Probabilistic Forecasting of Wave Height
for Offshore Wind Turbine Maintenance

James W. Taylor

Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, UK

james.taylor@sbs.ox.ac.uk

Jooyoung Jeon[*]

School of Management, University of Bath, Bath, BA2 7AY, UK

Graduate School of Engineering Practice, Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul, 08826, South Korea

j.jeon@bath.ac.uk

*   Corresponding author

Abstract

Wind power continues to be the fastest growing source of renewable energy. This paper is concerned with the timing of offshore turbine maintenance for a turbine that is no longer functioning. Service vehicle access is limited by the weather, with wave height being the important factor in deciding whether access can be achieved safely. If the vehicle is mobilized, but the wave height then exceeds the safe limit, the journey is wasted. Conversely, if the vehicle is not mobilized, and the wave height then does not exceed the limit, the opportunity to repair the turbine has been wasted. Previous work has based the decision as to whether to mobilize a service vessel on point forecasts for wave height. In this paper, we incorporate probabilistic forecasting to enable rational decision making by the maintenance engineers, and to improve situational awareness regarding risk. We show that, in terms of minimizing expected cost, the decision as to whether to send the service vessel depends on the value of the probability of wave height falling below the safe limit. We produce forecasts of this probability using time series methods specifically designed for generating wave height density forecasts, including ARMA-GARCH models. We evaluate the methods in terms of statistical probability forecast accuracy, as well as monetary impact, and we examine the sensitivity of the results to different values of the costs.

1. Introduction

Wind power constitutes a significant part of the rapidly growing global renewable sector. In contrast to onshore wind farms, offshore locations provide stronger and steadier sources of energy, and are criticized less for blemishing the landscape. In Europe, several countries have ambitious plans for new offshore installations, most notably Denmark, Germany and the UK. In China, there are impressive targets for offshore development, with a plan to increase total installed offshore capacity from less than 1GW to 30GW by 2020. In the United States, although recent years have seen considerable growth in wind power generation, it has been entirely onshore. Staid and Guikema (2015) explain that, while many of these onshore locations have great wind resource, they tend to be far from the many densely populated urban areas on the coast. This implies a need for investment in new transmission lines, which presents an obstacle to the development of wind power, and motivates offshore wind power generation. The first commercial offshore farm in the United States began generating power near Rhode Island in December 2016, and many more installations are under construction, including a 1GW wind farm off Martha's Vineyard. In terms of construction, offshore locations are challenging and costly (Irawan et al., 2017b; Ursavas, 2017). This is also the case for offshore operations and maintenance, which has been described as a fast developing sector in its own right (GL Garrad Hassan, 2013). Although turbine design and manufacturing has improved, the harsh marine environment and increasing turbine size mean that reliability will continue to be a challenge for offshore wind farm operators (Caroll et al., 2016).

Irawan et al. (2017a) describe how offshore maintenance can be predetermined, condition-based, or corrective. Predetermined maintenance can include work performed at predetermined intervals. Condition-based maintenance is work carried out in response to the condition of equipment revealed by ongoing monitoring. Corrective maintenance is in response to an equipment failure that has already occurred. Our interest in this paper is in the timing of maintenance for an offshore turbine that is either unable to operate, or has been shut down due to some level of fault. Service vehicle access to the turbine is limited by the weather, with wave height being the important factor in deciding whether access can be achieved safely. Dinwoodie et al. (2013) present wave height limits for various forms of vehicle, including helicopters and various sea vessels. They explain that the wave height limits apply for the duration that the service vehicle is at sea. We refer to this period as the *mobilization window*. In this paper, we follow Catterson et al. (2016) by focusing on the simplest form of crew transport vessel, which is used to transfer crew and tools for common maintenance work. They consider the limit of 1.5m, and compare wave height point forecasts from different time series methods in terms of their ability to predict whether this limit will be exceeded during the mobilization window. The decision as to whether to send the vessel is dictated by these predictions. An appealing feature of the work of Catterson et al. (2016) is that they evaluate the resulting decisions in terms of monetary cost. A wasted trip by the service vehicle will have an associated cost, and an opportunity cost will be incurred if an opportunity to send the vessel and repair the turbine is missed.

In this paper, we extend the work of Catterson et al. (2016) to incorporate probabilistic forecasting, and to investigate whether a probabilistic approach to decision making should be preferred to the deterministic approach that they employ. Conveying forecast uncertainty through probability estimates is important to improve situational awareness, as well as to enable rational decision making. This is discussed by LeClerc and Joslyn (2015), Winkler (2015) and Ursavas (2017) in the context of decisions based on weather forecasts. We first show that, in terms of minimizing expected cost, the decision as to whether or not to send the service vessel depends on the probability of wave height being below 1.5m for the duration of the mobilization window.We then produce forecasts of the probability using time series methods for generating forecasts of the probability density function, which are termed *density forecasts*. We evaluate wave height density and probability forecast accuracy using statistical measures, as well as the monetary cost resulting from the decision based on the probability forecast. We also use monetary cost to compare decision making based on probabilistic and point forecasts.

The time series methods that we use to produce density forecasts include kernel density estimation (KDE), time-varying parameter (TVP) regression models, autoregressive moving average generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) models, and density forecast combining. Catterson et al. (2016) did not consider these methods in their analysis, as their focus was point forecasting. Although other time series methods have been proposed for wave height forecasting, such as artificial neural networks (see Reikard et al., 2015), we chose methods that are suited to density forecasting. A discrete choice model could perhaps be used to model directly the probability of wave height being below 1.5m. Taylor (2017) uses a model of this type to predict wind power exceedance probabilities. However, the use of such models for our application is not straightforward, with a separate model needed for each lead time within the mobilisation window of interest. Furthermore, the use of a discrete choice model is often motivated by the lack of an obvious distributional assumption, as in Taylor's (2017) study of wind power, but we show that this is not a concern for our wave height data, if standard transformations are applied. Several of the methods that we consider involve the modelling of wave height in terms of wind speed. For these methods, wind speed forecasts are essentially generated autoregressively within the model. This has practical advantages in terms of convenience and cost, as the forecasts can be generated subject simply to the condition that historical wind speed observations are available. However, wind speed forecasts could be generated from other approaches, such as a numerical weather prediction system.

In the next section, we describe how the decision as to whether or not to mobilize the service vessel should depend on the probability that the wave height will be below 1.5m for the duration of the mobilization window. We then describe our dataset. The section that follows presents methods for wave height density forecasting. The next section describes an empirical study in which we compare forecasts in terms of statistical measures, as well as monetary cost. The final section provides a summary and concluding comments.

2. The Need for Probability Forecasts

Catterson et al. (2016) use wave height point forecasts from a variety of time series methods to predict whether or not a limit of 1.5m will be exceeded during the mobilization window. If, and only if, the point forecasts for all periods in the window are below the limit of 1.5m, mobilizing the service vessel is considered to be the optimal decision. Catterson et al. (2016) explain that, in terms of monetary outcome, it is the forecasts that lead to decisions with poor outcomes that have negative consequences, while forecasts that lead to decisions with satisfactory outcomes carry no penalty. If the vessel is mobilized, and the wave height then exceeds the limit during the mobilization window, the trip will have been wasted, with an associated cost C_{trip}. If the vessel is not mobilized, and the wave height turns out not to exceed the limit during the mobilization window, an opportunity to send the vessel and repair the turbine would have been missed, and this can be viewed as carrying an opportunity cost C_{opp} equal to the revenue that has been lost due to power not being generated. Catterson et al. (2016) use these costs to evaluate the decisions resulting from the point forecasts from different methods. By contrast, we base our decision making on these costs and probability forecasts.

Let p be the probability that a wave height of 1.5m is not exceeded during the mobilization window. The expected cost of opting to mobilize the vessel is:

$$E_{Mobilize} = p \times 0 + (1-p) \times C_{trip}$$

The expected cost of the alternative of opting not to mobilize the vessel is:

$$E_{NoMobilize} = p \times C_{opp} + (1-p) \times 0$$

Using the criterion of minimizing expected cost, it is optimal to mobilize the vessel when $E_{Mobilize} < E_{NoMobilize}$, which is the case when probability p is greater than a critical value $p_{critical}$ given by:

$$p_{critical} = C_{trip} / (C_{opp} + C_{trip}) \tag{1}$$

This provides a threshold for the probability that a wave height of 1.5m is not exceeded during the mobilization window. Expression (1) shows that, if the cost of the trip C_{trip} is low relative to the opportunity cost C_{opp}, it will be optimal to mobilize even for quite low values of p, but that, if the cost of the trip is relatively high, p would have to be high for mobilizing to be optimal. Our proposal is to base the decision, as to whether to mobilize, on whether or not the forecast for p is more than $p_{critical}$.

The cost of the trip is calculated as (Catterson et al., 2016):

$$C_{trip} = MobilizationWindowHours \times (FuelPricePerHour + VesselHirePricePerHour) \tag{2}$$

Note that wind farm operators do not own vessels, and so have to hire them instead (Catterson et al., 2016). The opportunity cost is the revenue that would have been generated from selling electricity if the turbine had been repaired. Catterson et al. (2016) calculate this for the duration of the mobilization window, because the decision can be made afresh at the end of this window. The opportunity cost is, therefore, calculated as:

$$C_{opp} = MobilizationWindowHours \times Capacity \times CapacityFactor \times EnergyPrice \tag{3}$$

In their empirical work, Catterson et al. (2016) assume *MobilizationWindowHours* = 3 hrs, *FuelPricePerHour* = £50, *Capacity* = 5MW, *CapacityFactor* = 0.5 and *EnergyPrice* = £120. Their

value for *VesselHirePricePerHour* seems to be based on the vessel hire charge per day given by Dinwoodie et al. (2013) as £1750. Using these values in expressions (2) and (3) delivers C_{trip} = £368.75 and C_{opp} = £900. Substituting these costs into expression (1), gives $p_{critical}$ = 29.1%. We use this value in our empirical analysis of Section 5, but, as the various parameter values may well be different in practice, we also perform a sensitivity analysis by considering other values for $p_{critical}$.

