High-resolution offshore wind resource assessment at turbine hub height with Sentinel-1 SAR data and machine learning

Louis de Montera[1], Henrick Berger[1], Romain Husson[1], Pascal Appelghem[2], Laurent Guerlou[1], Mauricio Fragoso[1]

[1]CLS Collecte Localisation Satellites, Ramonville-Saint-Agne, France
[2]Atmosky, Talence, France

Correspondence to: Romain Husson (rhusson@groupcls.com)

Abstract. This paper presents a method to calculate offshore wind power at turbine hub height from Sentinel-1 Synthetic Aperture Radar (SAR) data using machine learning. The method is tested in two 70 km x 70 km areas off the Dutch coast where measurements from Doppler wind Lidars installed at the sea surface are available and can be used as a reference. Firstly, SAR wind speeds at surface level are improved with a machine learning algorithm using geometrical characteristics of the sensors and parameters related to the atmospheric stability extracted from a high-resolution numerical model. The SAR wind speed bias against Lidar measurements at 10 m above sea level is reduced from -0.42 m s^{-1} to 0.02 m s^{-1}, and its standard deviation from 1.41 m s^{-1} to 0.98 m s^{-1}. After improvement, SAR surface wind speeds are extrapolated at higher altitudes with a separate machine learning algorithm trained with the wind profiles measured by the Lidars and additional parameters from the high-resolution numerical model. Once the wind speed at turbine hub height is obtained, we assume the presence of an 10 MW turbine with a typical power curve. The extractible wind power is calculated by obtaining the wind speed Weibull distribution with the method of the moments, and then multiplying it by the turbine power curve. The accuracy of the extractible wind power derived from SAR data when compared with Lidar measurements is ±3%. The additional error due to SAR satellites low temporal sampling is estimated at ±2%, but this error can be easily removed by using a numerical model to simulate the satellites' passages and estimating it. Finally, wind power maps at 200 m are presented and compared with the outputs of the numerical model at the same altitude. The maps based on SAR data have a much higher level of detail, especially for the coastal wind gradient. The new revealed patterns show differences with the numerical model of as much as 10% in some locations over distances of the order of 20 km. We conclude that SAR data combined with a high-resolution numerical model and machine learning techniques can improve the wind power estimation at turbine hub height, and thus provide useful insights for optimizing wind farm siting and risk management.

1 Introduction

Estimating the extractible offshore wind power at turbine hub height is a challenging problem due to the difficulty in measuring the wind profile in the boundary layer over the sea. Currently, it is estimated by using numerical models and/or Doppler wind Lidars installed at the sea surface pointing upwards (NREL, 2020). Lidars provide the complete wind profile at one location with a high temporal sampling, but are very expensive to operate. Therefore, only one or two are typically used to sound large areas. Conversely, numerical models provide outputs over the entire area of interest. However, they are not capable of resolving small scale phenomena due to their physics and resolution. As a result, their errors are not precisely known and may vary in time and space. This is particularly problematic in coastal areas where processes are more complex and involve smaller scales. Due to these limitations, considerable uncertainty remains about actual offshore wind resources, which can affect wind farm project planning and management.

The need to improve wind speed assessment, and thus estimating more precisely wind power availability throughout wind farms' life cycle, has led to a growing interest in using satellite data to estimate wind resources (see, e.g., Hasager et al., 2015). Contrary to ground-based Lidars, spaceborne sensors have the advantage of sounding large areas with high spatial resolution. However, they are not perfect: their revisit period is typically long (a couple of days for Sentinel-1 in Europe, for example) and they use an indirect measurement based on the sea state Radar backscatter. Therefore, their measurements are impacted by several sources of potential error (low temporal sampling, sensor geometry, currents, algae, bright targets, rain cells, bathymetry, turbulence, etc.). Moreover, the extrapolation of their measurements from the sea surface to hub height is not an easy task due to the variety of meteorological conditions that may impact the wind speed extrapolation ratio.

Several studies have attempted to assess offshore wind power potential with spaceborne scatterometers, such as ERS-1, ERS-2, NSCAT, QuickSCAT, and ASCAT (Sánchez et al., 2007; Pimenta et al., 2008; Karagali et al., 2014; Bentamy and Croize-Fillon, 2014; Remmers et al., 2019). However, the resolution of these instruments is at best 12.5 km^2, which is not adapted to coastal areas due to land contamination. Synthetic Aperture Radar (SAR) satellites are an interesting alternative because wind products derived from their measurements have a much finer resolution of 1 km. The potential of SAR data has already been assessed by numerous studies (Hasager et al., 2002; Hasager et al., 2005; Hasager et al., 2006; Christiansen et al., 2006; Hasager et al., 2011; Hasager et al., 2014; Chang et al., 2014; Chang et al., 2015; Hasager et al., 2020). However, validating SAR measurements with in-situ data has been limited (Ahsbahs et al., 2017; Badger et al., 2019; de Montera et al., 2020; Ahsbahs et al., 2020) and these studies concluded that important biases remained (in the context of this study, the term 'in-situ instruments' includes profiling Lidars, although technically they use remote sensing).

Regarding the extrapolation of surface wind speeds to higher altitudes, the statistical theory of turbulence provides theoretical wind profiles (see, e.g., Grachev and Fairall, 1996). However, the problem has not been satisfactorily resolved and becomes increasingly critical as the typical height of windmills increases. Empirical evidence from offshore meteorological masts measurements suggests that a simple power law could be sufficient to model the wind profile (Hsu et al., 1994). Nevertheless, the analysis of Lidar data shows that, above 40 m, the power law is no longer accurate. This limitation has led some authors

to use numerical models to improve the extrapolation to higher altitudes (Badger et al., 2016). The advantage of numerical models is that they provide information about atmospheric stability through parameters like surface temperature and surface heat flux. In Badger et al. (2016), these surface parameters were averaged and combined with the similarity theory of Monin-Obukhov to extrapolate wind Weibull parameters. However, to our knowledge, this method was validated with only one meteorological mast in the Baltic Sea and not higher than an altitude of 100 m. Therefore, more research is needed to improve the estimation of wind resources at hub height with SAR data, and convince the industry to use them.

Due to the high number of possible sources of error in the retrieval of surface winds with SAR satellites, machine learning seems appropriate to improve the accuracy of SAR surface wind speeds. Regarding their extrapolation at higher altitudes, on land, machine learning has also been found to improve the accuracy of wind speeds extrapolated at turbine hub height compared to power laws or logarithmic laws (Türkan et al., 2016; Mohandes and Rehman, 2018; Vassallo et al., 2019). Optis et al. (2021) also found that machine learning was more efficient at extrapolating offshore winds than theoretical approaches. Moreover, it has been shown that, even if the algorithm is trained with a few in-situ instruments, it can be applied in a large area around them without significantly degrading the accuracy (Bodini and Optis, 2020; Optis et al., 2021). Therefore, we also chose to use machine learning in order to derive wind speeds at hub height from SAR surface winds. As in Badger et al. (2016), we take advantage of a numerical model to assess the atmospheric stability and provide the algorithm with the relevant meteorological parameters.

Section 2 describes the SAR data used in this study, the numerical model, the Lidar data used as a reference to train the algorithms, and the formulas used to compute the wind power. Section 3 describes the two machine learning algorithms designed to improve the accuracy of SAR surface winds and extrapolate them to hub height, respectively. The reason for separating the method into two algorithms is the scarcity of offshore Lidar data. Since the first algorithm correcting SAR surface wind biases depends on geometric properties of the sensor, it may be improved by using a large network of classical metocean buoys as a training dataset in the future. On the contrary, the algorithm extrapolating surface winds to higher altitudes only depends on meteorological parameters. Therefore, it can be trained with a few Lidars in one location and applied in other areas (if similar meteorological conditions are met). In Section 4, the method is tested in two areas off the Dutch coast where profiling Lidar data are available. The SAR wind speeds extrapolated at hub height are converted into a Weibull distribution, and the extractible power is obtained by simulating the presence of a typical 10 MW wind turbine operating at 200 m. The resulting maps are presented and compared with the output of the numerical model in order to estimate the benefit of using this method compared with a state-of-the-art technique.

