



**HAL**  
open science

# Probabilistic Forecasting of Regional Wind Power Generation for the EEM20 Competition: a Physics-oriented Machine Learning Approach

Kevin Bellinguer, Valentin Mahler, Simon Camal, Georges Kariniotakis

## ► To cite this version:

Kevin Bellinguer, Valentin Mahler, Simon Camal, Georges Kariniotakis. Probabilistic Forecasting of Regional Wind Power Generation for the EEM20 Competition: a Physics-oriented Machine Learning Approach. 17th European Energy Market Conference, EEM 2020, KTH, IEEE, Sep 2020, Stockholm (by visio), Sweden. hal-02952589

**HAL Id: hal-02952589**

**<https://hal.science/hal-02952589v1>**

Submitted on 29 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Probabilistic Forecasting of Regional Wind Power Generation for the EEM20 Competition: a Physics-oriented Machine Learning Approach

Kevin Bellinguer<sup>1</sup>, Valentin Mahler<sup>1,2</sup>, Simon Camal<sup>1</sup> and Georges Kariniotakis<sup>1</sup>

<sup>1</sup>MINES ParisTech - PSL University, PERSEE - Centre for Processes, Renewable Energies and Energy Systems, CS 10207, rue Claude Daunesse, 06904 Sophia-Antipolis Cedex, France

<sup>2</sup>French Environment and Energy Management Agency 20, avenue du Grésillé- BP 90406 49004 Angers Cedex 01 France  
Email: {kevin.bellinguer, valentin.mahler, simon.camal, georges.kariniotakis}@mines-paristech.fr

**Abstract**—Variable renewable energy has a growing impact on electricity markets and power systems in many regions of the world. In this context, the 17th International Conference on the European Energy Market EEM20 set up a competition to develop probabilistic forecasting tools of wind production at a regional level. This paper proposes an adaptive approach for regional wind power forecasting. A physics-oriented pre-processing of the data delivers analog weather patterns and wind-power-related variables, then a k-means clustering of wind farms further reduces the dimension of the problem. The generated representative features feed a Quantile Regression Forests model that produces sharp and reliable predictions. As a result, our model won the competition with a relative improvement of the average pinball loss of 6.7% and 14.7%, compared to the teams ranked second and third respectively.

**Index Terms**—Analog, Competition, Forecasting, Smart grid, Wind energy.

## I. INTRODUCTION

In recent years, the share of wind power in Sweden's total energy production has grown: the installed capacity increased from 1.5 GW up to 9 GW between 2009 and 2019 [1]. Sweden has reaffirmed its commitment by aiming to meet 100% of its electricity needs from Renewable Energy Sources (RES) by 2040 [2]. To address this challenge, the International Renewable Energy Agency, and the Swedish Energy Agency have proposed different innovative solutions, which include paying special attention to RES production forecasting [2].

It is known that the variability of wind power and the uncertainty associated with forecasting wind production have an influence on electricity spot prices and on power producers'

The authors would like to thank ADEME, the French Environment and Energy Management Agency, ARMINES and Compagnie Nationale du Rhône (CNR), France's leading 100% RES producer, for their supports in financing the PhDs of V. Mahler and K. Bellinguer respectively. The models of the paper were based on research work carried out by MINES ParisTech in the frame of the Horizon2020 project Smart4RES, which received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under grant agreement No 864337. The sole responsibility for the content lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA, the EC, ADEME and CNR are not responsible for any use that may be made of the information it contains.

978-1-5386-5541-2/18/\$31.00 ©2020 IEEE

revenue [3]. In both grid management and market applications, wind production forecasting at short-term horizons is crucial, particularly at day-ahead horizons (12h-48h horizon). At such horizons, Numerical Weather Predictions (NWP) are known to be essential explanatory variables for the prediction of renewable power production [4].

