WIND ENERGY
*Wind Energ.* 0000; **00**:1–16
DOI: 10.1002/we

RESEARCH ARTICLE

# Probabilistic maximum-value wind prediction for offshore environments

Andrea Staid[1], Pierre Pinson[2], Seth D. Guikema[1]

[1] Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, MD, United States
[2] Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

## ABSTRACT

High wind speeds can pose a great risk to structures and operations conducted in offshore environments. When forecasting wind speeds, most models focus on the average wind speeds over a given time period, but this value alone represents only a small part of the true wind conditions. We present statistical models to predict the full distribution of the maximum-value wind speeds in a three-hour interval. We take a detailed look at the performance of linear models, GAM's, and MARS models using meteorological covariates such as gust speed, wind speed, CAPE, Charnock, MSLP, and temperature, as given by ECMWF forecasts. The models are trained to predict the mean value of maximum wind speed, and the residuals from training the models are used to develop the full probabilistic distribution of maximum wind speed. Knowledge of the maximum wind speed for an offshore location within a given time period can inform decision-making regarding turbine operations, planned maintenance operations, and power grid scheduling in order to improve safety and reliability, and probabilistic forecasts result in greater value to the end-user. The models outperform traditional baseline forecast methods and achieve low predictive errors on the order of 1-2 meters per second. We show the results of their predictive accuracy for different lead times and different training methodologies. Copyright © 0000 John Wiley & Sons, Ltd.

**KEYWORDS**
probabilistic prediction models; maximum value winds; offshore wind

**Correspondence**
Andrea Staid, Johns Hopkins University, Department of Geography and Environmental Engineering, 3400 North Charles St, Ames Hall 313, Baltimore, MD 21218, USA. E-mail: staid@jhu.edu

Received . . .

## 1. INTRODUCTION

Many offshore commercial activities are heavily reliant on weather forecasts for safe and efficient operations. Offshore wind farms are an obvious example; wind predictions are heavily relied upon for planning both farm and power grid operation schedules, but other offshore operations (i.e. oil platforms and shipping) are just as dependent on weather forecasts, and wind forecasts in particular. Knowledge of future wind speeds can improve the reliability, safety, and profitability of many operations. High wind speeds in particular can pose a significant risk to offshore structures and to workers exposed to high-wind conditions. A study of helicopter crashes related to oil and gas operations in the Gulf of Mexico, for example, found that 16% of accidents were due to bad weather, and almost a third of those cases were attributed to high winds [1]. With wind turbines specifically, they are designed to shut off when wind speeds reach a certain threshold, but even below this threshold, frequent operation at high wind speeds can cause significant damage to a turbine. Turbine reliability decreases with increasing exposure to high wind speeds or high accelerations [2, 3]. Managing these risks can be done, in part, by integrating forecasts specifically for maximum wind speeds into the decision-making process when planning offshore operations, and these operations may extend to the fields of wind farm scheduling, wind turbine control, maintenance planning, offshore oil platforms, or shipping operations. In this paper, we take a risk-based view of wind variability and focus on this one aspect of wind behavior: very high wind speeds for offshore applications. We demonstrate the capabilities of simple statistical models to forecast maximum wind speeds, and these predictions can be used to improve safety and operations for wind farms, offshore platforms, and shipping traffic.

### 1.1. Background and Existing Literature

There is a large body of work on wind speed prediction, and it has only continued to grow as the penetration of wind energy increases worldwide. The economic viability of wind is directly dependent on high quality wind forecasts, and this has led to great interest in improving wind-forecasting techniques. Wind is highly variable, and this can make accurate predictions difficult, especially when forecasting for very short time periods. It is usually more straightforward to forecast mean-value wind speeds for larger time scales, and much of the work in this field has focused on these mean-value forecasts and building better models with higher degrees of accuracy. An excellent overview, which details the various methods and techniques used in this field, can be found in a literature review by Giebel et al. [4]. This paper details the different forecasting approaches, including persistence models, numeric weather prediction, statistical models, and combinations of the aforementioned methods. For those interested in the use and application of wind forecasts, a report by Argonne National Laboratory gives a broader look at these issues [5]. This report covers the full spectrum of challenges present when integrating wind power: wind forecasting, subsequent power forecasting, estimating and presenting uncertainty, unit commitment of wind, and power grid operation with wind generation. Of the many proposed methods for wind prediction, strong examples include artificial neural networks, linear models, ensemble methods, and nonparametric approaches [6, 7, 8, 9].

Average-value forecasts are needed for planning, but they fail to capture the variability occurring within the associated time periods being evaluated, which are often one hour or more. The wind speeds within that period can deviate drastically from the mean-values forecasted [10]. Of this variability, maximum wind speeds are particularly important when assessing the risks to structures, workers, and ongoing activities. This risk is more of a concern in offshore environments because the wind speeds are often higher in general.

The field of offshore wind in particular would benefit strongly from maximum-wind forecasts, since farm operations are often scheduled a day in advance and turbines are susceptible to damage from extremely high winds. Power systems are already designed to deal with uncertainties and variability, historically in the realm of load forecasts [11]. Adding in the variability of wind is within the experience of system operators. However, as with load forecasts, wind forecasts are also needed to manage the uncertainty. An accurate knowledge of wind speeds is critical for efficient planning and operation of wind farms [11, 12, 13]. Electric grid operations are scheduled based on predicted mean wind speeds and subsequent power output estimates. Knowledge of the mean wind speed, however, does not give a full depiction of wind behavior, which can often vary greatly within very short time periods [14]. Therefore, in addition to mean value wind predictions, there is also a benefit to be gained from knowing what the maximum value winds will be in a certain time period. A careful operation scheme, where, for example, turbines are preventatively shut down when very high maximum wind speeds are predicted, could greatly increase turbine life and improve the current operating condition of the grid.

Several studies have identified the importance of high-wind forecasting for purposes of informing decisions regarding grid operation and system safety [11, 12]. With the added high-wind forecasts, a wind farm can be operated in such a way to minimize the transients during time periods with expected forays into wind speeds above the turbine threshold, for example. In addition to farm operations, there are considerable safety concerns associated with high wind speeds. Wind farm construction and maintenance are all dependent on acceptable weather conditions, and the safety of the workers


*Wind Energ.* 0000; **00**:1–16 
DOI: 10.1002/we
*Prepared using weauth.cls*

can be severely jeopardized if wind speeds pick up when they werent expected to. The same applies to offshore drilling platforms. Exposed maintenance operations, for example, should not be conducted if wind speeds rise above certain safety thresholds. Scheduled maintenance can be planned for periods with low predicted maximum wind speeds, thus reducing the exposure of the crew and equipment to high, and often dangerous, winds.

The body of work focusing on maximum winds, gust winds, or extreme value winds in short-term wind applications (i.e. on the order of several hours to a day) is much smaller than the work focused on mean-value winds. Gusts are usually defined as a three-second average above a threshold. Maximum winds are similar, but represent only the highest value recorded in a given time-period, without averaging and without a threshold. While the definitions of maximum wind speed and gust wind speed differ, the motivation for forecasting these high-speed wind events is similar. There are risks associated with both high maximum winds and strong gusts, and the literature looks at both.

