Towards Probabilistic Forecasting of Renewable Energy Sources using a Hierarchical Framework

Michael Vitali, Georgios Mitrentsis, and Lena Peter
Hitachi Energy Research
Mannheim, Germany

***Abstract*—The integration of renewable energy sources, such as photovoltaic (PV) and wind power, into the power grid presents challenges due to their intermittent nature, which is influenced by weather conditions. Accurate short-term forecasting of renewable energy generation is crucial for maintaining grid stability, optimizing energy trading, and minimizing imbalance costs. While significant progress has been made in probabilistic forecasting to account for uncertainty, the combination of probabilistic and hierarchical forecasting techniques remains underexplored. This paper proposes a new probabilistic forecasting model for virtual power plants (VPPs) that integrates multiple renewable energy sources, including PV and wind, within a hierarchical framework. The model leverages Probabilistic Gradient Boosting Machines (PGBM) with a tailored loss function to account for different aggregation levels. We evaluate the model using a real-world dataset from the IEEE Hybrid Energy Forecasting and Trading Competition, demonstrating that our approach improves forecast accuracy and provides more reliable uncertainty estimates.**

***Index Terms*—Hierarchical forecasting, wind power, solar power, probabilistic forecasting, renewable energy**

I. Introduction

Integrating renewable energy sources like photovoltaic (PV) and wind power into the power grid has brought new challenges to grid operators and energy traders. Due to the intermittent nature of renewable generation, which is highly influenced by weather conditions, precise short-term forecasting of PV and wind power is essential to maintain grid stability, optimize energy trading bids, and reduce imbalance costs. In recent years, significant efforts have been dedicated to advancing the accuracy of renewable energy forecasts and incorporating uncertainty into these predictions [1].

Probabilistic forecasts provide more comprehensive insights into the future power output of PV and wind farms, enabling system operators and traders to make better-informed decisions, such as minimizing financial risks and maximizing revenue [2], [3].

At the same time, hierarchical forecasting for renewable energy can be employed to increase the accuracy of individual forecasting models [4]. This involves generating forecasts at multiple levels of aggregation, typically across component, park, area, and country levels. This hierarchical approach enables predictions to be consistent across different levels, i.e., the aggregated forecasts at higher levels (e.g., total energy output of a wind park) should match the sum or expected contribution of the forecasts from the lower levels (e.g., energy output of individual wind turbines).

While there exists several approaches in probabilistic renewable energy forecasting as well as in hierarchical forecasting, the combination of them is not so widely studied in the literature. Nevertheless, a few notable research works have provided some inspiring contributions in this area. For instance, [5] introduced a probabilistic forecast reconciliation method that ensures coherence in wind power and electric load forecasts across different hierarchy levels. Reference [6] proposed a hierarchical probabilistic model to improve wind power forecasting accuracy by using turbine-level data, which contributes to more consistent forecasts across various aggregation levels. Similarly, [7] focused on solar energy by proposing a method that reconciles initial base forecasts to match higher-level aggregated forecasts within a geographical hierarchy. In this context, although some of those works combine hierarchical and probabilistic approaches, they focus on one renewable energy technology, either PV or wind, and they do not consider the combined PV and wind generation aggregation in the form of a virtual power plant (VPP).

To this end, we propose a new probabilistic forecasting model for VPPs using a hierarchical framework. Importantly, the VPP can include different types of renewable energy sources regardless of their technology, e.g., PV and wind power. To do so, we tailored the loss function of the Probabilistic Gradient Boosting Machines (PGBM) algorithm, introduced in [8], in order to account for the different aggregation levels.

To evaluate the efficacy of our proposed hierarchical probabilistic model, we employed a real-world dataset from the recent IEEE Hybrid Energy Forecasting and Trading Competition.

