

Influences in Forecast Errors for Wind and Photovoltaic Power: A Study on Machine Learning Models

Jens Schreiber Artjom Buschin Bernhard Sick¹

Abstract: Despite the increasing importance of forecasts of renewable energy, current planning studies only address a general estimate of the forecast quality to be expected and selected forecast horizons. However, these estimates allow only a limited and highly uncertain use in the planning of electric power distribution. More reliable planning processes require considerably more information about future forecast quality. In this article, we present an in-depth analysis and comparison of influencing factors regarding uncertainty in wind and photovoltaic power forecasts, based on four different machine learning (ML) models. In our analysis, we found substantial differences in uncertainty depending on ML models, data coverage, and seasonal patterns that have to be considered in future planning studies.

Keywords: uncertainty analysis, machine learning models, seasonal effects, data coverage

1 Introduction

With the further expansion of wind and photovoltaic (PV) energy, the power supply system will change significantly in the coming decades. The overall power supply will become more weather-dependent and solutions must be found to ensure a robust and inexpensive power supply that maintains grid stability.

The major challenges for the grid stability caused by the energy system's transformation can mainly be traced back to two aspects: Firstly, the actual power supply of wind and solar energy plants is directly dependent on the weather and thus not directly compatible to consumption. Secondly, the expected power of the next hours and days is uncertain due to the strong dependence on the weather and must be predicted by power forecasts based on numerical weather predictions (NWP).

Despite the increasing importance of forecasts for renewable power supply, current planning studies only address the forecast quality to be expected in the future for the whole of Germany based on representative forecasts (see, e.g., the dena grid study [KAS10]). Further, often, these studies only consider a limited number of forecasts horizons. However, these estimates allow only for limited and highly uncertain use in the planning of the electricity supply system. More reliable planning processes require considerably more information about future forecast quality.

¹ University of Kassel, Intelligent Embedded Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany
{j.schreiber, artjom.buschin, bsick}@uni-kassel.de

Therefore, the article provides a comprehensive study on influences in forecast uncertainty, which has to be taken into consideration for future planning studies. The article investigates uncertainty in four types of common ML models for wind and PV: Least absolute shrinkage and selection operator (LASSO), gradient boosting regression tree (GBRT), support vector regression (SVR), and multi-layer perceptron (MLP). Models are trained to forecast the estimated day-ahead power generation based on NWP features as input. By repeated training with different test datasets, we create forecasts over the entire datasets for later analysis.

In the next step, we compare error distributions for each model concerning known influencing factors: Amount of training samples, forecast horizon, the terrain of wind farms, and a comparison of the uncertainties between the different machine learning (ML) models. By comparing binned forecasts errors, e.g., for different forecast horizons, with the Kullback-Leibler Divergence (KLD), we measure similarities and differences of these distributions. This comparison allows estimating when a bin is substantially different compared to a baseline and therefore gives insights to influential factors. Further, bins are compared with the Kruskal-Wallis [HS18] hypothesis test to verify a significant difference. The main contributions of this article are:

- We utilize common ML models, the grid search algorithm to find optimal model parameters, and common feature engineering techniques to provide forecasts results for wind and PV farms. By repeated training of the models on different training sets, we create forecasts for the complete dataset.
- For wind power forecasts, we show that the amount of training samples influences the forecast errors up to a certain threshold of data coverage (where data coverage is the proportion of the maximum number of data samples in the dataset to the actual amount of data samples within the historical data).
- Analyzing seasonal patterns reveal different influences for wind and PV that are related to the specific weather conditions within the individual seasons. Interestingly, forecast errors of adjacent seasons are not necessarily similar to each other in PV forecasts.
- The comparative study of ML forecasts models shows, that wind forecast errors within a similar terrain are more alike than for a similar amount of training samples. This relation suggests that forecasts models are more alike when external influences are excluded.

