Journal of Physics: Conference Series

Using a Langevin model for the simulation of environmental conditions in an offshore wind farm

To cite this article: Helene Seyr and Michael Muskulus 2018 *J. Phys.: Conf. Ser.* **1104** 012023

IOP Conf. Series: Journal of Physics: Conf. Series **1104** (2018) 012023 doi:10.1088/1742-6596/1104/1/012023

Using a Langevin model for the simulation of environmental conditions in an offshore wind farm

Helene Seyr and Michael Muskulus

Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, NTNU 7491 Trondheim, Norway

E-mail: `helene.seyr@gmail.com`

Abstract. For the planning of operations and maintenance in offshore wind farms, many simulation models exist. Many rely on artificially generated weather time series to test different strategies. In this paper, we present a novel approach to modeling both the significant wave height and wind speed based on measurements from the site. We use a stochastic process called the Langevin process. First, equations are fitted to the available data, which are then used to generate the artificial weather data. The properties of these artificial weather time series are very close to the properties of the actual weather. Mean and standard deviation as well as the overall distribution and seasonality can be captured by the new model. Additionally, the persistence of waves and winds is replicated. This is especially important, as the length of weather windows is an important factor in operation and maintenance planning.

1. Introduction

Both in research and the wind industry, simulation models are often used to improve the operations and maintenance for offshore wind farms. There are research groups looking into optimal vessel routing, preventive maintenance strategies, optimization of corrective maintenance and condition monitoring among other topics. In order to provide simulations of an offshore wind farm, many of the existing models use weather time series to model the weather conditions, with significant wave heights and wind speeds. In order to model a specific location, without the risk of finding an optimal solution for a specific historical weather dataset, some researchers want to use artificial weather data. This artificial weather data should represent the given location and have the same properties, such as annual mean wind speeds or persistence of wave heights. The advantage with artificial weather data over historic weather data is that a natural variability of the weather can be achieved, without loss of site specific properties. Even if a model can theoretically be used with historic weather data, this data is not available of an appropriate length and quality for some locations. Also in this case, researchers benefit from artificial weather time series. In order to generate this artificial weather data, different methods can be used. Today, the three main choices for the simulation of weather conditions are Gaussian statistics, Auto RegressiveMoving-Average (ARMA) processes and Markov processes [1]. Different weather generation models have been developed using these methods and are being used in decision support tools. Dinwoodie et al. [2] use a multivariate autoregressive (MAR) process, an improved MAR process is described in detail by Dalgic et al. [3]. Scheu et al. [4] use a Markov chain approach which is described and analyzed in [5]. This approach was developed further to include more weather parameters by Hagen et al. [6]. Hersvik and

Endrerud [7] present an improved Markov chain process. In this paper, we want to introduce a novel approach to generating artificial weather series, based on the Langevin equations. In Section 2, this method is explained in more detail. Then we briefly explain the data sources used in this paper in section 3. An analysis of the generated weather is conducted in section 4, before we discuss the results in section 5. Finally, we conclude and give an outlook to further work in section 6.

2. Methodology

In this paper, we investigate a new approach to weather simulation, a Langevin process. Czechowski and Telesca [8] already fitted a Langevin type equation to wind speed data and the Langevin approach [9] has already been used to model turbulent wind velocities by Reinke et al. [10]. To our knowledge, the Langevin approach has not been applied to model the behavior of wind speeds on the scales used in offshore wind farm operation and maintenance models. Hadjihosseini et al. [11, 12] applied the approach to ocean waves, studying rogue wave phenomena, using data from Japan. They have shown that it is possible to use the Langevin approach to generate surrogate data sets and even forecast extreme wave events. However, the data used in their analysis had a sampling frequency of 1 Hz and the approach has not yet been used on data with a lower sampling frequency. As Hadjihosseini et al. [12] were interested in the study of rogue waves, they have not investigated other properties of the surrogate data, like the persistence of significant wave heights.

The Langevin process is a stochastic process governed by the Langevin equation.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X,t) + G(X,t,\Gamma) \tag{1}$$

This is a stochastic differential equation, including a deterministic term $F(X,t)$ and a stochastic term $G(X,t,\Gamma)$. Those terms are often referred to as drift and diffusion function and depend on the first two Kramers-Moyal coefficients $D^{(1)}(X)$, $D^{(2)}(X)$

$$F(X,t) = D^{(1)}(X)\tau \tag{2}$$

$$G(X,t) = \sqrt{D^{(2)}(X)\tau}\Gamma_t \tag{3}$$

where τ is a time-increment. It is important to note, that the coefficients depend on the observation X (in our case wave height or wind speed) as well as time t. Γ_t represents the stochastic forces. Hadjihosseini et al. [11] and Reinke et al. [10] have shown that the Langevin process can be described by the so-called Fokker-Planck equation, shown in Appendix A.1. The Kramers-Moyal coefficients can be defined as an expression of the conditional moments of the trajectory, as presented by [13], see Appendix A.2. The interested reader can find the derivations of the equations for the Langevin process and a detailed mathematical description in e.g. [11, 10].

When fitting the model to wave height data the two functions describing the dependence of the deterministic and stochastic contribution on the observation are estimated and later used for modeling. We will refer to these functions as the "parameter functions", the "drift-" ($D^{(1)}(X)$) and "diffusion- ($D^{(2)}(X)$) polynomial" and will use $D^{(1)}$ and $D^{(2)}$ as notation in the remainder of the paper.

