Quarterly Journal of the
Royal Meteorological Society   RMetS

Quality of wind characteristics in recent wind atlases over the North Sea

Peter C. Kalverla | Albert A. M. Holtslag | Reinder J. Ronda | Gert-Jan Steeneveld

Meteorology and Air Quality Section, Wageningen University, Wageningen, The Netherlands

Correspondence
G.-J. Steeneveld, Meteorology and Air Quality Section, Wageningen University, PO Box 47, 6700AA, Wageningen, The Netherlands.
Email: gert-jan.steeneveld@wur.nl

Abstract

Offshore wind energy production is rapidly growing as an essential element in the sustainable energy share. Wind energy siting studies require accurate wind data, and in particular the knowledge of extreme wind events (low-level jets, wind ramps, extreme shear and high wind speeds) is crucial for resource and load assessment. This study evaluates the skill of three relatively new wind atlases, i.e. ERA-5, DOWA and NEWA on the representation of extreme wind events using observations taken at the Met Mast IJmuiden over the North Sea. Overall, DOWA appears to best represent the wind speed profile with virtually no bias. ERA-5 underestimates the mean wind speed profile though the wind shear is well represented, while NEWA correctly represents the near surface wind but underestimates the wind shear. The frequency of low-level jets are also best represented by DOWA. Wind speed ramps and direction ramps are best represented by ERA-5, while DOWA appears to outperform the others concerning wind shear.

KEYWORDS

DOWA, ERA-5, low-level jets, Meteo Mast IJmuiden (MMIJ), NEWA, North Sea, offshore wind, wind energy

1 | INTRODUCTION

Ongoing climate change and the demand for more sustainable energy production have raised the interest in wind energy resources. For instance, the Netherlands generated 17.0 PJ of electrical energy via wind energy in 2011 (onshore and offshore) and the generation has increased to 35.7 PJ in 2018. Offshore wind energy was responsible for ~10 PJ of the increased energy production. Hence, it is clear that insight into the wind characteristics over potential wind park sites is crucial; not only mean wind characteristics, but also the special dynamics of the wind, are crucial for resource and load assessment studies, i.e., low-level jets, wind ramps, extreme wind shear, etc. (Smedman *et al.*, 1996). However, offshore wind observations are usually relatively scarce, especially at hub heights. Hitherto wind atlases have been important sources of wind information (e.g., Olauson, 2018), either as a direct source or to drive small-scale models for wind energy purposes (e.g., Witha *et al.*, 2019a). Wind atlases build upon the data assimilation technique, that is, determining the most probable atmospheric state that is con-

sistent with both theory and observations, by merging numerical weather prediction (NWP) model fields and observations.

Compared to mere observations, wind atlases offer a better spatial coverage and usually a longer time frame. Therefore, wind atlases have become crucial for wind energy applications (Olauson, 2018). However, NWP models are used to "fill in the gaps", and since these models are fundamentally limited in their representation of physical processes, wind atlases are subject to uncertainty as well. In fact, as argued by Parker (2016), the lack of uncertainty information may be their largest weakness. Cross-validation with observations that were not assimilated into a wind atlas may provide an intuitive means to appreciate its value for practical purposes.

This paper evaluates three relatively new wind atlases, that is,

(a) ERA-5 (C3S, 2017), a global reanalysis dataset produced by the European Centre for Medium-range Weather Forecasts (ECMWF) using their Integrated Forecasting System (IFS),
(b) DOWA, the Dutch Offshore Wind Atlas produced by the Dutch national weather service KNMI using their regional NWP model HARMONIE (Bengtsson *et al.*, 2017), and
(c) NEWA, the New European Wind Atlas (Petersen *et al.*, 2014; Dörenkämper *et al.*, 2020; Hahmann, 2020), produced by a consortium of European research institutes using the community Weather Research and Forecasting model (WRF; Powers *et al.*, 2017) for a multitude of partly spatially overlapping domains together covering the EU and Norway/Switzerland.

Kalverla *et al.* (2019b) evaluated the performance of these three NWP models in operational forecast mode. Although (short) forecasts and reanalyses are not the same, systematic biases may point out model weaknesses noticeable in both products. All models tended to underestimate the wind speed by up to $0.5\,\mathrm{m\cdot s^{-1}}$, with a typical root mean square error of up to $2\,\mathrm{m\cdot s^{-1}}$. Stable boundary layers proved to be challenging conditions, despite recent efforts to improve the turbulent mixing formulation for these conditions (e.g. Tastula *et al.*, 2012; Sandu *et al.*, 2013; Valkonen *et al.*, 2014).

Moreover, Kalverla *et al.* (2019a) extensively compared the ERA-5 wind speed data against observations at multiple sites over the North Sea and found that the overall representation of wind speed was quite good, with a maximum root mean square error of $1.5\,\mathrm{m\cdot s^{-1}}$. The superior performance of ERA-5 as compared to short

forecasts in Kalverla *et al.* (2019b) is likely due to the data assimilation. As compared to the single location and relatively small number of forecasts (30 days) evaluated in Kalverla *et al.* (2019b), this provided more significant results.

To enable a climatological description of local wind structures, Kalverla *et al.* (2017) introduced methods to systematically study various anomalous wind events. An anomalous event describes one type of local structure, for example, the presence of a wind speed maximum in the wind speed profile (a low-level jet, LLJ) and the corresponding fall-off (the difference between the maximum wind speed and the subsequent wind speed minimum aloft), the difference in wind speed or direction between two neighbouring vertical levels (wind shear, wind veer) or between consecutive time slots (i.e., wind ramps). Even in the absence of a characteristic local structure, a wind event can be anomalous just because it is rare. Therefore, wind extremes – strong wind speeds with long return periods – were also included in Kalverla *et al.* (2017). In their validation of ERA-5, Kalverla *et al.* (2019a) focused on one of these events: the LLJ. They showed that the representation of LLJs in ERA-5 was mediocre: one-to-one correspondence was poor, the LLJs seemed to be vertically displaced (too high) and their magnitude underestimated, but the climatological frequency representation of LLJ characteristics was reasonable.

In this study, we present a first evaluation of wind and anomalous wind events in the aforementioned three wind atlases against observations from a prospective wind farm site in the North Sea, 85 km off the Dutch coast – met mast IJmuiden (MMIJ). MMIJ is located far enough from the coast and spans a long enough period of time to show reasonable agreement with the ERA-5 data (Kalverla *et al.*, 2019a). The relatively long time span of the MMIJ dataset allows for reliable statistics (better than other platforms). However, MMIJ will only partly reflect effects of small-scale coastal processes, and may not completely obviate validation with other, near-shore observations. Further validation of DOWA against observations from other sites and with satellite data are reported in Duncan *et al.* (2019a; 2019b). Initial validation and sensitivity studies that were performed for the NEWA may be found in Witha *et al.* (2019b).

Section 2 briefly describes the datasets and the procedure to align the data spatiotemporally. Then Sections 3 and 4 present a general model evaluation for wind speed and direction, followed by Section 5 which concerns the evaluation of anomalous events. Section 6 features a new spatial climatology of LLJs based on DOWA. Conclusions and perspectives are discussed in Section 7.

TABLE 1 Summary of the most important characteristics of the three wind atlases and references to further documentation

	ERA-5	DOWA	NEWA
Full name	ECMWF Retrospective Analysis 5th generation	Dutch Offshore Wind Atlas	New European Wind Atlas
Time span	1950–present	2008–2017	1989–2018
Horizontal resolution	31 km	2.5 km	27, 9, 3 km (1-way nests)
Vertical resolution	137 levels up to 0.01 hPa.	65 levels up to 10 hPa	61 levels up to 50 hPa
Hydostatic assumption	Yes	No	No
Domain	Global	The Netherlands, Southern North Sea, Wadden Sea	Outer domain Europe, with 10 regional sub-domains
Data assimilation	4D-Var with 12 hr assimilation windows	3 hr 3D-Var	—
Data sources	Satellites and *insitu* observations (see documentation)	Routine observations, ASCAT (satellite), MODE-S EHS (aircraft)	—
Initialization	From analysis	From previous forecast with additional data assimilation	ERA-5
Strategy	Hourly surface analysis (2D optimal interpolation)	Concatenated 3 hr forecasts.	8-day forecasts including 24 hr spin-up
Lateral boundaries	—	ERA-5	ERA-5
Nudging	—	—	Spectral nudging in outer domain, above PBL and level 20
Uncertainty	10-member ensemble at 62 km resolution	Uncertainty assessment report	Sensitivity experiments and multi-physics ensemble
Model	IFS, Cycle 41r2	HARMONIE-AROME, Cycle 40h1.2.tg2	WRF V3.8.1 (with modifications)
Website	https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation	https://www.dutchoffshorewindatlas.nl	https://map.neweuropeanwindatlas.eu
Reference/documentation	ECMWF (2016); C3S (2017)	Bengtsson *et al.* (2017)	Gonzalez Rouco *et al.* (2019); Witha *et al.* (2019a)

Note: Websites accessed on 21 January 2020.

2 | DESCRIPTION OF THE DATASETS

Three wind atlases are used in this study. ERA-5 has been developed on a horizontal grid spacing of ∼30 km in mid-latitudes. Compared to ERA-5, DOWA assimilates additional regional observations and uses a fine grid spacing of 2.5 km. Every 3 hr, data assimilation is applied to initialise a new forecast cycle. The assimilated observations include ASCAT satellite sea-surface wind fields and MODE-S EHS aircraft wind profile measurements. NEWA is a wind atlas covering the entire EU. Thus its performance cannot be expected to be comparable to a wind atlas that was tailored for a certain region (like DOWA). It is rather a trade-off between many different model settings. NEWA also has a fine grid spacing of 3 km, but was produced with a slightly different procedure: it consists of 8-day runs (with the first day considered as spin-up) in which some model fields are nudged towards the ERA-5 reanalysis data to prevent the simulations from drifting away from the synoptic situation. Table 1 further summarises relevant characteristics of the wind atlases .

