https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Minimum Open Data Subset for Wind Power Prediction

Elizabeth von Zuben[1] and Kristen R. Schell[1]

[1]Department of Mechanical & Aerospace Engineering, Carleton University, 1125 Colonel By Dr., Ottawa, ON, Canada K1S 5B6

Correspondence: Kristen R. Schell (kristen.schell@carleton.ca)

Abstract. Accurate wind power prediction is required for grid integration of renewables, minimizing curtailment of renewable energy, and performing resource assessments. Prior research has explored the use of numerical weather prediction, reanalysis datasets, and observational data in power prediction and resource assessment applications. Observational data is spatially limited and often proprietary. Reanalysis datasets are available globally, but have a large spatial resolution and therefore do not capture the effects of complex geography well. Numerical weather prediction simulations allow for high spatial resolution flow models, but require significant processing resources and computational time. This work combines historical wind power production data, observational data, MERRA-2 reanalysis, and WRF model data at three wind farms in Ontario, Canada to determine the optimal data source, combination of data sources, and variables for prediction of wind power using a random forests model. Results show that a model combining select data from all three data sources, including a combination of wind speed, time, and other weather variables, improves predictive performance by up to 57% over the benchmark power curve model. Analysis of feature importance shows that aggregating wind speed allows the model to make better use of additional weather features. The minimum subset of input data for the best performing model, which achieves a mean absolute error (MAE) of 0.071 across all sites, consists of averaged wind speed, temperature, wind direction, pressure, air density, and time variables (hour, day and month).

1 Introduction

Development of wind energy is becoming increasingly important in the decarbonization of the electricity sector and mitigation of climate change. Successful grid scale wind energy projects require accurate wind energy assessments during the development phase and accurate wind power forecasts during operation to optimize wind turbine layouts, aid in transmission planning, and maximize integration of wind power into the electricity grid. The intermittent nature of wind energy can introduce challenges to grid operators for balancing electricity supply and demand, often leading to curtailment or the reduction in renewable generation below generating capability. The electricity grid operator in Ontario, Canada reports the percentage of curtailed renewable generation in year-end reports - reported curtailment ranges from a minimum of 5.3% reported in 2014 to a maximum of 26% in 2017 (Independent Electricity System Operator, 2021). When curtailment is estimated by comparing recorded forecasts to actual output, curtailment is much higher at 13% in 2014 and 30% in 2017. Accurate wind energy predictions are required to minimize wasting available energy at wind farms.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

In order to make such assessments, meteorological towers are typically installed with sensors close to hub height to characterize local wind conditions. Historical weather data combined with manufacturer provided data is used to estimate the annual energy production (Milan et al., 2010). A minimum of one year of data collection is typically required to capture seasonal variability (Brower et al., 2012). This data is temporally limited and typically not accessible to the public. Alternative datasets, including weather stations, reanalysis data, and numerical weather prediction (NWP) may be able to make up for the limitations of on site meteorological towers.

Large amounts of data are publicly available from weather stations, which often contain multiple years to decades of historical weather data. However, weather stations are geographically sparse outside of populated areas, wind speeds are typically only measured at a height of 10 meters (m), and data is often only available at an hourly resolution (Environment and Climate Change Canada, 2011). There may also be periods with significant amounts of missing data due to weather events, communication errors, and sensor failures.

Reanalysis data, which assimilates weather station data with global atmospheric simulations, is increasingly being used to simulate wind power generation (Gruber et al., 2022; Staffell and Pfenninger, 2016; González-Aparicio et al., 2017; Gruber et al., 2019) due to global coverage, extensive historical records and accessibility. However, there is a lack of validation for reanalysis datasets outside of Europe, the United States, and China (Gruber et al., 2019) and studies have found significant spatial biases when compared to observed wind speeds or wind power. Reanalysis data also has coarse spatial resolution that is unable to capture the effects of complex topography and turbine wake effects.

High resolution NWP modelling is often used as a tool in wind speed and power forecasting. This allows for accurate flow field simulations (Prósper et al., 2019) as well as the simulation of turbulence and turbine wake effects (Fitch et al., 2012). Resolution and physics parameterizations can be customized by application and geographic location. However, NWP computations require considerable computer memory and processing resources, can be time consuming to compute, and requires quality observational data for initial and boundary conditions of simulations.

1.1 Literature review

In Staffell and Pfenninger (2016), MERRA-2 reanalysis data was used to simulate wind power in various European countries using extrapolated MERRA-2 wind speeds and manufacturer provided power curves and found that the dataset led to both under and overestimates of wind power, particularly in regions with complex orography. This finding was reinforced in Gruber et al. (2022), which aimed to validate both MERRA-2 and ERA5 reanalyses across various countries (US, Brazil, New Zealand, South Africa) and found a large range of wind power bias depending on location. Various studies that have used reanalysis datasets for wind power prediction have noted significant bias in topographically complex areas and the inability to capture wind speed variations due to the coarse spatial resolution of reanalysis (Gruber et al., 2019; Morales-Ruvalcaba et al., 2020). Limited validation of reanalysis datasets has been completed within Canada, outside of Dolter and Rivers (2018) where MERRA data was used to estimate decarbonization pathway costs, with a validation of the dataset against weather stations provided in their work's supplementary material.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 1. Input variables, data sources, and power prediction methods used in previous wind power prediction papers reviewed in this work. Other* variables include geometric properties determined from turbine spacing (Staid, Yan), lapse rate and rotor equivalent wind speed (Sasser), turbulence intensity (Pang, Özen, Pombo), wind sheer (Pang), and other NWP variables (Özen). Özen, Shi, and Pombo differ from other included studies in that they are forecasting models, however are included to show the use of extended variables.

	Variables									Data source				Method		
Paper	wind speed	wind direction	temperature	air density	pressure	humidity	radiation	time variables	Other*	Reanalysis	WRF	Met data	SCADA	Power curve	Machine learning	Other
This paper	•	•	•	•	•	•	•	•		•	•	•		•	•	
Staffel et al., 2016 Staffell and Pfenninger (2016)	•									•				•		
Gonzalez-Aparicio et al., 2017 González-Aparicio et al. (2017)	•									•				•		
Dolter et al., 2018Dolter and Rivers (2018)	•									•				•		
Gruber et al., 2019 Gruber et al. (2019)	•									•				•		
Morales-Ruvalcaba et al., 2020 Morales-Ruvalcaba et al. (2020)	•									•				•		
Gruber et al., 2021 Gruber et al. (2022)	•									•				•		
Hayes et al., 2021 Hayes et al. (2021)	•									•				•		
Lee et al., 2017 Lee and Lundquist (2017)	•										•			•		
Giannaros et al., 2017 Giannaros et al. (2017)	•										•					•
Tomaszewaki et al., 2020 Tomaszewski and Lundquist (2020)	•										•			•		
Staid et al., 2018 Staid et al. (2018)	•	•	•						•			•		•	•	•
Bilal et al., 2018 Bilal et al. (2018)	•	•	•			•	•					•			•	
Sasser et al., 2021 Sasser (2021)	•								•			•			•	
Pang et al., 2021 Pang et al. (2021)	•	•							•			•			•	
Yan et al., 2019 Yan and Ouyang (2019)	•	•							•				•		•	
- Özen et al., 2021 Özen et al. (2021)	•	•	•	•	•	•	•		•		•			•	•	
- Shi et al., 2018 Shi et al. (2018)	•	•	•		•	•						•			•	
- Pombo et al., 2021 Pombo et al. (2021)	•	•						•	•			•	•		•	

NWP models such as the WRF (Weather Research and Forecasting) model are able to capture the effects of complex terrain and wake effects from wind turbines at higher resolutions (Skamarock et al., 2019). The WRF model has been used for wind energy assessments (Giannaros et al., 2017; Carvalho et al., 2012) and wind power forecasting (Prósper et al., 2019; Tomaszewski and Lundquist, 2020) and has demonstrated the ability to simulate spatial and temporal variations in wind speeds and power.

Machine learning methods are commonly used in renewable energy forecasting (Sweeney et al., 2020). The random forests (RF) method has been found to perform similarly to or outperform other machine learning algorithms for wind power prediction due to the ability to handle noisy data and outliers in measured meteorological data (Singh et al., 2021; Staid et al., 2018; Shi et al., 2018; Pombo et al., 2021; Arrieta-Prieto and Schell, 2024).

Previous studies have found that combinations of datasets and forecasting methods can improve wind power and wind speed predictions. In Özen et al. (2021) and Wang et al. (2021), machine learning models were use to improve WRF wind power and wind speed forecasts. Previous works, however, have not analyzed the effectiveness of combining all three aforementioned data sources using machine learning methods. This is the main research question posed by this study.

