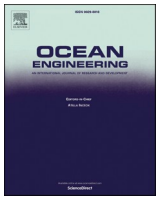
Ocean Engineering 285 (2023) 115265

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Effect of weather forecast uncertainty on offshore wind farm availability assessment

A. Kolios [a,*], M. Richmond [b], S. Koukoura [b], B. Yeter [b]

[a] Department of Wind & Energy Systems, Technical University of Denmark, Roskilde, Denmark
[b] Department of Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, Glasgow, United Kingdom

ARTICLE INFO

Handling Editor: A.I. Incecik

Keywords:
Offshore wind energy
Markov chains
Gradient boosting
Hybrid regression approach
Operation and maintenance modelling

ABSTRACT

With the growing demand for offshore wind energy and the continued drive for reduced levelised cost of energy, it is necessary to make operation and maintenance activities more effective and reduce related costs. A key factor in achieving this aim is to more representatively model operation and maintenance activities, and to do this, simulation models should include more accurate weather forecasting algorithms. In this paper, three weather forecast modelling methods are used to generate projections of wind and wave values which are then used as inputs in an operation and maintenance simulation model. These methods include Markov Chains, gradient boosting and a novel hybrid regression/statistical approach which has been developed and is presented herein. The change in key performance indicators after the wind farm lifespan is simulated using the forecasting methods and then compared to one another. It is shown that the Markov Chain and hybrid models numerically perform similarly, although the hybrid method has some additional desirable features. Finally, it is shown that the effect of this type of modelling uncertainty leads to significantly differing performance estimates through the operation and maintenance model.

1. Introduction

Offshore wind energy constitutes a significant portion of the exponentially growing European Union (EU) renewable sector. The offshore wind energy market has grown substantially due to factors such as the EU's increasingly competitive electricity production costs, the use of advanced technologies and limited lifecycle carbon emissions. In the EU, both the United Kingdom (UK) and Germany have taken the lead by accounting for 85% of all installations: 1.3 GW and 969 MW, respectively, in 2018. For example, European wind energy recorded 2649 MW of net additional capacity by connecting 409 new offshore wind turbines (OWTs) to the grid across 18 projects in 2018 (Wind Europe, 2019). According to the 'Offshore Wind in Europe Key Trends and Statistics' report published in 2019, Europe now has a total installed offshore wind capacity of 18,499 MW which is produced by 4543 grid-connected OWTs across 11 countries. As per the 'Wind Europe Outlook to 2023 report' (Fraile et al., 2018), the EU could install 90 GW of new wind energy capacity over the next five years. This projection suggests that clear government policy on National Energy & Climate Plans will be adopted and current issues concerning Operation & Maintenance (O&M), grid infrastructure and wind farm permitting will be resolved. Addressing these issues may allow the installation, combined onshore and offshore, of 277 GW of installed wind capacity in the EU by 2023.

Costs related to O&M constitute a significant part of the overall levelised cost of energy (LCoE) of wind energy (Ioannou et al., 2018), (Pandit et al., 2020). According to the European Wind Energy Association (EWEA) (European Wind Energy Association), O&M costs typically account for 20%–25% of the total LCoE of current wind power, assuming a twenty-year life span. Moreover, it is worth noting that the O&M costs are not evenly distributed over time; in fact, they fluctuate considerably over time, even more so towards the end of the service life. Offshore maintenance activities can be planned, condition-based or corrective (British Standards, 2010), (Scheu et al., 2019); hence, the O&M costs can be divided into fixed or variable types. Although both fixed and variable O&M costs contribute to a significant proportion of LCoE, the variable O&M costs (unexpected failures and unplanned maintenance) constitute the largest share of the O&M cost (Morthors and Awerbuch, 2009).

Offshore wind turbines can be installed in locations with harsh marine environmental conditions, which increases O&M costs due to the challenge of accessing and conducting maintenance activities on the OWTs, especially for cabling and towers (Ioannou et al., 2019), (Mytilinou and Kolios, 2019). Although turbine design and manufacturing processes have improved significantly over the last decade, severe sea

* Corresponding author.
E-mail address: atko@dtu.dk (A. Kolios).

https://doi.org/10.1016/j.oceaneng.2023.115265
Received 22 August 2022; Received in revised form 6 June 2023; Accepted 25 June 2023
Available online 17 July 2023

Nomenclature

EU	European Union
UK	United Kingdom
EWEA	European Wind Energy Association
OWT	Offshore Wind Turbine
LCoE	Levelised Cost of Energy
O&M	Operation and Maintenance
SCADA	Supervisory Control and Data Acquisition
SVM	Support Vector Machine
AR	Autoregressive
MPR	Multivariate Polynomial Regression
ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
XGBoost	eXtreme Gradient Boosting
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional LSTM
GRU	Gated Recurrent Units
MC	Markov Chain
LightGBM	Light Gradient-boosting Machine
IEC	International Electrotechnical Commission
ABS	American Bureau of Shipping
API	Application Programming Interface
PoF	Probability of Failure
i_n	State i at time n
$i_{n\text{-}1}$	State i at the previous time n-1
P_{ij}	Probability of a certain state i at time j
$F_m\ (X)$	New iterated model for gradient boosting
$F_{m\text{-}1}\ (X)$	Previous model for gradient boosting
$L(\ )$	Loss function
$h(X,a_m)$	constrained negative gradient
ρ	Line search along the direction of the gradient
u10	Wind speed at 10 m distance from the surface
Hs	Significant wave height
mean, μ	Mean value
std, σ	Standard deviation
NPV	Net Present Value
CTV	Crew Transfer Vessel
JUV	Jack-up Vessel
KPI	Key Performance Indicators

environments and increasing rotor blade size continue to pose a threat to the reliability of offshore wind assets and their operations (Carroll, McDonald, McMillan). Multiple factors influence offshore maintenance-related activities. These factors are restrictive and random maintenance windows due to weather conditions, transport and logistic issues, poor and restricted reliability-oriented field data, the massive scale of offshore deployment, and the assessed probability of different component failures (Carroll, McDonald, McMillan), (Pandit and Infield, 2018).

Regarding severe weather conditions (e.g., hurricanes), offshore O&M activities can be life-threatening, and they cause, such as unexpected catastrophic failures, unavailability and harsh weather conditions that significantly delay the access of the turbine for inspection and maintenance purposes, leading to revenue loss (Cevasco et al., 2021) (Stetco et al., 2019). For the given reasons above, offshore maintenance & reliability activities are more challenging than onshore; hence, they are more costly, which can cause significant financial loss to the turbine manufacturer (Horn, Leira). Nevertheless, more accurate prediction models of these factors can substantially reduce offshore O&M costs (Koukoura et al., 2021), (Pandit et al., 2020).

It is vital to direct the research efforts on optimising offshore maintenance strategies by developing robust weather forecasting tools enabling further O&M cost reduction. Consequently, O&M cost reduction also helps the cost-effectiveness of wind turbine technology. To this end, considerable research has been carried out in recent years in order to address the most pressing O&M issues (Leimeister, Kolios), which promises significant cost reduction and higher return on investment. The most relevant works are briefly reviewed below.

In the past, several data-driven models for time-series forecasting techniques have been proposed for prediction-related issues of OWTs, such as deep learning (Torres, Aguilar, Zuñiga-Meneses), autoregressive (AR) (Cavalcante et al., 2017), autoregressive integrated moving average (ARIMA) (Shukur and Lee, 2015), support vector machine approach (SVM) (Mohandes et al., 2004), and artificial neural networks-based approaches (Chang et al., 2017). Based on these fundamental approaches, many hybrid approaches have also been proposed in recent years (Yan et al., 2016), (Shi, Guo, Zheng). However, OWT data, like wind speed, power, and wave height, are generally non-linear and can also be non-stationary, which jeopardise the validity of these forecasting techniques.

One of the most prominent limitations of these techniques is extracting enough sequence data features to obtain accurate time series forecasting results (Pandit et al., 2019). These data are stored in SCADA systems and are critical for offshore wind in improving condition monitoring, optimisation and uncertainty (Pandit, Kolios 2020; Martinez-Luengo et al., 2016; Pandit et al., 2019). Furthermore, some applications specific to O&M models have included regression-based approaches, predicting values using gradient boosting as well as predicting direction using clustering regression (Gilbert et al., 2021). A key aspect of O&M is the vessel's operating condition, and it is critical to predict the short-term wind and wave conditions so that well-informed decisions can be made regarding marine operations. In this regard, Wu et al. (2019) employed an adaptive network-based fuzzy inference system to predict short-term wind and wave conditions for marine operations, whilst Taylor et al. (Taylor and Jeon, 2018) suggested using ARMA-GARCH to decide whether the vessels should be dispatched or not.

Emmanouil et al. suggested that the Bayesian Network model can improve the accuracy and precision of the wave height predictions, leading to better weather window characterisation. The performance of the Bayesian Network models was found satisfactory in comparison to the well-known wind-generated wave model SWAN given that the model gave valuable information about the uncertainty and the relationships between the considered variables (Emmanouil et al., 2020). Loake et al. also acknowledged the modelling uncertainty associated with physics-based weather modelling and employed a Bayesian (hierarchical) model. The model offered an update on numerical forecasts of met ocean conditions and examined the corresponding error. The hierarchical model involved a linear regression to remove systematic forecast biases and generalised autoregressive conditional heteroscedasticity models to remove residual error structure evolving in time (Loake et al., 2022).

Time-series analysis and forecasting are major subjects within the ship and offshore industry. The prediction models developed over the years were applied to deal with different issues. For instance, Papandreou and Ziakopoulos predicted the fuel oil consumption of a very large crude oil carrier based on different machine learning techniques using data from sensors and simple weather data. The study compared the predictive accuracy of Multivariate Polynomial Regression (MPR), Artificial Neural Networks (ANN) and eXtreme Gradient Boosting (XGBoost) regression models, whilst XGBoost demonstrated the best performance with a high accuracy rate of 86% (Papandreou and Ziakopoulos, 2022). Pandit et al. also compared three sequential data-driven weather forecasting models, namely, data-driven models

long short-term memory (LSTM), bidirectional LSTM (BiLSTM) and gated recurrent units (GRU) for long-term weather forecasting. The performance of these sequential machine learning techniques was satisfactory, and the difference between the trained models was not found to be statistically significant (Pandit et al., 2022).

Just like O&M operations, the offshore wind installation is impacted by the uncertainty in weather forecast and time-domain numerical model. Wu et al. and Wu and Gao showed that the sea states allowing safe operation differed significantly when considering weather forecast uncertainty (Wu et al., 2022), (Wu and Gao, 2021). As expected, it was stated that allowable sea states decreased as the forecast lead time increased. A case study of an offshore wind turbine transition piece installation was reported by Guachamin-Acero and Li, where the uncertainty related to the frequency domain energy content of a sea state was considered for the assessment of operational limits (Guachamin-Acero and Li, 2018).

