Ocean Engineering 206 (2020) 107249

Contents lists available at ScienceDirect

Ocean Engineering

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

# Statistical models for improving significant wave height predictions in offshore operations

Stergios Emmanouil [a,b,*], Sandra Gaytan Aguilar [c], Gabriela F. Nane [d], Jan-Joost Schouten [c]

[a] Department of Civil Engineering and Geosciences, TU Delft, Delft, the Netherlands
[b] Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA
[c] Deltares, Delft, the Netherlands
[d] Department of Applied Mathematics, TU Delft, Delft, the Netherlands

ARTICLE INFO

*Keywords:*
Bayesian networks
Offshore operations
Real-time predictions
Statistical techniques
Improved forecasts

ABSTRACT

Installation and maintenance strategies regarding offshore wind farm operations involve extensive logistics. The main focus is the right temporal and spatial placement of personnel and equipment, while taking into account forecasted meteorological and ocean conditions. For these operations to be successful, weather windows characterized by certain permissive wave conditions are of enormous importance, whereas unforeseen events result in high cost and risk of safety. Numerical modelling of waves, water levels and current related variables has been used extensively to forecast ocean conditions. To account for the inherited model uncertainty, several error modelling techniques can be implemented for the numerical model forecasts to be corrected. In this study, various Bayesian Network (BN) models are incorporated, in order to enhance the accuracy of the significant wave height predictions and to be compared with other techniques, in conditions resembling the real-time nature of the application. The implemented BN models differ in terms of training and structure and provide overall the most satisfying performance. Supplementary, it is shown that the BN models illustrate significant advantages as both quantitative and conceptual tools, since they produce estimates for the underlying uncertainty of the phenomena, while providing information about the incorporated variables' dependence relationships through their structure.

## 1. Introduction

Marine structures like offshore wind turbines can ensure safety and serve their main function adequately, in both reliability and economy terms, when most – if not all – of the variables involved in their design are modelled as accurately as possible. The specification of the uncertainties related to the environmental variables describing the ocean conditions is continuously gaining importance and interest by the offshore, coastal, and the emerging renewable energy industries. Several studies have been conducted in order to describe, classify, or quantify the uncertainties and errors related to meteorological and ocean climate variables (see e.g. Haver and Moan, 1983). Simplistically, as proposed by Bitner – Gregersen and Hagen (1990) and Bitner – Gregersen et al. (2014), the uncertainty can be classified as: (a) Phenomenon related uncertainty, which is a product of the natural randomness and stochastic nature of the variables incorporated and cannot be reduced, (b) data related uncertainty, which surfaces either from the measuring devices' accuracy, or from the assumptions adopted while post-processing the data, or due to the insufficient number and quality of the observations, (c) model related uncertainty, which constitutes a product of inaccurate idealisations, crude assumptions, or even insufficient use of either the meteorological or the hydrodynamic model, and (d) statistical uncertainty, which arises from random variations of the data originating from measurements, numerical simulations or laboratory tests. It is obvious that the true nature of any phenomenon cannot be modelled exactly and that even if the probability distributions of some variables are known a priori, the extreme complexity of the met-ocean environment makes the distributions of the rest completely unknown. The estimation of the bias, or systematic error, and the random error evaluation are the first steps to quantify the uncertainty of any variable.

In the case of offshore wind farms, the installation and maintenance strategies involve extensive logistics. The main focus is the right placement, in time and space, of both the personnel and the equipment, while taking into account forecasted meteorological conditions and the wave

* Corresponding author. 261 Glenbrook Road, Unit 3037, Storrs, CT, 06269, USA.
*E-mail address:* stergios.emmanouil@uconn.edu (S. Emmanouil).

https://doi.org/10.1016/j.oceaneng.2020.107249
Received 20 June 2019; Received in revised form 9 March 2020; Accepted 13 March 2020
Available online 16 April 2020

characteristics. In order for the aforementioned procedures to be carried out successfully, weather windows, interwoven with certain permissive wave, wind and current conditions, are of major importance, while unforeseen weather or sea climate events result in high cost and risk, primarily in terms of safety. Subsequently, successful operations require accurate and representative data for the wind farm sites, which unfortunately are inadequately - if at all - provided by surrounding stations.

In order to produce forecasts of the ocean conditions in a specific area, numerical models can be used. Wind speeds, as well as the air and water temperatures, resulting from a meteorological model serve as an input for numerical modelling of waves, water levels and current related variables. In that regard, SWAN (see Booij et al., 1999; Ris et al., 1999) is a third-generation wave model, developed at Delft University of Technology, which computes random, short-crested wind-generated waves in coastal regions and inland waters and provides output quantities in numerical files containing tables, maps and timeseries. Comparison of the wave model forecasts with observations is essential for characterizing the model deficiencies, identifying systematic and random model errors, thus providing areas for improvement.

Several techniques exist and can be implemented in order for the numerical model forecasts to be corrected. The Artificial Neural Networks (ANNs), which are information processing paradigms composed of a large number of highly interconnected processing elements (neurons) working together, have been used extensively in offshore and coastal applications (see e.g. Makarynskyy, 2004; Londhe et al., 2016; Krishna Kumar et al., 2017). Supplementary, Copulas (see e.g. Genest and Favre, 2007; Embrechts et al., 2001; Nelsen, 2006; Schmidt, 2006; Vanem, 2016) have been utilized in various occasions to model the dependency of ocean related variables and predict their behavior, as it has been done in the works of Leontaris et al. (2016) and Jane et al. (2016). Simpler but equally effective methods are the linear regression and the stochastic interpolation. Both of these techniques have been used extensively in a variety of engineering applications, including offshore and coastal (see e.g. Asma et al., 2012; Scotto and Guedes Soares, 2007). They do not require substantial training and pose serious advantages in terms of computational time and load. A variety of methods which can be used to model the joint probabilities of multivariate met-ocean parameters can also be found in the review of Bitner – Gregersen (2015).

All of the aforementioned techniques constitute soft computing methods and ensure a reasonable computational load. A number of them require training using historical or present time data, while others can be incorporated forthwith. Some studies have tried to produce valid met-ocean climate forecasts using coupled (hybrid) method (e.g. Deshmukh et al., 2016), as the ones discussed in this paper, or incorporate solely one of the techniques discussed previously to predict the environmental conditions therewithal. By "coupled" or "hybrid" methods the use of more than one error modelling techniques, or a combination of a soft computing method and a numerical model, is implied. Certainly, the use of a single soft computing method for prediction reduces the computational time significantly, but often at the expense of accuracy.

In this study, special attention is given to the implementation of the Bayesian Networks (BNs), graphical models which allow the representation of a probability distribution over more than one variable and whose use has not been that widespread in offshore applications (an example can be found in Malekmohamadi et al., 2011), but has been tested effectively in other engineering problems, such as coastal morphology (see e.g. Poelhekke et al., 2016; Kroon et al., 2017; Wilson et al., 2015; Plant and Holland, 2011), environmental modelling (see Chen and Pollino, 2012; Aguilera et al., 2011), construction reliability (Morales-Nápoles and Steenbergen, 2014), or flood risk analysis (Sebastian et al., 2017). Supplementary, many applications of the BNs on dependability, risk analysis and maintenance can be found in Weber et al. (2012) and Medina Oliva et al. (2009). An overview of many BN applications is given in the work of Hanea et al. (2015). Many of the applications, however, use networks consisting of nodes that represent discrete random variables. Those networks are characterized as discrete BNs and suffer from serious limitations, since the provided discrete representation of variables for many important problems is inadequate.

The perspective of this research deviates from providing a forecast, accompanied with a desired level of accuracy. The aim is to use automated tools to quantify the possible errors present in numerical model forecasts of the significant wave height ($H_s$), learn from these errors while understanding and quantifying the underlying relations induced by certain phenomena to eventually improve the predictions of the numerical model, which is solely based on empirically and theoretically derived formulas. The consideration of Bayesian Networks aims to the description and representation of the underlying uncertainty in nature's behavior, as accurately as possible. While most models, such as Copulas or ANNs, would just need past measurements, numerical model data and/or numerical model forecasts of the significant wave height, to produce a possible correction, the nature of Bayesian Networks imposes the use of more variables (e.g. wind velocity, wave period, etc.), whose dependency with the variable of interest can produce a forecast of enhanced accuracy.

