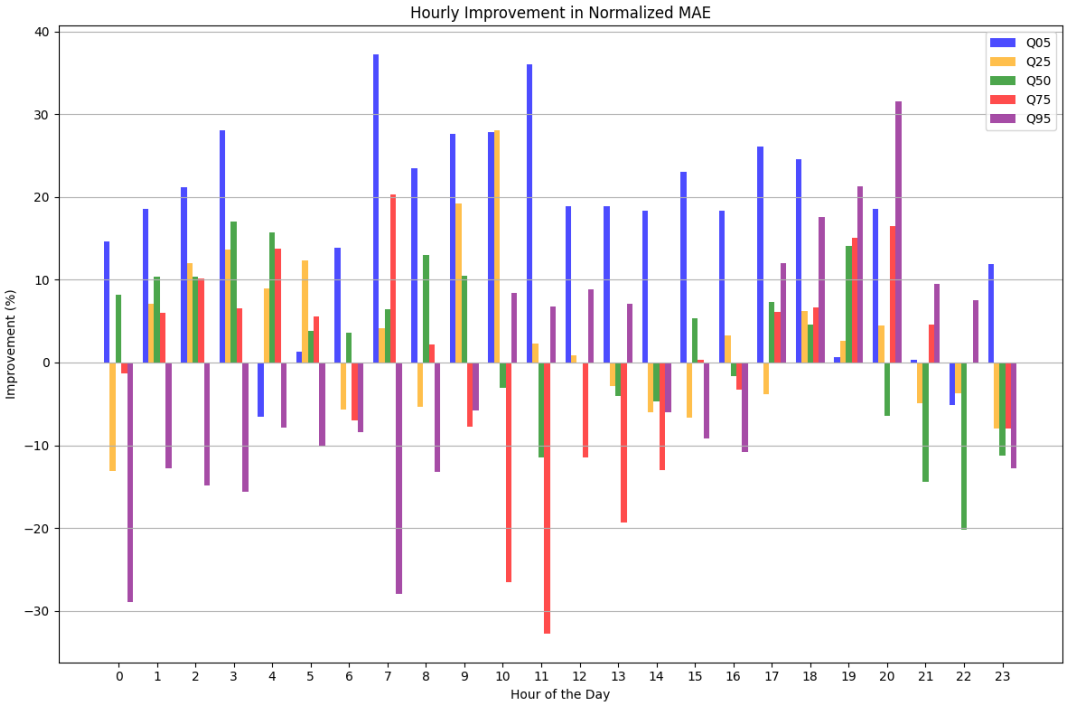

Figure 4.26: Bessakerfellet hourly MAE improvements for lead time t+49|t+06

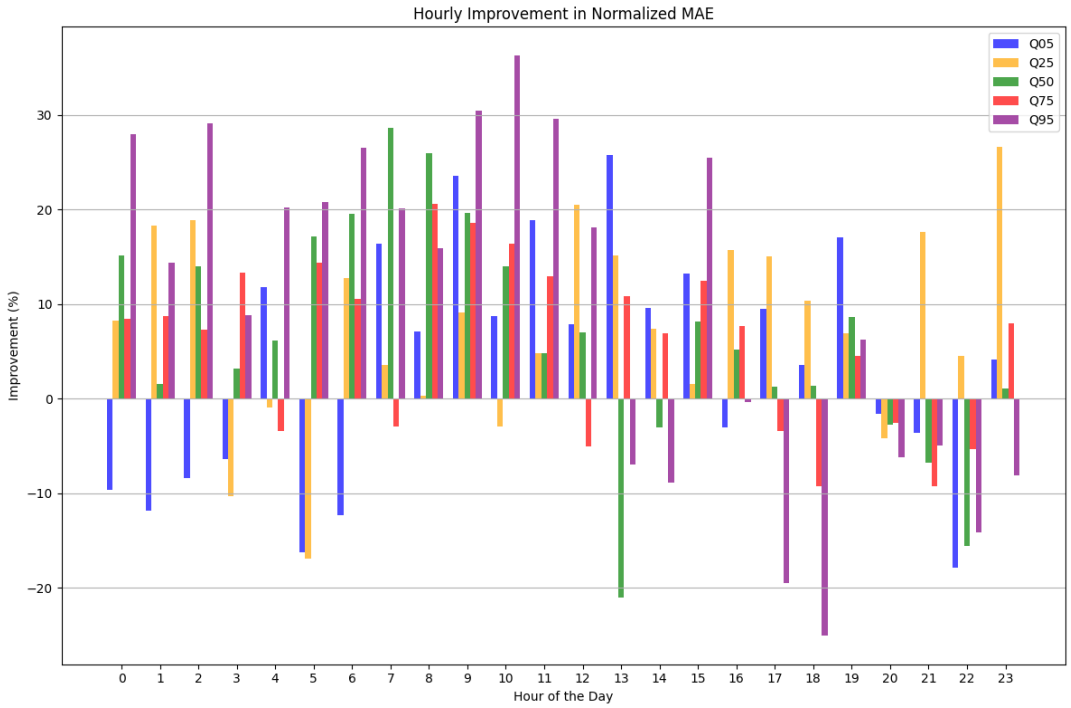

Figure 4.27: Bessakerfellet hourly MAE improvements for lead time t+25|t+06

The data depicted in the figure 4.26 illustrates that the Q05 quantile exhibited substantial improvements across most hours, with an average enhancement of 18.49% compared to the benchmark, as indicated in the figure 4.22. The other quantiles showed comparatively lesser improvements, with Q95 displaying only marginal enhancements between 10h00 and 13h00, and 17h00 and 22h00, resulting in a minor 0.61% deterioration compared to the benchmark.

When analyzing the hourly improvements for the second level forecast for lead time t+25|t+06 in figure 4.27, it was observed that the forecasts for all the quantiles generally showed improvements in most hours. However, the Q05 quantile between 00h00 and 06h00 and between 20h00 and 22h00 showed negative improvements compared to the benchmark. This resulted in a more modest overall improvement of 4.15%, as shown in figure 4.22.

4.4.4 Results for Windfarm Fakken

As before, 5 second-level forecasts were launched for a 48-hour time horizon generated with lead times of t+49|t+06, t+49|t+18, t+25|t+06, t+19|t+12 and t+25|t+18, obtaining improvements of 2.81%, 5.56%, 12.51%, 9.68% and 15.27% respectively as illustrated in figure 4.28.

The improvements over the benchmark models seen at the Fakken wind farm were similar to those at the Bessakerfjellet wind farm, indicating that the quality of the wind forecasts is not dependent on the wind farm's characteristics.