II. Methodology Proposed

In this paper the problem of renewable energy forecasting is solved using Probabilistic Gradient Boosting Machines to solve a hierarchical forecasting problem, where at the bottom level there are different generation plants. The proposed approach is shown in Fig. 1. In the following sections, all the blocks will be explained in more details.

A. Probabilistic Forecasting

This study focuses on the challenge of probabilistic forecasting, which falls under the broader domain of probabilistic

979-8-3315-1278-1/25/$31.00

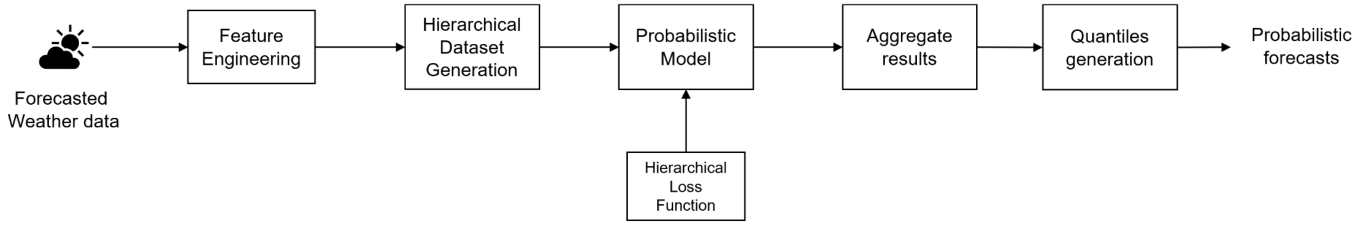


Fig. 1: Proposed Methodology

regression. The core objective of probabilistic regression is to derive a conditional probability distribution $P(y|\boldsymbol{x})$ for a scalar target variable y, given a set of input features x. In the specific context of probabilistic forecasting, the feature set $\boldsymbol{x}$ typically comprising both historical values of the target variable (referred as lagged variables) and additional relevant factors (known as covariates) [9].

Our research aims to develop a model $f(\boldsymbol{x})$ capable of estimating two key parameters of the target distribution: its mean μ and variance σ^2. By accurately estimating these parameters, we can generate predictions $\hat{y}$ by sampling from a specified distribution D post-training. This process can be formally expressed as:

$$f(x) \rightarrow (\mu_{\hat{y}}, \sigma^2_{\hat{y}}) \tag{1}$$

$$\hat{y} \sim D(\mu_{\hat{y}}, \sigma^2_{\hat{y}}) \tag{2}$$

This approach allows us to capture not only the expected value of the target variable but also the uncertainty associated with our predictions, which is crucial in many real-world applications of forecasting.

B. Hierarchical Forecasting

Hierarchical Forecasting involves a collection of time series that are interconnected through a set of linear aggregation constraints, forming a hierarchical structure. To formalize this concept, we introduce the following notation: Let $y_{i,t} \in \mathbb{R}$ represent the value of the *i*-th series at time t, corresponding to a feature vector $\boldsymbol{x}_{i,t} \in \mathbb{R}^p$, with p features and t ranging from 1 to T. We denote the collective values of all n series at time t as $\boldsymbol{y_t} \in \mathbb{R}^n$, with the associated feature matrix $\boldsymbol{x_t} = [\boldsymbol{x}_{1,t}, ..., \boldsymbol{x}_{n,t}]^T \in \mathbb{R}^{n \times p}$. Within this framework, we distinguish between bottom-level series and aggregated series. Let $\boldsymbol{y}_{t,b} \in \mathbb{R}^m$ represent the m bottom-level series at time t, and $\boldsymbol{y}_{t,a} \in \mathbb{R}^r$ denote the r aggregated series at various levels, such that $n = m + r$. The relationship between these series is defined by an aggregation matrix $\boldsymbol{S} \in \{0,1\}^{n \times m}$. The hierarchical structure can be expressed mathematically as:

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{y}_{t,b} \iff \begin{bmatrix} \boldsymbol{y}_{t,a} \\ \boldsymbol{y}_{t,b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_a \\ \boldsymbol{I}_m \end{bmatrix} \boldsymbol{y}_{t,b}, \ \forall t \in \{1, ..., T\} \tag{3}$$

In this equation, $\boldsymbol{S}_a \in \{0,1\}^{r \times m}$ is a matrix that indicates how to aggregate the bottom-level series to form the higher-level series, and $\boldsymbol{I}_m$ represents an m-dimensional identity matrix. This formulation encapsulates the hierarchical relationships between different levels of the time series structure.