Some previous works have explored the impact of environmental variables other than wind speed on wind power output. Shi et al. included temperature, pressure, and humidity, however these were filtered out by the feature selection process (Shi et al., 2018). Pang et al. explored the impact of other environmental variables using Shapley values and found that turbulence intensity, air density, and wind direction moderately impacted power output (Pang et al., 2021). Özen et al. included multiple WRF output variables as input to a power prediction model, however the importance of variables other than wind speed was unclear due to the use of multiple wind speed variables at various heights (Özen et al., 2021). This paper further demonstrates that wind speed measurements alone are not sufficient for wind power predictions.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

1.2 Contributions

This study advances the state-of-the-art through the assessment of the importance of key input variables, derived from open access datasets favored by the field, to the improvement of wind power prediction. Here, weather station, MERRA-2 reanalysis, and WRF data are assessed for their ability to predict normalized wind power using the random forest machine learning algorithm. A new machine learning interpretation method, Model Class Reliance (Smith et al., 2020), is used to determine an optimal combination of variables from all three data sources as well as to quantify how much wind power prediction models rely on different meteorological variables (wind speed, wind direction, temperature, pressure, air density, humidity, and irradiance) and time variables (month, day, and hour).

2 Data

2.1 Site Description

Three wind farms in Ontario were chosen based on power output data availability and proximity to public meteorological (met) stations. Chosen wind farms, surrounding wind farms, and met stations used in this analysis are indicated in Figure 1. All three sites are located on land primarily classified as cropland in Ontario, Canada. Ontario experiences a four-season continental climate that ranges between hot and humid summers to cold winters with moderate snowfall. Location, commission year, capacity, turbine model, and turbine hub heights for each site are outlined in Table 2.

Erieau (WF1) is located on the western coast of Lake Erie in an area that is highly saturated by other wind farms. Elevation at this site ranges between approximately 170 and 210 meters above sea level. This site is 7 – 19 km away from three meteorological stations.

East Lake (WF2) is located on the north-east coast of the significantly smaller Lake St. Clair, approximately 40 km away from WF1. WF2 is centered between one large and one smaller wind farm, both within 10 km of WF2. Elevation ranges between 170 and 180 meters. The closest three meteorological stations to WF2 are 33 – 49 km away from the site and are the same three stations surrounding WF1.

Wolfe Island (WF3) is located on an island in northern Lake Ontario at the entrance of the St Lawrence River, where elevation ranges between 70 and 90 meters. WF3 contains the greatest variation in land cover, with mostly cropland, some barren and some temperate or deciduous forest. WF3 is 15 km away from two meteorological stations.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

WIND ENERGY SCIENCE DISCUSSIONS

Table 2. Wind farm specifications

	WF1	WF2	WF3
Location (lat, lon)	42.3, -82	42.5, -82.4	44.2, -76.4
Commissioned (Year)	2013	2013	2009
Capacity (MW)	99	99	197.8
Turbine Model	Vestas V90/1800	Vestas V90/1800	Siemens SWT-2.3-101
Hub Height (m)	80	80	80

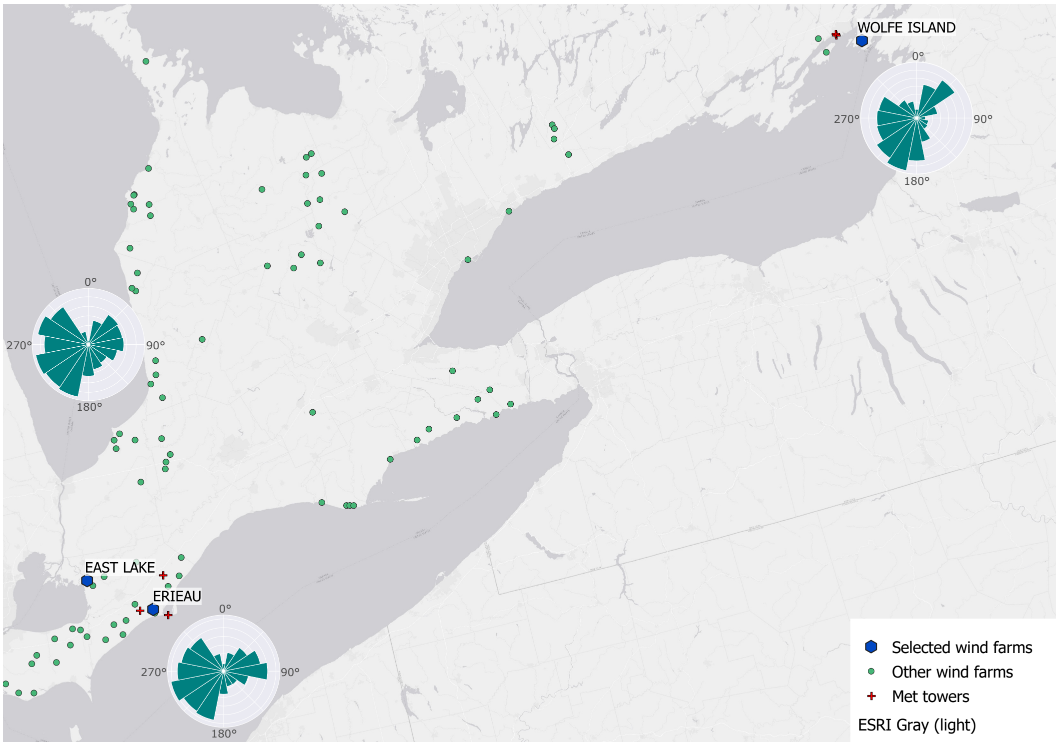


Figure 1. Locations of selected wind farms, surrounding wind farms, and met stations used in analysis.

2.2 Weather station data

Historical climate data was obtained from meteorological stations operated and maintained by Environment and Climate Change Canada (ECCC) (Environment and Climate Change Canada, 2011). Data at the stations utilized in this study was available at an hourly resolution. Wind speeds are measured at a height of approximately 10 m and all other climate variables are measured at a height of 2 m (Environment and Climate Change Canada, 2022). Wind speed, wind direction, pressure, temperature, and relative humidity were collected from five met stations, three of which were located near WF1 and WF2 (19, 7, 9.7 km and 42, 33, 48.8 km away respectively), and two of which were located near WF3 (14.7 and 15.2 km away).

Linear interpolation was used to fill in missing data points if the number of consecutive missing data points was below a threshold of three. If consecutive missing data points were greater than this threshold, rows with missing data were removed. 4% of data points were excluded from WF1 and WF2, and 3% were excluded from WF3.

Wind speed was interpolated to hub height using the logarithmic wind speed profile (Equation 1), where v_1 and v_2 are wind speeds at measured and extrapolated height, h_1 and h_2 are measurement and extrapolation heights, and z_0 is the roughness length as obtained via the Global Wind Atlas (noa).

$$v_2 = v_1 \frac{log(h_2/z_0)}{log(h_1/z_0)} \tag{1}$$

2.3 Reanalysis data

Reanalysis data aims to assimilate historical meteorological data using a single global simulation. MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications), developed by NASA's Global Modeling and Assimilation Office, has a spatial resolution of approximately 50 km in the latitudinal direction at the sites of interest (0.5 °× 0.625° latitude x longitude) and a temporal resolution of 1 hour (Gelaro et al., 2017; Molod et al., 2015). The Renewables.Ninja (RN) project aims to provide access to hourly simulations for wind and solar power using data from MERRA reanalysis (Staffell and Pfenninger, 2016; Pfenninger and Staffell, 2016). RN uses a Virtual Wind Farm and MERRA-2 data to spatially interpolate wind speeds to specific coordinates, extrapolates wind speeds to hub height using a logarithmic wind profile, and then converts speeds to power output using manufacturer provided power curves. Hourly wind speed, temperature, air density, and irradiance data was obtained from RN.

2.4 Wind farm production data

Generator output and capability data for transmission connected generation facilities greater than 20 megawatts (MW) are made available by the IESO (Independent Electricity System Operator). Hourly output and capability data is available from 2010 up to the current date (Independent Electricity System Operator, 2022).

Wind farms were chosen based on those within the reported data that had minimum estimated curtailment. While the IESO does not report curtailment of individual power generating stations, reported data includes both forecast and actual power output for all variable generating facilities. IESO forecasts and output were used to estimate curtailment. Chosen wind farms had an estimated curtailment of 9% for WF1 and WF2, and 11% for WF3 between 2014 and 2021, compared to an overall average of 20% over the same time period for all IESO connected wind farms within the province. WF1 and WF2 have production data available starting late May 2013, and WF3 has production data available starting January of 2010.

IESO data was cross-referenced with the Natural Resources Canada wind turbine database for all wind turbines installed in Canada based on open access data (Natural Resources Canada, 2021). The database contains location data, including latitude and longitude, and technology used, including turbine model, capacity and hub height.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Reported IESO output data was normalized according to total wind farm capacity allowing for comparison between wind farm sites (measured power output at each time step/total site capacity). While previous studies have found that a generalized-logit transformation of the normalized data (Pinson, 2012; Arrieta-Prieto and Schell, 2022) stabilizes variance and improves predictive accuracy, such an improvement was tested here, but found not to apply to the wind farms studied here.

