Evaluating the Impact of Inter-Annual Variability on Long-Term Wind Speed Predictions

Johanna Borowski[1,3], Sandra Schwegmann[1], Kerstin Avila[2,3], and Martin Dörenkämper[1]

[1]Fraunhofer IWES, Fraunhofer-Institut für Windenergiesysteme IWES, Küpkersweg 70, 26129 Oldenburg, Germany
[2]ForWind - Center for Wind Energy Research, Küpkersweg 70, 26129 Oldenburg, Germany
[3]Carl von Ossietzky Universität Oldenburg, School of Mathematics and Science, Institute of Physics, Ammerländer Heerstraße 114-118, 26129 Oldenburg, Germany

Correspondence: Johanna Borowski (johanna.borowski@iwes.fraunhofer.de)

Abstract. Assessing the wind resource and its associated uncertainties is essential for the profitability of a wind farm, with inter-annual variability in wind speed being a key factor. To estimate the wind resource at a potential wind farm site, a year-long wind measurement campaign is typically conducted and combined with long-term - often numerical - reference data using the Measure-Correlate-Predict (MCP) approach. This process accounts for systematic errors in the reference data and captures the long-term wind variability of wind speed. Since wind conditions vary from year to year, the selection of a single measurement year within the MCP framework can significantly influence the predicted wind resource. In this study, we systematically evaluate the impact of the measurement year on wind speed predictions using long-term met mast measurements. We also investigate whether classical and advanced machine learning methods can mitigate this sensitivity. Our results reveal that the variation in predicted wind speed due to the chosen measurement year ranges from 1 % to 14 %, depending on the site and correlation method, with an average of 6.5 %. Excluding years with exceptional wind conditions reduces the mean to 4.2 %. Among the methods selected, the correlation method SpeedSort, along with the advanced machine learning models Random Forest and AdaBoost, most effectively mitigates the influence of inter-annual wind variations in long-term referencing compared to classic linear regression. Additionally, the findings indicate that AdaBoost and Random Forest are especially beneficial for sites with heterogeneous and complex terrain. Furthermore, the study highlights the need for quality-controlled, long-term datasets across a variety of sites with differing terrain complexities to better understand and manage the effects of inter-annual wind variability in diverse wind climates.

1 Introduction

Accurately estimating the wind resource at a potential wind farm site is a critical component of the planning phase and plays a key role in ensuring project profitability. This includes both assessing the wind speeds at the potential site and estimating associated uncertainties (Rohrig et al., 2019).

To characterize the wind climate at a target site, a wind measurement campaign is usually carried out as a short-term measurement series at least one year as part of the site assessment process (MEASNET, 2022; FGW e.V., 2023).

Since wind conditions vary not only from year to year (inter-annual variability) but also over longer time scales (multi-decadal variability), it is necessary to relate the short-term measurement time series to longer periods (MEASNET, 2022; FGW e.V., 2023). To achieve this, short-term measurement time series are extended using long-term reference time series which typically span over at least one decade (FGW e.V., 2023).

To extend the one-year short-term measurement with the long-term reference data, a statistical relationship is typically determined based on the overlap time period between the two datasets. The derived correction function is then applied to adjust the long-term data to reflect the wind conditions at the potential site. The resulting ~~hindcast~~ prediction can then be used to evaluate the wind climate at the target site. This "Measure - Correlate - Predict (MCP)" methodology is widely used in wind resource assessment, with various correlation methods of varying degree of complexity available (Carta et al., 2013).

In industrial wind resource assessments, the MCP methodology is often based on simple linear models (Carta et al., 2013; Basse et al., 2021), but advanced machine learning (ML) applications gain popularity in context of wind energy, (e.g., Optis et al., 2021; Schwegmann et al., 2023; Velázquez et al., 2011; Bass et al., 2000; Zhang et al., 2014; Stetco et al., 2019; Bakhoday-Paskyabi, 2020; Barber and Nordborg, 2020; Bodini and Optis, 2020; Daniel et al., 2020). Optis et al. (2021) reveal that ML approaches, especially the Random Forest method, enhance the vertical interpolation of wind speeds from surface to hub heights and Schwegmann et al. (2023) conclude that ML applications are beneficial for gap filling with MCP in wind resource assessment.

When estimating the wind resource, it is necessary to account for various uncertainties as they impact the reliability of assessments. The wind speed uncertainty is a key component and encompasses the uncertainty components such as long-term referencing, MCP methodology, wind measurements, site environment and resolution, wind variability (~~both~~ diurnal, inter-annual and inter-monthly, as well as future projections), and other considerations as outlined in the International Electrotechnical Commission (IEC) 61400-15 proposed framework discussed in Lee and Fields (2021). Understanding and mitigating these uncertainties enhances decision-making and risk management in the development and operation of wind energy systems.

Inter-annual wind variability is influenced by various factors - some of which vary significantly by geography - including oceanic and atmospheric oscillations such as El-Niño Southern Oscillation, the North Atlantic Oscillation, and the Pacific Decadal Oscillation. Additionally, the geographical environment, such as mountains or the distribution of land and water masses, shapes wind patterns and can also have an impact on the inter-annual wind variability by altering atmospheric stability or wind direction. These influences can result in significant regional differences in wind climate and its variability. Long-term datasets, such as reanalysis data that integrate model outputs with actual measurements, can capture and reflect these effects. Consequently, long-term data are crucial for analyzing historical trends, variability, and anomalies in climate patterns, making them essential for assessing long-term wind conditions.

Numerous methods for determining inter-annual variability can be found in the literature; Lee et al. (2018) compared over 20 of these methods and evaluated their advantages and disadvantages. ~~Depending~~ They recommend using the robust and resilient coefficient of variation (RCoV) to quantify the fluctuations in wind resources and energy production. Since RCoV is a normalized spread metric and - compared to other metrics tested - more effectively captures the relationship between large wind speed fluctuations in a wind farm and the resulting variability in wind energy production, it is an advantageous spread

metric compared to other metrics. However, depending on the calculation method, datasets study area and the time period considered, the inter-annual variability vary between 1.3 % and 10–15 % (e.g., Lee and Fields, 2021; Wohland et al., 2019; Pullinger et al., 2017; Hamlington et al., 2015; Watson et al., 2015; Früh, 2013; Martin, 2010; Pryor et al., 2006; Klink, 2002; Baker et al., 1990; Justus et al., 1979; Corotis, 1976). The aim of the study is to investigate and mitigate the impact of inter-annual variability on wind speed predictions in the long-term referencing process of wind resource assessment using MCP methodology.

The first objective is to quantify the impact of inter-annual variability in long-term referencing across multiple sites with varying terrain complexities using classic linear regression as the correlation method in MCP. Furthermore, we examine the sensitivity of both classical and advanced ML methods to inter-annual variability. Based on this analysis, we assess which methods are most effective in reducing the impact of inter-annual variability in long-term referencing. The study benefits from quality-controlled, long-term measurement time series from twelve wind-energy-relevant sites with diverse terrain and environmental complexities.

The sites, data and climatology along with the MCP methodology and correlation methods used are introduced in Sect. 2. Key findings ~~related to site climatology,~~ of the intercomparison of correlation methods and time range variability are presented in Sect. 3. Finally, Sect. 4 discusses the results, draws conclusions, and provides a brief outlook.

2 Data and Methods

This section provides an overview of the site characteristics, the measurement and reanalysis data used in the study. Additionally, it outlines the procedure for defining the impact of inter-annual wind variability in long-term referencing.

2.1 ~~Data~~Selection of met mast and reanalysis data

2.1.1 Tall met mast data

To investigate the inter-annual variability of the wind climate, long-term continuous wind measurements are required. Numerous wind measurements exist at a height of 10 m (Ramon et al., 2020); but vertically interpolating these to wind turbine hub heights can introduce additional uncertainties. Therefore, this study focuses on measurement data from tall met masts that provide wind measurements at around 100 m - close to the hub heights of modern wind turbines. Additionally, the selected sites have a yearly average wind speed of more than $4\,\mathrm{m\,s^{-1}}$ at wind energy relevant heights and provide long-term wind measurements – each more than seven consecutive overlapping years (2010–2016).

A total of eight met mast locations from the Tall Tower Dataset (Ramon et al., 2020) were identified as suitable for the study (Table 1): Megler (US), Goodnoe Hills (US), Chinook (US), Butler Grade (US), Boulder (NWTC M2 (Jager and Andreas, 1996), US), Park Falls (WLEF (Davis et al., 2003), US), Hegyhátsál (HU) and Obninsk (RU). Additionally, data from four further met masts were considered with the sites: Hamburg (Hamburg Weather Mast, GER), Falkenberg (GER), Karlsruhe (Kohler et al. (2018), GER), and Cabauw (NL).
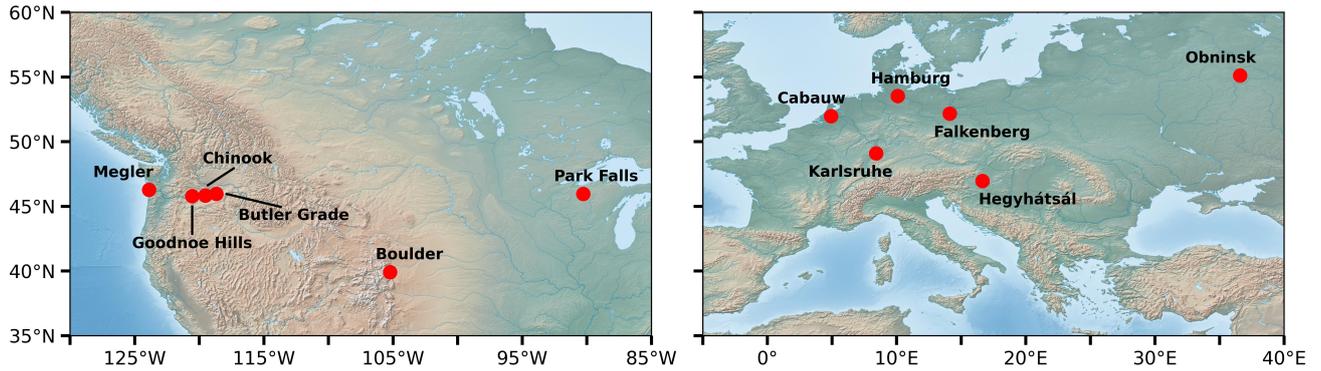
Figure 1. Distribution of the met mast positions from the Tall Tower Dataset (Ramon et al., 2020), including additional masts at the sites Cabauw, Hamburg, Falkenberg, and Karlsruhe.

Table 1. Overview of the met mast positions and measurement parameters. NWTC: National Wind Technology Center; WS: wind speed; WD: wind direction.