Deriving forecasts for regional or national RES production is challenging due to the high dimensionality of the problem (e.g. if deciding to upscale forecasts at the level of individual plants) and the possible redundancy of weather information for neighbouring sites. Direct forecasting of the total production is appealing because it removes the need to have access to production time series at each plant [5]. An intermediate approach consists in deriving forecasts on clusters of wind farms and then summing these middle-level forecasts to obtain regional forecasts [6].

In many applications for grid management or market bidding, a probabilistic forecast is needed to characterize the uncertainty associated with predictions, and probabilistic parametric approaches have been proposed in the context of regional RES production forecasting [5], [7]. The probabilistic approach developed in this work contributes to the scope of the Horizon2020 Project Smart4RES, which proposes next-generation solutions for renewable energy forecasting over a large range of scales, from the wind turbine level to large region-wide aggregations [8].

This paper presents the approach and the choices made by the authors during the EEM20 competition. The goal was to generate day-ahead quantile forecasts of the aggregated wind power production for the four Swedish bidding areas (Fig. 1(a)). The challenge started on March 4<sup>th</sup> 2020 and lasted six weeks. For each task, i.e. each week, the participating teams submitted hourly day-ahead probabilistic forecasts of wind production (10% to 90% quantiles) covering a period of 2 months. In total a year of forecasts was delivered. A particular challenge of the competition was that power production observations were only provided at the price region scale. The rules and results of the competition are given in [9], [10]. Over the course of the competition, our team incrementally upgraded the model to improve forecasting performances which enabled

us to rank first [9]. For simplicity’s sake, this paper presents only the last version of the model as it was used for the last step (see also [11]).

Considering the state of the art in day-ahead regional wind power forecasting and existing works in related forecasting competitions, the approach proposed is based on several original contributions that are going to be detailed in a forthcoming journal paper. Among them we can highlight the following:

- A physics-oriented machine learning model for day-ahead direct forecasting of regional wind production is proposed: a QRF model learns from a combination of wind-power-driven variables, analog weather conditions and wind farms clustering.
- The adaptivity and robustness of the approach is validated by having won a challenging forecasting competition.

The approach is qualified as *physics-oriented* because it captures weather patterns at a regional scale, and because it extracts information on the power generation of multiple wind farms by means of physics-inspired models (e.g. wind speed vertical profiles and wind-to-power conversion). The paper is organized as follows. Section II describes the data used for competition. Then section III presents the forecasting approach and details the methodology. IV discusses the outcomes. Finally, section V draws the conclusions of this study.

## II. AVAILABLE DATA

### A. Overview and Timeline

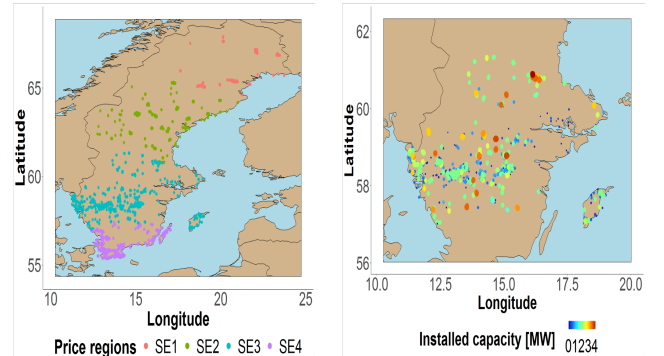
The datasets considered throughout this study are composed of regional wind power production observations and NWP at an hourly resolution. In addition, standing and technical data on wind energy converters (WEC) are also provided. At the beginning of the competition, a one-year long training dataset was provided (thereafter, this year is referred to as the year 2000). Then, new sets of NWP data corresponding to the next two months were successively provided on a weekly basis, in order for the participants to submit production forecasts. After each submission, the aggregated wind power observations, provided for validation purposes, were added to the training data set. At the end, the available data ranges from 2000-01-01 to 2001-12-31.