C. Probabilistic Gradient Boosting Machines

The Probabilistic Gradient Boosting Machine (PGBM) extends the traditional Gradient Boosting Machine framework to incorporate probabilistic predictions. This enhancement allows PGBM to provide full predictive distributions rather than just point estimates, making it particularly valuable for uncertainty quantification and probabilistic forecasting tasks.

At its core, PGBM utilizes an ensemble of decision trees, where each tree attempts to correct the errors of the preciding ones, but with the goal of modeling the entire predictive distribution rather than just the mean. The key contribution lies in how these trees are constructed and updated. Instead of focusing solely on minimizing prediction error, PGBM employs a probabilistic loss function and incorporates distributional assumptions about the target variable.

A distinctive feature of PGBM is its use of stochastic leaf weights. For each leaf in every tree, the model learns both a mean and a variance. This approach enables PGBM to estimate a distribution for each data point, rather than a single value. The underlying assumption is that the gradient and Hessian of the loss function can be treated as random variables with finite mean and variance. During training, PGBM approximates these statistical properties for each instance set within the trees. By modeling the expectation and variance of the leaf weights and using a specific update equation, the model estimates the overall distribution of the target variable. Once trained, PGBM can generate probabilistic predictions for new samples by sampling from a distribution parameterized by the learned quantities.

D. Model Training

The model employs a decision tree-based approach and is trained exclusively on weather-related data to predict aggregate energy production for the day ahead. This setup simulates a scenario where real-time power production data are unavailable. To generate forecasts at time t, the model utilizes weather forecasts along with their lagged features, represented as $\boldsymbol{x_t} = [x_{1,t-1}, \ldots, x_{1,t-l}, \ldots, x_{k,t-1}, \ldots, x_{k,t-l}]$, where k denotes the total number of weather variables and l represents the total lags employed, i.e., how many time steps in the past we look at to make a prediction. The aggregated power production is forecasted in the form of quantiles, ranging from 10% to 90% with 10% steps. The model is trained using a loss function that considers all the different levels in the hierarchy:

$$L_t = \boldsymbol{w}^T(\boldsymbol{y_t} - \hat{\boldsymbol{y}}_{\boldsymbol{t}})^2 = \boldsymbol{w}^T(\boldsymbol{S}\boldsymbol{y}_{t,b} - \boldsymbol{S}\hat{\boldsymbol{y}}_{t,b})^2 \quad (4)$$

where $\boldsymbol{w}$ is the weights' vector, $\boldsymbol{S}$ the aggregation matrix, $\boldsymbol{y}_{t,b}$ the real production for the plants, and $\hat{\boldsymbol{y}}_{t,b}$ the predicted production. In case, for instance, that two levels are considered, the loss function is defined as:

$$L_t = w_1(\boldsymbol{S}_a\boldsymbol{y}_{t,b} - \boldsymbol{S}_a\hat{\boldsymbol{y}}_{t,b})^2 + \boldsymbol{w_2}(\boldsymbol{I}_m\boldsymbol{y}_{t,b} - \boldsymbol{I}_m\hat{\boldsymbol{y}}_{t,b})^2 \quad (5)$$