In the analysis presented in this paper, we used the software R and the package 'Langevin' developed by Rinn et al. [14]. First, the parameter functions ($D^{(1)}$, $D^{(2)}$) were estimated based on the data. The software package, developed by Rinn et al., estimates the parameters for discrete points and reports back the estimated coefficients, their respective estimation errors and other information such as the mean and density of observations bin (interval) of the data used to estimate the coefficients.

In the second step a weighted linear regression taking into account the uncertainty of the estimate was conducted. The point estimates provided in step one where weighted with the inverse of the square of their respective error. For the wave heights, quadric polynomials were fitted for both parameter functions. Quadratic functions were chosen, based on the fit of the linear models to the estimated parameters. For the wind speed, a linear drift ($D^{(1)}$) and a quadratic diffusion ($D^{(2)}$) function were chosen based on best fit to the estimated parameters.
Following these two steps, the obtained parameter functions were used to generate artificial weather time series. This is done by using the quadraticlinear functions estimated previously as an input to the data generation. The timeseries generation function from Rinn et al. [14] had to be modified in order to generate realistic results for our specific case. The original function 'timeseries1D' generates non-negative as well as negative observations. Since a negative wave height is not meaningful in this context, the function was modified in order to assure only non-negative observations. In the modified function, the Langevin process is allowed to continue into the negative domain, but observations are only added to the simulated weather time series once the process crosses back into non-negative values again. The artificial time series and properties are compared to the original time series in Section 4.

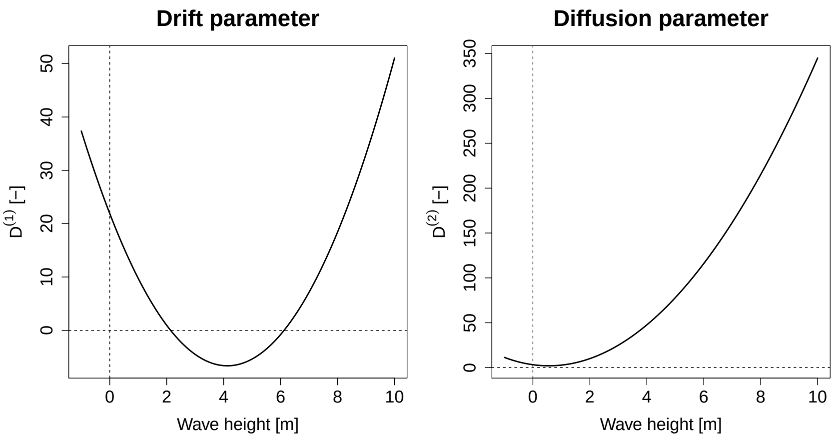

Figure 1. The drift and diffusion polynomial for the significant wave height in January, based on the data from FINO 1. The parameters $D^{(1)}$ and $D^{(2)}$ in the Fokker-Planck equation depend on the wave height.

3. Data

For our analysis, we investigated two different publicly available datasets. The re-analysis data from the ECMWF [15] is available in different resolutions, and we used the data that we already used for a different study [16]. These data have a resolution of six hours, providing one measured wave height and wind speed in the center of each six hour interval.
Additionally, data from the FINO measurement campaign [17] was used for the measurement platform FINO1 next to the Alpha Ventus wind farm in the North Sea. Here the significant wave height is provided in 30 min mean values and wind speed measurements are available in 10min mean wind speed steps.

4. Analysis of the artificial weather data

In this section the properties of the artificial weather series are compared to those of the actual weather data to see how well the Langevin approach captures the site specific weather properties.

IOP Conf. Series: Journal of Physics: Conf. Series **1104** (2018) 012023 doi:10.1088/1742-6596/1104/1/012023

Results are presented for both datasets. Not every investigation is shown for both sites, the performance of the model is similar in both cases and we have chosen to show plots from different sites in order to represent different weather conditions.

Table 1. Statistics of the wave height simulations based on the FINO 1 data.

Wave height	Mean	Maximum	SD
Data	1.44	9.77	0.93
Simulation	1.51	7.49	0.92
Simulation without seasonal effect	1.44	8.62	0.93
Simulation based on six-hour data	1.44	9.90	1.0

Table 1 shows some of the statistics of the data for significant wave height for the measurement at FINO 1. We included the mean, maximum and standard deviation from the original data and three different simulations based on the Langevin process. The first simulation, referred to as "simulation" later, is based on one Langevin process. The equations for this process were fitted to the whole data set, therefore the model is not able to capture any seasonality in the weather. The second simulation is based on Langevin equations that have been estimated based on the data for each month, we will refer to this as the "seasonal model" or "seasonal simulation" later in the paper. There are two equations for January, two equations for February and so on. An example of these parameter equations can be seen in Figure 1 for the month of January for the FINO 1 data. The third model is based on six-hour data and is obtained by fitting monthly Langevin equations to data that was previously filtered to have one observation every six hours. This model will also be referred to as "six-hour simulation" in the remainder of the paper. Table 1 shows that mean and standard deviation are replicated quite well by all different simulation models.

Table 2. Statistics of the wind speed simulations based on the FINO 1 data.