In Section 2 of this paper some information on the data used for the analysis, as well as a description of the theoretical background and functionality of the BN models, are outlined. To grant the desired corrections, several models that differ in terms of their training, their structure, and the incorporated variables were created and tested. A comparison of the performance of all the implemented statistical and stochastic techniques took place, to ascertain which one performs better, employing widely used evaluation metrics and more specific indicators created for the purposes of the application under consideration. Additionally, the ability of the error correction techniques to perform in operational (real-time) conditions was investigated, to evaluate their performance even in possible absence of measurements. The results and comparison of the different techniques can be found in Section 3, along with a discussion on the influence of different BN structures on the quality of the outcome. Finally, Section 4 contains the conclusions of this study, supplemented by future research paths.

## 2. Materials and methods

The error correction models described here, are essentially forecasting tools, which attempt to predict the wave conditions in open seas more accurately than a numerical model (in this case SWAN), while using the results provided by the latter as an input. Hence, they are referred to as "error correction" models, since their nature and behavior deviates slightly from a pure predictive tool (see e.g. Emmanouil, 2018).

In general, the models are able to perform both in non-operational (offline) and operational (online) situations. By operational situations, the continuous flow of the required data in real-time is implied, while in non-operational mode, the model interacts with data stored in the computer's memory. Nevertheless, in both cases the nature of the data, and the number of variables included in each simulation, are the same. The error correction models require three types of data: (1) on-site measurements (observations), which are processed before used (2) numerical model hindcast[1] data for a time interval prior to the one under consideration. Instead of using hindcast data for the analysis, one could alternatively use past forecast data of the numerical model, which of course will be less accurate, due to the input of wind data produced also by a numerical model (e.g. HIRLAM; see Cats and Wolters, 1996), incorporating and transferring uncertainties of its own, and (3) numerical model forecast data for the time interval under consideration (48 h ahead of current time). In a real-time scenario, the numerical model forecasts is produced every 6 h, so there would be 4 forecasts per

[1] The numerical model hindcast data are produced by incorporation of observational wind data as input to the model and a reverse procedure to obtain the results (i.e. the opposite of a forecast procedure).

day, each one for 48 h ahead. Depending on the error correction method some of the above data may or may not be used.

### 2.1. Bayesian Networks (BN) model

#### 2.1.1. Brief theoretical background

Bayesian Networks (BNs) are graphical models, which allow the representation of a probability distribution over a set of random variables (see Jensen and Nielsen, 2007; Morales-Napoles et al., 2013; Hanea et al., 2015; Weber et al., 2012). They consist of a directed acyclic graph (DAG) built on discrete (discrete networks), continuous (continuous networks), or both kinds (hybrid networks) of random variables ($X_1$, $X_2$, …, $X_n$), and a set of (conditional) distributions. A DAG is constituted by a set of nodes, that represent random variables, and a set of arcs, in a way that a directed cycle cannot be created. Within the graph, an ordering of the variables can be established, given the directionality, which provides information on the sampling order, i.e. the order which has to be followed so that a sample can be taken from this joint distribution. As a result, some of the nodes are characterized as "parents" and others as "children", depending on whether they precede or success the node of interest. A marginal distribution is assigned to each node with no parent, and a conditional distribution is associated with each child node, providing quantitative information about the dependences between the variables, which can be either retrieved from data or from expert judgment (see e.g. Cooke, 1991).

Denoting the parent nodes of $i$ as $Pa(i)$, the joint density of $X_1$, $X_2$, …, $X_n$ is given by:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=\prod_{i=1}^{n} f_{X_i|X_{Pa(i)}}(x_i|x_{Pa(i)}) \tag{2.1}$$

where $f_{X_i|X_j}$ denotes the conditional densities. The factorization of the joint distribution relies on the local Markov property of conditional independence.

BNs are quantitative tools, able to evaluate conditional probabilities between variables, and at the same time constitute valuable conceptual models, since they visually represent independent and dependent variables in causation relationships (see Chen and Pollino, 2012; Palmsten et al., 2014; Stewart-Koster et al., 2010). The principles of BNs as a modelling tool are described thoroughly in Pearl (1988) and Jensen (1996). The main property of the BNs is inference, which constitutes their ability to provide updated distributions, given observations, but also characterization of the relationship between the variables. Generally, the simple visualization of the complicated relationships between the random variables, as well as their polyvalence, i.e. the ability to deal with issues such as prediction, diagnosis, optimization, data analysis of feedback experience, and model updating, makes the use of BNs appealing.

#### 2.1.2. Training methodology

The Bayesian Networks, as most of the data driven techniques, need a sufficient amount of data in order to be trained sufficiently and be able to represent the desired relations. When the BN structure is acquired through the data, then a significant amount of data is needed. In every application the characterization of a training procedure as "sufficient" depends largely on the type and behavior of the data. A sensitivity analysis would be in place to determine what "sufficient amount" actually means for the application. The significant wave height, for instance, is a variable whose behavior is highly dynamic, i.e. it can change radically in short time intervals (e.g. hours). As a result, the more training the model has the better, since it can assimilate to, and later reflect a larger range of behaviors.

Here, the training techniques are divided into two major categories; (1) the long training, which involves past observational and numerical data, even from 3 years prior to the current date, and (2) the short training, which only involves measurements and numerical model data from 48 h prior to the start of the forecast.

In order to obtain the structure of the Bayesian Network, the bnlearn package[2] (see Scutari and Denis, 2014) of the R programming language is used. In general, there are two broad categories of algorithms to learn the structure of a BN, the score-based and the constraint-based. The constraint-based case employs conditional independence tests to identify a set of edge constraints for the graph and then finds the best DAG that satisfies these constraints; see e.g. Scutari (2015). The score-based approach (see Russell and Norvig, 2009; Korb and Nicholson, 2010) first defines a criterion to evaluate how well the BN fits the data, and then searches over the space of DAGs for a structure with maximal score.

For this study, a hill climbing (HC) score-based structure learning algorithm was used to train the network, which made use of the Akaike Information Criterion (AIC). The package also assumes a multivariate normal distribution for continuous variables (such as the wave characteristics in hand). This assumption can be considered restricting in many occasions, but as it will become obvious, the results of such an analysis are quite reasonable. In case the assumption of multivariate normality is violated, the non-parametric Bayesian Networks could produce a more accurate conditional distribution and possibly more accurate forecasting results; see e.g Hanea et al. (2015). Nevertheless, the assumption of multivariate normality was considered sufficient to test the BN behavior and performance, and the open-source bnlearn package as the most suitable one for this particular application.

For the case of long training, the training dataset is continuously enriched with new measurements, as well as with past numerical model data for the variable of interest only. Certainly, this requires a relatively large part of the computer's memory. This effect can be impugned by incorporation of new variables and deletion of older, or with smaller training sets, i.e. in the order of months instead of years.

In general, the user can impute his/her own structure, by whitelisting or blacklisting certain relations, i.e. providing a custom fit. This, certainly, creates large differences in the results, since in many occasions the whitelisted arc is not supported by the BN structure in representing the joint density. Thus, it is suggested by the writers that the procedure should be carried out using data-driven structure learning and fitting techniques, even if a given relation might not be supported intuitively.