Site	Country	Lon	Lat	Measurement height [m]	Mean wind speed [m s^{-1}]	Measuring devices	Resolution of available data
Megler	US	-123.88	46.27	53	5.22	unknown	10 min
Goodnoe Hills	US	-120.55	45.78	59	5.97	unknown	10 min
Chinook	US	-119.53	45.83	50	4.64	unknown	10 min
Butler Grade	US	-118.68	45.95	62	7.23	unknown	10 min
Boulder (NWTC M2)	US	-105.23	39.91	80	4.72	cup/wind vane	10 min
Park Falls (WLEF)	US	-90.27	45.95	122	6.07	unknown	hourly
Cabauw	NL	4.93	51.97	80	6.77	cup/wind vane	10 min
Karlsruhe	GER	8.43	49.09	100	4.67	cup/wind vane	10 min
Hamburg (Hamburg Weather Mast)	GER	10.10	53.52	110	6.36	sonic	10 min
Falkenberg	GER	14.12	52.17	98	5.91	sonic	10 min
Obninsk	RU	36.6	55.11	121	5.48	unknown	hourly
Hegyhátsál	HU	16.65	46.96	82 (WS), 115 (WD)	4.84	sonic	hourly

The selected met masts are distributed across the Northern Hemisphere (Fig. 1), primarily concentrated in Central Europe and
90 the northern states of the United States. They are located in various types of terrain complexities. To classify the complexity, the
standard deviation of terrain height ($\sigma_{terrain}$) within a 10 km x 10 km grid surrounding the met masts was calculated. The fol-

4
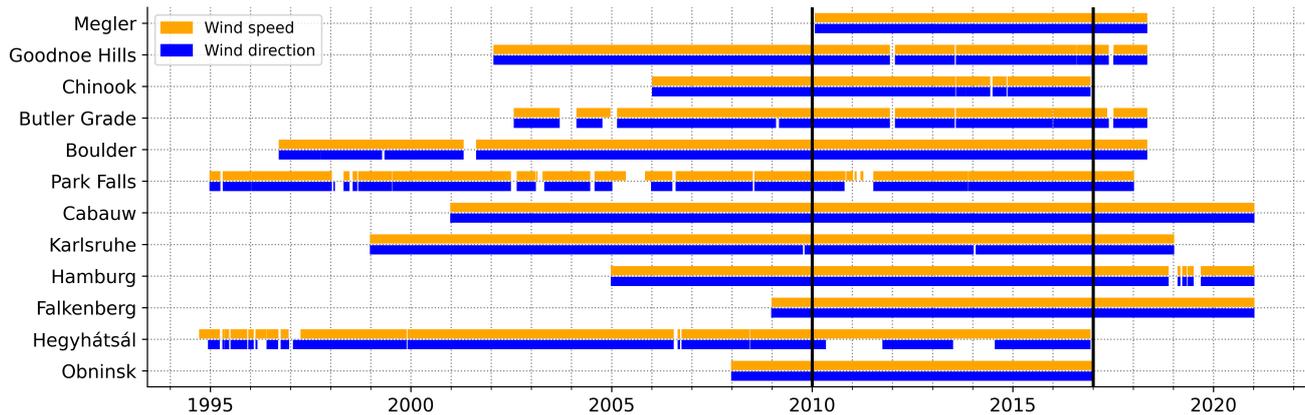
Figure 2. Measured wind speed and direction availability at the selected sites from end of 1994 to 2021, with an overlapping period from 2010 to 2016 highlighted by black lines. Within this overlapping period the individual sites exhibit at least 70 % wind data availability per year, except for Park Falls, which falls below 50 % in 2010 and 2011 and Hegyhátsál with less than 50 % wind direction data in 2010, 2011, 2013, and 2014.

lowing categories of terrain complexity were established: simple ($\sigma_{terrain} < 10$ m), heterogeneous (10 m $\leq \sigma_{terrain} < 20$ m), complex (20 m $\leq \sigma_{terrain} < 100$ m) and very complex ($\sigma_{terrain} \geq 100$ m). Based on this classification, the sites were assigned to the appropriate categories according to their terrain complexity (Table 2).

95   Some masts, e.g. Butler Grade and Goodnoe Hills have wind turbines in the surrounding but not from relevant wind direction sectors. So, the majority of sites are not impacted by wakes. Furthermore, the measurement data for this study are available at various averaging intervals (Table 1). Since the reanalysis data are provided at hourly resolution, the measurements are aggregrated aggregated to hourly intervals unless they are already available in hourly format.

2.1.2 Model data

100 Meteorological reanalysis data provide long-term datasets of relevant meteorological quantities by combining historical observation data with advanced numerical weather models. Modern Current generation of reanalysis datasets provide data over periods of more than 50 years thus enable to analyze and therefore multi-decadal records of atmospheric conditions, enabling the study of large-scale climate patterns and their impact on (regional) wind variability. In wind resource assessment, reanalysis data from global circulation models are used as long-term reference datasets to provide a basis for the

105 inter-annual and decadal variability of the wind conditions. In this study, the reanalysis data ERA5 (Fifth Generation European Reanalysis, Hersbach et al. (2020)) of the European Centre for Medium-Range Weather Forecasts (ECMWF) are used. Due to its high time resolution of one hour and While mesoscale dowscaling approaches might provide even more accurate average wind speeds at sites, several studies have pointed out the suitability of ERA5 for wind energy applications (Olauson, 2018; Hahmann et al., 2020; Dörenkämper et al., 2020; Gottschall and Dörenkämper, 2021) due to its relatively high

5

Table 2. Description of site characteristics and terrain complexity based on the standard deviation ($\sigma_{terrain}$) [m] of terrain height variation within a 10 km x 10 km area around each site.

Site	Landscape	$\sigma_{terrain}$ [m]	Terrain complexity category
Cabauw	coastal area	1.8	simple
Park Falls	boreal lowland, wetland forests, flat	9.8	simple
Hamburg	urban, flat	14.7	heterogeneous
Falkenberg	forested, fields, small lakes, flat	17.7	heterogeneous
Obninsk	forested, fields, urban, flat	19.9	heterogeneous
Karlsruhe	forested, Rhine plain	25.0	complex
Hegyhátsál	rural, fields, forest patches, flat	25.6	complex
Chinook	near river/valley, fields	43.2	complex
Megler	North: forested and hilly, East to South: hilly with a river and estuary, West: coastal	86.2	complex
Butler Grade	near valley, mountainous	177.6	very complex
Goodnoe Hills	near valley, mountainous	195.1	very complex
Boulder	West: Rocky Mountains, East: flat/open	266.9	very complex

spatial resolution of ~0.28° (31 km) and ~~an availability of the wind speed components at 100 m, the reanalysis~~ - compared to other reanalyses - high resolution in time of one hour. In particular, the correlation between measurements and ERA5 has been proven to be ~~well suited for several wind energy applications (Olauson, 2018; Hahmann et al., 2020; Dörenkämper et al., 2020; Gottschall~~ high, making it in particular suitable for long-term referencing applications (Meyer and Gottschall, 2022; Gottschall and Dörenkämper, 202 . The ERA5 reanalysis data cover the time period from 1940 till present and thus, it gives the possibility to capture the decadal changes in the atmosphere within this period.

In order to compare ERA5 with measurements and to use the data for further analysis, ERA5 data have been extracted from the nearest grid point to the site of interest, i.e. the met mast locations described above. An exception is made for the Megler site, where the nearest grid point is over the sea, and the dynamics can differ significantly. Therefore, the next grid point that has land surface properties in the reanalysis was selected instead. The analysis covers the time range from 1950 to 2020.
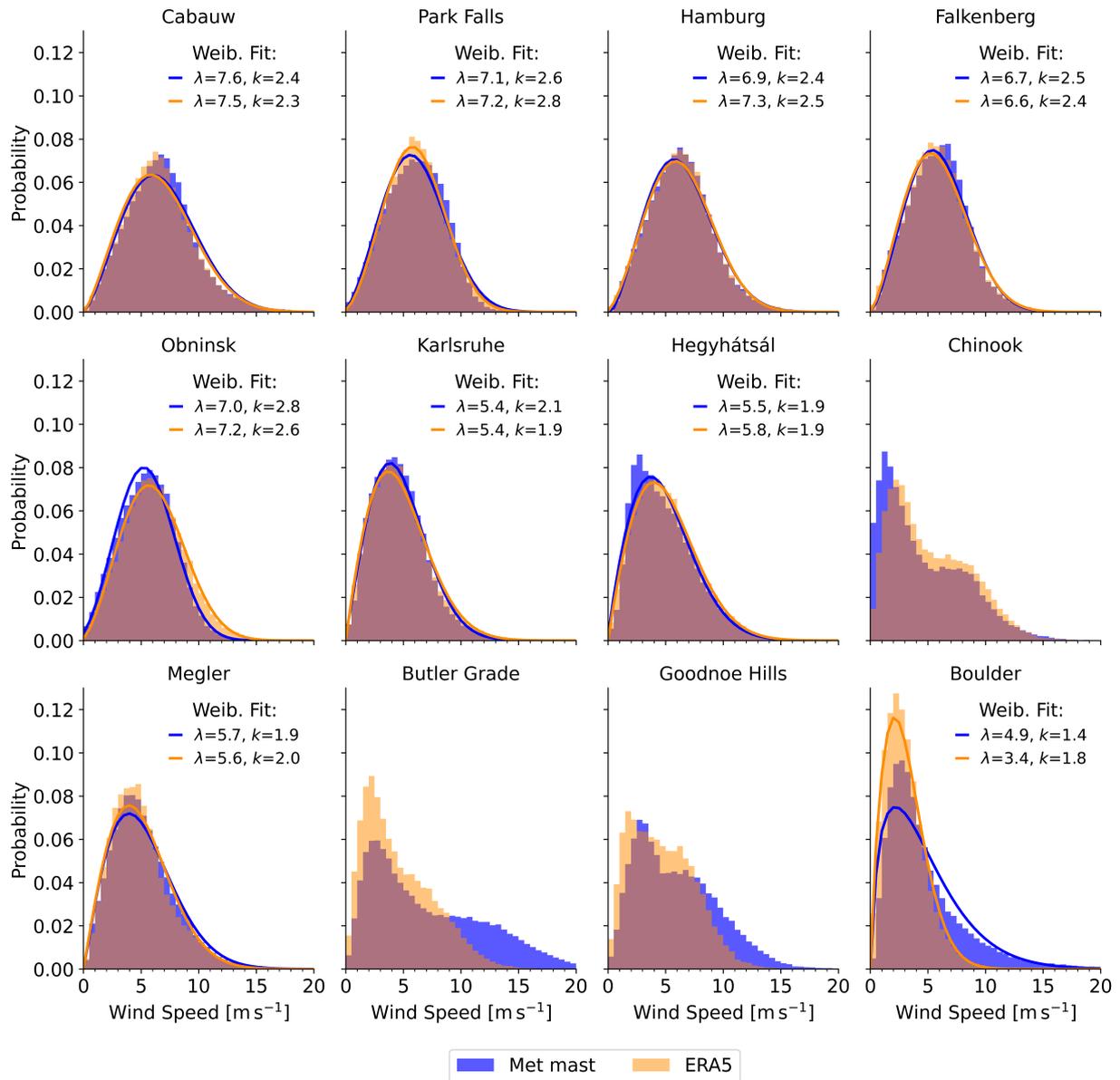
2.2 Wind climatology at selected sites: A comparison with ERA5

The selected sites exhibit varying wind climates, which are outlined below. In addition, the differences between the reference time series (ERA5) and the measurements are examined. The wind characteristics of the measurement data and the ERA5 reanalysis data for the multi-year overlap time period of 2010–2016 for all sites used in this study are illustrated in Figure 3 and Figure 4. At most sites, the measured wind direction is predominantly westerly (Fig. 3) with varying intensity, except in
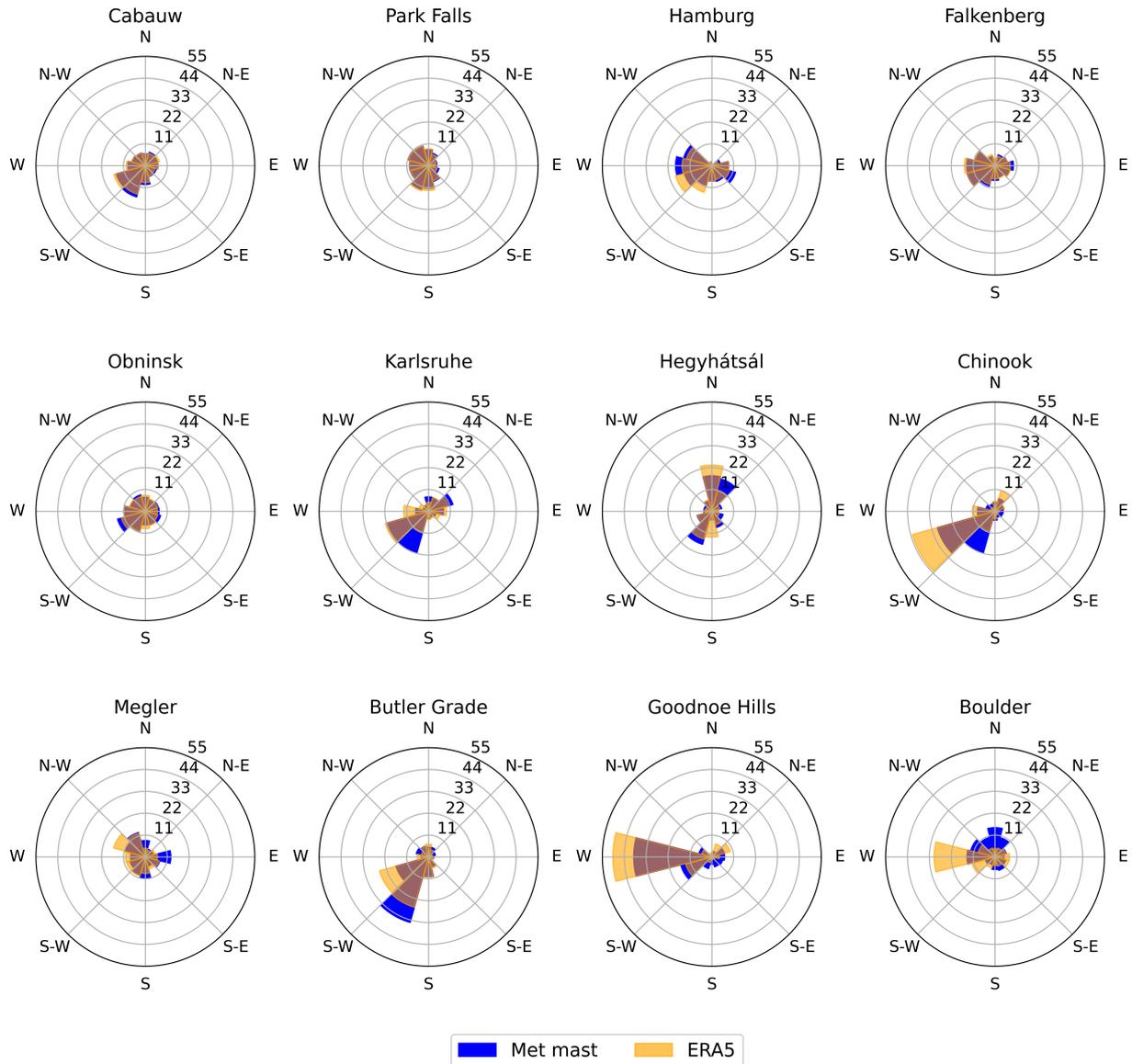
Figure 3. Wind direction distribution of measurement (blue) and ERA5 reanalysis (orange) data for the time period 2010–2016. Note: Overlapping sectors lead to a brownish color. Sites are sorted by terrain complexity from simple to very complex from top left to bottom right.

