NORDIC MACHINE INTELLIGENCE

WindAI: Wind power forecasting in Norway

Kristian Authen[1], Signe Riemer-Sørensen[2], Sabry Razick[3], and Kushtrim Visoka[4,*]

[1]Statnett
[2]SINTEF
[3]University of Oslo
[4]Norwegian Artificial Intelligence Research Consortium
[*]E-mail any correspondence to: kushtrim.visoka@nora.ai

May 28, 2025

Abstract

Keywords: artificial intelligence; machine learning; deep learning; robotics

Introduction

Operators of Nordic transmission systems like Statnett need to provide wind power forecasts that span two days ahead. These forecasts are used by the day ahead power exchange through the flow-based market coupling. The flow-based market coupling ensures that power trading does not overload critical network elements in the power grid.

In this hackathon, we will create such wind power forecasts for each of the Norwegian bidding zones. The forecasts shall contain the wind power production per hour for the entire day, 2-days ahead. That is, the forecast for '2025-03-15' shall contain the wind power between '2025-03-17 00:00' and '2025-03-17 23:00' CET.

The following datasets are included in the hackathon (described in Section):

- Weather forecasts (MET) - MEPS weather forecast containing 15 forecast simulations for each wind park.
- Weather observed (MET) - Now-cast weather for each wind park.
- Wind power observed (Statnett) - Observed wind power for each of the Norwegian bidding areas.
- Wind power metadata (Statnett) - Metadata for mapping wind park weather to bidding area.

The observed weather and wind power data will be available up until the issue time of the weather forecast, and the weather forecast will span 65 hours ahead.

Evaluation setup and criteria

Provided data

The available data are provided as a bundle from the competition website. The weather and power data spans from 2020-01-01 to 2025-03-31. Updated data will be provided before the evaluation process begins. All timestamps are given as UTC. In addition, the competitors get a Jupyter notebook showing how to load the data and explaining the content.

MET forecast

This dataset comes from the MEPS weather forecast provided by the Norwegian Meteorological Institute[1]. The dataset contains individual values for all the different wind parks in Norway. The dataset has 15 different forecast members (individual simulations) with slightly different perturbations in the weather simulation. The spread between these members indicate uncertainty in the weather forecast and potential outcomes. The issue time "time_ref" is the time when the forecast become available. The variables are listed in Table 1.

MET nowcast

The now-cast weather is the weather forecast zero hours into the future. This is the closest we get to the observed weather for locations without dedicated weather measuring stations. This data is not available ahead of time. The variables are listed in Table 1.

Statnett Wind Power Data

The wind power data contains the sum of all wind power production within each bidding area in MW. As shown in Figure 1, there are five bidding areas in Norway named NO1-NO5. Bidding area NO5 does not contain any wind parks and is omitted. Keep in mind that not all wind parks are in service at the beginning of the dataset. The variables are listed in Table 1.

[1]https://www.met.no/en/free-meteorological-data

Name	Description
MET Forecast	
sid	Name of wind park
time_ref	Time when forecast was generated [yyyy-mm-dd hh:mm:ss+hh:mm]
time	Time when forecast is valid [yyyy-mm-dd hh:mm:ss+hh:mm]
lt	Lead time in hours (time - time_ref) [hours]
ws10m_x	Wind speed at 10m for ensemble member x [m/s]
wd10m_x	Wind direction at 10m for ensemble member x [m/s]
t2m_x	Temperature at 2m for ensemble member x [degrees Celcius]
rh2m_x	Relative humidity at 2m for ensemble member x [%]
mslp_x	Mean sea level pressure for ensemble member x [hPa]
g10m_x	Gust at 10m for ensemble member x [m/s]
MET Now-casting	
air_temperature_2m	Air temperature 2m above ground [degrees Celcius]
air_pressure_at_sea_level	Air pressure at sea level [hPa]
relative_humidity_2m	Relative humidity at 2m above ground [%]
precipitation_amount	Precipitation [mm]
wind_speed_10m	Wind speed at 10m above ground [m/s]
wind_direction_10m	Dominating wind direction at 10m above ground [degrees]
wind park	Which wind park the now-casting relates to
Statnett Wind Power Data	
ELSPOT NOx	Wind power production for the hour where x is 1-4 [MW]
Statnett Wind Park Metadata	
bidding_area	Name of Norwegian bidding area. The challenge only considers NO1-4
substation_name	Name of wind park
operating_power_max	Maximum operating power [MW]
prod_start_new	Date of initial operation of the wind park [yyyy-mm-dd hh:mm:ss]
eic_code	Energy Identification Code (for linking parks to NordPool REMIT UMM)

Table 1: Description data variables.