3. Wave and Wind Data

In our empirical work, we used wave height, wave period and wind speed data recorded at the FINO3 research platform, which is situated in the North Sea, approximately 80km to the west of Sylt, Germany. FINO3 is one of three research platforms, introduced by the German Government to provide facilities where research and development projects can be performed for offshore wind power generation. Meteorological and oceanographic data are recorded for research purposes. Offshore wind farms located in the North Sea, such as Bard 1 and Borkum West, face harsh weather conditions, which present a challenge for vessel scheduling for installation and maintenance (Ursavas, 2017).

Our wave height and wave period observations were based on minute-by-minute observations for significant wave height and mean wave period, respectively. These variables are recorded for use in wave energy research. Significant wave height is the average height of the highest third of the waves, where height is measured in meters from the trough to the crest of the wave. Significant wave height is, therefore, a summary measure of the larger waves observed during a chosen time interval, rather than a record of the largest wave (Catterson et al., 2016). Mean wave period is recorded in seconds, and is the average time taken for successive wave crests to pass a given point.

Wind speed data is measured in m/s, and recorded at FINO3 every minute at various heights. We used measurements from the lowest available height, which was 30m. For the 4% of these measurements that were missing, we used measurements at the next lowest height available, with a form of adjustment that involves linear scaling to reduce the bias from the different dynamics of wind speed at different heights. The wind speed data was available to us at a 10-minute frequency, where each value was the average of the minute-by-minute measurements within each 10-minute interval.

For each of the three variables, we averaged observations within each hour to produce hourly data. In doing so, we are following Catterson et al. (2016), and are implicitly assuming that the decision as to whether or not to send the vessel is dictated by forecasts regarding the exceedance of hourly significant wave height above 1.5m. We acknowledge that this decision could instead be based on higher frequency data, or on a different measure of wave height instead of significant wave height.

To avoid long periods of missing observations, we chose to use the 18-month period from 1 March 2014 to 31 August 2015, inclusive. For this period, just 10 hourly observations were missing. For these missing values, we used linear interpolation to generate values. Of the total sample of 18 months, we used the first 12 months for model estimation, and the last six months for post-sample

evaluation. We produced wave height probability forecasts from each method using each period of the evaluation sample as forecast origin.

**Fig. 1.** Hourly time series of wave height, wind speed and wave period.
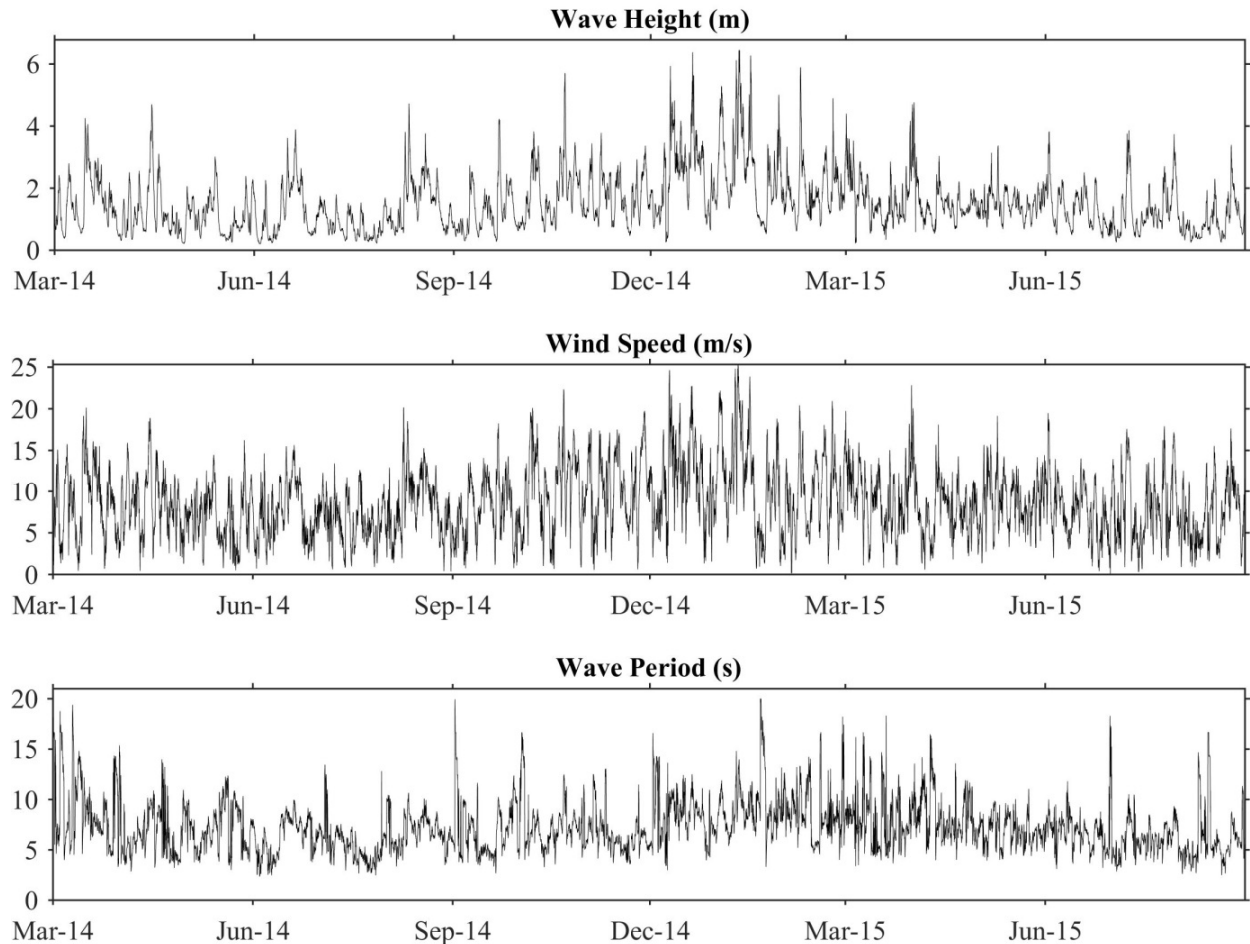
In Fig. 1, we present the hourly time series of wave height, wind speed and wave period. All three series show high variability, and are bounded below by zero. Positive skewness is apparent in the two wave series, and this is confirmed by the corresponding skewness values in the first column of values in Table 1, and the corresponding histograms in the first column of Fig. 2. The kurtosis values in the first column of values in Table 1 also confirm that the two wave series are not Gaussian. This table and figure show some level of positive skewness in wind speed. Table 1 and Fig. 2 also consider two transformations of the variables, and we discuss these in Section 4, in the context of ARMA-GARCH modelling. Despite the high volatility in the wave height time series of Fig. 1, the series is certainly not pure noise, which motivates some form of autoregressive modelling. The existence of autocorrelation is confirmed by Fig. 3(a), which shows significant autocorrelation in the wave height series for many lags. Another interesting feature of the time series plot of wave height is that there are clusters of periods of high volatility followed by clusters of relatively tranquil periods. This autoregression in the volatility motivates GARCH modelling.

Visual inspection of the wave height and wind speed time series plots of Fig. 1 suggests the two series are correlated, which is intuitive. This is confirmed by the scatter plot of Fig. 4(a), as well as Fig. 3(b), which supports the finding of Miller (1958) and Rieder (1997) that there can be significant correlation between wave height and even quite long lags of wind speed. This motivates the consideration of a bivariate model involving these two variables. Fig. 1 gives no clear sign of correlation between wave height and wave period, and this is confirmed by the scatter plot for these two variables in Fig. 4(b), which shows quite a complex relationship. It is, therefore, not surprising to see in Fig. 3(c) that the correlations between wave height and lags of wave period are quite low. Nevertheless, these correlations are statistically significant, and so there is some motivation for a bivariate model of wave height and wave period.

Table 1. Skewness and kurtosis for wave height, wind speed and wave period, and log and Box-Cox transformations of each variable. Statistics calculated for the 12-month estimation sample.

		Original	Log	Box-Cox
wave height	skewness	1.22	-0.30	-0.02 (λ= 0.19)
	kurtosis	4.88	2.58	2.49 (λ= 0.19)
wind speed	skewness	0.51	-1.10	-0.04 (λ= 0.57)
	kurtosis	3.02	5.16	2.68 (λ= 0.57)
wave period	skewness	1.14	0.06	0.00 (λ= -0.06)
	kurtosis	4.99	2.95	2.93 (λ= -0.06)

Fig. 2. Empirical distribution of wave height, wind speed and wave period, and their log and Box-Cox transformed distributions.