Extreme-value theory is an often-used method for analyzing extremes of a parameter, but it is usually applied on long time scales, on the order of years [15, 16, 17]. Several studies have analyzed the annual maximum wind values using a generalized extreme value distribution [18, 19]. This type of analysis is useful when deciding on design criteria to ensure that structures can withstand the fifty or one-hundred-year winds, but these values can often be underestimated [20]. In addition, these longer time scales are not useful for many planning operations, since decisions are made for time periods of hours or days ahead. This method has been successfully applied to daily wind speeds, however [21]. Research in the realm of gust predictions has been done by Brasseur using purely physical factors and by Ágústsson and Ólafsson using atmospheric models and highly localized terrain data, among others [22, 23]. Thorarinsdottir and Johnson have developed a model to predict wind gusts using a gust factor and a probabilistic forecast for the maximum wind speed and appearance of gusts [24]. They use nonhomogeneous Gaussian regression to predict the distribution of daily wind speed and gust speed. The probabilistic nature of their models conveys a lot of information to the user, but the forecast is issued as the distribution for a given day in the future, and this time interval is often too long to be of great use when planning operations and maintenance. With a more risk-centric application, Petroliagis and Pinson evaluate the relationship between extreme wind events and medium-term (i.e. on the order of several days) warnings of extreme events [25]. The motivation for their work is similar to the research presented here; advance knowledge of high wind speeds can result in better decision-making regarding the safety of many operations, both onshore and offshore.

### 1.2. Paper Objectives and Structure

In this paper, we build upon the techniques used for predicting average-value wind speeds and instead focus on predicting, probabilistically, the maximum-value winds for a given location with forecast lead times reaching from zero hours out to five days (120 hours). We use simple statistical models and focus on maximum-winds that are not necessarily extreme, according to statistical properties. We compare several models and assess their predictive performance at various lead times using different training methods. The probabilistic predictions accurately capture the variability of the maximum wind speed and convey more useful information to the end user. We present the methodology in section 2. We apply these methods to a dataset of measured and modeled meteorological parameters for a chosen location in the North Sea, as discussed in section 3, and the results are shown in section 4.

## 2. PROBABILISTIC FORECAST METHODOLOGY

Our goal is to produce accurate probabilistic predictions, for both short- and medium-range forecast windows, for the maximum wind speed in a given time interval and a given location. For our purposes, we define maximum wind speed as the highest value sampled at a rate of 1 Hz during a ten-minute interval. By issuing probabilistic forecasts, we capture the variability of the maximum wind speeds and allow the user to factor in this uncertainty when making decisions about offshore operations, whether for a wind farm, offshore platform, or shipping activities. Although we are developing statistical models for the maximum-value wind speeds, our proposed method first determines the expected value, or mean, of the maximum wind speed and then determines the full probabilistic distribution around this mean (of the maximum) wind speed. The models used are trained to predict the expected value of the maximum wind speed. We assume that the maximum wind speed values are Gaussian of the form $\mathcal{N}(\mu, \sigma^2)$ with mean $\mu$ and variance $\sigma^2$. In order to build up the full distribution of wind speed, we estimate the mean and variance parameters separately. The models are developed to predict the mean value of the maximum speed, $\mu$, and we use the parameters of the residual training errors to determine the variance, $\sigma^2$ (or standard deviation, $\sigma$.) The standard deviation of the residual training errors is used as the standard deviation of the new predicted distribution for each individual point forecast issued. The resulting predictions portray a normal distribution for the maximum wind speed in each time interval, so that any probabilistic prediction intervals can easily be calculated.

The model training and prediction process is conducted independently for each lead time. The full dataset is subset according to lead time. This subset is then split into two sequential groups; the first is used to train the models and the second is used to issue predictions. The models are trained on each training set for each lead time and then applied to predict new data points for the same lead time. The models, and therefore the model residuals, are specific to each lead time. The relationships among the covariates and the relative importance of the covariates in the models will likely differ for predictions issued at different lead times. Our method allows for the models to capture these changing interactions and variable importance.

### 2.1. Model Development

For preliminary model evaluation, we tried a large number of statistical model types and variable combinations. Based on the models that performed well in a few sample test cases, we were able to narrow down the field of feasible options. The model types initially tested included linear models, generalized additive models (GAM), random forests, multivariate adaptive regression splines (MARS), bagged classification and regression trees, and support vector machines [26]. For each of the models, variable selection was used to determine the best combination of parameters to be used; the combination of variables included differs for the different model types as a result. The models were compared based on their predictive errors: mean absolute error and root mean squared error. Based on the preliminary testing, three models were identified for further testing. A linear model, a GAM, and a MARS model all performed well. These three were then tested extensively for different forecast times, lead times, and training methods. The formulation of these three models is as follows. Given a vector of inputs, $\boldsymbol{X}_t^T = (X_{1t}, X_{2t}, \ldots, X_{p_t})$ and a response variable, $Y_t$ for lead time $t$ and the number of covariates $p$, the three model types discussed are given by the following formulations:

*Linear Model Formulation:*

$$Y_t = \beta_{0_t} + \sum_{j=1}^{p} X_{j_t}\beta_{j_t} + \epsilon_t$$

where $\beta_{0_t}$ is the intercept term for lead time $t$, the $\beta_{j_t}$'s are the coefficients of each input variable for lead time $t$, and $\epsilon_t$ is the error term of the form $\mathcal{N}(0, \sigma_t^2)$, a normal distribution with zero mean and finite variance, $\sigma_t^2$. This assumes a purely linear relationship between each of the input variables and the response variable, maximum wind speed.

*GAM Formulation [27]:*

$$Y_t = \beta_{0_t} + f_{1_t}(X_{1t}) + f_{2_t}(X_{2t}) + \cdots + f_{p_t}(X_{p_t}) + \epsilon_t$$

where $\beta_{0_t}$ is the intercept term, each $f_{i_t}(X_{it})$ specifies a function of each $X_{it}$, and $\epsilon_t$ is the error term of the form $\mathcal{N}(0, \sigma_t^2)$. The smoother functions take the form of cubic regression splines. The model is fitted by simultaneously estimating all p functions, allowing for nonlinear relationships between some or all of the input variables and the response.

*MARS Formulation [28]:*

$$Y_t = \beta_{0_t} + \sum_{m=1}^{M} \beta_{m_t} h_{m_t}(\boldsymbol{X}_t) + \epsilon_t$$


*Wind Energ.* 0000; **00**:1–16 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
*Prepared using weauth.cls*

where $\beta_{0t}$ is the intercept term, the $\beta_{mt}$'s are the coefficients associated with each basis function, $h_{mt}(\boldsymbol{X}_t)$, $M$ is the number of basis functions, and $\epsilon_t$ is the error term of the form $\mathcal{N}(0, \sigma_t^2)$. The basis functions can be formulated as hinge functions of the form $h(X_t) = (X_{jt} - c)_+$ with $c$ representing the location of the hinge, or as the product of two or more functions of the same form. These allow for nonlinearities of certain input variables or interactions between multiple input variables.

The models were trained using the R software environment, and the functions for linear models, GAMs, and MARS models can be found in the stats, mgcv, and earth packages, respectively [29, 30, 31]. The variable selection process is conducted separately for each model type. We initially assess variables using model fit as the guiding parameter, and well-performing variable combinations are then compared based on predictive accuracy.