145 ### 2.5 WRF model

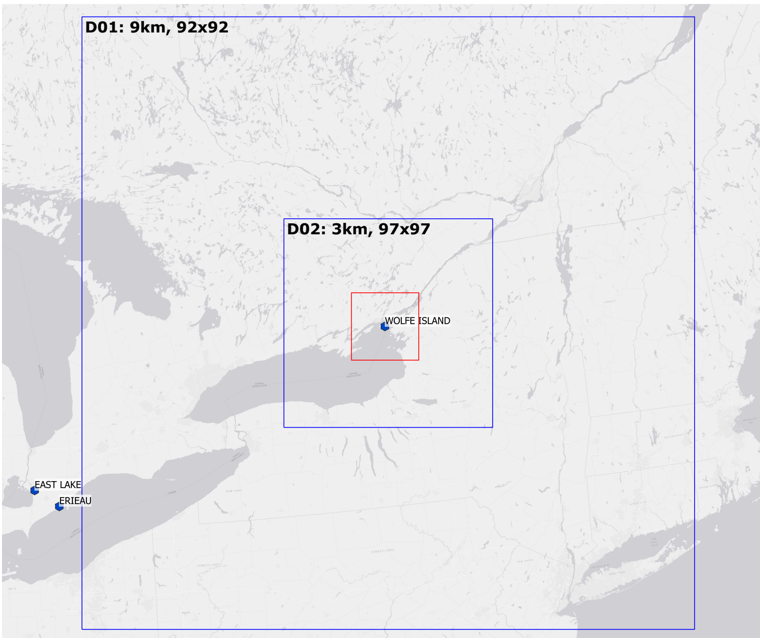


Figure 2. Wind farm locations and WRF domains for WF3 with horizontal grid indicated - inner 1 km domain (red) not used in analysis.

An NWP model was used to provide higher resolution wind speed and climate data than can be obtained from reanalysis data. The open access Weather Research and Forecasting (WRF) model, version 4.2.2, was used (Skamarock et al., 2019). WRF preprocessing and postprocessing was assisted by GIS4WRF, an open access plugin within QGIS (Meyer and Riechert, 2019).

Model configuration and physics options are outlined in Table 3 with configuration modelled after Tomaszewski et al.
150 (Tomaszewski and Lundquist, 2020). Three nested domains were initially tested with 9, 3 and 1 km resolutions. Wind speed output at the 1 km resolution was found to give negligible improvement over wind speeds from the 3 km domain when compared to observed meteorological data. Thus, the 1 km domain was removed in the interest of computational efficiency. The WRF domain configuration for WF3 is illustrated in Figure 2.

The wind farm parameterization presented in Fitch et al. (2012), which represents the impact of the turbulence from individ-
155 ual wind turbines, was also tested at the 1 km resolution. Including the wake model did not change the results significantly at the chosen resolution and was not used in the final model. National Centers for Environmental Prediction (NCEP) 0.25 degree

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

analysis data was used for initial and boundary conditions (National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, 2015).

The WRF model was computed for July and October of 2015 and January and April of 2016 to capture one month for each season for a total of 4 months of data. One WRF model was computed centered over WF1 and WF2, and a second model was computed centered over WF3. Wind speed, wind direction, temperature, pressure, and relative humidity were extracted from the WRF model from each unique cell that contained a wind turbine. Data were extracted from approximately 15 unique cells for each wind farm. WRF data from multiple cells were included in initial testing to determine whether multiple locations provided additional information to the model. As this was found to provide minimal improvement at the chosen resolution, WRF data from the central location of each wind farm only was used as input data to final models.

Table 3. Final WRF configuration

Land surface	unified Noah land-surface model Ek et al. (2003)
Surface layer	Revised MM5 Monin-Obukhov scheme
PBL	MYNN 2.5 level Nakanishi and Niino (2006)
Shortwave radiation scheme	Dudhia scheme Dudhia (1989)
Longwave radiation scheme	RRTM scheme Mlawer et al. (1997)
Microphysics	WSM 5-class scheme Hong et al. (2004)
Vertical levels	58
Time step (s)	30
Domains	2 nested domains (9 km, 3 km)

2.6 Other data

Time variables (month, day, and hour) were included in all models to capture diurnal and seasonal trends.

Power curve data was acquired from the and verified against wind farm and manufacturer documentation.

Wind farm power output data, reanalysis data, and observed met data were obtained for all hours computed in the WRF model for a total of 4 months of data from each data source, capturing each season. Summary statistics are outlined in Table 4.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

WIND ENERGY SCIENCE DISCUSSIONS

Table 4. Input variable summary statistics by location (WF1, WF2 or WF3) and data source (met, WRF or RN). Met stations for WF1 are the same met stations for WF2 (met 1, met 2, met 3). Average of all WRF locations is displayed.

	Location	WRF1				WF2				WF3			
Variable	Data source		mean	min	max		mean	min	max		mean	min	max
Power (MW)	IESO		34.6	0.0	97.0		36.3	0.0	97.0		61.5	0.0	193.0
Temperature (°C)	met	met 1	8.7	-16.3	30.6	met 1	8.7	-16.3	30.6	met 4	7.4	-18.9	28.1
	met	met 2	9.0	-14.7	30.6	met 2	9.0	-14.7	30.6	met 5	7.4	-19.2	28.2
	met	met 3	9.6	-11.9	29.1	met 3	9.6	-11.9	29.1				
	WRF		8.4	-13.0	28.6		8.4	-10.9	29.7		7.2	-16.4	26.9
	RN		9.6	-10.6	28.9		9.1	-11.9	30.4		7.9	-15.0	29.5
Wind speed (m/s)	met	met 1	5.07	0.00	16.93	met 1	5.07	0.00	16.93	met 4	6.37	0.00	22.58
	met	met 2	7.56	0.00	23.39	met 2	7.56	0.00	23.39	met 5	6.55	0.00	22.18
	met	met 3	6.42	0.40	20.56	met 3	6.42	0.40	20.56				
	WRF		8.11	0.55	23.07		7.81	0.56	21.77		7.84	0.53	27.25
	RN		7.88	0.31	20.16		7.70	0.08	18.91		6.50	0.63	19.20
Wind Direction (degrees)	met	met 1	192	0	360	met 1	192	0	360	met 4	199	0	360
	met	met 2	195	10	360	met 2	195	10	360	met 5	198	0	360
	met	met 3	189	0	360	met 3	189	0	360				
	WRF		193	2	356		191	4	357		185	3	356
Pressure (kPa)	met	met 1	99.2	96.8	101.3	met 1	99.2	96.8	101.3	met 4	100.4	97.5	102.7
	met	met 2	99.2	96.8	101.3	met 2	99.2	96.8	101.3	met 5	100.4	97.5	102.8
	met	met 3	99.5	97.0	101.6	met 3	99.5	97.0	101.6				
	WRF		99.3	96.9	101.3		99.4	97.2	101.4		100.5	97.7	102.7
Relative Humidity (%)	met	met 1	73	21	96	met 1	73	21	96	met 4	70	22	96
	met	met 2	75	25	100	met 2	75	25	100	met 5	70	21	100
	met	met 3	74	28	98	met 3	74	28	98				
	WRF		76	37	100		75	36	100		73	29	100
Air Density	RN		1.22	1.13	1.34		1.22	1.13	1.34		1.24	1.14	1.38
Irradiance	RN		0.07	0.00	0.45		0.07	0.00	0.46		0.07	0.00	0.45

3 Methods

3.1 Random forest model

The RF model was used to predict wind power output at each wind farm. RF is a regression ensemble learning method that trains a number of individual decision trees and outputs an average prediction of all trees (Breiman, 2001a). The RF method was chosen for its ability to handle outliers, a large number of variables, and correlated variables, with minimal negative impact on model performance (Jørgensen and Shaker, 2020). The RF method is described by equation (2), where B is the total number of trees, x is a test sample, and T_b is a single decision tree.

$$\hat{f}_{\mathrm{rf}}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2}$$

Grid search with cross validation was used to optimize hyperparameters. Final models contained 500 trees. All RF modelling and tuning was completed using the Scikit-learn (Pedregosa et al., 2011) package in Python version 3.9.7. Model performance was measured by mean absolute error (MAE) and root mean squared error (RMSE) of normalized wind power predictions, as described in equations (3) and (4), where $\hat{y}_i$ are predictions, y_i are observations, and n is the number of observations. Ten

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

random subsets of 70/30% train/test splits of data were used to test five scenarios (see Table 3.3 for scenario definitions), with the same ten random subsets used across scenarios. Reported MAE and RMSE scores are the average of these ten tests.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n}} \tag{4}$$

Percent improvement of RF models as compared to the benchmark power curve method was calculated as defined in equation (5), where EM represents the evaluation metric.

$$\%improvement = \frac{EM_{benchmark} - EM_{RF}}{EM_{benchmark}} \times 100\% \tag{5}$$

3.2 Benchmark Power Curves

The theoretical power P of a wind turbine is defined by Equation 6, where C_p is the coefficient of performance, v is the wind speed, ρ is the air density and A is the cross sectional area of the turbine.

$$P = C_p\frac{1}{2}\rho A v^3 \tag{6}$$

Annual energy production can be approximated using manufacturer provided power curves (**?**), which condense Equation 6 to a relationship between wind speed and power output, for a given turbine design. Power curves are determined using measurement methodology defined by IEC 61400-12-1, and historical wind speed data measured on site (Milan et al., 2010). Measured wind speeds from the three aforementioned data sources are used to find the expected power output given the power curve of the turbine (Figure 3). This expected wind power is used as a benchmark method to quantify the improvement when machine learning is introduced.