Wind speeds	Mean	Maximum	SD
Data	9.99	37.01	4.66
Simulation	9.83	35.57	4.38
Simulation without seasonal effect	10.03	36.25	4.34
Simulation based on six-hour data	10.0	25.68	4.01

The same can be seen in Table 2 for the wind speeds. Also here, the same three kinds of simulations are presented and the mean and standard deviation are again reproduced well. The maximum value in the simulation shows a greater discrepancy. The reason for this lies presumably in the generation of the six-hour data. When calculating the mean over six hours, the extrema are smoothed out. By using the six-hour data to fit the Langevin equations, the extreme values can then not be reproduced by the model.
An important aspect of the weather properties at a given site, is of course seasonality. The weather is generally harsher in winter months, with higher waves and wind speeds. In order to capture the seasonality, different Langevin equations where fitted for each month, as described

above. To see whether this is sufficient to capture the seasonality, we investigated the monthly means of both significant wave height and wind speed. Figure 2 shows box-plots for the original data and simulation of the significant wave height and wind speed at a location off the British coast, where data from the ECMWF was available. It can be seen that the seasonality in both the significant wave height and wind speed is captured well by the new model based on the Langevin process.

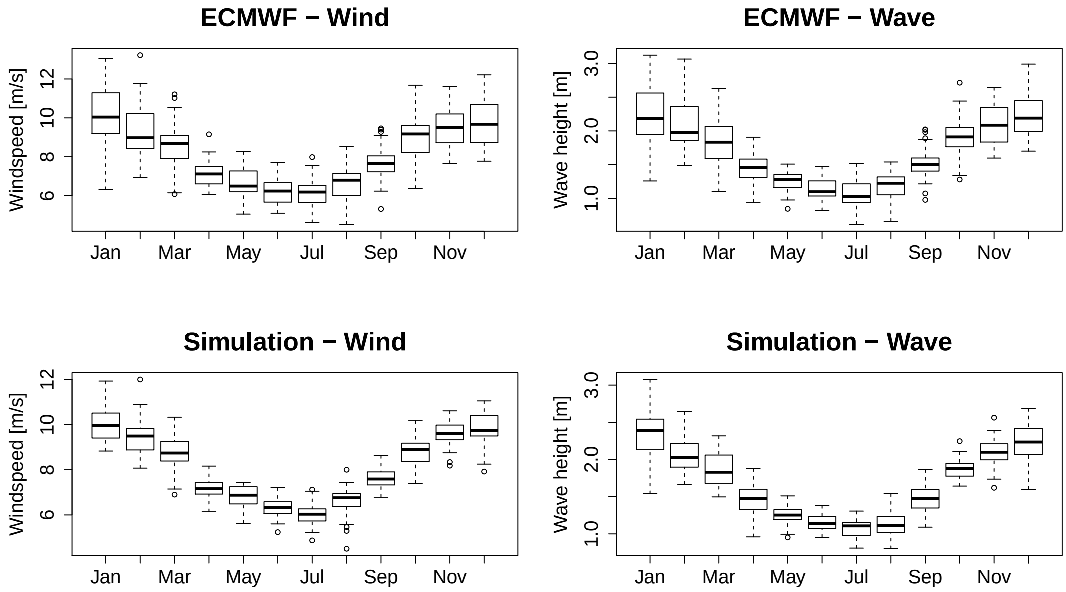

Figure 2. Monthly means of the significant wave height and wind speed at a British offshore location.

Not only the seasonality is an important characteristic of the local weather. Also the distribution of the wave heights and windspeeds throughout the year should be similar to the actual observations. In Figure 3, the distribution of the wave heights is shown for the location of FINO 1. Here the effect of using the simulation based on multiple Langevin equations can be observed. The simulation that was based on one separate set of equations for each month performs better than the simulation without seasonal effect. The same has been observed for the other location.

For the wind speed simulation, the distribution is not matched as well as for the wave heights as can be seen from Figure 4. The same observation of the distribution of simulated wave heights and wind speeds can be made for the British offshore site A1. One possible explanation is that the Langevin process is better at capturing the wave specific properties and it might not be the correct way to model wind speeds. Hadjihosseini et al. [11] have proven that the Langevin process can be used to describe ocean waves. In the absence of a similar investigation for wind speeds, we have in this study assumed that the Langevin process can also be used. It is possible that this assumption does not hold up, this will be the basis for new investigations. Even if the assumption holds, it might be that the type of equations that we fitted with the linear regression, were indeed not suitable to capture all properties of the wind speed.

To show the difference in wave height distribution between winter and summer, January and August were chosen as representative months in Figure 5. It can be seen that the wave height is subject to more variation in the winter, with a higher mean significant wave height. For both

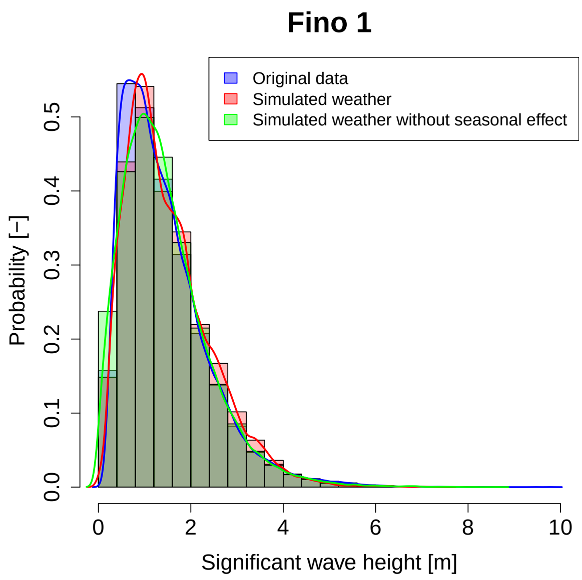

Figure 3. Distribution of the wave heights over one year for the FINO 1 data, seasonal simulation and simulation without the seasonal effect.