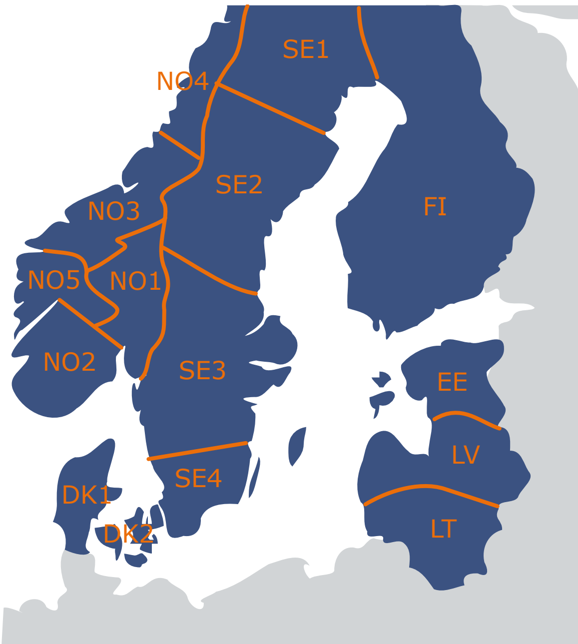

Figure 1: The Norwegian power market and neighbouring zones. Figure from [1].

Statnett Wind Park Metadata

The metadata table is used to map the wind parks to bidding areas. It also contains the production start of each of the wind parks in addition to the maximum power output for each park. The variables are listed in Table 1.

Open data

In addition, all public data can be used such as theme maps from The Norwegian Water Resources and Energy Directorate[2] or climate data from e.g. https://disc.gsfc.nasa.gov/ or https://globalwindatlas.info/en/. Climate data can say something about variability of local conditions. However, climate changes may lead to future larger variability. A useful list of open data can be found here https://iea-wind.org/task43/task-43-open-data/

NordPool REMIT UMM contains unavailability messages for planned maintenance and outages for windparks. These can be found here https://umm.nordpoolgroup.com.

Note that some companies (including Statnett) provide predictions for wind power production. The teams are free to explore such data, but should note that it most cases, the quality of the data is unknown and the predictions are only released retrospectively, so they may not be available at prediction time.

Simple example

The participants are provided with a simple example showing how the wind power generation in bidding area NO3 can be modelled.

First the data is pre-processed. The data are filtered to include only wind parks located in NO3 and the wind power generation corresponding to NO3. For input weather data, we use the mean observed weather data. This is computed by selecting the weather now-cast data for the wind parks in the bidding zone, dropping the 'wind park' column, and grouping the data by time to compute the mean values. Finally, we filter the combined dataset to include only the periods where all wind parks are operational. This is done by selecting data from the maximum production start date of the wind parks and dropping any rows with missing values.

To evaluate the performance of our model, we split the dataset into training and testing sets. The training set includes data up to January 1, 2025, while the testing set includes data from January 1, 2025, onwards.

Figure 2 left panel shows the scatter plot of produced power in NO3 as a function of wind speed.

The features used for training the model include wind speed at 10m, air pressure at sea level, air temperature at 2m, and relative humidity at 2m. The target variable is the wind power generation in NO3.

We create a linear regression model with a spline transformer on the input data. The spline transformer allows for capturing non-linear relationships between the features and the target variable.

Figure 2 right panel visualize the relationship between wind speed at 10m and wind power for observed and predicted values. The test RMSE with observed weather is 220.63 MW.