The forecasts produced at 00h00 to anticipate fluctuations at 18h00, specifically the t+49|t+18 and t+25|t+18 forecasts, outperformed those created at 00h00 for predicting fluctuations at 06h00, which were the t+49|t+06 and t+25|t+06 forecasts. This indicates that the variability of forecasts

between different updates throughout the day might influence the effectiveness of second-level forecasts. During periods of increased fluctuations in wind production, second-level forecasts become more valuable, offering greater potential to correct deviations between measured values and first-level forecasts.

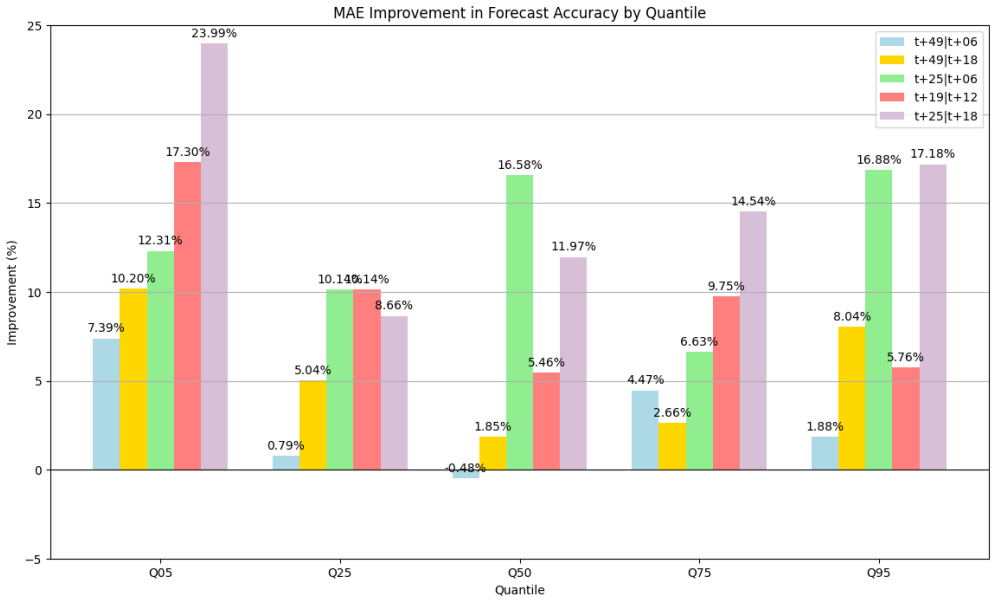

Figure 4.28: MAE improvements for windfarm Fakken

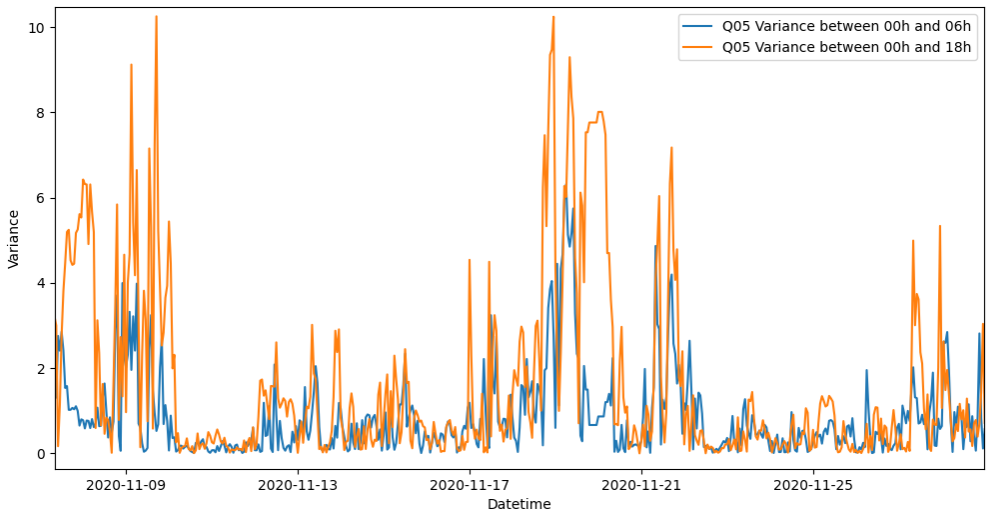

Figure 4.29: Q05 variance between 00h and 06h compared to the variance between 00h and 18h

The graph in 4.29 demonstrates that there is greater variability in wind forecasts for the quantile Q05 with a 48-hour lead time between 00h00 and 18h00 (t+49|t+18) compared to forecasts between 00h00 and 06h00 (t+49|t+06).

However, it was found that the improvement over the benchmark model (model comprised of wind forecasts generated at 00h00) is greater when there is more variability, with an improvement of 10.20% for t+49|t+18 and 7.39% for t+49|t+06. The features introduced into the forecast model for t+49|t+18 have proved to be effective in countering the deviations between the forecasts generated at 00h00 and the target values generated at 18h00, resulting in a significant improvement over the naive model.

The improvement in MAE values compared to the benchmark for the Fakken wind farm appears to rely more on the forecast horizon than the Bessakerfjellet wind farm. Shorter forecast horizons, specifically the 24-hour time horizons, display greater improvements compared to the 48-hour forecast horizons.

The models used for t+25|t+06, t+19|t+12, and t+25|t+18 forecasts are comprised of more reliable and accurate meteorological variables, produced with shorter lead times, providing greater predictive capacity for the models. These models also integrate meteorological variables from earlier instants, such as wind speeds at 10 and 100 meters altitude forecasted at 18h00 the previous day, which aid in reducing RMSE and MAE errors.

Conversely, models with lead times of t+49|t+06 and t+49|t+18, using meteorological variables generated earlier (NWP generated at 18h00 with a 72-hour lead time), do not positively contribute to the quality of the forecasts. Their use introduces noise and results in higher MAE and RMSE errors. Therefore, the t+25|t+06, t+19|t+12, and t+25|t+18 models are more complex but have greater predictive capacity and, for a 24-hour time horizon, there is a greater number of relevant variables for forecasting that positively contribute to reducing the deviations between the actual and forecasted values.

When analyzing the average hourly improvements for a lead time of t+49|t+06 shown in figure 4.30, significant improvements for the Q05 quantile forecasts were observed between 07h and 19h, resulting in an average improvement of 7.39%. However, the Q50 quantile shows a negative improvement of -0.48%, attributed to periods with notable MAE declining exceeding -10%, particularly between 09h and 11h.

The second-level forecasts with a lead time of t+25|t+06 displayed in figure 4.31, show improvement in MAE compared to the benchmark for almost every hour of the day. This improvement can be attributed to the use of variables with a shorter lead time, which, as mentioned earlier, may have greater predictive capacity, resulting in more accurate second-level forecasts.