2 Data and Methodology

2.1 High-resolution numerical model

The two zones of study are located off the Dutch coast (Figure 1). They have an approximate size of 70 x 70 km. Their geographic extent was defined in order to include offshore profiling Lidars and a part of the coastline in order to observe the

wind gradient. The WRF (Weather Research and Forecasting) non-hydrostatic meso-scale model (Skamarock et al., 2019) was run over these areas with a resolution of 1 km. The Planetary Boundary Layer (PBL) parametrization of the model was based on Hahmann etal., 2020. It was forced at its boundary limits by a larger-scale model, the reanalyzed GFS (Global Forecast System) having a resolution of 0.5° developed by NCEP (National Centers for Environmental Prediction). This larger-scale model was downscaled before using it to force the WRF model. This type of numerical model is representative of how wind resources are currently assessed by the industry. Here, it is used to estimate the atmospheric stability and extrapolate SAR surface winds to turbine hub height. The WRF model was run from January 2017 to December 2019. It provides the wind speed and direction from surface to 200 mwith increments of 20 m. It also provides other meteorological variables, such as air and sea surface temperature, surface heat flux, relative humidity, and pressure.

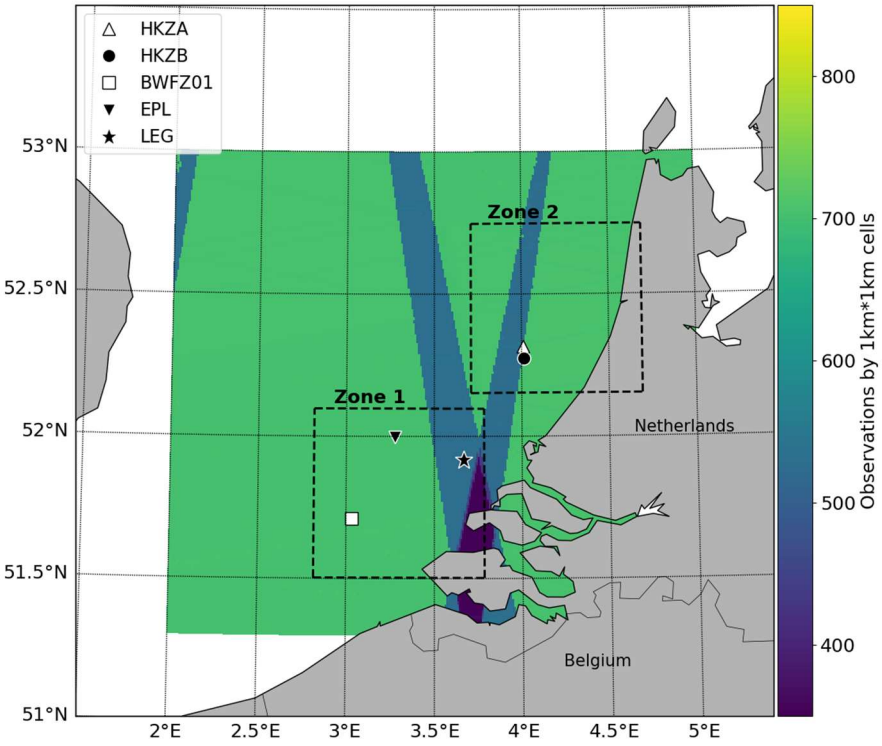

Figure 1: Locations of Zone 1 (bottom, latitude 51.50°N - 52.09°N / longitude 2.82°E - 3.77°E) and Zone 2 (top, latitude 52.15°N - 52.74°N / longitude 3.71°E - 4.68°E) with the positions of the profiling Lidars. The colour represents the number of Sentinel-1 SAR Level 2 wind observations during years 2017, 2018 and 2019.

2.2 In-situ instruments

The dataset used in this study comprises five ground-based profiling Lidars located off the Dutch coast (Figure 1). They are named HKZA, HKZB, BWFZ01, EPL and LEG. HKZ stand for Hollandse Kust Zuid wind farm, BWF for Borssele Wind Farm Zone, EPL for European Platform, and LEG for Lichteiland Goeree platform. Zone 1 includes the Lidars BWFZ01, EPL and LEG, and Zone 2 includes the Lidars HKZA and HKZB. Lidars HKZA, HKZB, BWFZ01 are floating. Lidars EPL and LEG are installed on platforms. The wind speed and direction are 10-minutes averaged around the observation times. The data were quality checked by our data provider C2WIND (for each time intervals, the minimum number of packets was set at 20 and the minimum availability at 80%). The vertical sampling and the duration of these Lidar measurements varies between observation campaigns and are displayed in Table 1.

Lidar	Longitude	Latitude	First date	Last date	Number of levels	Lowest altitude	Highest altitude
HKZA	4.011°E	52.309°N	2016-06-05	2018-06-05	11	30m	200m
HKZB	4.013°E	52.292°N	2016-06-05	2018-06-05	11	30m	200m
LEG	3.667°E	51.917°N	2014-11-17	2017-03-31	10	61m	300m
EPL	3.276°E	51.998°N	2016-05-30	2017-03-31	11	61m	290m
BWFZ01	3.033°E	51.71°N	2015-06-11	2017-02-27	10	30m	200m

Table 1: Main characteristics of the five profiling lidars

For each Lidar, the wind measured at the first altitude level is used to estimate the surface wind below at 10 m above sea level (a.s.l.), which is the altitude of SAR data. The extrapolation to 10 m a.s.l. is performed using a classical power law:

$$U_{10} = U_{min} \cdot \left(\frac{10}{Zmin}\right)^{\alpha} \quad \text{Eq. (1)}$$

where U_{10} is the wind speed at 10 m in m s^{-1}, U_{min} the Lidar wind speed at the first altitude level in m s^{-1}, and Z_{min} the altitude of the first level in m, and α a non-dimensional exponent. Hsu et al. (1994) recommend choosing an exponent of 0.11 over the sea. We checked this hypothesis with HKZA and HKZB Lidars that were equipped with anemometers measuring wind speed at 4 m a.s.l.. This exponent was found to be correct on average. However, in order to refine the wind speed values extrapolated

at 10 m a.s.l., we adapted the exponent depending on the current atmospheric stability. The instantaneous exponents obtained with HKZA and HKZB Lidars were compared with the air-sea temperature difference provided by the high-resolution numerical model WRF. The relation was fitted with a second-degree polynomial (Figure 2) and then the adaptive exponents were used to obtain the wind speed at 10 m a.s.l.. The anemometers located at the base of the Lidars at 4 m a.s.l. do not have a high precision and may add some uncertainty, however, since the final machine learning algorithm presented in this study is trained with Lidar measurements at hub height, this uncertainty is included in our results.

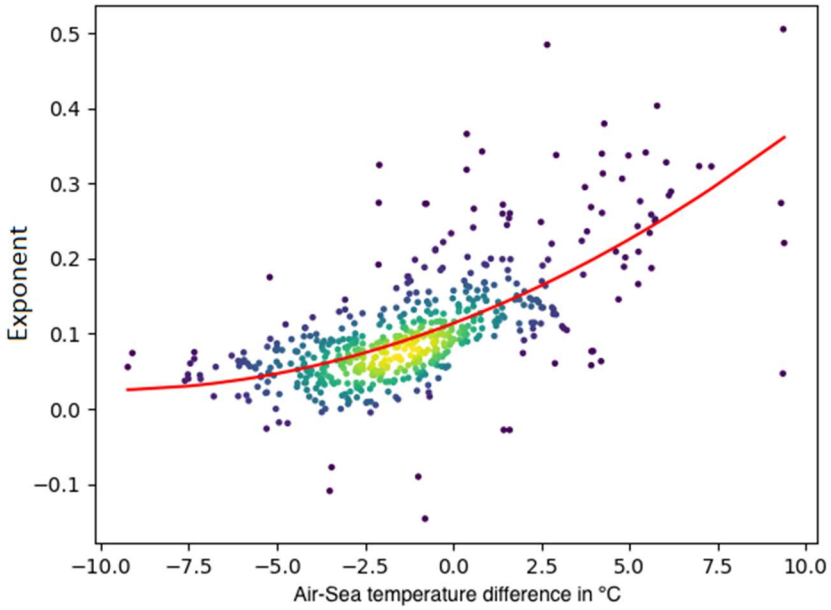

Figure 2: Exponent of the power law between the wind speeds at 4 m and 40 m as a function of the air-sea temperature difference fitted with a second-degree polynomial fit (red curve) with the following coefficients: $Y = 0.1137 + 0.0178\,X + 0.001\,X^2$. The colours represent the density of points.