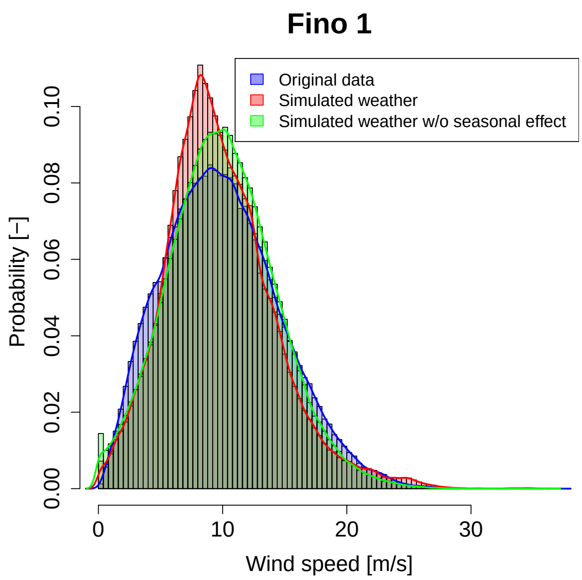

Figure 4. Distribution of the wind speeds over one year for the FINO 1 data, seasonal simulation and simulation without the seasonal effect.

months, the distribution is well matched by the artificial wave height series. Similar observations can be made for the simulations based on the FINO 1 data.

Figures 6 and 7 show the cumulative distribution functions (CDFs) of the wind speed and wave height respectively for each a summer and a winter month for the location of FINO 1.

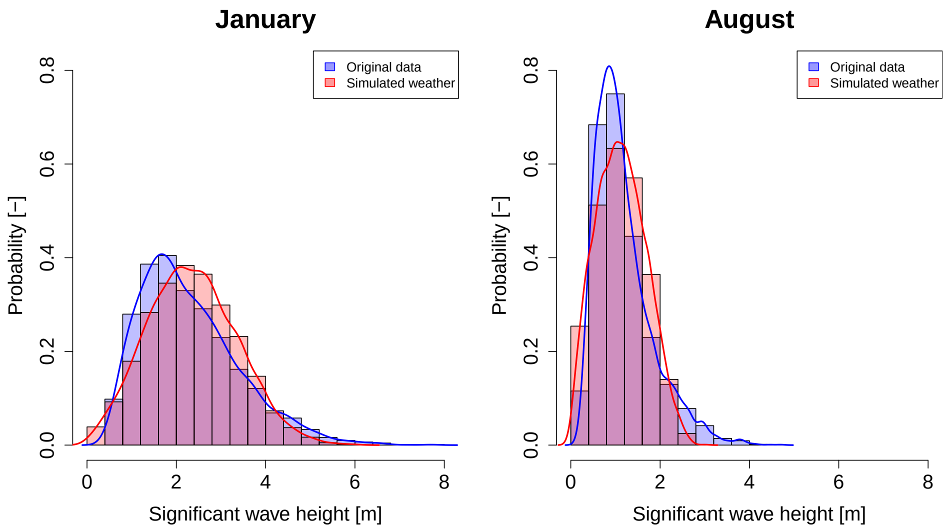

Figure 5. Distribution of the significant wave heights over one year for the ECMWF data. One month in summer and one month in winter were chosen. For both months, the distribution of the original data and of the artificial data are shown.

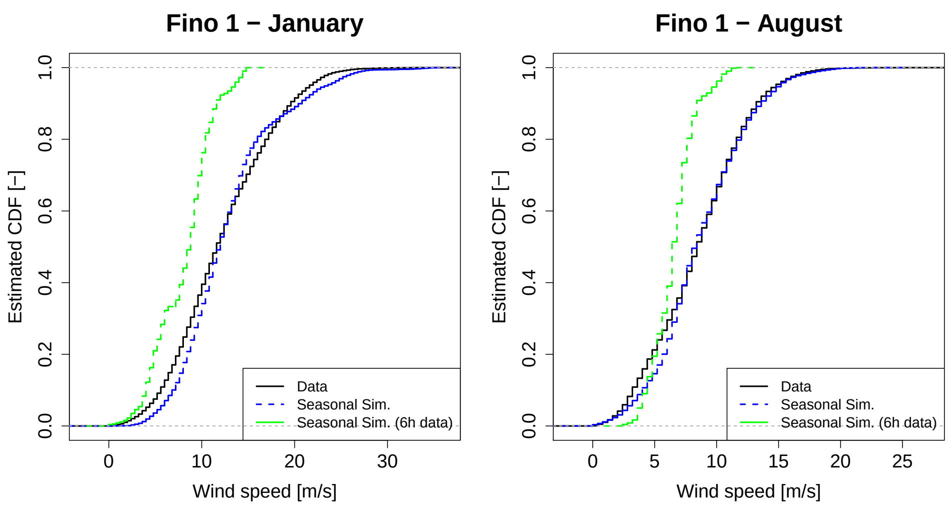

Figure 6. Cumulative distribution function of wind speeds for a winter and a summer month at the location of FINO 1.