Regarding the O&M activities, the operation threshold is generally stated by the characteristic of the sea states, such as significant wave height. Tomaselli et al. presented an alternative to this commonly used operation threshold: the significant vertical displacement of the vessel bow and motion sickness incidence; the study was conducted using an agent-based model framework (Tomaselli, Dixen, Bolaños Sanchez, Sørensen). Another simulation-based model offering an opportunistic maintenance strategy was reported by Papadopoulos et al. A two-stage stochastic mixed-integer linear program was developed to account for the uncertainties originating from weather and on-site maintenance resources (Papadopoulos et al., 2023). Such high-fidelity development cannot be seen as competition for the data-driven predictive models; instead, they should be considered complementary.

It is worth noting that the main goal of physics-based and data-driven models is to minimise financial loss that might occur due to weather window misprediction. One can argue that the accuracy of such proposed models is a key performance metric; however, the economic impact of an incorrect forecast above or below critical wave height boundaries can also be a key performance metric, according to Catterson et al. It is only reasonable to expect more research to be conducted adopting such a key performance metric. However, one should also be careful with calculating economic consequences (opportunity cost and maintenance costs) (Catterson et al., 2016).

In light of the studies given above, it can be deduced that there is a growing interest in developing physics-based and data-driven predictive models to make O&M activities more efficient, in turn, reduce O&M costs. However, the functional differences between the possible simulation methods lead to a numerical difference in the weather predictions, thus leading to considerable differences from an operations perspective. This difference cannot be neglected as it can add up, resulting in a significant financial loss. The present paper aims to address this issue by comparing different numerical weather simulation methods aiming to find the most cost-effective simulation method yielding accurate results. Another significant aspect of this study is that it shows how the choice of method impacts key performance indicators from an O&M standpoint. As it is shown in the current study, the consequence of model choice can easily result in tens of millions of revenue differences over the project lifecycle.

Furthermore, the present work proposes a novel method for simulating weather, which is a hybrid regression and statistical approach, combining existing ideas and tailoring them for this use case. Fig. 1 illustrates diagrammatically the approach used to carry out this study. The present study aims to tackle the variability resulting in time-series weather model choice using the lifecycle model that has already been presented and validated in (Kolios et al., 2019). To this end, outcomes should be interpreted mostly through the difference between the calculated key performance indicators through different forecasting models. The lifecycle model in this figure is a simplification of the diagram shown in Section 5.2.

The outcome of this study can be especially beneficial for those developing O&M models, whether in industry or academia, aiming to decide which weather time-series input model to use. The new time series method can be of value to wind and wave time-series forecasting as well as other time-series applications.

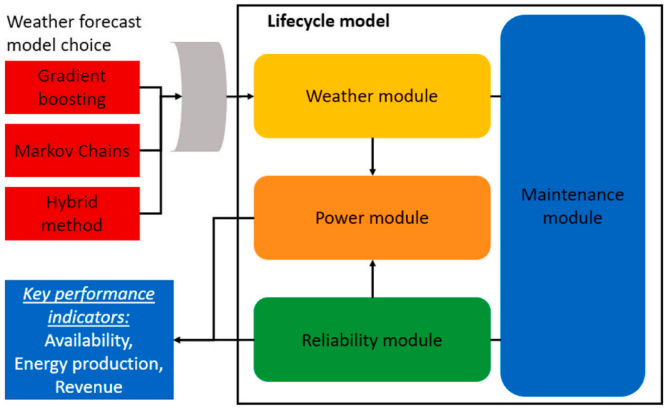


Fig. 1. Flowchart of study illustrating how weather forecasting models integrate into the lifecycle model.

The present paper is divided into six sections. Following the introduction, the paper continues in Section 2 with a literature review of the methods chosen, presenting how they have been previously used. The methods are then discussed in Section 3 in detail, including the specifics used for the current work. Section 4 presents the O&M assessment tool used for this study, explaining the individual modules and model outputs. The results are discussed in Section 5 by initially looking at the numerical differences between the forecasting method results, followed by the output from the simulation tool and a critical discussion. Finally, conclusions and future work are given in Section 6.

2. Literature review

2.1. Markov Chains

Markov chain (MC) is a widely accepted statistical tool for modelling a variety of natural phenomena, time-variant systems and signals (Grinstead and Snell, 1988), which makes it suitable for solving sequential time-series-related problems such as wind speed prediction (Pandit, Kolios, Infield). Long-term weather forecasting involves many datasets that lead to high computational and pre-processing costs. For example, in (Pandit, Kolios, Infield), LSTM (a deep learning method) and the Markov chain were compared, and it was found that the Markov chain technique is effective and fast in forecasting long-term weather conditions. In addition to that, environmental factors such as seasonality can easily be represented by a Markov model. However, representing such environments with decision trees would be confusing or intractable, if possible, and require major simplifying assumptions.

All these features given above make the Markov chain technique a suitable choice for this research. Other studies have also used Markov chains for both on and offshore wind (Tagliaferri et al., 2016), (Carpinone et al., 2015) and discussed the uses in reliability forecasting (Scheu et al., 2017), (Chen et al., 2009). There are some limitations to using Markov chains for wind speed prediction. It has been found that the autocorrelation plots for short intervals are often inaccurate due to a lack of persistence in the real data. Consequently, they do not perform well in ranges from 15 to 40 min, and it was found that the synthetic wind data fluctuated more rapidly in the first order MCs than the real data for these time scales (Brokish and Kirtley, 2009).

Nevertheless, modifications for Markov chains have been devised, which improve upon this aspect and others. For example, the aforementioned limitation was addressed by incorporating a second lag as well as a running-average filter (Pesch et al., 2015). Other

improvements include nested Markov Chains, which can better capture the temporal self-dependence in wind speed (Tagliaferri et al., 2016). Moreover, another way to further improve is related to accounting for the uncertainty in the transition matrix through Bayesian inference, which in a case study, was found to outperform traditional Markov Chains in a credible interval criterion. Another interesting approach is first to cluster the wind data as a method of discretisation, which was done on two-dimensional anemometer data and found to show the 'characteristic' wind behaviour of the site (Sánchez-Pérez et al., 2016).

There is a wide range of applications of Markov Chains in short-to-medium-term maintenance planning. The predictive models do not need to be limited to weather forecasting. However, instead, they incorporate future vessel costs and availability and the current condition of the turbine, which do not necessarily have to be a Gaussian and stationary process (Dawid et al., 2016).

It is essential to note that the key aspect of the present paper is seasonality. There is seasonality in the real wind data, which is to be represented in the model. Chen et al. (2009) trained separate models for summer and winter in order to incorporate seasonality. With only one year of training data, Karatepe et al. (Karatepe and Corscadden, 2013) trained a separate Markov Chain for each month. Not only did they find that this captured the seasonality present in the wind speed data, but they also found that one month of data per model was sufficient to capture the statistical properties.

2.2. Gradient boosting and trees

Gradient boosting is often used in stochastic value prediction (Esteoule et al., 2019). The method has been used to win two Global Energy Forecasting competitions (Hong et al., 2014), (Hong et al., 2016) where it was required to forecast the future and back-cast missing data. The method works well at forecasting at different time scales. In some research, it has been used to predict 1–6 h ahead (Persson et al., 2017), others up to one day (Verbois et al., 2018), and others have used it for even longer time scales (Browell et al., 2017); however, all found the method to perform very well compared to other methods tested in those studies. Boosted regression trees have also been used to predict maximum wind speed based on geographical considerations; though this is not forecasting, it shows the diversity of the approach (Fischer et al., 2015).

Developments have been found to improve upon the gradient-based boosting method for specific problems. For example, Cai et al. (Pesch et al., 2015) could incorporate related data sources into their training by incorporating instance-based transfer learning and improving against a benchmark approach. Chen et al. (2015) improved Gradient boosting by incorporating the Markov chain mixing rate to derive upper bounds in the loss function, enhancing the problem's convergence.

2.3. Probabilistic methods

Typically, statistical distributions are used for estimating sea states and wind speeds, which is the recommended practice in engineering standards as well. For instance, the prediction of annual energy production recommended by the International standards IEC 61400-12 is to be conducted with a two-parameter Weibull distribution (International Energy Agency, 1994). Similarly, the American Bureau of Shipping (ABS) recommends a two-parameter Weibull distribution for mean wind speed estimation; however, it finds that sometimes a Rayleigh distribution is appropriate (American Bureau of Shipping, 2011) for short-term sea states. Researchers have used a range of distributions to model significant wave height and mean wind speed, particularly for investigations at the design phase. In literature, various distributions have been used, including Weibull (Prevosto et al., 2000), (Satheesh et al., 2005), mentioned before, as well as Gumbel (Persson, 2010), (van Gelder and Vrijling, 2000), Lognormal (Teng and Palao, 1996), (Burrows and a Salih, 1982) and many other distributions.

Seasonality has been incorporated into the use of statistical distributions for wind speed. Jaramillo et al. (Jaramillo and Borja, 2004) fitted distribution for winter, spring, autumn and fall for a site in Mexico. This is the same approach used in numerous other works in which wind speed at sites is characterised using a Weibull distribution (Bilir et al., 2015), (Perea-Moreno et al., 2019). A challenge in addressing seasonality was discussed by Erikson et al. (Erickson and Taylor, 1989), who found that much of the wind speed can exhibit non-Weibull behaviour at different times of the year. To address this, Drobinski et el. (Drobinski et al., 2015). proposed an alternative that better fits both the diurnal and seasonal variability.

3. Weather forecasting methods

The methods in this study are trained using wind, wave and wind direction data recorded from 1992 to 2016 (24 years). These values are averages for each 3-h observation.

3.1. Markov Chains

A Markov Chain is a stochastic process which defines a set of probabilities for the next possible set of states, given the current state. The set of probabilities for moving from one state to the next is called transition probabilities and depends only on the state being moved from. Each possible discretised numeric value can be defined as a state in a regression problem. In the Markov chains method, the probability of any state at time t_n for a state, i, within a countable set of states, S, is independent of all previous states except for the last one, as shown (Tolver, 2016):

$$P\{X(t_n)=i_n|X(t_1)=i_1,\ldots X(t_{n-1})=i_{n-1}\}=P\{X(t_n)=i_n|X(t_{n-1})=i_{n-1}\} \quad (1)$$

Forecasting into the future is accomplished as a series of steps, and any state in the future depends on all the probabilities between those states. With a given start time i, the probability of a certain state at time j which is r steps from the value, is provided in the Chapman-Kolmogorov equation (Grinstead and Snell, 1988):

$$P_{ij}=\sum_{k=1}^{r}P_{ik}P_{kj} \quad (2)$$

Markov chains are described more in-depth in (Grinstead and Snell, 1988), (Tolver, 2016), (Serfozo, 2009). In this study, the Markov Chain model used is a first-order, observation-driven model which generates a probability vector of wave height given the previous (t_{n-1}) wave height, with a separate matrix for each month. A probability matrix of wind speed given for a wave height and a matrix of wind direction given for a wave height are constructed. These are constructed by counting the number of occurrences within each matrix position and then normalising them by the total number. In order to determine each next iteration, the model samples from the probability matrixes given the current state.