The remainder of this article is structured as follows. In Section 2 we detail related work. Section 3 outlines evaluation measures and applied ML models. Section 4 describes the experimental design and evaluation results w.r.t. data coverage, seasonal patterns, terrain, and model differences. Finally, we conclude our work and propose future work in Section 5.

2 Related Work

In current planning studies on future energy systems, considerations on the current and future uncertainty of power forecasts are only inconsiderably taken into account. The German Dena II study [KAS10], e.g., only considers forecasts error up to a horizon of two hours, neglecting that an increasing amount of renewable requires a larger forecast horizon such as day-ahead forecasts. Further, the study is missing an analysis of seasonal effects and forecast model specific uncertainties.

These (mostly) missing influential factors [Ya15] categorizes into the NWP input data, the power curve, and the prediction algorithm. The thesis of M. Lange [La03] relates forecasts error to the NWP data. In particular, the forecast uncertainty is assessed with respect to (w.r.t.) certain meteorological situations. However, the study only employs a physical model of the power curve with error correction and spatial refinement.

In the work of [Pi06], time series analysis techniques (e.g., ARIMA, ARX, Box-Jenkins) and a physical model are used to evaluate the forecasting skill. The author includes an analysis for different forecast horizons based on R^2 and root mean squared error (RMSE). Further, it contains a small subsection on the evaluation of the error distribution.

The results of [Mö04] include an analysis for time horizons between zero to nine hours and up to five days ahead. But the study is again focusing on the physical model and not considering machine learning models. Also [HMS13; KHP15] are focusing on time series analysis techniques (NARX) and (adapted) physical models, for uncertainty analysis.

More recently, in [Ge18] uncertainty in ML models, such as MLP, SVR, and an ensemble technique, are analyzed, but the thesis misses an evaluation for error distributions relating to different forecast horizons of wind power. Also, in [Re17] an analysis of ML models such as extreme gradient boosting technique, random forest, adaptive boosting, and persistence method is used to access the economic value for PV power generation.

[BT18] uses physical and semi-physical models for developing a forecast methodology for households that do not have access to solar irradiance information and are therefore limited to discrete weather information. The results are analyzed w.r.t. the weather features.

An interesting approach presents [NH18], in which the kriging method interpolate data with geographical properties for a location with no available data. A Naïve Bayes classifier along with a Gaussian probability distribution based on the overall data performs day-ahead forecasts of solar power based on the probability in one-hour intervals. The method is evaluated against the persistence model with mean absolute error (MAE) for different months of a year.

The simulation in [NR13] creates uncertainties for PV at different time scales to evaluate the economic and reliability effect for the grid. As it is a simulation tool, it is different from

ML models. The proposed method in [MOO18] allows for modeling the PV uncertainty based on past observations by using multivariate normal distributions.

The literature review shows that most of the work is focusing on models relating to time series analysis techniques and physical models. Further, the reviewed articles are missing a quantified comparison between the distribution of uncertainties or are even missing an in-depth analysis of the error distribution.

3 Method

This section gives a summary on common ML algorithms and presents their differences, to evaluate influences in forecast uncertainty. By using different ML algorithms, we assure to cover a broad spectrum in forecast errors. In the final section, we summarize error measures to estimate the deviation between actual and forecasted power generation.

3.1 Lasso

LASSO, also known as basis pursuit, is a linear model. Linear models typically provide a robust estimation, when NWP are uncertain. Further linear models allow measuring the contribution of individual features through their coefficient, hence, making them highly relevant for analysis on error origin [HTF01]. In contrast to other linear regression models, LASSO allows for automatic selection of essential features. This selection is achieved by L_1 penalty, that effectively causes the coefficient of features to be exactly zero and hence excluding individual features.

