Response to Referee comments — Second review

August 27, 2020

Thank you once again for very helpful comments.
The reviewers' comments are in black and our responses in blue.

The manuscript has improved substantially from the initial submission, and I would be more than happy to see it published. Especially the extensive tables with scores for each configuration are an excellent resource for future reference. I'm also happy with the authors' commitment to making (an excerpt of) the data publicly available. Thank you. The summary statistics from nearly all (a single run is missing due to corrupted data) simulations are now available from Zenodo, DOI: 10.5281/zenodo.4002351.

1. I just have a few small remarks that the authors might still want to address:

   - P4 L22 "not warranted" I would suggest to phrase this is bit more carefully. Agreed. The new sentence reads "Alternatively, a single continuous run can be performed. Although the continuous simulation may present certain advantages, like preserving the memory of land-atmosphere processes (Jimenez et al 2011), it is not necessarily more accurate. Additionally, this type of simulation increases the wall clock time required to complete the runs by at least an order of magnitude and thus make long-term high-resolution simulations not viable."

   - P12 L12: It would be good to end this section with a concluding sentence (e.g. move/use P13 L3). Agreed. Two sentences have been added before and after the start of the new section. "Summarising, the effect on the wind speed when one of those PBL schemes is replaced with another is nearly insensitive to the location of the domain (i.e., North versus South)" before and "We evaluate the mean wind profiles from the four model setups in Sec. 5.1 to observations in the NW domain to proceed in determining a suitable model configuration. " at the start of the next section.

   - P14 L4: It would be clearer to start this section with something like: "Having set our baseline configuration, we now start...". Currently you write "we start" while the reader feels like we've started 3 pages ago. Agreed. It has been changed to "Having set our baseline configuration, we now start testing other possible modifications to the model setup and parameterisation options. "

   - Colormaps: the colormaps are better than before, but sometimes the use of a diverging colormap for a continuous model score is still slightly confusing.

For a diverging score like bias, white is best, while for a continuous score like RMSE, purple is best. I think I would only use the brown part of the colormap here. We suggest a compromise here. The colour table for the biases is now brown for large positive and negative values and purple for values close to zero.

2. And some technical corrections:

   - P2 L20 "in two parts"? Done

   - P4 L21 "type of continuous ... may" Done

   - P4 L22 "its is not". Done

   - P9 L23 "the ERA-Interim ..." ? Done

   - P18 L17 difficult to read/wrong use of commas. Done

   - P19 L24 "similar than" Done

The Making of the New European Wind Atlas, Part 1: Model Sensitivity

Andrea N. Hahmann[1], Tija Sīle[2], Björn Witha[3,4], Neil N. Davis[1], Martin Dörenkämper[5], Yasemin Ezber[6], Elena García-Bustamante[7], J. Fidel González-Rouco[8,9], Jorge Navarro[7], Bjarke T. Olsen[1], and Stefan Söderberg[10,11]

[1]Wind Energy Department, Technical University of Denmark, Roskilde, Denmark
[2]Institute of Numerical Modelling, Department of Physics, University of Latvia, Riga, Latvia
[3]ForWind, Carl von Ossietzky University Oldenburg, Germany
[4]energy & meteo systems GmbH, Oldenburg, Germany
[5]Fraunhofer Institute for Wind Energy Systems, Oldenburg, Germany
[6]Eurasia Institute of Earth Sciences, Istanbul Technical University, Istanbul, Turkey
[7]Wind Energy Unit, CIEMAT, Madrid, Spain
[8]Dept. of Earth Physics and Astrophysics, University Complutense of Madrid, Madrid, Spain
[9]Institute of Geosciences, IGEO (UCM-CSIC), Madrid, Spain
[10]WeatherTech, Sweden
[11]Renewable Energy Analytics, DNV-GL Energy, Sweden

Correspondence: Andrea N. Hahmann (ahah@dtu.dk)

Abstract. This is the first of two papers that document the creation of the New European Wind Atlas (NEWA). It describes the sensitivity analysis and evaluation procedures that formed the basis for choosing the final setup of the mesoscale model simulations of the wind atlas. The suitable combination of model setup and parameterisations, bound by practical constraints, was found for simulating the climatology of the wind field at turbine-relevant heights with the Weather Research and Forecasting (WRF) model. Initial WRF model sensitivity experiments compared the wind climate generated by using two commonly used planetary boundary layer schemes and were carried out over several regions in Europe. They confirmed that the most significant differences in annual mean wind speed at $100\,\mathrm{m}$ above ground level mostly coincide with areas of high surface roughness length and not with the location of the domains or maximum wind speed. Then an ensemble of more than 50 simulations with different setups for a single year was carried out for one domain covering Northern Europe, for which tall mast observations were available. We varied many different parameters across the simulations, for example, model version, forcing data, various physical parameterisations and the size of the model domain. These simulations showed that although virtually every parameter change affects the results in some way, significant changes on the wind climate in the boundary layer are mostly due to using different physical parameterisations, especially the planetary boundary layer scheme, the representation of the land surface, and the prescribed surface roughness length. Also, the setup of the simulations, such as the integration length and the domain size, can considerably influence the results. We assessed the degree of similarity between winds simulated by the WRF ensemble members and the observations using a suite of metrics, including the Earth Mover's Distance (EMD), a statistic that measures the distance between two probability distributions. The EMD was used to diagnose the performance of each ensemble member using the full wind speed and direction distribution, which is essential for wind resource assessment. We identified the most

realistic ensemble members to determine the most suitable configuration to be used in the final production run, which is fully described and evaluated in the second part of this study (Dörenkämper et al., 2020).

Copyright statement.

1 Introduction

Wind atlases are defined as databases of wind speed and direction statistics at several heights in the planetary boundary layer. Wind atlases have been created for many regions of the world to help inform wind energy installations, but many other human activities benefit from the knowledge of the wind behaviour at its climatology. For example, wind atlases can be used in structural design for buildings, transportation infrastructure and operation, recreation and tourism. In 1989, the European Wind Atlas (EWA, Troen and Petersen, 1989) was released, which provided the first public source of wind climate data that covered the whole of Europe in a homogeneous way. The EWA was based on data from around 200 meteorological stations, and used the so-called *wind atlas method*, a collection of statistical models that are the core of the Wind Atlas Analysis and Application Program (WAsP) software package (Mortensen et al., 2011). This method allows for the transfer of the wind climate statistics from one location to another based on the surface characteristics. The EWA is now 30 years old and of limited usefulness. It has a very coarse spatial resolution, does not provide time series of the variables of interest, and was developed using a linearised method (Troen and Petersen, 1989), which limited its applicability in complex terrain.

In 2013, a team of 30 partners from 8 European countries started work on the "New European Wind Atlas" (NEWA) project (Petersen, 2017). The project had three main components: a series of intensive measuring campaigns, a thorough examination and redesign of the model chain (downscaling from global to mesoscale and to microscale models), and the creation of the wind atlas database. The scope of the modelling in NEWA is ambitious and too long for a single article. Therefore it is divided into two parts~~in this paper~~. The first one (this article) deals with sensitivity simulations and the second part (Dörenkämper et al., 2020) describes the production of the new database[a] and its evaluation.

Nowadays, wind atlases rely on the output from mesoscale model simulations, either sampling recurrent atmospheric states (e.g., Frank and Landberg, 1997; Pinard et al., 2005; Badger et al., 2014; Chávez-Arroyo et al., 2015) or by long-term simulations with Numerical Weather Prediction models (e.g., Tammelin et al., 2013; Nawri et al., 2014; Hahmann et al., 2015; Draxl et al., 2015; Wijnant et al., 2019). NEWA follows the latter approach and provides a unified high-resolution and publicly available dataset of wind resource parameters covering all of Europe and Turkey. The wind atlas is based on 30 years of mesoscale model simulations with the Weather Research and Forecasting (WRF, Skamarock et al., 2008) model at 3 km × 3 km spatial and 30 min temporal resolution and 7 vertical levels. The WRF model was chosen because of its open access, it is used by many wind energy research institutes and private companies, and it was familiar to all the NEWA partners. Wind statistics from further downscaling with a microscale model (see Dörenkämper et al., 2020, for more details) are provided for

---

[a]https://map.neweuropeanwindatlas.eu/

Europe and Turkey onshore and offshore up to 100 km from the coastline, plus the Baltic and the North Seas with a horizontal grid spacing of 50 m at three wind-turbine relevant heights.