2.3 Sentinel-1 SAR data

Sentinel-1 A and B are two polar-orbiting satellites equipped with C-band SAR. This sensor, which records surface roughness, has the advantage of operating day and night at wavelengths not impeded by cloud cover. The Sentinel-1 Level 1 GRD (Ground Range Detected) product has a grid spacing of a few tens of meters, whereas the Level 2 wind products typically have a resolution of 1 km. The two satellites are located on the same orbit 180° apart and at an altitude close to 700 km. In Dutch coastal waters, the acquisition mode is an Interferometric Wide (IW) swath using the TOPSAR technique, which provides a

better-quality product by enhancing the image homogeneity (De Zan and Guarnieri, 2006). The revisit rate is one passage every two days, which occurs usually in the morning around 5 AM or in the evening around 5 PM (UTC). The satellites pass in the morning or in the evening depending on the orbit orientation, descending or ascending, respectively. The exact acquisition time can vary by plus or minus 30 mn depending on the incidence angle under which the region of interest is observed. The constellation was only fully operational at the end of 2016. We collected all Sentinel-1 A and B images in IW acquisition mode for2017, 2018 and 2019 only to ensure a complete and homogenous annual sampling. The total number of samples over these three years for the areas of interest is shown in Figure 1, on which it can be seen that the coverage is not spatially uniform.

The Level 1 images were calibrated and corrected from the instrument noise provided as metadata. A dedicated bright target filtering was applied to remove Radar echoes created by ships, wind farms and other structures at sea. An additional filter (Koch, 2004) was used to identify heterogeneous signatures not related to wind, like currents, Radar interferences, and remaining bright targets. However, this filter has an increased sensitivity at low wind speeds, therefore, the identified pixels were not removed to avoid disrupting the wind speed Weibull distribution, which is necessary to estimate wind power. The information provided by this filter was only used to create maps of areas where wind power estimates are unreliable, typically due to dense regions of wind turbines or mooring areas, which are well identified on average by the heterogeneity filtering. Then, Level 1 SAR products were degraded to a 1 km resolution and Level-2 surface winds at 10 m a.s.l. were obtained using a Bayesian inversion scheme using as inputs the wind speed obtained by inverting the SAR backscatter with the CMOD5.N Geophysical Model Function (GMF) (KNMI, 2008; ECMWF, 2008; Hersbach, 2010) and the outputs of ECMWF (European Centre for Medium-Range Weather Forecasts) NWP (Numerical Weather Prediction) model to constrain the wind direction. The Level 2 product tiles were combined into a gridded map over the areas of interest, in order to form a data cube where each pixel corresponds to a time series of SAR measurements.

2.4 Wind power estimation

The average extractible wind power *P*, hereafter simply called wind power, is calculated by multiplying point-by-point the wind speed probability density function (pdf) by the power curve of a specific wind turbine. We chose to simulate an 10MW turbine with a typical power curve: the DTU 10 MW Reference Wind Turbine V1 (see DTU Wind Energy, 2017, and https://github.com/NREL/turbine-models/blob/master/Offshore/DTU_10MW_178_RWT_v1.csv, last accessed September 2, 2021). A simple histogram could be used to estimate the wind speed pdf. However, due to the low number of SAR samples, a more efficient technique consists in using the SAR data to fit a Weibull pdf, which usually describes the wind speed accurately. The Weibull pdf is given by:

$$pdf(U) = \frac{k}{\lambda}\left(\frac{U}{\lambda}\right)^{k-1} e^{-(U/\lambda)^k} \quad \text{Eq. (2)}$$

where λ is a scale parameter in m s^{-1} and k a dimensionless shape parameter. These parameters can be obtained by maximum likelihood, or by the method of the moments with the following formulas (Pavia and O'Brien, 1986):

$$k = (\sigma/\mu)^{-1.086} \quad \text{Eq. (3)}$$

$$\lambda = \frac{\mu}{\Gamma\left(\frac{1}{k}+1\right)} \quad \text{Eq. (4)}$$

where μ is the mean wind speed, σ the wind speed standard deviation, both in m s^{-1}, and Γ the Gamma function.

The accuracy of these two methods was assessed with simulations. A random variable following a Weibull law with known parameters was generated and the equivalent wind power computed. For both methods, the results were compared with the wind power computed with the original parameters. Figure 3 shows the wind power mean absolute error in percentage as a function of the number of samples. With 500 samples, which is approximately the amount of available SAR data available in the areas of interest, the accuracy of these methods is ± 2%. Both methods yield similar results, therefore the method of the moment, which is simpler and faster to run, was chosen.

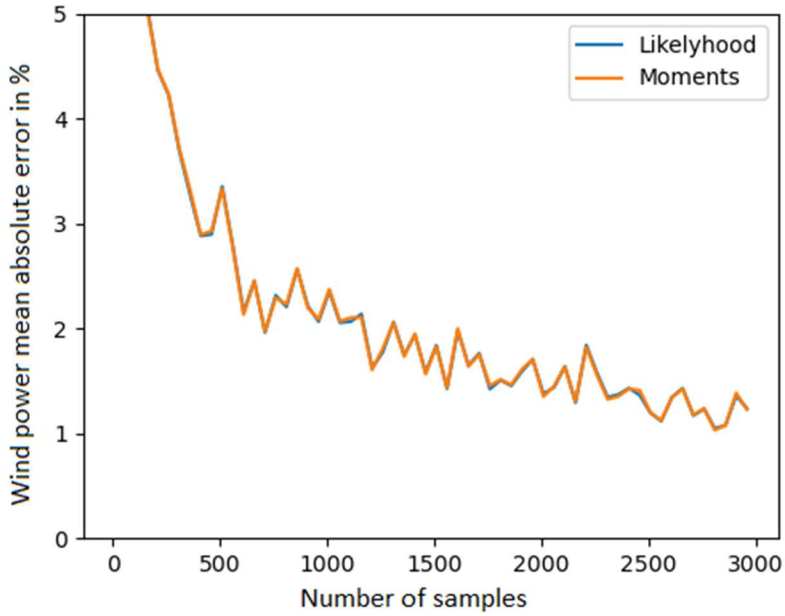

Figure 3: Wind power mean absolute error in percentage as a function of the number of samples, using maximum likelihood to fit the wind Weibull pdf (orange curve), or the method of the moments (blue curve).

2.5 Intra-diurnal variability

The main limitation of SAR satellites is their low temporal sampling (one passage every two days for Sentinel-1 in Europe). One advantage of this limitation, however, is that it guarantees the statistical independence of the measurements. Nevertheless, the satellites are on a sun-synchronous orbit, which means that they pass always at the same times of the day, in the morning or in the evening. As a result, they cannot fully see the intra-diurnal variability of the wind.

Figure 4 shows the mean wind speed at 10 m as a function of the hour of the day for each Lidar. It can be seen that the wind diurnal cycle is close to a 24 h period sinusoid. Therefore, since the satellites pass at two possible times of the day separated by 12 h, according to the Nyquist-Shannon sampling theorem, they should be able to capture the majority of the intra-day variability. In order to verify this, we simulated the satellites' passages over the Lidars by computing the mean wind speed and the wind power using only the Lidar measurements at the satellites' passage times. These values were compared to those obtained using all Lidar measurements at any time of day. For each Lidar, the differences were found to be below 1% and 2%, respectively, for the mean wind speed and the wind power. Therefore, this source of error is limited, and the intra-diurnal variability does not prevent the use of SAR data. However, this conclusion needs to be validated in geographical areas where thermic winds are stronger than in the North Sea.