#### 2.1.3. Predictions and uncertainty bounds

The predictions provided by the BN models are retrieved from the conditional distribution of the variable of interest, given the information about certain other variables. Since it is impossible to have future measurements for the incorporated variables, forecasted numerical model data for these variables are used to construct the conditional distribution for every point prediction. In other words, the network is trained and fitted with past observational data, as well as numerical model data for the variable of interest, subsequently providing a forecast based on forecasted numerical model data (essentially we are conditionalizing on forecast numerical model data). The point prediction is the expected value of the conditional distribution, which is assumed to be normal. Since the significant wave height ($H_s$) is not normally distributed (see e.g. Jasper, 1956; Kim, 2008; Shariff and Hadi Hafezi, 2012; Li et al., 2016), the assumption is in certain occasions not appropriate. Consequently, this assumption prevents us from retrieving realistic uncertainty bounds for the significant wave height. Nevertheless, the symmetrical uncertainty intervals can provide a fairly good coverage of the observations (more information and examples can be found in the following sections).

A fit test was carried out for the significant wave height ($H_s$) data by means of the "Find the Best Distribution" (FDB) tool in Matlab® language, which incorporates certain criteria to define the best parametric distribution for the data in hand (Aminov, 2020). Namely, the Akaike Information Criteria (AIC; see e.g. Akaike, 1974) and the Bayesian

[2] For more information the reader is referred to http://www.bnlearn.com/.

Information Criteria (BIC; see e.g. Schwarz, 1978) are taken into consideration. For both of the aforementioned criteria, the lower their values are, the better a distribution fits the data. As can be seen in Fig. 1, the log-normal distribution provides a good fit for the significant wave height data ($H_s$), displaying the overall lowest AIC and BIC scores when compared to the other tested distributions (namely the Inverse Gaussian and the Generalized Extreme Value distributions). This result is consistent with the information provided by the literature (see e.g. Jasper, 1956; Kim, 2008) regarding the distribution of $H_s$. This outcome will be proved useful in the simulations to follow.

The standard 95% are obtained from the 2.5th and 97.5th quantiles of the conditional distribution. Since the wave heights are modelled, in this case, based on the log-normal distribution, a log-transformation of the significant wave height ($H_s$) has been applied. The network was thus trained with the transformed data. The obtained predictions were converted back to their original form, which yield the log-normal intervals. Again the 2.5th and the 97.5th quantiles were used.

### 2.2. The data

The data were retrieved from stations deployed in the Irish Sea. The measurement stations, which are actually wave rider buoys and meteorological masts, are adjacent to the wind farms of Gwynt-y-Mor (53°27′N 03°35′W) and Rhyl Flats (53°22′N 03°39′W), located within the Liverpool Bay. The received datasets consist of measurements of meteorological and wave-related data, obtained between 01 and 09–2012 to 31-01-2018. It has to be stressed that the error correction techniques are suitable for any offshore environment, given the required training, and are not limited in the area of the Irish Sea. The case presented here serves as an example of the applicability of the models in real-life applications. The same procedures and techniques would have to be followed in any similar case, aiming to accurately predict the variables' behavior in mild offshore environments.

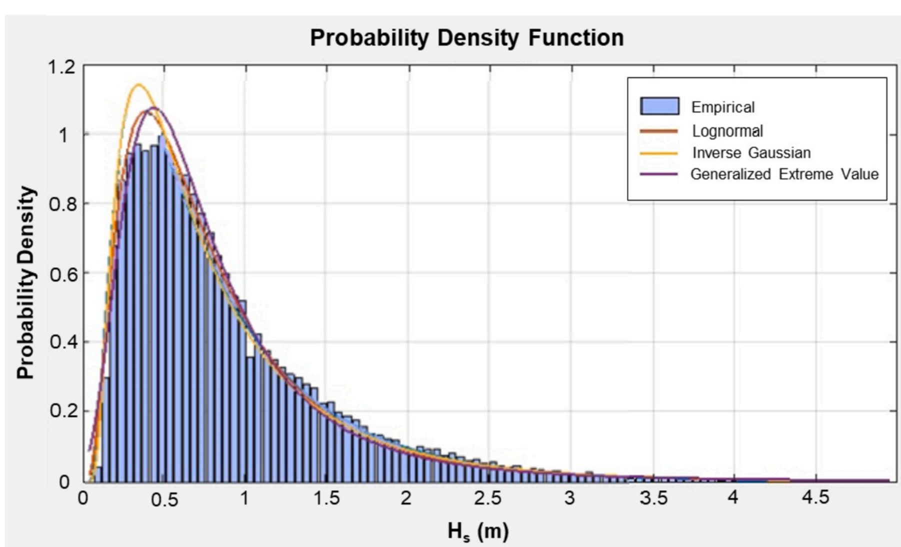
#### 2.2.1. Training and fitting datasets

Different error correction techniques require different sets for training, while some of them do not need substantial training at all. To be more exact, the simple linear regression and the Bernstein stochastic interpolation (see e.g. Kolibal and Howard, 2006, 2008; Seyfarth et al., 2006) utilized here, take as an input only numerical data and measurements corresponding to a time interval just 48 h prior to the forecast. The three-layered, feed forward ANN (see e.g. Deo and Sridhar Naidu, 1999; Mandal et al., 2005), which uses a back-propagation algorithm (see e.g. Tsai and Lee, 1999), as well as the bivariate Copula (chosen to be Gumbel based on a simple Cramér-Von Mises criterion test incorporating numerically modelled and observed data; see Anderson, 1962), were trained with 6 months of data corresponding to the period March–August 2015, and then used implementing the same input delineated for the aforementioned techniques. It has to be stressed that only $H_s$ data were used by all these techniques.

The BN models incorporate three different types of training; (1) long-training with data from 01 to 01–2014 to 31-12-2016, i.e. 3 years of training, (2) short-training with hourly data corresponding to 48 h prior to the forecast, i.e. 2 days of training, and (3) a fixed structure, produced by 3 years of training (2014–2016), and fitted with data tallying to 48 h prior to the respective 48-hr forecast, i.e. 3 years for training and 48 h for fitting and retrieving the required variable relations, necessary to produce a prediction. The term "fixed" was used to stress out that, while the power of the underlying relations between the variables constantly altered due to the dynamic behaviour of wave characteristics (i.e. the significant wave height, the zero-crossing wave period and the wave direction) and meteorological variables (i.e. the wind velocity and direction), the structure was not changing because of the significant amount of training.

#### 2.2.2. BN input data

When producing a prediction with the BN model, there should be an input of the variables based on which the conditional distribution is being produced (this is often referred to as *conditionalization*). The variables were selected to represent nodes in the network based on their relation to the significant wave height and their availability. In order to simulate a realistic scenario, where measurements and numerical model data exist, the following variables were selected: (1) the zero-crossing wave period ($T_z$), (2) the wave direction ($D_{irp}$), (3) the wind velocity 10 m above the sea level ($U_{10}$), (4) the wind direction ($U_{dir}$), and (5) the numerical significant wave height ($H_{s,num}$). As stated before, the




| Distribution | *Lognormal* | *Inverse Gaussian* | *Generalized Extreme Value* |
|---|---|---|---|
| **BIC** | 61674.71 | 61790.12 | 62942.80 |
| **AIC** | 61657.34 | 61772.75 | 62916.74 |

**Fig. 1.** Results of the parametric distribution fitting procedure to the significant wave height ($H_s$) data of Gwynt-y-Mor.

numerical model forecast data (48 h ahead) for the rest of the selected variables are used as conditionalizing values to generate accurate predictions for the variable of interest, namely the significant wave height ($H_s$).

#### 2.2.3. Model testing and validation datasets

For testing and comparison between the different incorporated techniques, data retrieved for the year of 2017 were used (01-01-2017 to 31-12-2017). In order to simulate effectively the real-time nature of the application, a forecast was corrected every 6 h of each day. Because SWAN produced 4 forecasts per day, one every 6 h, each one of the error correction techniques, generated a potential corrected (potentially more accurate) prediction an equal number of times. It can be realized that the extremely large amount of information makes it impossible for all the results to be presented. Thus, a collective set, encapsulating different types of behaviours, is going to be displayed.

## 3. Results - discussion

In this section, the summative results for simulations corresponding to the whole year of 2017 (from 01 to 01–2017 to 31-12-2017) are presented. As previously stated, the measurement stations were situated near the Gwynt-y-Mor (GyM) and Rhyl Flats (RF) offshore wind farms.