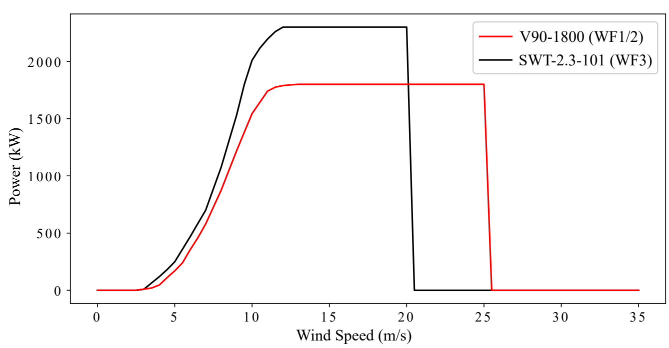


Figure 3. Manufacturer provided power curves for wind turbines at selected sites.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

3.3 Variable selection in random forest models

The "mean decrease impurity" (MDI) feature importance measure is often used as a variable reduction method, where X number of features identified as most important are kept, or features below a certain importance threshold are removed (Wenting et al., 2021; Shi et al., 2018). This default feature importance method defines importance as the mean reduction in mean squared error for each feature over all trees in the forest (Louppe et al., 2013). When used in this study, it was found that the multiple wind speed variables from different data sources were always classified as the most important to the model. This abundance of wind speed data resulted in the perfunctory removal of other weather and time variables that could improve model performance.

To ensure variables other than wind speeds are considered, an iterative random forward and backward selection process was used. To reduce the total number of features, highly correlated features (Pearson correlation coefficient > 0.97) were removed. Next, a baseline set of features was defined from well performing models from initial variable combination testing. Forward selection was used to iterate over the rest of the variables and the variable was kept if MAE improved. Backward selection was then used to iterate over selected variables – one feature at a time was removed, MAE was recalculated, and the feature was kept if performance decreased. Forward and backward selection was repeated until no new features were added or removed. The order of features in each process had some impact as some features may only be useful to the model in combination with another feature, therefore variables were selected randomly in both forward and backward selection.

Following forward and backward selection, approximately 30 variables were selected with multiples of wind speeds, directions, and other variable types from different met stations and WRF cells. Finally, MDI variable importance was used to select the most important variable for each variable type. The wind speed averaged across all data sources was found to be the most important and provided better results than models trained with WRF, RN, and met station wind speeds separately. A summary of the variable selection scenarios is outlined in Table 5.

The difference between the same variables as obtained from each data source is illustrated by the sample time series in Figure 4. Of all three datasets, only the meteorological stations are capable of capturing the short-term variations in the wind speed data, as exemplified by Figure 3a. Wind direction, humidity, pressure, and temperature were found to be very similar between meteorological measurements and WRF output and using one data source over the other was found to make negligible difference to the model output power. WRF data was therefore used over meteorological data to minimize missing data points and reliance on measured data.

Observed precipitation data from the met stations was found to improve model error metrics in early tests, however measured precipitation data was limited and not included in final models.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

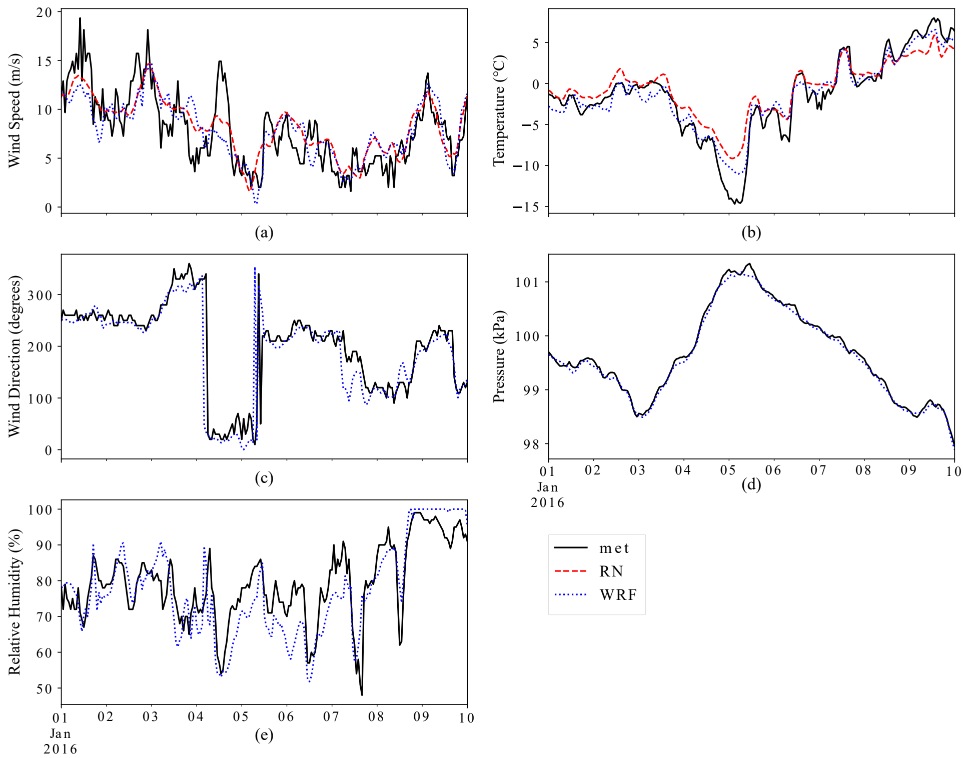


Figure 4. Sample time series for variables (Jan 1 - 10 2016 at one met station). Wind speeds interpolated to 80 m hub height. Air density and irradiance as obtained from RN not pictured.

3.4 Assessing variable importance across wind farms

Model Class Reliance (MCR), a recent method proposed by Fisher et al. (2019) and extended to the Random Forest algorithm (Smith et al., 2020), provides a more generalizable quantification of variable importance than traditional measures. MCR was used to quantify the importance ranges of the selected variables to the prediction of wind farm level power output. MCR (Fisher et al., 2019), provides a range of importance values for single variables to a *class* of models that provide equal predictive performance. The MCR for random forests python implementation as developed by Smith et al. (2020) was used for this analysis.

Model Reliance (MR) is defined as the decrease in model performance when a variable is disrupted and quantifies how much a single model relies on a variable to achieve accurate output. In Smith et al. (2020), this is defined as the difference between loss scores before and after a variable is disrupted.

Model reliance is defined in equation (7), where f is the predictive model, X_1 is the variable of interest, X_2 are other input variables, and Y is the model target variable. $\mathbb{E}L\left(f,\langle X_1,X_2,Y\rangle\right)$ and $\mathbb{E}L\left(f,\left\langle X_1^{\phi},X_2,Y\right\rangle\right)$ are the expected loss scores before and after X_1 has been rendered uninformative.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Table 5. Variable selection scenarios

Variable	Source	Met only	Reanalysis only	WRF only	WRF and reanalysis	Avg wind speed
Wind speed (m/s)	Average					•
	met	•				
	WRF			•	•	
	RN		•		•	
Temperature (°C)	met	•				
	WRF			•	•	•
	RN		•			
Wind Direction (degrees)	met	•				
	WRF			•	•	•
Pressure (kPa)	met	•				
	WRF			•	•	•
Relative Humidity (%)	met	•				
	WRF			•	•	•
Air Density	RN		•		•	•
Irradiance	RN		•		•	•
Total number of variables		5	4	5	8	7

$$MR_{X_1}(f,\Phi) = \mathbb{E}L\left(f, \left\langle X_1^{\phi}, X_2, Y \right\rangle\right) - \mathbb{E}L\left(f, \langle X_1, X_2, Y\rangle\right) \tag{7}$$

The MCR range is defined for each variable by MCR- and MCR+, which are the minimum and maximum MR values within the Rashomon set. Introduced by Breiman in Breiman (2001b), the Rashomon set is a set of different models with equal predictive performance. The MCR method for random forests searches over the Rashomon set by starting with a well-defined reference model, with hyperparameters fixed using cross-validation, and replaces trees within the model to minimize (MCR-) or maximize (MCR+) the new trees' reliance on the feature being studied. New trees are only added if they maintain the predictive accuracy achieved by the reference model, thus maintaining membership in the Rashomon set.

The MCR method is more useful than the node impurity variable importance measure used in the RF algorithm. This is because MCR provides an assessment of input variables across all RF models, making the final importance levels from MCR insensitive to the variances in the input dataset across single RF models.

4 Results and Discussion

Five scenarios of open access data combinations were tested as input to an RF model used to estimate normalized hourly production of three different wind farms. MAE and RMSE were used to evaluate different scenarios, the results of which are presented in Tables 6 and 7.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Table 6. Error metrics of random forest (RF) model and power curve (PC) predictions using met station data (MAE and RMSE of normalized wind power predictions). Percent improvement (*% imp*) of RF over PC (PC using average of all met data).

		met data only					
		RF	*PC met 1*	*PC met 2*	*PC met 3*	*PC avg met*	*% imp*
WF1	**MAE**	0.078	*0.184*	*0.173*	*0.190*	*0.122*	*36%*
	RMSE	0.112	*0.257*	*0.241*	*0.264*	*0.177*	*37%*
WF2	**MAE**	0.089	*0.193*	*0.206*	*0.217*	*0.153*	*42%*
	RMSE	0.126	*0.271*	*0.285*	*0.297*	*0.218*	*42%*
WF3	**MAE**	0.079	*0.155*	*0.158*		*0.155*	*49%*
	RMSE	0.115	*0.220*	*0.223*		*0.219*	*48%*

Table 7. Error metrics of random forest (RF) model and power curve (PC) predictions using RN, WRF, and combinations of WRF, RN, and met data. Percent improvement (*% imp*) of RF over PC.