3.2. Gradient boosting - LightGBM

Gradient boosting is applied in this study using LightGBM, which is an API for Python (Microsoft Corporation, 2020a). Gradient boosting is an ensemble of weak learners where new learners are added sequentially in such a way as to minimise the gradients of the loss function. A new model is added at each iteration, aiming to correct the previous model's error. Each new term fits the residual of the previous model, $R=y-F_m(x)$. By identifying this as the gradient of the squared error loss function, the method can be generalised with other loss functions. Each new iterated model, $F_m(x)$, is defined from the previous model, $F_{m-1}(x)$, in the following equations (Friedman, 1999):

$$F_m(x)=F_{m-1}(x)+\rho_m h(x;a_m) \quad (3)$$

where $h(x; a_m)$ is the constrained negative gradient. The 'line search', ρ_m, along the direction of the gradient is defined as:

$$\rho_m = \arg\, min_\rho \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m)) \tag{4}$$

$L()$ is the loss function between the value y_i and the previous model plus the new term.

In applying this method, the time variable, t, needs to be converted to lag variables. A starting point is variables of 'month', 'day', and 'hour'. However, this was found not to be enough to capture the time-series behaviour. In the same concept, generic lag variables were also created where a repeated duration is subdivided into a set of intervals. These durations were chosen from the optimised time lengths found through the hybrid statistical method discussed in the next section.

There are many parameters which can be set when implementing a LightGBM model (Microsoft Corporation, 2020b). A Gradient Boosting Decision Tree is used for training with the loss metric of mean absolute percentage error.

3.3. Hybrid probabilistic method

The fundamental premise of the proposed concept in terms of time-series forecasting is that any time series comprises three components: trend, seasonality and randomness. The trend is the gradual change over time, seasonality is repeated patterns expected to continue, and randomness is behaviour that cannot be fully explained based on past patterns. To an extent, trend and seasonality can be removed from time-series data by various approaches, leaving only the randomness of the data. This randomness can be parameterised and replicated through distribution fitting and sampling. This is insufficient if one is interested in predicting accurate values for any given observation. However, when the objective is to simulate rather than predict, then this can be appropriate, as shown in the current study.

The hybrid statistical/regression method proposed here combines sampling from statistical distributions with regression of parameters. The distribution parameters are fitted to the time series data through repeated polynomial regression. In the case of a univariate model, these parameters are the mean and standard deviation as they vary with time and are used to fit a normal distribution. In the case of a multivariate model, this is the mean and the covariance between the training data arrays, and this is used to fit a multivariate normal distribution. Both of these distributions are widely used in the literature (Gut, 2009). For wind speed prediction, normal distributions tend not to be used, and others are more common (Carrillo et al., 2014)– (Cakmakyapan, Kadilar). However, the data are much closer to being described by a normal distribution than when the seasonality is removed, and the needing a few parameters makes the normal distribution appealing.

The seasonality and trend are first removed from the data by the iterative fitting of polynomial regression curves, where the curve is subtracted from the training data before the next iteration is conducted. The first curve is a first-order polynomial to remove the trend, and the following sets of polynomials are all fourth order as this appears to fit the trends well. A fitting approach based on the least-squares-estimate is used.

The regression fitted to seasonality should cover each time span for which there is a significant repeated pattern. This may initially be taken as each year, each month, and each day; however, this is insufficient. The process for selecting the seasonality time spans is automated in this approach. At each iteration, a curve for all time lengths below a set value is fitted to and then subtracted from the data. The time span leading to the lowest standard deviation in the rolling average for the new, de-trended data is then used, and another iteration is conducted. The new maximum time length for the next iteration is a value just under the previously selected time length. The reduction in the error metric is shown in Fig. 2. These parameters can be visualised over the measurement values, as in Fig. 3.

A curve for the standard deviation (for a univariate model) or the covariance (for a multivariate model) is then fitted to the de-trended data using only an annual time scale. Afterwards, these fitted curves can be extended into the future. This whole process is shown diagrammatically in Fig. 4. One standard deviation is shown in Fig. 4 to represent the model, yet not how values are predicted. Further, quantiles are often used for such presentations; however, it opts not to use them here as they could imply some other behaviour of the model.

The model can be expressed by a probability density function $f(t)$ as in Eq. (5) (univariate) or Eq. (6) (multivariate).

$$f(t) = \frac{1}{\sigma(t)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu(t)}{\sigma(t)}\right)^2} \tag{5}$$

$$f(t) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma(t)|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-\mu(t))^T \Sigma^{-1}(z-\mu(t))} \tag{6}$$

where the parameters are time-dependent and are calculated in this model as follows:

$$\mu(t) = \sum_{n=0}^{N} \sum_{k=0}^{K} a_k t^k \tag{7}$$

$$\sigma(t)\ or\ \Sigma_{i,j}(t) = \sum_{k=0}^{K} a_k t^k \tag{8}$$

where N is the number of optimised timescales, i.e. the number of curves, and K is the polynomial order of each curve. The items i and j are

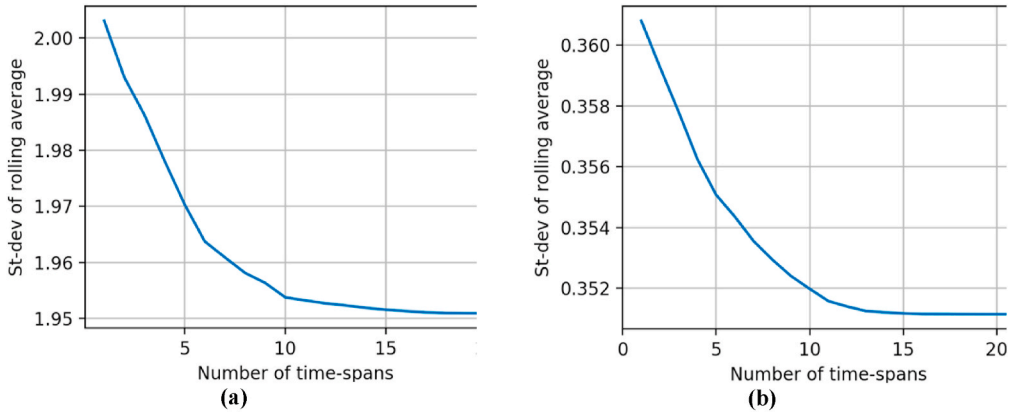


Fig. 2. Reduction in error metric, the standard deviation of the rolling mean, for the de-seasoned data at each iteration. (a) u10, (b) Hs.

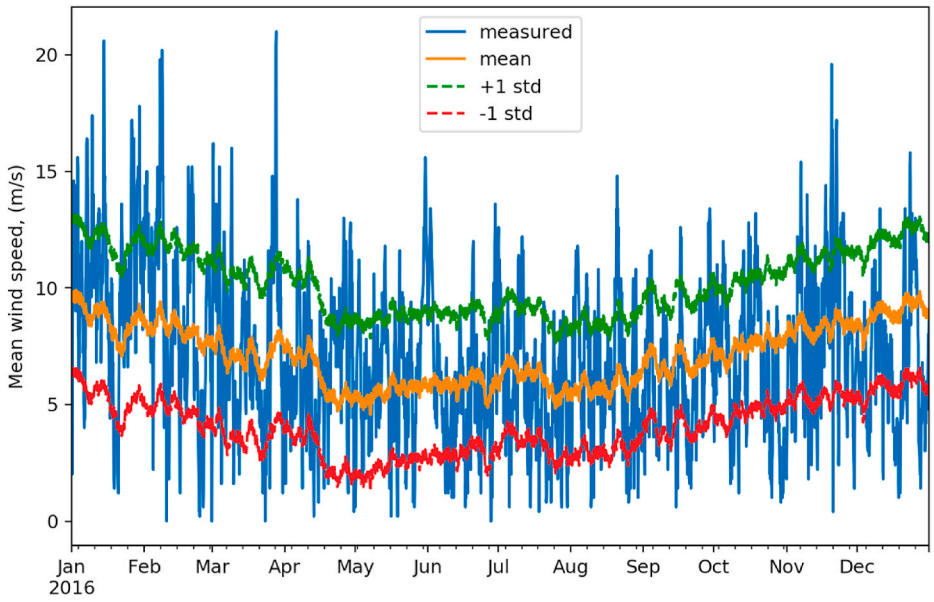


Fig. 3. Mean curve and the mean+/-1 standard deviation over the measured values for wind speed for 2016.

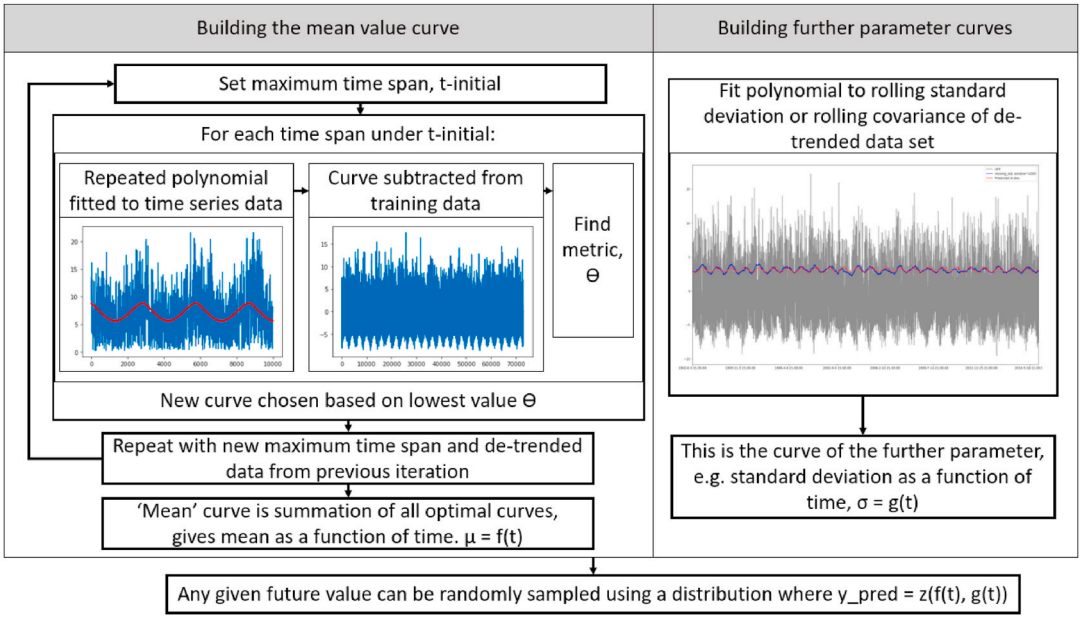


Fig. 4. Hybrid method process diagram showing the approach used to fit curves for distribution parameters from time series data.

indices in the covariance matrix. To match the rate of change of the real measurements, samples are taken every five intervals and linearly interpolated between these.