### 2.2. Probabilistic Forecasts

Any prediction made is associated with a certain degree of uncertainty; this is inherent in the process. Predictions are often given as one deterministic value; for example, a typical weather forecast may predict that tomorrow will have a high temperature of 36 degrees. The uncertainty in this forecast is not passed along to the public. The uncertainty around a deterministic forecast is often as important as the forecast itself. For this reason, we focus on developing statistical models to provide probabilistic predictions of the maximum-value wind speeds instead of a single, deterministic value. There is a real benefit to be gained through the use of probabilistic weather forecasts, and this has been demonstrated frequently in the literature. Reliable, and even moderately reliable, probabilistic forecasts outperform standard forecasting practices, resulting in lower costs and higher value to the user [32]. When dealing specifically with wind forecasts, it has been shown that including the uncertainty in forecasts results in an increased market value for the forecast itself [33]. This increase in value is important, as it represents one of the three main measures of forecast goodness as described by Murphy [34]. The statistical models aim to produce accurate (i.e. high quality) and consistent forecasts, and the probabilistic aspect of the predictions only serves to increase the value to the user.

As stated earlier, we assume that the maximum wind values are Gaussian of the form $Y_t \sim \mathcal{N}(\mu, \sigma^2)$. The models presented here offer mean-value predictions, $\mu$, for the maximum wind speed in a given time period. The predicted values represent the expected value of the probabilistic density function for the maximum wind speed in that time period. To estimate the distribution around this expected-value point prediction, we use the residuals to obtain the variance, $\sigma^2$ (and standard deviation, $\sigma$) in order to develop the prediction intervals around the expected value. We start by looking at the model residuals from the training set. The residuals are defined as $r_i = y_i - \hat{y}_i$ where $y_i$ is the actual value and $\hat{y}_i$ is the predicted value for all $i$ data points. We fit a normal distribution to the residuals, and determine the standard deviation based on the fitted curve. This standard deviation (of the residuals from each model) is then applied to each prediction made by the model to estimate the full distribution of the errors around the predicted expected value, $\mu$. With the assumption of normality in the errors, the residual standard deviation, and the predicted expected value for the response, the distribution is fully defined for each future observation being predicted. The probabilistic predictions for a given data point can then be issued either as a full distribution or as defined quantiles of the distribution. This method was compared to the standard method for obtaining predictive intervals and found to be in good agreement. The method used here has the advantage of estimating the full distribution directly, and it can be used for any type of statistical model for which the residuals can be easily calculated.

The usual methods for assessing model predictive performance, such as mean absolute error or (root) mean squared error, cannot be used for probabilistic forecasts, since there is no single value of the distribution to compare. Instead, alternative metrics need to be used for evaluating probabilistic forecasts; here, we use the continuous ranked probability score, or CRPS, as our method of comparison. The CRPS compares the cumulative distribution function (CDF) of the prediction with that of the actual. In this case, the actual is simply an observation, but it can still be represented as a CDF. The CRPS is defined as follows:

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 dy$$


DOI: 10.1002/we
*Prepared using weauth.cls*

where $F$ is the CDF of the probabilistic forecast, $x$ is the actual observation, and $\mathbb{1}\{y \geq x\}$ designates the function that takes a value of 1 if $y \geq x$ and 0 otherwise. The CRPS is exactly equivalent to the mean absolute error (MAE) in the case where the forecast is also deterministic.

## 3. DATA AND APPLICATION

In order to demonstrate the proposed models and assess the performance of the probabilistic predictions, we apply the methodology discussed in the previous section to a dataset of forecasted and measured wind and meteorological conditions for a location in the North Sea. We couple measured data from a meteorological tower with forecasts issued for the same location. The forecasted data is used to train and develop the models, and we assess the performance of the predictions against the actual measured data. Both the measured data and forecasted data were obtained for the time period starting in February 2010 and ending in May 2013. Wind data typically cycles on several different temporal scales. The most familiar cycles are daily and seasonal fluctuations. In addition to these, wind data sometimes has longer-scale cycles, lasting a year or more [35]. Having over three years worth of data allows for models that arent dependent on a single cycle, and this amount of data serves as a good representation of typical behavior in the designated area. The details of the data used for our application are discussed subsequently.

### 3.1. Measured Data

Germany has installed three large offshore meteorological towers in order to collect data in designated areas where they are planning for large amounts of future offshore wind development. The three towers are referred to as FINO1, FINO2, and FINO3, and are all located in the waters north of Germany. For the purposes of this paper, we chose to use the FINO1 data because of its long history of data collection and its location in a prime wind-development region. This tower is located in the North Sea waters off the coast of Germany, just north of the border between Germany and the Netherlands. The tower has been collecting data since 2003, and Germany is planning for projects adding up to almost two gigawatts of installed wind capacity in the area around the FINO1 tower in the near future [36].

These towers are collecting a wide variety of data, including measures of wave height, wave direction, pressure, temperature, humidity, lightning events, sea currents, shipping traffic, and, of course, wind. The wind available data includes measures of wind speed and direction at 33, 40, 50, 60, 70, 80, 90, and 100-meter heights. The data is given as ten-minute averages, with the minimum speed, maximum speed, and variance also given for each ten-minute interval [37]. This additional information about the range and variability of the wind speeds give a more detailed picture of the actual wind behavior. With wind turbines increasing in both capacity and size, we chose to use the 100-meter data for our model development and predictions. This is expected to be the closest to turbine hub-height for the next generation of offshore wind turbines. This actual, measured data from the FINO1 tower is used to test and train models in tandem with the forecast data, which is discussed below. Missing values make up 3% of the measured data, but this is a very small percentage overall, and dropping the missing values in model development is not expected to have much of an effect.

### 3.2. Input Weather Forecast Data

In addition to the measured data discussed above, the models also use a number of parameters from weather forecasts issued by the European Center for Medium-Range Weather Forecasts (ECMWF) [38]. These parameters are used as inputs to our models, and we use them to then issue our own predictions for maximum wind values. ECMWF runs their global meteorological models to issue forecasts twice a day, at 00 and 12 UTC. This data can be downloaded for a specific location, with the earths surface divided into small grid cells of 16 kilometers for which the forecast data is determined through a bilinear interpolation from the four points located at the corners of the grid cell as output by the global model; in this case, the data is taken from the grid cell that contains the FINO1 tower. Each of these forecasts contains data looking out five days (120 hours) in three-hour increments. The data used includes the u and v components of wind speed at 10


DOI: 10.1002/we
*Prepared using weauth.cls*

meters and 100 meters, gust wind speed at 10 meters, temperature at 2 meters, mean sea-level pressure, convective available potential energy (CAPE), and Charnock. CAPE represents a measure of atmospheric instability based on the buoyancy of air over a vertical reference frame [39]. Charnock is a means of characterizing sea surface roughness in relation to wind stress, which, in a way, represents part of the relationship between wind and waves [40]. The data available from ECMWF has been shown to have generally high forecast skill scores and significant usefulness when performing further analysis and evaluation [41].

## 4. RESULTS

The models and methods described in section 2 were applied to the data described in detail in section 3. We identify the importance of the training approach when developing models using the meteorological parameters taken from the ECMWF forecasts. We tested both static and sliding training windows of various sizes, and the size and type of training window can make a big difference in model performance. We also assess the performance of the probabilistic predictions for the different lead times and training windows and identify the variables that are the most influential for each model across all lead times. The skill scores of our probabilistic predictions are compared to traditional baseline measures. Our models outperform these baselines and, as expected, offer a higher skill than a deterministic prediction alone. Our probabilistic forecasts are shown to be highly skilled in terms of accuracy. We also provide reliability diagrams for the probabilistic predictions.