### 3.1. Method comparison

In order to establish a basis for comparison between the methods, certain well-known evaluation metrics, namely the Root-Mean-Square-Error (RMSE), the Bias, and the Unbiased-RMSE (URMSE) were employed. For reasons of brevity, only the Gwynt-y-Mor results are presented here (Table 1).

Both the long-trained and custom-fixed BNs displayed satisfying performance in terms of their error distribution, which is reflected on their bias values, while introducing an enhancement in accuracy, larger than any other method, with the exception of linear regression. Yet, even if the metrics of Table 1 are indicative of the general behavior of the models, it has to be stressed out that evaluating the techniques' performance solely based on them is impossible. This issue, regarding the robust and consistent validation of the predictions, can be resolved with the use of case specific metrics, i.e. indicators displaying the models' accuracy within and around the significant wave height boundaries of this specific application, i.e. $0.5 \le H_s \le 1.5$ m. Particular interest is given around the upper boundary of 1.5 m (see Table 2), which is commonly used in practice for offshore maintenance operations (see e.g. Asgarpour, 2016; Röckmann et al., 2017; Seyr and Muskulus, 2019; Stumpf and Hu, 2018; Hu et al., 2019) as an upper limit for equipment and personnel transport.

Consequently, three extra indicators were taken into account: (1) the percentage of the critically accurate predictions, i.e. the forecasts for which the measurements were higher than 1.5 m and the respective model managed to predict, (2) the false positive forecast percentage, which provides information on the amount of predictions above 1.5 m when the measurement was below, and (3) the percentage of the critically inaccurate forecasts, i.e. the amount of predictions below the 1.5 m upper boundary, when the measurement was above that limit. Notice that the percentages were calculated over the whole time interval, i.e. in terms of the whole dataset, hence their values are small. In any case, they provide the needed means for comparison in this stage.

An example of a correction to the numerical model's 48-hr forecast, given at critical values for an operation, is shown in Fig. 2. The BN model incorporating the so-called fixed structure managed to predict relatively accurate the offshore conditions while simultaneously prevented (hypothetically) any operation that might endanger the crews and the equipment.

### 3.2. Uncertainty estimates

One major advantage of the BN methods, in comparison to the rest of the techniques is their ability to provide estimates of the uncertainty (see also Section 1) governing the variable of interest; in this case the significant wave height $H_s$. The only one of the other techniques able to produce confidence intervals is the Gumbel Copula. Nevertheless, the assumption of a Gumbel Copula influences the confidence intervals' performance significantly.

Regarding the BN methods, the normality assumption for the conditional distribution of $H_s$ governs the predictions. As a result of the aforementioned supposition, the uncertainty boundaries given by the BN models are symmetrical. Despite the restrictive nature of this assumption, the predictions acquired by the BN models in our study are quite satisfying, providing a correction of the SWAN forecast in most of the cases. That of course might not influence their performance or their usefulness.

Since the $H_s$ data follow a log-normal distribution (see also Section 2.1.3), a log-transformation of the data has been considered for the BN methods. Note that the uncertainty bounds are no longer symmetric. Table 3 provides the results of uncertainty quantification from standard BN methods and BN methods applied to the log-transformation of the data, as well as from the Copula. The log-normal uncertainty bounds provide smaller coverage percentages (percentage of measurements in the test data within the confidence interval) with similar or larger average lengths of the confidence intervals or larger percentages accompanied with unrealistically large average lengths (approximately 1.18 m). As a result, the normal confidence intervals are more efficient and accurate. The most useful uncertainty boundaries seem to be the ones provided by the BN model incorporating the fixed structure, which have a reasonably high coverage percentage (86.1%) accompanied by a satisfying average length, in comparison to the bounds given by the long-trained BN model, which are 10 cm larger but only 3% more accurate.

Considering the overall performance in terms of the given uncertainty, in combination with the point predictions provided previously, it seems that the BN method incorporating a fixed structure, alongside with the respective normal confidence intervals, is the most suitable one for the Gwynt-y-Mor case study. The long-trained BN normal boundaries have also a steady and robust performance, which makes the corresponding model an attractive and satisfying alternative.

Finally, is has to be noted that the extremely large coverage percentage given by the log-normal uncertainty boundaries, for the case of the long-trained BN model, is justified by the similarly large average length of the intervals, which makes the solution less suitable. The log-normal boundaries have a more realistic form (i.e. only positive values and a match with the parametric distribution fitting the $H_s$), but in case the performance is taken into account the normal confidence intervals pose many advantages.

**Table 1**
Evaluation metrics for the year of 2017 (Gwynt-y-Mor).

| Method | SWAN | BN Long Training | BN Short Training | BN Fixed Structure | REG | ANN | Copula | SI |
|---|---|---|---|---|---|---|---|---|
| RMSE (m) | 0.231 | 0.218 | 0.253 | 0.209 | 0.206 | 0.225 | 0.246 | 0.325 |
| BIAS (m) | −0.046 | −0.011 | −0.051 | 0.005 | 0.004 | 0.0365 | −0.076 | −0.016 |
| URMSE (m) | 0.226 | 0.218 | 0.248 | 0.209 | 0.206 | 0.222 | 0.234 | 0.324 |

**Table 2**
Application specific metrics for the year of 2017 (Gwynt-y-Mor).

| Method | SWAN | BN Long Training | BN Short Training | BN Fixed Structure | REG | ANN | Copula | SI |
|---|---|---|---|---|---|---|---|---|
| Critically Accurate (%) | 19.72 | 21.16 | 20.27 | 22.31 | 22.00 | 23.05 | 16.89 | 20.83 |
| Critically Inaccurate (%) | 2.55 | 2.82 | 3.79 | 2.10 | 2.34 | 1.90 | 4.72 | 1.96 |
| False Positive (%) | 2.26 | 1.93 | 1.50 | 2.10 | 1.97 | 3.01 | 0.82 | 3.01 |

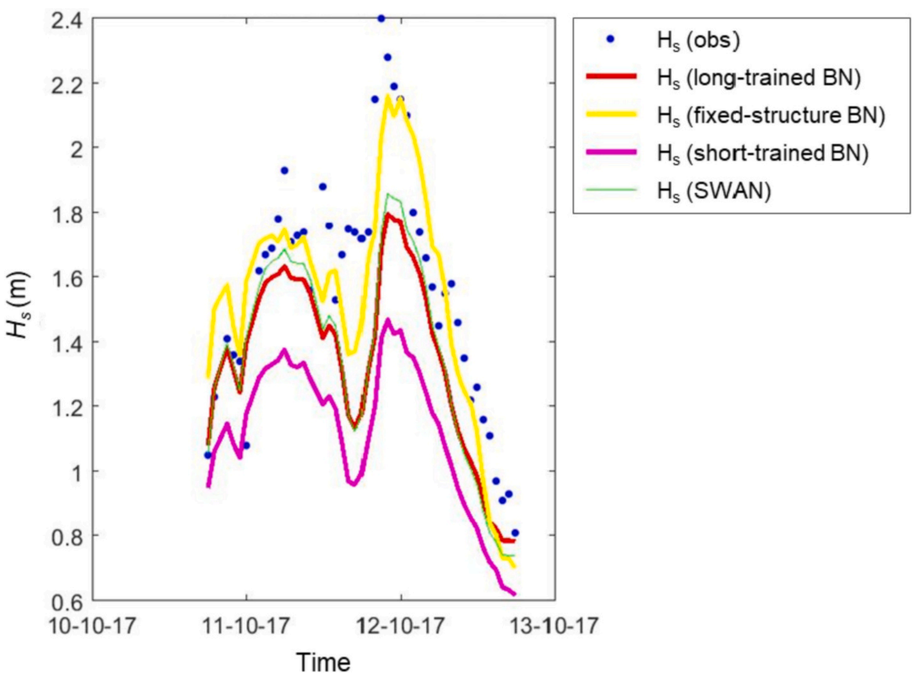



**Fig. 2.** Example of a correction to the SWAN forecast under critical conditions given by the BN models (Gwynt-y-Mor). The blue dots denote the $H_s$ in-situ observations, the red line the long-trained BN's prediction, the yellow line the forecast of the BN model incorporating the fixed structure, the magenta line the short-trained BN's prediction, while the green line indicates the forecast provided by SWAN.