### B. Wind Power Generation and Wind Turbines

The aggregated wind power production over the four Swedish price regions SE1, SE2, SE3 and SE4 were provided (cf. Fig. 1(a)). By the end of 2001, 4004 WEC constituting 8640 MW of installed capacity, had been put into service according to the dataset. However, the competition organizers mentioned that the Swedish Wind Power Association listed 4099 WEC, with a total installed capacity of 8984 MW.

We computed the installed capacity by price region from the technical data set. This data set gives WEC characteristics (e.g. rotor diameter, hub height), spatial position and commissioning dates. We observed some discrepancies between the installed capacity and the wind power production. For example, in November 2000, the observed wind production is higher than the installed capacity in SE1. A similar situation

is described in [12], the authors decided to excluded the concerned data. Here, we assumed that the commissioning dates were just approximations, therefore we chose another approach to mitigate the uncertainty regarding the validity of data, as detailed in section III-B.



(a) WEC distribution for the four price regions (b) Distribution of the installed capacity in SE3

Fig. 1. Spatial distribution of WEC

### C. Numerical Weather Predictions

As explanatory variables, we consider NWP from the Norwegian Meteorological Institute (MET Norway). No information regarding the NWP model specifications were communicated to the competitors. The inputs are made up of 10 ensembles and of the following variables: surface temperature, zonal and meridional 10 meter winds, wind gust, mean sea level pressure, relative humidity and total cloud cover. The NWP covers the entire Swedish territory with a  $10km \times 10km$  spatial resolution. These specifications induce a dimensionality burden that is tackled in Section III-C. It is worth mentioning that a few WECs are located outside the area covered by the NWP provided. Rather than extrapolating the NWP at these locations, we implemented a clustering approach assuming that close WECs are influenced by similar weather conditions.

## III. METHODOLOGY

### A. Overview of the Approach

Instead of investigating a model able to perform a regression in the large dimension of the complete original data set, we chose to focus on the creation of a relatively small set of information-rich features used as input into a machine learning algorithm, namely Quantile Regression Forests (QRF), which has proven its skills in probabilistic forecasting of aggregated renewable production [13]. Creating and selecting a few inputs has the advantage of shortening the computation time, thus giving us the possibility to try multiple models and combinations of inputs, and to better scale the influencing factors with the observed response. Due to the differences between prices regions (dominant weather patterns, characteristics of the installed capacity, etc.), we trained one model by price region (same approach, but different data points) without considering data from other regions, or spatio-temporal correlations between regions.

The first challenge for this forecasting objective is related to the evolution of the installed capacity and its uncertainty. In response, Section III-B presents the target engineering employed, involving a normalization of the total regional production. The next step involves crafting adequate model inputs. Data processing starts with the reduction of dimensionality in weather predictions and wind production sites (cf. Section III-C), which enables us to engineer features as presented in Section III-D. Finally, we explain why we chose the QRF model in Section III-E. The method is summarized in Fig. 2.

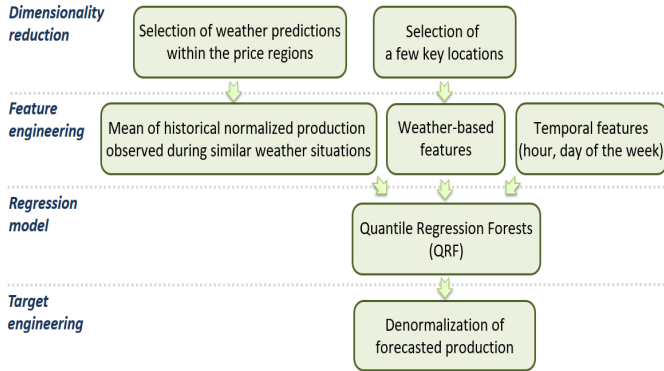


Fig. 2. Overview of the method

### B. Target Engineering

The installed wind capacity increases over the training and forecasting periods in all four regions at various rates (cf. Fig. 3). For this reason, we did not directly forecast the regional productions with the QRF model, but rather a normalized regional production also called load factor thereafter. This approach has the advantage of taking into account the temporal evolution of the WEC erections but does not reflect the evolution of the WECs' spatial distribution over time.