Figure 6 shows the CDF of the wind speeds for a winter month and a summer month. It can be observed that the simulation based on six-hour data under-estimates the wind speed at the given site for January. For August, the model also under-estimates the wind speeds. However, it can be seen that the occurrence of wind speeds below $5m/s$ is also under-estimated. The model based on the original data replicates the CDF of the wind speeds much better. here, in both months the occurrence of low wind speeds is slightly under-estimated and the CDF is replicated better for the summer month. For the wave heights, the CDF is shown in Figure 7, again for a summer and a winter month. Also here, we observe that the resolution in the data used for

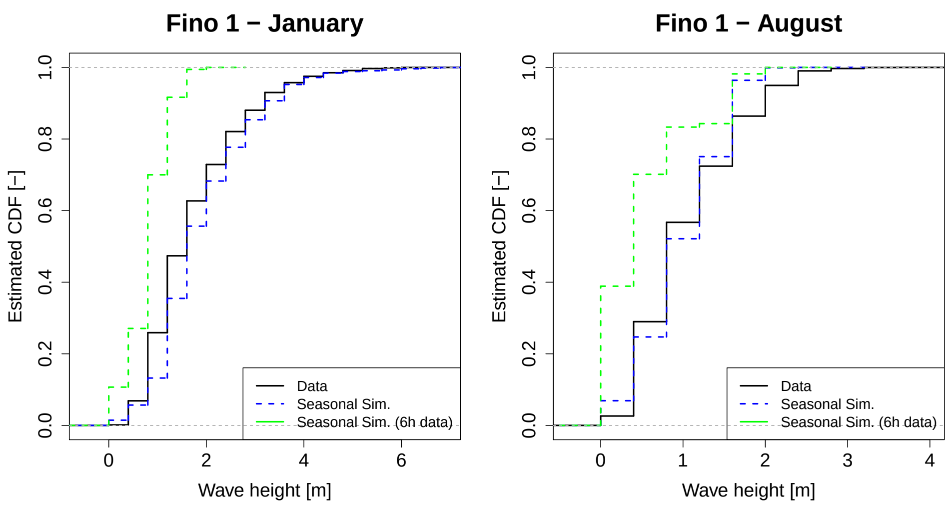

Figure 7. Cumulative distribution function of wave heights for a winter and a summer month at the location of FINO 1.

fitting the equations has an influence on how well the CDFs are represented. Contrary to the wind speed simulations, CDF of significant wave heights is better matched by the simulation for the winter month.

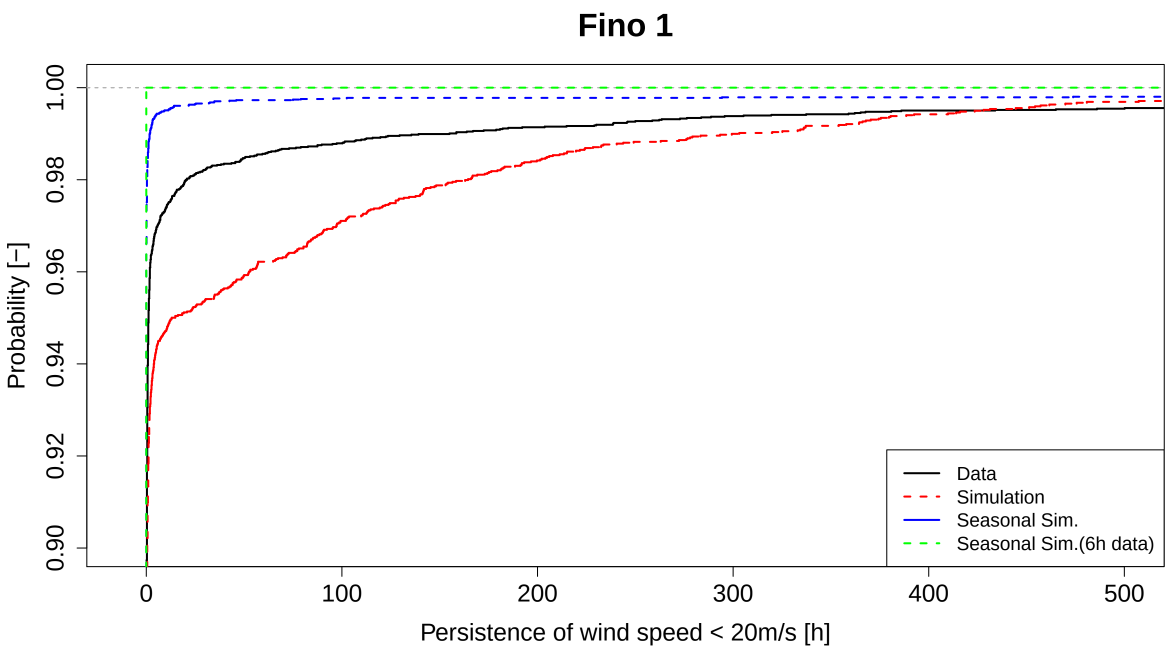

Figure 8. Persistence of wind speeds below $20m/s$ for the location of FINO 1. The data is compared to the simulation without seasonal effect, the simulation with seasonal effect and the simulation with seasonal effect, based on 6 hour data.