3.4. Strengths and weaknesses of methods

Table 1 presents the strengths and weaknesses of the methods introduced in Section 3.

4. O&M simulation model

A time-domain simulation model based on the Monte Carlo theory has been developed and deployed for offshore wind farm lifecycle O&M simulation and operational activity assessment. The model has been programmed by the authors, and its initial version has been used to assess the availability of various operating scenarios of current large-scale offshore wind farms (Chiachío-Ruano, Hermile, Kolios), (Kolios et al., 2019). A brief explanation of the tools' structure and individual modules is provided in the following and illustrated in Fig. 5. The tool is modular, consisting of a weather module, a power module, a reliability module and a maintenance module.

4.1. Weather module

The weather module aims to provide a forecast for wind and wave conditions, which is utilised in the power estimation and maintenance modules to assess accessibility constraints. Historic metocean data representative of local conditions can be obtained, and time-series forecasting models are trained. The different models utilised in this study have been presented in Section 3.

Table 1
Comparison of Markov Chains, Gradient Boosting, and the hybrid regression method.

Method	Strengths	Weaknesses
Markov Chains	• Accurately replicates the probability distribution (Brokish and Kirtley, 2009). • Can be generated with limited data (e.g. one year) (Karatepe and Corscadden, 2013) • Well developed with improved implementations. (Tagliaferri et al., 2016)	• Seasonality accomplished by having separate models, not unified in one, therefore limited. (Karatepe and Corscadden, 2013) • Persistence dependent on prediction rate and limited to certain timescales. (Brokish and Kirtley, 2009)
Gradient Boosting (LightGBM)	• Very low computational requirement and quick to train. • Can be useful at a range of timescales. (Persson et al., 2017)– (Browell et al., 2017) • Flexible and suitable for various problems (Hong et al., 2014), (Hong et al., 2016).	• Not a statistics-based method, so it will not inherently capture statistical distribution. • Deterministic method and so cannot be used as a stochastic input without modification.
Hybrid Regression	• Captures trends at all significant timescales. • Level of persistence is a tuneable parameter. • Uses statistical distributions and is therefore closer to recommendations from standards. (American Bureau of Shipping, 2011), (BSI British Standards Wind turbines, 2009)	• Slow to train in its current form as the code is not optimised. • Depends on linear interpolation between sample points. • New idea and so not developed yet.

4.2. Power module

The actual power output in each time step is calculated based on the wind speed, wind turbine hub height and power curve. Met mast measurements provide wind speed, which is extrapolated at hub height using the power law (Gualtieri and Secci, 2012). The produced power is calculated using the respective wind turbine power curve for wind speeds higher than the cut-in and lower than the cut-out.

4.3. Reliability module

As regards the distribution of unforeseen failures in time, this information is modelled from the reliability module based on the reliability data from the literature. The input failure rates are grouped into minor repair, major repair and major replacement, according to the material cost reported by Carroll et al. (Carroll, McDonald, McMillan). When a failure occurs, the turbine status varies depending on the failure type. In minor repairs, the turbine is assumed to continue operation even after the failure detection, shutting it down only during the repair time. For major repairs and replacements, the turbine is stopped after detecting a fault, returning to service only after the fault is restored. The time to failure associated with each failure mode for a particular subsystem i is assumed to be distributed by an exponential probability density function $f(t)$, reported in Eq. (9), with parameter $\lambda_{i,mode}$ being the failure rate for subsystem i under a particular failure mode.

$$f(t) = \lambda_{i,\mathrm{mode}}\, \mathrm{e}^{-\lambda_{i,\mathrm{mode}} t} \tag{9}$$

The cumulative distribution function is the probability of failure (*PoF*) of the subsystem according to the exponential reliability theory and is given in Eq. (10). *PoF* of the whole wind turbine covers all subsystems considering all failure mode classifications, as explained more in-depth in (Kolios et al., 2019). The probability of a subsystem failing is randomly generated.

$$PoF = 1 - \mathrm{e}^{-\lambda_{i,mode} t} \tag{10}$$

4.4. Maintenance module

The maintenance module takes into account the basic technical data of the wind turbine and the farm that the simulation uses in the analysis. The lifetime corresponds to the number of simulated years, and the final availability is calculated as an average over the entire lifetime. The number of vessels, their crew capacity and their wave-bearing capacity are included in the available means of transport considerations. If any requirements are not met, maintenance work is not completed, and all remaining work is planned for the next suitable weather window. There are two types of maintenance activities considered in this model: planned and unplanned.

Planned maintenance is a scheduled service, whereas unplanned maintenance takes place as soon as a failure occurs. Downtimes are calculated accordingly, based on the maintenance duration, the weather conditions and the resource availability. Workboats are assumed to be filled to their maximum capacity for planned maintenance, as they can simultaneously perform operations on multiple turbines.

For unplanned maintenance, the O&M tool differentiates between

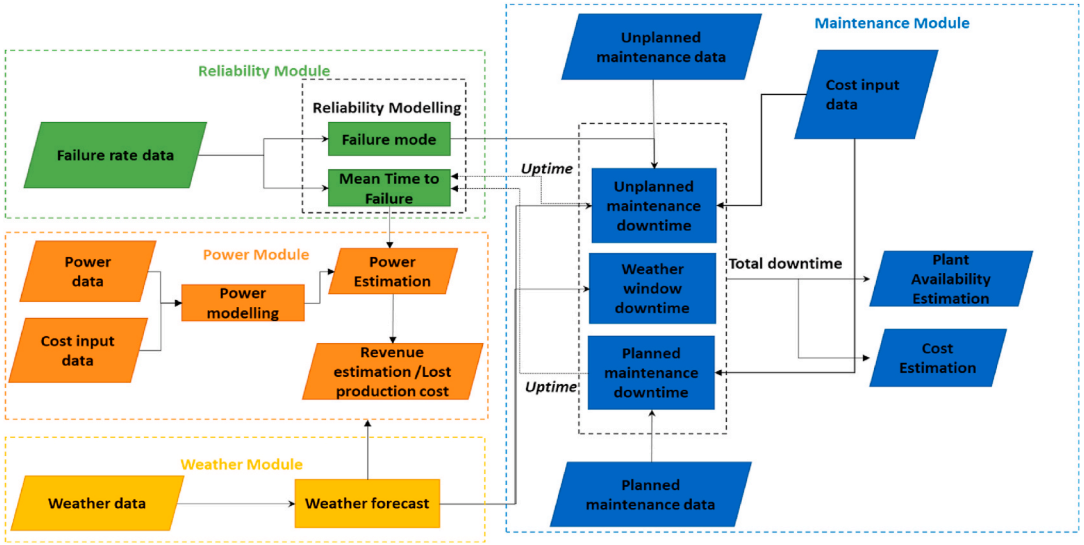


Fig. 5. O&M lifecycle assessment model flowchart.

failures requiring a JUV (jack-up vessel) and those requiring a CTV (crew transfer vessel). In order to decrease downtimes, respective maintenance campaigns are implemented, which do not only repair one turbine but store different maintenance tasks on a campaign list and follow this list during the campaign. While in one JUV campaign, all turbines are maintained for which a failure occurred within the lead time of ordering a JUV, the CTV campaign repairs all failures which occurred during the night when technicians rest. This difference is due to the usage of vessel type. A JUV is costly and needs to be ordered at the market, which takes time. Instead of just repairing one failure and ordering a JUV for another failure again, all pending turbines which need maintenance are served.

Moreover, JUV campaigns are performed in shifts to utilise the JUV to capacity. In comparison, CTV campaigns are only conducted during day shifts as no accommodation is available on this vessel type. All failures that occurred during the night are scheduled for the next day shift. In case not all turbines can be served, the campaign continues the next day.

The O&M strategy is based on a decision tree (as shown in Fig. 6) that follows a system failure in one or more wind turbines. In the event of a failure, it is first checked whether a crew and a ship suitable for the type of required repair are already on site. Component replacements are considered to require a crane vessel - all other system repairs are assumed to require a crew transfer vessel. The absence of a suitable crew-ship combination on-site leads to activating a ship or crane ship in port, if any are available. The activated vessel or barge will continue its transfer to the failed OWT as soon as weather conditions permit; environmental restrictions are limited to a certain wave height limit.

As soon as a failed system is put back into operation (status reached as soon as a crew ship combination has been placed on the failed OWT for the assigned repair duration), the subsequent failure for this system is determined similarly to how the original time-to-failure was generated. This process is repeated accordingly if a failed component is repaired or replaced.

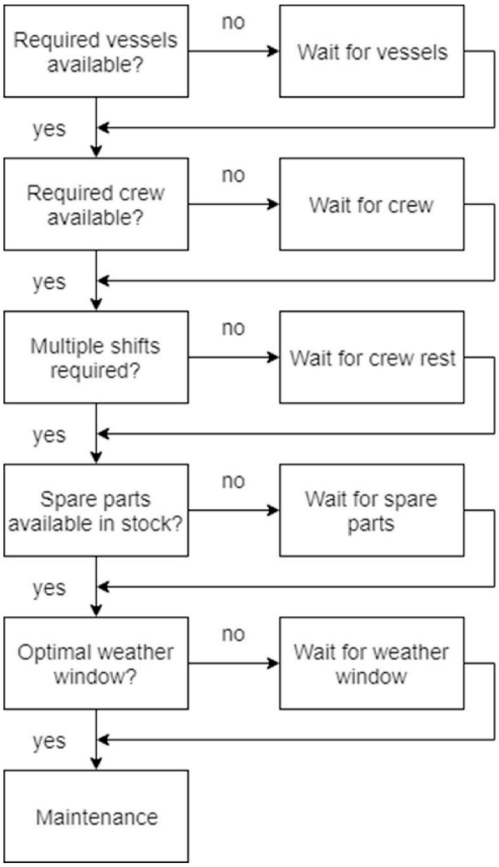


Fig. 6. Unplanned maintenance decision tree.

4.5. Cost module

The cost module is a subset of the maintenance module, and it calculates all monetary flows during the offshore wind farm lifecycle. The costs for each maintenance activity take into account the material costs of the replaced parts, the repair costs, the vessel's rent, fuel costs and the crew salaries based on their shifts. The revenue from the energy generated and sold is also calculated. The aforementioned cost estimations and revenue are integrated into cashflow analysis to calculate Net Present Value (NPV) as:

$$NPV = \sum_{r=1}^{n} \frac{R}{(1+i)^r} \# \tag{11}$$

where R is the net cash inflows-outflows during a period r and i is the discount rate.

4.6. Outputs

Maintenance activities are carried out until the end of the lifetime of all wind turbines in the wind farm. Various key performance indicators (KPI) are calculated thereafter. First, the downtimes of each turbine are added up, and the total wind farm availability is calculated, as shown below.

$$A = \frac{Lifetime_{wf} - Downtime_{wf}}{Lifetime_{wf}} \tag{12}$$

where A is the calculated wind farm availability, $Lifetime_{wf}$ is the cumulative lifetime of all wind turbines in the wind farm and $Downtime_{wf}$ is the cumulative downtime. Other key performance indicators calculated include the energy E generated from the wind farm as follows:

$$E = P \times t \tag{13}$$

where t is the time given in *hours*, and P is the power.