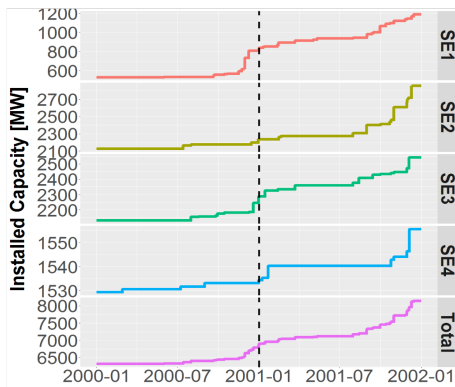


Fig. 3. Evolution of installed capacity

The most intuitive load factor would have been the ratio between the production ( $y_t$ ) and the installed capacity ( $P_{inst,t}$ ). However, to mitigate the uncertainty regarding the data quality of the wind turbines data set mentioned in the previous section, we considered a more robust denominator. For each price

region, we created a new time series ( $P_{obs,t}^{max}$ ) so that for any time  $t$ , its value is equal to the maximum production observed during the time period  $[0; t]$  (for the first six months, we simply assigned the maximum observed value over the period to avoid initialization issues). We then defined the load factor ( $y_t^*$ ) as:

$$y_t^* = \frac{y_t}{\text{mean}(P_{inst,t}, P_{obs,t}^{max})} \quad (1)$$

### C. Dimensionality Reduction

Considering the spatial extent of NWP as well as the number of ensembles and variables, nearly 840000 data points are available for each timestamp. To obtain a reasonable running time and to avoid using too much redundant or irrelevant data, we reduced the number of inputs in two dimensions, i.e. (1) the number of NWP ensembles, which contains potentially interesting information but dramatically increases the size of the problem; (2) the spatial extent of NWP points, in order to obtain a compact representation of weather conditions that has a level of detail appropriate for the aggregated regional production. On the other hand, a special focus was given to the relevance of weather variables. We chose to exclude the total cloud cover, but we processed other variables to create wind power related features.

1) *NWP Ensembles*: Although information can be retrieved from NWP ensembles (see [3], [14]), the study by [15] highlighted that production forecasts based on NWP ensembles have similar performances both when all ensembles are considered individually and when only the mean is used as input. Thus, we chose to reduce the data set by a factor of 10 by focusing on the mean of ensemble for each NWP variable.

2) *Spatial Extent*: Reducing the spatial extent not only decreases the computation time, but it also removes low-relevance NWP variables that take values in sites located far from WECs. Since we used different approaches for the feature engineering presented in the next subsection, we had to develop two methods in parallel to reduce the spatial extent and exploit the spatial distribution of WECs (cf. Fig. 1):

- **Price region areas**: we defined each price region as an area corresponding to a convex hull around its turbines using [16]. We then only considered the NWP's grid points within this area as shown in Fig. 4.
- **Key locations**: we selected a few key locations by clustering wind turbines. Inspired by [17], the key locations are the centroids ( $c$ ) of clusters ( $C$ ) defined with a weighted k-means approach, where weights depend on WEC installed capacity (a WEC belongs to  $C_i$  when  $c_i$  is the nearest centroid). A formulation of the weighted k-means clustering on a spherical surface adapted from [18] and [19] for this application is proposed in Eq. 2.

$$\arg \min_c \sum_{i \in [1..k]} \sum_{x \in C_i} w(x) \cdot d_g(x, c_i)^2 \quad (2)$$

Where:  $x$  = turbines coordinates  
 $w(x) = \frac{P_{inst}(x)}{P_{inst,C_i}}$  weight of WECs in  $x$  within  $C_i$   
 $d_g$  = geodesic distance, e.g. Haversine [20]