For validation we use the MMIJ dataset, which spans four years of observations (2012–2015) at several altitudes up to 315 m, spaced approximately 25 m apart. Observations at 27, 58, and 90 m are from mast-mounted cup and sonic anemometers and wind vanes, while the data beyond 115 m were obtained with an upward-pointing continuous-wave lidar. More details can be found in (e.g. Kalverla *et al.*, 2017, 2019a). The large temporal extent, distance to shore and vertical measurement range makes this dataset optimally suited to characterise the wind climate.
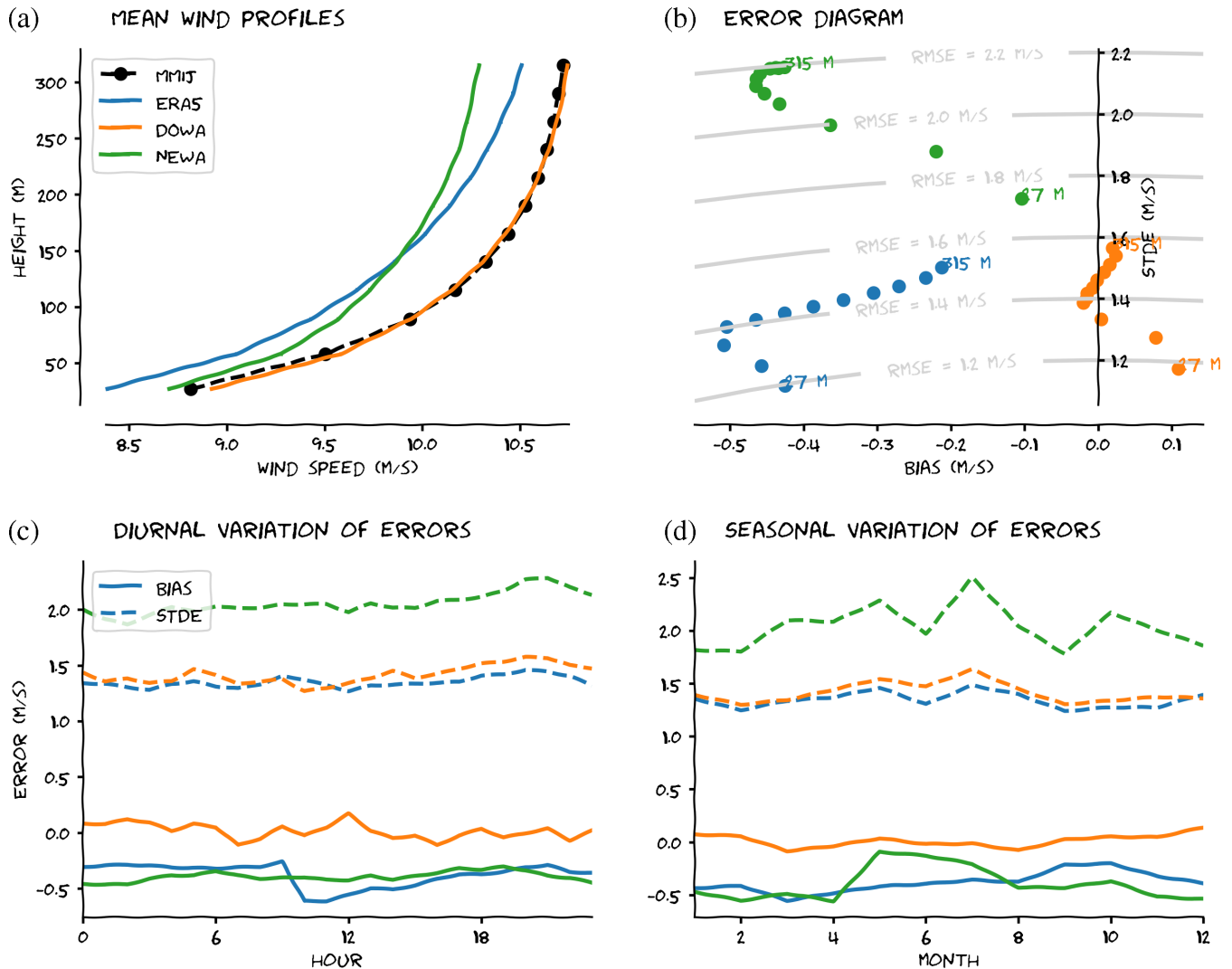
FIGURE 1 (a) Observed and modelled time-averaged vertical wind speed profile at met mast IJmuiden (MMIJ) and (b) error diagram showing the mean (bias) and standard deviation (STDE) of the error distributions ofthe three wind atlases compared to observations from MMIJ at multiple heights. (c, d) Diurnal and seasonal evolution of the mean and standard deviation of the same error distributions

Data quality was found to be very good (e.g., Poveda and Wouters, 2015), with only a few gaps (not shown). The observation data are available at 10 min intervals, and were hourly averaged to facilitate comparison with the reanalysis data. The wind atlas data were vertically (linearly) interpolated to, and temporally aligned with, the observations to obtain four collocated datasets.

3 | EVALUATION OF WIND SPEED AND THE ROLE OF ATMOSPHERIC STABILITY

Time-averaged wind speed profiles (Figure 1a) demonstrate a striking correspondence between DOWA and the observations, considering that MMIJ observations were not assimilated into the wind atlas. ERA-5 underestimates

the wind speed by $\sim 0.5\,\mathrm{m\cdot s^{-1}}$ through the whole profile, while NEWA is nearly unbiased near the surface and reaches a slow bias of $0.5\,\mathrm{m\cdot s^{-1}}$ at 300 m. As a consequence, NEWA appears to underestimate the wind speed shear within the layer. Parsons *et al.* (2018) showed that ERA-5 on average shows a good skill for wind speed, but that ERA-5 underestimates the wind speed for very extreme sea states. Overall, the underestimated wind speed is consistent with Couto *et al.* (2019) who found a similar bias close to the Portuguese coast. A more complete picture is obtained if we consider Figure 1b, which shows both the bias and RMSE due to phase differences. Although DOWA is nearly unbiased, it does not exactly align with the observations, leaving an overall RMSE of $\sim 1.5\,\mathrm{m\cdot s^{-1}}$. A small bias of $0.1\,\mathrm{m\cdot s^{-1}}$ is found only near the surface. For ERA-5 and NEWA, the negative bias is clearly present, and its altitude-dependence is also
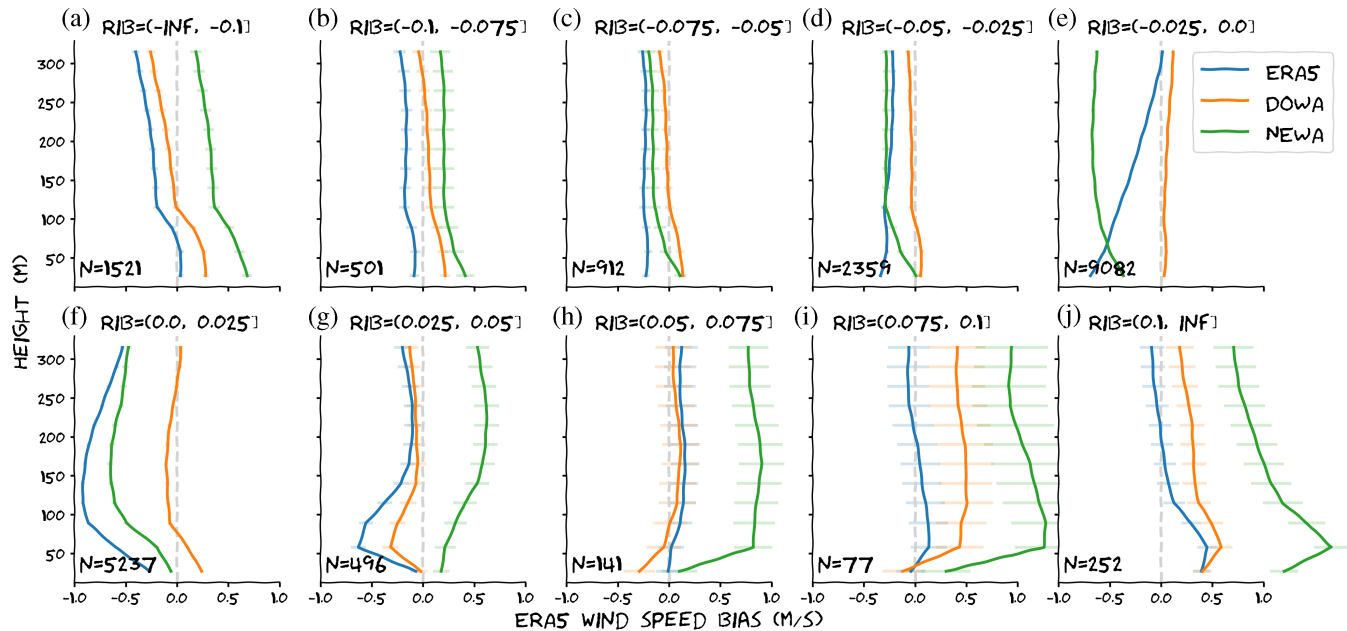
FIGURE 2 Vertical profile of the wind speed bias in all wind atlases for different stability intervals (based on the observed bulk Richardson number). (a)–(e) unstable stratification, and (f)–(j) stable stratification

apparent. For NEWA, the random errors are larger than for ERA-5 and DOWA, especially away from the surface.

Figure 2 reveals clearly that stability affects the bias in all datasets. Considering most unstable stratification, NEWA overestimates the near surface wind by $0.6\,\mathrm{m\cdot s^{-1}}$ and this bias decreases aloft. While ERA-5 and DOWA show the same shape of the bias profile, they mainly underestimate the wind in the upper part of the profile. For all datasets the bias reduces and appears more uniform with height for moderate unstable stratification ($-0.1 < Ri_b < -0.025$, where Ri_b is the bulk Richardson number). The near-neutral class contains the majority of datapoints ($-0.025 < Ri_b < 0.0$) and herein DOWA is nearly unbiased while NEWA underestimates the wind speed, which increases with height. Surprisingly ERA-5 shows a slow bias of $\sim 0.7\,\mathrm{m\cdot s^{-1}}$ near the surface, though its bias decreases with height. The most prominent biases and sensitivities occur for stable conditions, consistent with findings in Baas *et al.* (2016) for the HARMONIE model results over the North Sea. For near-neutral conditions, NEWA reveals a slow bias, which switches to a wind speed overestimation for $Ri_b > 0.025$, which increases for stronger stability, even to $1.0\,\mathrm{m\cdot s^{-1}}$ for ($0.075 < Ri_b < 0.1$). For that class DOWA also overestimates the wind by $\approx 0.5\,\mathrm{m\cdot s^{-1}}$, while ERA-5 represents this class rather well. The wind speed in NEWA is more accurate near the surface, which does not support a deficiency in the surface roughness formulation. Rather, it seems that too little momentum is transported downward to the surface. This could be a result of the large-scale nudging strategy employed in

the NEWA. Above the boundary layer, momentum fields were nudged towards the ERA-5 values. If wind speed is underestimated in ERA-5, it is thus very plausible that this error propagates to NEWA. In contrast, DOWA is completely free in the inner domain, except for the 3 hr data assimilation updates.