8
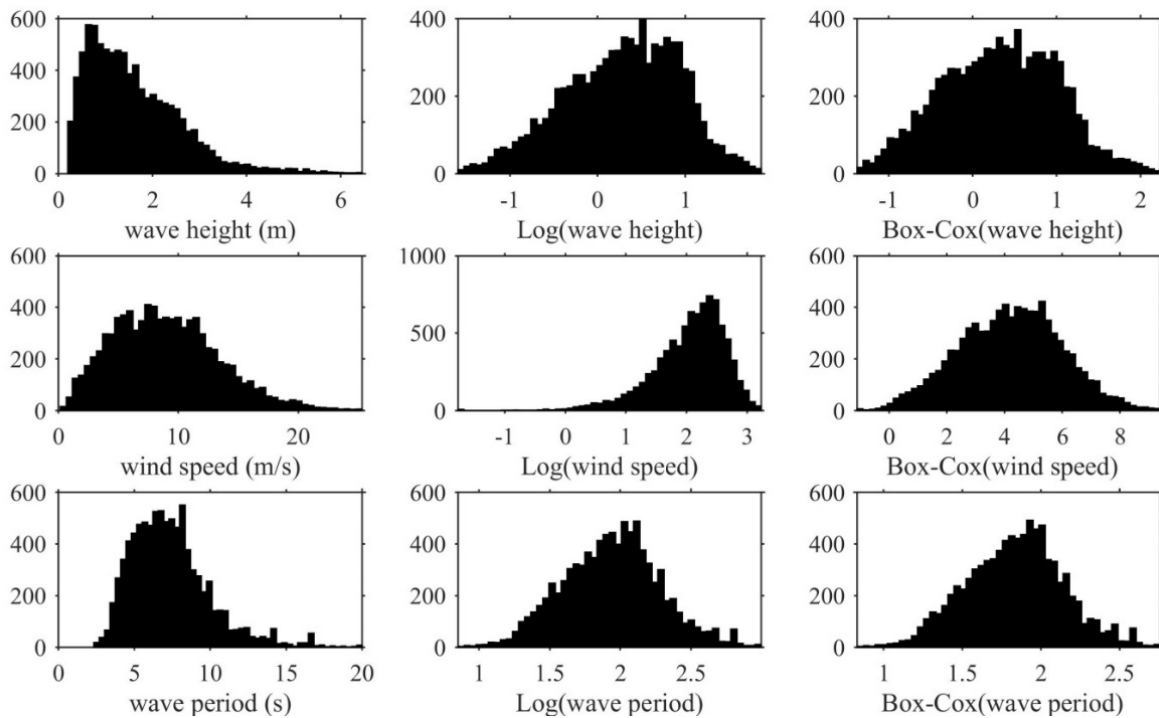
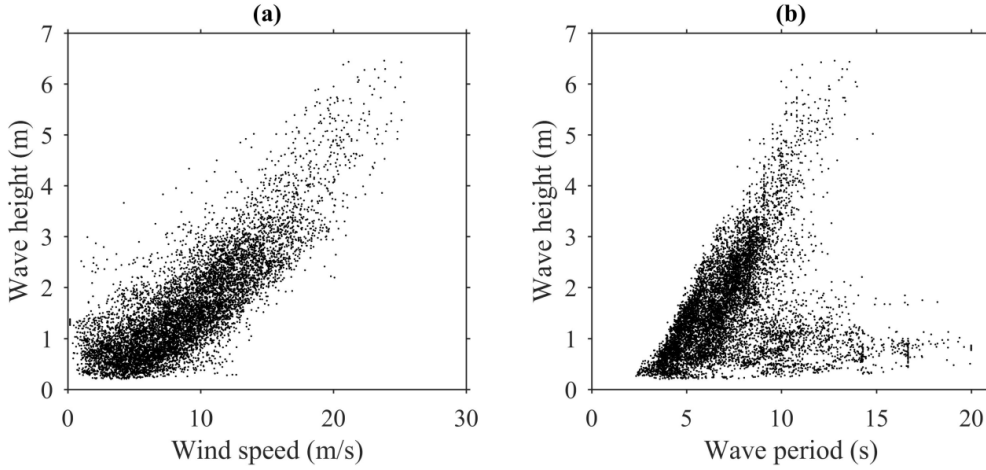**Fig. 3.** (a) Autocorrelations in wave height. (b) Correlations between wave height and lags of wind speed. (c) Correlations between wave height and lags of wave period. The upper bounds of 95% acceptance regions are indicated by the dotted lines.
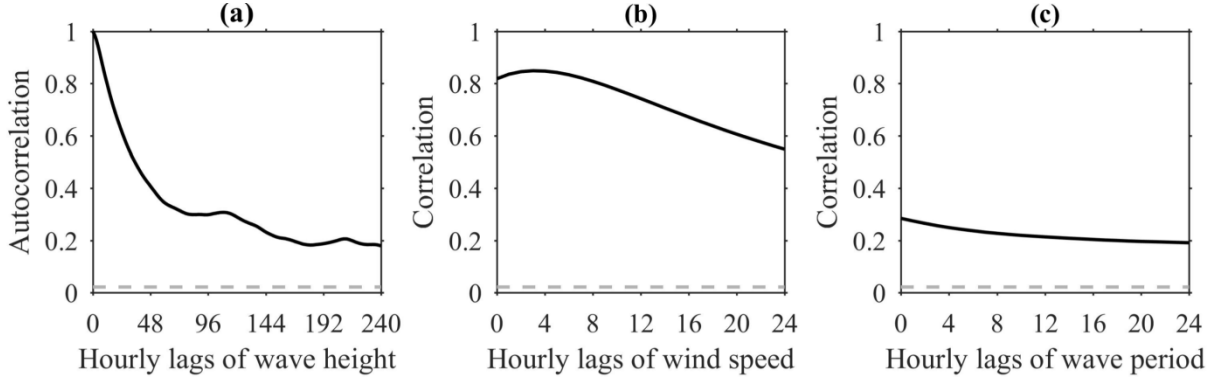
Fig. 4. (a) Scatter plot of wave height and wind speed. (b) Scatter plot of wave height and period.

4. Methods for the Probabilistic Forecasting of Wave Height

Probabilistic wave height forecasts can be produced using ensemble predictions produced by a physics-based model, such as a numerical weather prediction system (Roulston et al., 2005). However, generating ensemble predictions is computationally intensive, which limits their geographical coverage as well as the frequency of updating of the forecasts. Furthermore, for short lead times, statistical time series methods tend to be very competitive for meteorological variables, because recent observations can better describe the state of the atmosphere than a physical model (Yoder et al., 2014). For these reasons, in our forecasting of wave height, we implement only statistical time series methods. This is also the approach of Catterson et al. (2016). They consider a variety of models for point forecasting. In this paper, we consider a different set of models, as our focus is density forecasting. The first approach that we consider is KDE, which we view as a relatively simple nonparametric benchmark method. We then describe a regression method, and a variety of univariate and bivariate ARMA-GARCH models. We end this section by discussing the combination of density forecasts.

9

4.1. Kernel Density Estimation

KDE is a nonparametric approach, and so has the appeal of involving no distributional assumptions for the data. It applies kernel smoothing to the empirical distribution of historical wave height observations. The kernel density estimator of wave height is expressed as:

$$f(y) = \sum_{t=n-k+1}^{n} K_b(h_t - y)/k$$

where h_t is a wave height observation; y is the value of wave height for which a density is to be estimated; n is the forecast origin; k is the window length used for the estimation; and K_b is a Gaussian kernel function, with bandwidth b, which controls the degree of smoothing. We optimized the bandwidth by minimizing the continuous ranked probability score (CRPS) calculated for the 12-month estimation sample. This measure is a proper scoring rule for estimates of probability density functions, which means that the expected value of the CRPS is minimized when the actual density function is used as the density forecast (Gneiting et al., 2007). We considered the following three different window lengths: (a) 4 hours, (b) 24 hours, and (c) 12 months. Although the use of just 4 observations is unlikely to deliver good estimates of the spread of the density, it may be preferable to longer window lengths for the estimation of the mean of the density.

4.2. Time-Varying Parameter Regression

For the purposes of the short-term forecasting of wave energy flux, Reikard (2009) found success with a set of independently estimated regression models for wave height, wave period and wind speed, as in expressions (4)-(6). We implemented this approach in our study for the purpose of predicting wave height.

$$\ln s_t = \theta_{0t} + \theta_{1t} \ln s_{t-1} + \theta_{2t} \ln s_{t-2} + \theta_{3t} \ln s_{t-3} + \theta_{24t} \ln s_{t-24} + \epsilon_{1t} \tag{4}$$

$$\ln p_t = v_{0t} + v_{1t} \ln p_{t-1} + v_{2t} \ln p_{t-2} + v_{3t} \ln p_{t-3} + v_{4t} \ln p_{t-4} + \epsilon_{2t} \tag{5}$$

$$\ln h_t^2 = \delta_{0t} + \delta_{1t} \ln h_{t-1}^2 + \delta_{2t} \ln h_{t-2}^2 + \delta_{3t} \ln h_{t-3}^2 + \delta_{4t} \ln h_{t-4}^2 + \delta_{5t} \ln p_t + \delta_{6t} \ln s_t + \epsilon_{3t} \tag{6}$$

h_t, p_t and s_t are wave height, wave period and wind speed, respectively. The terms ϵ_{it} are assumed to be Gaussian white noise. θ_{it}, v_{it} and δ_{it} are parameters estimated separately for each expression and for each forecast origin using ordinary least squares. The lags in the three expressions are the same as those used by Reikard (2009), with lags up to 4 hours intended to capture short-run autocorrelation, while the lag 24 term aims to capture potential cyclicality -. As the parameters are updated for each forecast origin, we refer to this method as time-varying parameter (TVP) regression.

4.3. Univariate ARMA-GARCH Model of Wave Height

ARMA-GARCH models enable the modelling of autoregression in the mean and variance of a time series. When modelling wind speed, it is common to include terms to try to capture diurnal and annual cyclical patterns, and so it seems reasonable to do the same when modelling wave height.