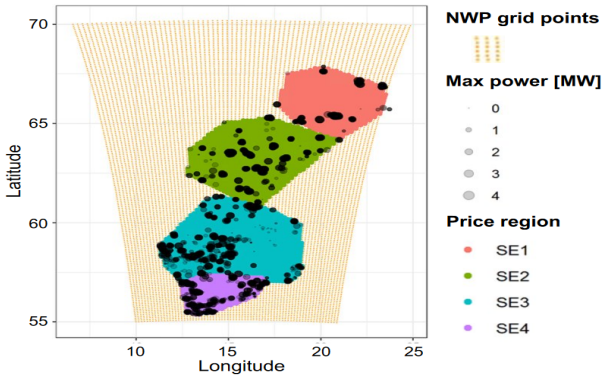


Fig. 4. Selection of NWP grid points by price region

The number of clusters was empirically determined (i.e. between 20 and 25, which appears high enough to define representative clusters and low enough for our dimensionality reduction goal). An example of the obtained clusters is shown in Fig. 5. During the competition, we only considered the final distribution of WEC as of 2001-12-31, even though the installation of WECs over time was likely to marginally impact the shape of the price region areas and the distribution of clusters.

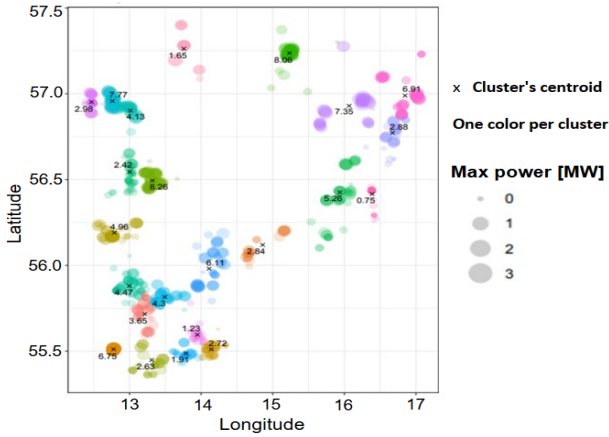


Fig. 5. Wind turbines clustering for the price region SE4

#### D. Feature Engineering

We combined three kinds of features used as inputs for the QRF, i.e. (1) features created thanks to the analog-based approach aimed at using past observed productions, (2) features related to wind power physics aimed at exploiting drivers of wind power production and (3) features related to operational constraints aimed at inferring the impact of additional power modulations by using proxies.

1) *Analog-Based Production Level*: A ranking of past weather patterns which are the most similar to the predicted weather conditions may be used to identify analog production levels. This is appealing in regional forecasting where synthetic regional indicators are needed. This idea of generating

analogues from a similarity ranking has proved to be efficient in wind forecasting [21] and solar forecasting [22], [23]. In this paper, we adapt an analog-based approach considering spatially distributed data [24] to estimate the future regional wind production. As a predictor variable for the analogy we focused on wind amplitude at hub height, but it could be relevant to add extra physical variables such as temperature or wind direction to capture wind production dependencies over weather. Then, we looked for similar spatial patterns over the price region areas defined in Section III-C. To do so, we applied a Principal Component Analysis on the analogy variable which generates a set of Principal Components (PC) representing the main patterns of wind amplitude. These PCs constitute a new coordinate system in which an Euclidean distance is used to measure similarity between situations and to rank them. Finally, the load factors corresponding to the  $n$  most analog weather patterns were averaged to provide an estimation of the future wind production (after empirical evaluation on the training set, we chose  $n = 200$ ).

2) *Features Related to Wind Power Physics*: To account for the wind energy conversion process, we defined new variables such as wind magnitude at hub height, air density and wind orientation. We also used some NWP variables directly (e.g. temperature, wind gust speed and relative humidity), since these factors could influence the efficiency of the conversion during extreme meteorological conditions.