Other indicators are related to costs, such as the direct costs involved with operational activities, the indirect costs involved with lost power production and the revenue from the energy sold to the grid. Table 2 summarises the input data for the O&M simulation model.

5. Results and discussion

The present section presents and discusses the results of the three simulation methods. Section 5.1 evaluates the time series forecasts from a numerical perspective, describing how similar the synthetic forecasts can be to the real data in terms of several statistical summary parameters. Section 5.2 presents the differences between the three methods regarding the KPIs predicted from the lifecycle model. These KPIs are availability, energy production and revenue. Final remarks are given in Section 5.3, discussing the results and why different methods lead to different KPIs.

5.1. Forecasting models

The time-series forecasting methods are compared here in terms of their numerical similarity to the real time series. This is to show which is a closer approximation of the real data as well as to help explain differences which will be seen in the operational KPIs. The methods are compared for the year 2016, which is the last year that real data is available. There is a large amount of scatter in the real wind and wave values, so it is unreasonable to compare individual 3-h observations. Instead, the results are compared on a monthly basis.

Fig. 7 shows box plots of the real as well as predicted values for the

Table 2
Input data to the O&M simulation assessment model.

Input	Data
Failures	• Failure rate data • Subsystems • Failure categories
Weather	• Wind speed • Wave height
Cost	• Energy price • Interest rates • Material costs • Vessel costs • Crew costs
Planned Maintenance	• Maintenance times • Subsystem grouping • Required crew • Required main vessel type • Required support vessel type
Unplanned Maintenance	• Repair times • Required crew number • Required main vessel type • Required support vessel type • Spare stock initial • Spare stock minimum • Spare wait time • Mission organization time

year 2016. This indicates that the general results of all models are similar and roughly comparable with the real values. However, there are months when the mean values of predictions and the standard deviation level do not match as closely with the real data, particularly in December, whilst all methods overestimate the mean value. To investigate this further, the difference in mean and standard deviation are calculated.

Fig. 8 shows this difference, from the real measurements, in monthly mean values for wind speed and significant wave height for 2016. In both u10 and Hs the methods all underpredict the early months and overpredict the later months. There are a few months where all but one method shows a low difference. For instance, Markov Chain underpredicts the mean values for the month of October, whilst all other methods slightly overpredict. These discrepancies are typically minor, roughly 10% of the mean values for the month. These differences are summarised in the following table.

The results presented in Table 3 conclude that the tree-based approach in LightGBM gives predictions with mean values furthest from the real mean and comparatively significant maximum differences. Markov Chains and the two hybrid methods provide the closest monthly mean values, with the hybrid multivariate approach yielding slightly better results. However, it is not close enough to say definitively that the hybrid methods outperform Markov Chains in this, only that they are all similarly good.

With a few exceptions, all methods underestimate the level of deviation in each month, as shown in Fig. 9. LightGBM particularly underpredicts the level of deviation, which is to be expected as tree-based approaches use an average of a leaf, so some level of scatter in the measurements is averaged out. All methods are closest to the real value of standard deviation during the mid-months, such as 4 and 6; this can simply be due to the fact that there is much less deviation in the real values for these months. These differences in standard deviation are summarised in the following table.

The conclusion from Table 4 is that LightGBM shows the largest difference from the real standard deviation. In contrast, Markov chains result in the smallest with the Hybrid multivariate values very close to the Markov Chain values. The values for Markov Chains and the two hybrid methods are very close; therefore, it can only be concluded that Markov Chains perform slightly better (see Table 5).

The final value which is essential in replicating the forecasts

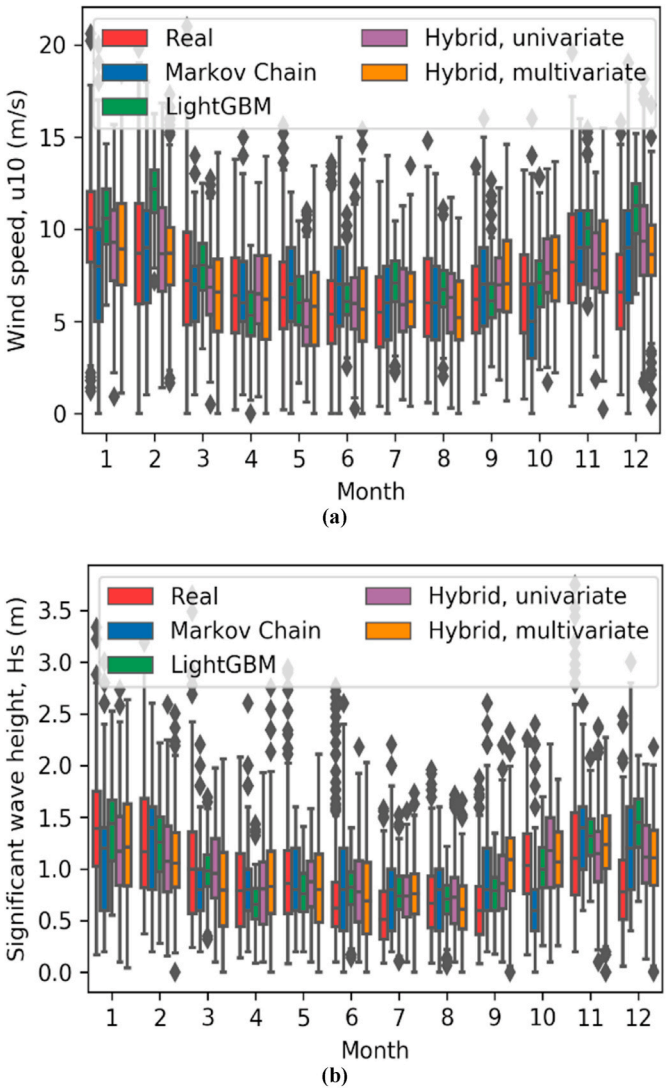


Fig. 7. Comparison of real monthly values with the predicted monthly values for wind speed (a) and significant wave height (b) for 2016.

compared to the real values is the level of correlation between the wind speed and the significant wave heights. This level of correlation is important for O&M modelling as some activities depend on both values. The Pearson correlation coefficients are given in the following table. This coefficient shows the level of linear correlation between two sets of variables in both magnitude and direction. A value of $+1$ indicates a perfect positive correlation, a value of -1 indicates a perfect negative correlation, and 0 indicates no relationship between the two variables.

The level of correlation in the real observations between u10 and Hs is roughly 0.8, which is also accomplished in Markov Chains and the hybrid multivariate approach. The tree-based LightGBM shows some level of correlation, and the hybrid univariate shows no correlation.

Based on the previous metrics, Markov chains and the hybrid multivariate approach appear to produce comparable estimates. However, looking within one month shows this not to be the case. There are several distinctions that can be seen in the plots for January 2016, as shown in Fig. 10.

The Markov Chain shows a higher persistence; it remains around a value for longer than the hybrid method. The persistence in Markov Chains is a fundamental quality of the model and is dependent on the training data. The hybrid method's persistence is determined by how frequently predictions are made, and so is a changeable parameter.

The Markov Chain captures seasonality within a year by training a

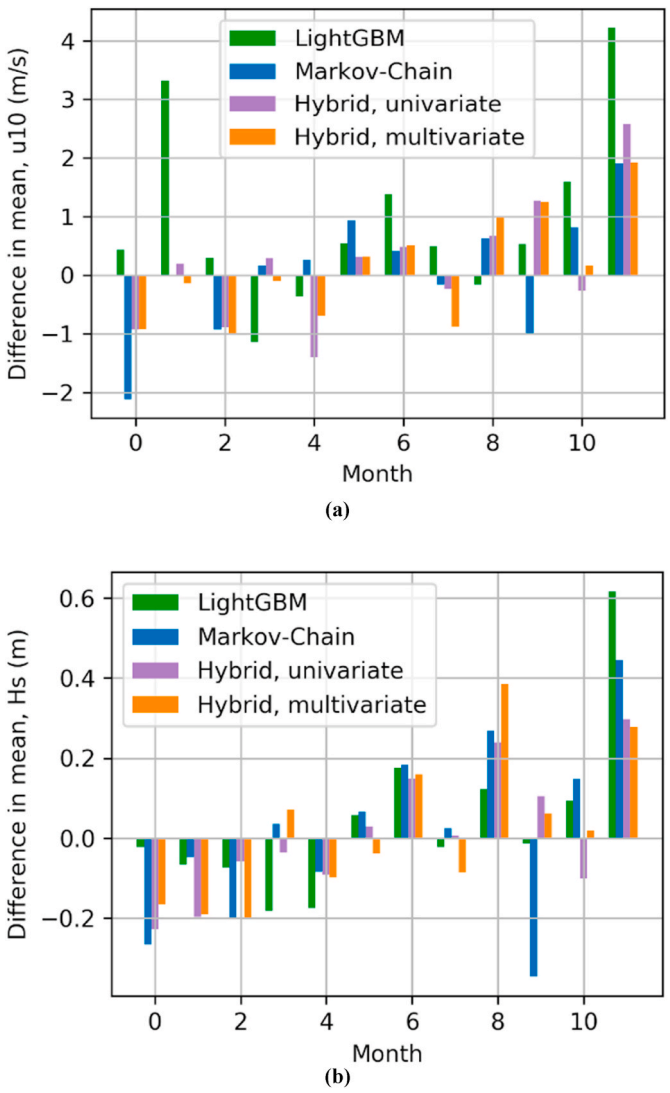


Fig. 8. Difference in mean, for each month, compared to the real data for 2016. (a) Wind speed, u10 and (b) Significant wave height, Hs, showing deviation from the real average.

separate model for each month. However, this approach does not capture any trend that occurs within a month. Conversely, the hybrid approach fits the parameters to trends at multiple timescales and shows trends within a month. This point about shorter trends is evident in Fig. 10, where the general movement of the hybrid method follows the real data more deliberately than the Markov Chains.

5.2. Availability assessment

The forecasting methods presented are compared using the lifecycle assessment model presented in Section 4. The comparison is based on the effects of each forecasting method on the respective KPIs calculated from the model.