In this case, the gradient and hessian are calculated as follows:

$$\frac{\partial L_t}{\partial \hat{\boldsymbol{y}}_{\boldsymbol{t,b}}} = -2w_1\boldsymbol{S}_{\boldsymbol{a}}^T(\boldsymbol{S}_{\boldsymbol{a}}\boldsymbol{y}_{t,b} - \boldsymbol{S}_{\boldsymbol{a}}\hat{\boldsymbol{y}}_{t,b}) - 2\boldsymbol{w_2}\boldsymbol{I}_{\boldsymbol{m}}^T(\boldsymbol{I}_{\boldsymbol{m}}\boldsymbol{y}_{t,b} - \boldsymbol{I}_{\boldsymbol{m}}\hat{\boldsymbol{y}}_{t,b}) \quad (6)$$

$$\frac{\partial^2 L_t}{\partial \hat{\boldsymbol{y}}_{\boldsymbol{t,b}}^2} = 2w_1\boldsymbol{S}_{\boldsymbol{a}}^T\boldsymbol{S}_{\boldsymbol{a}} + 2\boldsymbol{w_2}\boldsymbol{I}_{\boldsymbol{m}}^T\boldsymbol{I}_{\boldsymbol{m}} \quad (7)$$

Assuming that there are two plants in the bottom level, $\boldsymbol{S}_{\boldsymbol{a}} = [1, 1]$ will be the aggregation matrix, and the weights could be defined as: $w_1 = 0.5$ for the aggregate level and $\boldsymbol{w_2} = [0.25, 0.25]$ for the two power plants at the bottom level, respectively.

The model is trained on a dataset that accounts for hierarchical levels, differing significantly from typical time series datasets. To represent these hierarchical levels along with the various features, e.g., weather data, the different time series are stacked vertically and the different features are placed in different columns. Specifically, each row corresponds to a specific plant b and time step t, while each column represents the different lagged weather data for that time step. In this representation, each power plant has its own set of columns for features that are not shared with other plants, ensuring a clear separation of plant-specific data. This representation format ensures that data for different plants are concatenated one below the other, while columns for plant-specific features that are irrelevant to a particular plant are filled with zeros.

Table I provides the example of two levels (aggregate, bottom) and for two power plants at the bottom level:

TABLE I: Example of a hierarchical dataset representation considering two different power plants at the bottom level.

Index	$\boldsymbol{y}_{t,b}$	$\boldsymbol{x}^{p1}_{1,t-1}$	...	$\boldsymbol{x}^{p1}_{1,t-l}$	...	$\boldsymbol{x}^{p1}_{k,t-l}$	$\boldsymbol{x}^{p2}_{1,t-1}$	...	$\boldsymbol{x}^{p2}_{k,t-l}$
$\boldsymbol{p}_{1,1}$	*	*	...	*	...	*	0	...	0
$\boldsymbol{p}_{1,2}$	*	*	...	*	...	*	0	...	0
...	...	...	...	...	...	...	...	...	...
$\boldsymbol{p}_{2,1}$	*	0	...	0	...	0	*	...	*
$\boldsymbol{p}_{2,2}$	*	0	...	0	...	0	*	...	*
...	...	...	...	...	...	...	...	...	...

As illustrated in Table I, the dataset is organized such that the data for each power plant p_1 and p_2 are stacked vertically. For features not relevant to a particular plant, the corresponding columns are filled with zeros (e.g., $\boldsymbol{x}^{p_2}$ for p_1 rows and vice versa). The actual values for each plant are denoted by $*$, and these values populate the columns specific to that plant.

E. Forecast Aggregation and Quantiles generation

The probabilistic model implemented outputs forecasts drawn from the learned Gaussian distribution. The forecasts are related to hierarchy's the bottom-level, therefore they must be aggregated considering the defined aggregation matrix. In our specific case, the aggregate level considered is the highest one. Moreover, the model does not output quantiles directly, but it gives the possibility to draw d samples from the distribution. Hence, in this specific case $d = 1000$ is used, and then the quantiles are generated using these d samples.