- **Wind magnitude**: Wind speed is the most relevant variable for wind power generation. Wind speed is vertically interpolated (from NWP altitude  $z_{ref}$  to hub height  $z$ ) using the following wind shear model (or power law):

$$U_{wind}(z) = U_{wind}(z_{ref}) \cdot \left( \frac{z}{z_{ref}} \right)^\alpha \quad (3)$$

- $U_{wind}$  Horizontal wind speed,
- $\alpha$  Wind shear coefficient (or Hellman exponent).

For flat terrains, the one-seventh power law (i.e.  $\alpha = 1/7$ ) is usually used [25]. As a simplification, we considered a unique hub height (i.e.  $z = 100m$ ).

- **Wind direction**: Despite the rotation of the nacelle towards the wind direction to optimize wind production, we assumed that wind production is dependent on topographic characteristics and wake effects from neighboring WECs which are not represented by the NWP's model. As a result, the wind orientation was included in the forecasting model.
- **Air density**: Moist air density ( $\rho_{air}$ ) was estimated from temperature, pressure and relative humidity.
- **Wind power density**: We used the wind power density ( $WPD$ ) as an indicator of the maximum recoverable power [26].

$$WPD(z) = \frac{1}{2} \cdot \rho_{air} \cdot U_{wind}(z)^3 \quad (4)$$

- **Aggregated wind production**: For WECs the relation between wind speed and output power production is usually represented by S-shaped power curves of

wind turbines with a cut-in speed, a rated speed and a cut-out speed. As detailed WEC characteristics were not available, we considered a single wind power curve using typical values to process the wind magnitude. Then, we used the sum of installed capacity by cluster to calculate a weighted average in order to estimate aggregated wind production.

In addition to features corresponding to the time  $t$  to forecast, we also considered time-shifted features that had a positive impact on the model performance (i.e.  $t-1$  and  $t+1$ ).

Finally, in order to avoid adding too much noise to the most important information, we empirically differentiated features (e.g. wind power density was deemed as more relevant than relative humidity). For the main features, we kept the values for all key locations defined in Section III-C, whereas we selected a subset of the key locations for the other variables using the minimum Redundancy Maximum Relevance feature selection algorithm (mRMR) [27].

3) *Features Related to Operational Constraints:* Operational constraints can also impact wind power production (e.g. maintenance, curtailment). We had no information about such aspects for the competition, so we chose to add the hour of the day and day of week as inputs, considering that they can be proxies for such constraints (e.g. working hours for maintenance, very low demand in the middle of the night). Notably, [14] suggests that time-related information (e.g. time of day) is beneficial to forecasting performances.

### E. Regression Model

We selected the Quantile Regression Forests model (QRF) [28], [29] given its training speed, its robustness (i.e. good performance without advanced hyperparameter tuning), its capacity to account for non-linear interactions between features, its intrinsic ability to generate probabilistic forecasts for the desired quantiles and the proven track record of the random forests approach in other data science competitions (e.g. [14]). Also, since the QRF estimates variable importance, we chose to create multiple non-redundant features deemed as relevant for our forecasting application and leveraged the fact that, to some extent, the QRF can automatically select the most useful ones. Nevertheless, the feature selection was not only guided by this characteristic, but also by assessing the model performance on out-of-sample tests when incrementally adding new features.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

The method is implemented using R language [30]. An example of the forecasts obtained is displayed in Fig. 6. The different production regimes appear to be correctly anticipated in the four price regions, and forecasts are characterized by a high level of sharpness.

Forecasts are evaluated via a summary performance score based on the pinball loss function averaged over the price regions. The northern price region exhibits the highest forecasting error when considering normalized quantities. This