To predict ahead in time, we use the weather forecast data as weather input. We select the ensemble member 0, which is considered the best estimate, from the weather forecast. The weather forecast data is filtered to include only the relevant time period and wind parks in the bid zone. The predictions are shown in Figure 3, and the RMSE with forecasted weather is 280.77 MW.

Tasks

We recommend an exploratory approach when solving the suggested tasks and encourage innovation when developing solutions.

Provide a wind power forecast for each bidding zone in Norway on an hourly resolution for the extent of the MET MEPS weather forecast (1-61 hours). The model may use the recent power production up until the first timestamp of the weather forecast.

Additional tasks:

1. How can you demonstrate robustness of the model including robustness to changes in weather patterns/climate changes and installed wind power capacity?
2. How do you handle trustworthiness and explainability?
3. Can you provide reliable uncertainty predictions?

[2] https://www.nve.no/kart/

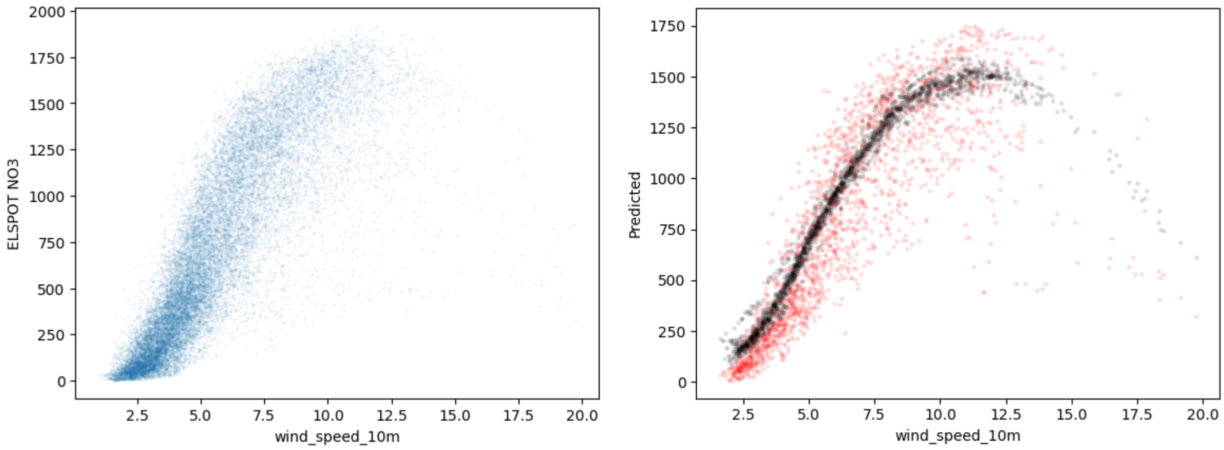

Figure 2: Left: Scatter plot of produced power in NO3 as a function of wind speed. Right: Predicted power as a function of wind speed for the test data.