		RN data only			WRF data only			RN and WRF			met, WRF, and RN (avg wind speed)		
		RF	*PC*	*% imp*	**RF**	*PC*	*% imp*	**RF**	*PC*	*% imp*	**RF**	*PC*	*% imp*
WF1	**MAE**	0.080	*0.167*	*52%*	0.076	*0.180*	*57%*	0.074	*0.167*	*56%*	0.065	*0.116*	*44%*
	RMSE	0.116	*0.235*	*51%*	0.110	*0.262*	*58%*	0.109	*0.241*	*55%*	0.097	*0.178*	*45%*
WF2	**MAE**	0.103	*0.224*	*54%*	0.080	*0.175*	*54%*	0.080	*0.175*	*54%*	0.076	*0.133*	*43%*
	RMSE	0.143	*0.301*	*52%*	0.115	*0.254*	*55%*	0.117	*0.247*	*53%*	0.111	*0.194*	*43%*
WF3	**MAE**	0.103	*0.177*	*42%*	0.074	*0.136*	*45%*	0.075	*0.135*	*45%*	0.067	*0.120*	*44%*
	RMSE	0.149	*0.249*	*40%*	0.109	*0.196*	*44%*	0.110	*0.187*	*41%*	0.101	*0.172*	*42%*

4.1 Analyzing the impact of data source

Power curve predictions at WF1 and WF2 were improved significantly by averaging available wind speeds (Table 6) and were further improved by using a Random Forest (RF) model and combining wind speed with other meteorological variables. The RF model is shown to be a significant improvement over predictions using only the power curve. Across input data, MAE improvement ranges from 36% for met data to 57% (Table 7) for WRF data.

At all three wind farm locations, output modelled using WRF data performed better than both reanalysis and met station data alone according to MAE and RMSE metrics. WRF data also provided the most consistent results between all locations. Models trained using RN data alone performed similarly to met station data at the WF1 but performed poorly compared to WRF and met data at both WF2 and WF3. Models trained using meteorological station data performed similarly well at both WF1 and WF3, both of which have met stations located within 20 km of the wind farm. The met data model did not perform as well as for WF2, which is 33 km away from the nearest met station. Including both RN and WRF data was found to make little change in error metrics when compared to the WRF only model – a small improvement can only be seen in the WF1 results. The best performing model for all three wind farms contained information from all three data sources, shown in Scenario 5 in Table 7.

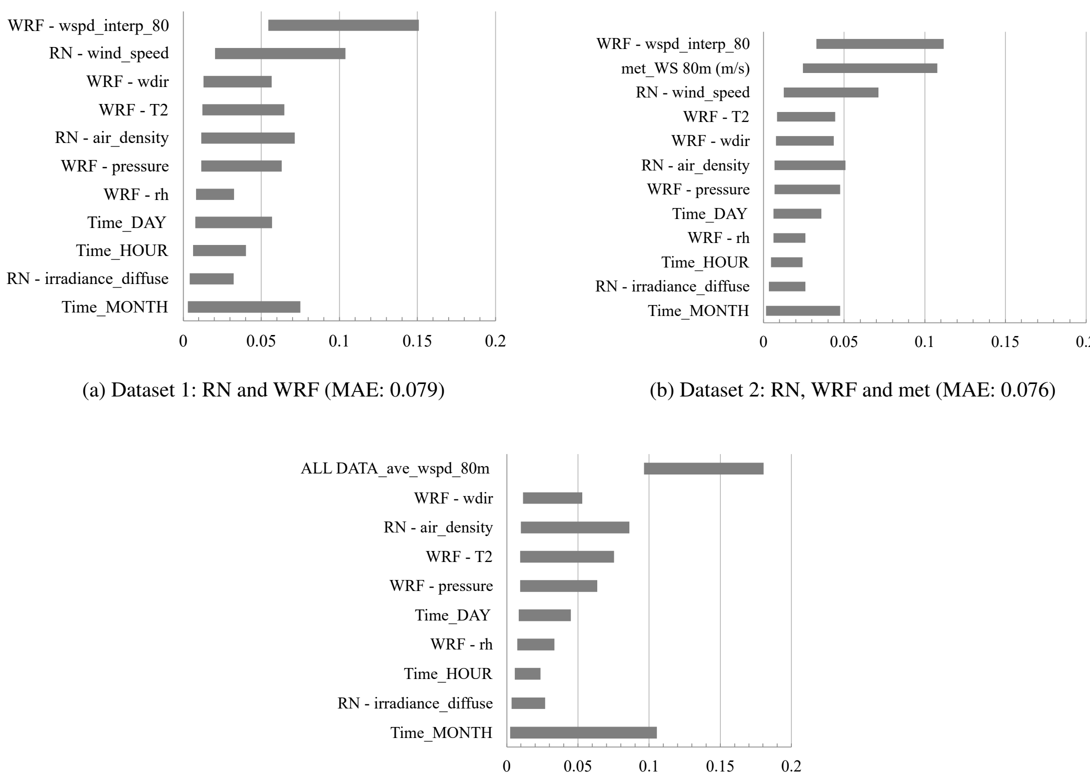
Across all three locations, including observed data from met stations improved results significantly, even at WF2 where the closest met stations are 33-49 km away from the wind farm. Reanalysis data may have been unable to capture local weather effects from terrain due to its coarse spatial resolution. Models trained using reanalysis data performed the best at WF1 which

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

has the least variation in geography and land use compared to the other locations. Models trained using WRF data provided the most consistent and accurate results out of weather station data, reanalysis data, and WRF data.

4.2 Analyzing the impact of feature importance

MCR was calculated for three input datasets, described in Figure 5, to quantify and understand the importance of different input variables. Datasets were designed to quantify the importance of wind speeds from different sources and the importance of other weather and time variables to each model. MCR was calculated on an RF model trained with input from all three locations, in order to capture the overall impact of each variable. All datasets included wind direction, temperature, pressure, and relative humidity data from the WRF models, irradiance and air density data from RN, and wind speed data from RN and WRF (Dataset 1), RN, WRF, and met (Dataset 2), or an average of wind speed data from all three data sources (Dataset 3). MCR ranges for each variable in each dataset are displayed in Figure 5.



(a) Dataset 1: RN and WRF (MAE: 0.079)

(b) Dataset 2: RN, WRF and met (MAE: 0.076)

(c) Dataset 3: average wind speed (MAE: 0.071)

Figure 5. MCR ranges by dataset (sorted by MCR- from lowest to highest)

No variables have an MCR- score of zero, therefore removing any variable from the model would decrease model performance to some degree (this result is further elaborated in Table 8). Month, day, hour, irradiance, and relative humidity

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

consistently have the lowest MCR- score, suggesting there is a model within each scenario where removing these variables would have minimal impact on model performance. The month variable has a high MCR+ score for all scenarios, indicating that some models rely heavily on the month to improve predictions. This also suggests month may be a good proxy for seasonality in wind speeds.

Pressure, temperature, air density, and wind direction have similar ranges of MCR- and MCR+, across all datasets. The overlapping MCR ranges suggest that these variables are correlated nonlinearly, and thus a model with similar predictive power but with one or more of the correlated variables dropped may exist.

As expected based from the theoretical power model, equation 6, wind speed variables are consistently the most important variable across all input datasets. Results from Datasets 1 and 2 indicate that removing WRF wind speed data would have the greatest impact on model performance, with both MCR- and MCR+ scores for WRF wind speed higher than the scores for RN and met wind speeds. In Dataset 3, MCR- and MCR+ scores for average wind speed are significantly higher than all other variables as the information from this variable can not be replaced by any other variable. In Datasets 1 and 2, MCR scores for wind speed variables are lower, as the models are able to obtain similar predictive accuracy by choosing to use one of the other available wind speeds.

The MAE for Dataset 3 was lower than Dataset 2, suggesting that the model may be able to make better use of other variables when variables of the same type are aggregated. This is also seen in MCR scores for the other weather variables, which are higher in Dataset 3 than in Dataset 2. Given the nature of decision trees and the RF model, removing correlated variables may allow the model to make better use of other variables as they are more likely to be selected as splits in a decision tree.

WRF wind speeds were more important to model performance than reanalysis and meteorological station wind speeds. An average of wind speeds from all data sources, however, was found to improve model performance. It was found that the month variable can improve model performance relatively significantly, and likely should be included for sites with seasonal variability.

4.2.1 Minimum subset of input data

The overlapping MCR ranges for these variables - particularly air density, temperature and pressure - indicate that there may be a smaller subset of variables within Dataset 3 that would give similar results. Subsets of Dataset 3 were tested to measure the impact on predictive accuracy when variables are removed. In each case, removing a variable increased MAE and RMSE by some percentage and the MAE and RMSE results were found to be significantly different when compared to the full Dataset 3 results via a 2-tailed t-test. Removing any single variable decreases model performance, with irradiance and humidity having the smallest impact in terms of percent decrease in MAE which was 0.5% and 1.1%, respectively (Table 8).