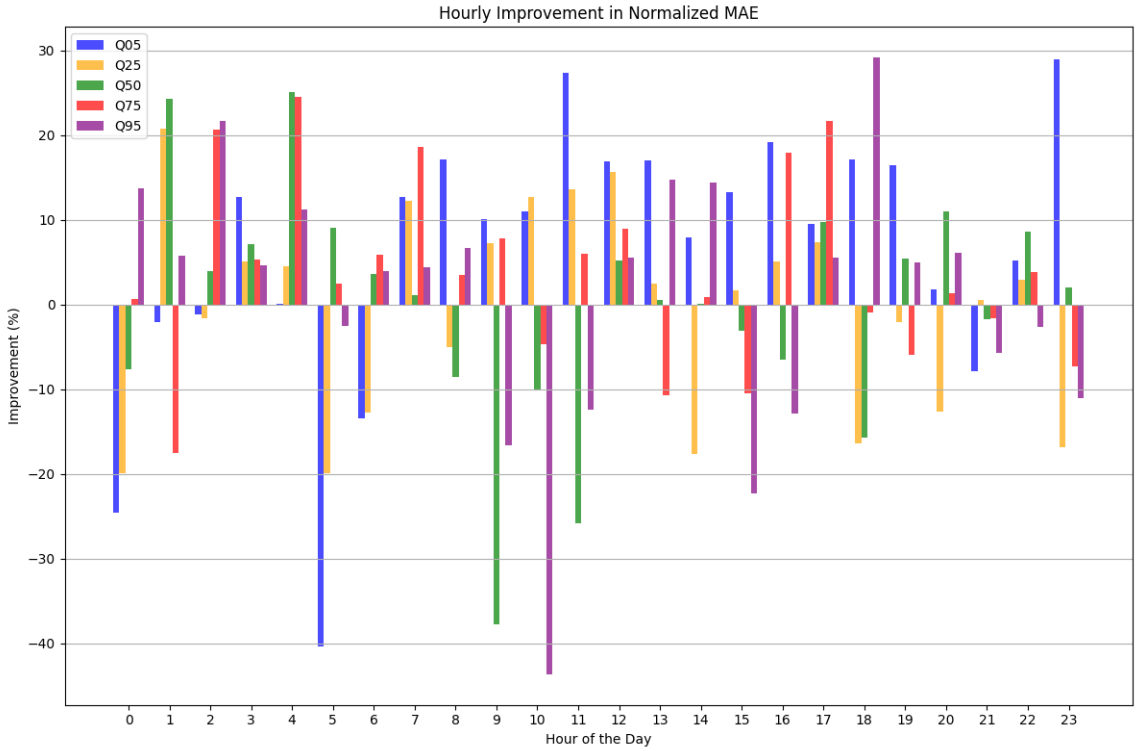

Figure 4.30: Fakken hourly MAE improvements for lead time t+49lt+06

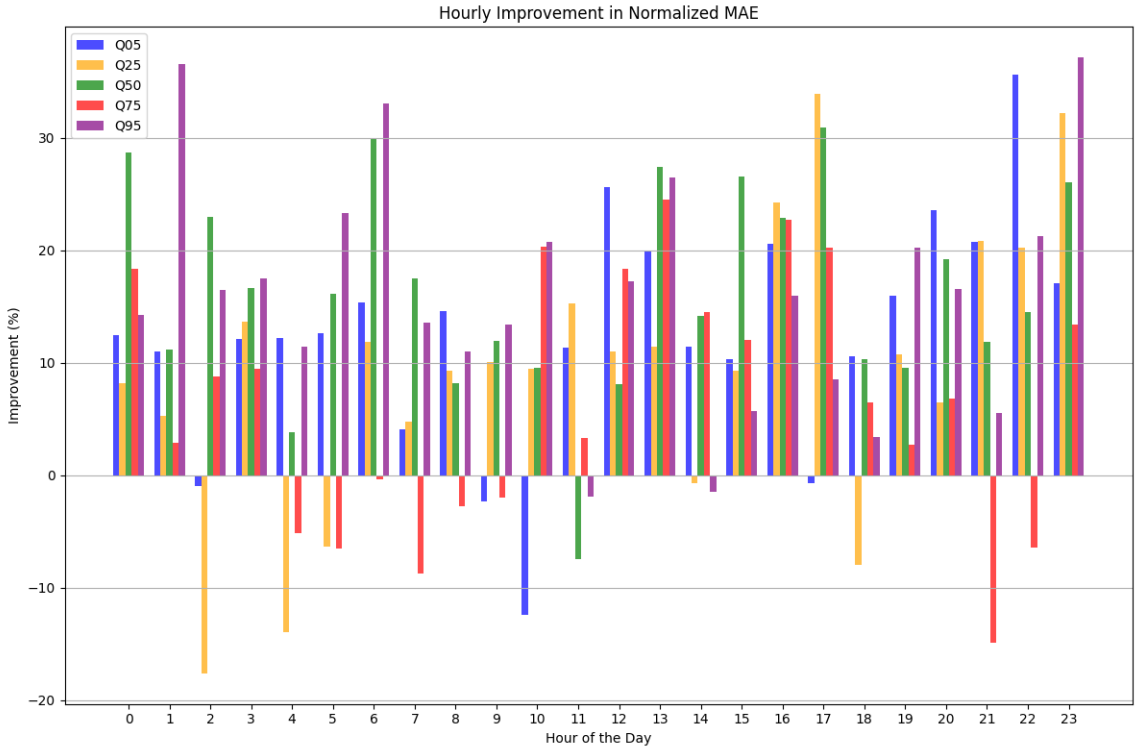

Figure 4.31: Fakken hourly MAE improvements for lead time t+25lt+06

4.4.5 Results for Windfarm Lista

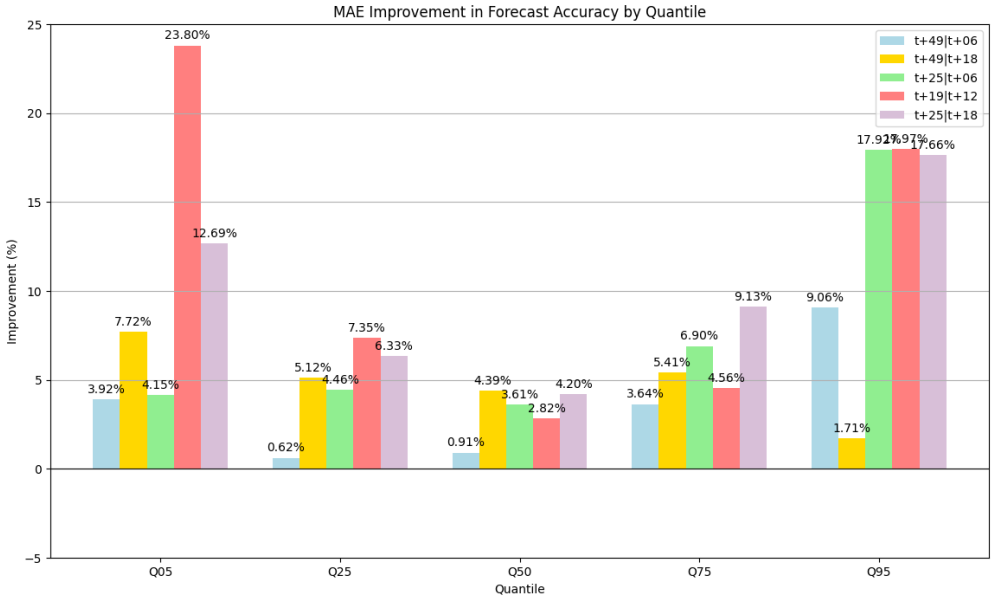

Figure 4.32: MAE improvements for windfarm Lista

Graph 4.32, shows forecasts with lead times t+49|t+06, t+49|t+18, t+25|t+06, t+19|t+18, and t+25|t+18 with improvements over their respective benchmark models of 3.68%, 4.87%, 7.41%, 11.3%, and 10.00% respectively, remaining consistent with the results obtained for the other wind farms.

Based on the results obtained, a similar behavior was observed as in the previous wind farms in which the improvements in MAE compared to the benchmark model were more significant for forecasts with a 24-hour time horizon.

It has been observed that the accuracy of second-level forecasts for extreme quantiles, such as Q05 and Q95, shows more significant improvement compared to their respective benchmarks than when compared to more central quantiles. This behavior could be due to the importance of the features included in the models. Certain features or groups of variables may be more relevant or indicative of extreme wind power values. If these features are highly predictive for extreme quantiles and less so for central ones, the model will naturally perform better on Q05 and Q95. For instance, specific meteorological conditions or patterns may be strongly correlated with extreme wind speeds, and as such, the inclusion of the wind speed can provide more value to predict the extreme quantiles.