3.2 Support Vector Regression

SVR is based on the concept of support vector machines (SVMs) for classification with changes in the definition of the optimization problem. One appealing property of SVMs is that the determination of parameters is locally and globally optimal due to the convex optimization [Bi06]. Further, by making use of the *kernel-trick* original NWP input features are transformed in a higher dimensional, even infinite dimensional, space. Transforming features into a higher-dimensional space provides features that are linearly separable [Va00]. The transformed features allow the SVR to achieve good results in many applications [Bi06], making them highly relevant for the evaluation of forecast uncertainty.

3.3 Multi-layer Perceptron

MLPs and more recently deep neural networks are a common technique for regression and classification tasks. In a feed forward MLP input features are transformed using

matrix multiplication and a subsequent (mostly) non-linear transformation. The former two operations are summarized as layers and successive applications of these layers, where the output of one layer is the input to the next layers, allows us to find a good representation of the data. In the final layer, the output layer, a simple linear combination can be used for renewable energy forecast. Primarily through their capability to find suitable representations of the NWP data, MLPs achieve state of the art performance in renewable power forecast [Ge16]. This performance makes them highly relevant for the evaluation of forecast uncertainty.

3.4 Gradient-Boosting-Regression-Tree

GBRT originate from the idea, that a combination of weak learners improves the overall performance. Therefore, the gradient boosting algorithm trains trees in regions of most substantial forecast error. The ensemble technique combines the individual trees improving the overall performance. A single tree partitions the features space in a set of rectangles and estimates a constant forecast value for each rectangle [HTF01]. The partitioning provides an interpretable structure to explain forecast decision, which is not feasible with SVR and MLPs. Further, the algorithm is not making use of any data representation techniques as with these approaches.

3.5 Error Measures

To assess influences in forecast uncertainty through forecast errors, it is essential to evaluate the error with $e = y - \hat{y}$. It gives insights between the actual power generation y compared to the forecasted power \hat{y}_i . In contrast to mean based measures, e provides the most detailed view on the error; combined with a visualization of the error distribution through a histogram or a boxplot it allows to assess skewness and other statistical measures of the error distribution. A comprehensive analysis of deterministic error measures in the field of renewable energy forecast is given in [Ge18]. The results can be summarized as follows

- The coefficient of determination R^2 assesses how much of the variance in the historical power data is explained by the model. As it is only capable of evaluating the amount of linear correlation, it is often used as a measure to compare different forecast techniques.
- To account for extreme errors of e , quadratic errors such as the mean squared error (MSE) are recommend.
- Absolute measures such as MAE are suited for monetary evaluation criteria (linear evaluation criteria).

In the following, we will stick to e^2 as it allows for comparison of overall forecast quality of the model by terms of mean (MSE) and median, especially when visualized via boxplot.

To compare distributions of errors with each other, we use the KLD. The KLD is a non-symmetric statistical measurement to determine the difference between two distributions allowing to quantify the similarity, e. g., between the error distribution from the GBRT and the SVR.

4 Experimental Evaluation

In the following, we provide analysis on error distributions from different ML models and measure their similarity to another for wind and PV. By estimating the KLD between distributions for different (external) factors, we get insights on how they relate to each other. Therefore, we first give details on the model training and the two datasets. The first study estimates influences caused by a limited amount of training samples for wind power forecasts. Results are evaluated w.r.t. the data coverage, where data coverage refers to the proportion between the maximum number of data samples to the actual amount of data samples within the dataset. In the next section, we analyze seasonal influences such as the hour of the day or season of the year for wind and PV as well as terrain specific influences in the wind dataset. As results for PV models suggest that there are strong seasonal patterns to consider - that are less present for wind models - we limit the final analysis to the WindFarm dataset. Limiting these and other external influences allow to compare the error distributions of the different power forecasting models.