Kendall's Tau correlation was used to help identify relationships among the input variables used in Dataset 3, with detailed results provided in B. Many of the highly related variable interactions can be described by known physical processes. For example, based on the theoretical wind power equation (Equation 6), wind speed and air density are known to impact power output. It is also known that air density is impacted by pressure, temperature, and humidity. Physical relationships between

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

atmospheric variables appear to be captured by the machine learning model as indicated by the decrease in MAE and RMSE when a variable is removed.

To further explore variable interactions within the model, all possible combinations of variables within Dataset 3 were tested as input variables to the RF model. A sample of the highest performing combinations for various variable subsets is provided in 9. Removing irradiance and humidity together was found to have effectively no impact on model performance, with a percent change in MAE of -0.4%. Removing irradiance, humidity, and month was found to increase MAE by 1.6%, which increases to 4.2% when hour is removed as well. It was found that the total number of input variables in the subset cannot be reduced to less than six without increasing the MAE by 10% or more.

Other combinations of variables are explored in Appendix A. Results in Table A1 demonstrate that RF models trained using wind speed alone have MAE comparable to the power curve results, emphasizing the need to include other weather or time variables.

Table 8. Impact of removing single variables on MAE and RMSE compared to RF model using all input data included in Dataset 3. (t-test used to compare baseline to subsets, * represents p<0.01, ** represents p<0.001)

Variable removed	MAE	RMSE	% change (MAE)
Full Dataset 3	0.071	0.106	
air density	0.073**	0.108**	2.1%
pressure	0.074**	0.110**	4.1%
wind direction	0.074**	0.110**	3.3%
temp	0.073**	0.108**	2.2%
humidity	0.072**	0.108**	1.1%
irradiance	0.072*	0.107**	0.5%
month	0.073**	0.108**	1.9%
day	0.075**	0.109**	4.7%
hour	0.072**	0.108**	1.4%

Table 9. Minimum subsets of Dataset 3. Impact of removing variables on MAE and RMSE compared to full Dataset 3. (t-test used to compare baseline to other variable scenarios, * represents p<0.01, ** represents p<0.001)

MAE	RMSE	ave wspd 80m	WRF - T2	WRF - wdir	WRF - pressure	WRF - rh	RN - irradiance	RN - air density	Hour	Day	Month	count	% change (MAE)
0.071	0.106	•	•	•	•	•	•	•	•	•	•	10 (Full Dataset 3)	
0.071	0.106	•	•	•	•			•	•	•	•	8	-0.4%
0.073**	0.108**	•	•	•	•			•	•	•		7	1.6%
0.075**	0.112**	•	•	•	•			•		•		6	4.2%
0.079**	0.118**	•	•		•				•	•		5	9.9%

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

4.3 Transferability: towards a generalized random forest model for wind power prediction

Staid et al. (2018) found that a RF model trained on one wind farm was able to predict power output of a second unrelated wind farm with reasonable accuracy. To test the transferability of the input dataset developed in this paper, an RF model was trained using data from one of the three wind farms then tested on each of the other two sites. Performance metrics for the RF model is compared to the power curve predictions using data from the site of interest. Results in Table 10 demonstrate the ability to achieve similar or better predictive performance than power curve predictions. Compared to power curve calculations in Table 6, a random forest model trained using WRF and RN data performs better than power curve predictions using data from a single nearby met station. Models trained using WF3 data perform noticeably worse than models trained and tested on WF1 and WF2. This can be explained by differences in site characteristics (Figure 6).

Transferable models have applications in power prediction at locations where no historical data exists. A statistical model can be trained using environmental variables and power output from an existing wind farm and potentially applied to a location with no historical weather or power data.

Table 10. Random forest models trained using data from other sites compared to power curve calculations using closest meteorological station data

		RN and WRF data			met, WRF, and RN (avg wind speed)		
		RF	**RF**	*PC*	**RF**	**RF**	*PC*
tested on:	trained on:	WF2	WF3		WF2	WF3	
WF1	**MAE**	0.107	0.144	*0.167*	0.094	0.122	*0.116*
	RMSE	0.152	0.186	*0.241*	0.135	0.162	*0.178*
	trained on:	WF1	WF3		WF1	WF3	
WF2	**MAE**	0.131	0.155	*0.175*	0.113	0.137	*0.133*
	RMSE	0.176	0.195	*0.247*	0.154	0.176	*0.194*
	trained on:	WF1	WF2		WF1	WF2	
WF3	**MAE**	0.143	0.137	*0.135*	0.117	0.120	*0.120*
	RMSE	0.192	0.183	*0.187*	0.157	0.158	*0.172*

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

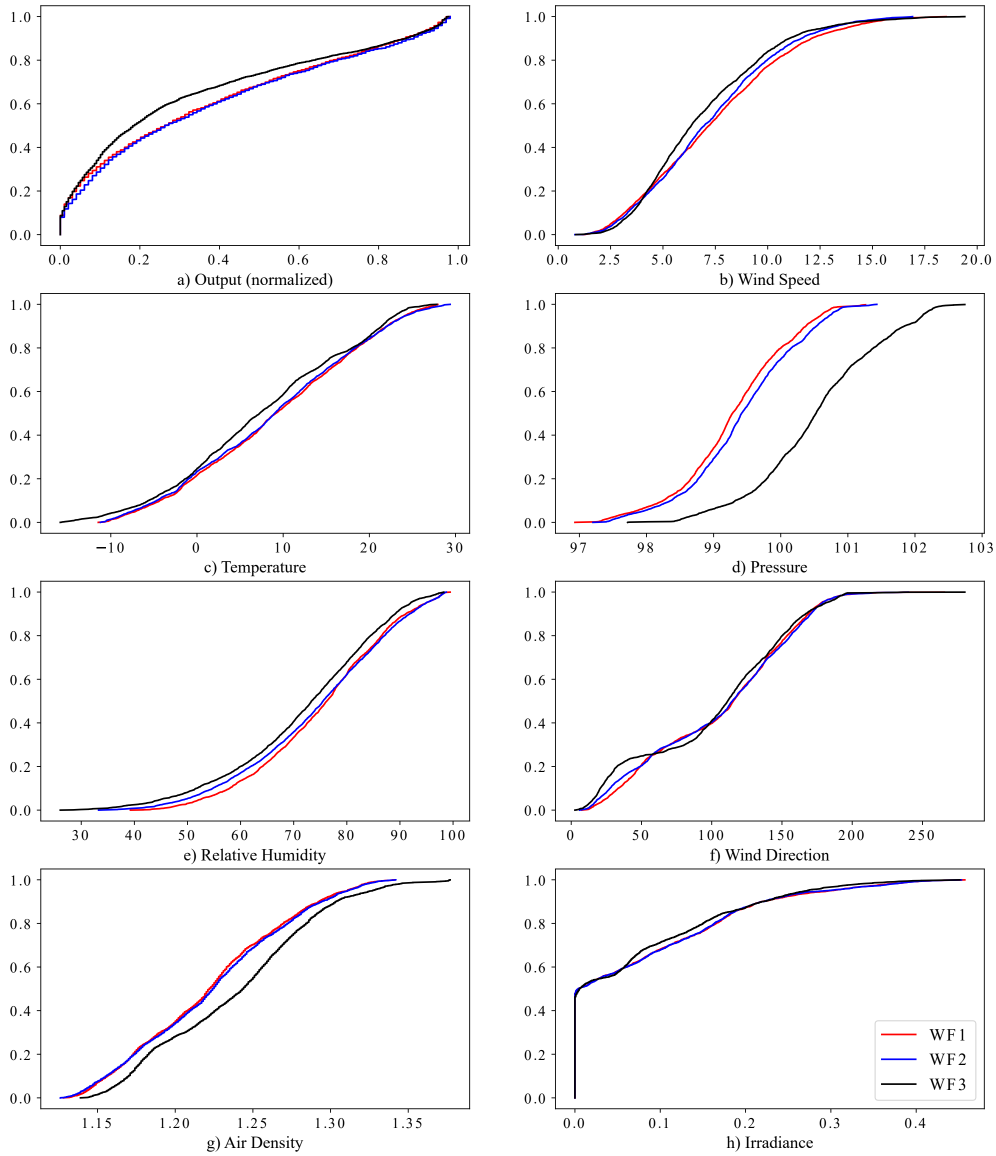


Figure 6. Empirical cumulative distributions of input and output variables from each site. Average of all data sources displayed. WF3 has noticeably different air pressure, density, temperature and power output characteristics.

340 ## 5 Conclusions

The random forest (RF) machine learning algorithm was used to predict wind power at three wind farms in Ontario using various combinations of weather station, reanalysis, and NWP data to determine the effectiveness of each open access input dataset, along with the optimal combination of datasets. MCR was used to quantify the impact of various meteorological variables from different sources on wind power predictions.

Models trained using data from a 3 km resolution in WRF consistently perform better than models trained with reanalysis or weather station data. In locations where weather stations have not been installed, the WRF model provides a viable alternative. While computationally expensive, the WRF model can provide historical time series data for meteorological variables faster than if a new weather station was installed.