Mesoscale models are, in general, not specifically developed for wind energy applications; however, over the last decade they have been extensively used for that purpose (see Olsen et al., 2017, for a review). Developing an optimal WRF model configuration for wind resource assessment is not a straightforward task, considering the large number of degrees of freedom in the model configuration, and the different choices of input data. Among the configuration options offered in the WRF model are, physical parameterisations such as planetary boundary layer (PBL), surface layer (SL), land surface model (LSM), cloud micro-physics, and radiation. Also numerical and technical options (e.g., domain layout, nudging options, time step), and the initial and boundary conditions of the atmosphere, sea surface, and land surface are relevant aspects to be explored before determining the set up that better fits a specific application. Arguably, an optimal configuration that performs best at all time and spatial scales cannot be expected, and we search herein for a configuration that tends to perform better at most instances within the ensemble of sensitivity experiments performed.

A large number of ensembles of model simulations using the WRF model are documented in the literature. The WRF model is also used for more general climate research purposes as a regional climate model (RCM), for example, within the context of the Coordinated Regional Climate Downscaling Experiment (CORDEX) project (Katragkou et al., 2015). However, in such cases the attention is typically focused on climate-relevant parameters, such as temperature or precipitation. A number of studies have reported sensitivity to the dependence of model-simulated temperature and precipitation on cloud microphysics, convection and radiation schemes (Katragkou et al., 2015), or PBL schemes (García-Díez et al., 2013) and all of the above (Strobach and Bel, 2019). Sensitivity studies with a large number of WRF model simulations for wind energy applications have also been reported in Lee et al. (2012); Siuta et al. (2017); Fernández-González et al. (2017); Fernández-González et al. (2018) for PBL-scheme ensemble prediction. Very few studies afforded an exhaustive sensitivity analysis of the model performance on the near-surface long-term wind climatology and many lack the verification at wind turbine heights. Two examples that looked at the sensitivity of the modelled wind climate are Hahmann et al. (2015), that investigated a limited number of model parameters over the sea in Northern Europe and Floors et al. (2018b), that concentrated on the impact of the model's spatial resolution on the atmospheric flow in the coastal region. This study expands on these earlier attempts with a much larger set of sensitivity simulations and the comparison to a limited set of observations. It is impossible to test every combination of the WRF model setup and possible parameterisations, as the number of such experiments would be in the thousands, which is unfeasible in terms of computational resources. Therefore, a compromise between available computational power and scientific soundness had to be found. The approach in NEWA was to first define a "best practice" setup using the vast and diverse experience of the mesoscale modellers in the project, and then to test the sensitivity of the results to changes in the model configuration that are presumably the most relevant for the simulation of the wind field. This includes some physical options, such as PBL schemes, but also included a wide range of other parameters, such as numerical options, for which sensitivity results are rarely reported in the literature. Those parameters that did not show an impact in the simulation of the wind field were fixed as in the best practice set up. It is impossible to claim that all existing sensitivities in the model were found and tested. However the large number of parameters that were tested and found not to be influential gives some credibility to our approach.

All simulations in this study covered one full year and used similar grid parameters and modelling setup, which will be briefly described in Section 2. The data used for the evaluation of the ensemble of simulations among the whole pool of cases that are best suited to provide a meaningful sensitivity range is presented in Section 3; the statistics used in the model assessment and comparison among ensemble members is introduced in Section 4. The process of finding the most adequate (in a sense that will be defined) combination of model setup and parameterisations occurred in several phases: (1) analysis of sensitivity dependence on the geographical domain (Section 5.1), (2) selection of the WRF model version (Section 5.3), (3) creation of a large multi-physics ensemble (Sections 5.4 and 5.5) and (4) the analysis of the model sensitivity to the size of the model domain (Section 5.6). A summary of the findings of the sensitivity experiments can be found in Section 5.7. The paper ends with a discussion of the limitations of the approach used and the outlook (Section 6).

2   Description of the WRF model simulations

The database of simulated winds and wind-energy relevant parameters for the model sensitivity tests was created by splitting the simulation period into a series of relatively short WRF model runs that, after concatenation, cover at least a year. The simulations overlap in time during the spin-up period, typically between 3 to 24 h, which is discarded, as described in Hahmann et al. (2010, 2015); Jiménez et al. (2010, 2013). Two approaches are tested: frequent re-initialisation, in our case daily 36-hour runs, versus several days long runs that are nudged towards the forcing reanalysis. In the first approach, the re-initialisation every day keeps the runs close to the driving reanalysis and the model solution is free to develop its own internal variability. In the second approach, the use of nudging prevents the model solution from drifting from the observed large-scale atmospheric patterns, and the multi-days simulation ensures that the mesoscale flow is fully in equilibrium with the mesoscale characteristics of the terrain (Vincent and Hahmann, 2015). Both methods have the added advantage that the simulations are independent of each other, and therefore, can be computed in parallel, reducing the total time needed to complete a multi-year climatology. Alternatively, a single ~~run of a continuous model experiment~~ continuous run can be performed. Although ~~this type of continuous~~ the continuous simulation may present certain advantages, like preserving the memory of land-atmosphere processes (e.g. Jiménez et al., 2011), ~~its is not warranted more accurate and it can significantly expand~~ it is not necessarily more accurate. Additionally, this type of simulation increases the wall clock time ~~and~~ required to complete the runs by at least an order of magnitude and thus make long-term high-resolution simulations not viable.

All mesoscale simulations in NEWA used three nested domains with a 3 km horizontal grid spacing for the innermost grid and a 1:3 ratio between inner and outer domain resolution, leading to 3 different resolutions: 27 km for the outer domain, and 9 km and 3 km for the inner nested domains. The model top was set to 50 hPa, following the best practices recommended by the WRF developers (Wang et al., 2019). The temporal coverage of the sensitivity simulations is one year (2015 or 2016), based on data availability. Other parameters common to all simulations are listed in Table A1. We explore the effect of changing various relevant parameters of the simulation set up from the base model configuration explained above to estimate the wind climatology over Europe.

3 Observed data

High-quality data from tall masts for the evaluation of the various sensitivity experiments is rare. In this study we used data from eight sites in northern Europe. The locations and names of the sites are shown in Fig 1; the details of the sites are summarised in Table 1. All sites are equipped with towers, and IJmuiden has an additional Zephir 300 continuous-wave lidar recording wind

5   speed and direction at 25 m intervals between 90 and 315 m (Kalverla et al., 2017).
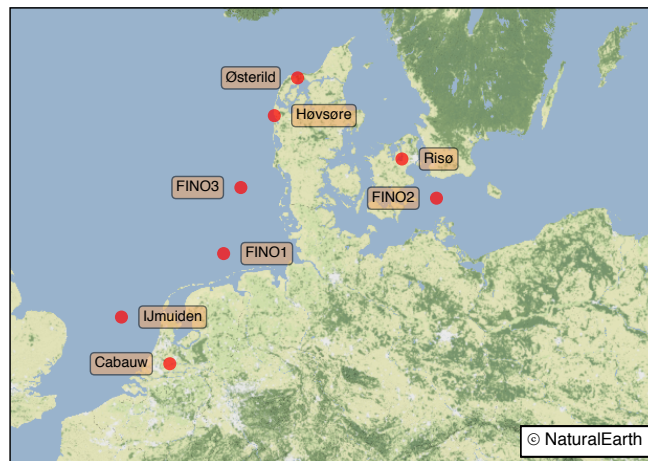
Figure 1. Location and name of the tall mast/lidar sites used in the model evaluation. Base map created with Natural Earth.

High quality data is essential for model evaluation (Lucio-Eceiza et al., 2018a, b). The mast data has been quality controlled and the mast flow distortion on the wind speed estimates was minimised by sub-setting the data. At FINO1, FINO2, Risø and Høvsøre, where wind speed measurements are available from only one boom, winds originating from $\pm 10°$ of the boom direction are filtered. A misalignment correction to the wind direction was applied to the wind vanes at FINO1 (Westerhellweg

10   et al., 2012). At FINO3, we use the data from the three heights where wind speed measurements from three booms are available. The wind speed value is taken from the boom where the wind direction is most perpendicular to the boom direction. At IJmuiden, the data was processed as discussed in Kalverla et al. (2017). At Cabauw, the data was processed and gap filled as described in Bosveld (2019).