Failure rates are based on the DTU 10 MW reference turbine (Bak et al., 2013). The components considered are the gearbox, generator,

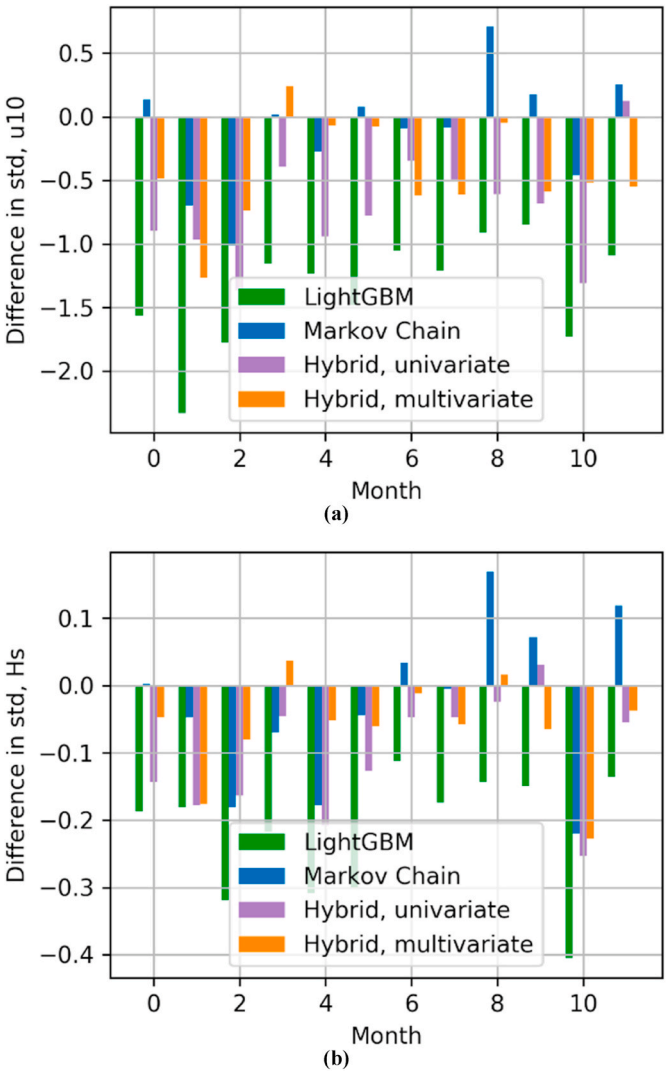


Fig. 9. Difference in standard deviation for each month compared to the real measurement values. (a) wind speed u10 and (b) significant wave height, Hs.

Table 3
Summary of the difference between the mean of real and forecast monthly values for 2016.

Table 3. Summary of the difference between the mean of real and forecast monthly values for 2016

	U10 average	U10 maximum	Hs average	Hs maximum
LightGBM	1.206	4.221	0.135	0.616
Markov Chains	0.776	2.119	0.176	0.445
Hybrid, univariate	0.793	2.579	0.128	0.297
Hybrid, multivariate	0.741	1.919	0.146	0.385

Table 4
Summary of the difference between the standard deviation of real and forecast monthly values for 2016.

Table 4. Summary of the difference between the standard deviation of real and forecast monthly values for 2016.

	U_{10} average	U_{10} maximum	H_s average	H_s maximum
LightGBM	1.367	2.334	0.220	0.406
Markov Chains	0.332	1.015	0.095	0.220
Hybrid, univariate	0.743	1.368	0.110	0.254
Hybrid, multivariate	0.485	1.267	0.072	0.228

Table 5
Pearson correlation coefficients between wind speed, u10, and significant wave height, Hs, for the real measurement observations and the forecasts.

Real	LightGBM	Markov Chains	Hybrid, univariate	Hybrid, multivariate
0.829	0.579	0.807	0.120	0.810

electrical system, pitch system, yaw system, blades and main shaft. More details on the wind farm layout can be found in (Carroll, McDonald, McMillan), whilst the repair information, including times and resources needed, is based on the data reported in (Bak et al.).

The O&M simulation model is run for the different weather prediction models, i.e., the decision trees, the multivariate hybrid and the Markov Chains. In total, 100 simulations were run for each of the three examined cases. The results are compared for the KPI outputs of each simulation. The KPIs examined are availability (as defined in Eq. (12)), the energy produced, calculated in the power module, and revenue, calculated in the cost module based on the energy sold. The effect of the various weather time-series prediction modelling methods presented on some KPIs is shown in the boxplots in Figs. 10–12. Each point of the boxplot represents a Monte Carlo simulation. Fig. 11 shows the availability, Fig. 12 the energy produced by the wind farm and Fig. 13 the revenue gained from the electricity generation.

5.3. Final remarks

Based on the results regarding the availability, tree-based methods result in the highest availability, followed by multivariate hybrid and Markov Chains. The tree-based approach captures the general trend reliably; however, it does not predict the extremely high values that sometimes occur. Tree-based algorithms are based on leaf averaging. As discussed earlier, the tree-based approach gives predictions with mean values farthest from the real mean value in the forecasting model results section. On the other hand, the hybrid approach manages to capture the level of scatter more realistically and sometimes predicts extremely high values. Even though Markov Chains and hybrid methods perform similarly in mean value and standard deviation, the difference in availability can be attributed to the difference in the level of persistence in the time series. Markov-Chains appear to change more slowly than the hybrid method, whilst the hybrid method is closer to the real data in terms of persistence.

Regarding the limitations of the study, it is worthwhile pointing out that the high availability of some methods might not sometimes be reflected in energy production since the energy produced on each simulation depends on a combination of factors, especially the randomness of failure simulations. Additionally, the hybrid method presented is new and should be developed further in future studies.

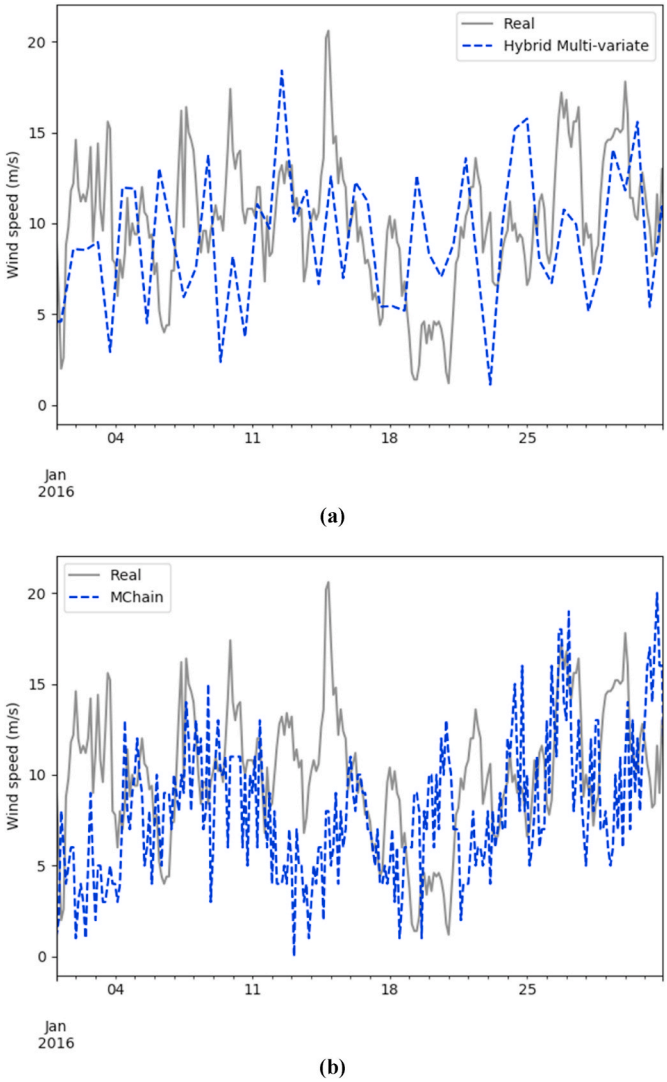


Fig. 10. Time series plots within the month of January 2016 comparing real values to predictions. The recorded wind speed time series is plotted in grey. (a) - prediction from the hybrid multivariate approach for that month, (b) - prediction from Markov Chain.

6. Conclusion

This paper compared numerical weather forecasting methods and their effect on lifecycle modelling. Three numerical methods were used: Markov Chains, gradient boosting and a proposed hybrid regression/statistical method. The forecasts from these methods were used as inputs

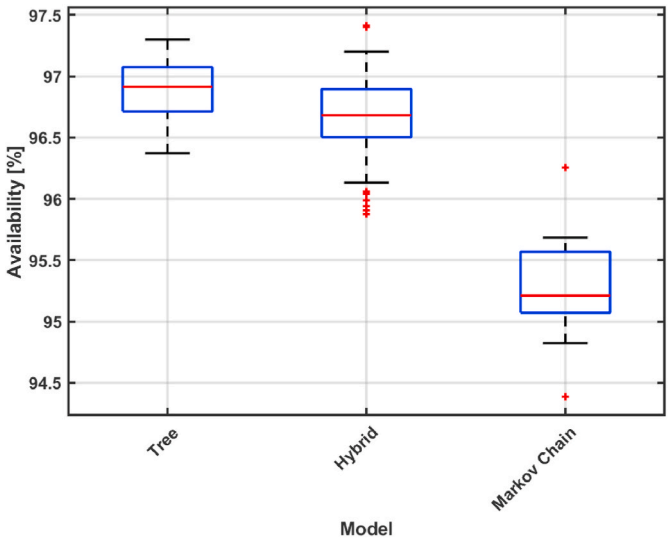


Fig. 11. Availability results of different forecasting methods.

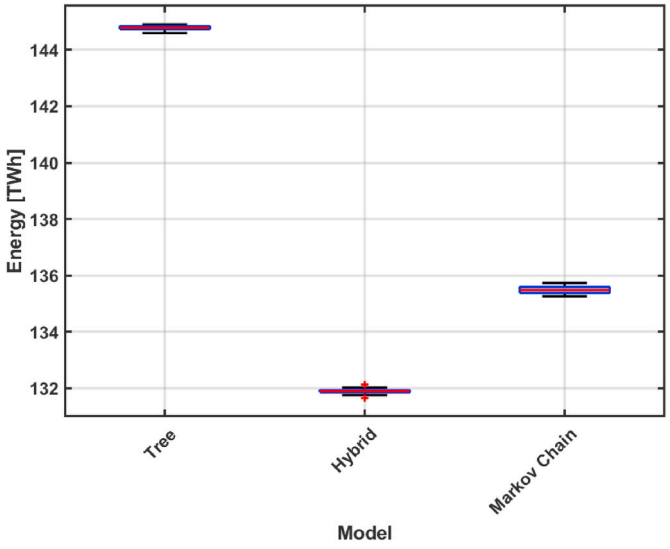


Fig. 12. Energy production results of different forecasting methods.

into a lifecycle assessment tool which models turbine performance, failure rates, and operational variables to determine key performance indicators in a Monte Carlo analysis. These key performance indicators are energy production, availability and revenue. The methods are compared both in terms of their numerical similarity to the measurement data and the difference in the results of the lifecycle model.

The O&M costs of offshore wind farms make up a significant portion of the total lifecycle cost, and unplanned maintenance activities O&M costs due to harsh marine environmental conditions. The predictive models, as described in the present study, allow for better-informed decision-making for these unplanned maintenance activities, which means significantly better returns from the offshore wind farm as well as safer operation. However, the precision of these predictive models is as important as the accuracy, as imprecise models can also lead to substantial differences in the availability of offshore wind assets, which is somewhat overlooked in the literature or has not been explicitly demonstrated using the relevant key performance indicators. The present study addresses this neglected yet vital issue described above by studying the uncertainties associated with these predictive models and their impact in monetary terms to support decision-making.