III. Input Data

The data used for the experiments are historical energy data of the past three years related the Horsea 1 wind park and the aggregated PV enery the East England region (PES Region 10). The East England region has a total installed capacity of 2.4GW, while the Hornsea 1 a capacity of 1.2GW.
Furthermore, historical weather forecasts are used, taken from the German Weather Agency [10]. The data are given in the netCDF format, including variables such as temperature, humidity, pressure, wind speed and direction. The available data set spans the period from September 2020 to January 2024 with an half-hour resolution. The period between 2020 and 2022 is used as training set, while the remaining is used as test set.

Different data analyses have been performed on the data. Starting from the PV historical data, the analyses highlighted a data distribution being right-skewed (see Fig. 2a), indicating more instances of lower PV energy. This is expected due to the limit imposed by the solar radiation and daily hours. Furthermore, there are seasonal trends.

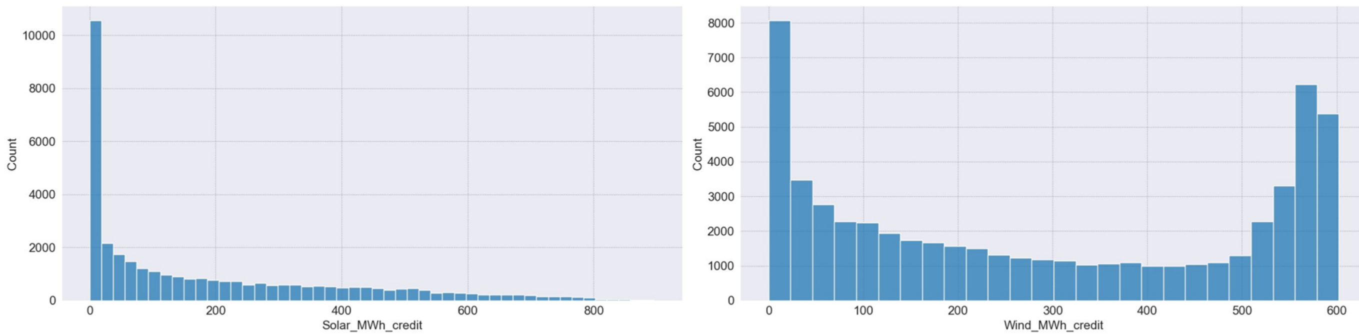
For what regards the wind energy, the distribution indicates a higher frequency of lower energy outputs (see Fig. 2b), a uniform characteristic across median power production, and higher frequency towards higher values. This pattern suggests that while low to moderate wind speeds are common, leading to lower energy production, there are significant periods of high wind speeds resulting in higher energy outputs.

The first step to create the training set to create the model was to merge the energy and weather data into a common set, which contains the weather forecasts of the day before. Then, the missing values have been linearly interpolated. As a next step, the information about time of the day and month is extracted by the timestamp and is stored in a separate column. Since those two time variables are cyclic, they are mapped onto a unit circle and two new features are created out of each one of them. In particular, the sine and cosine of time of the day and month of the year are estimated as:

$$month\ sin = \sin(\frac{2\pi \times month}{12}) \quad (8)$$

$$month\ cos = \cos(\frac{2\pi \times month}{12}) \quad (9)$$

The same is also done for the time of the day.



(a) PV data distribution

(b) Wind data distribution

Fig. 2: Energy data distribution

IV. Results

This section details the experiments and respective results. We examine the performances for day-ahead forecasting of wind and PV production combined, considering probabilistic forecasting with half-hour resolution.

A. Evaluation metrics

For the evaluation, the average pinball loss have been employed, using the predicted quantiles and the true aggregate power generation. For a forecasted quantile $\hat{q}_a$ and the real observation y, the pinball loss is given by

$$L(y, \hat{q}_a) = \begin{cases} (y - \hat{q}_a)a & if\ y \geq \hat{q}_a \\ (\hat{q}_a - y)(1 - a) & if\ y < \hat{q}_a \end{cases} \tag{10}$$

where $\alpha \in (0, 1)$ is the quantile level.