In addition to mast flow distortion, the wind speed estimates at some of the measurement sites are impacted by nearby wind

15   farms. At Høvsøre and Østerild, test turbines are located north of the mast, in the sector with the least frequent wind directions. At FINO1, the wind farm *Alpha Ventus* impacts the eastern sector with the nearest turbine only 405 m away in the direction of 90°. The wind farm *EnBW Baltic 2* started operation in September 2015 and lies directly to the southeast of the FINO2 mast. At FINO3, the *DanTysk* wind farm went into operation in December 2014 to the east of the mast. Due to the difficulty in understanding the impacts of the wind turbines on the measurements (e.g., times of operation at the onset can vary due to

20   adjustments and testing), the data has not been filtered or corrected for the turbine wakes. However, the presence of the wind farm can impact the evaluation of the model results and should be kept in mind.

5

Table 1. Tall mast sites and wind measurement (WS: wind speed, WD: wind direction) heights available at each site. The values indicated in bold are used in the evaluation of wind profiles; the values underlined are those used in the evaluation of temporal variability. The centre of the sector where the mast can cause flow distortion and influence the wind speed is also listed. The last three columns show the data availability for 2015 for profiles, time-series and wind direction, respectively.

site	type	measurement heights (m AGL/AMSL)	flow distortion centre (°)	data availability (%)		
				profile	time series	wind dir
FINO1[a]	Offshore	WS: 32.8, **40.3**, **50.3**, **60.3**, **70.3**, **80.3**, **90.3**, 101.2	315	78.9	82.2	88.3
		WD: 40.7, 50.3, 60.7, 70.3, 80.7, 90.3				
FINO2[a]	Offshore	WD: 32.4, **42.4**, **52.4**, **62.4**, **72.4**, **82.4**, **92.4**, 102.5	15	77.7	82.5	97.8
		WD: 31.8, 51.8, 71.8, 91.8				
FINO3[a]	Offshore	WS: 30.5, 40.5, **50.5**, 60.5, **70.5**, 80.5, **90.5**, 100, 106	–	97.3	97.3	97.6
		WD: 60.5, 100.5				
IJmuiden[b]	Offshore	WS: **27**, **58**, **89**, **115**, **140**, **165**, **190**, **215**, **240**, **265**	–	81.5	98.5	88.5
		WD: 27, 58, 89, 115, 140, 165, 190, 215, 240, 265				
Høvsøre[c]	Coastal	WS: 10, **40**, **60**, **80**, **100**, **116.5**	0	96.6	96.8	96.9
		WD: 10, 60, 100				
Risø[d]	Land	WS: **44.2**, **76.6**, **94**, **118**, **125.2**	225	90.9	94.1	100.
		WD: 76.5, 94.				
Østerild[e]	Land	WS: 10, **40**, **70**, **106**, **140**, **178**, **210**, **244**	0	78.3	78.3	78.6
		WD: 40, 244				
Cabauw[f]	Land	WS: 10, **20**, **40**, **80**, **140**, **200**	–	99.1	100.	100.
		wind direction: 10, 20, 40, 80, 140, 200				

[a] https://www.fino-offshore.de/en/, [b] Kalverla et al. (2017), [c] Peña et al. (2015), [d] http://rodeo.dtu.dk/rodeo/ProjectOverview.aspx?&Project=5&Rnd=674271, [e] Peña (2019)
[f] http://www.cesar-database.nl

All measurement data is 10 min mean values. It was filtered to one period per hour, using the period that was closest to 00 minutes. The additional filtering for mast flow distortion process removes a small number of samples. For the time-series evaluation, we used measurements from the level that was closest to 100 m above mean sea level and had a good data availability (underlined values, Table 1). The choice of the wind direction height for the model evaluation and data filtering uses the same procedure.

4 Model evaluation metrics

The main goal of the NEWA sensitivity study was the evaluation of the wind climate, which is usually understood as the probability distribution of wind speed and direction at a specific point. Thus, we used several metrics to evaluate the accuracy of the model simulations when compared to tall mast observations tailored to this purpose.

We calculate the temporal mean of each modelled distribution, $\overline{u}_m$, and the observed distribution, $\overline{u}_o$, for identical time periods. The bias herein is defined as difference between the two means, $\overline{u}_m - \overline{u}_o$. If the bias is positive, the model overestimates the observed wind speed.

The bias is a popular error statistic for comparing the wind speed distributions between observations and model-simulated fields. However, the shape of the wind speed distribution is more important in wind energy applications, because the wind power density is proportional to the cube of the wind speed. Thus, small changes in the wind speed distribution are amplified when converted to power. Accordingly, we introduce the Earth Mover's Distance to evaluate the differences in the shape of two frequency distribution. The EMD, also known as the first Wasserstein distance, is popular in image processing (EMD; Rubner et al., 2000). The EMD can be interpreted as the amount of physical work needed to move a pile of soil in the shape of one distribution to that of another distribution. More discussion about the EMD properties can be found in Lupu et al. (2017). For one-dimensional distributions the EMD is equivalent to the area between two cumulative distribution functions, and, this interpretation with slight modifications, can be applied also to circular variables (Rabin et al., 2008). The EMD was calculated using the Pyemd package (Pele and Werman, 2008).
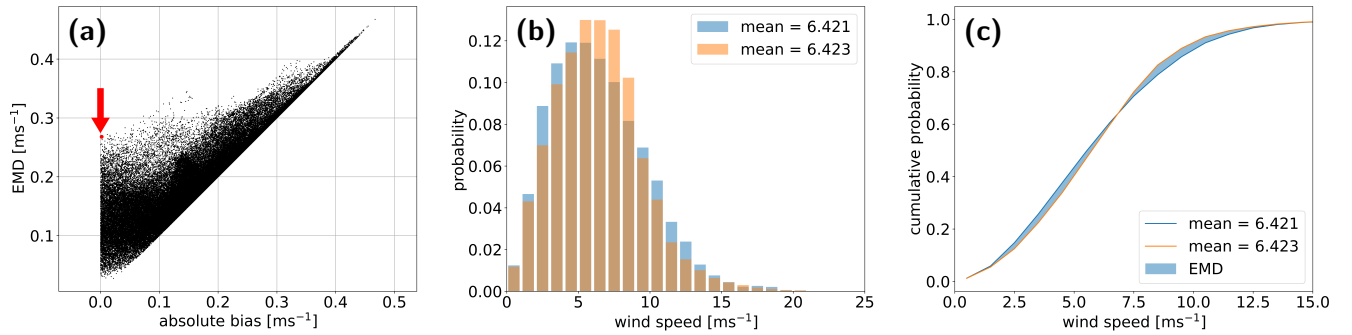
Figure 2. (a) Relationship between EMD values and absolute value of bias between the wind speed in two WRF model setups for all grid-points in the domain. The red dot represents an example where the EMD and the absolute value of bias are different. (b) Wind speed distributions of the two ensemble members corresponding to the red dot. (c) Cumulative probability distribution together with the EMD value calculated as the area between them. Distributions are the same as in panel (b).

Figure 2 illustrates the differences between the EMD and the absolute value of the bias. Figure 2a shows the relationship between the EMD and the absolute value of bias for two WRF model simulations for all points in the domain. Figure 2b shows modelled wind speed distributions for two separate grid points. This highlights the case where two differently shaped distributions can have the same mean. The EMD helps clarify such occurrences. For one-dimensional distributions the EMD can be calculated as the area between the cumulative distribution curves, as illustrated in Figure 2c.