Examining the seasonal cycle shows that the slow bias in ERA-5 is present throughout the lowest 300 m, though is most prominent from September to February with a maximum negative bias at the surface. From March to August the bias profile shows a maximum at $\sim$100 m (not shown). The wind speed underestimation in ERA-5 might be explained by the surface roughness, or the Charnock parameter, which dictates the relation between wind and waves. The wind speed seems to be mainly underestimated near the surface, which may point to an overestimation of the surface roughness. However, a comparison of the modelled significant wave height against wave height observations at the nearby K13 platform indicates that ERA5 slightly underestimates the wave height by 9% (not shown). Unravelling the wave height biases further in classes of atmospheric stability, we find the largest underestimation for $Ri_b > 0.05$ of 33%. For $Ri_b < -0.05$ the wave heights are underestimated by 26%. In near-neutral conditions, the wave heights in ERA-5 do show smaller biases. Although roughness and wind speed are interdependent over the sea, it seems the wind speed underestimation is not triggered by an overestimated roughness here. The representation of atmospheric stability and the turbulent mixing under stable conditions are more likely explanations.
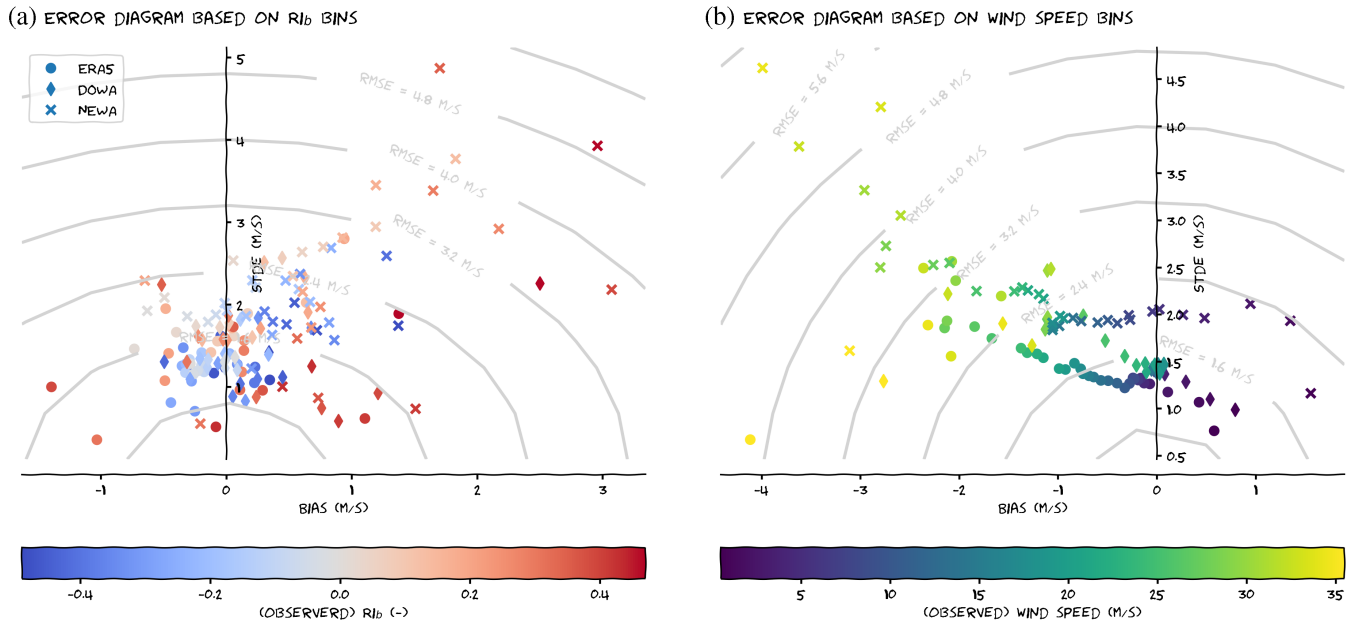
FIGURE 3 Error diagrams illustrating the performance of the three wind atlases for mean wind speed below 300 m in the period 2012 to 2015 as a function of (a) atmospheric stability and (b) wind speed. Observation data were used to aggregate the error statistics

Alternatively, the relatively coarse resolution of ERA-5 may induce a smoothing effect, especially for high wind speed events.

The smoothing effect described above can effectively suppress the random errors shown in Figure 1. Phase differences such as a delayed front passage will lead to a double penalty. The inability to reproduce small-scale features thus prevents a double penalty, which explains why the ERA-5 data perform relatively well in terms of the standard deviation of the error (STDE; Figure 1b). While one expects that higher-resolution models are generally more subject to this problem, DOWA performs similarly to ERA-5, which presumably has the merit of frequent data assimilation in DOWA (Duncan *et al.* 2019a).

A remarkable discontinuity in the diurnal cycle of wind speed was revealed in the ERA-5 data by Kalverla *et al.* (2019a). At 1000 UTC, the wind speed bias suddenly strengthens. To verify whether this artefact propagated to the other wind atlases, the diurnal and seasonal cycle of the wind speed bias and STDE are shown in Figure 1c,d. The discontinuity in the diurnal cycle occurs only for ERA-5. However, in the seasonal cycle, we find another remarkable feature in the NEWA data – a smaller bias in spring and early summer. The reason becomes clear upon inspection of Figure 3, which shows the wind speed bias and STDE as a function of (observed) wind speed and stability. Stable conditions lead to a substantial positive bias, while high wind speeds lead to a large negative bias. In other words, all models but especially NEWA tend to underestimate very strong winds, while they overestimate the wind speed during stable conditions. Since winds are

generally stronger in winter and stable conditions occur more frequently in spring and summer, this helps to explain the seasonal cycle of the bias in NEWA.

The results in Figure 3a are consistent with Kalverla *et al.* (2019b), who found a reduced model performance in stable conditions. Hence our current results are a substantial corroboration of this earlier result. Apparently, stable conditions are in general still challenging (Holtslag *et al.*, 2013; Sandu *et al.*, 2013; Steeneveld, 2014; Tsiringakis *et al.*, 2017), despite recent efforts to improve the turbulent mixing formulation (Sandu *et al.*, 2014; Bengtsson *et al.*, 2017; Olson *et al.*, 2019a; 2019b). Furthermore, the results in Figure 3b support the hypothesis that the slow speed bias results from a smoothing effect, as this would manifest itself most clearly for distinct wind speed maxima.

4 | EVALUATION OF WIND DIRECTION

Wind direction is critical for offshore wind energy purposes for determining the directional shear on wind turbines, understanding the model's representation of boundary-layer friction, and its representation of advection of onshore atmospheric phenomena towards offshore wind parks, for example (Dörenkämper *et al.*, 2015; Wagner *et al.*, 2019). Here we discuss two methods to evaluate wind direction, that is, (a) the bias as the difference between the means of the wind direction of two samples (Figure 4a), and (b) the bias as the mean of wind direction differences (Figure 4b). These definitions are equivalent
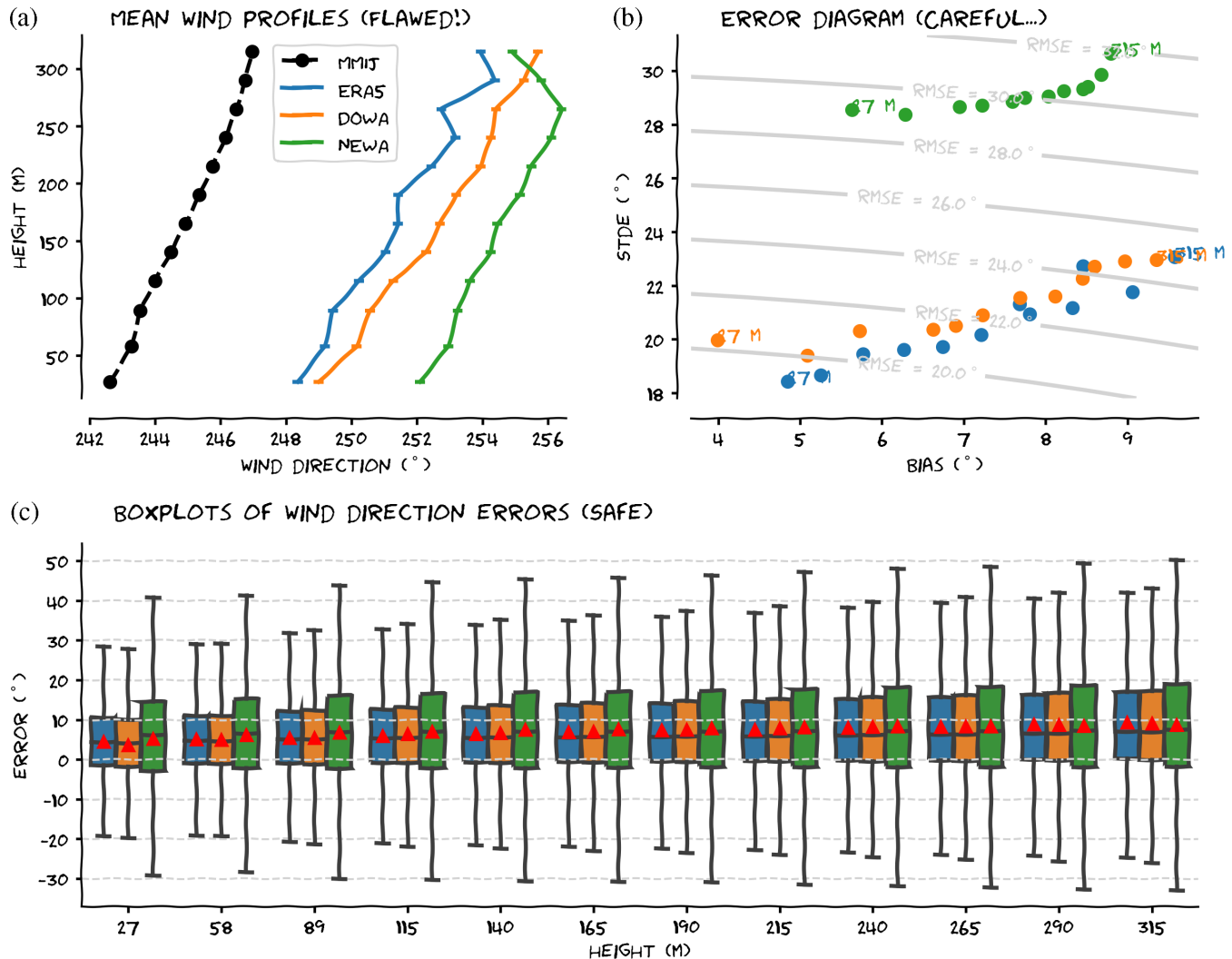
FIGURE 4 Modelled and observed wind direction using (a) circular statistics and (b) mean of difference in wind direction. (c) Boxplots show the wind direction error distributions for the three wind atlases as compared to MMIJ observations at multiple levels. Red triangles denote the arithmetic means. Outliers are not drawn, because that would require axis limits of up to $\pm 180°$

for arithmetic means. Figure 4a suggests that the bias is more or less constant with height, but Figure 4b suggests it increases steadily. This apparent inconsistency stems from the use of vector means instead of arithmetic means in Figure 4a – a common method to compute angular statistics to avoid artefacts like averaging 355 and 5°. The difference between the vector mean and the arithmetic mean is greater when the angles are widely distributed (Jammalamadaka and Sengupta, 2001). Figure 4b shows that the standard deviation increases with height, thus the apparent wind veer with height might represent a statistical artefact rather than a physical effect.

To circumvent the pitfalls of circular statistics, the performance for wind direction can be inferred directly from the error distribution of the wind direction (Figure 4c). Indeed, both the width and the mean of the error distribution increase with height, consistent with Figure 4b.