However, we were not able to model an annual cycle because our time series is not sufficiently long. The wave height autocorrelation plot of Fig. 3(a) indicates that, if there is a diurnal cyclicality in the series, it is not particularly strong. Nevertheless, we opted to allow for a possible diurnal cycle by including trigonometric terms in the ARMA and GARCH specifications. Our model is written as:

$$\varphi(L)y_t = s(\boldsymbol{\mu}, t) + \psi(L)\varepsilon_t \tag{7}$$

$$\alpha(L)\sigma_t^2 = s(\boldsymbol{\omega}, t) + \beta(L)\varepsilon_t^2 \tag{8}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{9}$$

$$s(\boldsymbol{\mu}, t) = \mu_0 + \sum_{i=1}^{N_\mu} \left[\mu_{i,1} \sin\left(2i\pi \frac{h(t)}{24}\right) + \mu_{i,2} \cos\left(2i\pi \frac{h(t)}{24}\right) \right] \tag{10}$$

$$s(\boldsymbol{\omega}, t) = \omega_0 + \sum_{i=1}^{N_\omega} \left[\omega_{i,1} \sin\left(2i\pi \frac{h(t)}{24}\right) + \omega_{i,2} \cos\left(2i\pi \frac{h(t)}{24}\right) \right] \tag{11}$$

where $\varphi(L)$, $\psi(L)$, $\alpha(L)$ and $\beta(L)$ are polynomial functions of the lag operator L, corresponding to the AR, MA, GARCH and ARCH components, respectively; y_t is the observation at time t; ε_t is an error term; η_t is white noise; σ_t is the conditional standard deviation; $\boldsymbol{\mu}$ and $\boldsymbol{\omega}$ are vectors of parameters; and $s(\boldsymbol{\mu}, t)$ and $s(\boldsymbol{\omega}, t)$ capture diurnal cyclical patterns in the mean and variance, respectively, where $h(t)$ is the hour of the day. In order to satisfy stationarity and invertibility conditions, all the roots of the polynomials $\varphi(L)$, $\psi(L)$, $\alpha(L)$ and $\beta(L)$ must lie outside the unit circle. As with all the ARMA-GARCH models in this paper, we used maximum likelihood to estimate the model parameters, and we used the Schwarz Bayesian Information Criterion (SBC) to select N_μ, N_ω and the orders of the four lag polynomials. For more than half of the ARMA-GARCH models considered in this paper, this led to N_μ and N_ω both equalling zero, implying no diurnal cyclicality.

The use of maximum likelihood, and indeed the generation of density and probability forecasts, relies on a distributional assumption. In Section 3, we noted from Fig. 2 and Table 1 that the unconditional distribution of our wave height observations was not Gaussian. This was also the case with the residuals from various ARMA-GARCH models applied to this data. This motivates consideration of data transformations and alternative distributional assumptions. For each of our three variables, we applied a log transformation and a Box-Cox transformation (Box and Cox, 1964), which is written in the following two expressions:

$$\left(y^\lambda - 1\right)/\lambda \quad \text{if } \lambda \neq 0, y > 0$$

$$\log(y) \quad \text{if } \lambda = 0, y > 0$$

where y is the original variable and λ is a parameter, which we estimated by maximizing a Gaussian log-likelihood. Looking at Fig. 2 and Table 1, we see that, for wave height and wave period, both the log and Box-Cox transformations were effective in reducing the skewness and bringing the kurtosis closer to 3, which is the value for a Gaussian distribution. This was also the case for the Box-Cox transformation applied to the wind speed series, but not for the log transformation applied to this series. In our empirical work, we implemented the ARMA-GARCH model described in this subsection, as well as the bivariate models of the next subsection, with the data in its original form,

and also after the application of the transformations. For all models, we considered the following three distributional assumptions for η_t: Gaussian, Student t distribution and skewed t distribution.

In addition to fitting an ARMA-GARCH model to wave height, we implemented an ARMA model with constant variance, as we were interested to investigate whether there is benefit in modelling autoregression in the variance. The first two sections of Table 2 summarise the model specifications selected using the SBC for the ARMA and ARMA-GARCH models with Gaussian and Student t distributional assumptions, and log and Box-Cox transformations. The remainder of this table summarises selected specifications for two bivariate ARMA-GARCH models, which we discuss in the next section.

Table 2. Summary of the lags and diurnal terms in the ARMA, ARMA-GARCH and bivariate ARMA-GARCH models selected by the SBC using the in-sample data.

Lags	AR	MA	Diurnal in mean	ARCH	GARCH	Diurnal in volatility
ARMA models of **wave height**						
Gaussian log	[1,2,3,4,5,6]	[1,2,3]	[2,4]	no	no	no
Gaussian Box-Cox	[1,2,3,4]	[1,2,3]	[2]	no	no	no
Student t log	[1,2,3]	[1,2,3,24]	no	no	no	no
Student t Box-Cox	[1,2,3,24]	[1,2,3,24]	no	no	no	no
ARMA-GARCH models of **wave height**						
Gaussian log	[1,2,3,4,5,6]	[1,2]	[2,4]	[1,2]	[1,2,3]	[2,4]
Gaussian Box-Cox	[1,2,3,4]	[1,2,3]	[2]	[1,2]	[1,2,3,24]	no
Student t log	[1,2,3]	[1,2,3,24]	no	[1]	[1,2,24]	no
Student t Box-Cox	[1,2,3,24]	[1,2,3,24]	no	[1,2,24]	[1,2,24]	no
Bivariate ARMA-GARCH models of **wave height** and **wind speed**						
Gaussian log	[1,24]	[1,2,3,24]	no	no	no	no
Gaussian Box-Cox	[1,2,3]	[1,2,24]	[2]	[1]	[1]	no
Student t log	[1,2,3,4,5,6]	[1,2]	[2,4]	[1,2,3,4]	[1,2,3]	[2]
Student t Box-Cox	[1,2,3,4,5,6]	[1,2,3,4]	no	[1,2,3,4]	no	No
Bivariate ARMA-GARCH models of **wave height** and **wave period**						
Gaussian log	[1,2,3,4,5,6]	no	[2,4]	no	no	no
Gaussian Box-Cox	[1]	[1,2,3,4]	no	no	[1,2,3,4]	no
Student t log	[1,2,3,24]	[1,2,3,24]	[2,4,6]	[1,2]	[1,2,3,24]	[2]
Student t Box-Cox	[1,2,3,4,5,6]	[1,2,3,4,5,6]	[2,4,6]	[1]	[1]	[2]

4.4. Bivariate ARMA-GARCH Models Involving Wave Height

As wave height is influenced by the strength of the wind, it seems natural to build a model that tries to capture this dependency, along with the autoregression in the wave height series. We, therefore, fitted a bivariate ARMA-GARCH model to the wave height and wind speed series. We are not aware of previous studies that have applied this model to this pair of variables. Jeon and Taylor

(2016) consider a bivariate ARMA-GARCH model for wave energy flux and wind speed, but this provides little insight for our study, as wave energy flux and wave height are notably different variables, with energy flux being a nonlinear function of wave height and wave period. With the energy flux function as the focus, Jeon and Taylor (2016) also consider a bivariate ARMA-GARCH model for wave height and wave period. This model has relevance for our work, as the joint modelling with wave period may improve the explanatory and predictive power of the modelling of wave height. In summary, we implemented a bivariate ARMA-GARCH model for two pairs of variables: (a) wave height and wind speed, and (b) wave height and wave period.

We used the VEC bivariate ARMA-GARCH model of Bollerslev et al. (1988) with cyclical terms to capture the diurnal cyclicality. This model is written as:

$$\boldsymbol{\varphi}(L)\boldsymbol{y}_t = \boldsymbol{S}(\boldsymbol{\mu}, t) + \boldsymbol{\psi}(L)\boldsymbol{\varepsilon}_t$$

$$vech(\boldsymbol{H}_t) = \boldsymbol{S}(\boldsymbol{\omega}, t) + \sum_{i=1}^{p} \boldsymbol{P}_i vech(\boldsymbol{H}_{t-i}) + \sum_{j=1}^{q} \boldsymbol{Q}_j vech(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}'_{t-j})$$

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t$$

where $\boldsymbol{y}_t$ is a (2×1) vector (for example, consisting of wave height and wind speed); $\boldsymbol{\varphi}(L)$ and $\boldsymbol{\psi}(L)$ are (2×2) matrices with lag polynomials; $\boldsymbol{\varepsilon}_t$ is a (2×1) vector of error terms; $\boldsymbol{H}_t$ is the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$; $vech(\cdot)$ denotes the column stacking operator of the lower triangular part of its argument symmetric matrix; $\boldsymbol{P}_i$ and $\boldsymbol{Q}_j$ are (3×3) matrices of parameters; p and q are the orders of $\boldsymbol{P}_i$ and $\boldsymbol{Q}_j$ respectively, as selected by the BIC; $\boldsymbol{S}(\boldsymbol{\mu}, t)$ and $\boldsymbol{S}(\boldsymbol{\omega}, t)$ are (2×1) and (3×1) vectors with elements that each differ and are of the form of $s(\boldsymbol{\mu}, t)$ and $s(\boldsymbol{\omega}, t)$, as defined in expressions (10) and (11); and $\boldsymbol{\eta}_t$ is a (2×1) vector of white noise. $\boldsymbol{P}_i$ and $\boldsymbol{Q}_j$ are restricted by the sufficient condition for the positivity of $\boldsymbol{H}_t$ proposed by Gourieroux (1997). For $\boldsymbol{\eta}_t$, we considered bivariate Gaussian, Student t and skewed t distributions. For the bivariate skewed t distribution, we used the definition of Azzalini and Genton (2008).