4.1 Design of Experiment

For the following two datasets we train the LASSO, SVR, MLP, and GBRT to forecast the power generation optimized through grid-search for the following parameters^{2,3}:

- **LASSO:** Alpha (0.1 - 1) and maximum number of iterations (1000-4000).
- **SVR:** Penalty (0.01-0.001) and gamma (0.01-1000).
- **GBRT:** Max depth (12 -20), minimum samples split (2-128), number of estimators (1200-2000), and learning rate (0.05-0.2).
- **MLP:** Maximum number of iterations (10000), early stopping (true or false), learning rate ($0.01-1 \times 10^{-6}$), learning rate type (constant or adaptive), and hidden layer sizes ((100, 50, 30, 15, 5), (50, 100, 200, 50), (20, 10, 5)).

Solar Farm Dataset: The *SolarFarm* dataset consists of 114 PV facilities in Germany. Their installed nominal power ranges between 7.2kW and 12573kW. The dataset has a

² Other model parameters are the default parameters by scikit-learn, last-accessed June 2015.

³ Parameters are selected to obtain a similar training time for all models (except LASSO as this is a linear model) while achieving a good performance based on initial testing.

three-hour resolution and is recorded from the beginning of 2016 to the end of April 2017, resulting in a maximum of 3880 data points. In total the dataset has 51 input features as input. Features with correlation to the power generation (e.g., sun position, solar height, clear sky, and radiation) are shifted in time by three hours to take future and past effects of the weather into account for prediction.

Wind Farm Dataset: The *WindFarm* dataset contains the power generation taken from 54 wind farms that are distributed throughout Germany. These values were recorded hourly over two years (2016 and 2017), resulting in a maximum of 17520 data points. The dataset contains information about the terrain of each farm (flatland, forest, and offshore). In total the dataset has 7 NWP features as input. Features of wind speed and wind direction influencing the power generation [SS18] are time-shifted by two hours to take future and past effects of the weather into account for prediction.

Both datasets were manually filtered to remove outliers, e.g., caused by maintenance. Depending on the number of outliers and maintenance the amount of data coverage ranges between 50 and 100 percent for wind data w.r.t. recorded period, where data coverage refers to the proportion between the maximum number of data samples to the actual amount of data samples within the dataset. The data coverage for PV is mostly above 90%.

To compare forecast errors, we normalize the generated power by the maximum power generation. Input features are standardized for zero mean and unit variance based on the training dataset in each run. We optimize each model through a grid search on the validation dataset. To make the best use of the full data range, we use different runs of the experiment to shift the test data throughout the recorded period: Six months for the wind and four months for the PV dataset resulting in four runs for each dataset. In each run, the remaining data is used for training (80%) and validation (20%). After completing all training runs, combining predictions from all test datasets provides an *evaluation dataset* for estimating influences in the entire period of the original data. To account for extreme errors and measure the quality between a single forecast and the historical power, we use the squared error. We fit distributions of the squared error with the χ^2 distribution to compare them with the KLD

4.2 Influence of the Amount of Training Data

The digitalization of the current and future energy market will provide an increasing amount of training data. To determine the extent to which the amount of training data influences the forecast error, we analyze it in this section.

Therefore we estimate the data coverage of a farm in percent compared to the maximum number of data points. It turns out that the data coverage in PV farms is consistently above 92% except for one farm, respectively, we do not consider it in further analysis. However, the data coverage from wind farm range between 49 and 99% allowing for clustering them in ten percent steps, see Figure 1. As the size of the test dataset is constant in each run, the

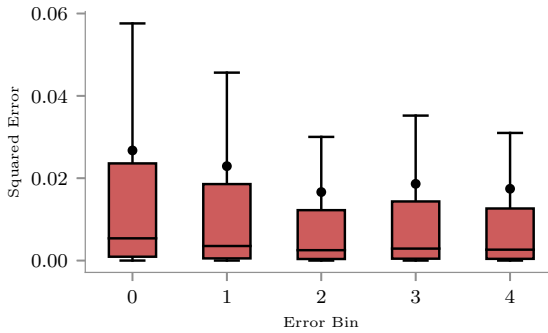


Fig. 1: Boxplot on squared error for bins of data coverage for the WindFarm dataset based on the MLP model. From left to right the data coverage is as follows: 50 – 60%, 60 – 70%, 70 – 80%, 80 – 90%, and 90 – 100%. The mean of the error is visualized as a dot.