The circular EMD (CEMD; Rabin et al., 2008) extends the EMD concept to one-dimensional circular histograms, such as the frequency distribution of wind directions. Two examples of the value of the CEMD are given in Fig. 3 for FINO3 and FINO1. At FINO3, the observed and simulated wind direction distributions are very similar and the value of CEMD is 2.18°,
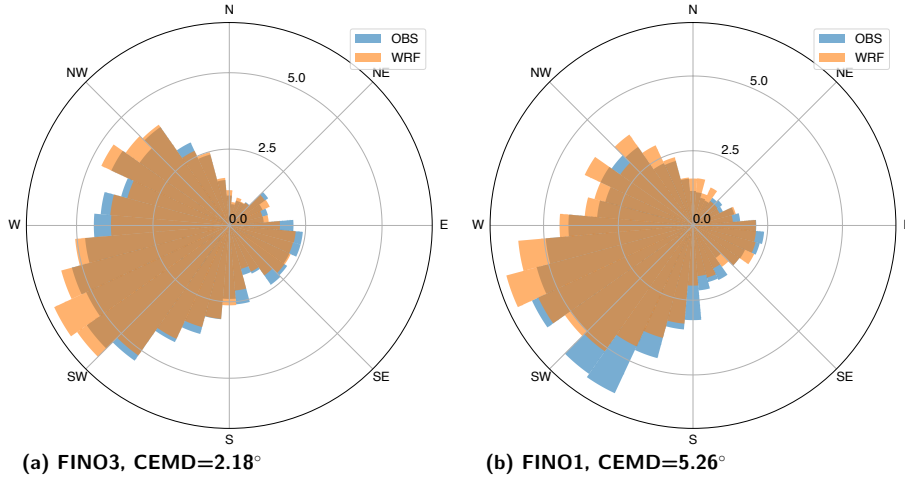
Figure 3. Wind direction distributions for (a) FINO3 and (b) FINO1 and using measurements (OBS) and model simulated (WRF) data.

mainly due to differences in frequency within the same sector. A higher value of CEMD (=5.28°) is obtained when a rotation in the wind direction is found between the observed and simulated wind directions. The CEMD is used throughout the paper, and is to our knowledge the first time that it is used for evaluating wind directions in meteorology or climate science.

The information about temporal co-variability is provided herein by root mean square error (RMSE), which estimates of
5   systematic biases in model skill (von Storch and Zwiers, 1999). The RMSE is calculated over all i time steps, from

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(u_m^i - u_o^i)^2},\tag{1}$$

with u_o^i and u_m^i being the i-th modelled and observed values in the time series of length n.

In this study, we compared a large ensemble of WRF model setups against the observations at several sites. One of the ensemble members was designated to be the baseline or "BASE". The aim is to evaluate if a certain model set up from the
10   pool performs better or worse than the baseline configuration. For this purpose we define a general skill score (von Storch and Zwiers, 1999):

$$\text{SS} = 1 - M_j/M_B,\tag{2}$$

where M_j is the value of the metric for the j-th ensemble member and M_B is that of the baseline. The metric M can be the absolute value of the bias, the RMSE, EMD, or CEMD. If $\text{SS} > 0$ the ensemble member j "improves" the metric with
15   respect to the baseline case, if $\text{SS} < 0$ it "worsens" it. A value of $\text{SS} = 1$ means that the new simulation is perfect. The SS is easily understood and is applicable to all our evaluation metrics. However, when the BASE simulation evaluates extremely well against observations (e.g. when the bias is close to zero), the skill score can become very large. Therefore, the SS is a useful quantity for the RMSE, which is rarely close to zero, but can be misleading when used for the absolute bias or the EMD,

indicating large improvements when the differences in metrics themselves are small. Accordingly, we suggest using both SS and the original metric when interpreting the results.

5   Sensitivity analysis of WRF simulations

In this section, the results from the different sensitivity experiments are presented and discussed. These are grouped into six subsections: Section 5.1 presents the results from five different domains for a small number of experiments to see how the results differ depending on the simulated region; Section 5.2 evaluates the results from one of these domains against observations; Section 5.3 highlights the impact of the WRF version on the wind speed results by investigating four different versions of the WRF model; Section 5.4 presents the results from a 25 simulation sensitivity study addressing the impact of the SL, PBL and LSM schemes, including changes to the surface roughness length; Section 5.5 documents the impact of other parameterizations and forcing data; and finally Section 5.6 focuses on the impact of the size of the domain. It should be noted that these sections do not necessarily follow the chronological order that they were performed in the NEWA project, but provide a logical progression of the decisions taken during the project.

5.1   Sensitivity to geographical domain

In the initial stage of the evaluation of the WRF model setup, we designed four numerical experiments (top four simulations in Table 2) over five different regions in Europe (Fig. 4), mostly located near countries represented in the NEWA team. The experiments explore the impact of using different PBL schemes: the MYNN (Mellor and Yamada, 1982) or the YSU (Hong et al., 2004) and the effect of different initialisation strategy (short, S1, or weekly, W1, simulation length and exclude or include nudging). The YSU and MYNN schemes are non-local-K mixing and TKE 1.5 order closure models, respectively. They were chosen because they are popular among the WRF model users and have shown good skill in previous studies (Draxl et al., 2014; Hahmann et al., 2015). The main objective was to clarify whether the sensitivity of the mean wind speed to these changes is similar in different geographic regions or there were regional differences. All other settings were fixed. The simulations were carried out with the WRF model version 3.6.1, released in August 2014. The basic WRF model setup includes using the ERA-Interim reanalysis (Dee et al., 2011) for initial and boundary conditions, NCEP optimal interpolation sea surface temperature (SST, Reynolds et al., 2007) and 61 vertical levels. Other details are given in Table A1.
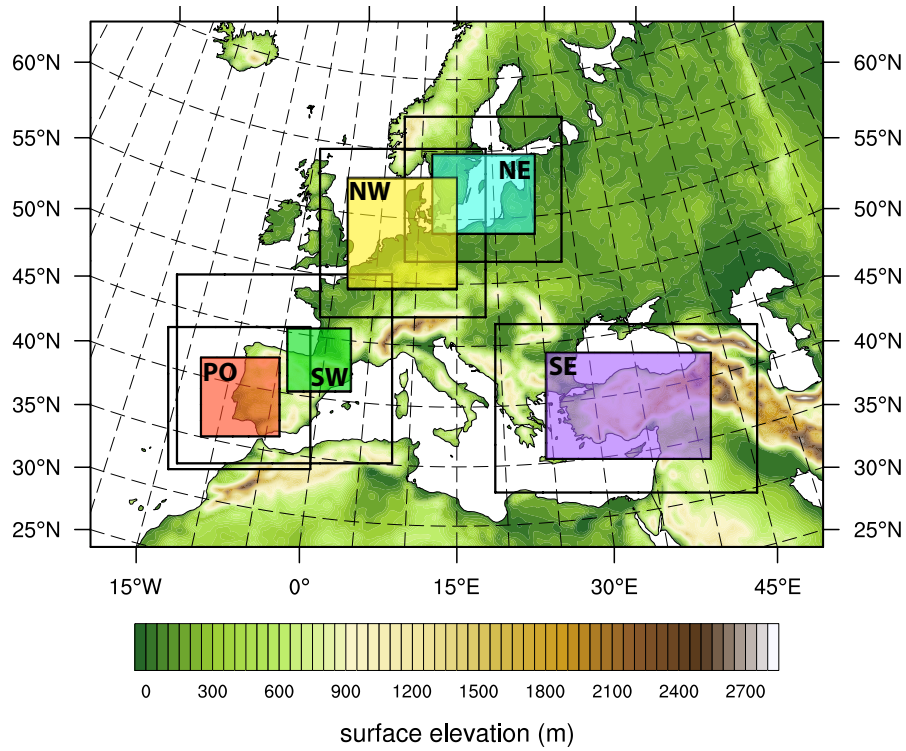
Figure 4. The location of the five inner domains (D3; NW, NE, SE, SW, and PO in coloured boxes) used in the geographic similarity experiments. The surface elevation of the WRF model outer domain (D1) is shown with colour; the black lines show the extent of D2. All inner domains share the same outer domain (D1), which corresponds to the area of the base map.