4.5. Forecast Combining

Rather than select from among a set of competing forecasting methods, empirical results across a vast array of applications have shown that accuracy can be improved by combining the forecasts from the individual methods (see, for example, De Menezes et al., 2000). This motivates the use of combining in our work. However, the focus of the combining literature has largely been on point forecasting, with the simple average frequently proving to be very competitive, and often the most accurate combining method (see, for example, Genre et al., 2013). Catterson et al. (2016) combine point forecasts for wave height using an artificial neural network. Despite this being an unusual combining approach, they report that it delivered greater accuracy than the individual methods. Although combining methods have been proposed for density forecasting (see, for example, Gneiting and Raftery, 2013; Kapetanios et al., 2015), the literature is still in its infancy, and there is little empirical evidence to support any one particular approach. In view of this, we opted to consider

simple averages in our analysis. Lichtendahl et al. (2013) point out that distributional forecasts can be averaged by either averaging quantiles or by averaging probabilities, with the former approach being preferable theoretically. We found the results to be very similar for the two approaches, and so, to save space, in the remainder of the paper, we describe the results for only the quantile averaging approach. We considered the following two combinations: (i) TVP regression and the four different types of ARMA-GARCH models that we implemented, and (ii) TVP regression and the bivariate ARMA-GARCH model of wave height and wind speed, which was, as we describe in the next section, the most accurate of the four types of ARMA-GARCH models.

5. Empirical Comparison of Methods

In Section 5.1, we provide an initial comparison of the various methods of Section 4 in terms of their density forecast accuracy. Section 5.2 focuses on the specific problem of interest in this paper, by evaluating the methods in terms of their accuracy in forecasting the probability that is needed as the basis of the decision as to whether or not to mobilize the maintenance vessel. In Sections 5.3 and 5.4, we calculate the costs that would result from this decision, and use this as a basis for comparing the forecasting methods.

5.1. Statistical Evaluation of Density Forecasts

In this section, we evaluate the methods of Section 4 in terms of their ability to produce wave height density forecasts for lead times from 1 to 12 hours ahead. For each method, we used the first 12 months of the data to estimate parameters, and the final 6 months for post-sample evaluation. We rolled the forecast origin forward through each hour of the post-sample evaluation period to produce a density forecast from each method for each lead time. For KDE, the density forecast is identical for each lead time. For the KDE method, and for the ARMA and ARMA-GARCH models, we opted not to re-estimate method parameters as we rolled the forecast origin forward, because we felt it would be unlikely to be done in practice, as it would be computationally demanding. By contrast, we did re-estimate the TVP regression models using a growing window for each forecast origin, as this is computationally simple, and because it seems to us fundamental to Reikard's (2009) intended implementation of the method.

For the ARMA-GARCH models, it is not possible to generate density forecasts using an analytical approach beyond 1 step-ahead. This is because the inclusion of a GARCH component leads to the distribution of standardised forecast errors differing for successive lead times (see, for example, Baillie and Bollerslev, 1992, Section 6). An analytical approach is also not possible beyond 1 step-ahead for an ARMA model with Student t or skewed t distributions. This is because analytical expressions do not exist for the sum of random variables that have these distributions. Therefore, we used simulation to generate density forecasts from the ARMA and ARMA-GARCH models. From each forecast origin, we projected 10^3 simulated sample paths across the 12 lead times, and from this

we constructed a density forecast for each lead time. For each simulated observation, the value of the noise term (η_t or $\boldsymbol{\eta}_t$) is sampled from the chosen distribution (i.e. univariate or bivariate Gaussian, Student t or skewed t). A similar simulation approach is described by Hyndman et al. (2008, Section 6.1) for generating prediction intervals from exponential smoothing models.

The wave height density forecasts are evaluated using the CRPS averaged over the post-sample period. Table 3 reports the mean CRPS values. The unit of the CRPS values is millimetres (mm), and lower values of this measure are better. Each of the first four columns of values group three lead times together, and the final column averages the CRPS across all 12 lead times.

Of the three KDE implementations, Table 3 shows that using a window length of just 4 hours was preferable. As we noted in Section 4.1, this window length will not enable reasonable estimation of the spread of the density, and so its relative success must be due to its ability to swiftly adapt to changes in the mean of the series. Comparing the results for KDE with the other values in Table 3, we see that KDE is uncompetitive with the better of the other models. For example, the KDE methods are comfortably outperformed by TVP regression at all lead times.

For the ARMA and ARMA-GARCH models, for conciseness, we report results for only the models with Gaussian and Student t distributional assumptions, and log and Box-Cox transformations. Application of the models to the original untransformed data led to uncompetitive results, and the skewed t distribution led to poorer results than the Student t distribution. Comparing the results for the univariate ARMA models with and without the GARCH component, we see that there was some benefit in including the autoregressive model for the variance. If we then compare these results with those of the bivariate models, the overall picture is that the bivariate models are more accurate. The three best performing methods are all bivariate ARMA-GARCH models for wave height and wind speed. Our preliminary analysis of the data in Section 3 revealed that there was strong potential for a bivariate modelling of these two variables. In Section 3, we also noted that Fig. 2 and Table 1 showed that the skewness in wind speed was reduced by applying the Box-Cox transformation, but not by the use of the log transformation. It is, therefore, not surprising that, for the models involving wind speed, Table 3 shows that the Box-Cox transformation was preferable. With regard to the distributional assumption, Student t led to the greatest accuracy.

Table 3 shows that the bivariate ARMA-GARCH models for wave height and wind speed outperforms the TVP regression method at all lead times. Turning to the combining methods, we see that combining these two methods led to results that are reasonable, but no better than the bivariate ARMA-GARCH model.

Although our focus in this paper is probabilistic forecasting, it is interesting also to consider point forecasting, as this conveys the accuracy of the estimation of the mean of the density forecasts. Table 4 summarizes point forecast accuracy using the mean absolute error (MAE), which has mm as its units. The table shows a similar ranking of methods to the CRPS results of Table 3, with the best of the individual methods being the bivariate ARMA-GARCH model for wave height and wind speed

with Box-Cox transformation and Student *t* distribution. The results for this model were matched by combining it with TVP regression.

Table 3. Evaluation of post-sample wave height density forecast accuracy using mean CRPS (mm).

Lead time (hours):	1-3	4-6	7-9	10-12	1-12
KDE of wave height					
4 hours	256	283	306	325	293
24 hours	306	324	340	354	331
12 months	452	452	451	450	451
Regression model of wave height in terms of wind speed and wave period					
TVP regression	88	154	212	255	177
ARMA models of wave height					
Gaussian log	90	188	277	355	228
Gaussian Box-Cox	95	216	337	451	275
Student *t* log	85	171	246	309	203
Student *t* Box-Cox	93	210	327	435	266
ARMA-GARCH models of wave height					
Gaussian log	85	164	224	268	185
Gaussian Box-Cox	95	212	328	435	267
Student *t* log	85	171	245	309	202
Student *t* Box-Cox	93	208	323	432	264
Bivariate ARMA-GARCH models of wave height and wind speed					
Gaussian log	121	209	261	295	222
Gaussian Box-Cox	75	141	193	231	160
Student *t* log	76	142	193	229	160
Student *t* Box-Cox	**73**	**137**	**187**	**223**	**155**
Bivariate ARMA-GARCH models of wave height and wave period					
Gaussian log	86	165	217	248	179
Gaussian Box-Cox	83	155	206	241	171
Student *t* log	84	158	212	248	176
Student *t* Box-Cox	92	170	231	275	192
Combination of TVP regression and all four ARMA-GARCH models					
Gaussian log	87	157	209	246	175
Gaussian Box-Cox	81	154	212	256	176
Student *t* log	79	148	201	239	167
Student *t* Box-Cox	81	156	218	266	180
Combination of TVP regression and bivariate ARMA-GARCH model of wave height and wind speed					
Gaussian log	101	165	209	239	179
Gaussian Box-Cox	79	141	189	224	158
Student *t* log	79	142	191	225	159
Student *t* Box-Cox	78	139	188	**223**	157

Note: Lower values are better. The best value in each column is in bold.

Table 4. Evaluation of post-sample wave height point forecast accuracy using MAE (mm).

Lead time (hours):	1-3	4-6	7-9	10-12	1-12
KDE of **wave height**					
4 hours	159	250	315	362	272
24 hours	316	360	397	428	375
12 months	604	603	602	601	602
Regression model of **wave height** in terms of **wind speed** and **wave period**					
TVP regression	109	205	275	325	228
ARMA models of **wave height**					
Gaussian log	117	238	342	430	282
Gaussian Box-Cox	128	287	448	598	365
Student *t* log	112	221	309	379	255
Student *t* Box-Cox	124	275	420	546	342
ARMA-GARCH models of **wave height**					
Gaussian log	111	209	279	327	232
Gaussian Box-Cox	124	270	407	530	333
Student *t* log	112	221	309	380	255
Student *t* Box-Cox	123	268	404	527	330
Bivariate ARMA-GARCH models of **wave height** and **wind speed**					
Gaussian log	127	261	345	396	282
Gaussian Box-Cox	101	191	265	321	219
Student *t* log	103	196	270	324	223
Student *t* Box-Cox	**99**	**186**	256	309	213
Bivariate ARMA-GARCH models of **wave height** and **wave period**					
Gaussian log	115	223	292	327	239
Gaussian Box-Cox	110	211	288	340	237
Student *t* log	113	212	284	334	236
Student *t* Box-Cox	126	230	313	369	259
Combination of TVP regression and all four ARMA-GARCH models					
Gaussian log	106	199	265	309	220
Gaussian Box-Cox	105	200	272	323	225
Student *t* log	105	197	265	311	219
Student *t* Box-Cox	107	204	281	336	232
Combination of TVP regression and bivariate ARMA-GARCH model of wave height and wind speed					
Gaussian log	112	215	285	331	236
Gaussian Box-Cox	102	191	260	309	215
Student *t* log	104	194	263	311	218
Student *t* Box-Cox	101	188	**255**	**304**	**212**

Note: Lower values are better. The best value in each column is in bold.