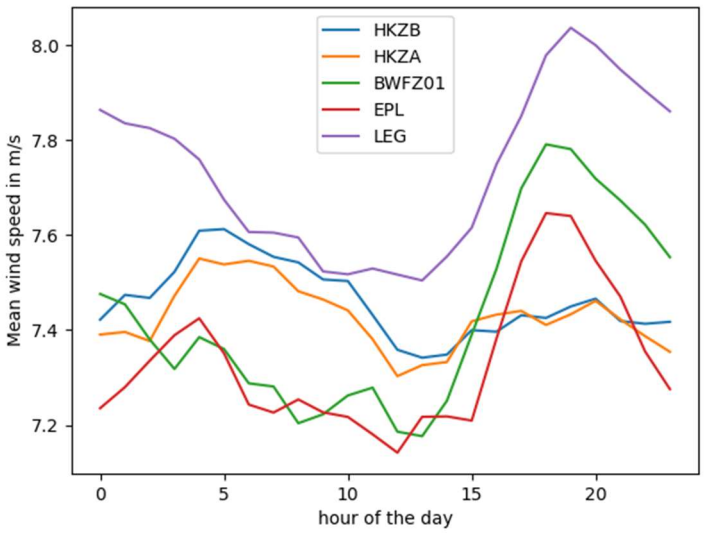

Figure 4: Intra-diurnal variability of the mean wind speed at 10 m for each Lidar. The time is given in UTC, which is close to the local time since Zone 1 and Zone 2 are located near Greenwich meridian.

3 Improvement of SAR surface wind speeds and extrapolation at hub height

3.1 Machine learning at surface level

SAR surface winds are obtained by inverting the backscatter over a given pixel with a GMF (see Section 2.3). Originally, GMFs were designed to retrieve the wind from spaceborne scatterometers. However, SAR is a specific sensor and differences between the SAR and scatterometers backscatter in C-band may occur. In addition, GMFs were empirically designed using global numerical models as a reference, but they are not as reliable as real data, especially in coastal areas. Moreover, GMFs may not fully capture the complex relation between the sea state and the wind, in particular because they assume a neutral atmosphere. As a result, SAR surface winds are typically biased when compared with in-situ buoys (see, e.g., de Montera et al., 2020). Therefore, it is necessary to improve the accuracy of SAR wind speeds obtained with a GMF. This is particularly important because the wind power is related to the cube of the wind speed, and therefore very sensitive to wind speed estimation errors.

Given the complex relation between the sea state and the wind speed, and the number of factors able to influence it, machine learning was found to be an appropriate technique to improve the accuracy of SAR surface wind speeds and remove their biases. We used a Random Forest algorithm (Breiman, 2001), which is known to perform well in regression tasks. It was implemented with the RandomForestRegressor function of Scikit-learn Python toolbox and its architecture was chosen by using cross-validation. The default hyperparameters were found to be the most appropriate ones, except the number of trees set to 240 and the maximum depth set to 20. The algorithm was trained with the wind measured by the Lidars extrapolated to 10 m (the first Lidar level was extrapolated to this altitude with a power law, see Section 2.2). Combining all measurement sites, more than 1000 collocated data points between the Lidars and Sentinel-1 SAR could be found. The algorithm was trained with 50% of the data points randomly chosen, and the rest of them were used as a test dataset.

In order to select the input parameters, we made a list of interesting parameters and looked for the ones related to the differences between SAR and Lidars. This was done visually by plotting scatterplots of these parameters against the errors of the SAR compared to Lidar measurements. The following parameters were selected: the SAR surface wind, the SAR wind direction, the azimuth angle (i.e., the angle between the North and the satellite track), the incidence angle (i.e., the angle between the radar illumination and the zenith of the target), the elevation angle (i.e., the angle between the radar illumination and the nadir of the satellite), the backscatter, the thermal noise of the instrument, and the difference between the azimuth angle and the wind direction (an important parameter in the inversion of the backscatter). Since the SAR surface winds are given assuming a neutral atmosphere, the atmospheric stability also needs to be taken into account. Therefore, air-sea temperature difference and surface heat flux were extracted from the high-resolution numerical model and added as input parameters. The relative importance of these parameters was measured after the training stage using the feature_importances_ attribute of Scikit-learn Python toolbox (Figure 5).

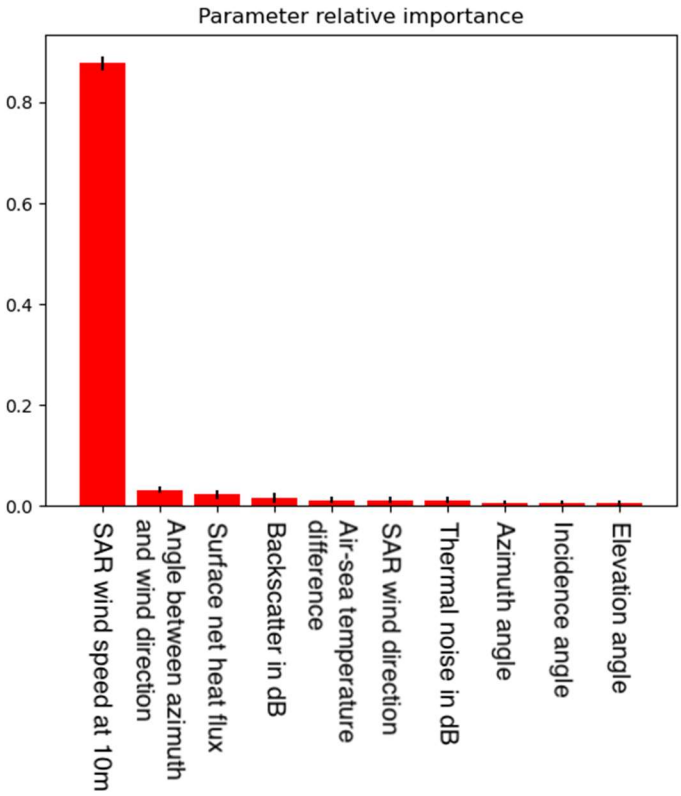

Figure 5: Relative importance of the input parameters used to correct the SAR surface wind at 10m with the Random Forest algorithm.

The Random Forest algorithm was able to reduce the SAR wind speed bias from -0.42 m s^{-1} to 0.02 m s^{-1} and its standard deviation from 1.41 m s^{-1} to 0.98 m s^{-1}. Figures 6 shows the scatterplots of the SAR wind speeds against Lidars measurements before and after applying machine learning. It can be seen that the bias is indeed reduced and that the cloud of points is thinner after machine learning. However, the resulting wind speeds are still biased at very low and very high wind speeds. These two ranges are more difficult to estimate because low wind speeds have little effect on the sea state, and because the relation between the sea state and the backscatter saturates at high wind speeds. A multi-expert algorithm using three different Random Forest algorithms to process respectively low, middle and high wind speeds was tested. However, this approach did not improve the results.

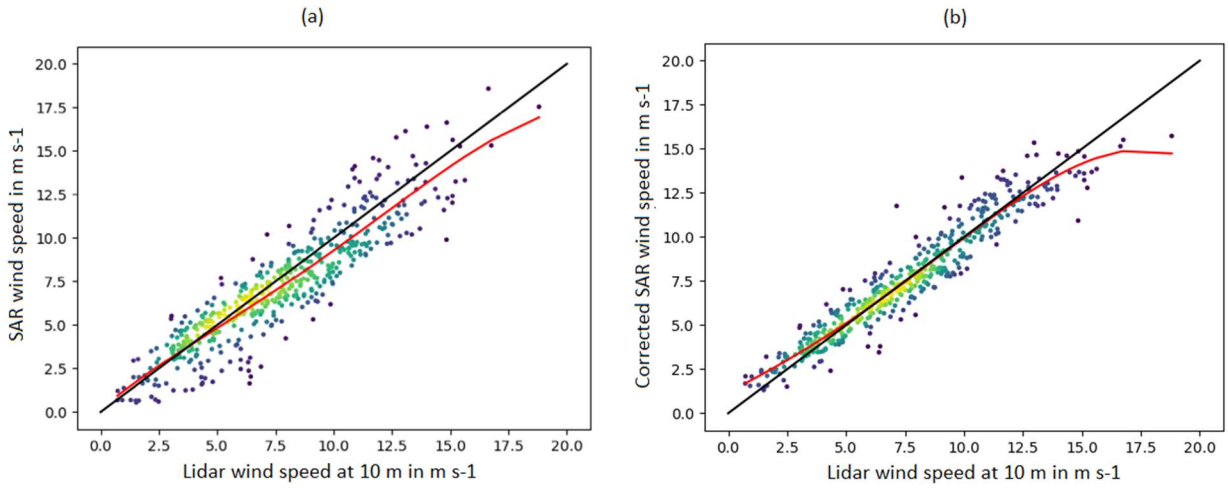

Figure 6: Scatterplots between the SAR and Lidar wind speeds at 10 m before machine learning (a) and after machine learning (b) using the test dataset. The colours represent the density of points. The black curve is the identity line and the red curve a fourth-degree polynomial fit illustrating the bias.