Forecasts for the Q05 quantile generally depict lower wind production values. Consequently, a strong correlation is expected between the production values of the Q05 quantile and wind speeds ranging from 0 m/s up to the cut-in speed, which typically stands at 3 m/s.

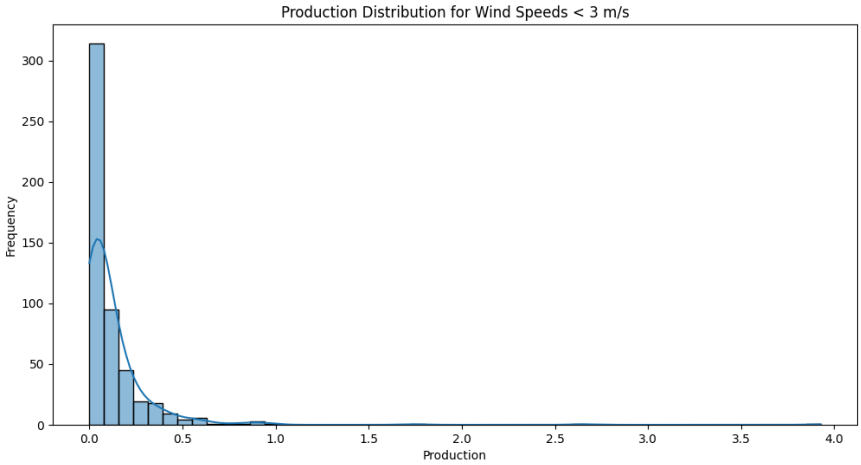

Figure 4.33: Power distribution for wind speeds lower than 3 m/s of Q05 target value for lead time t+19|t+12

Upon analyzing graph 4.33, it is evident that for wind speeds below the cut-in speed, there is minimal variation in production values for the Q05 target quantile at lead time t+19|t+12. This indicates low uncertainty surrounding production for this range of wind speeds, suggesting that the wind speed variable has a strong predictive capacity for low production values, justifying why there was 23.80% MAE improvement over the benchmark.

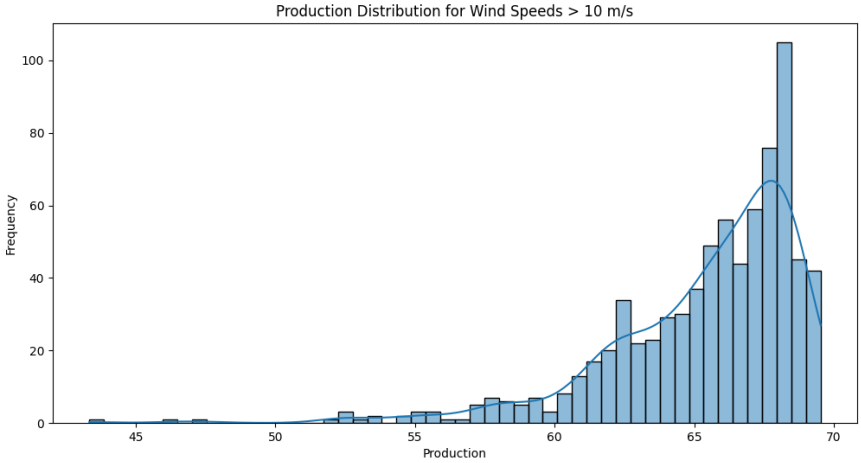

Figure 4.34: Power distribution for wind speeds greater than 10 m/s of Q95 target value for lead time t+19|t+12

The Q95 quantile generally indicates higher wind production values, particularly for wind speeds exceeding 10 m/s and around the rated speed. While there is a greater disparity in wind production values for the Q95 target quantile at lead time t+19|t+12, there is a strong correlation between speed values and predicted productions for Q95, as seen in 4.34, resulting in a 17.97%

improvement in mean absolute error compared to the benchmark model.