B. Model results

In this section, the model results are shown. The hierarchical model is compared with a Quantile Regression (QR) model [11], a well-established approach in probabilistic forecasting. QR provides a robust benchmark for the proposed method by enabling the evaluation of forecasting performance across different quantiles. Additionally, it allows a detailed assessment of the model's ability to capture uncertainty and tail behavior in the data distribution. Both models are built using the same features. The models have been tested on the 2023 data for each different season. However, for the winter season only data until January 2024 were available.

In Fig. 3, the results for both models in the case of point predictions are shown. The performances are calculated using the 50% quantile as prediction. It can be observed that the results for the hierarchical model are better across all seasons. Fig. 4 shows the probabilistic results for both models for each season. As we can see, the hierarchical model has lower pinball loss during each season. Moreover, Fig. 5 shows the daily average pinball loss for both models. In this case, we can notice that both models have peaks in the pinball loss, possibly caused by bad weather forecasts. However, QR during these peaks has better performance due to wider quantiles. This result can be explained by how PGBMs are trained, minimizing the point prediction loss during training.

To better understand the model's performance, the Fig. 6-9 show the predicted quantiles plotted against the true value for different days during different weather conditions. The red line indicates the true values, whereas the blue lines indicate the different predicted quantiles. In Fig. 6-7, the true values are within the range 10%-90% predicted quantiles. The pinball losses for these days are quite low, especially between 9:00-15:00. The wind and sun profile vary across the days presented indicating the ability of the model to show good performance across different weather conditions.

However, in Fig. 8 and Fig. 9, days for higher pinball loss are shown. As can be seen from the figures, the true values were not within the range of the 10%-90% quantiles, hence leading to higher pinball losses in the range of 60-65. As the model is highly dependent on weather forecast for the day ahead, the accuracy of weather forecasts plays a crucial role. One of the potential reasons for the larger errors could be that the weather forecasts were not accurate or consistent. To summarize, the hierarchical model has better results during each season considered. However, the pinball loss during some days is quite high, mainly caused by an over- or underestimation of the power generation, that can be mainly attributed to bad weather forecast.