In terms of specific model formulation, the linear model is simplest; maximum wind speed is a function of the forecasted 10-meter gust speed, CAPE, and the u and v components of wind speed at both 10 and 100 meters. The GAM includes the same parameters listed for the linear model, with the addition of the Charnock parameter. The MARS model adds another two more variables to those included in the GAM: mean sea-level pressure and temperature at 2 meters. The addition of these other meteorological variables to the wind speed data gives the models more to work with when it comes to predicting high-value wind speeds.

### 4.1. Model Training

With any model development and prediction process, the choice of training data is critical. The training set should be large enough to capture enough information to accurately model any future data complexities that arise in the testing data; larger is generally better if sufficient data is available, but the marginal benefit may be minuscule after reaching a certain size. The size of the training set also depends on model complexity. As the model complexity increases, the model learns to capture even small perturbations in the training data, and the training error decreases. When using the same model to instead look at the test set error, or predictive error, errors tend to decrease only up to a certain point. Beyond that point, additional model complexity leads to an increase in predictive error—the classic bias-variance tradeoff problem [26]. The large dataset that we are working with, over three years worth of data, allows us to evaluate whether there is a significant performance improvement for very large training windows and to answer the question of what an optimal window size would be for this particular case of wind data. It should be noted, however, that even three years of data might not be enough to capture the very large time-scale variations that may exist for wind in certain regions [35]. Our models work by evaluating the relationships of the variables, and if these relationships were to change significantly due to large temporal fluctuations, the models may need to be retrained on more recent data in order to ensure optimal performance.

Here, we focus on test, or predictive, errors as a performance metric and analyze the models predictive accuracy for a wide range of training window sizes under both static and sliding training conditions. Static training windows work by setting aside a subset of the data, training the model on that subset, and then using that model to predict the response variable for the remainder of the data. In the case of weather data, a static window tends to work best with large amounts of data spanning at least a year, since many climatological parameters cycle on a seasonal or annual time scale. Static windows are very simple to use, and the models never need to be retrained. Sliding training windows use only the most


DOI: 10.1002/we
*Prepared using weauth.cls*

recent data when predicting a given response data point. For a given training window size, $n$, a model is retrained for each individual data point to be predicted, using only the $n$ most recent observations for that point in the training set. This results in a custom model for each separate data point. Sliding windows tend to work well with meteorological data when the windows are small, since there is a tendency for the present weather to resemble the recent past weather. For issuing predictions on large amounts of data, the model has to be retrained for each individual data point. This is extremely time consuming, especially when the models become more complex and more computationally intensive to run.

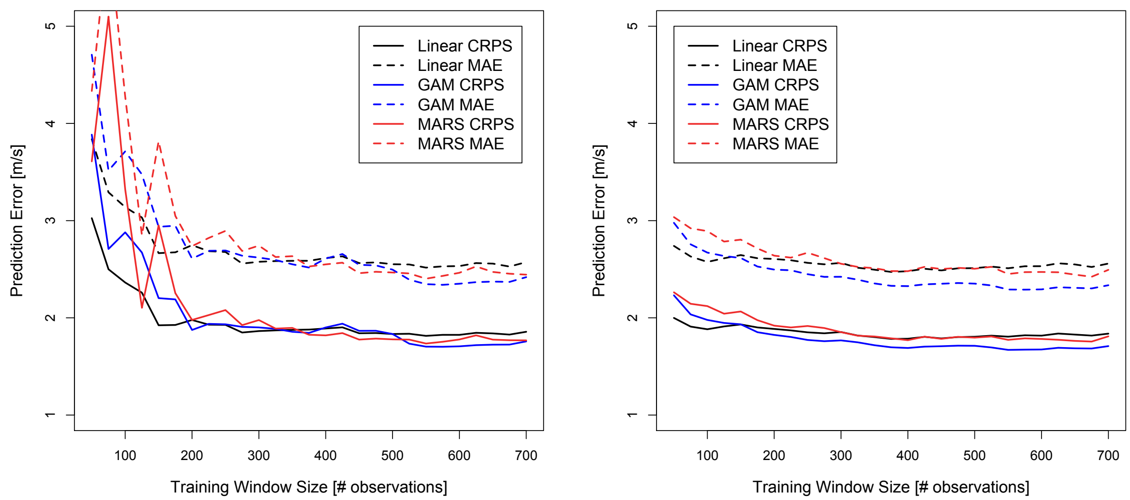
Figure 1 shows the CRPS and MAE for the three models as a function of training window size for both static and sliding windows. Shown here are the plots for a 72-hour lead time. The behavior of the curves is similar for all other lead times, but the values differ as expected: lower errors for shorter lead times and higher errors for longer lead times. In the case of static training windows, the model errors do not settle into any sort of pattern until the training window reaches a size of 200 observations. After 200 observations, the model errors stay fairly stable for increasing window sizes. The plot of the sliding training windows has several distinct differences when compared to the static training plot. The models do not behave erratically for even very small training windows. The smallest window tested here is 25 observations—just shy of a month when using daily predictions for one specific lead-time—and the errors are relatively low and lack the erratic behavior seen in the small static training windows. The errors of the three models are grouped very closely, regardless of training window size, but there is a subtle minimum value for windows of around 600 observations. Although not all lead times are plotted here, the 600-observation minimum is consistent across lead times. For the dataset under evaluation, we show that a sliding training window of 600 observations results in the smallest prediction errors on average. The relatively flat CRPS curve for these models demonstrates the extent to which these models are generalizable. With a sliding training window, the models can accurately predict maximum value wind speeds even in areas where data collection has only been ongoing for a couple of months. The predictive errors are less than 2 meters/second on average for predictions up to three days in the future and close to 1 meter/second on average for day-ahead predictions.

For all lead times, the dashed lines representing the MAE values lie above the solid lines representing the CRPS values. Issuing the predictions as probabilistic distributions, instead of deterministic point predictions, improves the overall value of the information provided in the prediction. The probabilistic predictions are able to convey more information regarding the uncertainty of each prediction, and this additional information results in lower overall errors and greater value when predicting the maximum-value wind speed.

We have identified three different models that all perform well for predicting maximum-value wind speeds in time intervals stretching out to five days. The specific model performance depends on the training method and training window size, and the strength of this dependence depends on model type. In general, if large amounts of data are available, a training window of 600 observations will outperform other window sizes, both for static and sliding windows. The errors, as measured by the CRPS, are shown in Figure 2 for a training window of 600 observations for all models. It should be noted that the spike in the CRPS for the linear model with a lead time of zero is a result of a discrepancy in the dataset. The input forecast data does not include the 10-meter gust parameter for the zero-hour forecasts. This causes the higher-than-expected error, since the linear model would otherwise be heavily dependent on the gust parameter for issuing accurate predictions.