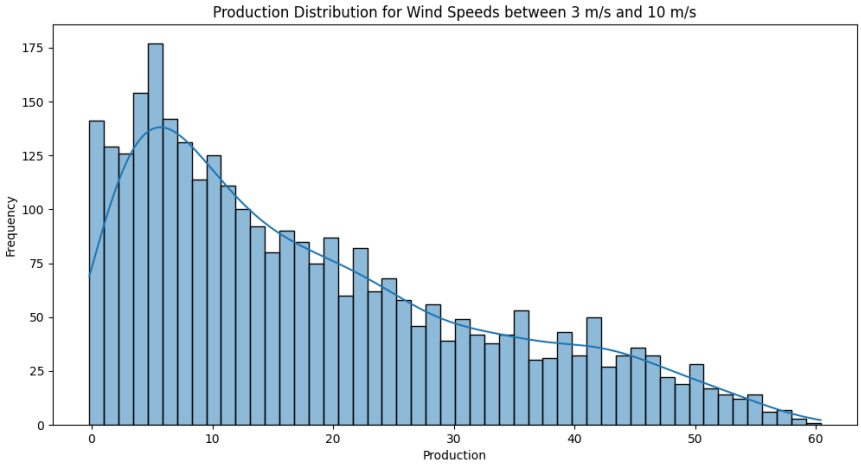

Figure 4.35: Power distribution for wind speeds greater than 3 m/s and lower than 10 m/s of Q50 target value for lead time t+19|t+12

For the quantile Q50 in figure 4.35, the distribution of wind output is concentrated more around the intermediate values, which typically fall between wind speeds of 3 m/s and 10 m/s. Upon analyzing the graph, it is evident that there is a wide dispersion of production values for the target quantile Q50 with a lead time of t+19|t+12, indicating a significant level of uncertainty regarding production values for this speed range.

Although using wind speed in the forecast model for quantile Q50 has positively contributed to reducing the MAE and RMSE, it is less relevant for predicting wind production for the more intermediate percentiles. As such the MAE for the second level forecast of quantile Q50 with lead time t+19|t+12 improved by 2.82% when compared to its respective benchmark model.

It is also worth noting that there is the risk regarding the central quantiles being more prone to overfitting due to the relatively higher volume of data points around the median. This can cause the model to perform worse on unseen data, while extreme quantiles, with less frequent but more distinct patterns, can be modeled more robustly.

When examining the hourly MAE improvements for the forecasts with lead time t+49|t+06 in figure 4.36, extensive periods with substantial improvements above 10% were observed for the Q95 quantile, particularly between 03h00 and 11h00. This resulted in an average improvement of 9.06%, as illustrated in the figure 4.32. In contrast, the Q50 quantile shows several instances without improvement as well as several instances with improvement compared to its respective benchmark, resulting in a modest overall improvement of 0.91%.

For the forecasts with a lead time of t+19|t+12, all quantiles performed well, especially the Q05 and Q95 quantiles, which showed significant improvements above 20% at various times during the day, with average improvements of 23.8% and 17.92%, respectively. Although the Q50 and Q75 quantiles generally performed well throughout the day, there were some instances of notable negative improvements. Specifically, the Q50 quantile showed negative improvements at 17h, 18h,