5.2. Statistical Evaluation of Probability Forecasts

In Section 2, we described how the decision as to whether or not to send the service vessel depends on the probability that the wave height does not exceed 1.5m for any period in the mobilization window. In this subsection, we evaluate the accuracy of the various methods for forecasting this probability. As in the study of Catterson et al. (2016), we consider mobilization windows of 3 hours, and we forecast the probability for 10 lead times, with the longest of these lead times corresponding to the 3-hour mobilization window that opens 10 hours in the future.

Forecasts are required for the probability p of the joint event that the wave height does not exceed 1.5m in the 3 hours of the mobilization window. For the ARMA-GARCH models, density forecasts are generated by simulating 10^3 future sample paths from the model. We estimated the required probability p as the proportion of the 10^3 simulated sample paths for which the wave height did not exceed 1.5m in any of the 3 hours of the mobilization window. In simulating sample paths, we are able to capture the potential dependence between the likelihood of exceedance in each of the 3 successive hours. For the KDE approach, we have to treat them as 3 independent events. As KDE does not model autocorrelation, and the KDE density forecast is the same for each period in the forecast horizon, we estimated p using the probability forecast for the wave height not exceeding 1.5m in an individual hour, raised to the power of 3. It is an unappealing limitation of our use of KDE that we must assume independence for the probabilities of exceedance in the 3 successive periods. This aspect deterred us from directly modelling the exceedance probability using the autoregressive discrete choice models of Taylor (2017), although they could perhaps be adapted for modelling the probability of the joint event.

Our use of the CRPS in the previous subsection gave an indication of the overall quality of a method's density forecasts. In this subsection, rather than looking at the whole density forecast, we are interested in evaluating just the forecasts of the probability p. To evaluate the forecasts of p, we used the Brier score, which is a widely used proper scoring rule for probability forecasts of a variable with a binary outcome. Table 5 summarizes the results, showing the Brier scores for the 3-hour mobilization windows that open 1, 4, 7 and 10 hours in the future. Lower values of the score are better. The final column averages the score across all the mobilization windows. Although the rankings of the methods are broadly similar to the rankings for the CRPS and MAE results, it is interesting to see some differences. For example, KDE based on a window length of 4 hours outperformed some of the ARMA-GARCH models in terms of CRPS and MAE, but not in terms of the Brier score. This demonstrates how the CRPS can be quite heavily influenced by the mean of the density forecast. As with the CRPS and MAE, the best Brier score results were produced by the bivariate ARMA-GARCH models for wave height and wind speed with Box-Cox transformation and Student t distribution.

Table 5. Evaluation of post-sample wave height probability forecast accuracy using Brier score ($\times 10^3$).

Mobilization window (hours):	1-3	4-6	7-9	10-12	Mean of all windows
KDE of wave height					
4 hours	215	224	232	242	228
24 hours	247	255	263	270	259
12 months	320	320	319	319	319
Regression model of wave height in terms of wind speed and wave period					
TVP regression	125	166	199	228	179
ARMA models of wave height					
Gaussian log	46	112	165	198	132
Gaussian Box-Cox	95	176	240	281	201
Student *t* log	45	102	148	178	120
Student *t* Box-Cox	91	171	236	279	197
ARMA-GARCH models of wave height					
Gaussian log	70	118	157	183	134
Gaussian Box-Cox	94	173	235	276	197
Student *t* log	75	134	183	219	154
Student *t* Box-Cox	91	171	236	279	197
Bivariate ARMA-GARCH models of wave height and wind speed					
Gaussian log	94	189	234	255	197
Gaussian Box-Cox	60	91	118	138	102
Student *t* log	58	91	117	137	101
Student *t* Box-Cox	**39**	**78**	**107**	**130**	**90**
Bivariate ARMA-GARCH models of wave height and wave period					
Gaussian log	74	120	157	181	135
Gaussian Box-Cox	62	101	132	153	113
Student *t* log	63	106	138	162	119
Student *t* Box-Cox	70	120	168	207	142
Combination of TVP regression and all four ARMA-GARCH models					
Gaussian log	78	120	151	174	131
Gaussian Box-Cox	79	125	161	188	138
Student *t* log	72	114	146	171	126
Student *t* Box-Cox	80	128	166	195	142
Combination of TVP regression and bivariate ARMA-GARCH model of wave height and wind speed					
Gaussian log	79	113	137	156	121
Gaussian Box-Cox	66	100	126	148	110
Student *t* log	65	100	127	148	110
Student *t* Box-Cox	65	99	125	148	109

Note: Lower values are better. The best value in each column is in bold.

Table 6. Evaluation of post-sample wave height probability forecast accuracy using monetary cost resulting from the decision based on the probability forecast. Values are total cost (£000s) calculated using the following values introduced in Section 2: C_{trip} = £368.75, C_{opp} = £900 and $p_{critical}$ = 29.1%.

Mobilization window (hours):	1-3	4-6	7-9	10-12	Mean of all windows
KDE of wave height					
4 hours	418	454	491	522	472
24 hours	528	555	584	604	569
12 months	1060	1061	1062	1064	1062
Regression model of wave height in terms of wind speed and wave period					
TVP regression	109	164	218	270	190
ARMA models of wave height					
Gaussian log	59	163	222	244	177
Gaussian Box-Cox	62	177	254	283	200
Student *t* log	56	142	204	227	161
Student *t* Box-Cox	60	168	244	275	192
ARMA-GARCH models of wave height					
Gaussian log	57	139	209	246	166
Gaussian Box-Cox	62	180	256	284	202
Student *t* log	56	144	218	253	171
Student *t* Box-Cox	60	168	242	274	192
Bivariate ARMA-GARCH models of wave height and wind speed					
Gaussian log	57	163	319	412	239
Gaussian Box-Cox	52	115	159	192	132
Student *t* log	52	116	**155**	**189**	131
Student *t* Box-Cox	**51**	**110**	156	**189**	**129**
Bivariate ARMA-GARCH models of wave height and wave period					
Gaussian log	60	140	207	231	162
Gaussian Box-Cox	57	127	178	203	144
Student *t* log	62	134	181	210	150
Student *t* Box-Cox	64	131	204	249	164
Combination of TVP regression and all four ARMA-GARCH models					
Gaussian log	100	174	227	255	189
Gaussian Box-Cox	91	165	219	247	181
Student *t* log	85	148	195	225	163
Student *t* Box-Cox	89	161	213	244	177
Combination of TVP regression and bivariate ARMA-GARCH model of wave height and wind speed					
Gaussian log	113	168	207	235	181
Gaussian Box-Cox	83	138	176	201	149
Student *t* log	83	136	174	199	148
Student *t* Box-Cox	82	136	174	197	147

Note: Lower values are better. The best value in each column is in bold.

5.3. Evaluation of Forecasts in Terms of Monetary Cost Resulting from Proposed Decisions

Catterson et al. (2016) compare forecasting methods using the costs incurred as a result of decisions made on the basis of point forecasts. In this subsection, we do the same, except that we evaluate decisions made on the basis of density forecasts. In Section 2, we described how there are two important costs to consider. C_{trip} is the cost of a wasted trip by the service vessel, caused by the vessel being sent and then wave height being too high for the vessel to dock safely. C_{opp} is the opportunity cost, calculated as lost income due to the service vessel not being sent when it should have been. The final cost outcome C_{final} can, therefore, be summarized as:

$$C_{final} = \begin{cases} C_{trip} & \text{if vessel has a wasted trip} \\ C_{opp} & \text{if vessel should have been sent, but was not} \\ 0 & \text{otherwise} \end{cases}$$

In Section 2, we used the values of Catterson et al. (2016) to calculate C_{trip} = £368.75 and C_{opp} = £900. From these values, we calculated $p_{critical}$ = 29.1%, implying that the service vessel is sent if and only if the forecast of the probability p is above 29.1%. Using this decision making rule, along with C_{trip} and C_{opp}, we calculated the final cost C_{final} resulting from decisions made at each forecast origin for each lead time. We then summed this cost over the post-sample period to give a total cost. The total cost values resulting from the forecasts of the various methods are presented in Table 6. Reassuringly, the lowest values in Table 6 correspond to the same methods that delivered the lowest CRPS, MAE and Brier scores. The best performing method was again the bivariate ARMA-GARCH model based on wave height and wind speed with Box-Cox transformation and Student t distribution.

5.4. Sensitivity Analysis and Comparison of Probabilistic and Deterministic Forecasts

To produce the results in Table 6, we used values for C_{trip}, C_{opp} and $p_{critical}$ that were calculated in Section 2 using assumptions for fuel price, the cost of hiring the vessel, the capacity and capacity factor of the wind turbine, and the energy price. As these parameters will vary over time, and from one wind farm to another, it is important to investigate the robustness of our results to variations in the assumed values. We do this by performing sensitivity analysis on the critical probability $p_{critical}$. This involves varying $p_{critical}$, and noting the resulting impact on C_{final}. However, expression (1) shows that $p_{critical}$ is a function of C_{trip} and C_{opp}, and this implies that the only way to vary $p_{critical}$ is to vary C_{trip} and/or C_{opp}, which by definition forces a change in C_{final}. In view of this, in our sensitivity analysis, we analyse the impact of varying $p_{critical}$ on the ratio of C_{final} to the sum of C_{trip} and C_{opp}. Results of our sensitivity analysis are presented in Figs. 5-7.