Hegyhátsál and Boulder, where northern winds are also more frequent. Sites with simpler terrain exhibit a more uniform wind direction distribution (e.g. Cabauw) compared to those with complex terrain (e.g. Butler Grade). The measured wind speed for most sites (e.g. Cabauw, Falkenberg, Boulder) follows a Weibull distribution, while some (e.g. Goodnoe Hills) display bimodal distributions, which are common in highly complex terrain due to redirection of the larger-scale flow.

Figure 4. Wind speed distribution of measurement (blue) and ERA5 reanalysis (orange) data for the time period 2010–2016. Note: Overlapping bars lead to a brownish color. Sites are sorted by terrain complexity from simple to very complex from top left to bottom right. A Weibull fit (solid lines) is performed and shape (λ [m s^{-1}]) and scale (k) parameters are provided in the legend except for cases with distinct non-Weibull, e.g. bimodally distributed met mast winds.

When comparing reanalysis data with the measurements, the main wind direction distribution is generally captured by the ERA5 data. But, in some cases, deviations occur between wind direction sectors, with wind directions shifted to an adjacent

130

8

sector (e.g., Karlsruhe, Chinook, Butler Grade). At the Megler site, an underrepresentation (overrepresentation) of the easterly (northwesterly) winds can be observed in the ERA5 data. This could be attributed to the highly variable wind direction sectors that are influenced by terrain characteristics (Table 2). At the Boulder site, the ERA5 data indicate a shift of nearly 90° compared to the measured wind direction. This can be attributed to the sites's location at the foot of the Rocky Mountains, where local geographical features strongly influence wind patterns. Furthermore, the ERA5 reanalysis data overrepresent lower wind speeds and underrepresent higher wind speeds at certain sites (Fig. 4). At the three most complex sites, lower wind speeds are overrepresented, while the highest wind speeds are almost completely absent in the ERA5 data. Conversely, at sites Hegyhátsál and Chinook lower wind speeds are slightly underrepresented, while higher wind speeds occur more frequently in ERA5 than in the measurements. The differences between ERA5 and measurement data may arise from choosing the nearest ERA5 grid point to the mast location, which may not accurately reflect the site characteristics due to ERA5's resolution.

2.3 Methods

2.3.1 Measure-Correlate-Predict (MCP)

In the wind energy industry, the state-of-the-art procedure for correcting reference data to site-specific conditions (mostly based on measurement data) is known as the **M**easure-**C**orrelate-**P**redict (MCP) method. In MCP, on-site measurements are correlated with a long-term (typically numerical) reference dataset, and predictions are then made based on this correlation, accounting for the systematic errors that long-term datasets often exhibit. In this study, several MCP methods are used and compared with each other, and are introduced in the following.

Classical approaches

In the classical approaches, the correlation between the short- and long-term dataset is done based on the full overlapping period of both datasets. Three correlation methods using linear regression are described in the following and are used in this study:

- Classic linear regression (Clas-LinReg): It is commonly used in wind resource assessment to calculate the correlation in the MCP methodology (Carta et al., 2013).

- Sector-wise linear regression (Sec-LinReg): Modification of the classic linear regression method, which considers a division of the wind data into individual wind direction sectors. In this study, a binning in 30° sectors is applied.

- SpeedSort (King and Hurley, 2005): Wind speed values are sorted in ascending order step-wise before performing classic linear regression. The size of the steps are set to $1\,\mathrm{m\,s^{-1}}$ in this study.

Advanced approaches

The advanced approaches in this study for predicting wind speed conditions at a specific site use ML models as regression tools within the MCP framework. The input data are divided into training, testing, and validation datasets. In this case, the data from

9

the overlapping time period are randomly selected within a year to create a split of 70 % training and 30 % test data. Although this reduces the amount of sample size, it still captures all relevant characteristics - such as daily and seasonal cycles - due to the randomness of the selection process. The choice of sub-sample regression models is based on the findings of Schwegmann et al. (2023) using the Python package *scikit-learn (sklearn)* (Pedregosa et al., 2011), and includes:

165   – ML linear regression (ML-LinReg): Similar to classic linear regression, but the training data are randomly selected from the overlapping time period. No additional features are included.

   – Random Forest (RForest) Regressor (Grömping, 2009): An ensemble-based regression method that combines multiple uncorrelated decision trees. Each tree has a limited depth and is trained on a different subset of the training dataset. The final prediction is obtained by averaging the outputs of all individual trees, each given equal weight.

170   – AdaBoost Regressor (Solomatine and Shrestha, 2004; Freund and Schapire, 1997; Drucker, 1997): The AdaBoost Regression is an ensemble-based method that uses decision trees. The main difference to the RForest method is that each tree has only one node and two leaves. Additionally, individual weights are applied to combine the results from each tree into a single final prediction. Data points whose prediction appears more complex than others are weighted stronger and the next tree grows on the weighted previous tree, taking its errors into account. The goal of this method is to combine

175   multiple weak learners into a more accurate and robust model.

   – K-Nearest-Neighbours (KNN) Regressor (Fukunaga and Narendra, 1975): KNN is a simple ML approach, which is based on the average of the k-nearest neighbors of all features in the reference dataset, weighted by their distance of similarity.

In order to find the best combination of defined model parameters - such as the number (k) of nearest neighbors (KNN) or the
180 optimal tree depth (RForest), we use the GridSearchCV function provided by the *scikit-sklearn* package (LaValle et al., 2004) for model training. The negative mean squared error is used as the evaluation score. The parameter combination with the lowest score is then used for performing the predictions.

~~As~~ The wind speed related and atmospheric state descriptive ~~features for the advanced ML approaches,~~ variables wind direction, temperature, pressure, time of day~~and seasonality of the ERA5 dataset are used partly decomposed into sine and~~
185 ~~cosine to account for the circular nature of the quantity~~, and seasonality are used as features for the advanced ML approaches. The variables with circular nature (wind direction, time of day, months) are decomposed into sine and cosine components (Table 3).

2.3.2   Hindcast ensemble range

The impact of inter-annual variability on the wind resource is analyzed by considering the maximum ensemble range of wind
190 speed predictions. This is done as follows (Fig. 5): In a first step, for each individual year within the multi-year overlap period (red hatched) a wind speed prediction with MCP is performed. The resulting wind speed predictions (hindcasts; Fig. 5a, gray

Table 3. Input features derived from the reference dataset for the KNN, Random Forest, and AdaBoost regression models.

Input feature
100 m wind speed
Sine and cosine of 100 m wind direction
2 m air temperature
Mean sea level pressure
Sine and cosine of time of day
Sine and cosine of month
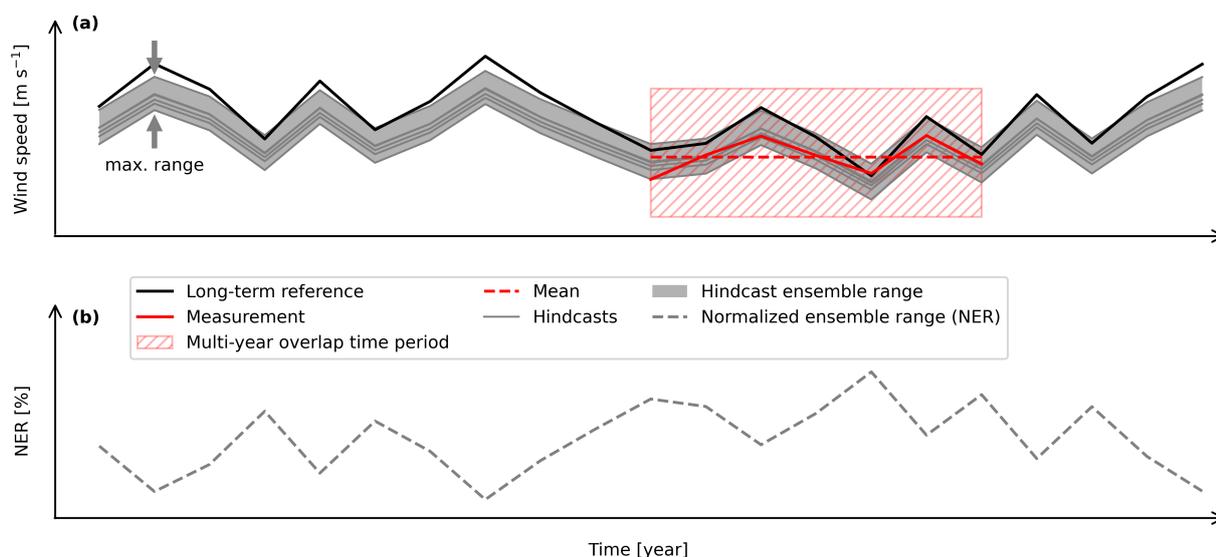
Figure 5. Schematic illustration of the max. hindcast ensemble range: (a) Annual mean of measurement (red) and long-term reference (black) data with multi-year overlap time period (red hatched). Annual mean wind speed predictions (hindcasts, gray lines) based on individual years of the multi-year overlap period, along with the hindcast ensemble range (gray shaded area). (b) By the measured mean wind speed data (red dashed) normalized hindcast ensemble range (NER [%], gray dashed line).

lines) spanning the time range of the long-term reference data, vary due to year-to-year wind variability, with higher variability at a site leading to a wider spread.

Note, that the long-term correction covers periods both before and after the measurement period, we use the term "hindcast" because the entire corrected time series lies in the past from today's perspective. To fully capture the impact of inter-annual wind variability on the hindcast, the maximum range of the hindcast ensemble is analyzed (Fig. 5a). This maximum hindcast

195

ensemble range (ER) is defined as the difference between the maximum and minimum wind speeds in the ensemble for each year. The time average of these over all years is referred to as MER (**m**ean max. **e**nsemble **r**ange).

To ensure comparability between sites, the maximum **e**nsemble **r**ange is **n**ormalized (NER [%], Fig. 5b) by the mean measured wind speed of the multi-year overlap period (Fig. 5a, red dashed). Accordingly, the **m**ean **n**ormalized max. **e**nsemble **r**ange (MNER [%]) is the time average of the NER values. A high (low) MER, NER, or MNER value indicates a broader (narrower) maximum ensemble range, which corresponds to greater (lower) inter-annual variability at the site. This entire procedure is repeated for each site and each correlation method used in the MCP approach. The MCP itself is based on hourly ERA5 data and measured wind speed data.

The wind power density (WPD) is calculated by $WPD = 0.5\rho v^3$, where v represents wind speed. The air density ρ is determined by $\rho = p_0(RT)^{-1}$, with p_0 being the surface pressure, R as the specific gas constant of dry air ($R = 287.058$ J(kg·K)$^{-1}$), and T is the the surface air temperature at 2 meters. The WPD-MNER is calculated similarly to the MNER of wind speed (as described in the previous paragraph). The factor between WPD-MNER and MNER of wind speed is calculated by the quotient: WPD-MNER / MNER of the wind speed.