The best performing models contain wind speed data from all three data sources. Weather station data improves power predictions even at sites where the nearest weather station is 33 km away. Reanalysis data is likely unable to capture local variability in wind speeds due to its coarse spatial resolution, but has demonstrated utility in combination with other data sources. It is also able to provide additional meteorological variables that are not often measured at weather stations, including irradiance.

All meteorological variables (wind speed, wind direction, temperature, pressure, air density, humidity, and irradiance) and time variables (month, day, and hour) were found to be useful to prediction performance to some degree based on MCR ranges and resulting changes to MAE and RMSE when variables were removed. Wind speeds from WRF were found to be most important to performance, followed by met and then reanalysis data. Removing each variable alone decreased performance, however removing both humidity and irradiance had very little impact on predictive performance. Time variables can increase predictive performance significantly and can be included without the need for additional on-site measurements. The best performing open access input dataset consisted of eight variables: 1) wind speed averaged from WRF, RN and weather stations sources; 2) temperature, 3) wind direction, and 4) air pressure from the WRF model; 5) air density from reanalysis data; and time variables associated with the input data, including 6) hour of the day, 7) day of the week and 8) month of the year. It is hypothesized that these time variables help the RF model capture the seasonality of wind speed.

It is important to note that the weather station data used in this study is measured well below hub height. Data measured at tall on-site stations would be expected to provide improved power predictions and should be considered in future work. WRF data computed at a 3 km resolution was unable to capture the variability seen in weather station wind speeds - higher resolution WRF models may be able to capture more variability and further prove WRF models as an alternative to local observational data. ERA5 reanalysis has been shown to perform better than MERRA2 reanalysis in other locations (Gruber et al., 2022) and should be considered in future work. Geometric variables based on wind direction and turbine spacing such as in (Yan and Ouyang, 2019) could be explored to better understand the importance of wind direction and expand the spatial transferability of wind power prediction models.

Appendix A: Other variable combination scenarios (wind speed, weather, and time variables)

RF models trained using only wind speed data produce error similar to the power curve model. Including time variables decreased MAE by up to 52%. MAE was further decreased by including weather variables. Time variables can be included by expanding wind speed time series and do not require additional on-site measurements.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Table A1. Error metrics of RF model trained using combinations of wind speed, time, and weather variables from each data source. % change compared to using wind speed alone. Time variables include hour, day, and month. Weather variables include temperature, air density, and irradiance for RN and temperature, wind direction, relative humidity, and pressure for WRF and met data. RF models trained on all three sites.

		RN	% change	WRF	% change	met	% change
Wind speed only	**MAE**	0.213		0.185		0.146	
	RMSE	0.299		0.268		0.195	
Wind speed & time	**MAE**	0.120	-44%	0.102	-52%	0.111	-48%
	RMSE	0.172	-42%	0.149	-50%	0.161	-46%
Wind speed & weather	**MAE**	0.161	-25%	0.093	-57%	0.108	-50%
	RMSE	0.224	-25%	0.134	-55%	0.151	-49%
Wind speed & weather & time	**MAE**	0.115	-46%	0.079	-63%	0.094	-56%
	RMSE	0.162	-46%	0.116	-61%	0.133	-55%

Appendix B: Kendall's Tau test

Kendall's Tau correlation coefficient was used to detect any monotonic relationships among the variables used in Scenario 3 (Tables B1 and B2). All variables were shown to have some degree of positive or negative monotonic relationship with the actual power output, with wind speed speed having the highest positive correlation followed by air density and wind direction. Temperature and pressure were shown to have a small negative correlation.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Table B1. Kendall's Tau output - 20 highest KT scores (* represents p<0.02, ** represents p<0.001)

x	y	kt
Temperature	Air Density	-0.818**
Output (normalized)	Wind Speed	0.642**
Temperature	Month	0.47**
Air Density	Month	-0.42**
Irradiance	Hour	0.386**
Relative Humidity	Irradiance	-0.275**
Pressure	Air Density	0.266**
Temperature	Irradiance	0.231**
Wind Speed	Pressure	-0.225**
Pressure	Relative Humidity	-0.215**
Air Density	Irradiance	-0.178**
Pressure	Wind Direction	-0.174**
Wind Speed	Temperature	-0.173**
Relative Humidity	Hour	-0.168**
Output (normalized)	Temperature	-0.155**
Output (normalized)	Pressure	-0.14**
Temperature	Pressure	-0.14**
Wind Speed	Air Density	0.131**
Output (normalized)	Air Density	0.122**
Relative Humidity	Month	-0.105**

Table B2. Kendall's Tau output - output vs all other variables (* represents p<0.02, ** represents p<0.001)

x	y	kt
Output (normalized)	Wind Speed	0.642**
Output (normalized)	Temperature	-0.155**
Output (normalized)	Pressure	-0.14**
Output (normalized)	Air Density	0.122**
Output (normalized)	Month	-0.099**
Output (normalized)	Wind Direction	0.079**
Output (normalized)	Day	-0.036**
Output (normalized)	Hour	0.035**
Output (normalized)	Irradiance	-0.027**
Output (normalized)	Relative Humidity	0.017*

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Appendix C: Kolmogorov-Smirnov test

Table C1. KS test output (* represents p<0.05, ** represents p<0.02, *** represents p<0.001)

comparison between:	WRF1 - WRF2		WRF1 - WRF3		WRF2 - WRF3	
variable	ks	equal?	ks	equal?	ks	equal?
Output (normalized)	0.044**	FALSE	0.268***	FALSE	0.266***	FALSE
Wind Speed	0.045**	FALSE	0.098***	FALSE	0.079***	FALSE
Temperature	0.02	TRUE	0.082***	FALSE	0.07***	FALSE
Pressure	0.089***	FALSE	0.534***	FALSE	0.478***	FALSE
Relative Humidity	0.043**	FALSE	0.083***	FALSE	0.07***	FALSE
Wind Direction	0.039*	FALSE	0.106***	FALSE	0.071***	FALSE
Air Density	0.026	TRUE	0.169***	FALSE	0.148***	FALSE
Irradiance	0.006	TRUE	0.042**	FALSE	0.042**	FALSE

Author contributions. EVZ: conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), software (lead), visualization (lead), writing - original draft and preparation (lead), writing - review and editing (equal). KRS: conceptualization (equal), formal analysis (equal), funding acquisition (lead), investigation (lead), project administration (lead), resources (lead), supervision (lead), writing - review and editing (equal).

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2021-04238].

Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [numéro de référence RGPIN-2021-04238].

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

References

Global Wind Atlas, https://globalwindatlas.info.

Wind energy database, https://www.thewindpower.net/.

Arrieta-Prieto, M. and Schell, K. R.: Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model, International Journal of Forecasting, 38, 300–320, https://doi.org/10.1016/j.ijforecast.2021.05.013, 2022.

Arrieta-Prieto, M. and Schell, K. R.: Spatially transferable machine learning wind power prediction models: v- logit random forests, Renewable Energy, 223, 120 066, 2024.

Bilal, B., Ndongo, M., Adjallah, K. H., Sava, A., Kebe, C. M. F., Ndiaye, P. A., and Sambou, V.: Wind turbine power output prediction model design based on artificial neural networks and climatic spatiotemporal data, in: 2018 IEEE International Conference on Industrial Technology (ICIT), pp. 1085–1092, https://doi.org/10.1109/ICIT.2018.8352329, 2018.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/a:1010933404324, num Pages: 5-32 Place: Dordrecht, Netherlands Publisher: Springer Nature B.V., 2001a.

Breiman, L.: Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author), Statistical Science, 16, 199–231, https://doi.org/10.1214/ss/1009213726, 2001b.

Brower, M., Bernadett, D. W., Elsholz, K. V., Filippelli, M. V., Markus, M. J., Taylor, M. A., and Tensen, J.: Wind Resource Assessment: A Practical Guide to Developing a Wind Project, John Wiley & Sons, Incorporated, Somerset, UNITED STATES, ISBN 978-1-118-24984-0, 2012.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., and Santos, C.: A sensitivity study of the WRF model in wind simulation for an area of high wind energy, Environmental Modelling & Software, 33, 23–34, https://doi.org/10.1016/j.envsoft.2012.01.019, 2012.

Dolter, B. and Rivers, N.: The cost of decarbonizing the Canadian electricity system, Energy Policy, 113, 135–148, https://doi.org/10.1016/j.enpol.2017.10.040, 2018.

Dudhia, J.: Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model, Journal of the Atmospheric Sciences, 46, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2, 1989.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD003296, 2003.

Environment and Climate Change Canada: Historical Data - Climate, https://climate.weather.gc.ca/historical_data/search_historic_data_e.html, last Modified: 2022-03-02, 2011.

Environment and Climate Change Canada: Technical Documentation - Digital Archive of Canadian Climatological Data, https://climate.weather.gc.ca/doc/Technical_Documentation.pdf, 2022.

Fisher, A., Rudin, C., and Dominici, F.: All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously, Journal of Machine Learning Research, 20, 1–81, http://jmlr.org/papers/v20/18-760.html, 2019.