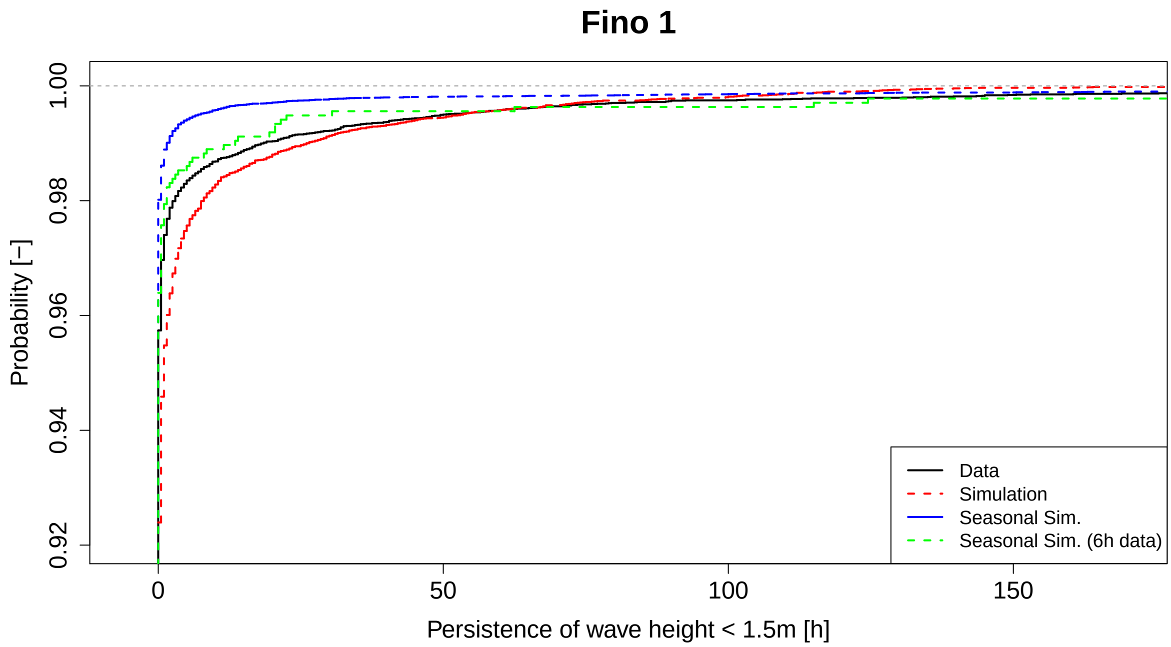

Figure 9. Persistence of wave heights below $1.5m$ for the location of FINO 1. The data is compared to the simulation without seasonal effect, the simulation with seasonal effect and the simulation with seasonal effect, based on 6 hour data.

For the planning of operations and maintenance, weather windows play an important role. In order to see whether the length of these windows is adequately represented in the artificial weather series, we also investigated the persistence of waves and wind speeds. Anastasiou and Tsekos [18] define the persistence of wave height below a threshold level as "the time interval between a down-crossing of that threshold and the first subsequent up-crossing". For the wave height, a threshold of $1.5m$ height was chosen and for the wind speeds a threshold of $20m/s$ was used to investigate the persistence. These values were chosen based on typical wave height and wind speed limits for different vessels used for operation and maintenance (e.g. crew transfer or lifting operations).

Figure 8 shows the persistence of wind speeds below $20m/s$ for the location of the FINO 1 measurement platform. It is necessary to note that the distribution of the persistence could not be calculated for the simulation based on the six-hour data. In the simulated time series, no value above $20m/s$ was present. Therefore, the probability of having a persistence lower than $20m/s$ of the length of the dataset it one. The length of suitable weather windows is over predicted by the seasonal simulation model and under-predicted by the model without seasonal variation. The Kolmogorov-Smirnov-(KS-)distance of the distribution of the persistence is slightly bigger for the non-seasonal model (0.16) than for the seasonal model (0.14), but both are in the same magnitude and it cannot be rejected that the samples come from the same distribution as the data (p-values > 0.8 for all three simulations). Depending on the application (e.g. turbine performance calculation, scheduling a lift operation), the more conservative model might be chosen.

Figure 9 shows the persistence statistics for wave heights below $1.5m$ for the location of FINO 1. Also for the waves, the simulation without the seasonal effect under-predicts the lengths of weather window, whereas the model with seasonal variation over-predicts the lengths.

Calculating the KS-distance shows that the maximum distance between the persistence-curves of the two seasonal simulations (0.04 for both models) is not significantly different (p-values > 0.95) from the persistence of the original data. This means that the over- and under-prediction of the two models are of the same magnitude. The KS-distance of the simulation without seasonal effect (0.08) is about double the KS-distance of the two seasonal simulations. Also this model does not show a significant (p-value > 0.9) difference from the original persistence distribution.

5. Discussion

The popular Markov Process uses discrete time steps. One challenge when using Markov chains is the change from one transition probability matrix to the next. This happens e.g. when switching from one month to the next in the model presented by [4]. The last value generated in a month needs to be used for the generation of values in the next month. However, if this value has an occurrence probability of zero in the new matrix, the process cannot select a first value for the new month.

The Langevin process however is a continuous process and does not have any issues with different starting values for the process. Here, the last generated value for a month can be directly input as a starting value for the next month's process. Additionally, for the Langevin process fewer parameters need to be estimated and therefore less data is needed to fit the model. While for the Markov chain, many transition probabilities need to be calculated and large matrices handled, the Langevin equations can be describes with a handful of parameters.

Having fewer parameters that describe the Langevin model compared to the Markov model, has the disadvantage that some of the properties of the data cannot be replicated as well as with the Markov model. Usually, models with fewer parameters need fewer observations to estimate these parameters. In the given application this would mean that a shorter weather observation period can be used to base the site weather model on. This might especially be useful to optimize the operation and maintenance for new wind farm projects, where data is collected for a short period of time. Additionally, an advantages of the Langevin model is that the model does not require the handling of large matrices and random sampling with discrete probabilities. The analysis of the properties of the artificial weather data generated based on the Langevin process showed that the properties of the site specific weather conditions are conserved well enough to use the model for operation and maintenance optimization. It is necessary to note that the Markov model with its many more parameters is (for the location and dataset used in this study) superior in replicating the site properties of the weather. A comparison of the distribution of the significant wave height and wind speed for the data from the British offshore site and the simulations of this site with the (seasonal) Langevin model and Markov model can be found in the Appendix in Figure A2.

The possible correlation between wave heights and wind speeds are missing in both applications. Previously, a correlation matrix has been used by [4] to generate the wind speeds from the wave height simulation. Using a multidimensional (2D) Langevin process could also solve the issue of correlation.