**Table 3**
Uncertainty comparison for the Gwynt-y-Mor case study.

| Method | BN Long Training | BN Fixed Structure | BN Short Training | Copula | BN Long Training (Log-N) | BN Short Training (Log-N) | BN Fixed Structure (Log-N) |
|---|---|---|---|---|---|---|---|
| Coverage (%) | 89.2 | 86.1 | 75.3 | 68.5 | 95.4 | 73.1 | 76.5 |
| Average Length (m) | 0.630 | 0.531 | 0.356 | 0.375 | 1.185 | 0.550 | 0.594 |

### 3.3. BN structures and configurations

Up until now, the incorporated BN structures involved 6 nodes. Fig. 3 displays the long-trained structure, which has also been used for the fixed BN model. The simulations were carried out using data driven structures, i.e. structures acquired by the nature of the data and not imposed a priori. In general, it was noted that trying to create a structure using general knowledge on the incorporated variables (i.e. knowledge on the underlying relations procured by the literature or by experts) only hindered the prediction/correction procedure instead of enhancing its accuracy (see also Emmanouil, 2018).

Some of the relations governing the structures are anticipated, when others oppose what would be expected by the common knowledge on the variables at hand. The most distinctive examples here are the relations between the observed significant wave height ($H_s$) and the wind velocity ($U_{10}$), as well as the wind ($U_{dir}$) and wave ($D_{irp}$) directions. In a situation represented by the dependencies described in the literature (see e.g. Hasselmann and Olbers, 1973), one would expect the wind direction to influence the wave direction, i.e. the arc connecting those two nodes to have a direction from $U_{dir}$ to $D_{irp}$. Nevertheless, the data-driven analysis conducted in this study implies that the wind direction depends on the wave direction, something which is certainly not the case. But a reasonable explanation exists, justifying this kind of behaviour. The wind and wave directions are measured at the same locations, a fact that insinuates that the variables influence one another in one specific area. Still, waves are created by storms occurred many kilometres (or miles in the nautical language) away from the location of the measurement. As a result, the measured wind directions might indeed not have any influence on the wave directions. Further, the wave direction is influenced by many effects, such as diffraction due to islands or other obstacles, so it can be totally irrelevant to the values given by the wind direction. That of course raises the question on whether the wind direction could be omitted by the analysis, which will be addressed hereupon.

On the other hand, the significant wave height and wave direction relation is a different story. For the case of the long training (3 years of

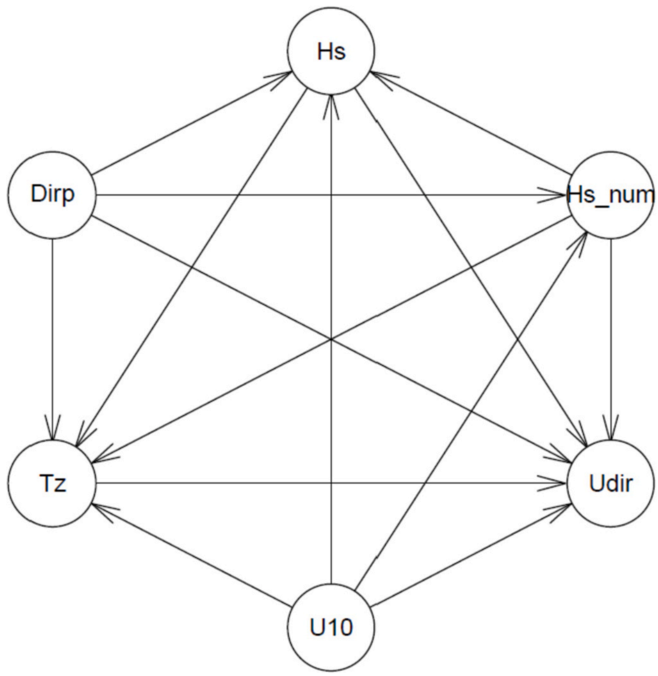



**Fig. 3.** Structure for the long-trained and fixed BN models, incorporating 6 variables (Gwynt-y-Mor).

data), presented in Fig. 3, the relation is the one expected by the descriptions available in the literature, corresponding to the experts' opinions; see e.g. Pierson and Moskowitz (1964), Hasselmann and Olbers (1973), as well as Phillips (2006). To be more exact, the wind velocity influences the significant wave height, a dependence which is highlighted by the high correlation between the variables (correlation coefficient equal to 0.795), shown in Table 4. In the same table other relations are also visible, as for instance between the wind and wave direction, which justifies the structure's form. Also visible is the extremely high dependency between the observed and numerically derived wave heights, which gives the character of correction instead of pure prediction, since the quality of the numerical model (SWAN) results influence highly the long-trained models' accuracy.

Contrarily, the short-trained BN model provides a variety of relations between the wind velocity and the observed significant wave height, due to the dynamic nature of the offshore events, which force the data to rapidly change behaviour. There is no clear relation between the two aforementioned variables, since the direction of the connection changes repeatedly, and in some occasions becomes even inexistent. That of course is again explained by the wave creation by distant storms, or secondary effects like diffraction or reflection, since also those two variables are measured in the same location.

It is interesting to examine how different configurations of the BN structures (see Fig. 4), i.e. a different number of nodes with a selection of variables, influence the predictions and the provided uncertainty. This comparison will shed some light on whether one or more of the incorporated variables influence the models' accuracy positively and will reveal if the erratic behaviour of the models incorporating short-term past data can be casted off.

Testing was conducted with a 5-variable BN structure, incorporating only the wind velocity ($U_{10}$) as a meteorological variable. Examples of the arc directions for the case of Gwynt-y-Mor are shown in Fig. 4, where the relations between the meteorological variables and the wave characteristics are again varying depending on the training of the BN model (long or short training). The explanation here is similar to the case of the 6-variable structure, since for the largest part of the year the wind velocity can in general influence the significant wave height, while in certain occasions this might not happen due to the origin of the waves. The performance of the models is only enhanced slightly (approximately 0.5%), while being more consistent for the BNs incorporating short-term past data. Even so, the RMSE values were in general smaller for all BN models, with the one provided by the fixed structure being the smallest in comparison to the rest of the error correction techniques (0.208). The accuracy in predictions close to the critical boundary also increased, particularly in terms of the false positive percentages (nearly 8%; a value of 1.95% for the case of the fixed structure).

Regarding the uncertainty estimates, the coverage percentages and the average lengths were similar to the 6-variable BN models' figures, without any improvement to the length of the long-trained log-normal confidence intervals. It is truly difficult to determine which boundary is the most suitable and it always depends on the applications needs. Nevertheless, for this case both kinds of confidence intervals display superiority when compared to the uncertainty estimates given by the Gumbel Copula (see Table 7). Of particular interest are the results produced for the case of Rhyl Flats. As shown in Table 5, there is a significant improvement in terms of all metrics. Table 6 illustrates that also in terms of critical performance, around the 1.5 m upper boundary, the fixed structure BN model's performance is enhanced. Moreover, the behaviour of the 5-variable structures regarding models which include short-term past data (i.e. 48 h prior to the forecast), is quite consistent and robust in comparison to the structures incorporating 6 variables. Here, the point that the wind direction causes unsteadiness to the predictions is proved.