275

3.2 Extrapolation at hub height

The instantaneous SAR wind speeds were also extrapolated using a Random Forest algorithm. This algorithm uses parameters extracted from the high-resolution numerical model as input parameters. The most relevant parameters were found to be the air-sea temperature difference and the surface heat flux. In order to increase the accuracy and adapt to the current
280 meteorological conditions, the model extrapolation ratio between the surface wind and the wind at hub height was also added. However, comparisons to Lidar measurements showed that the numerical model outputs were less accurate in the lower boundary layer, and that the mean wind speed was strongly biased below 40 m. Therefore, we decided to use the ratio between the wind speed at 40 m and higher altitudes, which was more accurate (for each Lidar, the ratio bias was lower than 1%). These parameters were used together with the corrected SAR wind speeds at 10 m as input to the Random Forest algorithm, which
285 was trained to learn the Lidar wind speed at several altitude levels until 200 m using the same training dataset as previously. The algorithm was also implemented with the RandomForestRegressor function of Scikit-learn Python toolbox. We used the default hyperparameters, except the number of trees set to 340, the maximum depth set to 50 and the maximum number of features set to 'sqrt'. The relative importance of the parameters after the training phase is shown in Figure 7.

Figure 8 shows the bias of the extrapolated SAR wind speeds against each Lidar in percentage as a function of the altitude.
290 The algorithm was successful in extrapolating SAR wind speeds because these biases are stable with altitude and remain low (comprised within ± 3%). At 200 m, the SAR wind speed bias against all Lidars was -0.04 m s^{-1} and its standard deviation 1.69 m s^{-1}. Thus, this method provides an almost unbiased estimate of the wind speed at hub height. We also attempted to follow the same approach as Badger et al. (2016), in which the extrapolation is performed on wind statistics. However, the

extrapolation of the wind power with the corresponding ratio provided by the numerical model was not as accurate as when the instantaneous winds were extrapolated first.

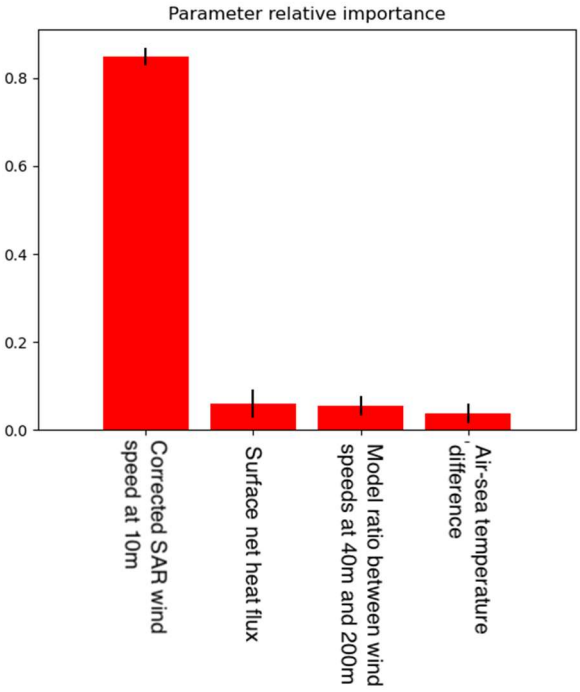

Figure 7: Relative importance of the input parameters used to extrapolate SAR surface winds to 200 m.

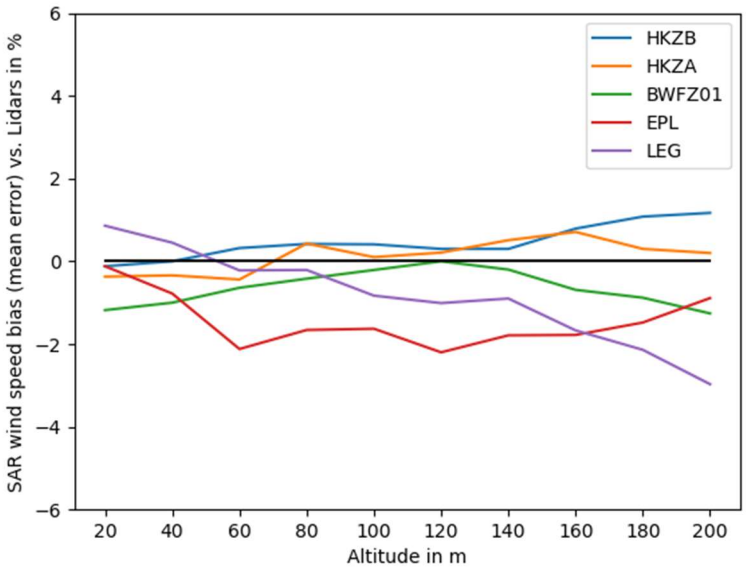

Figure 8: Bias of the extrapolated SAR wind speed against each Lidar in percentage.

4 Results

4.1 Correction of the standard deviation

As explained in Section 2.4, in order to estimate the extractible wind power, the Weibull parameters of the wind are needed. These parameters are directly linked to the first two moments of the wind speed pdf, which are the mean wind speed and the wind speed standard deviation (see Eqs. 3 and 4). Therefore, an accurate estimation of these two moments alone is enough to guarantee a low error of the extractible wind power. The extrapolation method presented above provide an unbiased estimation of the mean wind speed. However, the wind speed standard deviation was found to be biased. This occurs because machine learning estimates the expected value of the wind speed, which tends to reduce the variability of the wind speed: it squeezes its distribution because the errors are not reproduced. As a consequence, the original distribution of the data is not conserved. Its tails are lighter, and the standard deviation is underestimated. Figure 9 shows an example of the wind speed distribution obtained at 200 m after machine learning compared to the one obtained with Lidar HKZA. In that case, the error of the wind speed standard deviation was - 9%.

In order to compensate for this effect, we reintroduced artificially the original variability of the data. This was done by analysing the distribution of the SAR wind speed errors compared to Lidar measurements and adding a similar random variable to the SAR wind speed obtained after machine learning. The appropriate random variable was found to be a Gaussian with the

standard deviation of the SAR wind speed errors. For each data point, at least five additional artificial datapoints needed to be created for the wind speed standard deviation to converge. After this bootstrap, the wind speed standard deviation error was 1.5% when considering all Lidars together in the test dataset. Thus, the result of this correction is an almost unbiased estimation of the wind speed standard deviation.

Figure 9 also shows the wind speed distribution provided by the high-resolution numerical model. In this case, the distribution was shifted to the left, which means that the numerical model underestimates the wind speed compared to Lidar measurements. The wind speed bias was found to be - 4%. It was corrected by adding the opposite quantity to the model outputs.