6. Conclusion and further work

In this paper we have shown, that for the type of application that was the focus of our investigation, namely generating artificial weather time series for operations and maintenance simulations, a Langevin process can be used. The properties of the waves are represented well, both in terms of distribution of the significant wave height and in terms on persistence of waves. As with most data-driven models, the performance of the Langevin process improves with the quality and sampling frequency of the data used to fit the equations. Higher data sampling frequencies lead to a better representation of the site conditions. Especially the persistence is an important property that is needed for O&M simulations, since the length of weather windows

plays an important role in deciding e.g. a maintenance strategy.
In the future, the Langevin approach might be a tool that can assist in propagating site specific properties of the waves without relying on simulations. It might also be used as an alternative input to closed form models like the one from Feuchtwang and Infield [19] or the simulations models mentioned above. The Langevin model presented here models the mean wind speeds and wave heights independently. Hagen et al. [6] have presented a multivariate approach for the Markov chain model to capture correlations between different weather parameters. Capturing the correlation between wave heights, wind speeds (and other weather parameters) should be possible by using a two (multi) dimensional Langevin process. This is another topic for further research.

Acknowledgments
The work leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska- Curie grant agreement No 642108.
For the data collected in the FINO measurement campaign, we would like to thank the "Bundesministerium für Wirtschaft und Energie" (the German Federal Ministry for Economic Affairs and Energy) and "Projektträger Jülich", PTJ.

7. References

[1] Monbet V, Ailliot P and Prevosto M 2007 *Probabilistic Engineering Mechanics* **22** 113–126
[2] Dinwoodie I, McMillan D, Revie M, Lazakis I and Dalgic Y 2013 *Energy Procedia* **35** 157 – 166
[3] Dalgic Y, Lazakis I, Dinwoodie I, David M and Revie M 2015 *Ocean Engineering* **101** 211–226
[4] Scheu M, Matha D, Hofmann M and Muskulus M 2012 *Energy Procedia* **24** 281–288
[5] Scheu M N, Matha D and Muskulus M 2012 Validation of a markov-based weather model for simulation of o&m for offshore wind farms *The 22nd Int. Offshore and Polar Engineering Conf.* (Int. Society of Offshore and Polar Engineers)
[6] Brede Hagen Ingve Simonsen M H M M 2013 *Energy Procedia* **35** 137–147
[7] Hersvik K and Endrerud O E V 2017 *IOP Conf. Series: Materials Science and Engineering* **276** 012003
[8] Czechowski Z and Telesca L 2013 *Physica A: Statistical Mechanics and its Applications* **392** 5592 – 5603
[9] Friedrich R, Peinke J, Sahimi M and Tabar M R R 2011 *Physics Reports* **506** 87 – 162 ISSN 0370-1573
[10] Reinke N, Fuchs A, Medjroubi W, Lind P G, Wächter M and Peinke J 2015 *The Langevin Approach: A Simple Stochastic Method for Complex Phenomena* (Springer International Publishing) pp 125–141
[11] Hadjihosseini A, Hoffmann N P and Peinke J 2014 *New Journal of Physics* **16** 053037
[12] Hadjihosseini A, Wächter M, Hoffmann N P and Peinke J 2016 *New Journal of Physics* **18** 013017
[13] Risken H 1996 Fokker-planck equation *The Fokker-Planck Equation* (Springer) pp 63–95
[14] Rinn P, P L, M W and J P 2016 *Journal of Open Research Software* **4**(1)
[15] Dee D P, Uppala S M, Simmons A J, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M A, Balsamo G, Bauer P, Bechtold P, Beljaars A C M, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer A J, Haimberger L, Healy S B, Hersbach H, Hólm E V, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally A P, Monge-Sanz B M, Morcrette J J, Park B K, Peubey C, de Rosnay P, Tavolato C, Thépaut J N and Vitart F 2011 *Quarterly Journal of the Royal Meteorological Society* **137** 553–597 ISSN 1477-870X
[16] Seyr H and Muskulus M 2016 *Journal of Physics: Conference Series* **753** 092009
[17] DEWI - Deutsches Windenergie-Institut 2012 FINO1 measurement platform - Installation Documentation for customers, Technical Report
[18] Anastasiou K and Tsekos C 1996 *Applied Ocean Research* **18** 187 – 199 ISSN 0141-1187 URL `http://www.sciencedirect.com/science/article/pii/S0141118796000302`
[19] Feuchtwang J and Infield D 2013 *Wind Energy* ISSN 10954244

Appendix

$$\frac{\partial P(X)}{\partial t} = \left(-\frac{\partial}{\partial X} D^{(1)}(X) + \frac{\partial^2}{\partial X^2} D^{(2)}(X) \right) P(X) \tag{A.1}$$

Equation A.1: Fokker-Planck equations, describing a Langevin process.

$$D^{(n)}(X) = \lim_{\tau \to 0} \frac{1}{n!\tau} M^{(n)}(X, \tau) \tag{A.2}$$

Equation A.2: Equation for the Kramers-Moyal coefficients, where τ is a time increment and $M^{(n)}(X, \tau)$ is the conditional moment of the Langevin processes' trajectory in time, with respect to τ.