The positions of the means relative to the medians, and the upward shift of the 75percentile as compared to the relatively constant location of the 25percentile indicate a changing skewness with height. From a physical point of view, the wind in the wind atlas veers with respect to the observations, and this veering increases with height. These results confirm findings in Kalverla *et al.* (2019b) and previous literature, where the models' inability to represent a realistic wind veer with height was related to excessive mixing in stable conditions and to strong baroclinity (Brown *et al.* 2005; Holtslag *et al.*, 2013; Sandu *et al.*, 2013). Despite the remaining biases, wind direction in ERA-5 and DOWA is better represented than in the 30 operational forecasts evaluated in Kalverla *et al.* (2019b), probably due to data assimilation, or model improvements discussed in Sandu *et al.* (2014) and Bengtsson *et al.* (2017). Again, the relative wide error distributions in NEWA may be the
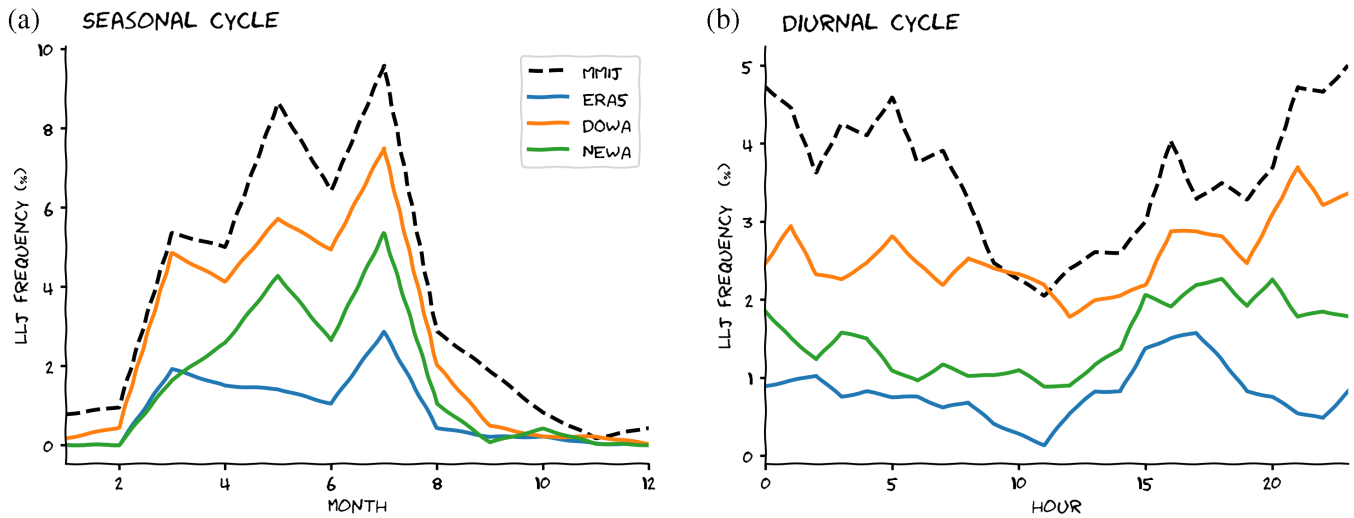
FIGURE 5 (a) Seasonal and (b) diurnal cycle of low-level jets, based on observations and wind atlases. Low-level jets are defined as all hourly wind speed profiles with a maximum exceeding a fall-off threshold of $2\,\mathrm{m\cdot s^{-1}}$ at the location of met mast IJmuiden, expressed as a percentage of the total number of wind profiles in that month/hour

result of a substantial double penalty, considering that NEWA consists of 8-day simulations without data assimilation – as compared to the 3 hr update cycles for DOWA.

5 | ANOMALOUS WIND EVENTS IN WIND ATLASES

Building upon the model representation for general wind characteristics, this section discusses the representation of anomalous events in these datasets. The methodology used to assess model performance is explained after the subsection about LLJs, using the LLJ data as illustration.

5.1 | Low-level jets

LLJs are wind profiles with a wind speed maximum near the surface, as illustrated, for example, in Kalverla *et al.* (2019a). LLJs over MMIJ occur primarily in spring and early summer, often appear at the end of the afternoon and persist until the next morning. Wagner *et al.* (2019) studied the LLJ climatology at the FINO1 site in the German Bight and found LLJs occur for 14.5% of the time and on 64.8% of days, mostly from directions between east and south. They are formed by a variety of mechanisms, but baroclinic effects, orographic effects (e.g., flow forced through the Dover Strait; Capon 2003), and the combination with a stable boundary layer explain most of their characteristics (also Wagner *et al.*, 2019). Recently, Kalverla *et al.* (2019a) demonstrated that LLJs are present in the ERA-5 data, although they tend to be located too high above the surface. Consequently, when the ERA-5 data are interpolated to

observation heights, the number of LLJs is grossly underestimated. However, the seasonality could still be faintly recognised at MMIJ.

To investigate whether the refined datasets improve upon the representation of LLJs, we study the seasonal and diurnal cycles for all three wind atlases (Figure 5). The observations exhibit a pronounced seasonal cycle, the erratic nature of which has been discussed at length in Kalverla *et al.* (2019a). In the four-year observation period (2012 to 2015), May and July saw more LLJ events than April and June.

Indeed, ERA-5 grossly underestimates the amplitude of the seasonal cycle, and completely misses the peak in May. The other two datasets, especially DOWA, demonstrate considerably better skill. The diurnal cycle is characterised by a distinct dip around noon, and peaks in the afternoon and the early hours of the morning. The afternoon peak, presumably related to the adjustment of the sea breeze, appears to be best represented in the wind atlases. If two different mechanisms are responsible for LLJ formation, one of these mechanisms might be better resolved than the other. Alternatively, the formation mechanism might be relatively well represented, but the jets' propagation through the night proves challenging. Further investigation of individual LLJ events could provide a definitive answer in this matter, but that exercise is beyond the scope of this evaluation.

Often, a fall-off threshold is used to distinguish between "real" LLJs and "normal" conditions that happen to show a weak wind speed maximum by chance. In line with previous studies, a fall-off threshold of $2\,\mathrm{m\cdot s^{-1}}$ was used for Figure 5. Alternatively, the absolute fall-off may be inspected directly. This is shown in Figure 6, where
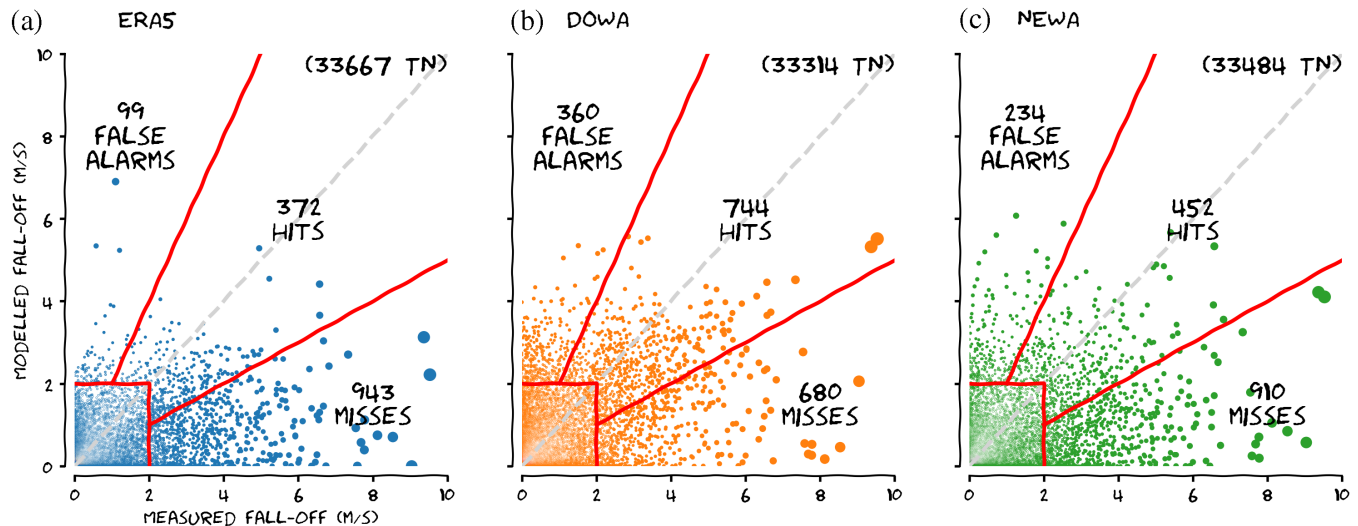
FIGURE 6 Scatter plots of absolute fall-off of low-level jets as represented in the wind atlases versus observed fall-off. A $2\,\mathrm{m\cdot s^{-1}}$ fall-off threshold is indicated by the red box. Outside this region, red lines indicate the region where wind atlas and observations agree on the absolute fall-off value to within a factor of 2. The dashed grey line indicates a 1:1 correspondence. The point size increases when further away from the origin to expose the structure both in dense and sparse regions of the graph. TN is an abbreviation for true negatives, relating to the bottom left corner in each panel

the red box near the origin indicates the fall-off threshold of $2\,\mathrm{m\cdot s^{-1}}$. All points outside this box may be regarded as significant LLJ events – either observed, modelled, or (preferably) both.

5.2 | Quantification of model performance for anomalous events

Usually model performance is expressed in summary statistics or as a variety of skill scores. With the current data the signal will be dominated by non-significant events, and since the differences are subtle, mismatches in timing will lead to very low correlation coefficients (notice the dense clustering of scatter points along the zero lines of both axes of Figure 6). It is reasonable to suppose that a stronger LLJ event is more likely to be picked up by the wind atlas data, and as it occurs, these most anomalous events are the main focus of this section.

To quantify model performance, we establish the following contingency "rules":

(a) If either observations or wind atlas data report on the presence of a significant LLJ event, and the absolute fall-off in both datasets is comparable to within a factor of 2, then the model performance is satisfactory and the data point is counted as a hit.
(b) If a significant LLJ is observed, but not present in the wind atlas data, or if it is present in the wind atlas data but much weaker than observed (less than half as strong), then this data point is regarded a miss.

(c) If a significant LLJ is present in the wind atlas , but it is not observed, or it is observed but the wind atlas overestimates its strength by at least a factor of 2, then it is labelled as a false alarm. All other events are true negatives.

The contingency rules allow for estimating skill scores, such as the probability of detection, false alarm rate, or critical success index (CSI; Schaefer, 1990). The CSI, defined as

$$\mathrm{CSI} = \frac{hits}{(hits + misses + false\ alarms)}$$

is a simple and intuitive parameter to compare the performance of several models: the score increases if more events are correctly predicted, and it decreases as more events are missed or falsely predicted.

An alternative and more robust (but less intuitive) statistic than the CSI is the symmetric extreme dependency score, defined as (Hogan *et al.*, 2009):

$$\mathrm{SEDS} = \frac{\ln[(hits + false\ alarms)/n] + \ln[(hits + misses)/n]}{\ln(hits/n)},$$

where n is the total number of events (hits+misses+false alarms+true negatives). The SEDS varies between -1 and 1, where 1 indicates a perfect forecast, a random forecast would receive a skill score of 0, and a forecast that actually degrades the quality of a random forecast tends to -1.

The CSI and SEDS both penalise phase errors, which is desirable in forecast verification. However, for climatological studies for resource assessment, phase

TABLE 2 Critical success index (CSI), symmetric extreme dependency score (SEDS) and frequency bias (FBIAS) for the representation of LLJs in ERA-5, DOWA and NEWA

	ERA-5	DOWA	NEWA
CSI	0.26	0.42	0.28
SEDS	0.67	0.73	0.65
FBIAS	0.36	0.78	0.50

errors are not essential. An alternative score that measures the climatological model performance is the (frequency) bias:

$$\text{FBIAS} = \frac{total\ predicted}{total\ observed} = \frac{hits + false\ alarms}{hits + misses}.$$

A frequency bias of ∼1 means that the total number of LLJ events is more or less correct, even if the timing is wrong. FBIAS < 1 represents an underestimation of the number of LLJ events and vice versa. A downside of this score is that correct model forecasts for the wrong physical and dynamical reasons are counted as successes.