3   Results

3.1   ~~Wind climatology of investigated sites~~

~~The selected sites exhibit varying wind climates, which are outlined below. In addition, the differences between the reference time series (ERA5) and the measurements are examined. Wind direction distribution of measurement (blue) and ERA5 reanalysis (orange) data for the time period 2010–2016. Note: Overlapping sectors lead to a brownish color. Sites are sorted by terrain complexity from simple to very complex from top left to bottom right. Wind speed distribution of measurement (blue) and ERA5 reanalysis (orange) data for the time period 2010–2016. Note: Overlapping bars lead to a brownish color. Sites are sorted by terrain complexity from simple to very complex from top left to bottom right. A Weibull fit (solid lines) is performed and shape (λ m s^{-1}) and scale (k) parameters are provided in the legend except for cases with distinct non-Weibull, e.g. bimodally distributed met mast winds. The wind characteristics of the measurement data and the ERA5 reanalysis data for the multi-year overlap time period of 2010–2016 for all sites used in this study are illustrated in Figure 3 and Figure 4. At most sites, the measured wind direction is predominantly westerly (Fig. 3) with varying intensity, except in Hegyhátsál and Boulder, where northern winds are also more frequent. Sites with simpler terrain exhibit a more uniform wind direction distribution (e.g. Cabauw) compared to those with complex terrain (e.g. Butler Grade). The measured wind speedfor most sites (e.g. Cabauw, Falkenberg, Boulder) follows a Weibull distribution, while some (e.g. Goodnoe Hills) display bimodal distributions, which are common in highly complex terrain due to redirection of the larger-scale flow.~~

~~When comparing reanalysis data with the measurements, the main wind direction distribution is generally captured by the ERA5 data. But, in some cases, deviations occur between wind direction sectors, with wind directions shifted to an adjacent sector (e.g., Karlsruhe, Chinook, Butler Grade). At the Megler site, an underrepresentation (overrepresentation) of the easterly (northwesterly) winds can be observed in the ERA5 data. This could be attributed to the highly variable wind direction sectors~~

230 ~~that are influenced by terrain characteristics (Table 2). At the Boulder site, the ERA5 data indicate a shift of nearly 90°~~
~~compared to the measured wind direction. This can be attributed to the sites's location at the foot of the Rocky Mountains,~~
~~where local geographical features strongly influence wind patterns. Furthermore, the ERA5 reanalysis data overrepresent lower~~
~~wind speeds and underrepresent higher wind speeds at certain sites (Fig. 4). At the three most complex sites, lower wind~~
~~speeds are overrepresented, while the highest wind speeds are almost completely absent in the ERA5 data. Conversely, at~~
235 ~~sites Hegyhátsál, Chinook and Megler, lower wind speeds are slightly underrepresented, while higher wind speeds occur more~~
~~frequently in ERA5 than in the measurements. The differences between ERA5 and measurement data may arise from choosing~~
~~the nearest ERA5 grid point to the mast location, which may not accurately reflect the site characteristics due to ERA5's~~
~~resolution.~~

3.1 Wind speed hindcast based on classic linear regression

240 To assess the impact of inter-annual wind variability on long-term referencing, a wind speed hindcast ensemble derived from individual years spanning 2010 to 2016 is analyzed using classical linear regression as the correlation method in the MCP process. The hindcast ensemble is illustrated in Figure 6 for several sites in Europe (a) and in the US (b), respectively.

Those sites are affected by different weather regimes due to their geographical locations, which shape the inter-annual variations in the wind climate. Central European sites follow a similar pattern, while eastern sites like Obninsk and Hegyhátsál

245 diverge due different i.e. more continental climate. Similarly, in the US, Butler Grade, Goodnoe Hills, and Chinook located west of the continental divide share a common pattern that is different to sites further east.

The dispersion of the individual hindcasts within the ensemble range varies depending on the site (Fig. 6a and b, shaded areas). The density distribution of the individual hindcasts within the ensemble is displayed to the right of the corresponding time series for each site (Fig. 6). At the sites Falkenberg and Cabauw, there are only small differences between the individual

250 hindcasts, resulting in a narrow density distribution (Fig. 6a). However, at most remaining sites broader density distributions are found. While, e.g., at site Chinook the broad density distribution is caused by evenly distributed hindcasts. At other sites such as Megler and Boulder in the US as well as Hegyhátsál and Obninsk in Europe, individual outliers dominate the ensemble spread. In most cases these outliers are related to the years 2010 and 2011 in the MCP process (see ~~below~~Section 3.3.1). Furthermore, some sites, e.g. Butler Grade and Hamburg have a bimodal density distribution with individual hindcasts clustering at the upper

255 and lower limits of the ensemble. At the site Hamburg, the first (last) years of the multi-year overlap period are related to the upper (lower) hindcasts, which could be connected to a decreasing trend in the measurements for the period 2010 to 2016 (not shown).

Considering the ensemble range of the sites, the mean max. hindcast ensemble range (MER [$\mathrm{m\,s^{-1}}$]) over the period 1950 to 2020 is analyzed and varies between 0.1 and 0.7 $\mathrm{m\,s^{-1}}$ depending on the site (Table 4, MER). To enhance the comparison

260 of the sites, the normalized hindcast ensembles range (NER [%]) for each site from 1950 to 2020 is illustrated in Figure 6c and the corresponding time averaged values are presented in Table 4 (MNER [%]). The MNER has values between 0.9 % to ~~14~~13.6 % with a certain clustering of sites (Fig. 6c). The lowest values of less than 2 % can be found at the sites Cabauw (simple terrain) and Falkenberg (heterogeneous terrain). Most of the sites exhibit ranges between 3 % and 7 % (simple to
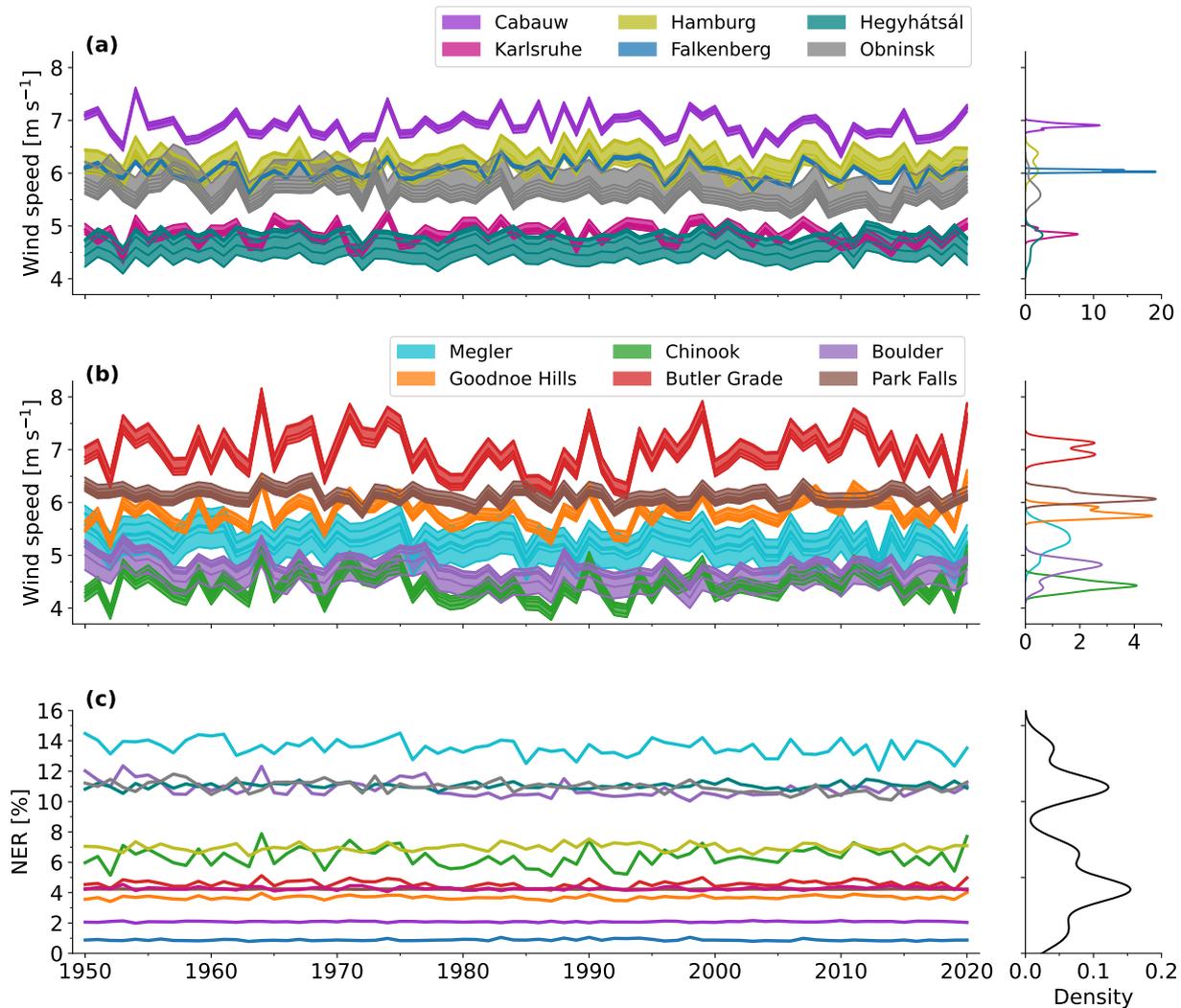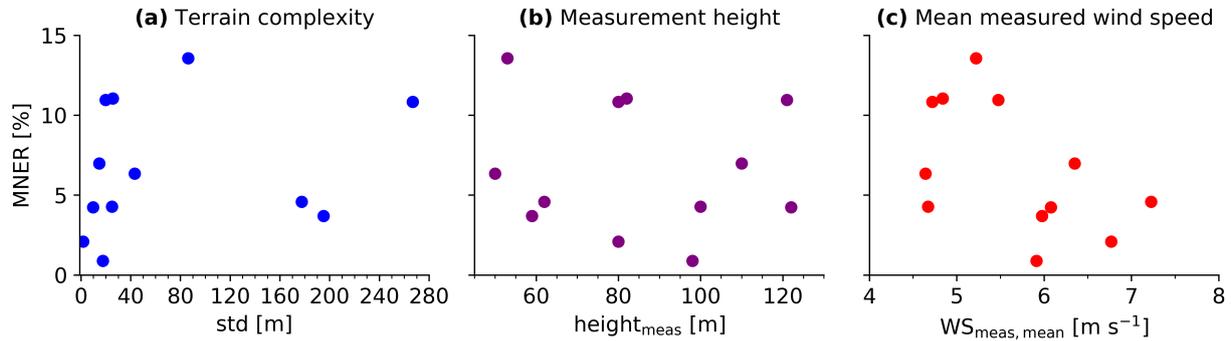
Figure 6. Maximum wind speed hindcast ensemble range (shaded) and individual hindcasts (solid lines) for sites in Europe **(a)** and the United States **(b)** are based on individual years from 2010 to 2016 using classical linear regression in the MCP method. **(c)** The normalized max. hindcast ensemble range (NER [%]) is normalized with the averaged measured wind speed from 2010 to 2016.

(very) complex), grouped into a cluster from 3 % to 5 % and another from 6 % to 7 %. Three of the remaining sites (Boulder

265   (very complex terrain), Hegyhátsál (complex terrain) and Obninsk (heterogeneous terrain) cluster around 11 %, while Megler (complex terrain) exhibits the highest value of approx. ~~14~~13.6 %. Regarding a potential connection between terrain complexity and MNER, it appears that a clear statistical relationship remains unclear (Fig. 7a). Furthermore, there is no indication of a correlation between the MNER and the used measurement height (Fig. 7b) or the measured average wind speed (Fig. 7c). Excluding the years 2010 and 2011, which were mentioned above as being responsible for a larger ensemble range at certain

Table 4. Overview of the mean max. hindcast ensemble range (MER [m s^{-1}]) ~~for the time period from 1950 to 2020~~ and its normalization (MNER [%]) by the mean measured wind speed ~~normalized max. hindcast~~ The ensemble ~~range (MNER %) for~~ is based on the individual years within the time ~~period from~~ intervals 2010 to 2016 and 2012 to 2016.