For both static and sliding training methods, the models outperform traditional baseline forecasting methods for large training windows (greater than 200 observations.) The baselines used for comparison are climatology and persistence forecasts. Using climatology as a means of predicting the future is done by looking at the historical distribution of the parameter of interest, and making the assumption that the future will mirror the past. In our case, the climatology forecast, which is used as a baseline, requires the assumption that the distribution of maximum wind speed for each future time period being predicted will match that of the historical distribution of maximum wind speeds for our chosen location. This is an extremely simplistic means of prediction as it ignores any hourly, daily, or seasonal variation but it serves as a useful point of comparison because of its simplicity. The persistence forecast is another simple prediction tool that is often used as a baseline for comparison. A persistence forecast assumes that the conditions in the future will be the same as the conditions at the present moment, i.e. the maximum wind speed that is observed now is the prediction for

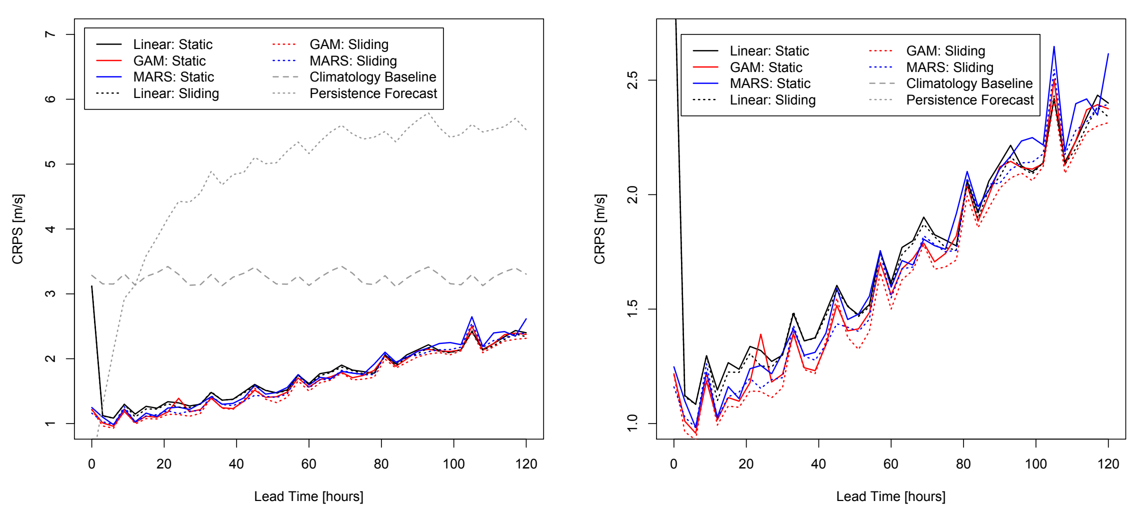

DOI: 10.1002/we
*Prepared using weauth.cls*




(a) static (b) sliding

**Figure 1.** A comparison of maximum wind speed prediction errors for static 1(a) and sliding 1(b) training windows for the three models, shown with a lead time of 72 hours. The shape and behavior of the curves is consistent across lead times, but the values will differ.

the maximum wind speed in the next time period. Persistence forecasts are often very accurate in the very short term (a few hours or less), but the errors tend to become high when trying to predict values even one day in the future. For the case of comparison presented here, the persistence forecast is gathered for each separate five-day issued ECMWF forecast. We take the maximum wind speed observed at the time that the forecast is issued and use that value as the persistence prediction of the maximum wind speed for each time period during the following five days.




(a) full view with baselines (b) a closer look at the model errors

**Figure 2.** CRPS of models trained with 600 observations and baseline measures (climatology and persistence) for all lead times are shown in 2(a). Figure 2(b) gives a closer look at the individual model errors. Note the different y-axis values.

All three of the models presented here outperform the two baseline comparisons, climatology and persistence forecasting, for training windows of 600 observations. Not shown are plots for smaller training windows, but the models all outperform the baseline measures for training windows larger than 200 observations for both static and sliding windows. The linear model, which is the most consistently behaved for the range of training window sizes, outperforms both baseline measures when using sliding training for all window sizes tested. Even with 50 observations and 5-day lead times, the linear model trained with sliding windows results in a lower CRPS than both the climatology and persistence forecasts. In fact, a small sliding window outperforms the baseline measures for all three models. In the case of static windows, however, the training window size becomes significant. Small, static training windows (i.e. 100 observations or less) are not reliable; errors are especially erratic for the GAM and MARS, and they are higher than the climatology forecast at certain lead times.

The simplicity of a linear model makes it much more resilient when it comes to changes in training methods or training window sizes. The GAM and MARS models are slightly more complex, and the higher degree of model complexity results in poor performance when the models are developed using very little data (100 observations of less), and they end up being highly specific to the training data and not as generalizable as a simple linear model when applied to the unknown test data. For the most part, the sliding errors are lower than the static errors for each model. The GAM with a sliding window of 600 observations has the lowest errors over the most lead times. For this reason, the GAM model will be used to demonstrate the predictive performance in the following section.

### 4.2. Prediction Performance

Two sample prediction plots are shown in Figures 3 and 4. Both have been plotted using the GAM with a sliding window of 600 observations, meaning that for each lead time, the previous 600 observations that came prior to the start of the period for which the prediction is being issued were used to train the models. It should be noted that in actuality, the training window size varies slightly due to occasional missing data in the dataset of measured values. Depending on the forecast period, training windows nominally set for 600 observations usually ended up containing between 545 and 595 observations. Figure 3 plots the probabilistic forecasts issued by the GAM for a randomly selected five-day period in April 2013. The shaded areas depict the 10-90% probabilistic predictions for each three-hour time interval. In this instance, the 10-90% prediction interval encapsulates all of the actual measured values for maximum wind speed. The spread of the 10-90% prediction interval varies over time, and the variance tends to increase as the lead-time of the prediction increases. Intuitively, this makes sense as predictions should be more accurate for smaller lead times, since the very near future is more likely to resemble most recent past data available than periods further out will.

Figure 4 shows another set of GAM predictions with a nominal 600-observation sliding training window, but this prediction interval was chosen specifically because it includes the highest observed maximum wind speed value in the entire dataset that we are working with. This was done to test the predictive skill of our models for the most extreme cases and to verify that they still maintained a reasonable level of accuracy far out into the tail of the distribution. For reference, the highest observed maximum-value wind speed is 40.52 meters per second, the median value for maximum wind is 12.61 meters per second, and the 99th percentile is 28.50 meters per second. For the extreme wind values in Figure 4, the 10-90% prediction interval misses only three of the peak values in the actual wind curve. The general shape of the predictions follows the curve of the actual data well, and only three data points, or 7.3% of the measured data, fall outside of the 10-90% prediction interval. Our models show that the prediction accuracy still holds for the most extreme wind values.

Instead of looking at individual five-day forecasts, we can assess the overall performance of the models by looking at their calibration, or reliability. Figure 5 shows a reliability diagram for the predictive performance of the GAM at two different lead times. With a large enough sample size, such as the one analyzed here, the reliability of the predictions can be estimated accurately. While most of the literature on reliability diagrams is applied to ensemble forecasts or binary events, the same principle can be applied here, where we have a parametric distribution for the predictions [42, 43]. For the reliability diagrams in Figure 5, the x-axis represents probability intervals of the density distribution for the predicted