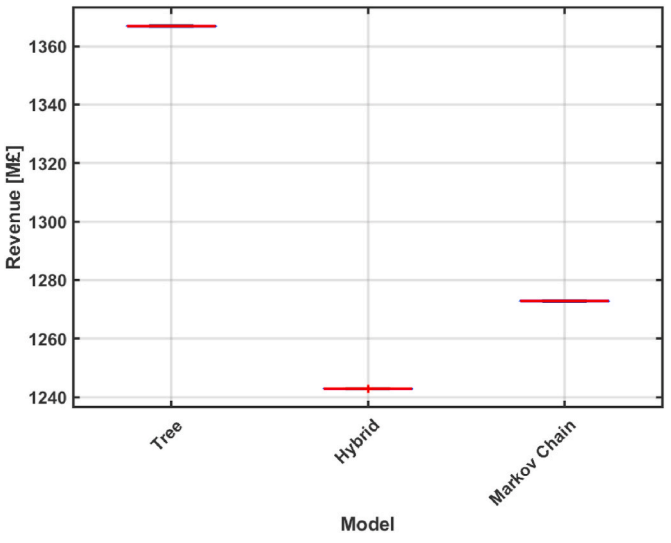


Fig. 13. Revenue results of different forecasting methods.

The results showed that the choice of weather forecasting method could easily lead to a difference in availability of nearly 2 per cent, in turn, a difference in revenue of tens of millions of revenue. Numerically, both Markov Chains and the hybrid method performed similarly to the measurement data. Markov Chains and the hybrid method are roughly equal in terms of prediction of monthly mean and deviation values as well as the correlation between wind and wave vectors. The most considerable difference numerically resulted from the gradient boosting method. The difference between Markov Chains and the hybrid method comes down to the level of persistence in the two methods and capturing trends at multiple timescales, where the hybrid method excels. The consequence in the lifecycle model is that Markov Chains predict the lowest availability, whilst the hybrid method predicts the lowest revenue.

The novelty of the present work comes from the fact that it calculates the consequence of numerical weather model choice on lifecycle modelling prediction. Such knowledge is a significant contribution to the literature because the prospective modeller may initially assume that two forecasting methods predicting the same monthly mean value and standard deviation would have the same key performance indicator. However, the present shows that this assumption does not necessarily have to be the case.

Moreover, this study also presented a novel hybrid forecasting approach that combines existing ideas tailored to this problem. The novel hybrid forecasting approach validated by the measurement data showed a strong predictive power. Future work will further focus on developing numerical forecasting models and advancing other aspects such as uncertainty propagation over the course of service life, which is critical for offshore wind lifecycle modelling. This work will be of value to practitioners who are developing maintenance strategies for offshore wind farms, as contractual relationships between developers/operators and the grid often depend on ensuring adequate levels of availability; accurate prediction of such value is critical towards reducing risks of such agreements. Further, the study can be of value to researchers and developers of numerical tools that utilise forecasting algorithms as it shows how different algorithms can be modelled and simultaneously discusses performance characteristics for each method applied.

CRediT authorship contribution statement

A. Kolios: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **M. Richmond:** Data curation, Formal analysis, Validation, Visualization, Writing – original draft. **S. Koukoura:** Data curation,

Formal analysis, Validation, Visualization, Writing – original draft. **B. Yeter:** Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 745625 (ROMEO) ("Romeo Project" 2018) (https://www.romeoproject.eu). The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

Furthermore, this work was also supported by the sponsorship of the University of Strathclyde for pursuing Doctoral Studies within the Centre for Doctoral Training in Renewable Energy Marine Structures - REMS (http://www.rems-cdt.ac.uk/).

References

American Bureau of Shipping, 2011. Design Standards for Offshore Wind Farms. Houston Texas.

Bak, T., et al., 2017. Baseline layout and design of a 0.8 GW reference wind farm in the North Sea. Wind Energy 20 (9), 1665–1683. https://doi.org/10.1002/we.2116.

Bak, C., Zahle, F., Bitsche, R., Kim, T., 2013. The DTU 10-MW reference wind turbine Report. Danish Wind power 22.

Bilir, L., Imir, M., Devrim, Y., Albostan, A., 2015. Seasonal and yearly wind speed distribution and wind power density analysis based on Weibull distribution function. Int. J. Hydrogen Energy 40 (44), 15301–15310. https://doi.org/10.1016/j.ijhydene.2015.04.140.

British Standards, 2010. BSI standards publication maintenance — maintenance terminology. BS EN 13306. https://doi.org/10.1007/s00168-003-0173-6.

Brokish, K., Kirtley, J., 2009. Pitfalls of modeling wind power using Markov chains. In: 2009 IEEE/PES Power Syst. Conf. Expo. PSCE 2009. https://doi.org/10.1109/PSCE.2009.4840265.

Browell, J., Gilbert, C., McMillan, D., 2017. Use of turbine-level data for improved wind power forecasting. In: 2017 IEEE Manchester PowerTech, Powertech 2017. https://doi.org/10.1109/PTC.2017.7981134.

BSI British Standards Wind turbines, 2009. Part 3: design requirements for offshore wind turbines. BS EN 61400–3:2009.

Burrows, R., a Salih, B., 1982. Statistical modelling of long-term wave climates. Coast. Eng. 42–56.

Cakmakyapan, S., Kadilar, G., 2016. The Poisson Gamma distribution for wind speed data. AIP Conf. Proc. 1726 https://doi.org/10.1063/1.4945896.

Carpinone, A., Giorgio, M., Langella, R., Testa, A., 2015. Markov chain modeling for very-short-term wind power forecasting. Elec. Power Syst. Res. 122 (April), 152–158. https://doi.org/10.1016/j.epsr.2014.12.025.

Carrillo, C., Cidrás, J., Díaz-Dorado, E., Obando-Montaño, A.F., 2014. An approach to determine the weibull parameters for wind energy analysis: the case of Galicia (Spain). Energies 7 (4), 2676–2700. https://doi.org/10.3390/en7042676.

Carroll, J., McDonald, A., McMillan, D., 2016. Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. Wind Energy 19 (6), 1107–1119. https://doi.org/10.1002/we.1887.

Catterson, V.M., et al., 2016. An economic impact metric for evaluating wave height forecasters for offshore wind maintenance access. Wind Energy 19 (2), 199–212. https://doi.org/10.1002/we.1826.

Cavalcante, L., Bessa, R.J., Reis, M., Browell, J., 2017. LASSO vector autoregression structures for very short-term wind power forecasting. Wind Energy 20 (4), 657–675. https://doi.org/10.1002/we.2029.

Cevasco, D., Koukoura, S., Kolios, A.J., 2021. Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications. Renew. Sustain. Energy Rev. 136 (January 2020), 110414 https://doi.org/10.1016/j.rser.2020.110414.

Chang, G.W., Lu, H.J., Chang, Y.R., Lee, Y.D., 2017. An improved neural network-based approach for short-term wind speed and power forecast. Renew. Energy 105, 301–311. https://doi.org/10.1016/j.renene.2016.12.071.

Chen, P., Berthelsen, K.K., Bak-Jensen, B., Chen, Z., 2009. Markov model of wind power time series using Bayesian inference of transition matrix. IECON Proc. (Industrial Electron. Conf (January 2015), 627–632. https://doi.org/10.1109/IECON.2009.5414993.

Chen, T., Singh, S., Taskar, B., Guestrin, C., 2015. Effcient second-order gradient boosting for conditional random fields. In: Proceedings of the 18th International Conference on Arti Cial Intelligence and Statistics (AISTATS), vol. 38.

Chiachío-Ruano, J., Hermile, M., Kolios, A., 2019. A Sensitivity Study for Operational Availability of Offshore Wind Energy Assets. https://doi.org/10.1115/OMAE2019-96323.

Dawid, R., McMillan, D., Revie, M., 2016. Time series semi-Markov decision process with variable costs for maintenance planning. In: European Safety and Reliability Conference ESREL 2016, p. 183.

Drobinski, P., Coulais, C., Jourdier, B., 2015. Surface wind-speed Statistics modelling: alternatives to the weibull distribution and performance evaluation. Boundary-Layer Meteorol. 157 (1), 97–123. https://doi.org/10.1007/s10546-015-0035-7.

Emmanouil, S., Aguilar, S.G., Nane, G.F., Schouten, J.-J., 2020. Statistical models for improving significant wave height predictions in offshore operations. Ocean Eng. 206, 107249 https://doi.org/10.1016/j.oceaneng.2020.107249.

Erickson, D.J., Taylor, J.A., 1989. Non-Weibull behavior observed in a model-generated global surface wind field frequency distribution. J. Geophys. Res. 94 (C9), 12693 https://doi.org/10.1029/JC094iC09p12693.

Esteoule, T., Bernon, C., Barthod, M., 2019. Improving wind power forecasting through cooperation: a case-study on operating farms. Proc. Int. Jt. Conf. Auton. Agents Multiagent Syst. AAMAS 4, 1940–1942.

European Wind Energy Association, "Wind at Work.".

Fischer, P., Etienne, C., Tian, J., Krauß, T., 2015. Prediction of wind speeds based on digital elevation models using Boosted Regression Trees. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 40 (1W5), 197–202. https://doi.org/10.5194/isprsarchives-XL-1-W5-197-2015.

Fraile, D., Komusanac, I., Walsh, C., 2018. Wind Energy in Europe: Outlook to 2020. no. September, p. 44.

Friedman, J.H., 1999. Greedy Function Approximation: A Gradient Boosting Machine.

Gilbert, C., Browell, J., McMillan, D., 2021. Probabilistic access forecasting for improved offshore operations. Int. J. Forecast. 37 (1), 134–150. https://doi.org/10.1016/j.ijforecast.2020.03.007.

Grinstead, C., Snell, J., 1988. Introduction to Probability.

Guachamin-Acero, W., Li, L., 2018. Methodology for assessment of operational limits including uncertainties in wave spectral energy distribution for safe execution of marine operations. Ocean Eng. 165, 184–193. https://doi.org/10.1016/j.oceaneng.2018.07.032.

Gualtieri, G., Secci, S., 2012. Methods to extrapolate wind resource to the turbine hub height based on power law: a 1-h wind speed vs. Weibull distribution extrapolation comparison. Renew. Energy 43, 183–200. https://doi.org/10.1016/j.renene.2011.12.022.

Gut, A., 2009. An Intermediate Course in Probability. Springer.

Hong, T., Pinson, P., Fan, S., 2014. Global energy forecasting competition 2012. Int. J. Forecast. 30 (2), 357–363. https://doi.org/10.1016/j.ijforecast.2013.07.001.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J., 2016. Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Int. J. Forecast. 32 (3), 896–913. https://doi.org/10.1016/j.ijforecast.2016.02.001.