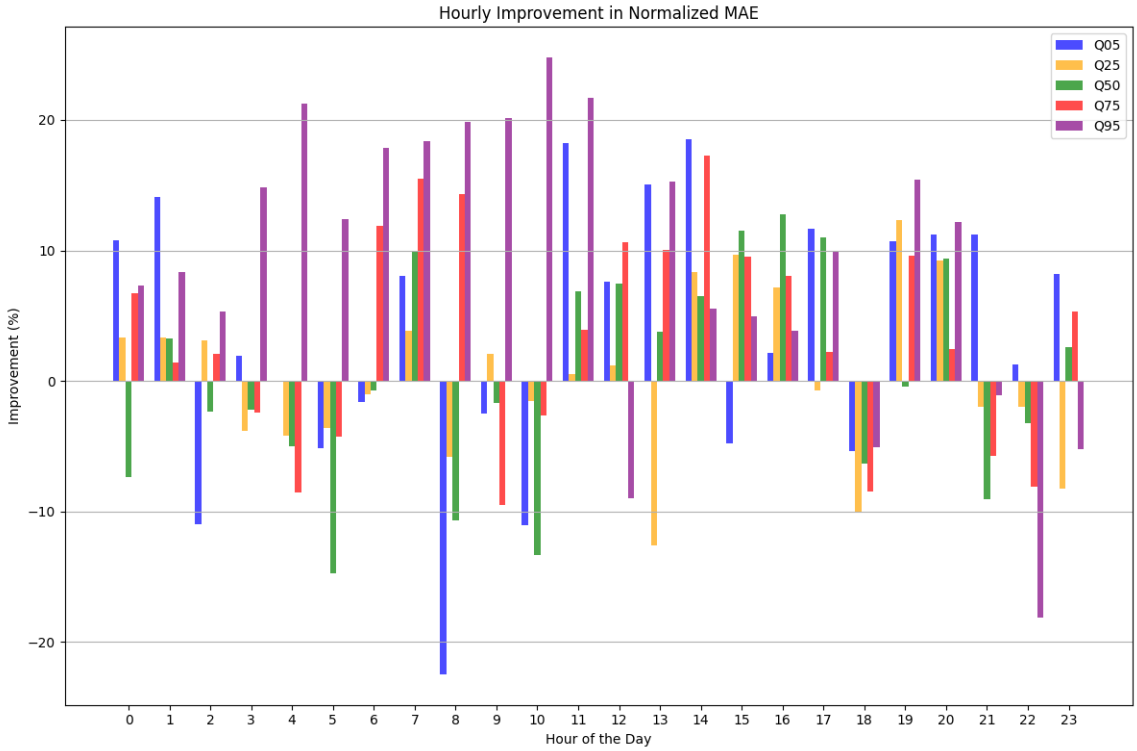

Figure 4.36: Lista hourly MAE improvements for lead time t+49|t+06

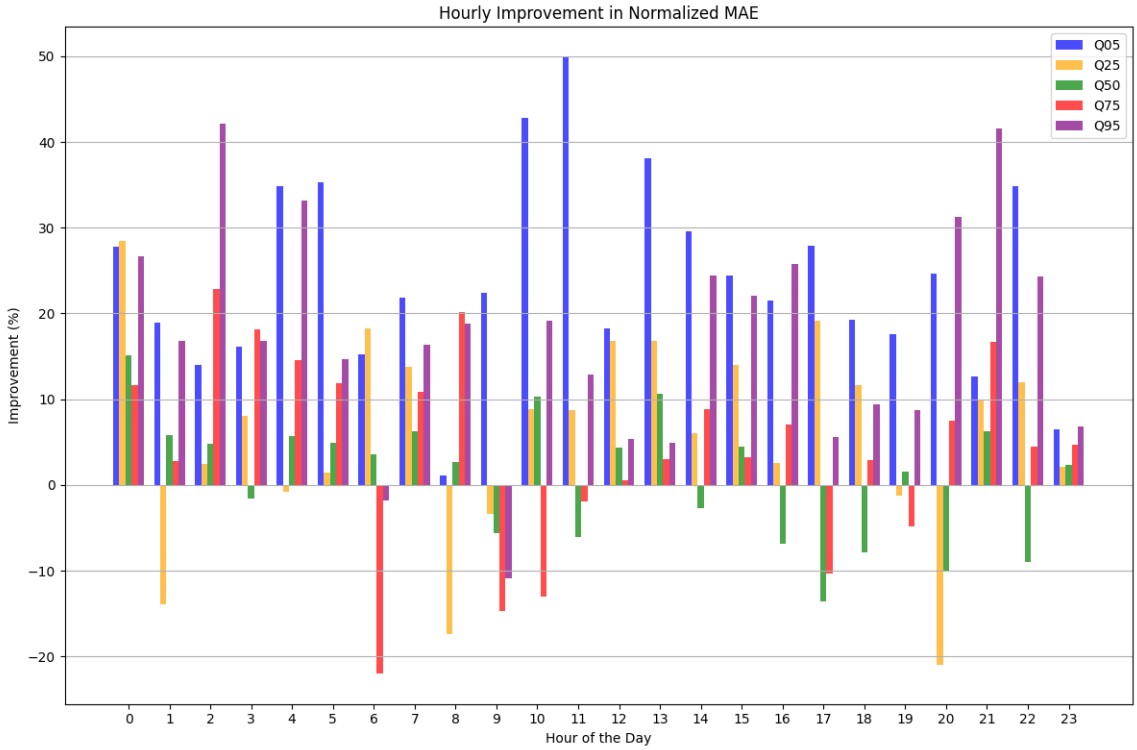

Figure 4.37: Lista hourly MAE improvements for lead time t+19|t+12

20h, and 22h, while the Q75 quantile showed negative improvements at 06h, 09h, 10h, and 17h. This affected the overall average improvement for the day for these quantiles, which was 2.82% and 4.56% respectively.

Chapter 5

Conclusions

The dissertation examined second-level wind energy forecasts with a 48-hour time horizon to predict fluctuations in wind energy forecasts occurring during each NWP forecast update at 00h, 06h, 12h, and 18h for three different wind farms. Second-level prediction models were created to anticipate how wind energy forecasts vary at 06h and 18h using first-level data and NWP generated at 00h with a 48-hour lead time (i.e., t+49|t+06 and t+49|t+18, respectively). For a 24-hour time horizon, second-level forecasts were developed to anticipate wind production fluctuations at 06h and 18h using data generated at 00h (t+25|t+06 and t+25|t+18) and the fluctuations at 12h were forecasted using data generated at 06h (t+19|t+12). The result analysis focused on improvements in Mean Absolute Error (MAE) over benchmark models across different quantiles and lead times

The research findings suggest that shorter lead times (24-hour horizons) result in more accurate forecasts compared to longer lead times (48-hour horizons). Shorter lead times allow for more reliable and accurate meteorological variables, enhancing predictive capabilities. Forecasts with longer lead times, especially when using NWP variables with extended forecast periods, showed less accuracy and reliability due to the propagation of errors.

It was noted that because the performance of each quantile varied, different second-level models had to be developed for each quantile to achieve more accurate forecasting. The gradient boosting regressor identified different relevant features for each quantile, resulting in varying improvements. Additionally, the improvements of the second-level forecasts over the benchmark were more significant when the variability of wind energy production between the different NWP updates was greater. This suggests that second-level forecasts are more valuable to the user when there is greater variability in wind production, as assuming that forecasts remain constant throughout the day is not a suitable prediction for such cases.

Additionally, second-level forecasts for extreme quantiles (e.g., Q05 and Q95) demonstrated more significant improvements over their respective benchmark models compared to second-level forecasts for more central quantiles, such as Q50.

Future Work

In order to improve second-level forecasting, it is recommended to extend the training period to capture the trends of the first-level wind energy forecasts more accurately. This means that the testing period for the first level should also be extended to gather more data for future use for second-level forecasts. Furthermore, it would be beneficial to explore better features that have a stronger correlation with the target values and that can contradict the deviations between forecasted and actual values in order to enhance the accuracy of second-level predictions, particularly for forecasts of central quantiles, which did not show substantial improvements compared to benchmark models. It's also suggested to test different machine learning approaches, such as encoder-decoder or artificial neural networks which the authors in [3] found to have greater improvements when compared to the gradient boosting trees.

Appendix

Below are the complete results of the second-level forecasts for the various wind farms across different lead times.

Bessakerfjellet Second-Level forecasts for t+49|t+06

Table 1: Results of Second-Level forecasts for t+49|t+06

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G1, G3	2.96	2.12	2.60	18.46
Q25	G1, G3, G4, G5	5.01	3.39	3.51	3.42
Q50	G1, G3, G4	5.67	3.92	4.01	2.00
Q75	G1, G3, G4	5.37	3.84	3.87	0.78
Q95	G3, G4, G5	4.67	3.29	3.27	-0.61

Bessakerfjellet Second-Level forecasts for t+49|t+18

Table 2: Results of Second-Level forecasts for t+49|t+18

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G3, G4	6.37	4.19	7.48	20.34
Q25	G1, G3, G4	10.09	6.04	6.61	8.02
Q50	G1, G3, G4	9.82	6.43	6.58	2.27
Q75	G1, G3, G4	8.77	6.06	3.08	0.32
Q95	G3, G4	6.29	4.44	4.50	1.33

Bessakerfjellet Second-Level forecasts for t+25|t+06

Table 3: Results of Second-Level forecasts for t+25|t+06

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G1, G3, G4	4.62	3.00	3.13	4.15
Q25	G1, G3, G4	4.42	2.58	2.82	8.51
Q50	G1, G3, G4	4.79	2.93	3.16	7.28
Q75	G3, G4	4.47	2.92	3.11	6.11
Q95	G1, G3, G4	3.75	2.58	2.92	11.64

Bessakerfjellet Second-Level forecasts for t+19|t+12

Table 4: Results of Second-Level forecasts for t+19|t+12

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G3, G4	2.61	1.88	1.94	3.09
Q25	G1, G3, G4	2.04	1.46	1.70	13.53
Q50	G1, G3	1.72	1.25	1.36	8.09
Q75	G1, G3, G4	1.80	1.30	1.39	6.47
Q95	G1, G2, G3, G4	2.46	1.71	1.83	6.56

Bessakerfjellet Second-Level forecasts for t+25|t+18

Table 5: Results of Second-Level forecasts for t+24|t+06

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G1, G3, G4, G5	6.45	4.11	4.51	9.07
Q25	G1, G3, G4	7.48	4.54	4.88	22.92
Q50	G1, G3, G4	7.89	4.75	4.86	2.26
Q75	G1, G2, G3, G4	7.54	4.80	5.01	4.00
Q95	G1, G3, G4, G5	5.95	4.06	4.59	11.55