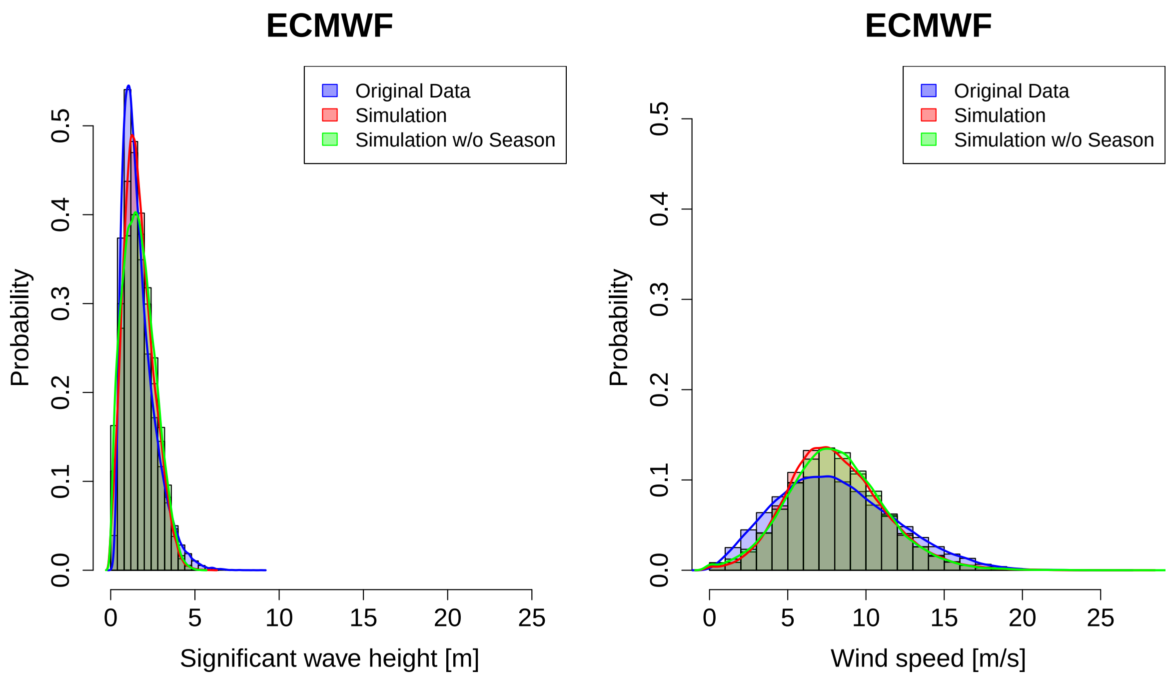

Figure A1. Distributions of the significant wave height and wind speed for the British offshore site. Blue line and shading show the distribution of the original data, red shows the simulated weather based on the seasonal Langevin model introduced in this paper and the green line and shading show the distribution of the Langevin model without taking into account seasonal effects.

IOP Conf. Series: Journal of Physics: Conf. Series **1104** (2018) 012023 doi:10.1088/1742-6596/1104/1/012023

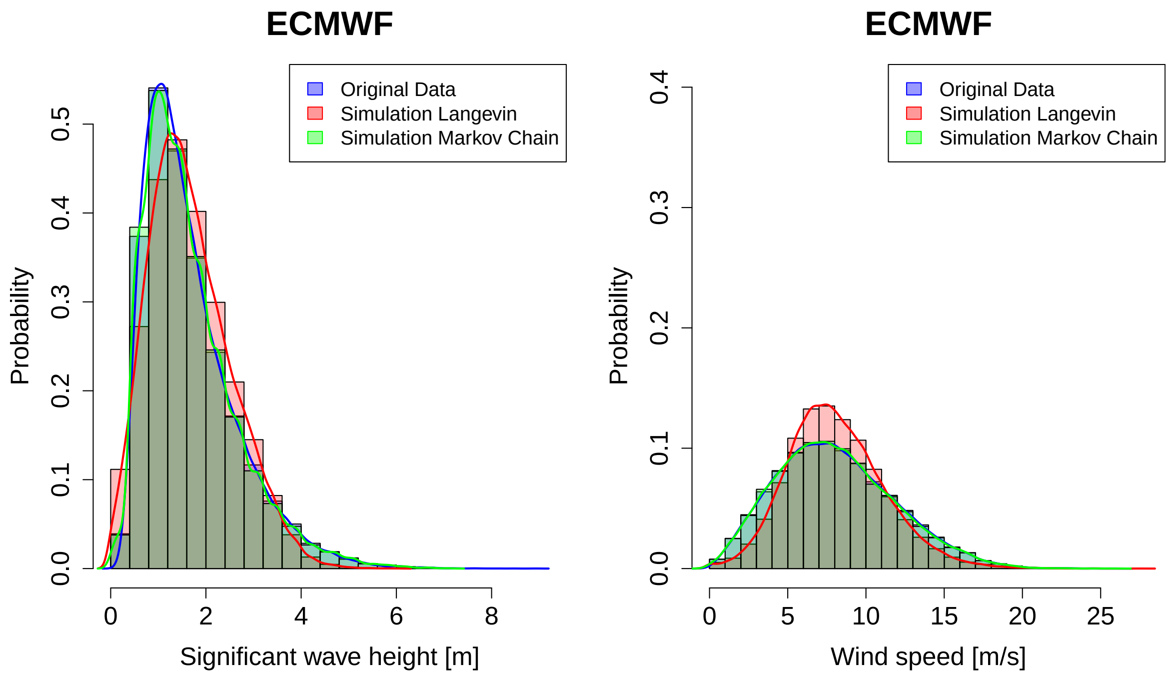

Figure A2. Distributions of the significant wave height and wind speed for the British offshore site. Blue line and shading show the distribution of the original data, red shows the simulated weather based on the Langevin model introduced in this paper and the green line and shading show the distribution of the artificial weather generated by a Markov chain model.