Bins	0	1	2	3	4
0	0.000	0.003	0.044	0.037	0.065
1		0.000	0.068	0.059	0.094
2			0.000	0.000	0.002
3				0.000	0.004

Tab. 1: KLD measuring the similarity between different amounts of data coverage for the error of the GBRT model on the WindFarm dataset. From one to five the data coverage is as follows: 50 – 60%, 60 – 70%, 70 – 80%, 80 – 90%, and 90 – 100%.

size of the training data is as well, respectively, the data coverage is directly linked to the amount of training data and will be treated equally in the following.

Figure 1 shows the relation between the number of training samples and the error: With the increasing amount of data, the median as well as the mean decrease. The spread of the error is similar for bin two, three, and four. Bin zero and one have a broader spread of the error. The decreasing mean, median, and spread show that there is a relation between the amount of data for training and the forecast error.

To verify a significant difference between these bins, we compare them with the KLD and the Kruskal-Wallis hypothesis test. Kruskal-Wallis hypothesis shows that the forecast error for all ML models and all bins of data coverage are significantly different at a significance level of $\alpha = 5\%$. The exemplary results in Table 1, highlight the previous observation: A decreasing data coverage, causes an increased spread, median, and mean resulting in larger values of the KLD, e.g., when comparing bin zero with bin four. Bin two, three, and four are quite similar to each other nonetheless.

As expected, there is a relation between the amount of data available and the forecast error.

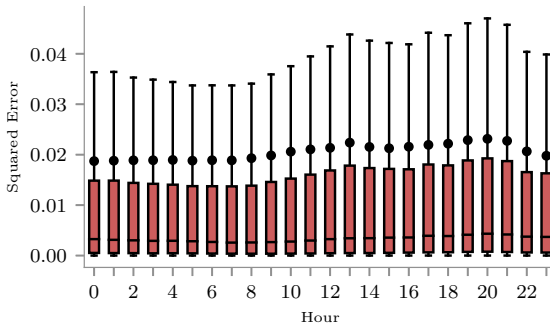


Fig. 2: Boxplot on squared error for bins of the hour of the day for the WindFarm dataset based on the MLP model. The mean of the error is visualized as a dot.

With an increasing amount of available data, the ML model tends towards a minimum error, the NWP input data probably cause that.

4.3 Influences by Seasonal Patterns and Terrain

Seasonal influences that are present in seasons of a year or hours of a day are well known. Nonetheless, there is limited research on how these patterns affect the error distribution in wind and PV forecast based on ML techniques. More common is the analysis of forecast error w.r.t. their terrain, which this section also covers.

In the following, we address season of a year and the hour of the day. In terms of PV, the hour of the day has two meanings. First, due to the daily pattern of the sun, we can observe patterns within the power generation. Second, with the rising time of the day, the forecast horizon of the NWP model increases (as the so-called NWP model run typically originates from 12 UTC). As the horizon increases, the error of the weather forecast increases and respectively that of the power forecast model. The latter also holds for wind power forecasts.

In the sample boxplot, Figure 2 for wind errors, we can observe this pattern. The median and mean errors for different hours of the day do not increase drastically due to the absence of seasonal weather patterns in the wind; detailed observations exist when measuring similarity through KLD in Table 2. The errors at the end of the day are more similar to each other employing the KLD, compared to the origin of the weather forecast due to the increased forecast error of the NWPs. Nonetheless, all errors, when comparing different hours of the day, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) except: Four cases in the linear model, one for the MLP, and two comparisons for the GBRT model.