	MER [m s^{-1}]			MNER [%]		
Site	2010–2016	2012–2016	Difference	2010–2016	2012–2016	Difference
Cabauw	0.14	0.1	0.04	2.1	1.5	0.6
Park Falls	0.26	0.11	0.15	4.2	1.8	2.4
Hamburg	0.44	0.41	0.03	7.0	6.6	0.4
Falkenberg	0.05	0.04	0.01	0.9	0.7	0.2
Obninsk	0.6	0.23	0.37	11.0	4.3	6.7
Karlsruhe	0.2	0.15	0.05	4.3	3.3	1.0
Hegyhátsál	0.53	0.34	0.19	11.0	7.0	4.0
Chinook	0.29	0.29	0.00	6.3	6.4	-0.1
Megler	~~0.74~~0.71	~~0.38~~0.35	0.36	~~14.1~~13.6	~~7.2~~6.5	~~6.9~~7.1
Butler Grade	0.33	0.32	0.01	4.6	4.5	0.1
Goodnoe Hills	0.22	0.19	0.03	3.7	3.3	0.4
Boulder	0.51	0.26	0.25	10.8	5.5	5.3

Figure 7. Correlation between MNER [%] (mean max. hindcast ensemble range) and terrain complexity **(a)** as the standard deviation (std [m]) for a 10 km x 10 km grid around the met mast, measurement height **(b)** and mean measurement wind speed **(c)**, respectively, based on the overlapping time period 2010 to 2016.
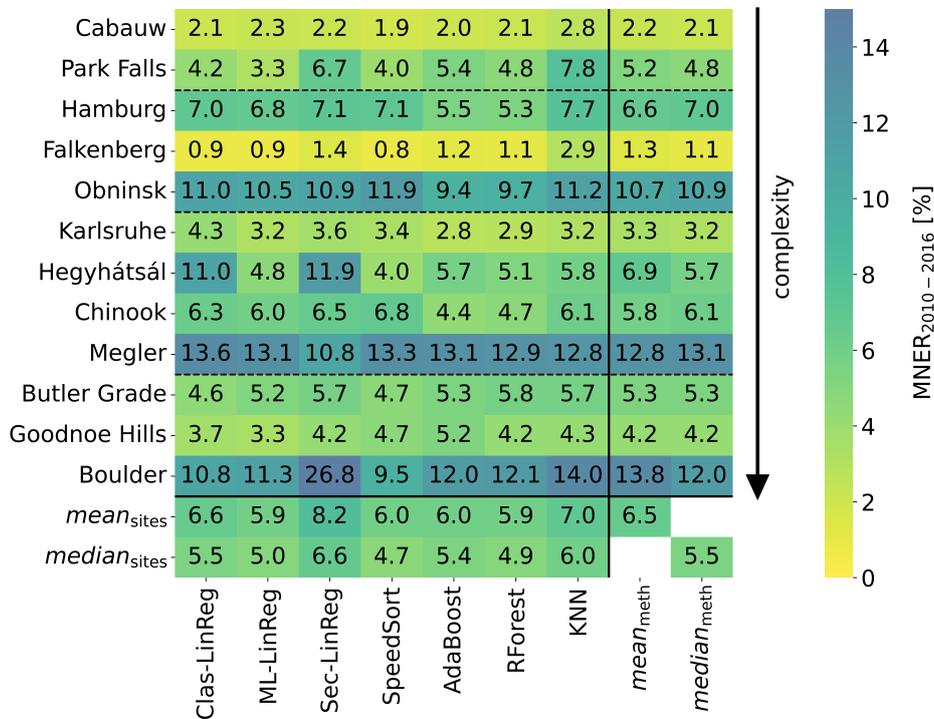
270 sites, the MER reduces to 0.1 to ~~0.4~~0.35 m s^{-1} and the MNER reduces to values from 0.7 % to ~~7.2~~7.0 %, whereas the four sites Megler, Boulder, Hegyhátsál, Obninsk decrease most to values between 4 and 7 % (Table 4). Especially at the site Megler, the

two years ~~almost~~ more than doubled the MNER [%]. The impact of the years 2010 and 2011 is further analyzed and discussed in greater detail in Section 3.3.1 using various MCP approaches.

The site's wind climate and the deviation of an individual year from the wind climate of a site emerge as significant factors on the wind speed hindcast based on classic linear regression. In the following sections, further classical and complex regression methods are analyzed whether they can reduce the influence of inter-annual variability in the long-term referencing process.

3.2 Intercomparison of MCP approaches

In addition to classical linear regression, other regression models in MCP are used for long-term referencing. Moreover, advanced machine learning (ML) methods become increasingly attractive as correction methods in MCP. This section expands the previous analysis to include six additional regression models, including advanced ML methods. The aim is to evaluate the performance of the various methods and assess whether any of these methods, particularly the advanced ML methods, have the potential to reduce uncertainty in estimating inter-annual variability. Furthermore, the analysis compares the methods with respect to their sensitivity across different wind climatologies.

	Clas-LinReg	ML-LinReg	Sec-LinReg	SpeedSort	AdaBoost	RForest	KNN	$mean_{meth}$	$median_{meth}$
Cabauw	2.1	2.3	2.2	1.9	2.0	2.1	2.8	2.2	2.1
Park Falls	4.2	3.3	6.7	4.0	5.4	4.8	7.8	5.2	4.8
Hamburg	7.0	6.8	7.1	7.1	5.5	5.3	7.7	6.6	7.0
Falkenberg	0.9	0.9	1.4	0.8	1.2	1.1	2.9	1.3	1.1
Obninsk	11.0	10.5	10.9	11.9	9.4	9.7	11.2	10.7	10.9
Karlsruhe	4.3	3.2	3.6	3.4	2.8	2.9	3.2	3.3	3.2
Hegyhátsál	11.0	4.8	11.9	4.0	5.7	5.1	5.8	6.9	5.7
Chinook	6.3	6.0	6.5	6.8	4.4	4.7	6.1	5.8	6.1
Megler	13.6	13.1	10.8	13.3	13.1	12.9	12.8	12.8	13.1
Butler Grade	4.6	5.2	5.7	4.7	5.3	5.8	5.7	5.3	5.3
Goodnoe Hills	3.7	3.3	4.2	4.7	5.2	4.2	4.3	4.2	4.2
Boulder	10.8	11.3	26.8	9.5	12.0	12.1	14.0	13.8	12.0
$mean_{sites}$	6.6	5.9	8.2	6.0	6.0	5.9	7.0	6.5	
$median_{sites}$	5.5	5.0	6.6	4.7	5.4	4.9	6.0		5.5

Figure 8. Comparison of the mean max. normalized ensemble range based on the multi-year overlap time period 2010–2016 ($MNER_{2010-2016}$ [%], illustrated by the color scale) as obtained by different methods (indicated below) and for varying sites (indicated on the left). The sites are sorted by the terrain complexity from lowest to highest (black arrow); terrain complexity categories (simple, heterogeneous, (very) complex) are separated by dashed horizontal black lines.

The MNER based on the years 2010 to 2016 is illustrated for all sites and correlation methods in Figure 8, where the Clas-
LinReg column matches that in Table 4. Across all sites and the six additional methods, the MNER varies from less than 1 %
to approximately 14 %, with an outlier at Boulder with 27 % for sector-wise linear regression. Aside from this outlier, the same
MNER is observed as for classical linear regression of the previous section. Averaged across all sites and correlation methods
(including Clas-LinReg), the MNER is 6.5 % (median: 5.5 %).

Regardless of the correlation approach, the MNER tends to be higher at certain sites, with no clear dependence on terrain
complexity (Fig. 8): The lowest $mean_{\mathrm{meth}}$ (mean across all methods) is found for the sites Falkenberg (heterogenous terrain),
Cabauw (simple terrain) and Karlsruhe (complex terrain) with 1 % to 3 %. Most of the other sites reveal a MNER from approx.
4 % up to 7 % including all complexity types (except simple) and measurement heights. The highest $mean_{\mathrm{meth}}$ values are found
for Obninsk (heterogeneous terrain), Boulder (very complex terrain) and Megler (complex terrain) with MNERs of approx.
11 % to 14 %, what is similar to the results from the previous section (Sect. 3.1) for the Clas-LinReg except for Hegyhátsál.

Inter-comparing the seven methods, the simple methods ML-LinReg and SpeedSort as well as the advance models RForest
and the AdaBoost exhibit a MNER averaged across all sites ($mean_{\mathrm{sites}}$) of about 6 %. But, considering the $median_{\mathrm{sites}}$, the
SpeedSort method exhibits the lowest value of 4.7 %, followed by ~~ML-LinReg (5~~RForest (4.9 %), ~~RForest (5.2~~ML-LinReg
(5 %) and AdaBoost (5.4 %). The Clas-LinReg as well as KNN and the Sec-LinReg are most affected by the inter-annual
variability, resulting in a $mean_{\mathrm{sites}}$ MNER between ~~6.7~~6.6 % to ~~8~~8.2 %.

More specifically, *KNN* indicates the highest MNER across multiple sites. While the KNN method is fast and simple, there
is no indication that it effectively reduces the impact of inter-annual wind variability in hindcast wind speed predictions at any
site.

The *Sec-LinReg* indicates varying affects on inter-annual variability, depending on the site, data quality and availability. There
are indications that the high values in the MNER result from a limited sample size of wind data for specific wind directions
related to site-specific climate conditions or general data loss due to measurement failures rather than inter-annual variability.
The Megler and Boulder sites exhibit significantly different terrain characteristics for varying wind directions (Table 2). But,
while the MNER decreases for Megler, the Boulder site, including the wind direction, results in a MNER of 26.8 %. Details
reveal an insufficient sample size at Boulder to establish a representative correlation function within the MCP process for certain
sectors (not shown). There is strong evidence that this issue is linked to the nearly 90-degree shift between the measurements
and ERA5 data (Fig. 3), arising from the complex terrain that is not accurately represented by the nearby ERA5 data point
used. At the sites Park Falls and Hegyhátsál, there is a loss of wind data (Fig. 2) due to measurement failures in 2010/11 and
2013/14 (Hegyhátsál only), which adversely impacted the accuracy of wind speed predictions.

The *SpeedSort* has a normalized ensemble range averaged over all sites, which is about ~~0.7~~0.6 percentage points lower than
that of Clas-LinReg. At the sites such as Hegyhátsál, Park Falls and Boulder, where the ensemble range for Sec-LinReg is
higher, SpeedSort provides lower values and could be a suitable method in such cases.

The advanced ML approaches *RForest and AdaBoost* have the potential to reduce the MNER compared to the Clas-LinReg
method, with the reduction varying by site and depending on terrain complexity. The reductions occur in heterogeneous (except

17

Falkenberg) and complex terrain, while no reduction is noted in simple and very complex terrain. This could be due to the limited number of sites; a larger number of sites per complexity category would be beneficial to gain further insights.