Table 2. Acronyms and relevant set up parameters of the WRF model sensitivity experiments oriented to address the influence of the geographic domain and WRF model version. The mod MYNN scheme is described in section 5.3.

experiment	WRF model version	PBL scheme	run length [d]	spin-up length [h]	nudging
MYNL61S1	3.6.1	MYNN	1.5	12	no
MYNL61W1	3.6.1	MYNN	8	24	yes
YSUL61S1	3.6.1	YSU	1.5	12	no
YSUL61W1	3.6.1	YSU	8	24	yes
MYNL61W1_V381	3.8.1	MYNN	8	24	yes
MYNL61W1_V381_MOD	3.8.1	mod MYNN	8	24	yes
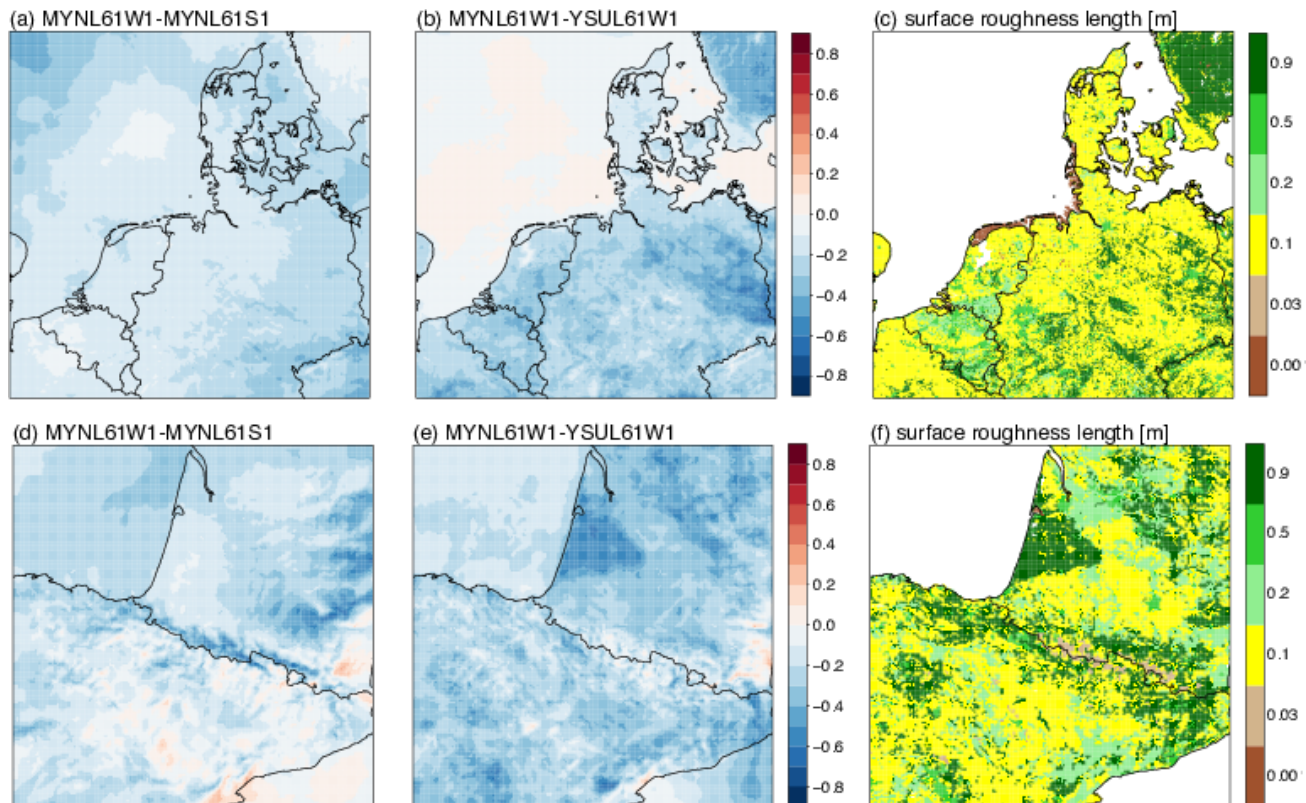
Figure 5. Annual mean difference in wind speed $[\mathrm{m\,s^{-1}}]$ between the simulations with: different initialisation strategy for the MYNN simulations (W1 minus S1; a and d), different PBL scheme (MYNN minus YSU; b and e), and the surface roughness length [m] used in the simulations (c and f). (a), (b) and (c) are for the NW domain, (d), (e) and (f) are for the SW domain shown in Fig. 4.

The analysis of the experiments presented in Fig. 5 show that on average the differences in annual mean wind are small and those that arise from the choice of PBL scheme are larger and more extensive than those from the initialisation strategy. The largest differences in wind speed that arise from the initialisation strategy (Fig. 5a and Fig. 5d) coincide with areas of elevated terrain in the western French Alps and the Pyrenees in the SW domain and the north-west corner of the NW domain. In these

5   areas the daily runs (S1) have on average stronger winds that the weekly nudged simulations (W1). The largest differences in wind speed that arise from the choice of PBL scheme (Fig. 5b and Fig. 5e) coincide with the regions with particularly large surface roughness length, namely forests in southern Sweden and south-western France. There, the experiment using the MYNN scheme provides wind speeds that are on average more than $0.5\,\mathrm{m\,s^{-1}}$ lower than in the experiment using the YSU scheme. Over the sea, no significant difference is seen in the NW domain, and only slight differences (less than $0.3\,\mathrm{m\,s^{-1}}$) exist

10   above the Atlantic and Mediterranean (SW domain).

11

Similar analysis was carried out for the other three domains in Fig. 4. All five domains show the same pattern of higher wind speed at 100 m for simulations using the YSU scheme over land, with the largest differences occurring over rougher terrain (e.g. forests). Over water, the differences are much smaller. The winds simulated using the YSU scheme are slightly higher than those simulated using the MYNN scheme, but only in regions dominated by unstable stratification (e.g. the Atlantic Ocean,

5   Mediterranean and Black Seas). In the nudged weekly simulations, winds speeds are overall reduced from those in the daily simulations, and to a higher magnitude over higher terrain and closer to the edge of the domain. Similar results were obtained in Hahmann et al. (2015) when the discarded spin-up time of the short simulations was varied. We hypothesise that as the model integration progresses, a balance is reached between the large-scale flow, the model physical parameterisations and the surface forcing supplied by the terrain elevation and surface roughness. This process takes some time and results in lower wind

10   speeds in the longer simulations. The effect is different on the northwest corner of the NW domain, which is the dominant inflow boundary, probably resulting from the nudging in the outer domain that is absent in the short runs. Summarising, the effect on the wind speed when one of those PBL schemes is replaced with another is nearly insensitive to the location of the domain (i.e., North versus South).

5.2   Evaluation against observations in the NW domain

15   We evaluate the mean wind profiles from the four model setups in Sec 5.1 to observations in the NW domain to proceed in determining a suitable model configuration. The evaluation of the mean wind speeds from the four WRF V3.6.1 sensitivity experiments against the mast measurements is shown in Fig. 6 and Fig. 7. The differences in mean wind speed between the various experiments are small. The mean wind profile from the MYNL61W1 simulation is closest to the observed profile at nearly all the sites, except at the offshore sites where the MYNL61W1 and YSUL61W1 are nearly indistinguishable from

20   each other. The evaluation metrics of the wind speed and direction for the experiments and sites are presented in Fig. 7. This confirms that the wind speeds from the MYNL61W1 simulation have: the lowest biases at Cabauw, Høvsøre, Østerild, and Risø; the lowest RMSE at all sites, except for FINO1 and IJmuiden, where the results from MYNL61W1 and YSUL61W1 are virtually the same. The EMD shows lowest values for the weekly nudged simulations compared to the short runs and for the coastal and land sites the lowest EMD is for the MYNL61W1 simulation. Similar conclusions apply to the CEMD of the wind