Fakken Second-Level forecasts for t+49|t+06

Table 6: Results of Second-Level forecasts for t+49|t+06

Quantile	Variable Groups	RMSE(%)	MAE(%)	$MAE_{Benchmark}$(%)	Improvement(%)
Q05	G1, G3, G4	2.43	1.63	1.76	7.39
Q25	G1, G3, G4	5.42	3.75	3.78	0.79
Q50	G1, G3, G4	5.77	4.20	4.19	-0.48
Q75	G1, G3, G4	5.32	3.85	4.03	4.47
Q95	G1, G3, G4	4.78	3.14	3.21	1.88

Fakken Second-Level forecasts for t+49|t+18

Table 7: Results of Second-Level forecasts for t+49|t+18

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G3, G4	4.16	2.64	2.94	10.20
Q25	G1, G3, G4	9.15	6.21	6.54	5.04
Q50	G1, G3, G4	9.75	6.90	7.03	1.85
Q75	G1, G3, G4	9.70	6.95	7.14	2.66
Q95	G1, G3, G4	7.92	5.49	5.97	8.04

Fakken Second-Level forecasts for t+25|t+06

Table 8: Results of Second-Level forecasts for t+25|t+06

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G2, G3, G4	2.48	1.71	1.95	12.31
Q25	G1, G3, G4	3.68	2.66	2.96	10.14
Q50	G1, G3, G4	4.14	3.07	3.68	16.58
Q75	G1, G2, G3, G4	4.41	3.24	3.47	6.63
Q95	G1, G3, G4	3.76	2.61	3.14	16.88

Fakken Second-Level forecasts for t+19|t+12

Table 9: Results of Second-Level forecasts for t+19|t+12

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G2, G3, G4	2.83	1.96	2.37	17.30
Q25	G1, G2, G3, G4	4.33	3.15	3.44	10.14
Q50	G1, G3, G4	5.22	3.81	4.03	5.46
Q75	G1, G2, G3, G4	5.92	4.26	4.72	9.75
Q95	G1, G2, G3, G4	5.99	4.09	4.34	5.76

Fakken Second-Level forecasts for t+25|t+18

Table 10: Results of Second-Level forecasts for t+25|t+18

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G2, G3, G4	4.48	2.82	3.71	23.99
Q25	G1, G2, G3, G4, G5	6.26	4.43	4.85	8.66
Q50	G1, G3, G4, G5	7.33	5.37	6.09	11.97
Q75	G1, G2, G3, G4	8.07	5.82	6.81	14.54
Q95	G1, G2, G3, G4, G5	7.10	4.87	5.88	17.18

Lista Second-Level forecasts for t+49|t+06

Table 11: Results of Second-Level forecasts for t+49|t+06

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G3, G4	4.83	3.019	3.33	3.92
Q25	G1, G3, G4	4.86	3.23	3.25	0.62
Q50	G1, G3, G4	4.85	3.28	3.31	0.91
Q75	G1, G3, G4	5.15	3.44	3.57	3.64
Q95	G1, G3, G4	4.82	3.01	3.31	9.06

Lista Second-Level forecasts for t+49|t+18

Table 12: Results of Second-Level forecasts for t+49|t+18

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G2, G3, G4	8.24	5.14	5.57	7.72
Q25	G1, G2, G3, G4	9.79	6.49	6.85	5.12
Q50	G1, G3, G4	8.98	6.10	6.38	4.39
Q75	G1, G2, G3, G4	9.87	6.30	6.66	5.41
Q95	G1, G3, G4, G5	8.84	5.18	5.27	1.71

Lista Second-Level forecasts for t+25|t+06

Table 13: Results of Second-Level forecasts for t+25|t+06

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G3,	5.02	3.23	3.38	4.15
Q25	G1, G3, G4	4.73	3.00	3.14	4.46
Q50	G1, G3, G4	4.73	2.94	3.05	3.61
Q75	G1, G3, G4	4.57	2.97	3.20	6.90
Q95	G1, G3, G4	4.25	2.84	3.46	17.92

Lista Second-Level forecasts for t+19|t+12

Table 14: Results of Second-Level forecasts for t+19|t+12

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G3, G4	4.07	2.53	3.32	23.80
Q25	G1, G3, G4, G5	3.95	2.52	2.72	7.35
Q50	G1, G3, G4, G6	4.32	2.76	2.84	2.82
Q75	G1, G3, G4, G5	4.74	2.93	3.07	4.56
Q95	G1, G3, G4	3.88	2.51	3.06	17.97

Lista Second-Level forecasts for t+25|t+18

Table 15: Results of Second-Level forecasts for t+25|t+18

Quantile	**Variable Groups**	**RMSE(%)**	**MAE(%)**	$MAE_{Benchmark}$**(%)**	**Improvement(%)**
Q05	G1, G3, G4	6.75	4.47	5.12	12.69
Q25	G1, G3, G4, G6	7.12	4.59	4.90	6.33
Q50	G1, G3, G4, G5, G6	6.99	4.57	4.78	4.20
Q75	G1, G3, G4, G5, G6	6.88	4.48	4.93	9.13
Q95	G1, G3, G4	6.37	4.15	5.04	17.66

References

[1] Corinna Möhrlen, R Bessa, Gregor Giebel, and J Jørgensen. Uncertainty forecasting practices for the next generation power system. Wind Integration Workshop, Berlin (DE), 2017.

[2] Jan Dobschinski, Ricardo Bessa, Pengwei Du, Kenneth Geisler, Sue Ellen Haupt, Matthias Lange, Corinna Möhrlen, Dora Nakafuji, and Miguel De La Torre Rodriguez. Uncertainty forecasting in a nutshell: Prediction models designed to prevent significant errors. *IEEE Power and Energy Magazine*, 15(6):40–49, 2017.

[3] Ricardo J. Bessa, Ferinar Moaidi, João Viana, and José R. Andrade. Uncertainty-aware procurement of flexibilities for electrical grid operational planning. *IEEE Transactions on Sustainable Energy*, pages 1–14, 2023. doi:10.1109/TSTE.2023.3305865.

[4] Allan H. Murphy and Qian Ye. Optimal decision making and the value of information in a time-dependent version of the cost-loss ratio situation. *Monthly Weather Review*, 118(4):939 – 949, 1990. URL: https://journals.ametsoc.org/view/journals/mwre/118/4/1520-0493_1990_118_0939_odmatv_2_0_co_2.xml, doi:https://doi.org/10.1175/1520-0493(1990)118<0939:ODMATV>2.0.CO;2.

[5] Ricardo J. Bessa, Corinna Möhrlen, Vanessa Fundel, Malte Siefert, Jethro Browell, Sebastian Haglund El Gaidi, Bri-Mathias Hodge, Umit Cali, and George Kariniotakis. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, 10(9), 2017. URL: https://www.mdpi.com/1996-1073/10/9/1402, doi:10.3390/en10091402.