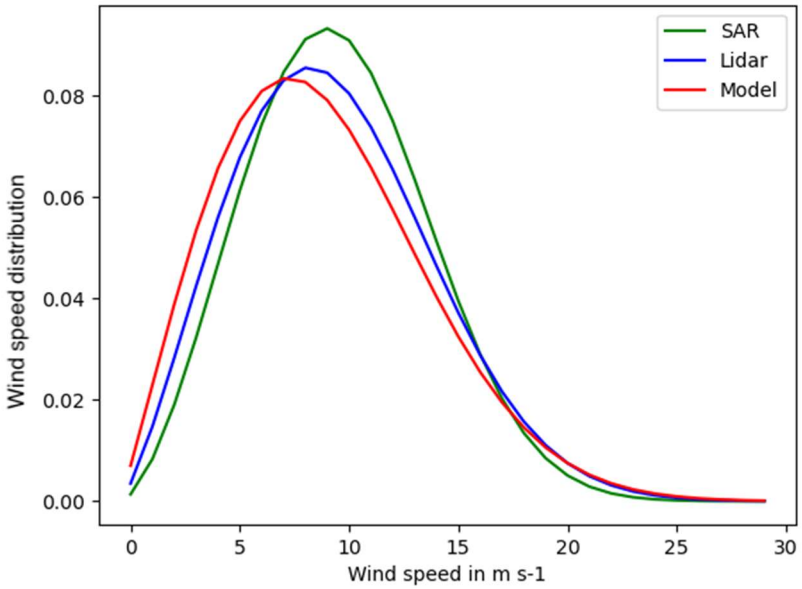

Figure 9: Wind speed distribution at 200m measured by Lidar HKZA compared to the ones obtained with SAR data combined with machine learning and the numerical model at the same location.

4.2 Extractible wind power at hub height

Once the wind speed standard deviation had been corrected for its bias, the estimation of the extractible power was done with the method presented in Section 2.4. The Weibull parameters were obtained with Eqs. 3 and 4, and then the wind power was obtained by multiplying the Weibull distribution (Eq. 2) by the typical 10 MW turbine power curve. Figures 10 shows the bias of the wind power between the SAR and the Lidars in percentage. The method accuracy is ± 3%. However, this result needs to be confirmed in other geographical locations than the North Sea. In areas where the machine learning algorithm trained in the North Sea would not be appropriate due to very different wind patterns and where Lidar measurements are not available, a simpler method could be used. We tested using directly the extrapolation ratio given by the high-resolution numerical model,

without machine learning. In that case, the wind power error was comprised between ± 7% for each Lidar, which is accurate enough to provide useful insights about the coastal wind gradient.

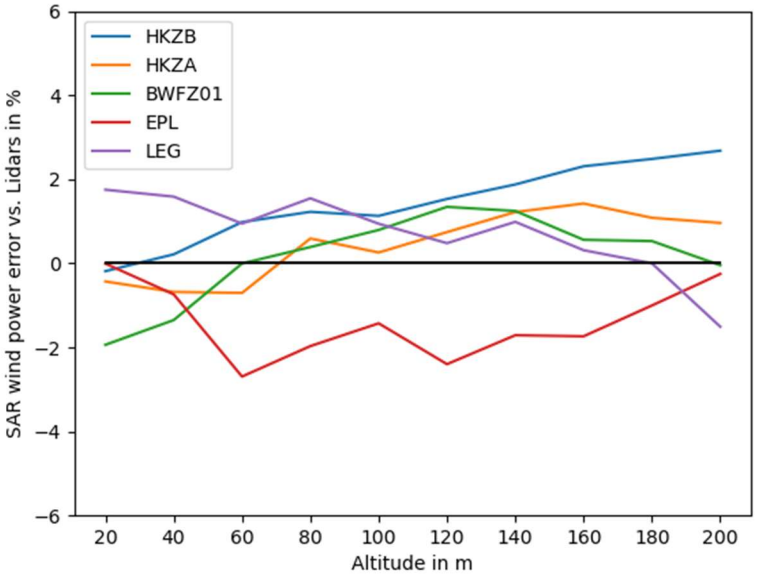

Figure 10: SAR wind power error in % compared to the one computed with Lidars measurements.

4.3 Wind power maps at hub height

Figures 11 and 12 show the extractible wind power maps at 200 m over Zone 1 and Zone 2 for the typical 10MW turbine. They present the wind power predicted by the numerical model, the wind power obtained with SAR data and machine learning, and the difference between these two in percentage. It can be seen that the use of SAR data significantly increases the level of detail compared to the numerical model outputs. The difference can reach as much as 10% of the wind power between sites separated by less than 20 km.

Some artefacts are still visible on the maps and need to be corrected in the future. For example, the swath edges can still be seen. Moreover, in some areas, the estimation was less reliable, mainly due to bright targets that could not be filtered. Some of these areas are linked to existing wind farms having a high density of turbines, while other areas had large numbers of stationary shipping vessels. The presence of these artefacts was measured by a Koch filter and a quality flag was created. Figures 11 and 12 also show the percentage of SAR data flagged as 'low quality', and therefore the areas where the assessment with SAR satellites is unreliable. In addition, in Zone 1, a series of three unrealistic 'waves' can be seen close to the coast. We could check that these patterns correspond to similar 'waves' of sand in the seabed. The bathymetry in these shallow waters seems to affect currents and, therefore, the SAR backscatter.

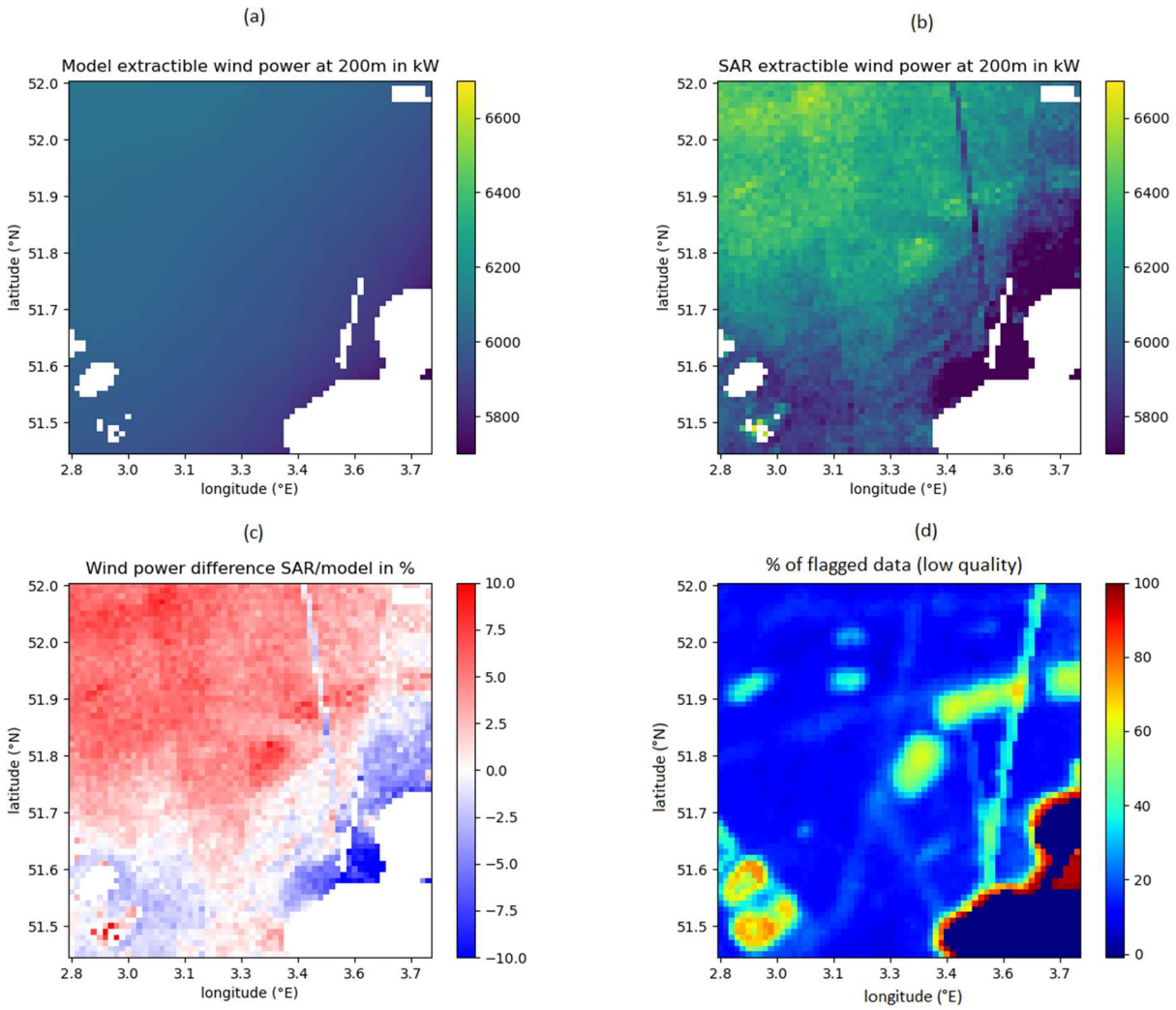

Figure 11: Extractible wind power over Zone 1 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).