Fitch, A. C., Olson, J. B., Lundquist, J. K., Dudhia, J., Gupta, A. K., Michalakes, J., and Barstad, I.: Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model, Monthly Weather Review, 140, 3017 – 3038, https://doi.org/10.1175/MWR-D-11-00352.1, 2012.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. d., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Giannaros, T. M., Melas, D., and Ziomas, I.: Performance evaluation of the Weather Research and Forecasting (WRF) model for assessing wind resource in Greece, Renewable Energy, 102, 190–198, https://doi.org/10.1016/j.renene.2016.10.033, 2017.

González-Aparicio, I., Monforti, F., Volker, P., Zucker, A., Careri, F., Huld, T., and Badger, J.: Simulating European wind power generation applying statistical downscaling to reanalysis data, Applied Energy, 199, 155–168, https://doi.org/10.1016/j.apenergy.2017.04.066, 2017.

Gruber, K., Klöckl, C., Regner, P., Baumgartner, J., and Schmidt, J.: Assessing the Global Wind Atlas and local measurements for bias correction of wind power generation simulated from MERRA-2 in Brazil, Energy, 189, 116 212, https://doi.org/10.1016/j.energy.2019.116212, 2019.

Gruber, K., Regner, P., Wehrle, S., Zeyringer, M., and Schmidt, J.: Towards global validation of wind power simulations: A multi-country assessment of wind power simulation from MERRA-2 and ERA-5 reanalyses bias-corrected with the global wind atlas, Energy, 238, 121 520, https://doi.org/10.1016/j.energy.2021.121520, 2022.

Hayes, L., Stocks, M., and Blakers, A.: Accurate long-term power generation model for offshore wind farms in Europe using ERA5 reanalysis, Energy, 229, 120 603, https://doi.org/10.1016/j.energy.2021.120603, 2021.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather Review, 132, 103–120, https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

Independent Electricity System Operator: 2021 Year in Review, https://www.ieso.ca/en/Corporate-IESO/Media/Year-End-Data, 2021.

Independent Electricity System Operator: Data Directory, https://www.ieso.ca/en/Power-Data/Data-Directory, 2022.

Jørgensen, K. L. and Shaker, H. R.: Wind Power Forecasting Using Machine Learning: State of the Art, Trends and Challenges, in: 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), pp. 44–50, https://doi.org/10.1109/SEGE49949.2020.9181870, iSSN: 2575-2693, 2020.

Lee, J. C. Y. and Lundquist, J. K.: Evaluation of the wind farm parameterization in the Weather Research and Forecasting model (version 3.8.1) with meteorological and turbine power data, Geoscientific Model Development, 10, 4229–4244, https://doi.org/10.5194/gmd-10-4229-2017, 2017.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P.: Understanding variable importances in forests of randomized trees, in: Advances in Neural Information Processing Systems, vol. 26, Curran Associates, Inc., https://proceedings.neurips.cc/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html, 2013.

Meyer, D. and Riechert, M.: Open source QGIS toolkit for the Advanced Research WRF modelling system, Environmental Modelling & Software, 112, 166–178, https://doi.org/10.1016/j.envsoft.2018.10.018, 2019.

Milan, P., Wächter, M., Barth, S., and Peinke, J.: Power curves for wind turbines, in: WIT Transactions on State of the Art in Science and Engineering, vol. 1, pp. 595–612, WIT Press, 1 edn., ISBN 978-1-84564-205-1, https://doi.org/10.2495/978-1-84564-205-1/18, 2010.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16 663–16 682, https://doi.org/10.1029/97JD00237, 1997.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2, Geoscientific Model Development, 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Morales-Ruvalcaba, C. F., Rodríguez-Hernández, O., Martínez-Alvarado, O., Drew, D. R., and Ramos, E.: Estimating wind speed and capacity factors in Mexico using reanalysis data, Energy for Sustainable Development, 58, 158–166, https://doi.org/https://doi.org/10.1016/j.esd.2020.08.006, 2020.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, Boundary-Layer Meteorology, 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce: NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids, https://doi.org/10.5065/D65Q4T4Z, 2015.

Natural Resources Canada: Canadian Wind Turbine Database, https://open.canada.ca/data/en/dataset/79fdad93-9025-49ad-ba16-c26d718cc070, 2021.

Pang, C., Yu, J., and Liu, Y.: Correlation analysis of factors affecting wind power based on machine learning and Shapley value, IET Energy Systems Integration, 3, 227–237, https://doi.org/10.1049/esi2.12022, 2021.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, : Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, http://jmlr.org/papers/v12/pedregosa11a.html, 2011.

Pfenninger, S. and Staffell, I.: Renewables.ninja, https://www.renewables.ninja/.

Pfenninger, S. and Staffell, I.: Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data, Energy, 114, 1251–1265, https://doi.org/10.1016/j.energy.2016.08.060, 2016.

Pinson, P.: Very-short-term probabilistic forecasting of wind power with generalized logit—normal distributions, Journal of the Royal Statistical Society. Series C (Applied Statistics), 61, 555–576, https://doi.org/10.1111/j.1467-9876.2011.01026.x, 2012.

Pombo, D. V., Göçmen, T., Das, K., and Sørensen, P.: Multi-Horizon Data-Driven Wind Power Forecast: From Nowcast to 2 Days-Ahead, in: 2021 International Conference on Smart Energy Systems and Technologies (SEST), pp. 1–6, https://doi.org/10.1109/SEST50973.2021.9543173, 2021.

Prósper, M. A., Otero-Casal, C., Fernández, F. C., and Miguez-Macho, G.: Wind power forecasting for a real onshore wind farm on complex terrain using WRF high resolution simulations, Renewable Energy, 135, 674–686, https://doi.org/10.1016/j.renene.2018.12.047, 2019.

Sasser, C.: Data-Driven Meteorological-Feature-Informed Wind Power Prediction with Machine Learning Decision Trees, Master's thesis, University of Maryland, Baltimore County, United States – Maryland, http://www.proquest.com/docview/2572607772/abstract/55B7E56DE934C4BPQ/1, iSBN: 9798535573427, 2021.

Shi, K., Qiao, Y., Zhao, W., Wang, Q., Liu, M., and Lu, Z.: An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness, Wind Energy, 21, 1383–1394, https://doi.org/10.1002/we.2261, 2018.

Singh, U., Rizwan, M., Alaraj, M., and Alsaidan, I.: A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments, Energies, 14, 5196, https://doi.org/10.3390/en14165196, 2021.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., and Huang, X.-Y.: A Description of the Advanced Research WRF Model Version 4, Tech. rep., UCAR/NCAR, https://doi.org/10.5065/1DFH-6P97, 2019.

https://doi.org/10.5194/wes-2025-29
Preprint. Discussion started: 20 March 2025

Smith, G., Mansilla, R., and Goulding, J.: Model Class Reliance for Random Forests, in: Advances in Neural Information Processing Systems, edited by Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., vol. 33, pp. 22 305–22 315, Curran Associates, Inc., https://proceedings.neurips.cc/paper/2020/file/fd512441a1a791770a6fa573d688bff5-Paper.pdf, 2020.

Staffell, I. and Pfenninger, S.: Using bias-corrected reanalysis to simulate current and future wind power output, Energy, 114, 1224–1239, https://doi.org/10.1016/j.energy.2016.08.068, 2016.

Staid, A., VerHulst, C., and Guikema, S. D.: A comparison of methods for assessing power output in non-uniform onshore wind farms, Wind Energy, 21, 42–52, https://doi.org/10.1002/we.2143, 2018.

Sweeney, C., Bessa, R. J., Browell, J., and Pinson, P.: The future of forecasting for renewable energy, WIREs Energy and Environment, 9, e365, https://doi.org/https://doi.org/10.1002/wene.365, 2020.

Tomaszewski, J. M. and Lundquist, J. K.: Simulated wind farm wake sensitivity to configuration choices in the Weather Research and Forecasting model version 3.8.1, Geoscientific Model Development, 13, 2645–2662, https://doi.org/10.5194/gmd-13-2645-2020, publisher: Copernicus GmbH, 2020.

Wang, A., Xu, L., Li, Y., Xing, J., Chen, X., Liu, K., Liang, Y., and Zhou, Z.: Random-forest based adjusting method for wind forecast of WRF model, Computers & Geosciences, 155, 104 842, https://doi.org/10.1016/j.cageo.2021.104842, 2021.

Wenting, Z., Jie, L., Yalong, L., and Yingyu, L.: Short-term wind power forecasting model based on random forest algorithm and TCN, in: 2021 40th Chinese Control Conference (CCC), pp. 5776–5781, https://doi.org/10.23919/CCC52363.2021.9550052, iSSN: 1934-1768, 2021.

Yan, J. and Ouyang, T.: Advanced wind power prediction based on data-driven error correction, Energy Conversion and Management, 180, 302–311, https://doi.org/10.1016/j.enconman.2018.10.108, 2019.

Özen, C., Dinç, U., Deniz, A., and Karan, H.: Wind power generation forecast by coupling numerical weather prediction model and gradient boosting machines in Yahyalı wind power plant, Wind Engineering, 45, 1256–1272, https://doi.org/10.1177/0309524X20972115, 2021.