Hour	0	3	6	9	12	15	18	21
0	0.000	0.000	0.057	1.905	7.619	5.150	0.831	0.437
3		0.000	0.067	1.975	7.815	5.294	0.873	0.467
6			0.000	1.212	5.636	3.707	0.431	0.173
9				0.000	1.003	0.449	0.168	0.423
12					0.000	0.094	2.190	3.168
15						0.000	1.244	1.932
18							0.000	0.054

Tab. 2: KLD measuring the similarity between hours of a day for the error of the MLP model on the PVFarm dataset.

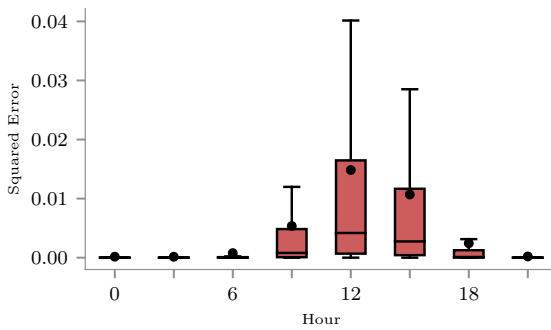


Fig. 3: Boxplot on squared error for bins of an hour of the day for the SolarFarm dataset based on the GBRT model. The mean of the error is visualized as a dot.

For PV however, see Figure 3, we can observe a strong seasonal pattern in the error distributions for different hours of the day. This observation is also present when estimating the KLD, resulting in large values when comparing 12 with 0 o'clock, see Table 2. This seasonal pattern is to expected, as during the night there is no power generation and respectively the difference in the error distribution is notable when compared to the day. Compared to wind, there are also more considerable differences in the error distributions during the daytime. The daily pattern of the sun causes these differences that result in different error distributions. Again, all errors, when comparing different hours of the day, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) except: Four cases in the SVR model and one case for the MLP model.

In the analysis for different seasons of a year for wind, we observe that in the third season all models and datasets have the lowest median, mean, and spread of the error for wind. In other seasons of the year, extreme weather conditions are more common, causing larger error values. The KLD, see Table 3, confirms our intuition, that error distributions for seasons close to each other are more similar to another than those far away.

Season	1	2	3	4
1	0.00	0.02	0.19	0.10
2		0.00	0.09	0.04
3			0.00	0.01

Tab. 3: KLD measuring the similarity between seasons of a year for the error of the SVR model on the WindFarm dataset. Season one equals winter, season two equals spring, etc..

Season	1	2	3	4
1	0.00	0.18	3.21	0.08
2		0.00	1.58	0.02
3			0.00	2.02

Tab. 4: KLD measuring the similarity between seasons of a year for the error of the SVR model on the SolarFarm dataset. Season one equals winter, season two equals spring, etc..

Contrary to wind forecasts, PV models have more substantial errors in the third season. In other seasons of the year, the different position of the sun causes a different amount of direct and diffuse radiation making it the forecast model easier to forecast the power generation. For instance, the solar radiation (direct and diffuse) is the smallest in season one in the dataset. The analysis of the KLD in Table 4 suggest that the difference of uncertainty is significant even for seasons of the year that are close to another. These differences are caused by the larger magnitude of the forecast error, especially in the third season. Only when comparing season one and three for errors of the GBRT model on the WindFarm dataset the Kruskal-Wallis ($\alpha = 0.05$) estimates no significant difference.

The analysis of the terrain in Figure 4 shows that the smallest errors are present for parks located in a farmland terrain. All errors, when comparing the different terrains, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$), where farmland has the smallest, forest the second smallest, and offshore the most substantial forecast error. Note that the terrains have a varying amount of farms. Farmland has 37, the forest has 11, and offshore includes four farms. Even though the terrains have a varying amount of farms, the larger error for offshore parks might be caused by large ramping events of the wind. Interestingly, when measuring the similarity, the error distribution of offshore farms is closer to the farmland, than farmland to the forest employing the KLD. This smaller KLD might be due to more complex weather conditions in the forest and offshore terrain. For instance, turbulence on the sea might be similarly present in forests (that are often also elevated) causing a similar uncertainty distribution.