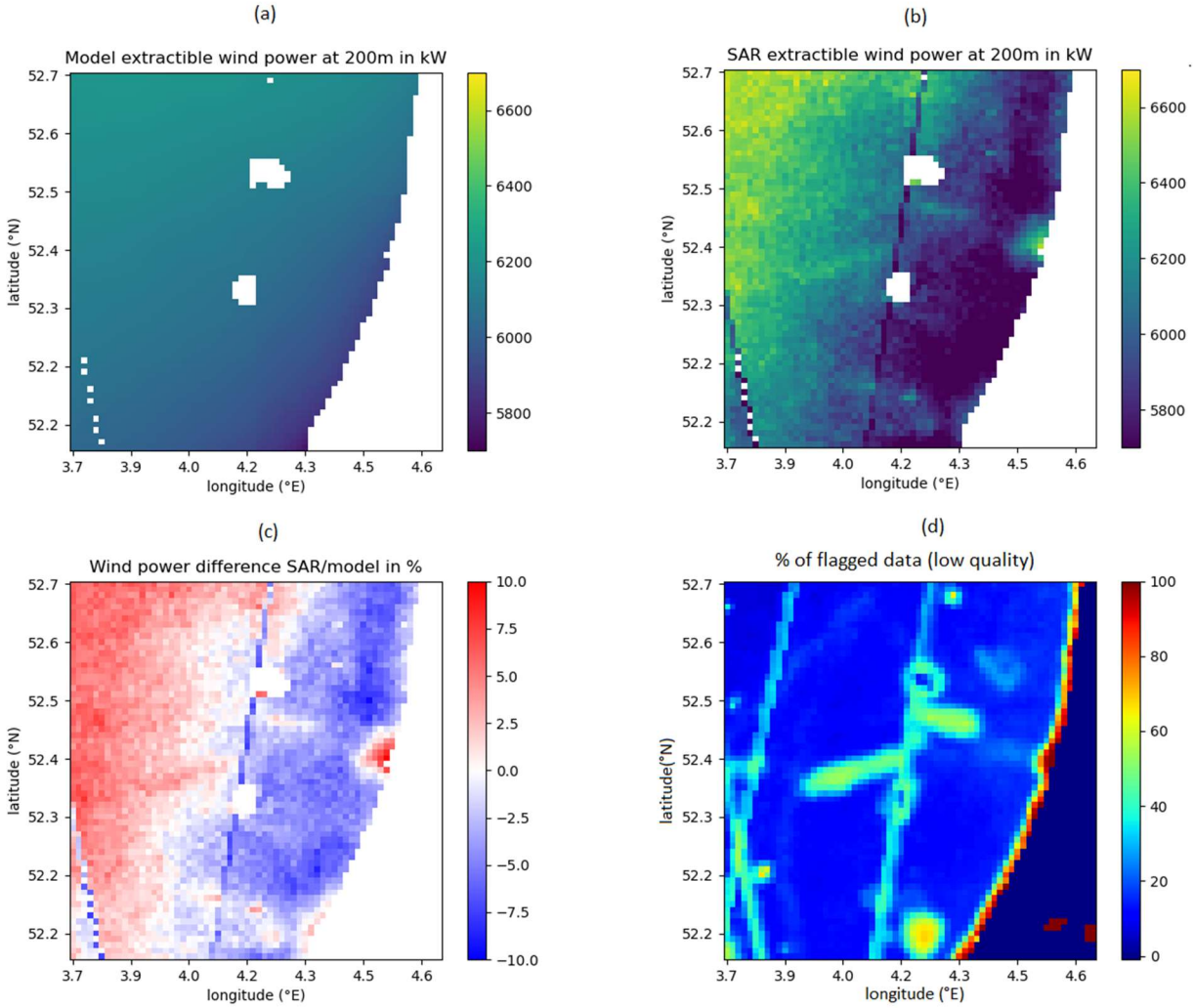

Figure 12: Extractible wind power over Zone 2 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).

365

5 Conclusion

A new method for estimating the offshore extractible wind power at turbine hub height based on SAR data and machine learning has been presented. The method has two steps: first correcting SAR surface wind speeds with a machine learning
370 algorithm fed with geometrical parameters of SAR sensor and meteorological parameters extracted from a high-resolution

numerical model, then extrapolating these winds to higher altitudes with another machine learning algorithm. The method was tested in two areas off the Dutch coast using data from 5 Doppler wind Lidars installed on the sea surface. The wind power maps were computed assuming a typical 10 MW turbine power curve. At 200m a.s.l., the accuracy of the method was 3% for both the wind speed and the wind power. One must add the error due to intra-diurnal variability, which is not seen by the satellites. This source of error was estimated to be less than 1% for mean wind speed and 2% for wind power. Note that this additional uncertainty could be easily removed in the future by simulating the passages of SAR satellites using the high-resolution numerical model, and precisely estimating the effect of their time sampling. In the areas affected by the coastal gradient, the difference between the SAR wind power maps and the outputs of the numerical model can reach 10% over short distances of less than 20 km. Compared to the maps provided by the numerical model, this method has the advantage of providing a much higher level of details. Therefore, using SAR data combined with a high-resolution numerical model and processing them with machine learning can improve the assessment of the offshore wind resource. It can provide useful insights to optimize wind farm siting and risk management.

Further research should focus on removing some artefacts remaining on the SAR wind power maps, such as the swath edges, bright targets, and the effect of bathymetry. The method could also be improved by identifying other useful input parameters for machine learning, like the cross-polarization backscatter, which is more sensitive to strong winds. Finally, the method needs to be generalized to other geographical areas and trained with a larger training dataset that could combine Lidars and classical metocean buoys measurements.

Author contribution

Louis de Montera designed the algorithm and wrote the paper, Henrick Berger processed the SAR raw data and created a Level 2 gridded wind product. Romain Husson provided his expertise on SAR satellite and wind measurement from space. Pascal Appelghem parametrized the WRF model and performed the runs. Laurent Guerlou and Mauricio Fragoso supervised the study, organised the funding, and gathered together the project team.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank ESA and the Dutch Ministry of Economic Affairs and Climate Policy for providing freely Sentinel-1 data for providing the Lidar data, respectively. We would also like to thank Rémi Gandoin from C2WIND for checking the quality of the Lidar data, Cynthia Johnson for correcting the spelling and improving the article style, and two anonymous referees for their contributions. This research was funded by the French Space Agency (Centre National d'Etudes Spatiales).

References

Ahsbahs, T., Badger, M., Karagali, I. and Guo Larsén, X.: Validation of Sentinel-1A SAR coastal wind speeds against scanning LiDAR. Remote sensing, 9(6), doi: 10.3390/rs906055, 2017.

Ahsbahs, T., MacLaurin, G., Draxl, C., Jackson, C., Monaldo, F. and Badger, M.: US East Coast synthetic aperture radar wind atlas for offshore wind energy. Wind Energy Science, 5, 1191-1210. doi: 10.5194/wes-5-1191-2020, 2020.

Badger, M., Peña, A., Hahmann, A. N., Mouche, A. A. and Hasager, C. B.: Extrapolating Satellite Winds to Turbine Operating Heights, J. Applied Meteorology and Climatology, 55(4), 975-991, doi: 10.1175/JAMC-D-15-0197.1, 2016.

Badger, M., Ahsbahs, T. T., Maule, P. and Karagali, I.: Inter-calibration of SAR data series for offshore wind resource assessment. Remote Sensing of Environment, 232, 111316, doi: 10.1016/j.rse.2019.111316, 2019.

Bentamy, A. and Croize-Fillon, D.: Spatial and temporal characteristics of wind and wind power off the coasts of Brittany. Renewable energy, 66, 670-679, doi: 10.1016/j.renene.2014.01.012, 2014.

Breiman, L.: Random Forests, Machine Learning, 45(1), 5-32, 2001, doi: 10.1023/A:1010933404324

Bodini, N. and Optis, M.: The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds. Wind Energy Science, 5(2), 489–501, doi:10.5194/wes-5-489-2020, 2020.