Because the uncertainty estimates display large improvement as well, it seemed fit to present them here in comparison to the results given by the 6-variable BN structure (see Table 7). The normal confidence intervals of the fixed-structured BN reach a coverage percentage of nearly 91% of the total observations, with an average length of just 49 cm. Certainly, the form of the boundaries is not ideal, since they are

**Table 4**
Correlation matrix for the long-trained BN models for the Gwynt-y-Mor case.

| Variable | $D_{irp}$ | $T_z$ | $U_{10}$ | $U_{dir}$ | $H_{s,num}$ | $H_s$ |
|---|---|---|---|---|---|---|
| $D_{irp}$ | 1.000 | 0.381 | 0.001 | 0.515 | 0.245 | 0.249 |
| $T_z$ | 0.381 | 1.000 | 0.596 | 0.359 | 0.842 | 0.874 |
| $U_{10}$ | 0.001 | 0.596 | 1.000 | 0.110 | 0.820 | 0.795 |
| $U_{dir}$ | 0.515 | 0.359 | 0.110 | 1.000 | 0.319 | 0.329 |
| $H_{s,num}$ | 0.245 | 0.842 | 0.820 | 0.319 | 1.000 | 0.964 |
| $H_s$ | 0.249 | 0.874 | 0.795 | 0.329 | 0.964 | 1.000 |

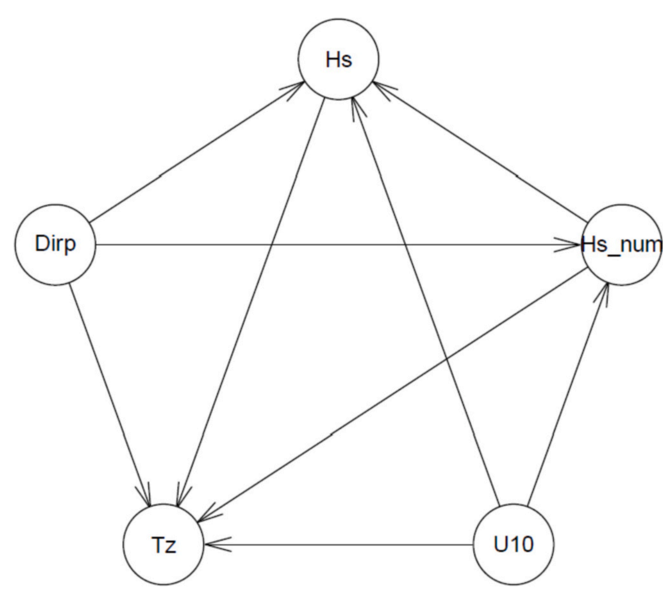



**Fig. 4.** BN structures incorporating 5 variables for the long-trained models (Gwynt-t-Mor).

**Table 5**
Evaluation metrics for the case of the 5-variable BN models (Rhyl Flats).

| Method | SWAN | BN Long Training | BN Short Training | BN Fixed Structure | BN Long Training | BN Short Training | BN Fixed Structure |
|---|---|---|---|---|---|---|---|
| | | 6 Nodes | | | 5 Nodes | | |
| RMSE (m) | 0.203 | 0.178 | 0.200 | 0.201 | 0.178 | 0.195 | 0.163 |
| BIAS (m) | −0.004 | −0.010 | −0.037 | 0.003 | −0.013 | −0.038 | 0.003 |
| URMSE (m) | 0.203 | 0.178 | 0.196 | 0.201 | 0.177 | 0.191 | 0.163 |

**Table 6**
Application-specific evaluation metrics for the case of the 5-variable BN models (Rhyl Flats).

| Method | SWAN | BN Long Training | BN Short Training | BN Fixed Structure | BN Long Training | BN Short Training | BN Fixed Structure |
|---|---|---|---|---|---|---|---|
| | | 6 Nodes | | | 5 Nodes | | |
| Critically Accurate (%) | 18.02 | 17.01 | 16.04 | 18.82 | 16.87 | 16.08 | 18.03 |
| Critically Inaccurate (%) | 1.05 | 2.28 | 2.50 | 1.34 | 2.30 | 2.45 | 1.47 |
| False Positive (%) | 2.55 | 1.16 | 1.03 | 1.58 | 1.14 | 0.84 | 1.14 |

symmetrical, but still their performance provides a significant enhancement in accuracy, making the BN models a valuable correction tool for this application. The long-trained BN model is equally good in terms of accuracy, regardless the number of incorporated variables, making it also a robust and reliable tool, which with the inclusion of its uncertainty bounds introduces a significant improvement of the significant wave height ($H_s$) predictions. As such, it can be concluded that the 5-variable BN models would need to be used for the case of Rhyl Flats, due to its robust behaviour, in comparison to similar techniques incorporating 6 variables.

## 4. Conclusions

The Bayesian Networks described in this study provide a useful tool for the decision making process of installation and maintenance operations in offshore wind farms. The applicability of the models in real-time scenarios could assure the right temporal and spatial placement of the personnel and equipment in dynamic circumstances, hence leading to an optimal utilization of the available resources. Since the success of offshore operations is based on the accurate prediction of specific weather windows, the improved $H_s$ forecasts provided by the BN models will lessen the risk of high cost, while ensuring the safety of the crews.

The basic goal was to manage to emulate the real-time nature of the application and draw conclusions for the applicability of the methods under consideration in operational environments. In that regard, the 5-variable fixed-structured BN model (incorporating $H_s$, $H_{s,num}$, $D_{irp}$, $T_z$, and $U_{10}$) outperforms any other technique. Certainly, this kind of model has one major disadvantage; the fact that it needs short-term past data (48-hrs prior to the forecast) makes it unable to produce corrected forecasts in the absence of recent observations. This is not an issue when it comes to the long-trained BN model, which is able to produce forecasts of enhanced accuracy constantly, even in the absence of recent observations, thus constituting an attractive alternative for real-time use.

The topology can induce secondary effects in terms of hydrodynamics (i.e. reflection, diffraction, etc.), with direct variable relations which are not obvious (e.g. between the wind $U_{dir}$ and wave direction $D_{irp}$). Graphical data-driven approaches, such as the Bayesian Networks developed here, proved to be able to reveal non-trivial dependencies when the morphology of the area, or the way the measurements were collected (e.g. with wave-rider buoys and met-masts), induce many uncertainties. After all, a major benefit of the BN models is the information acquired by the uncertainty estimates they supply, which can be either provided in normal or log-normal form and cover nearly 90% of the total number of measurements in the validation set. The normal confidence intervals seem to be the most suitable for this application, since they demonstrate good performance, especially in terms of the crucial 1.5 m boundary, introducing an acceptable average length of 50–60 cm.

It has to be stressed that the relatively small period of testing (i.e. 1 year) is a result of the limited data availability, due to the fact that the wind farms under consideration are new. For the future, extensive real-time testing for the sites under question, as well as for additional wind farms in completely different locations, would provide a more concise and consistent validation of the models' performance. Supplementary, some variables (e.g. wind direction) could be discretized rather than used as an additional continuous variable, leading to a hybrid network. As such the models' accuracy could be evaluated based on the type of events (e.g. for wind coming from NW in comparison to SE). Finally, it is evident that the discrepancies between the model results are in certain occasions small. An application-based impact assessment would highlight the contribution of each model (potentially in monetary terms) and could manifest the actual footprint of these small differences, when it comes to offshore maintenance operations.

**Table 7**
Uncertainty estimates' performance for the case of a 5-variable BN structure (Rhyl Flats).

| Method | BN Long Training | BN Fixed Structure | BN Short Training | Copula | BN Long Training (Log-N) | BN Short Training (Log-N) | BN Fixed Structure (Log-N) |
|---|---|---|---|---|---|---|---|
| **5 Variables** | | | | | | | |
| Coverage (%) | 89.6 | 90.8 | 77.2 | 70.9 | 95.0 | 77.1 | 73.2 |
| Average Length (m) | 0.527 | 0.489 | 0.430 | 0.327 | 1.024 | 0.505 | 0.460 |
| **6 Variables** | | | | | | | |
| Coverage (%) | 89.7 | 64.7 | 69.8 | 70.9 | 94.7 | 68.9 | 61.0 |
| Average Length (m) | 0.527 | 0.491 | 0.427 | 0.327 | 0.948 | 0.466 | 0.425 |

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.