We first describe Fig. 5, which concerns KDE based on a window length of 4 hours. In Fig. 5, we plot the results for four different lead times for the following four decision making approaches:
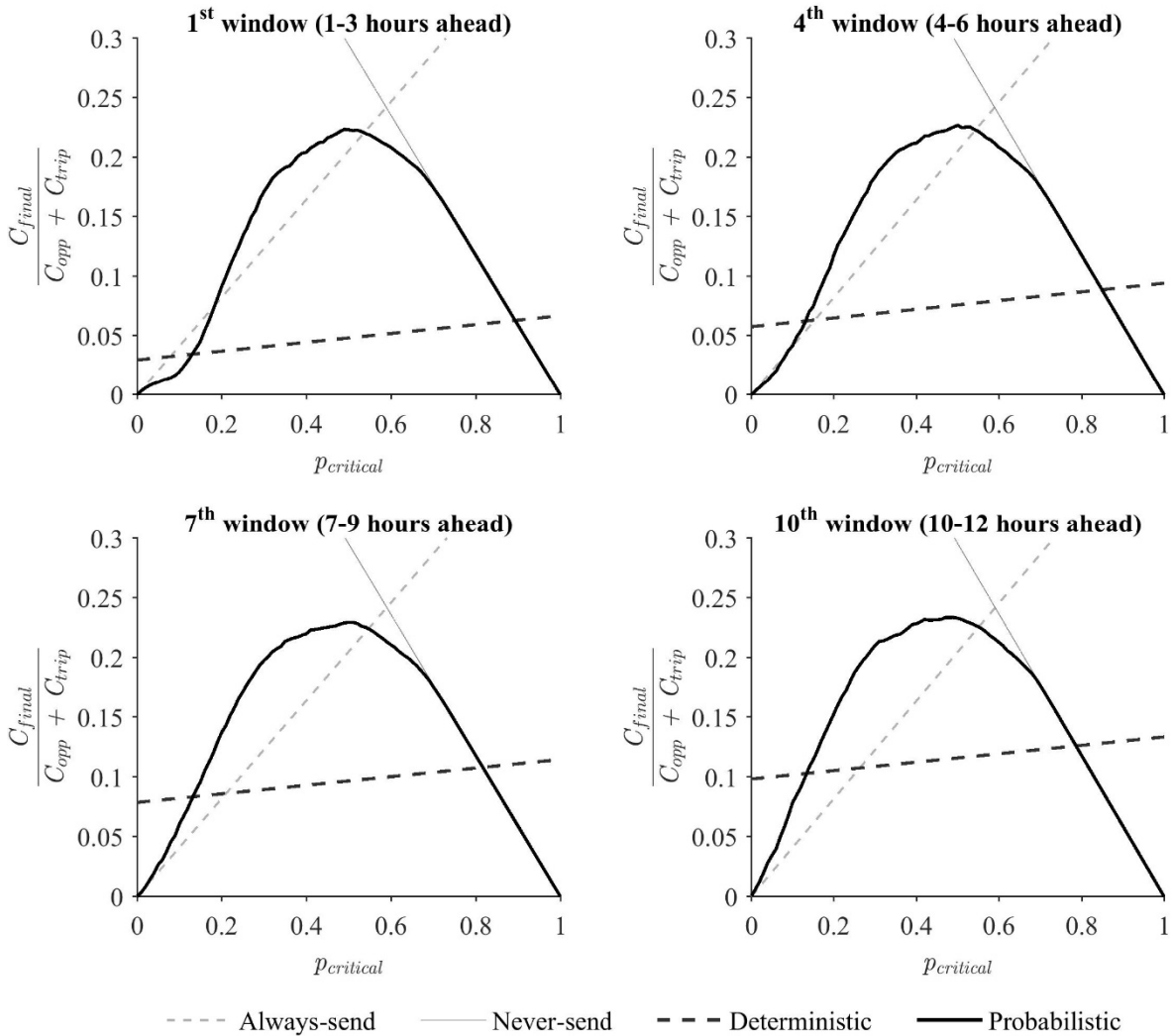
(a) Always send the service vessel.

(b) Never send the service vessel.

(c) Send the vessel if and only if the wave height point forecasts are less than 1.5m for all 3 hours within the mobilization window. The point forecasts are derived from density forecasts, but because the method does not use the probabilistic information in the density forecast, we refer to this as a *deterministic* approach. This deterministic approach was used in the work of Catterson et al. (2016).

(d) Send the vessel if and only if the forecast of p is greater than $p_{critical}$. This is our proposed approach, and we refer to it as the *probabilistic* approach. This approach was used with $p_{critical} = 29.1\%$ to produce the results of Table 6.

In Fig. 5, lower cost values are better. The figure shows that approach (d), which uses the KDE probability forecasts, was not the best for almost all values of $p_{critical}$, and almost all four lead times. The deterministic approach of basing the decision on the KDE point forecasts was the best approach for most values of $p_{critical}$ for all four lead times. Our explanation for this is that the KDE probability forecasts were relatively poor, as shown by the Brier score results of Table 5, while the KDE point forecasts were more competitive, as shown by the MAE results of Table 4.

Fig. 5. The comparison of the costs based on four approaches: (a) always send the vessel, (b) never send the vessel, (c) use the deterministic point forecasts of KDE with a window length of 4 hours, and (d) use probabilistic forecasts from this method.

Fig. 6. The comparison of the costs based on four approaches: (a) always send the vessel, (b) never send the vessel, (c) use the deterministic point forecasts from the TVP regression method, and (d) use probabilistic forecasts from this method.
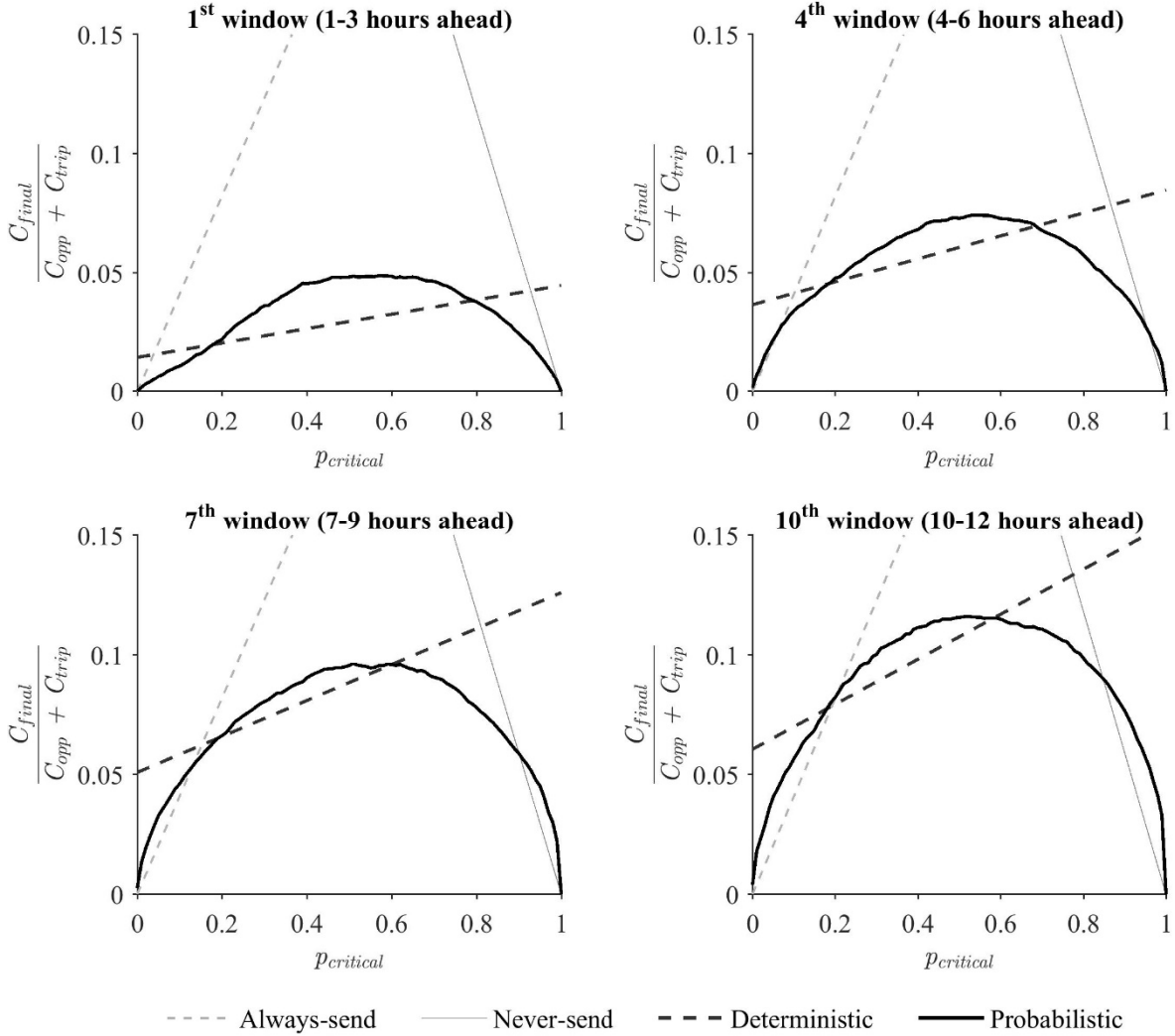

Fig. 6 shows the plots for the TVP regression method. The first point to note in this figure is that both the deterministic and the probabilistic approaches of the TVP regression method show smaller ratios of C_{final} to the sum of C_{trip} and C_{opp} than for KDE in Fig. 5. In Fig. 6, the probabilistic approach showed slightly enhanced performance, but is still not better than the deterministic approach for the values of $p_{critical}$ between 0.2 and 0.6 for all four lead times. This is not surprising given that the TVP regression probability forecasts were clearly poorer when judged by the Brier score in Table 5 than by the MAE in Table 4. The plots of Fig. 7 correspond to the bivariate ARMA-GARCH models fitted to wave height and wind speed, which employed the Box-Cox transformation and Student t distribution. This was the best method in Tables 3 to 6. Fig. 7 shows the probabilistic approach providing the best results for almost all values of $p_{critical}$ for all four lead times. This indicates that the probabilistic approach should be preferred to a deterministic approach based on point forecasts.