DOI: 10.1002/we
*Prepared using weauth.cls*

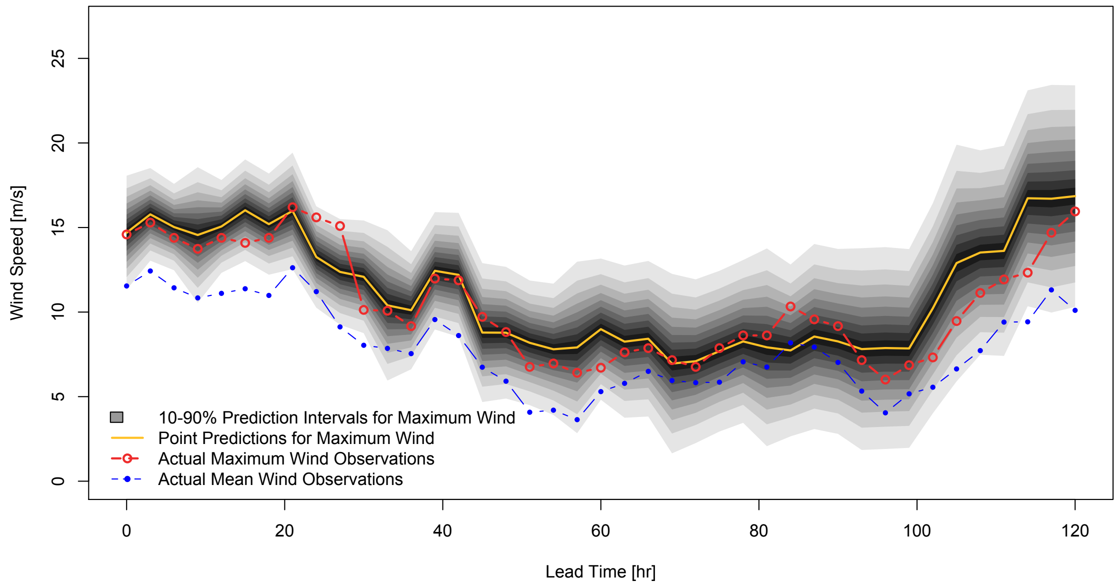



**Figure 3.** Probabilistic GAM predictions for the five-day period starting on 4 April, 2013 at 00 UTC. The shaded grey regions represent the 10–90% prediction intervals for each prediction. Also shown are the actual observations for both maximum and mean value winds for each time period.

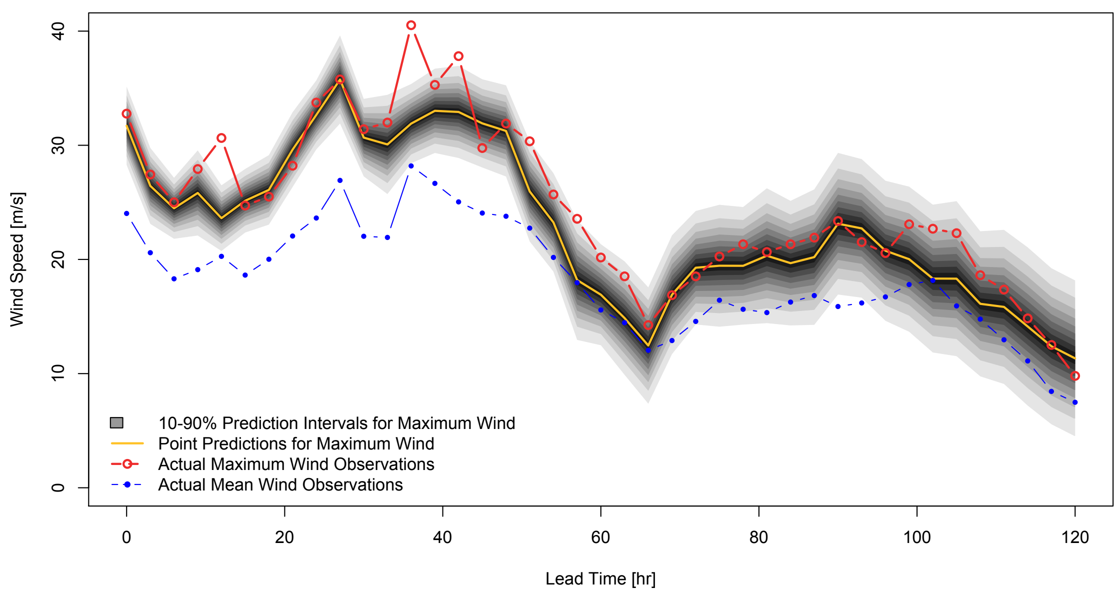
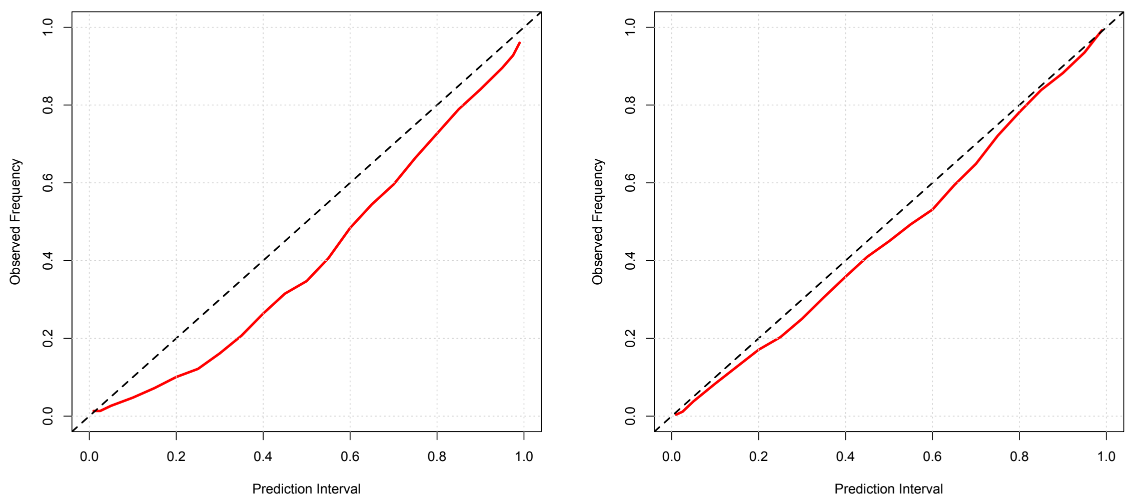



**Figure 4.** Probabilistic GAM predictions for the five-day period starting on 4 January, 2012 at 00 UTC. The shaded grey regions represent the 10–90% prediction intervals for each prediction. Also shown are the actual observations for both maximum and mean value winds for each time period. This interval contains the highest maximum wind speed observed in the available dataset.

values of maximum-wind. The y-axis measures the frequency with which the actual maximum-wind values fall below the associated predicted interval on the x-axis (i.e. what percentage of actual values correspond to the matching percentage of predictions). For perfect forecasts, the points lie along the diagonal, as shown by the dashed line.

There is a recognizable bias in the reliability plot for a 24-hour lead time. The prediction intervals capture a smaller portion of the true data than they ideally should. This bias likely stems from the parametric assumptions made throughout. The bias is significantly reduced for longer lead times, as seen in 5(b). The variance of the predicted distribution gets larger with increasing lead time, since there is more expected variability in predictions that are made further in advance. As shown, this results in a more reliable forecast for a lead time of 120 hours than for a lead time of 24 hours. The bias shown in the diagrams is present, but minimal for longer lead times. The models capture the expected behavior of the maximum-wind speeds well, but the accuracy of the predictions varies with lead time.