All wind atlas datasets underestimate the amount of significant LLJ events (Table 2). Over four years (∼35,000 hr wind profiles), only 372 LLJs have been correctly captured in ERA-5, against 943 missed events. With "only" 99 false alarms, this results in a frequency bias of 0.36. DOWA performs better and picks up approximately twice as many LLJ events, reflected in a much higher frequency bias of 0.78. The representation in NEWA is intermediate: more hits than ERA-5, but substantially more false alarms. Hence, (a) although NEWA does not seem to improve upon ERA-5 with respect to a general validation of the wind speed profiles, the increased resolution does favour the climatological LLJ representation, and (b) DOWA especially improves upon the representation of the dynamical conditions in coastal areas.

5.3 | Wind ramps

Wind ramps are rapid changes of the wind speed and/or direction in time. In climatologies for wind energy applications, the mean wind is often assumed to be stationary, or time-averaged statistics are considered. Therefore, wind ramps are interesting anomalous events that require additional, tailored evaluation. Kalverla *et al.* (2017) determined the wind speed and direction differences over various time intervals in the MMIJ dataset, and studied the frequency distributions to build some intuition about the magnitude of these difference. Naturally, the frequency distributions centres around zero, for in the absence of a

long-term trend, increasing wind speeds must be balanced by equivalent decreases. For wind direction, this is not necessarily true, but it was found in practice. Kalverla *et al.* (2017) used the 5 and 95percentile to obtain site-specific characteristic up-ramp and down-ramp thresholds, and analysed the sensitivity to this threshold.

Here we use the 2.5 and 97.5percentile instead, to put even more emphasis on the most extreme conditions. The cumulative probabilities of hourly differences in wind speed and direction based on MMIJ and the corresponding grid points in the wind atlas datasets are shown in Figure 7a,b. In general, only small differences appear between the datasets, though the distribution of wind speed differences in DOWA is slightly broader than observed, and the distribution is slightly too narrow in ERA-5 and DOWA. The differences between datasets are quantified through the 2.5 and 97.5percentile. A typical 1 hr down-ramp at MMIJ amounts to $-2.0 \, \text{m·s}^{-1}$, while ERA-5, DOWA and NEWA estimate the ramp intensity at -1.6, -2.2 and $-1.9 \, \text{m·s}^{-1}$, respectively. Typical up-ramp values amount to $2.0 \, \text{m·s}^{-1}$ according to observations, and 1.7, 2.3 and $2.0 \, \text{m·s}^{-1}$ for the respective wind atlas datasets. Thus, NEWA best captures the climatology of wind speed ramps. Before further quantification, some notes must be made about the evaluation of wind ramps.

Figure 7c shows the joint distribution of wind speed and subsequent hourly wind speed differences at MMIJ. Such a representation might be relevant for forecasting applications, where wind ramps within the cubic part of the power curve lead to the largest power fluctuations. A slightly negative correlation between wind speed and wind speed difference appears. This is in agreement with the analysis of MMIJ data in Kalverla *et al.* (2017) (their figure 7) who found the most severe down-ramps for high mean wind speeds.

Ramps over relatively short time intervals are probably more relevant for energy applications, but unfortunately ERA-5 and DOWA do not offer short intervals (NEWA is available every 30 min). To provide some intuition about the relevance of hourly ramps, Figure 7d depicts the typical wind speed up-ramps for MMIJ as a function of the time interval over which the ramp is considered . Herein, a moving average and a resampling (with corresponding time interval) were subsequently applied to the MMIJ data. The 97.5percentile increases almost linearly with the time interval of the underlying data, except below 20 min, where the acceleration appears smaller. To show the robustness of this result, and also to illustrate the sensitivity to the ramp threshold, three different percentiles are shown in Figure 7d. Both the typical ramp and the acceleration are smaller for lower ramp thresholds, which makes sense. The current results differ from Kalverla *et al.* (2017), since they only applied resampling (no moving average).
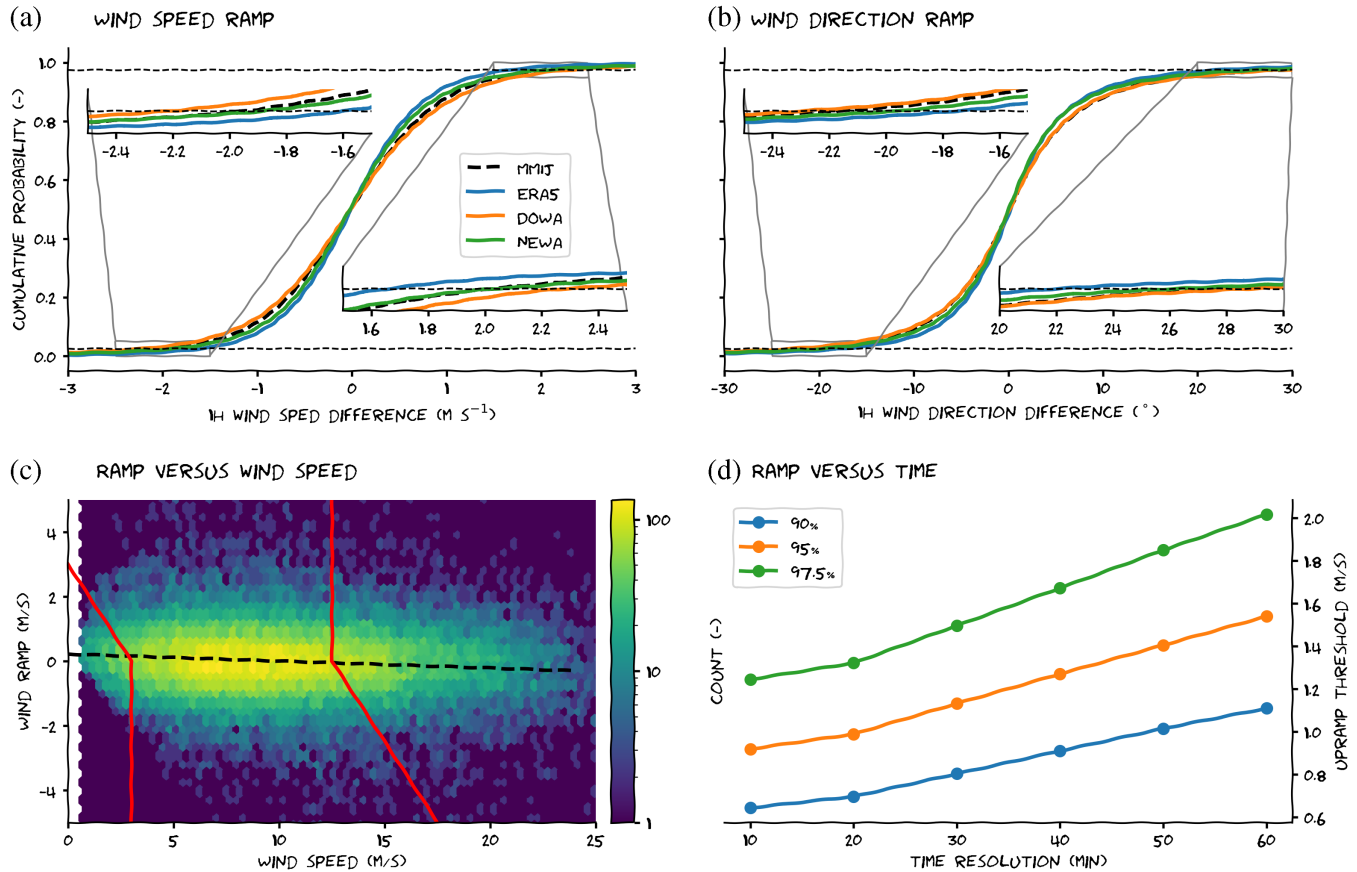
FIGURE 7 Cumulative probability distribution of 1-hourly (a) wind speed and (b) direction differences based on observations and wind atlas data. The dashed lines indicate the 2.5 and 97.5percentile as thresholds for the classification of typical up- and down-ramp events. (c) Hexbin visualisation of hourly MMIJ data illustrating the frequency of wind speed ramps as a function of wind speed. The dashed line represents a linear fit, and the red lines denote the area where the wind ramp causes a disturbance within the cubic part of a conceptual power curve. (d) shows wind speed ramp as a function of the time resolution of the underlying data, based on MMIJ data resampled at different intervals using a moving average to mimic a smoothing effect. Three different up-ramp thresholds are used to assess the robustness of the relation

The averaging appears to be responsible for the decreased acceleration between 10 and 20 min (not shown). These factors explain why the empirical square root relation between wind ramp magnitude and time interval found in Kalverla *et al.* (2017) does not correspond with the present results.

Upon visual inspection, the distribution of wind direction differences is well-captured by DOWA and slightly underestimated in the other datasets (Figure 7b). Typical 1 hr direction up- and down-ramps at MMIJ are −22° and +28°, respectively. ERA-5 underestimates both thresholds: −18° and +22°, and DOWA slightly overestimates them: −23° and +29°. NEWA reports thresholds of −20° and +26°, i.e. a small underestimation. The asymmetry of the distribution is probably related to frontal passages, which are accompanied by an abrupt wind veer. Wind direction ramps on hourly time-scales as investigated in this study are relevant for offshore wind power forecasting,

especially when below rated power, as the efficiency of a wind turbine array depends on the generated wakes, and the fixed layout strongly depends on the wind speed.