To examine the robustness of the results of Fig. 7 to post-sample periods from different times of the year, we estimated method parameters using 9 months of data, and evaluated the results using the following 3-month post-sample periods: 1 December 2014 to 28 February 2015, 1 March 2015 to 31 May 2015, and 1 June 2015 to 31 August 2015. We found that, for each of these three post-sample periods, the ranking of methods for all four lead times was similar to the ranking in Fig. 7, with the probabilistic approach outperforming the other three approaches for almost all values of the $p_{critical}$.

Fig. 7. The comparison of the costs based on four approaches: (a) always send the vessel, (b) never send the vessel, (c) use the deterministic point forecasts from the best bivariate ARMA-GARCH model of wave height and wind speed, and (d) use probabilistic forecasts from this method.
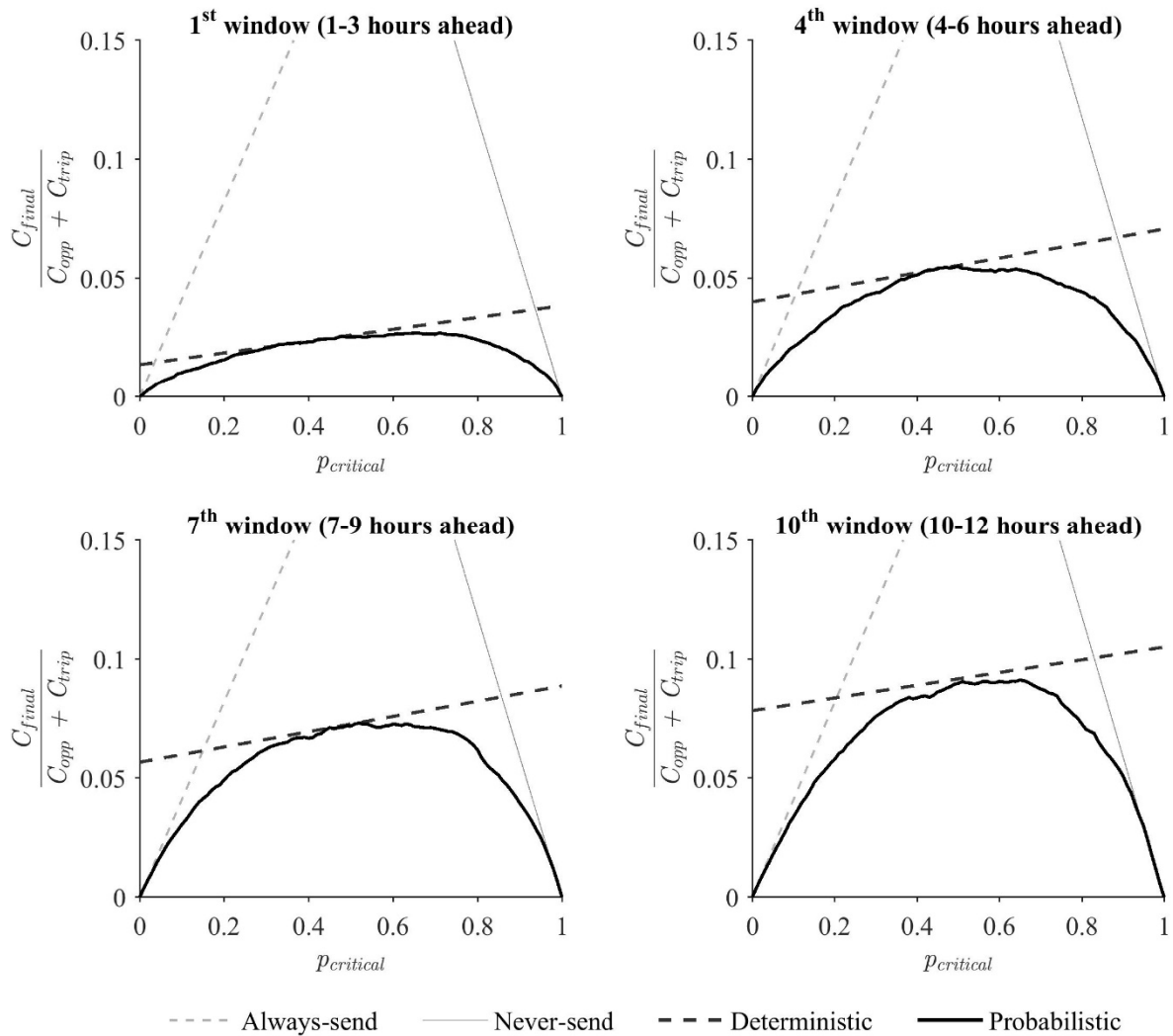
6. Summary and Concluding Comments

In this paper, we have described how wave height probabilistic forecasts should be used to make decisions regarding whether or not to launch service vehicles for offshore wind turbines. Our empirical analysis investigated probabilistic forecast accuracy using both statistical measures and the monetary cost resulting from decisions based on the probabilistic forecasts. The most accurate density

forecasting method was a bivariate ARMA-GARCH model of wave height and wind speed. The results also showed that there is monetary benefit in using a probabilistic approach to decision making, rather than a deterministic approach based on point forecasting. In future work, one possibility is to evaluate the use of weather ensemble predictions for forecasting at longer lead times. Also, the methods that we have considered could potentially be used in other offshore engineering contexts, such as the management of anchor handling operations (see Shyshou et al., 2010).

Acknowledgements

References

Azzalini, A., & Genton, M.G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76, 106-129.

Baillie, R.T., & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, 52, 91-113.

Bollerslev, T., Engle, R.F., & Wooldridge, J.M. (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96, 116-131.

Caroll, J., McDonald, A., & McMillan, D. (2016). Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*, 19, 1107-1119.

Catterson, V.M., McMillan, D., Dinwoodie, I., Revie, M., Dowell, J., Quigley, J., & Wilson, K. (2016). An economic impact metric for evaluating wave height forecasters for offshore wind maintenance access. *Wind Energy*, 19, 199-212.

De Menezes, L.M., Taylor, J.W., & Bunn, D.W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, 120, 190-204.

Dinwoodie, I.A., Catterson, V.M., & McMillan, D. (2013). Wave height forecasting to improve offshore access and maintenance scheduling. In 2013 IEEE Power & Energy Society General Meeting.

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29, 108-121.

GL Garrad Hassan. (2013). A Guide to UK Offshore Wind Operations and Maintenance. Scottish Enterprise and The Crown Estate.

Gneiting, T., Balabdaoui, F., & Raftery, A.E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 69, 243-268.

Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7, 1747-1782.

Gourieroux, C. (1997). *ARCH Models and Financial Applications*. New York: Springer.

Hyndman, R.J., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin, Heidelberg, Germany: Springer-Verlag.

Irawan, C.A., Ouelhadj, D., Jones, D., Stålhane, M., & Sperstad, I.B. (2017a). Optimisation of maintenance routing and scheduling for offshore wind farms. *European Journal of Operational Research*, 256, 76-89.

Irawan, C.A., Song, X., Jones, D., & Akbari, N. (2017b). Layout optimisation for an installation port of an offshore wind farm. *European Journal of Operational Research*, 259, 67-83.

Jeon, J., & Taylor, J.W. (2006). Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation. *International Journal of Forecasting*, 32, 991-1004.

Kapetanios, G., Mitchell, J., Price, S., & Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188, 150-165.

LeClerc, J., & Joslyn, S. (2015). The cry wolf effect and weather-related decision making. *Risk Analysis*, 35, 385-395.

Lichtendahl, K.C. Jr., Grushka-Cockayne, Y., & Winkler, R.L. (2013). Is it better to average probabilities or quantiles? *Management Science*, 59, 1594-1611.

Miller, A.R. (1958). The effects of winds on water level on the New England coast. Limnology and *Oceanography*, 3, 1-14.

Reikard, G. (2009). Forecasting ocean wave energy: Tests of time-series models. *Ocean Engineering*, 73, 168-178.

Reikard, G., Robertson, B., & Bidlot, J. R. (2015). Combining wave energy with wind and solar: Short-term forecasting. *Renewable Energy*, 81, 442-456.

Rieder, K.F. (1997). Analysis of sea-surface drag parameterizations in open ocean conditions. *Boundary Layer Meteorology*, 82, 355-377.

Roulston, M.S, Ellepola, J., von Hardenberg J., & Smith L.A. (2005). Forecasting wave height probabilities with numerical weather prediction models. *Ocean Engineering*, 32, 1841-1863.

Shyshou, A., Gribkovskaia, I., & Barceló, J. (2010). A simulation study of the fleet sizing problem arising in offshore anchor handling operations. *European Journal of Operational Research*, 203, 230-240.

Staid, A., & Guikema, S.D. (2015). Risk analysis for US offshore wind farms: the need for an integrated approach. *Risk Analysis*, 35, 587-593.

Taylor, J.W. (2017). Probabilistic forecasting of wind power ramp events using autoregressive logit models. *European Journal of Operational Research*, 259, 703-712.

Ursavas, E. (2017). A benders decomposition approach for solving the offshore wind farm installation planning at the North Sea. *European Journal of Operational Research*, 258, 703-714.

Winkler, R.L. (2015). The importance of communicating uncertainties in forecasts: overestimating the risks from Winter Storm Juno. *Risk Analysis*, 35, 349-353.

Yoder, M., Hering, A.S., Navidi, W.C., & Larson, K. (2014). Short-term forecasting of categorical changes in wind power with Markov chain models. *Wind Energy*, 17, 1425-1439.