25   direction, where the YSUL61W1 performs the best at six of the eight sites.
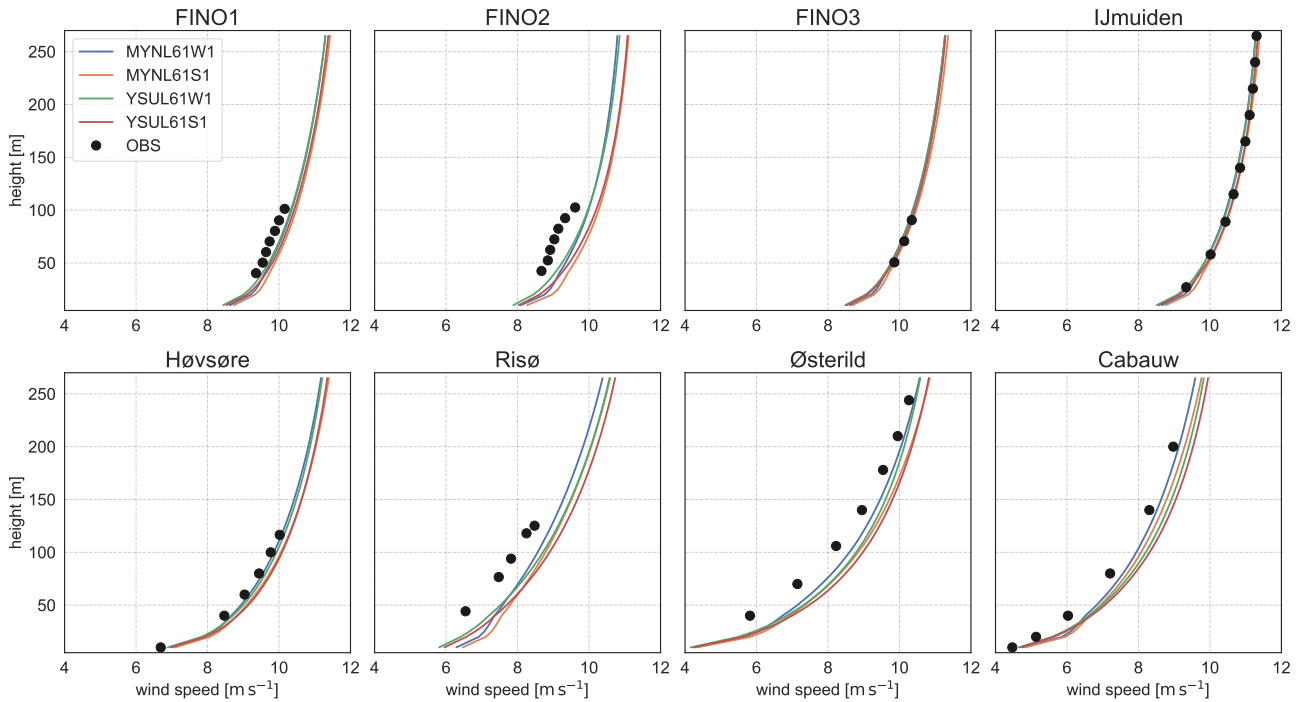
Figure 6. Comparison of the observed mean wind speed [m s^{-1}] as a function of height for the eight sites and the simulated mean wind speed from the four sensitivity WRF V3.6.1 runs in Table 2.

In summary, the weekly nudged simulations for both MYNN and YSU schemes result in lower biases and lower RMSE at all sites. Because the effect on the wind speed when one of those PBL schemes is replaced with another is nearly insensitive to the location of the domain (i.e. North versus South) as shown in Section 5.1, we continue under the assumption that it is valid for other regions in Europe, except for regions with more complex terrain, which have not been evaluated. The weekly nudged simulation setup was chosen for the remainder of the NEWA sensitivity simulations. The setup and constants used in the nudging will be re-evaluated in the sensitivity experiments in Section 5.5.

5.3 Sensitivity to properties of the MYNN scheme

We also tested the version of the WRF model used during the sensitivity analysis to evaluate whether changing the version implied any difference with respect to the baseline configuration described above. At the time of the development of this work, the latest version was WRF version 3.9.1, released in August 2017. The simulations using versions WRFV3.8.1 and WRFV3.9.1 use the same NW domain and the same model setup as in the MYNL61W1 simulations (Table 2). The results for WRFV3.6.1 and WRFV3.8.1 are presented in Fig. 7 and Fig. 8; the results for WRFV3.9.1 (not shown) are nearly identical to those from WRFV3.8.1.

The wind speed from simulations using WRFV3.8.1 presented an increased bias compared to observations for all sites and most levels, except for FINO1 and FINO2, which suffer from flow distortion and wind farm effects. The increased bias was

13
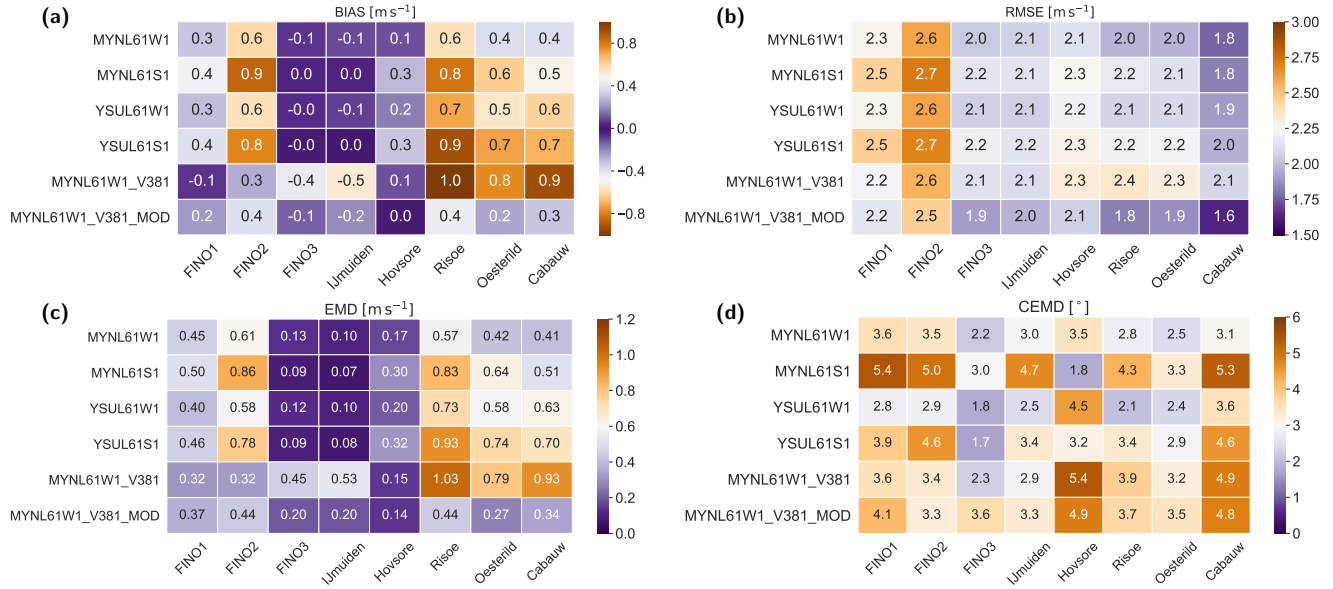
Figure 7. Evaluation metrics: (a) bias $[\mathrm{m\,s^{-1}}]$, (b) RMSE $[\mathrm{m\,s^{-1}}]$, (c) EMD $[\mathrm{m\,s^{-1}}]$ for the simulated wind speed and (d) CEMD $[°]$ for simulated wind direction for the eight sites and underlined heights in Table 1 and the sensitivity experiments in Table 2.

traced back to changes in two important equations in the MYNN SL and PBL scheme, which became defaults in WRF Version 3.7. The first is the scalar roughness length over water, which was changed from the formulation in Fairall et al. (2003) to that in Edson et al. (2013), thus affecting the wind speed over the ocean. The second is a change in the definition of the mixing length (Olson et al., 2016). Both of these options could be customised and set as in the baseline configuration. Results from such a

5   setup are labelled "WRFV3.8.1_MOD" in Fig. 8. Some of the previous characteristics of the profile are restored after these changes, and the simulation using WRFV3.8.1 with modifications improved the RMSE and EMD for all sites (Fig. 7). These changes are consistent with those found by Yang et al. (2017) for the MYNN scheme. Although above $\sim 100\,\mathrm{m}$, the simulation using the modified WRFV3.8.1 gives lower mean wind speeds than WRFV3.6.1 at all sites, we consider nonetheless that this difference is less relevant and based on the improvements of the scheme "WRFV3.8.1_MOD", this set up was selected as the

10   baseline, named "BASE" hereon. The WRF model setup of this baseline is summarised in Table A1. Unless explicitly labelled otherwise, when referring to the MYNN option in the remainder of this work we mean the modified version of the scheme. The unmodified MYNN PBL will be referred to as MYNN*.