While wind atlases are not actual forecasts, it is illustrative to inspect the 1:1 correspondence between observed and simulated wind ramp events (Figure 8a–c). The R^2 value corresponding to a linear fit is annotated for reference, but this parameter is mostly determined by the bulk of the data, while our main interest is in the extreme cases. Hence, the contingency rules from the previous section are used. The red box in the middle corresponds to the typical up- and down-ramp thresholds as observed. In this case, the distribution of wind speed differences is two-sided, that is, an additional possibility where the wind atlas would "predict" an up-ramp while a down-ramp is actually observed (or vice-versa) is present. Although this rarely occurs in practice, the possibility requires an additional rule to distinguish between false up-ramp
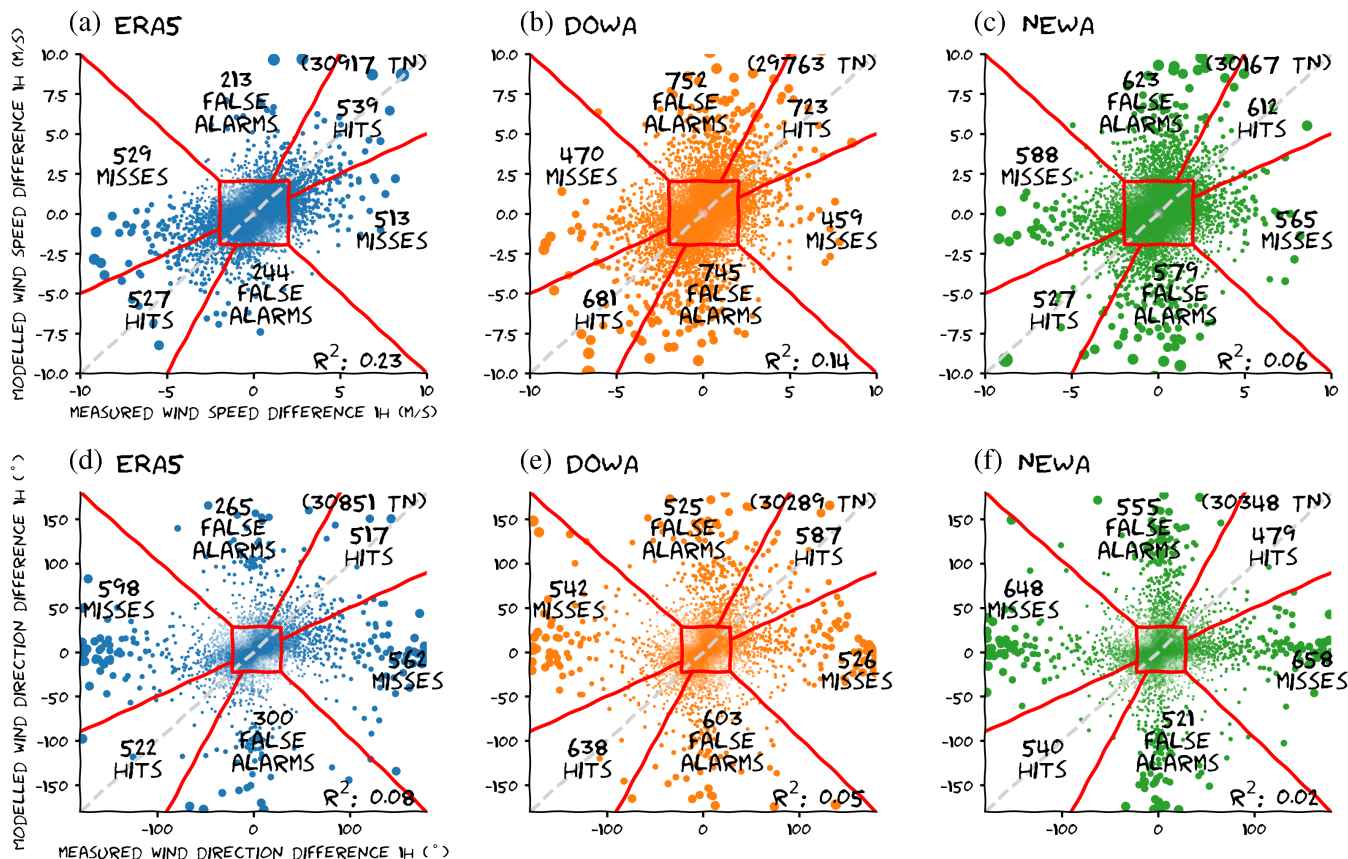
FIGURE 8 Scatter plots of 1-hr (a–c) wind speed and (d–f) direction difference as estimated from ERA-5, DOWA, and NEWA. In this figure, both positive and negative values are considered. The ramp thresholds in red have been based on the 2.5 and 97.5percentile of the observed ramps. Additional red lines have been inserted to separate missed up-ramp events from false down-ramp alarms and vice versa

alarms and missed down-ramp events (and vice versa). The rule employed here can be summarised as "whichever is greater". The corresponding skill scores are provided in Table 3.

The frequency bias again demonstrates that ERA-5 underestimates the number of wind ramps. This is expected, since the relatively coarse horizontal grid spacing of 30 km seriously limits the model representation of small-scale structures that are responsible for ramp events. However, the number of false alarms is also limited (at least, relative to the other datasets). Consequently, the CSI and SEDS are highest for ERA-5. The false alarm rate is much higher in DOWA, which indicates an overestimation of ramp events (FBIAS = 1.24). NEWA has almost no frequency bias, but the 1:1 correspondence with observations is particularly poor, as reflected by the high number of misses and false alarms compared to the number of hits. In other words, a more realistic climatology of ramps in NEWA comes with a deterioration in timing of these events. While the climatology is more important during the resource assessment and planning phase, correct timing is obviously quite relevant for forecasting applications.

Model performance for wind direction ramps is visualised and quantified in Figure 8d–f and Table 3. The interpretation is analogous to that of wind speed ramps: ERA-5 underestimates wind ramps, DOWA in this case performs best in a climatological sense, and NEWA especially struggles with the timing of events. Further investigation is needed to assess whether the physical characteristics of the wind ramps (both in wind speed and direction) are consistent between all four datasets.

5.4 | Extreme shear

It is illustrative to split the wind vector in a streamwise and a normal component (Kalverla *et al.*, 2017). If the wind would turn without a change in magnitude, a substantial wind shear would remain hidden if only wind speed was analysed. Besides, the energy in the lateral wind component would be falsely regarded in load and power calculations. Thus, the analysis of extreme shear starts with an evaluation of the longitudinal wind component (aligned with the 115 m wind, at approximately hub-height) in

TABLE 3 Skill scores for wind ramps and wind shear ("shear" and "veer" are used here to refer to the longitudinal and lateral components)

	ERA-5	DOWA	NEWA	ERA-5	DOWA	NEWA
	Speed ramps			Direction ramps		
CSI	0.42	0.37	0.33	0.38	0.36	0.30
SEDS	0.70	0.61	0.58	0.66	0.61	0.56
FBIAS	0.72	1.24	1.02	0.73	1.03	0.90
	Wind shear			Wind veer		
CSI	0.40	0.69	0.66	0.00	0.00	0.00
SEDS	0.73	0.88	0.86	−0.11	-0.05	−0.04
FBIAS	0.51	0.80	0.84	0.09	0.14	0.13

Note: Up- and down-ramps have been combined in the calculation of skill scores, although it would also be possible to calculate them separately.
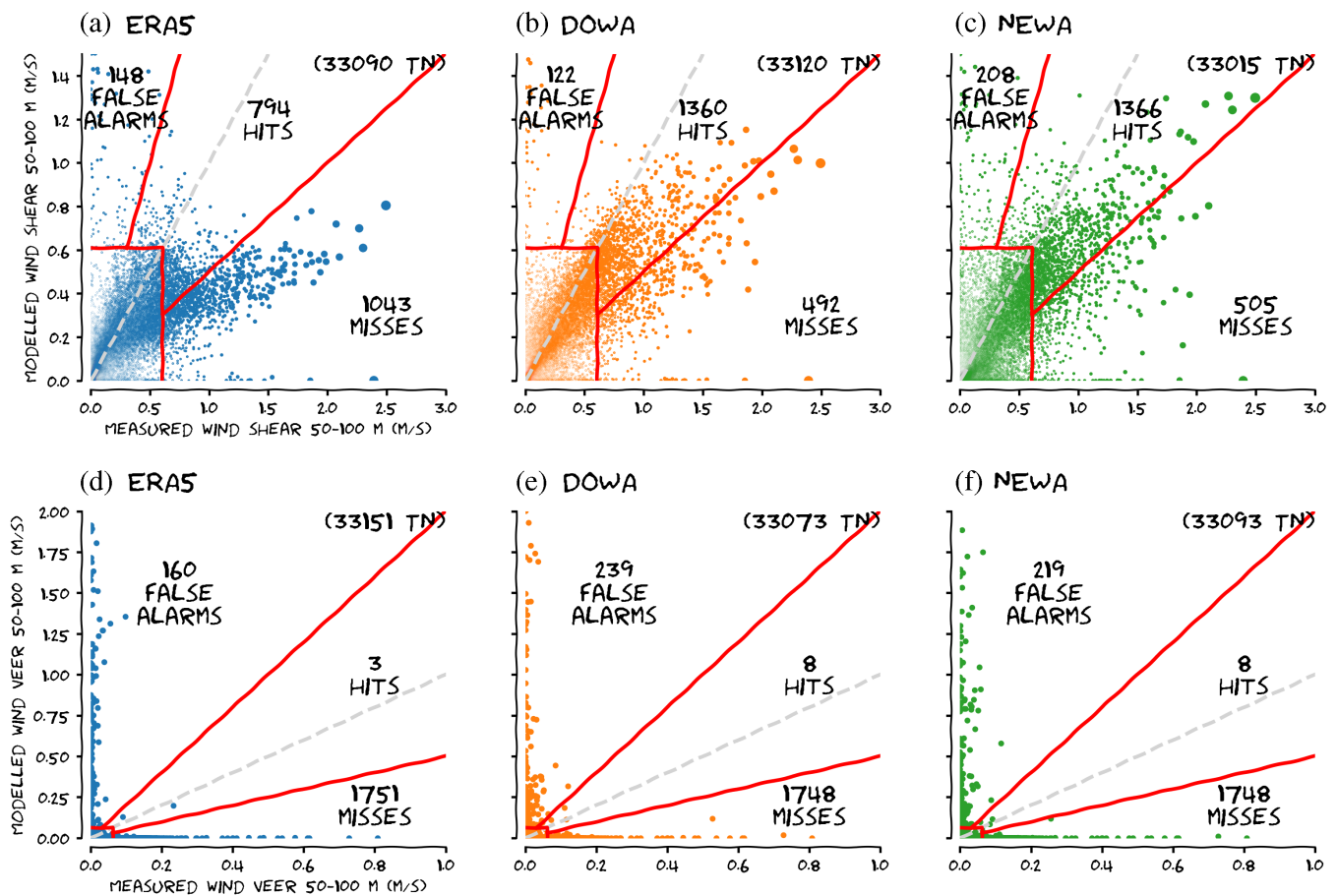

FIGURE 9 Scatter plots of modelled and observed integrated wind difference in the (a–c) longitudinal and (d–f) lateral wind components between 50 and 100 m. The red lines delimiting the extreme/anomalous shear thresholds are based on the 95percentile of the observations

Figure 9a–c, which is based upon accumulated wind shear between 50 and 100 m.

Clearly, the degree of wind shear is underestimated by all wind atlases datasets. Following the convention of Kalverla *et al.* (2017), the extreme shear threshold is defined as the 95percentile. This value is 0.61 m·s^{-1}

according to the MMIJ data, while ERA-5, DOWA and NEWA estimate it at 0.41, 0.49 and 0.47 m·s^{-1}, respectively. Especially for ERA-5, where the timing is quite well-represented, the underestimation of the absolute wind shear is so large that most extreme events (exceeding the 95percentile of the observations) are classified as

missed events. Both DOWA and NEWA perform much better in this respect, with slightly better performance for DOWA.

It is interesting to explore the causes behind general underestimation of wind shear and the difference between the datasets. Earlier we found that the surface roughness in ERA-5 is slightly underestimated. Stable stratification suppresses turbulent mixing and thereby supports the development of strong shear (and LLJs). As shown in Section 4 and in Kalverla *et al.* (2019b), NWP models still struggle to adequately represent these conditions. An alternative rendering of Figure 9, in which the observed Richardson number was used to further categorise the data in the scatter points (not shown), revealed that the majority of the extreme shear cases are indeed characterised by stable stratification, while unstable and neutral cases with little wind shear make up the bulk of the data. This explains why the underestimation of extreme shear is not reflected in the mean wind profile in Figure 1a.

The lateral wind shear, which may be interpreted as a measure of vertical wind veer, is evaluated in Figure 9d–f. The 95percentile threshold of the extreme accumulated shear over the layer 50–100 m is very low, 0.06 m·s^{-1}, and underestimated in all wind atlas datasets (0.001, 0.002 and 0.001 m·s^{-1}; but these values are hardly significant). Almost all extreme events are missed by the wind atlas data, and the fact that these misses are barely compensated by false alarms confirms that, also in a climatological sense, the wind atlas provide a poor impression of lateral (extreme) shear. Since turbulence generally tends to destroy vertical gradients, it is likely that these errors are the result of excessive mixing, either due to inadequate representation of the physics, including insufficient resolution, or due to misrepresentation of atmospheric stability, as discussed previously.