Chang, R., Zhu, R., Badger, M., Hasager, C.B., Zhou, R., Ye, D. and Zhang, X.: Applicability of synthetic aperture radar wind retrievals on offshore wind resources assessment in Hangzhou bay, China. Energies, 7, 3339-3354. doi: 10.3390/en7053339, 2014.

Chang, R., Zhu, R., Badger, M., Hasager, C.B., Xing, X. and Jiang, Y.: Offshore wind resources assessment from multiple satellite data and WRF modeling over South China sea. Remote Sensing, 7, 467-487, doi: 10.3390/rs70100467, 2015.

Christiansen, M. B., Koch, W., Horstmann, J., Hasager, C. B. and Nielsen, M.: Wind resource assessment from C-band SAR. Remote Sensing of Environment, 105, 68-81, doi: 10.1016/j.rse.2006.06.005, 2006.

De Zan, F. and Guarnieri, A. M.: TOPSAR: Terrain Observation by Progressive Scans. IEEE Trans. on Geoscience and Remote Sensing, 44(9), 2352-2360, doi: 10.1109/TGRS.2006.873853, 2006.

DTU Wind Energy. HAWC2 Model for the DTU 10-MW Reference Wind Turbine. 2017. https://www.hawc2.dk/Download/HAWC2-Model/DTU-10-MW-Reference-Wind-Turbine, last accessed September 2, 2021.

ECMWF. CMOD5.N: A C-band geophysical model function for equivalent neutral wind. Hersbach, H., 2008, https://www.ecmwf.int/en/elibrary/9873-cmod5n-c-band-geophysical-model-function-equivalent-neutral-wind, last accessed 24 March 2021, 2008.

Grachev, A. A. and Fairall, C. W.: Dependence of the Monin-Obukhov stability parameter on the bulk Richardson number over the ocean, J. Applied Meteorology, 36, 406–414, doi: 10.1175/1520-0450(1997)036<0406:DOTMOS>2.0.CO;2, 1996.

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., & Söderberg, S.: The making of the New European Wind Atlas – Part 1: Model sensitivity. Geoscientific Model Development, 13(10), 5053-5078, doi: 10.5194/gmd-13-5053-2020, 2020.

Hasager, C.B., Frank, H.P. and Furevik, B.R.: On offshore wind energy mapping using satellite SAR. Canadian Journal of Remote Sensing, 28(1), 80-89, doi: 10.5589/m02-008, 2002.

Hasager, C.B., Nielsen, M., Astrup, P., Barthelmie, R., Dellwik, E., Jensen, N.O., Jørgensen, B.H., Pryor, S.C., Rathmann, O. and Furevik, B.R.: Offshore wind resource estimation from satellite SAR wind field maps. Wind energy, 8, 403-419, doi: 10.1002/we.150, 2005.

Hasager, C.B., Barthelmie, R.J., Christiansen, M.B., Nielsen, M. and Pryor, S.C.: Quantifying offshore wind resources from satellite wind maps: study area the North Sea. Wind energy, 9, 63-74. doi: 10.1002/we.190, 2006.

Hasager, C.B., Badger, M., Peña, A., Larsén, X.G. and Bingöl, F.: SAR-Based Wind Resource Statistics in the Baltic Sea. Remote Sensing, 3, 117-144, doi: 10.3390/rs3010117, 2011.

Hasager, C. B., Mouche, A., Badger, M., Bingöl, F., Karagali, I., Driesenaar, T., Stoffelen, A., Peña, A. and Longépé, N.: Offshore wind climatology based on synergetic use of Envisat ASAR, ASCAT and QuikSCAT. Remote sensing of environment, 156, 247–263, doi: 10.1016/j.rse.2014.09.030, 2015.

Hasager, C.B., Hahmann, A.N., Ahsbahs, T., Karagali, I., Sile, T., Badger, M. and Mann, J.: Europe's offshore winds assessed with synthetic aperture radar, ASCAT and WRF, Wind Energy Science, 5(1), 375-390, doi: 10.5194/wes-5-375-2020, 2020.

Hersbach, H.: Comparison of C-band scatterometer CMOD5.N equivalent neural winds with ECMWF. J. Atmospheric Ocean Technology, 27(4), 721-736, 2010, doi: 10.1175/2009JTECHO698.1, 2010.

Hsu, S. A., Meindl, E. A. and Gilhousen, D. B.: Determining the power-law wind-profile exponent under near-neutral stability conditions at Sea, Journal of Applied Meteorology and Climatology, 33(6), 757-765, doi: 10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2, 1994.

Karagali, I., Peña, A., Badger, M. and Hasager, C. B.: Wind characteristics in the North and Baltic Seas from the QuikSCAT satellite. Wind energy, 17(1), 123-140. doi:10.1002/we.1565, 2014.

KNMI. CMOD5.n—The CMOD5 GMF for neutral winds. Verhoef, A., Portabella, M., Stoffelen, A. and Hersbach, H., 2008, http://projects.knmi.nl/publications/fulltexts/cmod5_neutral_winds_1.0.pdf, last accessed 22 March 2021.

Koch, W.: Directional Analysis of SAR Images Aiming at Wind Direction. IEEE Transactions on Geoscience and Remote Sensing, 42(4), 702–710, doi: 10.1109/TGRS.2003.818811, 2004.

Mohandes, M. A. and Rehman, S.: Wind speed extrapolation using machine learning methods and LiDAR measurements. IEEE Access, 6, 77634-77642, doi: 10.1109/ACCESS.2018.2883677, 2018.

de Montera, L., Remmers, T., O'Connell, R. and Desmond, C.: Validation of Sentinel-1 offshore winds and average wind power estimation around Ireland, Wind Energy Science, 5(3), 1023-1036, doi: 10.5194/wes-5-1023-2020, 2020.

NREL. Best Practices for the Validation of U.S. Offshore Wind Resource Models. Optis, M., Bodini, N., Debnath, M., and Doubrawa, P., 2020, https://www.nrel.gov/docs/fy21osti/78375.pdf, last accessed 24 August 2021.

Optis, M., Bodini, N., Debnath, M., and Doubrawa, P.: New methods to improve the vertical extrapolation of near-surface offshore wind speeds, Wind Energ. Sci. Discuss. [preprint], doi: 10.5194/wes-2021-5, in review, 2021.

Pavia, E. G. and O'Brien, J. J.: Weibull Statistics of Wind Speed over the Ocean. Proceedings of the Royal Society of Mathematical, Physical and Engineering Sciences., 25(10), 1324-1332, doi: 10.1175/1520-0450(1986)025<1324:WSOWSO>2.0.CO;2 , 1986.

Pimenta, F., Kempton, W. and Garvine, R.: Combining meteorological stations and satellite data to evaluate the offshore wind power resource of Southeastern Brazil. Renewable energy, 33(11), 2375-2387, doi: 10.1016/j.renene.2008.01.012, 2008.

Remmers, T., Cawkwell, F., Desmond, C., Murphy, J., and Politi, E.: The Potential of Advanced Scatterometer (ASCAT) 12.5 km Coastal Observations for Offshore Wind Farm Site Selection in Irish Waters. Energies, 12(2), 206, doi: 10.3390/en12020206, 2019.

Sánchez, R. F., Relvas, P., and Pires, H. O.: Comparisons of ocean scatterometer and anemometer winds off the southwestern Iberian Peninsula. Continental shelf research, 27(2), 155-175, doi: 10.1016/j.csr.2006.09.007, 2007.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang: A Description of the Advanced Research WRF Version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., doi:10.5065/1dfh-6p97, 2019.

Türkan, Y. S., Yumurtacı Aydoğmuş, H. and Erdal, H.: The prediction of the wind speed at different heights by machine learning methods. Inter. J. of Optimization and Control: Theories & Applications, 6(2), 179–187. Doi: 10.11121/ijocta.01.2016.00315, 2016.

Vassallo, D., Raghavendra K., and Harindra J. S. F.: Decreasing wind speed extrapolation error via domain-specific feature extraction and selection. Wind Energy Science, 5(3), doi: 10.5194/wes-5-959-2020, 2020.