(a) 24 hour lead time

(b) 120 hour lead time

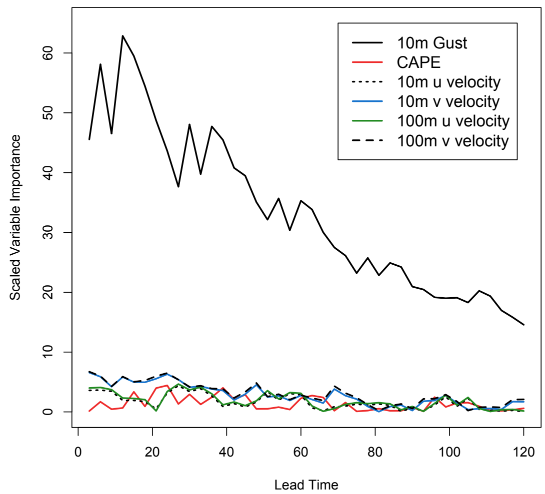
**Figure 5.** Reliability diagram of GAM predictions for 24-hour 5(a) and 120-hour 5(b) lead times. These plot the frequency with which the actual maximum wind values fall below a given prediction quantile.

### 4.3. Variable Importance

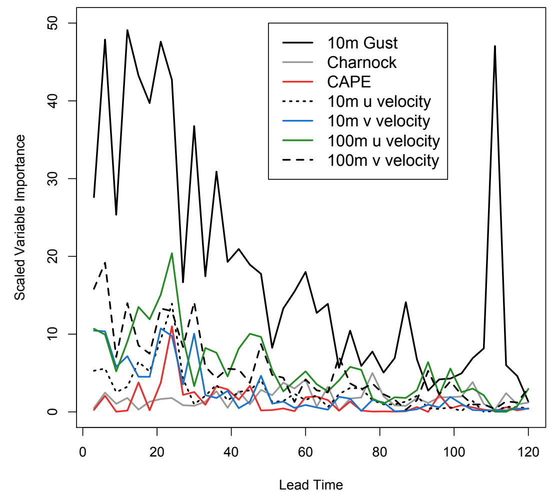
The three models each include a set of covariates, made up of meteorological parameters, that were used in training the models and issuing predictions for new observations. The covariates in each model were chosen based on a variable selection process to minimize errors in model fit and predictive accuracy. The linear model makes use of the following forecasted values: wind velocities at 10 and 100 meter heights, the gust wind velocity at 10 meters, and CAPE. The gust speed is expected to be especially important here, since we are predicting maximum wind speeds, which are more similar to gust speeds than to the average wind speeds given by the other forecasted values. In addition, gust wind forecasts have been previously shown to be significant when predicting extreme weather events [25]. CAPE is a measure of atmospheric instability and can be useful in predicting severe weather. The GAM also includes the Charnock parameter. This is a rough equivalent to a measure of surface roughness specific to offshore conditions, and it is used to characterize the near-surface wind speeds over water. The inclusion of Charnock as a covariate can show the influence of roughness if it is present in the data [44, 45]. Finally, the MARS model adds two more covariates: mean sea-level pressure and temperature at 2 meters. These standard meteorological measures give the MARS model more to work with. The model does not necessarily use all of the variables available; it performs variable selection during the fitting process in order to select the best combination


DOI: 10.1002/we
*Prepared using weauth.cls*

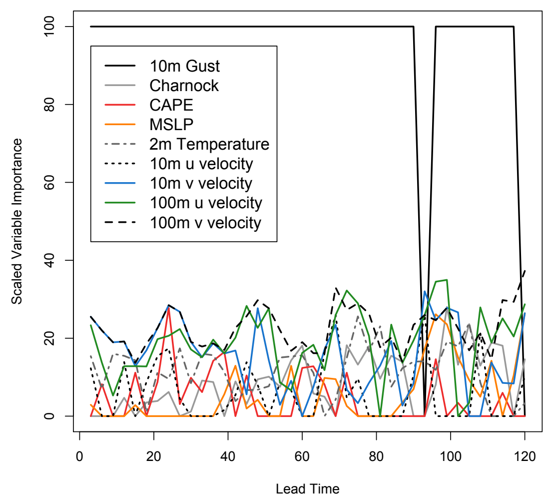
of variables which result in the lowest GCV (generalized cross validation) score. This allows the model to use only those variables that help to improve the performance for each lead time, and the remaining variables are left out.




(a) Linear Model




(b) GAM




(c) MARS

**Figure 6.** Variable importance, as measured relative to other variables in each model, plotted against lead time for linear 6(a), GAM 6(b), and MARS 6(c) models.

As discussed previously, the models are re-trained for each lead-time and training window. This results in different models with different levels of variable importance, and the importance is calculated differently based on model type. For the linear model, the importance is taken as the absolute value of the t-statistic for each variable. For the GAM, it is calculated using the log of the p-values. The MARS model calculates the total reduction in GCV error for each individual variable. Figure 6 shows the variable influence plotted against lead-times for each of the three models. The parameter plotted is the importance of each variable in model fit, not in prediction. There is a general decreasing pattern for the linear model and GAM as lead time increases. This is because predictions become less accurate at longer lead times, and the individual covariates do not contribute as much to error reduction. The MARS model calculates importance quite differently than the linear model or GAM, and the decreasing pattern doesn't show up for this reason. The MARS model can choose to exclude variables in each formulation, and Figure 6(c) shows that this exclusion occurs regularly for some variables, such as CAPE and MSLP. The 10-meter gust variable is the most important overall for all of the models, especially at shorter lead-times. Even with the MARS, the gust variable was of maximum importance in all but two lead times.

## 5. CONCLUSION

The need for high-wind forecasts in offshore environments is clear; turbine safety, worker safety, and efficient power grid operations all stand to benefit greatly from accurate short- and medium-range forecasts of maximum winds, in addition to the established practice of forecasting mean wind speeds. A probabilistic prediction of maximum winds will help to fill in a portion of the knowledge gap regarding the inherent variability and uncertainty of wind. We show that this can be done with a high degree of accuracy using relatively simple models. Three different types of models — a linear model, a GAM, and a MARS model — perform very well for predicting maximum wind speeds, and they are also able to convey the intrinsic variability of the predictions by describing the full parametric distribution around the expected value of maximum wind speed. The choice of the 'best' model depends greatly on the conditions used, available data, and the level of simplicity desired by the user. The CRPS (prediction errors) for the three models are low, but the values do depend on training method and lead-time. Our models achieve errors of less than 1 meter per second at lead times of six hours; even for lead times of five days, the errors are as low as 2.31 meters per second. For short-term planning of offshore operations, errors of this level are small enough to accurately inform decision-makers and to ensure operational safety of structures, components, and workers. Day-ahead predictions are essential for power system operation decisions, and predictions on the order of several days would be useful in the planning of maintenance or construction projects, whether for wind farms or other offshore operations.

We recognize the weakness of the prediction reliability for shorter lead times. The bias introduced is a function of the parametric assumptions made throughout the modeling process. Although the normal assumption is reasonable, it introduces biases. These biases could potentially be reduced by using a different distribution or by not assuming a parametric distribution at all and using non-parametric models. The lack of calibration would likely be improved, but the extent of improvement is not known. Even with this bias, however, the models have great value. They offer accurate predictions with an extremely high degree of simplicity. Our methods are highly skilled and cheap to implement, both in terms of complexity and computational cost. The simplicity presented here allows the methods to be adopted by a larger audience of users, and we leave the alternative modeling techniques as an interesting avenue for future work.