5.5 | Wind extremes

Finally, we address wind speed extremes as anomalous events because the conventional statistics may not adequately capture them. For example, the Weibull fit is strongly determined by the bulk of the data, but especially rare events in the tail may be relevant for structural loads. Therefore in Kalverla *et al.* (2017), extreme value theory was applied to estimate the 50-year extreme wind speed, based on the IJmuiden observations. Because four years is too short to select only annual maxima, the method of independent storms (Palutikof *et al.*, 1999) was used to select ∼ 40 unrelated events within the measurement period. Here this analysis is repeated for the wind atlas (Figure 10), though with hourly-averaged observations

rather than 10 min observations. Thus, here we find much lower estimates of the 50-year wind speed extreme. Since design standards are based on the 10 min estimate (Burton *et al.*, 2011), this section mostly serves as model validation and intercomparison.

The smoothing effect of a relatively coarse model resolution has been used in Section 3 to explain the underestimation of high wind speeds in ERA-5 and NEWA, since the latter is nudged towards the ERA-5 momentum fields. A similar effect is introduced by the time-averaging of the observations. The impact is substantial: while an extreme value of $42.7 \pm 2.4 \text{ m·s}^{-1}$ was reported in Kalverla *et al.* (2017), here we find $36.7 \pm 2.1 \text{ m·s}^{-1}$, a difference of 6 m·s^{-1}. Consequently, the wind atlas data should not be used directly to estimate wind extremes. Nevertheless, the difference between the three wind atlases can be compared to the uncertainty related to the spatio-temporal characteristics. With a 50-year extreme value of $35.9 \pm 2.5 \text{ m·s}^{-1}$, ERA-5 closely approaches the estimate based on hourly-averaged observations. DOWA actually overestimates it with a value of $38.5 \pm 2.2 \text{ m·s}^{-1}$, and NEWA underestimates it at $34.2 \pm 1.6 \text{ m·s}^{-1}$.

Thus, the difference between the three models is small compared to the impact of time-averaging. A more systematic investigation of the 50-year extreme as a function of the spatio-temporal characteristics of the underlying data could provide the additional information required to obtain reliable estimates of wind extremes from the wind atlases . Coupling between weather models and large-eddy simulations has received much attention lately, and is rapidly becoming accepted (e.g. Muñoz-Esparza *et al.*, 2014; Sanz Rodrigo *et al.*, 2017; Hewitt *et al.*, 2018). These developments make such a study possible in the near future.

Yet, even if reliable data about past extremes is available (including information about their validity), there is another potential pitfall that should be considered when applying extreme value theory. The fundamental assumption is that all extreme events in the data are drawn from the same parent distribution. The parent distribution, in this case, is the long-term wind climate at MMIJ. Using only four years of data is already pushing the limits of this assumption, for not all physical extremes that may occur within the current climate may be represented in the subset. Moreover, since the theory is used to make prediction far into the future, the assumption that the climate does not change may be violated. In 2017, *Ophelia* set a record for the easternmost Atlantic major hurricane. Scientists at the Dutch national weather service warn that such storms, which can get considerably stronger than other types of storms in this area, may occur more often as the ocean warms by climate change (Haarsma *et al.*, 2013; Baatsen *et al.*, 2015; Dekker *et al.* 2018). Hence, present results on
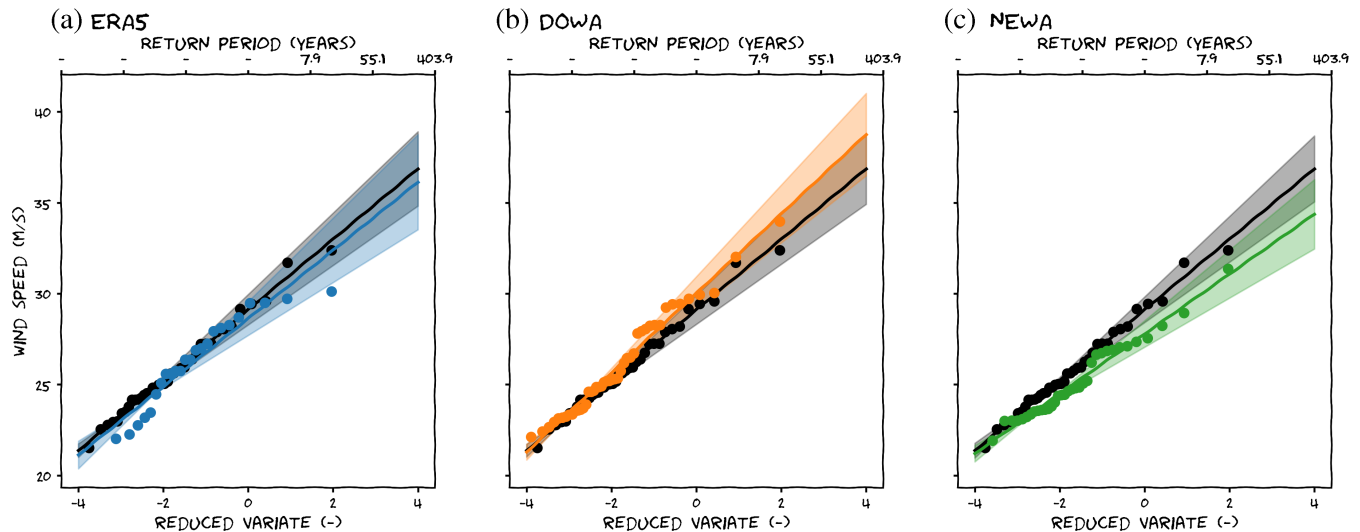
FIGURE 10 Gumbel plots for extreme value analysis based on the three wind atlases (colour) as compared to observations (black) from met mast IJmuiden. Return periods are shown on the top axis. Shaded regions represent uncertainty estimates of the mean plus or minus one standard deviation based on a Monte Carlo procedure
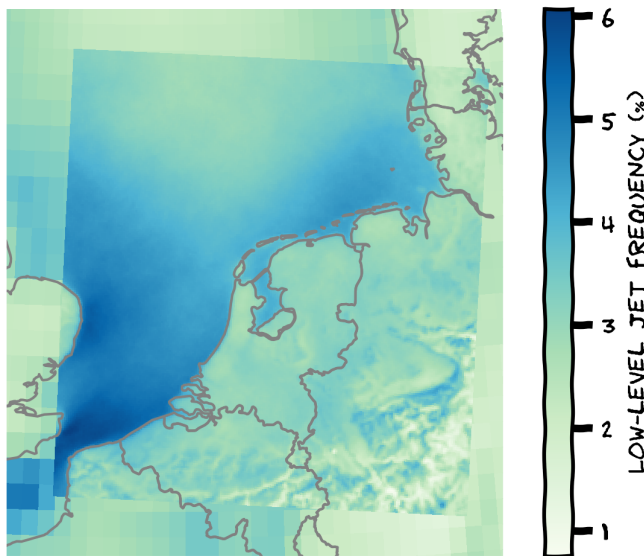

FIGURE 11 Spatial climatology of low-level jets up to a height of 600 m as represented in the Dutch Offshore Wind Atlas (2008–2017), overlaid on the corresponding ERA-5 visualisation

wind extremes only act as an illustration of their uncertainty, and interested readers are strongly advised to turn their attention to the dedicated literature referenced above.

6 | A SPATIAL CLIMATOLOGY OF LOW-LEVEL JETS BASED ON THE DOWA

Now that the performance of the three wind atlases have been evaluated, it is instructive to briefly highlight

their potential regarding spatial analysis of anomalous events. To illustrate the refinement achieved by downscaling, the ten-year (2008–2017) mean LLJ frequency over the DOWA domain is overlaid on a similar visualisation of the ERA-5 data (Figure 11). The detailed orographic structure in the southeast of the domain especially stands out, but also the coastal morphology is represented much more truthfully.

Some striking features are revealed in Figure 11. For example, the shape of the eastern coastline of East Anglia clearly favours LLJ formation. A band of preferred LLJ occurrence appears which more or less follows the shape of this coastline. Furthermore, the impact of the Dover Strait is clearly visible in the climatology, and the easternmost extremity of Kent leaves a LLJ "wake" towards the northeast. Some aspects of this specific jet are discussed by Capon (2003).

A fixed fall-off threshold of $2\,\mathrm{m\cdot s^{-1}}$ was used to produce Figure 11. Alternatively, it would be possible to map the mean, median or 95percentile of the absolute fall-off. That would provide additional information about the spatial distribution of LLJ characteristics. Moreover, this approach can be used for other anomalous events as well. However, a comprehensive spatial analysis of a variety of anomalous events is left for future work.

7 | SUMMARY AND CONCLUSIONS

This study evaluates three state-of-the-art wind atlas datasets used in the wind energy industry, that is, ERA-5, DOWA and NEWA against four years of high-quality

wind profile observations over the North Sea. Exceptional performance was found for the Dutch Offshore Wind Atlas, which was nearly unbiased in terms of wind speed. ERA-5 demonstrates comparable root mean square errors ($\sim 1.4\,\mathrm{m\cdot s^{-1}}$), but it generally underestimates the wind speed, probably due to the smoothing effects due to its relatively coarse resolution. NEWA, despite its increased resolution, does not improve upon ERA-5, which seems to inherit the wind underestimation from ERA-5, and an increase of the random errors suggest that the model is considerably more sensitive to the double-penalty problem. The fact that DOWA performs much better in this respect, even though its resolution is comparable to NEWA, illustrates the impact of the modelling strategy and additional data assimilation.

The wind in the wind atlas products is typically veered with respect to the observations, and this veering increases with height. ERA-5 and DOWA performed very similarly, while NEWA again exhibited a wider range of wind direction errors. Also a potential pitfall in using summary statistics for the evaluation of wind direction was illustrated.

The wind atlases' representation of anomalous wind events was evaluated. Generally, the relatively high-resolution models are able to represent more fine-scale structures, but this comes at the cost of considerable mismatches in the timing of events. For LLJs, DOWA outperforms the two other datasets. In a climatological sense, wind ramps are best represented in NEWA, but one-to-one correspondence is slightly better in DOWA. Extreme wind shear is best represented by the higher-resolution models, though they still underestimate the vertical wind shear, which has been linked to deficiencies in the representation of stable conditions. The representation of lateral shear, or wind veer is very poor in all datasets. For wind extremes, the differences between the models are nullified by the uncertainties related to spatiotemporal characteristics of the underlying data and about changes in future climate.

Finally, a climatological map of LLJ frequency based on the DOWA data was briefly discussed. Compared to ERA-5, the enhanced resolution reveals much more detail of the LLJ climatology; orography-related features can clearly be distinguished. The high-resolution data reveal many interesting aspects of the LLJ climatology, such as the role of orography and coastal effects. This opens up a wealth of possibilities for further investigations, and it is advised that climatological maps of anomalous events are incorporated in future standards.

ACKNOWLEDGEMENTS

ORCID

Gert-Jan Steeneveld https://orcid.org/0000-0002-5922-8179

REFERENCES

Baas, P., Bosveld, F.C. and Burgers, G. (2016) The impact of atmospheric stability on the near-surface wind over sea in storm conditions. *Wind Energy*, 19, 187–198.