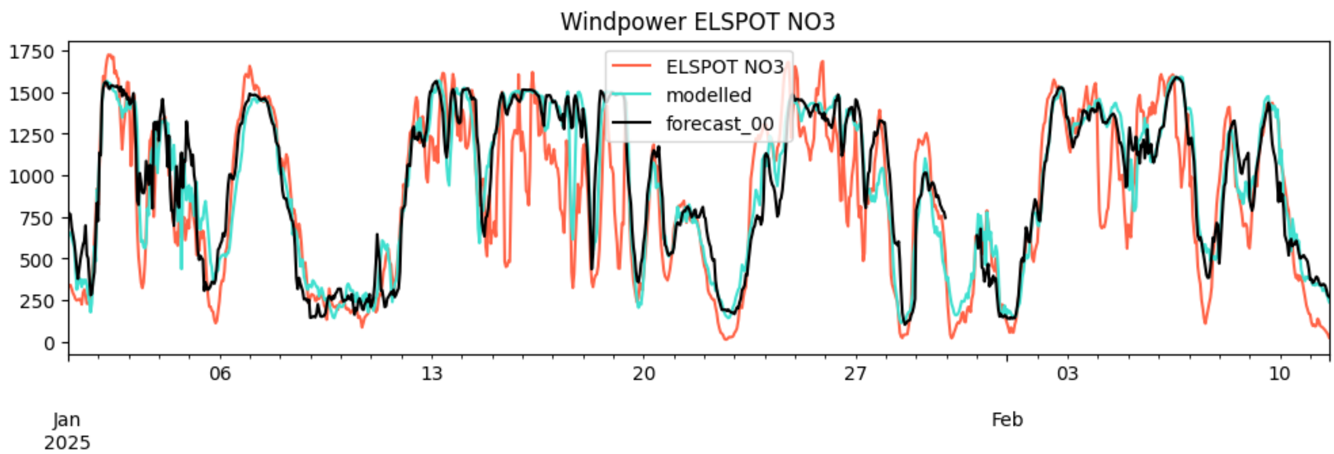

Figure 3: Forecasting from weather data. Red line is true data, cyan is the model predictions given the actual weather, and black is the forecasting using the weather forecast ahead in time.

4. How lightweight can the model be in terms of computational resources?

Submission and evaluation criteria

The submission will mimic a production setting. For the two weeks following the deadline [Sep 22-26 and Sep 29-Oct 3], all teams will deliver a daily forecast on workdays (10 in total, one can be missed).

The forecast should be uploaded to the allocated disk before midnight CET containing predicted wind power for NO1-NO4 for the entire day (24 hours) two days ahead. Input data (weather, recent power production) will be updated on a shared container around 14:00 CET every day. Submissions should be uploaded as csv-files following the convention in example provided with the data (see Table 2).

The forecasts will be scored using the root mean square error (RMSE) metric. The predictions will be shown on a public dashboard (RMSE and plots). After the 10 submissions, the top 10 performing teams on accuracy will be evaluated on usefulness for Statnett according to the following criteria:

1. Accuracy [65%]. Firstly, used to rank all contributions, after which the other factors will also contribute.

2. Trustworthiness and explainability [20%]. Has the team understood the task and provide meaningful models? Does the model provide uncertainty on the predictions? Does the team provide a benchmark against other models? Does the team provide explanations for the model predictions? Is it simple to understand how the data influence the model? Is there an obvious simpler model with similar performance?

3. Implementation and presentation [10%]. Is the model easy to use? Can it easily be retrained? Does the team explain the overall concepts in a good way? Is it feasible for Statnett to implement the model in their daily operations?

4. Robustness to changes [5%]. Is the model setup robust to changes in installed wind power capacity, climate changes etc.).

5. Other criteria that can be used to break a tie are computational requirements such as prediction time of the model and required computational to train and run the model.

Summary

This paper presents the WindAI data challenge.

Acknowledgements

Thanks to Statnett for providing the challenge and data. The WindAI competition is a collaboration between multiple initiatives, some of which are funded by the Research Council of Norway (RCN). In particular, we mention Center for Research Based Innovation NorwAI (RCN 309834), FORSKNINGSINFRA Norwegian Artificial Intelligence Cloud (RCN 322336), INPART Collaborative Research Based Education for Optimized Performance of Wind Farms (RCN 337301), UTFORSK Global Collaborative Research Based Education for Enhancing Performance of Wind Farms (EduWind) (UTF-2021/10169), NORA Norwegian AI Research School (INSERT).

Time	ELSPOT NO1	ELSPOT NO2	ELSPOT NO3	ELSPOT NO4
2025-03-25 00:00:00+01:00	0	0	946.317551	0
2025-03-25 01:00:00+01:00	0	0	761.134424	0
2025-03-25 02:00:00+01:00	0	0	509.700472	0
...				
2025-03-25 22:00:00+01:00	0	0	605.197394	0
2025-03-25 23:00:00+01:00	0	0	778.815429	0

Table 2: Specification of submission format for an example date.

References

1. Wikipedia. File:Hou710 ElectricityPriceArea.svg. en. 2015. Available from: `https://commons.wikimedia.org/wiki/File:Hou710_ElectricityPriceArea.svg` [Accessed on: 2022 Dec 30]