5.4   Effect of surface and planetary boundary layer and land surface model

Having set our baseline configuration, we now start testing other possible modifications to the model setup and parameterisation

15   options. The first series of sensitivity studies tested the sensitivity of the near surface wind to various combinations of LSM, PBL, SL schemes and the specification of surface roughness length. The combinations tested are listed in Table 3. All the
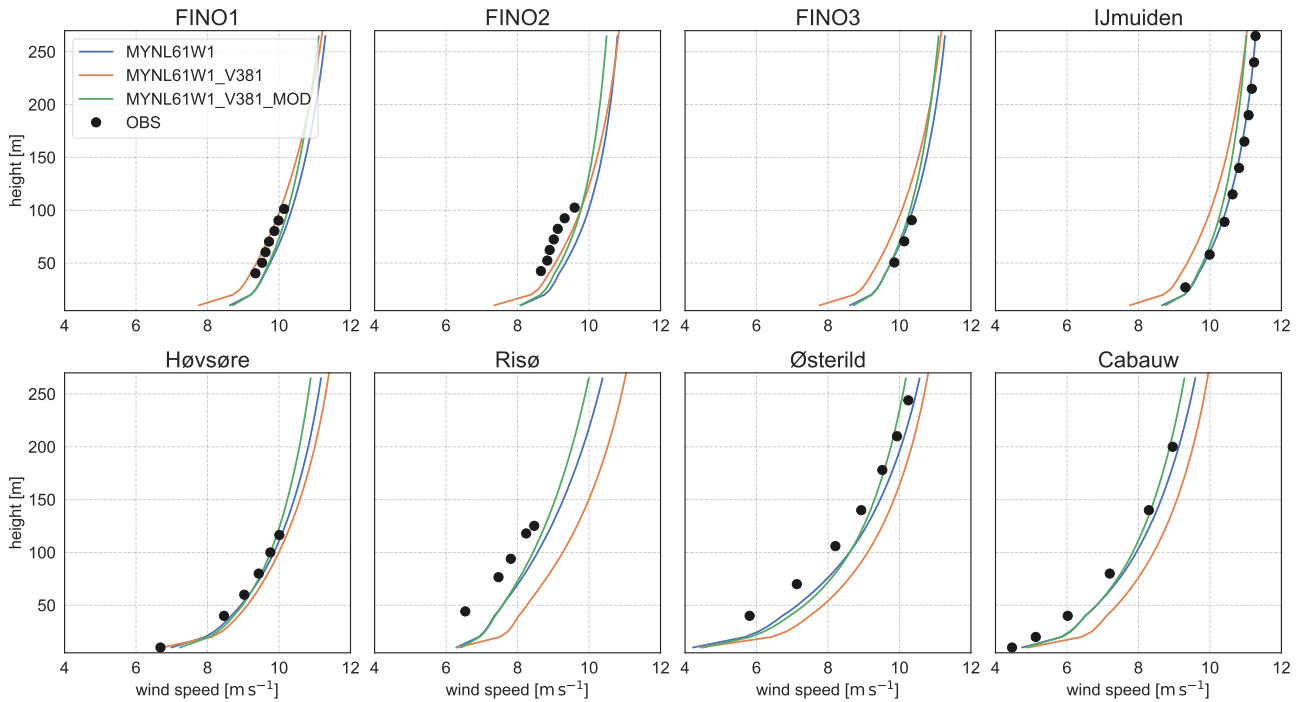
Figure 8. Comparison of the mean observed wind speed [m s^{-1}] as a function of height for the eight sites and the mean wind speed in the simulations using WRFV3.6.1, WRFV3.8.1 and the modified version of WRFV3.8.1_MOD described in Table 2.

simulations in this and most of those in Section 5.5 use the ERA5 reanalysis (Hersbach et al., 2020) as the source of initial and boundary conditions.

The large number of schemes and their potential combinations led to a large number of possible combinations. In this work a total of 25 different combinations were tested, listed in Table 3, including changes in parameters in the schemes themselves.

5 We also included the simulation using the MYNN* scheme. Our original table of sensitivity experiments contained many other LSM/PBL/SL combinations, but some of these suffered from diverse technical issues and did not complete the runs. In some cases, small adjustments were needed. For example, in the experiments using the MM5 or MO surface layer schemes the fractional sea-ice option had to be turned off.

An important aspect of the LSM/PBL/SL sensitivity studies was the use of an alternative look-up table for the surface

10 roughness length as a function of land-use class. A custom NEWA lookup table was created since many values used in the WRF-distributed tables do not match the aerodynamic characteristics of European vegetation, especially over forests (Hahmann et al., 2015; Floors et al., 2018a). The new lookup table was created by polling wind energy resource assessment experts from the NEWA consortium. Both the new and old values of surface roughness length for each roughness class are shown in Table 4. Some of the larger changes include, "Herbaceous Wetland", which has an original value of $z_0 = 0.20$ m, but in the NEWA

15 region represents the tidal zone in coastal Holland, Germany, and Denmark, which is much smoother (Wohlfart et al., 2018),

15

Table 3. Overview of the ensemble of WRF model simulations varying LSM/PBL/SL scheme and vegetation lookup table carried out for the NW domain. The names of the schemes are: LSM: NOAH (Tewari et al., 2004), RUC (Benjamin et al., 2004), PX (Noilhan and Planton, 1989), SLAB (Dudhia, 1996), NOAH-MP (Niu et al., 2011); PBL/SL: YSU (Hong et al., 2006), MYJ (Mellor and Yamada, 1982), MYNN (Nakanishi and Niino, 2006), modified MYNN see Section 5.3), ACM2 (Pleim, 2007), MM5 (Jiménez et al., 2012), M-O surface layer (Janjic and Zavisa, 1994). The number in parenthesis is the respective namelist value in the WRF model configuration file. The "simulation code" is explained in Appendix A2.

run name	simulation code	LSM (#)	PBL (#)	SL (#)	veg table
BASE	EES81_2551040004	NOAH (2)	MYNN (5)	MYNN (5)	NEWA
MYNN*	EES81_2550040004	NOAH (2)	MYNN[a] (5)	MYNN[a] (5)	NEWA
MM5	EES81_2511040004	NOAH (2)	MYNN (5)	MM5 (1)	NEWA
MO	EES81_2521040004	NOAH (2)	MYNN (5)	M-O (2)	NEWA
MYJ-MO	EES81_2220040004	NOAH (2)	MYJ (2)	M-O (2)	NEWA
YSU-MM5	EES81_2110040004	NOAH (2)	YSU (1)	MM5 (1)	NEWA
RUC	EES81_3551040004	RUC (3)	MYNN (5)	MYNN (5)	NEWA
RUC-WRF	EES81_3551040004_A	RUC (3)	MYNN (5)	MYNN (5)	WRF
RUC-MO	EES81_3521040004	RUC (3)	MYNN (5)	M-O (2)	NEWA
RUC-YSU-MM5	EES81_3110040004	RUC (3)	YSU (1)	MM5 (1)	NEWA
RUC-ACM2-PX	EES81_3770040004	RUC (3)	ACM2 (7)	ACM2 (7)	NEWA
PX-ACM2-PX	EES81_7770040004	PX (7)	ACM2 (7)	ACM2 (7)	NEWA
PX-ACM2-MM5	EES81_7710040004	PX (7)	ACM2 (7)	MM5 (1)	NEWA
SLAB	EES81_1551040004	SLAB (1)	MYNN (5)	MYNN (5)	NEWA
SLAB-MYJ-MO	EES81_1220040004	SLAB (1)	MYJ (2)	M-O (2)	NEWA
SLAB-YSU-MM5	EES81_1110040004	SLAB (1)	YSU (1)	MM5 (1)	NEWA
SLAB-ACM2-PX	EES81_1770040004	SLAB (1)	ACM2 (7)	ACM2 (7)	NEWA
NOAHMP	EES81_4550040004	NOAH-MP (4)	MYNN (5)	MYNN (5)	NEWA
NOAHMP2[b]	EES81_4550040004_B	NOAH-MP (4)	MYNN (5)	MYNN (5)	NEWA
NOAHMP-WRF	EES81_4550040004_A	NOAH-MP (4)	MYNN (5)	MYNN (5)	WRF
NOAHMP-MYJ-MO	EES81_4220040004	NOAH-MP (4)	MYJ (2)	M-O (2)	NEWA
NOAHMP-YSU-MM5	EES81_4110040004	NOAH-MP (4)	YSU (1)	MM5 (1)	NEWA
ANNZ0[c]	EES82_2551040004	NOAH (2)	MYNN (5)	MYNN (5)	WRF[d]
ANNZ0N[c]	EES82_2551040004_A	NOAH (2)	MYNN (5)	MYNN (5)	WRF
AGGZ0[c]	EES83_2551040004	NOAH (2)	MYNN (5)	MYNN (5)	NEWA

[a] is the unmodified MYNN scheme, [b] uses opt_sfc = 2 in the NOAH-MP scheme. [c] differs in surface roughness, see text for details. [d] WRF table, but low roughness for tidal zone.

and thus was changed to $z_0 = 0.001$ m. The forest classes were also significantly changed, with the $z_0 = 0.50$ m value in the default table being changed to $z_0 = 0.90$ m, which is more representative of forests in, for example Sweden (Dellwik et al.,

Table 4. Vegetation look-up table for the surface roughness length as a function of the USGS land use category (Anderson et al., 1976) in the NEWA and default NCAR WRF model configuration. Only values changed from default are shown.