## 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.


## Acknowledgements

This research was supported by the Joint Industry Projects (JIPs) taking place within the TKI Wind at Sea program. Data used in the project was kindly provided within the JIPs framework. The authors would also like to thank two anonymous reviewers for their constructive comments.


## References


Aguilera, P.A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian networks in environmental modelling. Environ. Model. Software 26, 1376–1388. https://doi.org/10.1016/j.envsoft.2011.06.004.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19 (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Aminov, Y., 2020. FBD - "find the best distribution" tool, MATLAB central file exchange. Retrieved: February 13, URL. https://www.mathworks.com/matlabcentral/fileexchange/36000-fbd-find-the-best-distribution-tool.

Anderson, T.W., 1962. On the distribution of the two-sample cramer-von Mises criterion. Ann. Math. Stat. 33 (3), 1148–1159. https://doi.org/10.1214/aoms/1177704477.

Asgarpour, M., 2016. Assembly, transportation, installation and commissioning of offshore wind farms. In: Ng, Chong, Ran, Li (Eds.), Offshore Wind Farms: Technologies, Design and Operations. Elsevier, Woodhead Publishing Series in Energy, ISBN 9780081007792, pp. 527–541. https://doi.org/10.1016/B978-0-08-100779-2.00017-9. No. 92.

Asma, S., Sezer, A., Ozdemir, O., 2012. MLR and ANN models of significant wave height on the west coast of India. Comput. Geosci. 49, 231–237. https://doi.org/10.1016/j.cageo.2012.05.032.

Bitner-Gregersen, E.M., 2015. Joint met-ocean description for design and operations of marine structures. Appl. Ocean Res. 51, 279–292. https://doi.org/10.1016/j.apor.2015.01.007.

Bitner-Gregersen, E.M., Hagen, O., 1990. Uncertainties in data for the offshore environment. Struct. Saf. 7, 11–34. https://doi.org/10.1016/0167-4730(90)90010-M.

Bitner-Gregersen, E.M., Bhattacharya, S.K., Chatjigeorgiou, I.K., Eames, I., Ellermann, K., Ewans, K., Hermanski, G., Johnson, M.C., Ma, N., Maisondieu, C., Nilva, A., Rychlik, I., Waseda, T., 2014. Recent developments of ocean environmental description with focus on uncertainties. Ocean. Eng. 86, 26–46. https://doi.org/10.1016/j.oceaneng.2014.03.002.

Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions, Part I, Model description and validation. J. Geophys. Res. 104 (C4), 7649–7666. https://doi.org/10.1029/98JC02622.

Cats, G., Wolters, L., 1996. The Hirlam project. IEEE Comput. Sci. Eng. 3 (4) https://doi.org/10.1109/99.556505.

Chen, S.H., Pollino, C.A., 2012. Good practice in Bayesian network modelling. Environ. Model. Software 37, 134–145. https://doi.org/10.1016/j.envsoft.2012.03.012.

Cooke, R., 1991. Experts in Uncertainty: Opinion and Subjective Probability in Science, Environmental Ethics and Science Policy Series. Oxford University Press, 336 pages, ISBN: 0195362373.

Deo, M.C., Sridhar Naidu, C., 1999. Real time wave forecasting using neural networks. Ocean. Eng. 26, 191–203. https://doi.org/10.1016/S0029-8018(97)10025-7.

Deshmukh, A.N., Deo, M.C., Bhaskaran, P.K., Balakrishnan Nair, T.M., Sandhya, K.G., 2016. Neural-network-based data assimilation to improve numerical ocean wave forecast. IEEE J. Ocean. Eng. 41 (4) https://doi.org/10.1109/JOE.2016.2521222.

Embrechts, P., Lindskog, F., McNeil, A., 2001. Modelling Dependence with Copulas and Applications to Risk Management. Department of Mathematics, ETHZ, CH-8092 Zurich, Switzerland. URL. https://people.math.ethz.ch/~embrecht/ftp/copchapter.pdf.

Emmanouil, S., 2018. Error Correction for Wave Modelling, MSc Thesis. Delft University of Technology, Department of Civil Engineering and Geosciences, Delft, the Netherlands, pp. 1–193. http://resolver.tudelft.nl/uuid:e712c0c1-3c85-4cff-8455-443a84ff7537.

Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. 12 (4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347).

Hanea, A., Morales-Napoles, O., Ababei, D., 2015. Non-parametric Bayesian networks: improving theory and reviewing applications. Reliab. Eng. Syst. Saf. 144, 265–284. https://doi.org/10.1016/j.ress.2015.07.027.

Hasselmann, K., Olbers, D., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (JONSWAP). Ergnzungsheft zur Deutschen Hydrographischen Zeitschrift Reihe A 8 (12), 1–95. HDL:10013/epic.20654.

Haver, S., Moan, T., 1983. On some uncertainties related to the short term stochastic modelling of ocean waves. Appl. Ocean Res. 5 (2), 93–108. https://doi.org/10.1016/0141-1187(83)90021-4.

Hu, B., Stumpf, P., van der Deijl, W., 2019. Offshore wind access 2019, TNO, petten (Netherlands), TNO 2019 R10633, 40 pages, [Available online], URL. http://resolver.tudelft.nl/uuid:e8f05155-aa5a-4aad-a7ba-8bed2e9b08fe.

Jane, R., Dalla Valle, L., Simmonds, D., Raby, A., 2016. A copula-based approach for the estimation of wave height records through spatial correlation. Coast Eng. 117, 1–18. https://doi.org/10.1016/j.coastaleng.2016.06.008.

Jasper, N.H., 1956. Statistical distribution patterns of ocean waves and wave induced ship stresses and motions with engineering applications. Soc. Nav. Archit. Mar. Eng. Trans 64, 375–432. https://doi.org/10.5962/bhl.title.47257.

Jensen, F.V., 1996. Introduction to Bayesian Networks, first ed. Springer-Verlag. 186 pages, ISBN: 0387915028.

Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs, Information Science and Statistics. Springer Verlag, New York, ISBN 978-0-387-68282-2. https://doi.org/10.1007/978-0-387-68282-2, 448 pages.

Kim, C.H., 2008. Nonlinear Waves and Offshore Structures, first ed. World Scientific Publishing Company, ISBN 978-981-02-4884-0, p. 540. https://doi.org/10.1142/4906.

Kolibal, J., Howard, D., 2006. MALDI-TOF baseline drift removal using stochastic Bernstein approximation. EURASIP J. Appl. Signal Process. 2006, 1–9. https://doi.org/10.1155/ASP/2006/63582. Article ID 63582.

Kolibal, J., Howard, D., 2008. Alternative parametric boundary reconstruction method for biomedical imaging. J. Biomed. Biotechnol. 2008 https://doi.org/10.1155/2008/623475. Article ID: 623475, 7 pages.

Korb, K., Nicholson, A.E., 2010. Bayesian Artificial Intelligence, Chapman & Hall/CRC Computer Science & Data Analysis (Book 2), second ed. Chapman & Hall/CRC. 491 pages, -10: 1439815917.

Krishna Kumar, N., Savitha, R., Mamun, A.A., 2017. Regional ocean wave height prediction using sequential learning neural networks. Ocean. Eng. 129, 605–612. https://doi.org/10.1016/j.oceaneng.2016.10.033.

Kroon, A., de Schipper, M., den Heijer, K., Aarninkhof, S., van Gelder, P., 2017. Uncertainty assessment in coastal morphology prediction with a bayesian network. In: Aagaard, T., Deigaard, R., Fuhrman, D. (Eds.), Proceedings of Coastal Dynamics 2017: Helsingør, Denmark, pp. 1909–1920. Paper No. 254, URL. https://repository.tudelft.nl/islandora/object/uuid%3A424c4f19-4bd2-48cd-9221-b6e3e8426636.