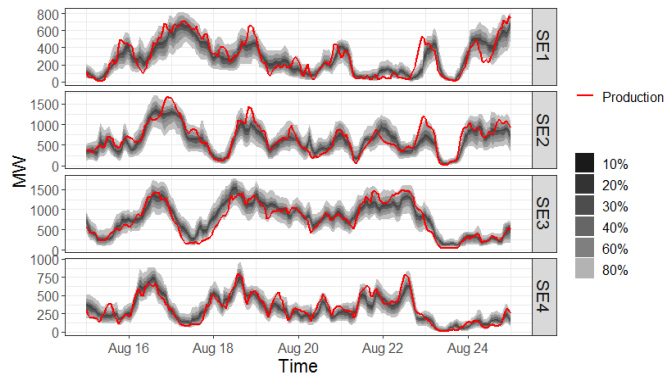


Fig. 6. Example of obtained forecasts for all four price regions

	Task.1	Task.2	Task.3	Task.4	Task.5	Task.6
<i>Submitted forecasts</i>	58.96	52.59	38.39	34.73	42.92	55.97
<i>Forecasts with latest model</i>	58.36	52.11	37.56	33.07	43.03	55.97
<i>Best scores from other teams</i>	57.17	53.84	42.53	34.31	46.18	58.54
<i>Mean scores from other teams</i>	104.01	86.95	60.91	59.02	65.42	70.57

Fig. 7. Forecasting scores based on the pinball loss function [9]

assessment is in line with [12] who consider that this phenomenon results from a low installed capacity, a complex terrain and icing losses. The last version developed during the competition was applied on the different tasks for backtesting: the results in Fig. 7 indicate that the continuous development of the approach leads to improvements compared to previous versions of the approach on Tasks 1 to 4, where pre-processing steps were less developed and not all hyperparameters were not adjusted.

## V. CONCLUSION

This paper proposes a probabilistic forecasting model for regional wind production at a day-ahead horizon, developed in the context of the EEM20 forecasting competition. The approach is based on a physics-oriented pre-processing of regional data, which captures analog weather patterns and reduces the dimensionality of wind production sites by means of a k-means clustering. The features obtained are augmented with a selection of variables including wind-power-driven quantities and temporal lags, and then fed into a Quantile Regression Forests model to obtain probabilistic forecasts for each price region. Our team incrementally improved its approach throughout the contest, with the result that we ranked in the top 2 teams for all 6 tasks, and ultimately won the forecasting competition.

Perspectives for future work include a further development of physics-oriented features (e.g. modelling the ageing effects of WECs), integration of spatio-temporal correlations, and the consideration of other analog predictors. Lastly, the approach based on an optimal combination of forecasting models proposed by [31], which ranked second in the competition, could benefit from the sharp predictions generated by the present model.

## ACKNOWLEDGMENT

The authors would like to thank Mikhail Skalyga, Priyanka Shinde and Evelin Blom for organizing of the EEM20 forecasting competition. We also thank MET Norway and Greenlytics AB for providing and archiving the data.