Conclusively, we showed similarity and dissimilarity in seasonal and terrain specific patterns. Interestingly, the difference in error distribution is one of the largest for different seasons of the year for wind and PV. Wind errors are more significant in the winter and autumn, while PV models have larger values in the spring and summer time. Finally, the uncertainty distribution in offshore terrain is like that of the forest.

4.4 Influences by Models

After analyzing external influences to the error distributions, in this section, we are interested in comparing the similarity between the ML models. As results for PV models suggest that

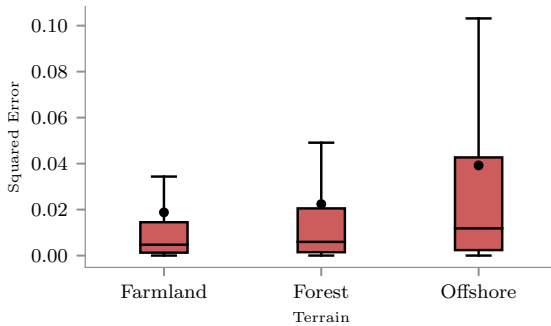


Fig. 4: Boxplot on squared error for different terrains for the WindFarm dataset based on the GBRT model. The mean of the error is visualized as a dot.

	GBRT	LASSO	SVR	MLP
GBRT	0.00	1.53	0.05	0.14
LASSO		0.00	0.98	0.66
SVR			0.00	0.02

Tab. 5: KLD measuring the similarity between ML models with data coverage between 90 and 100%.

	GBRT	LASSO	SVR	MLP
GBRT	0.00	1.46	0.05	0.15
LASSO		0.00	0.94	0.61
SVR			0.00	0.03

Tab. 6: KLD measuring the similarity between ML models for a farmland terrain.

there are strong seasonal patterns to consider - that are less present for wind models - we limit further analysis to the WindFarm dataset.

In previous results from wind models, we show that the error is the smallest for the farmland terrain and when training the ML model on a data coverage between 90 to 100%.

Limiting the analysis to the smallest errors gives us insights, more similar to an absence of external influences, as those influences must have a smaller effect on the error distribution compared to distribution with larger mean, median, and spread. Ultimately, allowing to access the differences in the ML models and not those caused by external influences.

In both analyses, we observe that GBRT achieves the smallest error, SVR the second, MLP the third lowest and LASSO has the most substantial error. Table 5 and 6 summarize the the difference in their error distributions for the experiment with maximum data coverage and the farmland terrain.

Results suggest that distributions within a terrain are more similar to each other than within maximum data coverage caused by the relation that specific weather conditions, are individual for different terrains, resulting in terrain specific forecast errors. Nonetheless,

estimates of the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) shows that they are still substantially different.

5 Conclusion and Future Work

In this article, we presented an in-depth analysis and comparison for influencing factors of uncertainty in wind and PV power forecasts based on four different ML models. In our analysis, we found substantial influences and differences between compared bins of uncertainty that reveal the need to consider them in future planning studies.

For instance, the study reveals strong seasonal patterns in the uncertainty for wind and PV. For wind power forecasts, neighboring seasons and hours are similar to each other. For seasonal patterns within a year, these forecasts will benefit from optimizing NWP forecasts for extreme weather situations that cause substantial errors in the winter time. Due to significantly larger errors for the third season adjacent seasonal bins in PV forecasts are not necessarily similar to each other. Similar results are obtained for daily patterns. For daily patterns, we recommend to use NWP forecasts that are closer to the time (noon) of most substantial error.

By analyzing the relation between the amount of training data and the uncertainty, we showed that models improve when using additional data up to a data coverage of about 70%. Reducing this error further is, e.g., possible with deep learning models that have a higher capacity to learn the relation between NWP features, e.g., they can be trained across several parks in a multi-task-learning setting. However, even with an increasing amount of data, the minimum forecast error will be limited to that error caused by the NWP.