Leontaris, G., Morales-Napoles, O., Wolfert, A.R.M. (Rogier), 2016. Probabilistic scheduling of offshore operations using copula based environmental time series – an application for cable installation management for offshore wind farms. Ocean. Eng. 125, 328–341. https://doi.org/10.1016/j.oceaneng.2016.08.029.

Li, W., Isberg, J., Waters, R., Engström, J., Svensson, O., Leijon, M., 2016. Statistical analysis of wave climate data using mixed distributions and extreme wave prediction. Energies 9 (6), 396. https://doi.org/10.3390/en9060396.

Londhe, S.N., Shah, Shalaka, Dixit, P.R., Balakrishnan Nair, T.M., Sirisha, P., Jain, R., 2016. A coupled numerical and artificial neural network model for improving location specific wave forecast. Appl. Ocean Res. 59, 483–491. https://doi.org/10.1016/j.apor.2016.07.004.

Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean. Eng. 31, 709–724. https://doi.org/10.1016/j.oceaneng.2003.05.003.

Malekmohamadi, I., Bazargan-Lari, M.R., Kerachian, R., Nikoo, M.R., Fallahnia, M., 2011. Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. Ocean. Eng. 38, 487–497. https://doi.org/10.1016/j.oceaneng.2010.11.020.

Mandal, S., Rao, S., Raju, D.H., 2005. Ocean wave parameters estimation using backpropagation neural networks. Mar. Struct. 18, 301–318. https://doi.org/10.1016/j.marstruc.2005.09.002.

Medina Oliva, G., Weber, P., Simon, C., Iung, B., 2009. Bayesian networks applications on dependability, risk analysis and maintenance. In: 2nd IFAC Workshop on Dependable Control of Discrete Systems DCDS`09, Bari, Italy, June 10-12, 2009. https://doi.org/10.3182/20090610-3-IT-4004.0040.

Morales-Nápoles, O., Steenbergen, R.D.J.M., 2014. Analysis of axle and vehicle load properties through Bayesian Networks based on Weigh-in-Motion data. Reliab. Eng. Syst. Saf. 125, 153–164. https://doi.org/10.1016/j.ress.2014.01.018.

Morales-Napoles, O., Worm, D., van den Haak, P., Hanea, A., Courage, W., Miraglia, S., 2013. Reader for Course: Introduction to Bayesian Networks, TNO Report, Reference: TNO-060-DTM-2013-01115.

Nelsen, R.B., 2006. An Introduction to Copulas, second ed. *Springer Series in Statistics*, Springer-Verlag, New York, ISBN 978-0-387-28678-5. https://doi.org/10.1007/0-387-28678-0.

Palmsten, M.L., Splinter, K.D., Plant, N.G., Stockdon, H.F., 2014. Probabilistic estimation of dune retreat on the Gold Coast, Australia. Shore Beach 82 (4), 35–43. https://pubs.er.usgs.gov/publication/70159345.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Series in Representation and Reasoning, first ed. Morgan Kaufmann Publishers, ISBN 978-0-08-051489-5.

Phillips, O.M., 2006. On the generation of waves by turbulent wind. J. Fluid Mech. 2 (5), 417–445. https://doi.org/10.1017/S0022112057000233.

Pierson, W.J., Moskowitz, L., 1964. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. J. Geophys. Res. 69 (24), 5181–5190. https://doi.org/10.1029/JZ069i024p05181.

Plant, N.G., Holland, K.T., 2011. Prediction and assimilation of surf-zone processes using a Bayesian network Part II: Inverse models. Coast Eng. 58, 256–266. https://doi.org/10.1016/j.coastaleng.2010.11.002.
Poelhekke, L., Jäger, W.S., van Dongere, A., Plomaritis, T.A., McCall, R., Ferreira, Ó., 2016. Predicting coastal hazards for sandy coasts with a Bayesian Network. Coast Eng. 118, 21–34. https://doi.org/10.1016/j.coastaleng.2016.08.011.
Ris, R.C., Booij, N., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions, Part II, Verification. J. Geophys. Res. 104 (C4), 7667–7681. https://doi.org/10.1029/1998JC900123.
Röckmann, C., Lagerveld, S., Stavenuiter, J., 2017. Operation and maintenance costs of offshore wind farms and potential multi-use platforms in the Dutch north sea. In: Buck, B., Langan, R. (Eds.), Aquaculture Perspective of Multi-Use Sites in the Open Ocean. Springer, Cham, ISBN 9783319511573, pp. 97–113. https://doi.org/10.1007/978-3-319-51159-7_4.
Russell, S.J., Norvig, P., 2009. Artificial Intelligence: A Modern Approach, third ed. Prentice Hall. 1152 pages, 10: 0136042597.
Schmidt, T., 2006. Coping with copulas. In: Rank, J. (Ed.), Copulas-From Theory to Applications in Finance, first ed. Incisive Media Risk Books, London, UK, ISBN 190433945X, pp. 3–34. 350 pages.
Schwarz, G.E., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464. https://doi.org/10.1214/aos/1176344136.
Scotto, M.G., Guedes Soares, C., 2007. Bayesian inference for long-term prediction of significant wave height. Coast Eng. 54, 393–400. https://doi.org/10.1016/j.coastaleng.2006.11.003.
Scutari, M., 2015. Bayesian network constraint-based structure learning algorithms: parallel and optimised implementations in the bnlearn R package. J. Stat. Software 10 (2), 20 pages, arXiv:1406.7648v2.
Scutari, M., Denis, J.-D., 2014. Bayesian Networks: with Examples in R, first ed. Chapman and Hall/CRC, ISBN 978-1482225587. 241 pages.
Sebastian, A.G., Dupuits, E.J.C., Morales-Napoles, O., 2017. Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed. Coast Eng. 125, 42–50. https://doi.org/10.1016/j.coastaleng.2017.03.008.
Seyfarth, R., Kolibal, J., Howard, D., 2006. New mathematical method for computer graphics. In: 2006 International Conference on Hybrid Information Technology, 9-11 Nov. 2006, Cheju Island, South Korea, ISBN 0-7695-2674-8. https://doi.org/10.1109/ICHIT.2006.253457.
Seyr, H., Muskulus, M., 2019. Decision support models for operations and maintenance for offshore wind farms: a review. Appl. Sci. 9 (2) https://doi.org/10.3390/app9020278, p. 278.
Shariff, A.A., Hadi Hafezi, M., 2012. Modelling significant wave height data of north sea: Rayleigh vs weibull distribution. Appl. Mech. Mater. 157–158, 652–657. http://www.scientific.net/AMM.157-158.652.
Stewart-Koster, B., Bunn, S.E., Mackay, S.J., Poff, N.L., Naiman, R.J., Lake, P.S., 2010. The use of Bayesian networks to guide investments in flow and catchment restoration for impaired river ecosystems. Freshw. Biol. 55 (1), 243–260. https://doi.org/10.1111/j.1365-2427.2009.02219.x.
Stumpf, H.P., Hu, B., 2018. *Offshore Wind Access 2018*, Energy Research Center (ECN), Petten (Netherlands), ECN-E–17-071, 46 pages, [Available online], URL. https://publications.ecn.nl/ECN-E-17-071.
Tsai, C.P., Lee, T.L., 1999. Back-propagation neural network in tidal-level forecasting. J. Waterw. Port, Coast. Ocean Eng. 125, 195–202. https://doi.org/10.1061/(ASCE)0733-950X(1999)125:4(195).
Vanem, E., 2016. Copula-based bivariate modelling of significant wave height and wave period and the effects of climate change on the joint distribution. In: Proc. OMAE 2016, 19-24 June, Busan, South Korea. https://doi.org/10.1115/OMAE2016-54314.
Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Eng. Appl. Artif. Intell. 25, 671–682. https://doi.org/10.1016/j.engappai.2010.06.002.
Wilson, K.E., Adams, P.N., Hapke, C.J., Lentz, E.E., Brenner, O., 2015. Application of Bayesian Networks to hindcast barrier island morphodynamics. Coast Eng. 102, 30–43. https://doi.org/10.1016/j.coastaleng.2015.04.006.