Horn, J.-T., Leira, B.J., 2019. Fatigue reliability assessment of offshore wind turbines with stochastic availability. Reliab. Eng. Syst. Saf. 191, 106550 https://doi.org/10.1016/j.ress.2019.106550.

International Energy Agency, 1994. Estimation of Cost of Energy from Windenergy Conversion System.

Ioannou, A., Angus, A., Brennan, F., 2018. A lifecycle techno-economic model of offshore wind energy for different entry and exit instances. Appl. Energy 221, 406–424. https://doi.org/10.1016/j.apenergy.2018.03.143.

Ioannou, A., Angus, A., Brennan, F., 2019. Informing parametric risk control policies for operational uncertainties of offshore wind energy assets. Ocean Eng. 177, 1–11. https://doi.org/10.1016/j.oceaneng.2019.02.058.

Jaramillo, O.A., Borja, M.A., 2004. Bimodal versus Weibull wind speed distributions: an analysis of wind energy potential in La Venta, Mexico. Wind Eng. 28 (2), 225–234. https://doi.org/10.1260/0309524041211404.

Karatepe, S., Corscadden, K.W., 2013. Wind speed estimation: incorporating seasonal data using Markov chain models. ISRN Renew. Energy 2013 (January 2014), 1–9. https://doi.org/10.1155/2013/657437.

Kolios, A., Walgern, J., Koukoura, S., Pandit, R., Chiachio-Ruano, J., 2019. Robust O&M open access tool for improving operation and maintenance of offshore wind turbines. In: 29th European Safety and Reliability Conference. ESREL 2019), pp. 452–459. https://doi.org/10.3850/981.

Koukoura, S., Scheu, M.N., Kolios, A., 2021. Influence of extended potential-to-functional failure intervals through condition monitoring systems on offshore wind turbine availability. Reliab. Eng. Syst. Saf. 208, 107404 https://doi.org/10.1016/j.ress.2020.107404.

Leimeister, M., Kolios, A., 2018. A review of reliability-based methods for risk analysis and their application in the offshore wind industry. Renew. Sustain. Energy Rev. 91 (M), 1065–1076. https://doi.org/10.1016/j.rser.2018.04.004.

Loake, M.C.A., Astfalck, L.C., Cripps, E.J., 2022. Modelling sea surface wind measurements on Australia's North-West Shelf. Ocean Eng. 244, 110308 https://doi.org/10.1016/j.oceaneng.2021.110308.

Martinez-Luengo, M., Kolios, A., Wang, L., 2016. Structural health monitoring of offshore wind turbines: a review through the statistical pattern recognition paradigm. Renew. Sustain. Energy Rev. 64 (August), 91–105. https://doi.org/10.1016/j.rser.2016.05.085.

Microsoft Corporation, 2020a. Welcome to LightGBM's Documentation.

Microsoft Corporation, 2020b. Parameters.
Mohandes, M.A., Halawani, T.O., Rehman, S., Hussain, A.A., 2004. Support vector machines for wind speed prediction. Renew. Energy 29 (6), 939–947. https://doi.org/10.1016/j.renene.2003.11.009.
Morthors, P.-E., Awerbuch, S., 2009. The Economics of Wind Energy. https://doi.org/10.1111/j.1745-6622.2009.00231.x. Belgium.
Mytilinou, V., Kolios, A.J., 2019. Techno-economic optimisation of offshore wind farms based on life cycle cost analysis on the UK. Renew. Energy 132, 439–454. https://doi.org/10.1016/j.renene.2018.07.146.
Pandit, R.K., Infield, D., 2018. SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes. IET Renew. Power Gener. 12 (11), 1249–1255. https://doi.org/10.1049/iet-rpg.2018.0156.
Pandit, R.K., Infield, D., Carroll, J., 2019a. Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy. Wind Energy 22 (2), 302–315. https://doi.org/10.1002/we.2285.
Pandit, R., Astolfi, D., Tang, A.M., Infield, D., 2022. Sequential data-driven long-term weather forecasting models' performance comparison for improving offshore operation and maintenance operations. Energies 15 (19), 7233. https://doi.org/10.3390/en15197233.
Pandit, R.K., Infield, D., Kolios, A., 2020a. Gaussian process power curve models incorporating wind turbine operational variables. Energy Rep. 6, 1658–1669. https://doi.org/10.1016/j.egyr.2020.06.018.
Pandit, R.K., Infield, D., Kolios, A., 2019b. Comparison of advanced non-parametric models for wind turbine power curves. IET Renew. Power Gener. 13 (9), 1503–1510. https://doi.org/10.1049/iet-rpg.2018.5728.
Pandit, R., Kolios, A., 2020. SCADA data-based support vector machine wind turbine power curve uncertainty estimation and its comparative studies. Appl. Sci. 10 (23), 8685. https://doi.org/10.3390/app10238685.
Pandit, R.K., Kolios, A., Infield, D., 2020b. Data-driven weather forecasting models performance comparison for improving offshore wind turbine availability and maintenance. IET Renew. Power Gener. 14 (13), 2386–2394. https://doi.org/10.1049/iet-rpg.2019.0941.
Papadopoulos, P., Coit, D.W., Aziz Ezzat, A., 2023. STOCHOS: stochastic opportunistic maintenance scheduling for offshore wind farms. IISE Trans 1–15. https://doi.org/10.1080/24725854.2022.2152913.
Papandreou, C., Ziakopoulos, A., 2022. Predicting VLCC fuel consumption with machine learning using operationally available sensor data. Ocean Eng. 243, 110321 https://doi.org/10.1016/j.oceaneng.2021.110321.
Perea-Moreno, A.J., Alcalá, G., Hernandez-Escobedo, Q., 2019. Seasonal wind energy characterization in the Gulf of Mexico. Energies 13 (1), 1–21. https://doi.org/10.3390/en13010093.
Persson, K., 2010. Exponentiated Gumbel distribution for estimation. J. Environ. Stat. 1 (3), 1–12.
Persson, C., Bacher, P., Shiga, T., Madsen, H., 2017. Multi-site solar power forecasting using gradient boosted regression trees. Sol. Energy 150, 423–436. https://doi.org/10.1016/j.solener.2017.04.066.
Pesch, T., Schröders, S., Allelein, H.J., Hake, J.F., 2015. A new Markov-chain-related statistical approach for modelling synthetic wind power time series. New J. Phys. 17 https://doi.org/10.1088/1367-2630/17/5/055001.
Prevosto, M., Krogstad, H.E., Robin, A., 2000. Probability distributions for maximum wave and crest heights 40 (4).
Sánchez-Pérez, P.A., Robles, M., Jaramillo, O.A., 2016. Real time Markov chains: wind states in anemometric data. J. Renew. Sustain. Energy 8 (2), 1–22. https://doi.org/10.1063/1.4943120.
Satheesh, S.P., Praveen, V.K., Jagadish Kumar, V., Muraleedharan, G., Kurup, P.G., 2005. Weibull and Gamma distributions for wave parameter predictions. J. Indian Geophys. Union 9 (1), 55–64.
Scheu, M.N., Kolios, A., Fischer, T., Brennan, F., 2017. Influence of statistical uncertainty of component reliability estimations on offshore wind farm availability. Reliab. Eng. Syst. Saf. https://doi.org/10.1016/j.ress.2017.05.021.
Scheu, M.N., Tremps, L., Smolka, U., Kolios, A., Brennan, F., 2019. A systematic Failure Mode Effects and Criticality Analysis for offshore wind turbine systems towards integrated condition based maintenance strategies. Ocean Eng. 176 (February), 118–133. https://doi.org/10.1016/j.oceaneng.2019.02.048.
Serfozo, R., 2009. Probability and its Applications.
Shi, J., Guo, J., Zheng, S., 2012. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew. Sustain. Energy Rev. 16 (5), 3471–3480. https://doi.org/10.1016/j.rser.2012.02.044.
Shukur, O.B., Lee, M.H., 2015. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. Renew. Energy 76, 637–647. https://doi.org/10.1016/j.renene.2014.11.084.
Stetco, A., et al., 2019. Machine learning methods for wind turbine condition monitoring: a review. Renew. Energy 133, 620–635. https://doi.org/10.1016/j.renene.2018.10.047.
Tagliaferri, F., Hayes, B.P., Viola, I.M., Djokić, S.Z., 2016. Wind modelling with nested Markov chains. J. Wind Eng. Ind. Aerod. 157, 118–124. https://doi.org/10.1016/j.jweia.2016.08.009.
Taylor, J.W., Jeon, J., 2018. Probabilistic forecasting of wave height for offshore wind turbine maintenance. Eur. J. Oper. Res. 267 (3), 877–890. https://doi.org/10.1016/j.ejor.2017.12.021.
Teng, C., Palao, I., 1996. Wave height and period distributions from long-term wave measurement. Coast. Eng. Proc. 368–379.
Tolver, A., 2016. An introduction to Markov chains. In: An Introduction to Markov Chains, p. 15. Copenhagen.
Tomaselli, P.D., Dixen, M., Bolaños Sanchez, R., Sørensen, J.T., 2021. A decision-making tool for planning O& M activities of offshore wind farms using simulated actual decision drivers. Front. Mar. Sci. 7 https://doi.org/10.3389/fmars.2020.588624.
Torres, J.M., Aguilar, R.M., Zuñiga-Meneses, K.V., 2018. Deep learning to predict the generation of a wind farm. J. Renew. Sustain. Energy 10 (1), 013305. https://doi.org/10.1063/1.4995334.
van Gelder, PHAJM., Vrijling, JK, 2000. On the distribution function of the maximum wave height in front of reflecting structures. Proc. Coast. Struct 1, 37–46.
Verbois, H., Rusydi, A., Thiery, A., 2018. Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting. Sol. Energy 173 (October), 313–327. https://doi.org/10.1016/j.solener.2018.07.071.
Wind Europe, 2019. "Offshore Wind in Europe Key Trends and Statistics 2018,".
Wu, M., Gao, Z., 2021. Methodology for developing a response-based correction factor (alpha-factor) for allowable sea state assessment of marine operations considering weather forecast uncertainty. Mar. Struct. 79, 103050 https://doi.org/10.1016/j.marstruc.2021.103050.
Wu, M., Stefanakos, C., Gao, Z., Haver, S., 2019. Prediction of short-term wind and wave conditions for marine operations using a multi-step-ahead decomposition-ANFIS model and quantification of its uncertainty. Ocean Eng. 188, 106300 https://doi.org/10.1016/j.oceaneng.2019.106300.
Wu, M., Gao, Z., Zhao, Y., 2022. Assessment of allowable sea states for offshore wind turbine blade installation using time-domain numerical models and considering weather forecast uncertainty. Ocean Eng. 260, 111801 https://doi.org/10.1016/j.oceaneng.2022.111801.
Yan, J., Li, K., Bai, E.-W., Deng, J., Foley, A.M., 2016. Hybrid probabilistic wind power forecasting using temporally local Gaussian process. IEEE Trans. Sustain. Energy 7 (1), 87–95. https://doi.org/10.1109/TSTE.2015.2472963.