The model details associated with this finding are specific to the dataset used here, but the techniques and methodology are highly generalizable. The ECMWF forecasts provide high-quality data that allows us to accurately model the maximum wind speed, which is not represented in the forecasts. However, the success of these models, or any type of wind-forecast model, depends on the appropriate integration of the predictions into the decision-making process. The information should be communicated early enough and often enough so as to allow planned operations and any resulting dependencies to adapt. The cost savings realized by using such maximum-wind predictions could be calculated for the various industries mentioned that stand to benefit, but it is not assessed here. Such calculations would be interesting extensions and are left for future work.

## ACKNOWLEDGEMENTS

This work was conducted with the support of the WindINSPIRE Project, NSF Grant OISE 1243482. We would also like to acknowledge ECMWF for the input forecast dataset and the German Maritime and Hydrographic Agency and the FuE-Zentrum FH Kiel GmbH for operating the FINO 1 meteorological platform and maintaining the database of measurements that were used to develop and test our models.

## REFERENCES

1. Baker SP, Shanahan DF, Haaland W, Brady JE, Li G. Helicopter crashes related to oil and gas operations in the gulf of mexico. *Aviation, Space, and Environmental Medicine* 2011; **82**(9):885–889.
2. Tavner P, Edwards C, Brinkman A, Spinato F. Influence of wind speed on wind turbine reliability. *Wind Engineering* 2006; **30**(1):55–72.
3. Dueñas-Osorio L, Basu B. Unavailability of wind turbines due to wind-induced accelerations. *Engineering Structures* 2008; **30**(4):885–893.
4. Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS. plus* 2011; .
5. Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. Wind power forecasting: state-of-the-art 2009. *Technical Report ANL/DIS-10-1*, Argonne National Laboratory (ANL) 2009.
6. Alexiadis M, Dokopoulos P, Sahsamanoglou H, Manousaridis I. Short-term forecasting of wind speed and related electrical power. *Solar Energy* 1998; **63**(1):61–68.
7. Riahy G, Abedi M. Short term wind speed forecasting for wind turbine applications using linear prediction method. *Renewable Energy* 2008; **33**(1):35–41.
8. Pinson P, Nielsen H, Madsen H, Kariniotakis G. Skill forecasting from ensemble predictions of wind power. *Applied Energy* 2009; **86**(7):1326–1334.
9. Pinson P, Nielsen HA, Møller JK, Madsen H, Kariniotakis GN. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* 2007; **10**(6):497–516.
10. Trombe PJ, Pinson P, Vincent C, Bøvita T, Cutululis NA, Draxl C, Giebel G, Hahmann AN, Jensen NE, Jensen BP, *et al.*. Weather radars – the new eyes for offshore wind farms? *Wind Energy* 2013; doi:10.1002/we.1659.
11. Jones L, Zavadil R, Grant W. The future of wind forecasting and utility operations. *Power and Energy Magazine, IEEE* 2005; **3**(6):57–64.
12. Smith JC, Milligan MR, DeMeo EA, Parsons B. Utility wind integration and operating impact state of the art. *Power Systems, IEEE Transactions on* 2007; **22**(3):900–908.
13. Milligan MR, Miller AH, Chapman F. *Estimating the economic value of wind forecasting to utilities*. National Renewable Energy Laboratory Golden, CO, 1995.
14. Vincent CL, Pinson P, Giebel G. Wind fluctuations over the north sea. *International Journal of Climatology* 2011; **31**(11):1584–1595.
15. Cooley D, Nychka D, Naveau P. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* 2007; **102**(479):824–840.
16. Towler E, Rajagopalan B, Gilleland E, Summers RS, Yates D, Katz RW. Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research* 2010; **46**(11).
17. Gilleland E, Brown BG, Ammann CM. Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics* 2013; **24**(6):418–432.
18. Brabson B, Palutikof J. Tests of the generalized pareto distribution for predicting extreme wind speeds. *Journal of Applied Meteorology* 2000; **39**(9):1627–1640.
19. Holmes J, Moriarty W. Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics* 1999; **83**(1):1–10.
20. Larsén XG, Mann J. The effects of disjunct sampling and averaging time on maximum mean wind speeds. *Journal of Wind Engineering and Industrial Aerodynamics* 2006; **94**(8):581–602.
21. Friederichs P, Thorarinsdottir TL. Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics* 2012; **23**(7):579–594.

22. Brasseur O. Development and application of a physical approach to estimating wind gusts. *Monthly Weather Review* 2001; **129**(1):5–25.
23. Ágústsson H, Ólafsson H. Forecasting wind gusts in complex terrain. *Meteorology and Atmospheric Physics* 2009; **103**(1-4):173–185.
24. Thorarinsdottir TL, Johnson MS. Probabilistic wind gust forecasting using nonhomogeneous gaussian regression. *Monthly Weather Review* 2012; **140**(3):889–897.
25. Petroliagis TI, Pinson P. Early warnings of extreme winds using the ECMWF extreme forecast index. *Meteorological Applications* 2012; doi:10.1002/met.1339.
26. Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning*, vol. 1. Springer New York, 2001.
27. Hastie T, Tibshirani R. Generalized additive models. *Statistical science* 1986; :297–310.
28. Friedman JH. Multivariate adaptive regression splines. *The annals of statistics* 1991; :1–67.
29. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2012. URL `http://www.R-project.org/`, ISBN 3-900051-07-0.
30. Wood S. *Generalized additive models: an introduction with R*. CRC Press, 2006.
31. Stephen Milborrow Derived from mda:mars by Trevor Hastie and Rob Tibshirani. *earth: Multivariate Adaptive Regression Spline Models* 2012. URL `http://CRAN.R-project.org/package=earth`, r package version 3.2-3.
32. Murphy AH. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review* 1977; **105**(7):803–816.
33. Pinson P, Chevallier C, Kariniotakis GN. Trading wind generation from short-term probabilistic forecasts of wind power. *Power Systems, IEEE Transactions on* 2007; **22**(3):1148–1156.
34. Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and forecasting* 1993; **8**(2):281–293.
35. Sinden G. Characteristics of the UK wind resource: Long-term patterns and relationship to electricity demand. *Energy Policy* 2007; **35**(1):112–127.
36. FINO. Forshungsplattformen in nord- und ostsee nr. 1,2,3. `http://www.fino-offshore.de/en/` September 2013.
37. German Maritime and Hydrographic Agency. Fino1 database information. `http://www.bsh.de/en/Marine_data/Observations/MARNET_monitoring_network/FINO_1/index.jsp` September 2013.
38. ECMWF. European centre for medium-range weather forecasts. `http://www.ecmwf.int/` September 2013.
39. Moncrieff M, Miller M. The dynamics and simulation of tropical cumulonimbus and squall lines. *Quarterly Journal of the Royal Meteorological Society* 1976; **102**(432):373–394.
40. Charnock H. Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society* 1955; **81**(350):639–640.
41. Pinson P, Hagedorn R. Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteorological Applications* 2012; **19**(4):484–500.
42. Atger F. Estimation of the reliability of ensemble-based probabilistic forecasts. *Quarterly Journal of the Royal Meteorological Society* 2004; **130**(597):627–646.
43. Hamill TM. Reliability diagrams for multicategory probabilistic forecasts. *Weather and forecasting* 1997; **12**(4):736–741.
44. Hersbach H. Sea surface roughness and drag coefficient as functions of neutral wind speed. *Journal of Physical Oceanography* 2011; **41**(1).
45. Phillips OM. On the generation of waves by turbulent wind. *Journal of Fluid Mechanics* 1957; **2**(05):417–445.




*Wind Energ.* 0000; **00**:1–16 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
*Prepared using weauth.cls*