320   Comparing the *Clas-LinReg* with the *ML-LinReg* method, the normalized ensemble range results in almost all sites in similar values. Only at some sites, the values for the ML linear regression are slightly reduced (up to 1 %), while at the Hegyhátsál site, a reduction of 6 % is observed.
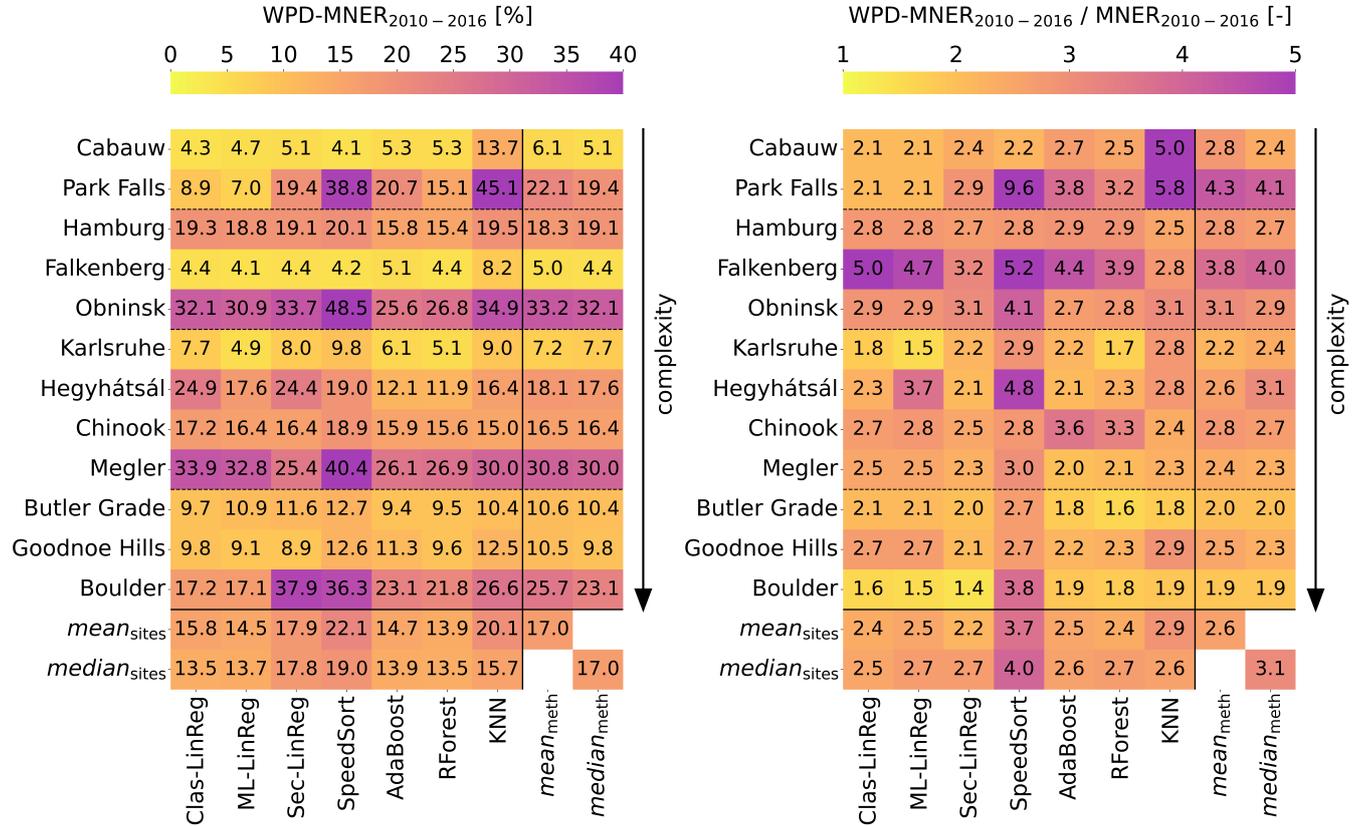
Figure 9. Left: Comparison of the mean max. normalized ensemble range based on the multi-year overlap time period 2010–2016 for wind power density (WPD-MNER$_{2010-2016}$ [%], illustrated by the color scale) as obtained by different methods (indicated below) and for varying sites (indicated on the left). The sites are sorted by the terrain complexity from lowest to highest (black arrow); terrain complexity categories (simple, heterogeneous, (very) complex) are separated by dashed horizontal black lines. Right: Figure setup like left, but for the factor between WPD-MNER$_{2010-2016}$ and MNER$_{2010-2016}$ for wind speed.

The relationship between wind power density and wind speed is cubic and therefore, uncertainty in estimating wind speed significantly impacts the uncertainty of wind power density and annual energy potential. The translation of wind speed 325   uncertainty to annual energy potential uncertainty is not consistently defined. Holtslag (2013) state that 1 % uncertainty in wind speed is related to 1.8 % in annual energy potential. EMD (2025) give the rough orientation that in regions with low wind speeds (6 - 7 m s^{-1}), wind speed uncertainty should be tripled for annual wind potential uncertainty while it is only doubled for

higher wind speeds (about $8\,\mathrm{m\,s^{-1}}$) and only 1.5 times at regions with even higher wind speeds (about $9\,\mathrm{m\,s^{-1}}$). To connect the MNER of wind speed (Fig. 8) to a more comprehensive energy parameter, the MNER for wind power density (Fig. 9, left) is
330   calculated. Additionally, the factor (Fig. 9, right) between the MNER of wind speed and wind power density is determined. The results indicate that, on average, a factor of approximately three (2.6) is achieved. Further, the factor between MNER of wind speed and wind power density does not increase with complexity. But, the studied sites have averaged wind speeds ranging from approximately $5\,\mathrm{m\,s^{-1}}$ to $7\,\mathrm{m\,s^{-1}}$ (Table 1), which suggests that the roughly tripled uncertainty aligns with the results from EMD (2025).

335   Overall, the impact of inter-annual wind variability vary between 1 % and 14 %, depending on the site and method. Advanced ML methods have the potential to reduce the MNER, but the magnitude depends on the site. The impact of inter-annual variability average across all sites and methods is 6.5 % (median 5.5 %). The observed tripled uncertainty in WPD-MNER underscores the importance of accurately estimating wind speed and reducing wind speed uncertainties.

3.3   Variation of multi-year overlap time period

340   Varying the multi-year overlap period is crucial for assessing result robustness across different time frames. Thus, the overlap period will be varied in the following subsections. First, it will be shortened to analyze the impact of years with deviating wind climate (Sect. 3.3.1), and second, it will be extended to examine result consistency over time (Sect. 3.3.2).

3.3.1   Impact of years with a deviating wind climate

In Sect. 2.2, it was observed that the hindcasts based on the measurement years 2010 and 2011 deviate at many sites. At the
345   beginning of 2010, a strong El Niño event took place, followed by a La Niña event towards the end of 2010 and throughout 2011 (NOAA/National Weather Service, 2025) which can affect wind speed and has a high impact on inter-annual wind variability (e.g., Mohammadi and Goudarzi, 2018; Li et al., 2010). Thus, in the further analysis both years, 2010 and 2011, are excluded, and the impact of their exclusion on long-term referencing is evaluated.

     After the exclusion of 2010/11, the MNER lies between approx. 1 % to 7 % depending on the site and method (Fig. 9 10).
350   Averaged across all methods, the lowest $mean_{\mathrm{meth}}$ values of less than 3 % are related to the simple and heterogeneous sites Cabauw, Falkenberg and Park Falls, but also to the complex site Karlsruhe. The sites Obninsk (heterogeneous terrain) and Goodnoe Hills (very complex terrain) follow with $mean_{\mathrm{meth}}$ values below 5 %, while the remaining sites have $mean_{\mathrm{meth}}$ MNER between 5 % and 6 %.

     Averaged across all sites and methods, the MNER is 4.2 % (Fig. 9 10) representing a decrease of 2.3 %-points (Fig. 10 11)
355   compared to the time period from 2010 to 2016 (Fig. 8). This can be especially attributed to the decrease of the MNER at sites with high to very high MNER for the time period 2010 to 2016 (Obninsk, Megler, Boulder). Averaged across all methods, the MNER decreases between 2.9 2.7 % (Park Falls) and 8 7.9 % (Boulder) (Fig. 10 11). The highest decrease is observed for the Sec-LinReg method at the site Boulder with a reduction of 21.5 %.

     Furthermore, excluding the years 2010/11 at the Hegyhátsál site significantly reduces the ensemble range, particularly for
360   the Clas-LinReg and Sec-LinReg methods. For clarification, the deviations between the difference values in Table 4 and Figure
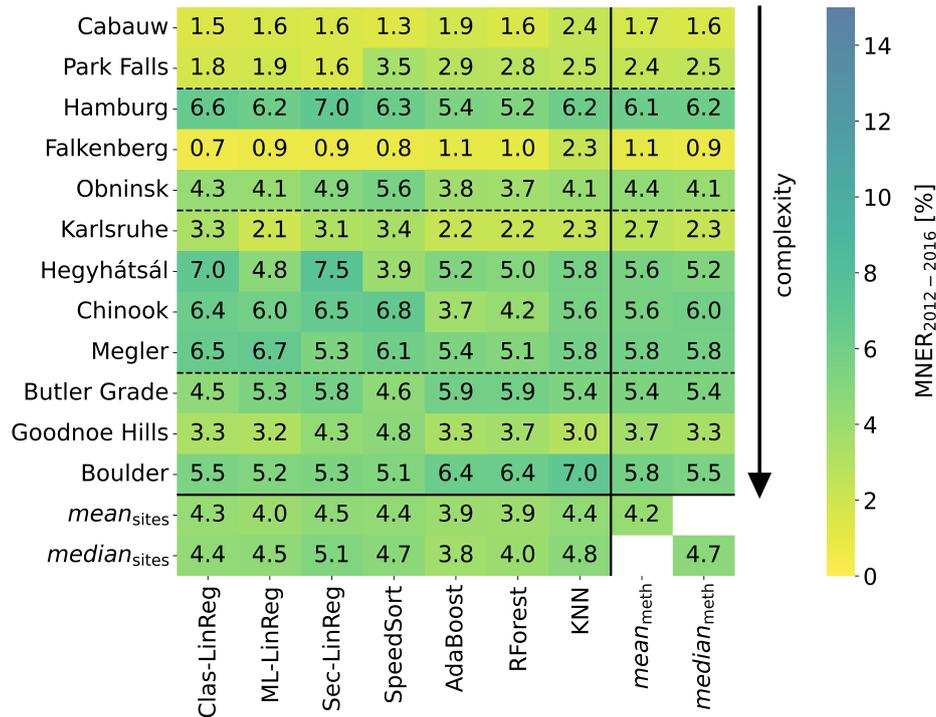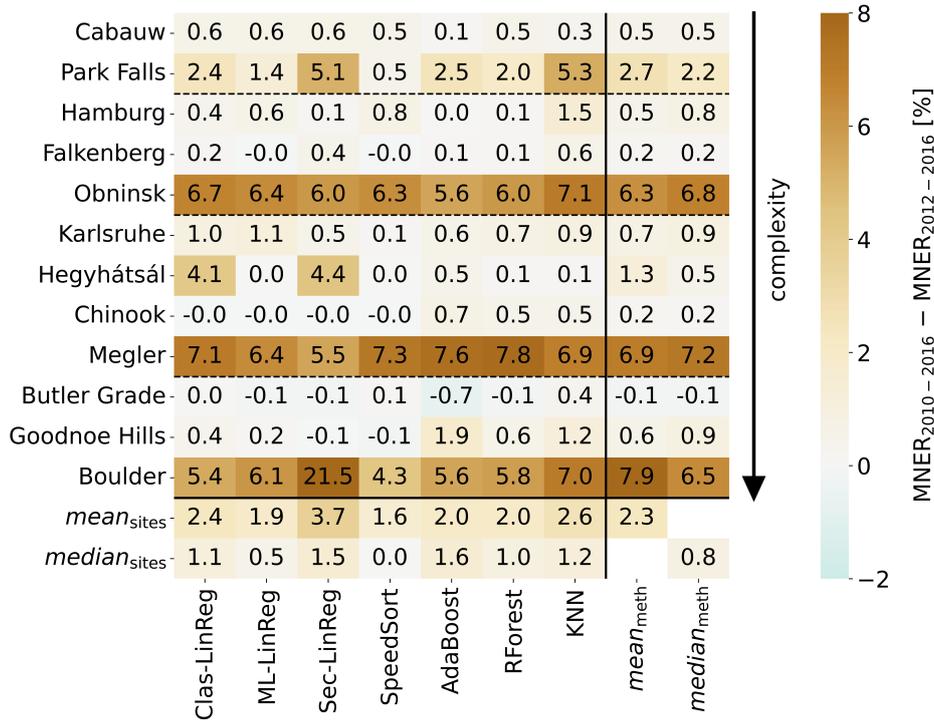
Figure 10. Comparison of the mean max. normalized ensemble range based on the multi-year overlap time period ~~2010–2016~~ 2012–2016 (MNER$_{2012-2016}$ [%], illustrated by the color scale) as obtained by different methods (indicated below) and for varying sites (indicated on the left). The sites are sorted by the terrain complexity from lowest to highest (black arrow); terrain complexity categories (simple, heterogeneous, (very) complex) are separated by dashed horizontal black lines.

~~10~~ 11 for the Clas-LinReg are related to rounding errors. Additionally, the previously noted difference between ML-LinReg and Clas-LinReg persists, with ML-LinReg demonstrating an approximately two percentage points lower MNER (4.8 % vs. 7.0 %).

The RForest and AdaBoost can reduce the impact of inter-annual variability on long-term prediction, especially at sites with
365 heterogeneous (e.g., Hamburg) and complex (e.g., Chinook) terrain complexity (Fig. ~~9~~10). At simple and very complex sites, ML-LinReg or classical approaches could be suitable options.