[6] Robert A. Stine. Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, 80(392):1026–1031, 1985. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1985.10478220, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1985.10478220, doi:10.1080/01621459.1985.10478220.

[7] C Monteiro, R Bessa, V Miranda, A Botterud, J Wang, G Conzelmann, Decision, Information Sciences, and INESC Porto. Wind power forecasting : state-of-the-art 2009. URL: https://www.osti.gov/biblio/968212, doi:10.2172/968212.

[8] Javier Antonanzas, Natalia Osorio, Rodrigo Escobar, Ruben Urraca, Francisco J Martinez-de Pison, and Fernando Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar energy*, 136:78–111, 2016.

[9] Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81:912–928, 2018.

[10] Yao Zhang, Jianxue Wang, and Xifan Wang. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32:255–270, 2014. URL: https://www.sciencedirect.com/science/article/pii/S1364032114000446, doi:https://doi.org/10.1016/j.rser.2014.01.033.

[11] Hans Bludszuweit, José Antonio Domínguez-Navarro, and Andrés Llombart. Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems*, 23(3):983 – 991, 2008. Cited by: 695. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-49249100812&doi=10.1109%2fTPWRS.2008.922526&partnerID=40&md5=d6752a73629059ccb4dde5c373a15223, doi:10.1109/TPWRS.2008.922526.

[12] Bri-Mathias Hodge and Michael Milligan. Wind power forecasting error distributions over multiple timescales. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–8, 2011. doi:10.1109/PES.2011.6039388.

[13] J Zack and Corinna Möhrlen. Iea wind task 36: Practical application examples from the recommended practices for forecast solution selection. In *Proceedings of the 20th International Workshop on Large-Scale Integration of Wind Power into Power Systems (virtual). Energynautics GmbH, Darmstadt Germany*, 2020.

[14] Kate Doubleday, Vanessa Van Scyoc Hernandez, and Bri-Mathias Hodge. Benchmark probabilistic solar forecasts: Characteristics and recommendations. *Solar Energy*, 206:52–67, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0038092X20305429, doi:https://doi.org/10.1016/j.solener.2020.05.051.

[15] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. URL: https://doi.org/10.1198/016214506000001437, arXiv:https://doi.org/10.1198/016214506000001437, doi:10.1198/016214506000001437.

[16] D.W. van der Meer, J. Widén, and J. Munkhammar. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews*, 81:1484–1512, 2018. URL: https://www.sciencedirect.com/science/article/pii/S1364032117308523, doi:https://doi.org/10.1016/j.rser.2017.05.212.

[17] Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(Volume 1, 2014):125–151, 2014. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-062713-085831, doi:https://doi.org/10.1146/annurev-statistics-062713-085831.

[18] Tao Hong and Shu Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016. URL: https://www.sciencedirect.com/science/article/pii/S0169207015001508, doi:https://doi.org/10.1016/j.ijforecast.2015.11.011.

[19] Ruosha Li and Limin Peng. Quantile Regression for Left-Truncated Semicompeting Risks Data. *Biometrics*, 67(3):701–710, 12 2010. URL: https://doi.org/10.1111/j.1541-0420.2010.01521.x, arXiv:https:

//academic.oup.com/biometrics/article-pdf/67/3/701/53151705/biometrics_67_3_701.pdf, doi:10.1111/j.1541-0420.2010.01521.x.

[20] Sven Kolkmann, Lars Ostmeier, and Christoph Weber. Modeling multivariate intraday forecast update processes for wind power. pages 1–28, June 6, 2022. URL: https://ssrn.com/abstract=413250, doi:http://dx.doi.org/10.2139/ssrn.4132502.

[21] Pierre Pinson, Henrik Madsen, Henrik Aa. Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.284, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.284, doi:https://doi.org/10.1002/we.284.

[22] Craig Wanke and Daniel Greenbaum. *Sequential Congestion Management with Weather Forecast Uncertainty*. URL: https://arc.aiaa.org/doi/abs/10.2514/6.2008-6327, arXiv:https://arc.aiaa.org/doi/pdf/10.2514/6.2008-6327, doi:10.2514/6.2008-6327.

[23] Stephen Jewson, Sebastian Scher, and Gabriele Messori. Decide now or wait for the next forecast? testing a decision framework using real forecasts and observations. *Monthly Weather Review*, 149(6):1637–1650, 2021.

[24] Allan H Murphy and Qian Ye. Optimal decision making and the value of information in a time-dependent version of the cost-loss ratio situation. *Monthly Weather Review*, 118(4):939–949, 1990.

[25] Stefanos Fafalios, Pavlos Charonyktakis, and Ioannis Tsamardinos. Gradient boosting trees. *Gnosis Data Analysis PC*, pages 1–3, 2020.

[26] Fabian Pedregosa. Scikit-learn: Machine learning in python fabian. *Journal of machine learning research*, 12:2825, 2011.

[27] Stephen Jewson, Sebastian Scher, and Gabriele Messori. Communicating properties of changes in lagged weather forecasts. *Weather and Forecasting*, 37(1):125 – 142, 2022. URL: https://journals.ametsoc.org/view/journals/wefo/37/1/WAF-D-21-0086.1.xml, doi:https://doi.org/10.1175/WAF-D-21-0086.1.

[28] Mark Landry, Thomas P. Erlinger, David Patschke, and Craig Varrichio. Probabilistic gradient boosting machines for gefcom2014 wind forecasting. *International Journal of Forecasting*, 32(3):1061–1066, 2016. URL: https://www.sciencedirect.com/science/article/pii/S0169207016000145, doi:https://doi.org/10.1016/j.ijforecast.2016.02.002.

[29] José R. Andrade and Ricardo J. Bessa. Improving renewable energy forecasting with a grid of numerical weather predictions. *IEEE Transactions on Sustainable Energy*, 8(4):1571–1580, 2017. doi:10.1109/TSTE.2017.2694340.

[30] Roger Schaer, Henning Müller, and Adrien Depeursinge. Optimized distributed hyperparameter search and simulation for lung texture classification in ct using hadoop. *Journal of Imaging*, 2(2), 2016. URL: https://www.mdpi.com/2313-433X/2/2/19, doi:10.3390/jimaging2020019.

[31] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[32] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, and Si-Hao Deng. Hyperparameter optimization for machine learning models based on bayesian optimizationb. *Journal of Electronic Science and Technology*, 17(1):26–40, 2019. URL: https://www.sciencedirect.com/science/article/pii/S1674862X19300047, doi:https://doi.org/10.11989/JEST.1674-862X.80904120.

[33] European centre for medium-range weather forecasts. URL: https://www.ecmwf.int/.