## REFERENCES

- [1] IRENA. (2020) Query Tool. [Online]. Available: <https://www.irena.org/Statistics/Download-Data>
- [2] ——. (2020) Innovative solutions for 100% renewable power in Sweden. [Online]. Available: <https://www.irena.org/publications/2020/Jan/Innovative-solutions-for-100-percent-renewable-power-in-Sweden>
- [3] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, Apr. 2014.
- [4] G. Juban, J., Fugon, L., Kariniotakis, "Probabilistic Short-Term Wind Power Forecasting Based on Kernel Density Estimators," *European Wind Conference and Exhibition*, pp. 1–11, 2007.
- [5] T. Thorarindottir, A. Løland, and A. Lenkoski, "Probabilistic forecasting of temporal trajectories of regional power production - part 1: Wind," 2019.
- [6] N. Siebert, "Development of methods for regional wind power forecasting," Ph.D. dissertation, École Nationale Supérieure des Mines de Paris, Mar. 2008.
- [7] E. Nuno, M. Koivisto, N. Cutululis, and P. Sorensen, "Simulation of regional day-ahead PV power forecast scenarios," in *2017 IEEE Manchester PowerTech, Powertech 2017*. IEEE, jun 2017, pp. 1–6.
- [8] "Smart4RES - data science for renewable energy prediction." [Online]. Available: <https://www.smart4res.eu>
- [9] "Forecasting Competition." [Online]. Available: <https://eem20.eu/forecasting-competition/>
- [10] "EEM20 Wind Power Forecasting Competition - Evaluation and Lessons Learned." [Online]. Available: <https://www.youtube.com/watch?v=pxkg48BBFYM>
- [11] "EEM20 Forecasting Competition: Probabilistic forecast of aggregated wind production with hourly resolution." [Online]. Available: <https://www.youtube.com/watch?v=n0m3SI8Zwtk>
- [12] J. Olauson and M. Bergkvist, "Modelling the Swedish wind power production using MERRA reanalysis data," *Renewable Energy*, vol. 76, pp. 717–725, Apr. 2015.
- [13] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, and L. Badesa, "Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications," *Applied Energy*, vol. Volume 242, pp. 1396–1406, May 2019.
- [14] G. I. Nagy, G. Barta, S. Kazi, G. Borbély, and G. Simon, "GEF-Com2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1087–1093, Jul. 2016.
- [15] J. Juban, L. Fugon, and G. Kariniotakis, "Uncertainty estimation of wind power forecasts: Comparison of Probabilistic Modelling Approaches," in *European Wind Energy Conference & Exhibition EWEC 2008*. EWEC, Mar. 2008, p. 10.
- [16] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The Quickhull algorithm for convex hulls," *Acm Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996.
- [17] P. Indarjo, "Using Weighted K-Means Clustering to Determine Distribution Centres Locations," Jan. 2020. [Online]. Available: <https://towardsdatascience.com/using-weighted-k-means-clustering-to-determine-distribution-centres-locations-2567646fc31d>
- [18] M. Ackerman, S. Ben-David, S. Brânzei, and D. Loker, "Weighted clustering," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, ser. AAAI'12. Toronto, Ontario, Canada: AAAI Press, Jul. 2012, pp. 858–863.
- [19] R. M. Rustamov, "Barycentric Coordinates on Surfaces," *Computer Graphics Forum*, vol. 29, no. 5, pp. 1507–1516, 2010.
- [20] V. Hegde, T. S. Aswathi, and R. Sidharth, "Student residential distance calculation using Haversine formulation and visualization through GoogleMap for admission analysis," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2016, pp. 1–5, iSSN: 2473-943X.
- [21] S. Alessandrini, L. Delle Monache, S. Sperati, and J. N. Nissen, "A Novel Application of an Analog Ensemble for Short-term Wind Power Forecasting," *Renewable Energy*, vol. 76, pp. 768–781, Apr. 2015.
- [22] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An Analog Ensemble for Short-term Probabilistic Solar Power Forecast," *Applied Energy*, vol. 157, pp. 95–110, Nov. 2015.
- [23] T. Carriere, C. Vernay, S. Pitaval, and G. Kariniotakis, "A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames," *IEEE Transactions on Smart Grid*, pp. 1–1, 2019.
- [24] K. Bellinguer, R. Girard, G. Bontron, and G. Kariniotakis, "Short-term Forecasting of Photovoltaic Generation based on Conditioned Learning of Geopotential Fields," in *55th International Universities Power Engineering Conference (UPEC)*, 2020, p. 6.
- [25] M. Ray, A. Rogers, and J. McGowan, "Analysis of wind shear models and trends in different terrain," Jan. 2006.
- [26] A. Kalmikov, "Chapter 2 - Wind Power Fundamentals," in *Wind Energy Engineering*, T. M. Letcher, Ed. Academic Press, Jan. 2017, pp. 17–24.
- [27] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [28] N. Meinshausen, "Quantile Regression Forests," *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 983–999, 2006.
- [29] ———, "quantregForest: Quantile Regression Forests," <https://CRAN.R-project.org/package=quantregForest>, 2017.
- [30] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: <https://www.R-project.org/>
- [31] J. Browell, C. Gilbert, R. Tawn, and L. May, "Quantile Combination for the EEM20 Wind Power Forecasting Competition," *International Conference on the European Energy Market, EEM*, vol. 2020, 2020.