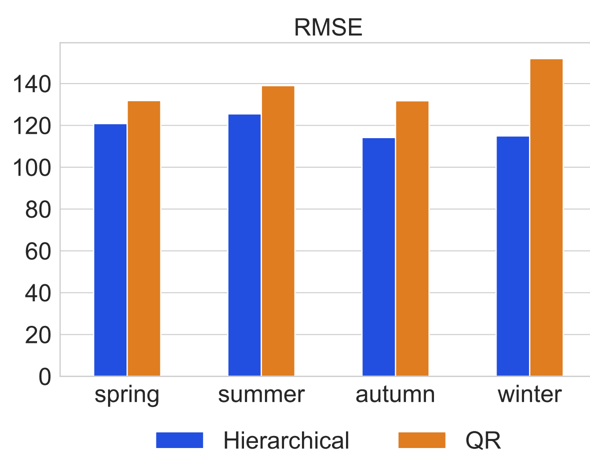


Fig. 3: Average RMSE loss for each season

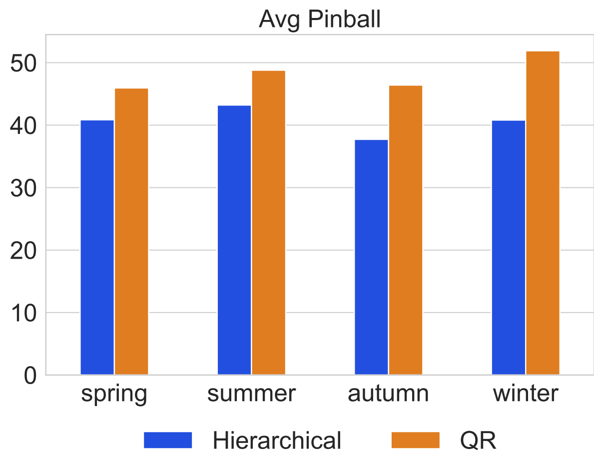


Fig. 4: Average Pinball loss for each season

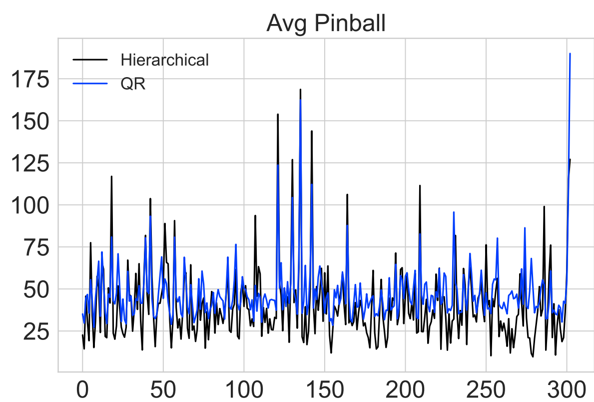


Fig. 5: Daily average pinball loss

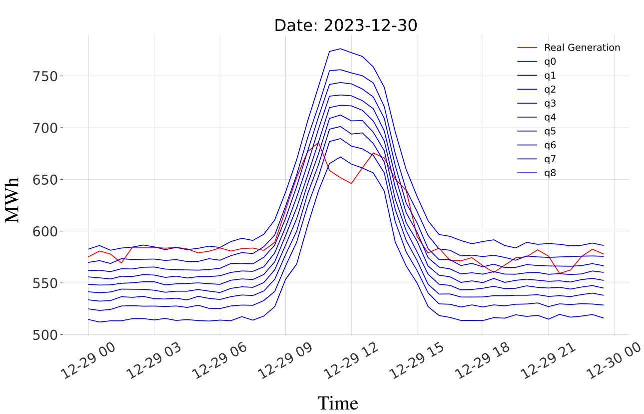


Fig. 6: High wind, sunny day

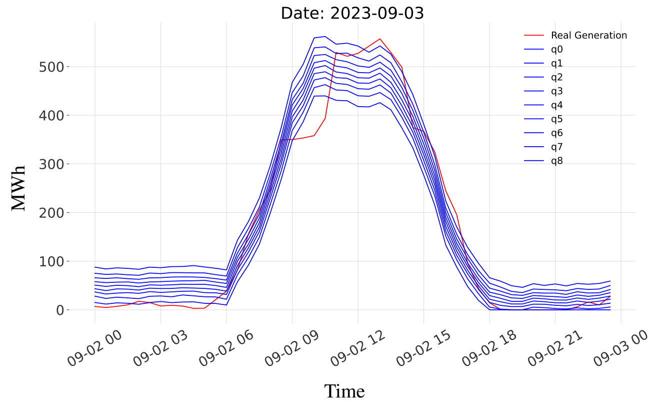


Fig. 7: Low wind, sunny day

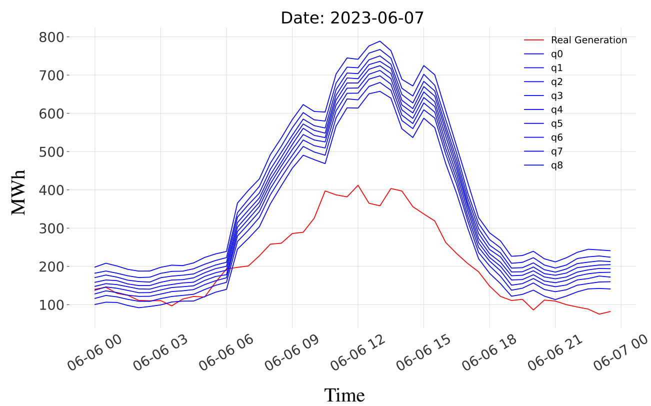


Fig. 8: Overestimation of predictions leads to higher pinball loss

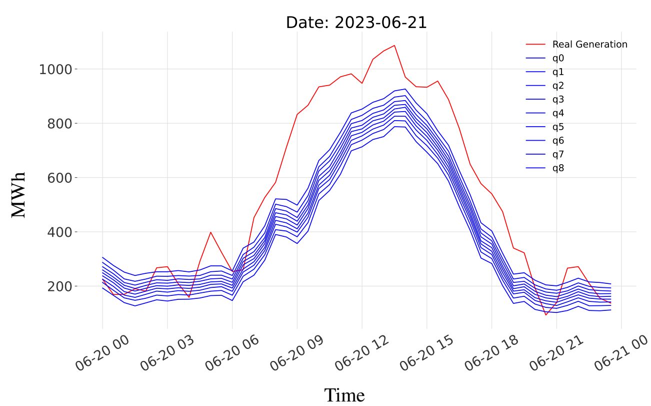


Fig. 9: Underestimation of predictions leads to higher pinball loss

V. Conclusions

In this paper, we investigated the use of Probabilistic Gradient Boosting Machines for probabilistic hierarchical forecasting. The proposed method is able to consider any hierarchy and different types of power generation. The predictions are in the form of quantiles for the day-ahead production, considering the case in which real-time power production data are not available. The presented approach was tested with a two-level hierarchy considering the aggregated generation of multiple PV farms and a wind park, predicting their joint production. The results show the ability of the model to well predict the day-ahead power production, based only on the weather-forecasts. For future work, the model can be extended to the case in which real-time power production data are available at prediction time. Furthermore, it can be extended to more complex hierarchies considering bigger portfolio of plants.

REFERENCES

[1] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890419307812

[2] P. Pinson and H. Madsen, "Ensemble-based probabilistic forecasting at horns rev," *Wind Energy*, vol. 12, no. 2, pp. 137–155, 2009.