Averaged across all sites ($mean_{sites}$), there are only slight deviations between the MNER for the various correlation methods (0.6 %-points) compared to the time period 2010 to 2016 (2 %-points). The lowest $mean_{sites}$ are achieved by the advanced ML methods AdaBoost and RForest (3.9 %), followed by the ML-linear regression with ~~4.1~~4 %. The Sec-LinReg reduces most
370 with approx. ~~3.6~~3.7 %-points excluding 2010/11, but is still the method with the highest value (4.5 %).

In general, it can be concluded that the impact of individual years with a differing wind climate is higher than the choice of the correlation method within the MCP approach. However, by selecting the correlation method in long-term referencing, the impact of inter-annual variability can be reduced, but the magnitude depends on the site.
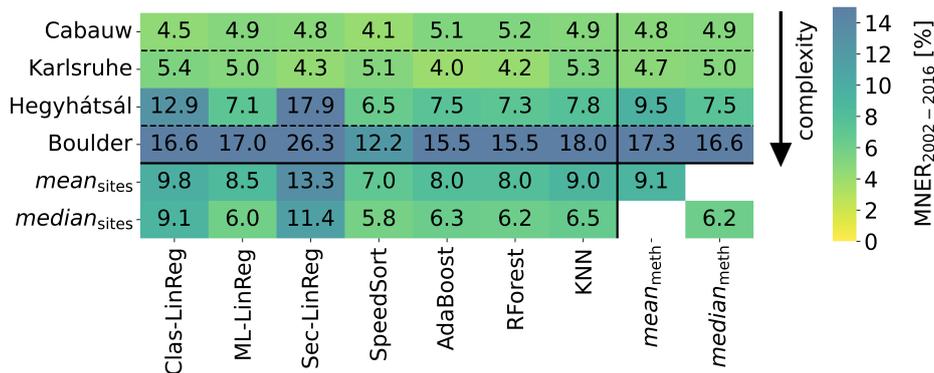
Figure 11. Comparison of the difference of the mean max. normalized ensemble range based on the multi-year overlap time period 2010–2016 ($MNER_{2010-2016}$ [%]) and ~~2010–2016~~ 2012–2016 ($MNER_{\sout{2010-2016}2012-2016}$ [%], illustrated by the color scale) as obtained by different methods (indicated below) and for varying sites (indicated on the left). The sites are sorted by the terrain complexity from lowest to highest (black arrow); terrain complexity categories (simple, heterogeneous, (very) complex) are separated by dashed horizontal black lines.

3.3.2 Long-term multi-year overlap

375 In the previous sections, time periods of five (Sect. 3.1) and seven (Sect. 3.3.1) years were analyzed. This section extends the multi-year overlap time period to 15 consecutive years from 2002 to 2016 which allows analyzing the robustness of the ensemble range due to inter-annual variability. But, extending the multi-year period by such extent limits the analysis to four sites, as not all provide data for such a long time period (Fig. 2). In Figure ~~11~~ 12 the MNER is presented for the four remaining sites.

380 Depending on the site and method, the MNER for the 15 year period results in values from 4 % up to 18 % (except for Sec-LinReg at Boulder). Comparing the MNER based on the seven and 15 years multi-year overlap periods, the MNER more than doubles at the simple terrain site Cabauw. At the complex site Karlsruhe, the MNER averaged over all methods increases by approx. 1.5 %-points. For the remaining sites Hegyhátsál (heterogeneous terrain) and Boulder (very complex terrain), the increase of the $mean_{\mathrm{meth}}$ is approx. 2 % and 4 % points, respectively. This increase in the MNER for the 15-year overlap,

21

	Clas-LinReg	ML-LinReg	Sec-LinReg	SpeedSort	AdaBoost	RForest	KNN	$mean_{meth}$	$median_{meth}$
Cabauw	4.5	4.9	4.8	4.1	5.1	5.2	4.9	4.8	4.9
Karlsruhe	5.4	5.0	4.3	5.1	4.0	4.2	5.3	4.7	5.0
Hegyhátsál	12.9	7.1	17.9	6.5	7.5	7.3	7.8	9.5	7.5
Boulder	16.6	17.0	26.3	12.2	15.5	15.5	18.0	17.3	16.6
$mean_{sites}$	9.8	8.5	13.3	7.0	8.0	8.0	9.0	9.1	
$median_{sites}$	9.1	6.0	11.4	5.8	6.3	6.2	6.5		6.2

Figure 12. Comparison of the mean max. normalized ensemble range based on the multi-year overlap time period 2002–2016 ($MNER_{2002-2016}$ [%], illustrated by the color scale) as obtained by different methods (indicated below) and for varying sites (indicated on the left). The sites are sorted by the terrain complexity from lowest to highest (black arrow); terrain complexity categories (simple, complex, very complex) are separated by dashed horizontal black lines.

indicating that the full inter-annual variability is not captured by the 7-year period and that the increase is likely linked to further years with deviating wind climate within the extended time period.

Comparing the methods at the Cabauw site, the SpeedSort has the lowest MNER at 4.1 % (Fig. 11 12). For the complex Karlsruhe site, the advanced ML methods AdaBoost and RForest have the lowest value, followed by the Sec-LinReg. At the remaining and (very) complex sites Hegyhátsál and Boulder, the SpeedSort method followed by the advance ML methods Random Forest and AdaBoost have the lowest MNER. In conclusion, the SpeedSort, Random Forest and AdaBoost methods have the lowest MNER for the extended 15 years multi-year overlap time period.

Bringing together the findings from the previous analysis, the results offer valuable insights into the inter-annual variability related sensitivities of classical and advanced ML methods in the long-term referencing (MCP) process. They illustrate that selecting the appropriate method can substantially reduce the uncertainties tied to inter-annual variability, where advanced ML methods can make a contribution. Moreover, the analysis highlights the crucial need for thorough exploratory data analysis of the measurements before the application of the long-term correction method. The choice of the appropriate MCP method(s) may then depend on:

– Data availability and correlation of measurement and reference dataset.

– The deviation of the wind climate in measurement period from the climate mean.

– The general wind climate.

– The complexity of the terrain at the individual site.

4 Discussion & Conclusions

This study contributes to an improved assessment and a reduction of uncertainty due to inter-annual variability in the long-term referencing process. It quantifies the impact of inter-annual - i.e. the year-to-year wind variability - on hindcast wind speed predictions using seven different correlation methods of varying complexity within the measure-correlate-predict (MCP) approach. This analysis benefits from quality-controlled, multi-year tall tower measurement data from twelve wind energy-relevant sites at the height of modern wind turbines. While these data are rare, they offer valuable insights into wind energy related questions. The study provides an overview of the uncertainty related to the inter-annual wind variability across different site complexities and emphasizes the substantial impact that a single measured wind year - as commonly done in the wind energy industry - can have on long-term referencing. Specifically, the following recommendations and conclusions can be made:

- **Quantification of inter-annual variability:** Inter-annual variability has been investigated in several studies. In the overview of Lee and Fields (2021), the uncertainties related to inter-annual variability of wind are indicated to be up to 10 %, with most studies reporting uncertainties around 5 %. This is generally consistent with our findings, which indicate an average of 6.5 %. Notably, the obtained variability across different sites and correlation methods spans from 1 % to 14 %, suggesting that relying solely on such an average estimate may be inadequate. A generalizable dependency of these values on terrain complexity could not be identified. This could be an indication that the variability is more dependent on other characteristics of the local wind climate such as the homogeneity of the wind direction distribution, the land-use in the surroundings or the quality of the measurement data (see "Database" below). Therefore, it is essential to investigate factors that can contribute to narrowing this range.

- **Translation into energy parameter:** The analysis of the relationship between wind power density and wind speed reveals a cubic dependency, indicating that even minor uncertainties in wind speed can lead to significant variations in annual energy potential. Transferring the results from wind speed uncertainty to wind power density, an average factor of 2.6 was found; but, it could be individually higher or lower. Our study, which encompasses sites with wind speeds between $5\,\mathrm{m\,s^{-1}}$ and $7\,\mathrm{m\,s^{-1}}$, aligns with the EMD (2025) assertion of approximately tripled uncertainty under these conditions. This significant factor highlights the critical need for accurate uncertainty estimations in wind speed predictions, with the goal of enhancing the reliability of energy yield prediction in the wind energy sector.

- **Time range dependency:** Excluding years with a wind climate deviating strongly from the climate mean reduces the average by approx. two percentage points to 4.2 %. Furthermore, based on a smaller subset of tall tower measurements, we investigated the impact of using individual yearly measurements from a dataset of 15 years of measurements vs. data from seven years of continuous measurements. These results indicated that the full variability was not covered in the shorter dataset even after the application of different methods in the MCP correction process. Other - purely model-based - studies indicate significant multidecadal variability Wohland et al. (2019).

- **Database:** Discrepancies between model based long-term reference and short-term measurement data can vary depending on factors such as horizontal and vertical resolution or model physics but also the measurement data quality at each site. ~~Data gaps~~ Long-term data from reanalyses enable the analysis of historical climate parameters and patterns, particularly their variations and temporal developments, and are therefore essential for estimating wind resources. Although reanalyses largely depict fluctuations and patterns in a uniform manner, differences in model physics and resolutions lead to discrepancies. Expanding the study to include various reanalyses could help quantifying the sensitivity to long-term data and reducing related uncertainties. Data gaps in measurement data can have an impact on the results with varying magnitude depending on the methodology. Thus, this study demonstrates the importance of long-term - i.e. multi-year - well documented observational data from tall meteorological towers across various regions (terrain complexity, geographic location). In their comprehensive review, (Carta et al., 2013) summarize that the database has a higher influence on uncertainties than the choice of the correlation method in the MCP process. Our study aligns with these findings; however, it also demonstrates that employing the appropriate correlation method can be beneficial to reduce uncertainties in the long-term referencing process associated with inter-annual variability.

- **Correlation method:** We show that the selection of the correlation method in the MCP process can have an influence on the variability of the long-term corrected wind speed. This is particularly true when the atmospheric conditions in the short-term measurement data are not representative of the wind climate, for example, due to extraordinary weather. Accordingly, we recommend not to rely just on one method, but rather test different methods and pay attention to the variability.

The sector-wise linear regression method exhibits high sensitivity to the input data. Insufficient data per sector can result in inadequate correction functions. To address this issue, a dynamic sector detection approach as proposed by (Riedel et al., 2001) and (King and Hurley, 2005) can be used. We did not investigate such corrections to enable a comparison of a simple implementation without any site specific or other manual adaptations.

On average, the hindcasts based on the more advanced ML methods Random Forest (Grömping, 2009) and AdaBoost (Drucker, 1997) as well as the SpeedSort (King and Hurley, 2005) method are less affected by the inter-annual variability in the measurement data. Reductions appear to occur in heterogeneous and complex terrain, while no reduction is noted in simple and very complex terrain. To further substantiate this, a larger number of sites per complexity category would be beneficial. ~~Further, in~~ When translating wind speed uncertainties to wind power density uncertainty, SpeedSort appears to be the most affected, with a factor of 3.7, while Random Forest and AdaBoost exhibit similar performance to the classic methods, with a factor of approximately 2.5. Further, changes in the implementation of machine learning methods can influence the outcome. The impact of the changes in the implementation is highly dependent on the specific site, resulting here in average values ranging from 0 to 0.3 for RForest and reaching up to 0.6 for AdaBoost. Additionally, in this study, the same set of features was used within the advanced ML approaches for all sites, enabling a comparison of the correlation methods. Site specific calibrations of ML methods could potentially contribute to further improving the individual results but reduce the comparability.

In summary, this study offers supportive insights into uncertainties in long-term referencing due to inter-annual variability, which is particularly relevant in the planning and operation phase of wind farms. Besides a thorough investigation of long-term hub height wind measurement data, only a small number of sites with suitable multi-year data could be investigated. In case further high-quality measurement data should become available, future studies could extend to further areas of various terrain complexity. The impact of climate change was not considered in this study, further studies could make use of climate model data in the MCP process to cover future wind variability.

References

Baker, R. W., Walker, S. N., and Wade, J. E.: Annual and seasonal variations in mean wind speed and wind turbine energy production, Sol. Energy, 45, 285–289, https://doi.org/10.1016/0038-092X(90)90013-3, 1990.

Bakhoday-Paskyabi, M.: Predictive Analysis of Machine Learning Schemes in Forecasting Offshore Wind Speed, in: J. Phys.: Conf. Ser., https://doi.org/10.1088/1742-6596/1669/1/012017, 2020.