Baatsen, M., Haarsma, R.J., Van Delden, A.J. and De Vries, H. (2015) Severe autumn storms in future Western Europe with a warmer Atlantic Ocean. *Climate Dynamics*, 45, 949–964.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelä, S., Nielsen, K.P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D.S., Subias, A., Tijm, S., Toll, V., Yang, X. and Køltzow, M.Ø. (2017) The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Monthly Weather Review*, 145, 1919–1935.

Brown, A.R., Beljaars, A.C.M., Hersbach, H., Hollingsworth, A., Miller, M. and Vasiljevic, D. (2005) Wind turning across the marine atmospheric boundary layer. *Quarterly Journal of the Royal Meteorological Society*, 131, 1233–1250.

Burton, T., Jenkins, N., Sharpe, D. and Bossanyi, E. (2011) *Wind Energy Handbook* (2nd ed.) Wiley, Chichester, UK.

C3S (2017). Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available at: https://cds.climate.copernicus.eu/cdsapp#!/home; accessed 21 January 2020.

Capon, R.A. (2003) Wind speed-up in the Dover Straits with the Met Office new dynamics model. *Meteorological Applications*, 10, 229–237.

Couto, A., Silva, J., Costa, P., Santos, D., Simões, T. and Estanqueiro, A. (2019) Towards a high-resolution offshore wind atlas – the Portuguese case. *Journal of Physics: Conference Series*, 1356. https://doi.org/10.1088%2F1742-6596%2F1356%2F1%2F012029

Dekker, M.M., Haarsma, R.J., de Vries, H., Baatsen, M. and van Delden, A.J. (2018) Characteristics and development of European cyclones with tropical origin in reanalysis data. *Climate Dynamics*, 50, 445–455.

Dörenkämper, M., Sïle, T. and Zagar, M. (2020) The making of the new European wind Atlas, Part 2: production and evaluation. *Geoscientific Model Development Discussions*. (personal communication, 5th January 2020).

Dörenkämper, M., Optis, M., Monahan, A. and Steinfeld, G. (2015) On the offshore advection of boundary-layer structures and the influence on offshore wind conditions. *Boundary-Layer Meteorology*, 155, 459–482.

Duncan, J.B., Marseille, G.J. and Wijnant, I.L. (2019a). DOWA validation against ASCAT satellite winds. Technical Report R11649, TNO, Utrecht, Netherlands. Available at: https://www.dutchoffshorewindatlas.nl/binaries/dowa/documents/reports/2019/01/18/tno-report-dowa-validation-against-ascatsatellite-winds/TNO+2018+R11649+DOWA+validation+ASCAT.pdf; accessed 21 January 2020.

Duncan, J.B., van der Werff, P.A. and Bot, E.T.G. (2019b). Understanding of the offshore wind resource up to high altitudes ($\leq 315\ m$). Technical Report R11592, TNO, Utrecht, Netherlands. Available at: https://www.dutchoffshorewindatlas.nl/binaries/dowa/documents/reports/2019/01/14/tno-report-offshore-wind-resource/TNO+2018+R11592+Profiles_rev1.pdf; accessed 21 January 2020.

ECMWF (2016). Data assimilation, in IFS Documentation – Cy41r2,. ECMWF, Reading, UK.

Gonzalez Rouco, F., García Bustamante, E., Hahmann, A.N., Karagili, I., Navarro, J., Tobias Olsen, B., Sïle, T. and Witha, B. (2019). NEWA Report on uncertainty quantification (Deliverable D4.4). Available at: https://doi.org/10.5281/zenodo.3382572; accessed 21 January 2020.

Haarsma, R.J., Hazeleger, W., Severijns, C., De Vries, H., Sterl, A., Bintanja, R., Van Oldenborgh, G.J. and van den Brink, H.W. (2013) More hurricanes to hit western Europe due to global warming. *Geophysical Research Letters*, 40, 1783–1788.

Hahmann, A.N. (2020) The making of the new European Wind Atlas, Part 1: model sensitivity. *Geoscientific Model Development Discussions*. (personal communication, 5th January 2020).

Hewitt, S., Margetts, L. and Revell, A. (2018) Building a digital wind farm. *Archives of Computational Methods in Engineering*, 25, 879–899.

Hogan, R.J., O'Connor, E.J. and Illingworth, A.J. (2009) Verification of cloud-fraction forecasts. *Quarterly Journal of the Royal Meteorological Society*, 135, 1494–1511.

Holtslag, A.A.M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A.C.M., Bosveld, F.C., Cuxart, J., Lindvall, J., Steeneveld, G.J., Tjernström, M. and Van De Wiel, B.J.H. (2013) Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. *Bulletin of the American Meteorological Society*, 94, 1691–1706.

Jammalamadaka, S.R. and Sengupta, A. (2001) *Topics in Circular Statistics*. World Scientific, Singapore.

Kalverla, P.C., Duncan, J.B., Steeneveld, G.-J. and Holtslag, A.A.M. (2019a) Low-level jets over the North Sea based on ERA5 and observations: together they do better. *Wind Energy Science*, 4, 193–209.

Kalverla, P.C., Steeneveld, G., Ronda, R., Holtslag, A.A.M. and M, A.A. (2019b) Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast IJmuiden in the North Sea. *Wind Energy*, 22, 34–48.

Kalverla, P.C., Steeneveld, G.-J., Ronda, R.J. and Holtslag, A.A.M. (2017) An observational climatology of anomalous wind events at offshore meteomast IJmuiden (North Sea). *Journal of Wind Engineering and Industrial Aerodynamics*, 165, 86–99.

Muñoz-Esparza, D., Kosović, B., Mirocha, J. and van Beeck, J. (2014) Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models. *Boundary-Layer Meteorology*, 153, 409–440.

Olauson, J. (2018) ERA-5: the new champion of wind power modelling?. *Renewable Energy*, 126, 322–331.

Olson, J.B., Kenyon, J.S., Angevine, W., Brown, J.M., Pagowski, M. and Sušelj, K. (2019a) *A description of the MYNN-EDMF scheme and the coupling to other components in WRF-ARW*. NOAA Technical Memorandum OAR GSD-61, ESRL, Boulder, CO.

Olson, J.B., Kenyon, J.S., Djalalova, I., Bianco, L., Turner, D.D., Pichugina, Y., Choukulkar, A., Toy, M.D., Brown, J.M., Angevine, W.M., Akish, E., Bao, J.-W., Jimenez, P., Kosovic, B., Lundquist, K.A., Draxl, C., Lundquist, J.K., McCaa, J., McCaffrey, K., Lantz, K., Long, C., Wilczak, J., Banta, R., Marquis, M., Redfern, S., Berg, L.K., Shaw, W. and Cline, J. (2019b) Improving wind energy forecasting through numerical weather prediction model development. *Bulletin of the American Meteorological Society*, 100, 2201–2220.

Palutikof, J.P., Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999) A review of methods to calculate extreme wind speeds. *Meteorological Applications*, 6, 119–132.

Parker, W.S. (2016) Reanalyses and observations: what's the difference?. *Bulletin of the American Meteorological Society*, 97, 1565–1572.

Parsons, M.J., Crosby, A.R., Orelup, L., Ferguson, M. and Cox, A.T. (2018). Evaluation of ERA5 reanalysis wind forcing for use in ocean response modeling. In: Waves in Shallow Environments (WISE) conference, 22–26 April 2018, Tel Aviv, Israel. Available at: https://oceanweather.com/news/_site/2018/04/20/era5-evaluation-at-2018-wise.html

Petersen, E.L., Troen, I., Ejsing Jørgensen, H. and Mann, J. (2014) The New European Wind Atlas. *Energy Bulletin*, 17, 34–39.

Poveda, J.P.M. and Wouters, D. (2015) *Wind Measurements at Meteorological Mast IJmuiden*. ECN, Petten, Netherlands. https://publicaties.ecn.nl/PdfFetch.aspx?nr=ECN-E--14-058

Powers, J.G., Klemp, J.B., Skamarock, W.C., Davis, C.A., Dudhia, J., Gill, D.O., Coen, J.L., Gochis, D.J., Ahmadov, R., Peckham, S.E., Grell, G.A., Michalakes, J., Trahan, S., Benjamin, S.G., Alexander, C.R., Dimego, G.J., Wang, W., Schwartz, C.S., Romine, G.S., Liu, Z., Snyder, C., Chen, F., Barlage, M.J., Yu, W. and Duda, M.G. (2017) The Weather Research and Forecasting model: overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, 98, 1717–1737.

Sandu, I., Beljaars, A. and Balsamo, G. (2014) Improving the representation of stable boundary layers. *ECMWF Newsletter*, 138, 24–29.

Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T. and Balsamo, G. (2013) Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models?. *Journal of Advances in Modeling Earth Systems*, 5, 117–133.

Sanz Rodrigo, J., Chávez Arroyo, R.A., Moriarty, P., Churchfield, M., Kosović, B., Réthoré, P.-E., Hansen, K.S., Hahmann, A., Mirocha, J.D. and Rife, D. (2017) Mesoscale to microscale wind farm flow modeling and evaluation. *Wiley Interdisciplinary Reviews: Energy and Environment*, 6(2). https://doi.org/10.1002/wene.214

Schaefer, J.T. (1990) The critical success index as an indicator of warning skill. *Weather and Forecasting*, 5, 570–575.

Smedman, A.-S., Högström, U. and Bergström, H. (1996) Low level jets – A decisive factor for off-shore wind energy siting in the Baltic Sea. *Wind Engineering*, 20, 137–147.

Steeneveld, G.-J. (2014) Current challenges in understanding and forecasting stable boundary layers over land and ice. *Frontiers in Environmental Science*, 2, 41. https://doi.org/10.3389/fenvs.2014.00041

Tastula, E.-M., Vihma, T. and Andreas, E.L. (2012) Evaluation of polar WRF from modeling the atmospheric boundary layer over Antarctic Sea ice in autumn and winter. *Monthly Weather Review*, 140, 3919–3935.

Tsiringakis, A., Steeneveld, G.J. and Holtslag, A.A.M. (2017) Small-scale orographic gravity wave drag in stable boundary layers and its impact on synoptic systems and near-surface

meteorology. *Quarterly Journal of the Royal Meteorological Society*, 143, 1504–1516.

Valkonen, T., Vihma, T., Johansson, M.M. and Launiainen, J. (2014) Atmosphere–sea ice interaction in early summer in the Antarctic: evaluation and challenges of a regional atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, 140, 1536–1551.

Wagner, D., Steinfeld, G., Witha, B., Wurps, H. and Reuder, J. (2019) Low-level jets over the southern North Sea. *Meteorologische Zeitschrift*, 28, 389–415. https://doi.org/10.1127/metz/2019/0948

Witha, B., Dörenkämper, M., Frank, H., García-Bustamante, E., González-Rouco, F., Navarro, J., Schneider, M., Steeneveld, G.-J. and Svensson, N. (2019a) The NEWA ferry lidar benchmark: comparing mesoscale models with lidar measurements along a ship route. https://doi.org/10.5281/zenodo.3372693

Witha, B., Hahmann, A., Sile, T., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J.F., Leroy, G. and Navarro, J. (2019b) WRF model sensitivity studies and specifications for the NEWA mesoscale wind atlas production runs. https://doi.org/10.5281/zenodo.2682604