[3] S. Alessandrini, F. Davò, S. Sperati, M. Benini, and L. Delle Monache, "Comparison of the economic impact of different wind power forecast systems for producers," *Advances in Science and Research*, vol. 11, pp. 49–53, 05 2014.

[4] A. Stratigakos, D. van der Meer, S. Camal, and G. Kariniotakis, "End-to-end learning for hierarchical forecasting of renewable energy production with missing values," in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2022, pp. 1–6.

[5] J. Jeon, A. Panagiotelis, and F. Petropoulos, "Probabilistic forecast reconciliation with applications to wind power and electric load," *European Journal of Operational Research*, vol. 279, no. 2, pp. 364–379, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221719304242

[6] C. Gilbert, J. Browell, and D. McMillan, "A hierarchical approach to probabilistic wind power forecasting," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018, pp. 1–6.

[7] G. M. Yagli, D. Yang, and D. Srinivasan, "Reconciling solar forecasts: Probabilistic forecasting with homoscedastic gaussian errors on a geographical hierarchy," *Solar Energy*, vol. 210, pp. 59–67, 2020, special Issue on Grid Integration. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X20306137

[8] O. Sprangers, S. Schelter, and M. de Rijke, "Probabilistic gradient boosting machines for large-scale probabilistic regression," *CoRR*, vol. abs/2106.01682, 2021. [Online]. Available: https://arxiv.org/abs/2106.01682

[9] G. Mitrentsis and H. Lens, "An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting," *Applied Energy*, vol. 309, p. 118473, 2022.

[10] "Wetter und klima - deutscher wetterdienst."

[11] R. Koenker and K. F. Hallock, "Quantile regression," *The Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143–156, 2001. [Online]. Available: http://www.jstor.org/stable/2696522