The study reveals that after minimizing external influences, differences in the uncertainty distributions from the four ML models are still present motivating the need to consider the underlying forecast model in future planning studies.

In the future, we aim to repeat the analysis for intra-day, different NWP models, and data with a 15-minute resolution. Further, we will investigate how *transfer learning* can be utilized to reduce forecast uncertainty when limited data is available.

Acknowledgement This work was supported within the project Prophecy (0324104A) funded by BMWi (Deutsches Bundesministerium für Wirtschaft und Energie / German Federal Ministry for Economic Affairs and Energy).

References

- [Bi06] Bishop, C. M.: Pattern Recognition and Machine Learning. Springer, 2006.
- [BT18] Brecl, K.; Topic, M.: Photovoltaics (PV) system energy forecast on the basis of the local weather forecast: Problems, uncertainties and solutions. *Energies* 11/5, p. 12, 2018.

- [Ge16] Gensler, A.; Henze, J.; Sick, B.; Raabe, N.: Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks. In: SMC. IEEE, pp. 002858–002865, 2016.
- [Ge18] Gensler, A.: Wind power ensemble forecasting, PhD thesis, University of Kassel, 2018, p. 204.
- [HMS13] Holttinen, H.; Miettinen, J.; Sillanpää, S.: Wind power forecasting accuracy and uncertainty in Finland. 2013.
- [HS18] Hedderich, J.; Sachs, L. 1.-.: Angewandte Statistik Methodensammlung mit R. Springer Spektrum, 2018.
- [HTF01] Hastie, T.; Tibshirani, R.; Friedman, J.: The Elements of Statistical Learning. Springer, 2001.
- [KAS10] Kohler, S.; Agricola, A.-C.; Seidl, H.: Integration erneuerbarer Energien in die deutsche Stromversorgung im Zeitraum 2015-2020 mit Ausblick 2025, tech. rep., 2010, p. 564.
- [KHP15] Ko, W.; Hur, D.; Park, J.-K.: Correction of wind power forecasting by considering wind speed forecast error. JICEE 5/1, pp. 47–50, 2015.
- [La03] Lange, M.: Analysis of the Uncertainty of Wind Power Predictions, PhD thesis, 2003, p. 139.
- [Mö04] Möhrle, C.: Uncertainty in wind energy forecasting, PhD thesis, University College Cork, 2004, p. 178.
- [MOO18] Murata, A.; Ohtake, H.; Oozeki, T.: Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals. Renewable Energy 117/, pp. 193–201, 2018.
- [NH18] Nam, S.; Hur, J.: Probabilistic Forecasting Model of Solar Power Outputs Based on the Naive Bayes Classifier and Kriging Models. Energies 11/11, p. 15, 2018.
- [NR13] NREL: Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales, tech. rep. May, 2013, p. 41.
- [Pi06] Pinson, P.: Estimation of the uncertainty, PhD thesis, Ecole des Mines de Paris, 2006, p. 266.
- [Re17] Reindl, T.; Walsh, W.; Yanqin, Z.; Bieri, M.: Energy meteorology for accurate forecasting of PV power output on different time horizons. In: Energy Procedia. Vol. 130, pp. 130–138, 2017.
- [SS18] Schreiber, J.; Sick, B.: Quantifying the Influences on Probabilistic Wind Power Forecasts. In: ICPRE. Vol. 3, p. 6, 2018.
- [Va00] Vapnik, V. N.: The Nature of Statistical Learning Theory. Springer-Verlag, 2000, ISBN: 978-0-387-98780-4.
- [Ya15] Yan, J.; Liu, Y.; Han, S.; Wang, Y.; Feng, S.: Reviews on uncertainty analysis of wind power forecasting. Renewable and Sustainable Energy Reviews 52/, pp. 1322–1330, 2015.