Barber, S. and Nordborg, H.: Improving site-dependent power curve prediction accuracy using regression trees, in: J. Phys.: Conf. Ser., vol. 1618, https://doi.org/10.1088/1742-6596/1618/6/062003, 2020.

Bass, J. H., Rebbeck, M., Landberg, L., Cabré, M. F., and Hunter, A.: An improved measure-correlate-predict algorithm for the prediction of the long term wind climate in regions of complex environment: Final Report JOR3-CT98-0295, in: Renewable Energy Systems Ltd (UK), Risø National Laboratory (Denmark), Ecotecnia (Spain), University of Sunderland (UK), https://hdl.handle.net/10779/lincoln.24373816.v1, 2000.

Basse, A., Callies, D., Grötzner, A., and Pauscher, L.: Seasonal effects in the long-term correction of short-term wind measurements using reanalysis data, Wind Energy Sci., 6, 1473–1490, https://doi.org/https://doi.org/10.5194/wes-6-1473-2021, 2021.

Bodini, N. and Optis, M.: How accurate is a machine learning-based wind speed extrapolation under a round-robin approach?, in: J. Phys.: Conf. Ser., vol. 1618, https://doi.org/10.1088/1742-6596/1618/6/062037, 2020.

Carta, J. A., Velázquez, S., and Cabrera, P.: A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site, Renew. Sustain. Energy Rev., 27, 362–400, https://doi.org/https://doi.org/10.1016/j.rser.2013.07.004, 2013.

Corotis, R. B.: Stochastic modelling of site wind characteristics, Tech. Rep. Final report, DOE - Department of Energy's, USA, https://doi.org/https://doi.org/10.2172/7257559, 1976.

Daniel, L. O., Sigauke, C., Chibaya, C., and Mbuvha, R.: Short-term wind speed forecasting using statistical and machine learning methods, Algorithms, 13, https://doi.org/https://doi.org/10.3390/a13060132, 2020.

Davis, K. J., Bakwin, P. S., Yi, C., Berger, B. W., Zhao, C., Teclaw, R. M., and Isebrands, J. G.: The annual cycles of CO2 and H2O exchange over a northern mixed forest as observed from a very tall tower, Glob. Change Biol., 9, 1278–1293, https://doi.org/https://doi.org/10.1046/j.1365-2486.2003.00672.x, 2003.

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N. N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Sastre-Marugán, M., Sīle, T., Trei, W., Žagar, M., Badger, J., Gottschall, J., Sanz Rodrigo, J., and Mann, J.: The Making of the New European Wind Atlas – Part 2: Production and Evaluation, Geosci. Model Dev., 13, 5079–5102, https://doi.org/10.5194/gmd-13-5079-2020, 2020.

Drucker, H.: Improving regressors using boosting techniques, in: ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning, vol. 97, pp. 107–115, ISBN 978-1-55860-486-5, 1997.

EMD: WindPRO 4.2, User Guide, EMD International A/S, https://help.emd.dk/knowledgebase/content/windPRO4.2/c3-UK_windPRO4.2_ENERGY.pdf, Page 41, Figure 40, 2025.

FGW e.V.: Technical Guidelines for Wind Turbines - Part 6 (TG6) Determination of Wind Potential and Energy Yield - Revision 12, Technical guideline, Fördergesellschaft Windenergie und andere Dezentrale Energien, https://wind-fgw.de/themen/richtlinienarbeit/, 2023.

Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, J. Comput. Syst. Sci., 55, 119–139, https://doi.org/10.1006/jcss.1997.1504, 1997.

Früh, W.-G.: Long-term wind resource and uncertainty estimation using wind records from Scotland as example, Renew. Energy, 50, 1014–1026, https://doi.org/https://doi.org/10.1016/j.renene.2012.08.047, 2013.

Fukunaga, K. and Narendra, P. M.: A branch and bound algorithm for computing k-nearest neighbors, IEEE Trans. Comput., 100, 750–753, https://doi.org/10.1109/T-C.1975.224297, 1975.

Gottschall, J. and Dörenkämper, M.: Understanding and mitigating the impact of data gaps on offshore wind resource estimates, Wind Energy Sci., 6, 505–520, https://doi.org/10.5194/wes-6-505-2021, 2021.

Grömping, U.: Variable importance assessment in regression: linear regression versus random forest, Am. Stat., 63, 308–319, https://doi.org/10.1198/tast.2009.08199, 2009.

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., and Söderberg, S.: The making of the New European Wind Atlas – Part 1: Model sensitivity, Geosci. Model Dev., 13, 5053–5078, https://doi.org/10.5194/gmd-13-5053-2020, 2020.

Hamlington, B. D., Hamlington, P. E., Collins, S. G., Alexander, S. R., and Kim, K.-Y.: Effects of climate oscillations on wind resource variability in the United States, Geophys. Res. Lett., 42, 145–152, https://doi.org/https://doi.org/10.1002/2014GL062370, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Holtslag, E.: Improved Bankability: The Ecofys position on LiDAR use, https://www.nrgsystems.com/assets/resources/Ecofys-2013-position-paper-on-lidar-use-Whitepapers.pdf, utrecht, the Netherlands, 2013.

Jager, D. and Andreas, A.: NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data); NREL Report No. DA-5500-56489, Tech. rep., National Renewable Energy Laboratory, http://dx.doi.org/10.5439/1052222, 1996.

Justus, C. G., Mani, K., and Mikhail, A. S.: Interannual and Month-to-Month Variations of Wind Speed, J. Appl. Meteorol. Climatol., 18, 913–920, https://doi.org/10.1175/1520-0450(1979)018<0913:IAMTMV>2.0.CO;2, 1979.

King, C. and Hurley, B.: The SpeedSort, DynaSort and Scatter Wind Correlation Methods, Wind Eng., 29, 217–241, https://doi.org/https://doi.org/10.1260/030952405774354868, 2005.

Klink, K.: Trends and Interannual Variability of Wind Speed Distributions in Minnesota, J. Clim., 15, 3311–3317, https://doi.org/10.1175/1520-0442(2002)015<3311:TAIVOW>2.0.CO;2, 2002.

Kohler, M., Metzger, J., and Kalthoff, N.: Trends in temperature and wind speed from 40 years of observations at a 200-m high meteorological tower in Southwest Germany, Int. J. Climatol., 38, 23–34, https://doi.org/https://doi.org/10.1002/joc.5157, 2018.

LaValle, S. M., Branicky, M. S., and Lindemann, S. R.: On the relationship between classical grid search and probabilistic roadmaps, Int. J. Robot. Res., 23, 673–692, https://doi.org/https://doi.org/10.1177/0278364904045481, 2004.

Lee, J. C. Y. and Fields, M. J.: An overview of wind-energy-production prediction bias, losses, and uncertainties, Wind Energy Sci., 6, 311–365, https://doi.org/https://doi.org/10.5194/wes-6-311-2021, 2021.

Lee, J. C. Y., Fields, M. J., and Lundquist, J. K.: Assessing variability of wind speed: comparison and validation of 27 methodologies, Wind Energ. Sci., 3, 845–868, https://doi.org/https://doi.org/10.5194/wes-3-845-2018, 2018, 2018.

Li, X., Zhong, S., Bian, X., and Heilman, W. E.: Climate and climate variability of the wind power resources in the Great Lakes region of the United States, J. Geophys. Res., 115, https://doi.org/https://doi.org/10.1029/2009JD013415, 2010.

570   Martin, L.: Wind Energy – The Facts: A Guide to the Technology, Economics and Future of Wind Power, J. Clean Prod., 18, 1122–1123, https://doi.org/10.1016/j.jclepro.2010.02.016, 2010.

MEASNET: Evaluation of Site Specific Wind Conditions, Technical Report, Measurement Network of Wind Energy Institutes, https://www.measnet.com/wp-content/uploads/2022/09/Measnet_Evaluation-of-Site-Especific-Wind-Conditions_v3-1.pdf, last accessed: 30.04.2025, 2022.

575   Meyer, P. J. and Gottschall, J.: How do NEWA and ERA5 compare for assessing offshore wind resources and wind farm siting conditions?, Journal of Physics: Conference Series, 2151, 012 009, https://doi.org/10.1088/1742-6596/2151/1/012009, 2022.

Mohammadi, K. and Goudarzi, N.: Study of inter-correlations of solar radiation, wind speed and precipitation under the influence of El Niño Southern Oscillation (ENSO) in California, Renew. Energy, 120, 190–200, https://doi.org/https://doi.org/10.1016/j.renene.2017.12.069, 2018.

580   NOAA/National Weather Service: Cold and Warm Episodes by Season, https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php, last access: 27.05.2024, 2025.

Olauson, J.: ERA5: The new champion of wind power modelling?, Renew. Energy, 126, 322–331, https://doi.org/10.1016/j.renene.2018.03.056, 2018.

Optis, M., Bodini, N., Debnath, M., and Doubrawa, P.: New methods to improve the vertical extrapolation of near-surface offshore wind
585   speeds, Wind Energy Sci., 6, 935–948, https://doi.org/https://doi.org/10.5194/wes-6-935-2021, 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, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, https://doi.org/https://dl.acm.org/doi/pdf/10.5555/1953048.2078195, 2011.

Pryor, S. C., Barthelmie, R. J., and Schoof, J. T.: Inter-annual variability of wind indices across Europe, Wind Energy, 9, 27–38,
590   https://doi.org/10.1002/we.178, 2006.

Pullinger, D., Zhang, M., Hill, N., and Crutchley, T.: Improving uncertainty estimates: Inter-annual variability in Ireland, J. Phys. Conf. Ser., 926, 012 006, https://doi.org/10.1088/1742-6596/926/1/012006, 2017.

Ramon, J., Lledó, L., Pérez-Zanón, N., Soret, A., and Doblas-Reyes, F. J.: The Tall Tower Dataset: a unique initiative to boost wind energy research, Earth Syst. Sci. Data, 12, 429–439, https://doi.org/https://doi.org/10.5194/essd-12-429-2020, 2020.

595   Riedel, V., Strack, M., and Waldl, H.: Robust approximation of functional relationships between meteorological data: alternative measure-correlate-predict algorithms, pp. 806–9, ISBN 3-936338-09-4, 2001.

Rohrig, K., Berkhout, V., Callies, D., Durstewitz, M., Faulstich, S., Hahn, B., Jung, M., Pauscher, L., Seibel, A., Shan, M., Siefert, M., Steffen, J., Collmann, M., Czichon, S., Do, M., Lange, B., Ruhle, A., Sayer, F., Stoevesandt, B., and Wenske, J.: Powering the 21st century by wind energy—Options, facts, figures, Appl. Phys. Rev., 6, 031 303, https://doi.org/https://doi.org/10.1063/1.5089877, 2019.

600   Schwegmann, S., Faulhaber, J., Pfaffel, S., Yu, Z., Dörenkämper, M., Kersting, K., and Gottschall, J.: Enabling Virtual Met Masts for wind energy applications through machine learning-methods, Energy AI, 11, 100 209, https://doi.org/10.1016/j.egyai.2022.100209, 2023.

Solomatine, D. P. and Shrestha, D. L.: AdaBoost. RT: a boosting algorithm for regression problems, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), vol. 2, pp. 1163–1168, https://doi.org/10.1109/IJCNN.2004.1380102, 2004.

605   Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., and Nenadic, G.: Machine learning methods for wind turbine condition monitoring: A review, Renew. Energy, 133, 620–635, https://doi.org/https://doi.org/10.1016/j.renene.2018.10.047, 2019.

Velázquez, S., Carta, J. A., and Matías, J. M.: Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: A case study in the Canary Islands, Appl. Energy, 88, 3869–3881, https://doi.org/https://doi.org/10.1016/j.apenergy.2011.05.007, 2011.

610 Watson, S. J., Kritharas, P., and Hodgson, G. J.: Wind speed variability across the UK between 1957 and 2011, Wind Energy, 18, 21–42, https://doi.org/https://doi.org/10.1002/we.1679, 2015.

Wohland, J., Omrani, N. E., Keenlyside, N., and Witthaut, D.: Significant multidecadal variability in German wind energy generation, Wind Energy Science, 4, 515–526, https://doi.org/10.5194/wes-4-515-2019, 2019.

Zhang, J., Chowdhury, S., Messac, A., and Hodge, B.-M.: A hybrid measure-correlate-predict method for long-term wind condition assess-
615 ment, Energy Convers. Manag., pp. 697–710, https://doi.org/https://doi.org/10.1016/j.enconman.2014.07.057, 2014.