USGS type	land-use land cover class	z_0 NEWA [m]	z_0 WRF orig range [m]
2	Dryland Cropland and Pasture	0.10	0.05–0.15
3	Irrigated Cropland and Pasture	0.10	0.02–0.10
4	Mixed Dryland/Irrigated Cropland and Pasture	0.10	0.05–0.15
5	Cropland/Grassland Mosaic	0.10	0.05–0.14
7	Grassland	0.10	0.10–0.12
8	Shrubland	0.12	0.01–0.05
9	Mixed Shrubland/Grassland	0.12	0.01–0.06
11	Deciduous Broadleaf Forest	0.90	0.5
12	Deciduous Needleleaf Forest	0.90	0.5
13	Evergreen Broadleaf Forest	0.90	0.5
14	Evergreen Needleleaf Forest	0.90	0.5
15	Mixed Forest	0.50	0.20–0.50
17	Tidal zone[a]	0.001	0.20

[a] Originally called "Herbaceous Wetland" in the default WRF model vegetation table.

2014). The new roughness values should be considered only as estimates and as such there might be some limitations in the representation of the roughness length, they are nevertheless much more realistic than the default ones. The experiments using the standard vegetation tables are labelled WRF vegetation in Table 3.

All NEWA setups use a constant value of surface roughness and have no annual cycle, except for two of the setups (ANNZ0 and ANNZ0N) that have annual cycle according to the default WRF vegetation table (except for the tidal zone). In WRF, the seasonality of the surface roughness length is controlled by the value of the green vegetation fraction (Refslund et al., 2014) and applies mostly for cropland classes in Table 4. The annual cycle of green vegetation fraction does not change from year to year and is spatially inconsistent with the ESA-CCI land-use dataset used in the NEWA simulations. Therefore, because of inherent uncertainties, in NEWA we have chosen to use a single constant value of roughness for land-use and land cover class. Also, since the WRF wind climatologies will be further downscaled, as described in Dörenkämper et al. (2020), using a constant value of surface roughness facilitates the process. Another roughness-related experiment that was included, AGGZ0, uses the sub-tiling option for NOAH (Li et al., 2013), with the NEWA vegetation table. The sub-tiling option generates more realistic values of surface roughness length in areas of mixed vegetation, which could reduce the biases in wind speed (Santos-Alamillos et al., 2015).

The vertical profiles of mean wind speed for all the LSM/PBL/SL sensitivity experiments for four of the eight evaluation sites are shown in Fig. 9 as an example. The spread in the wind speed among the simulations excluding outliers generally

increases with height reaching around $1\,\mathrm{m\,s^{-1}}$ at 100 m. Over land, for example at Østerild, the observed wind speed profile below ~ 150 m lies well below the simulated values for all experiments, signifying a lack of representativeness of the surface conditions in the WRF model at this site, since the biases are independent of the LSM/PBL/SL used.
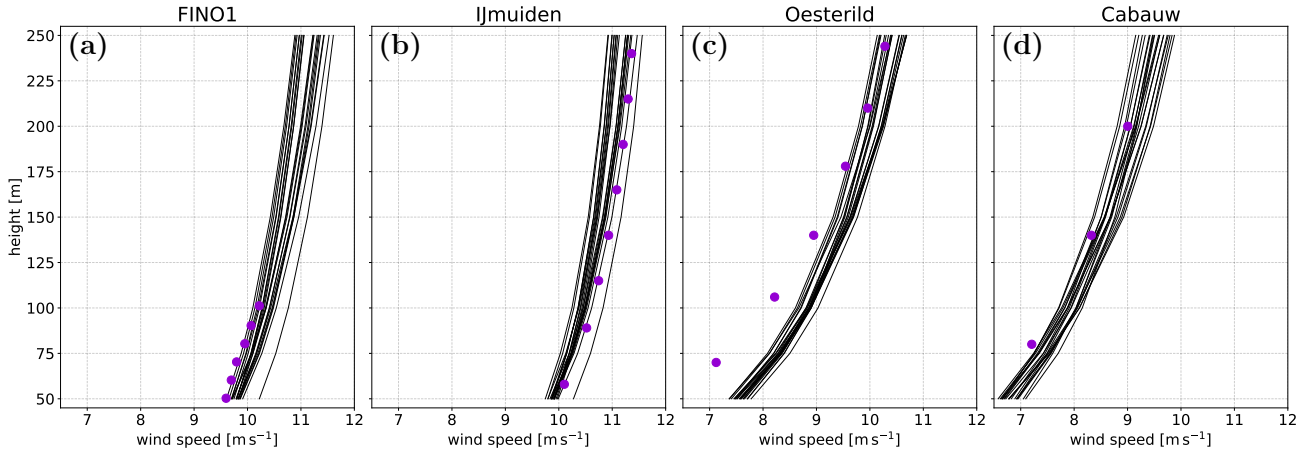
Figure 9. Mean wind speed $[\mathrm{m\,s^{-1}}]$ simulated by the various experiments in the LSM/SL/PBL ensemble as a function of height for four sites: (a) FINO1, (b) IJmuiden, (c) Østerild and (d) Cabauw. The observed values are shown as purple dots.

To facilitate the intercomparison among the ensemble members, we computed all the evaluation metrics of the wind speed and direction for each simulation (Fig. 10). The metrics compare the filtered wind speed observations and wind directions during 2015 at the levels underlined in Table 1 with the corresponding WRF-simulated time series interpolated to the same height. Figure 10a shows the model bias at all the sites. For this metric differences among stations are larger than differences between models in a single station, expressed in this figure as consistent colours for each column. On average the characteristic of the bias relates most directly to the the quality of the measurements and type of site, that is slightly negative bias over the sea and positive bias over land. The latter is likely a consequence of deficient representation of the land characteristics around each site. Some other general patterns are that the simulations using the MYJ scheme tend to have largest absolute biases, except at FINO3, and that the YSU-MM5-RUC simulation is an outlier, whose results differ from other setups, typically being among the worst of setups for any station.

To better quantify the differences between the simulations, Fig. 10b shows the skill score (SS) of the bias from the BASE simulation as defined in Section 4. Positive numbers (in purple) show a decrease in absolute bias, which point to a more accurate simulation. From these values some conclusions can be drawn. First, the differences ~~of~~ between the simulations are quite modest~~, with~~ : a maximum SS of 0.9 in the MM5 and MYJ-MO simulation at FINO1 and FINO3~~and large up to~~, and large (up to SS=−16.2) in several simulations at Høvsøre~~, which~~. These large negative SS are an artefact of the very low bias in the BASE simulation at this site. Second, no simulation is capable of improving or degrading the bias at all of the sites: the SLAB simulation improves the bias (i.e., SS > 0) at three of the eight sites, while the MO simulation improves the bias at five

18

of the eight sites. The MYNN* scheme considerably degrades the simulations at seven of the eight sites. The latter supports our decision to use the modified version of the MYNN scheme as baseline. It is relevant to note that the changes in z_0 only cause minor variations in the biases (see members ANNZ0, ANNZ0n, AGGZ0), however the sites are located in regions with vegetation classes that were not changed significantly from the WRF model standard table (Table 4).

5   Figures 10c and 10d provide further information about the sensitivity tests based on the EMD metric defined in Section 4 to evaluate the shape of the wind speed distributions. As with the bias, the EMD shows that the largest differences in total error are linked to the site location, with the best model performance at Høvsøre, with EMD between $0.1–0.4\,\mathrm{m\,s^{-1}}$, while at Risø the results fall between $0.4–1.0\,\mathrm{m\,s^{-1}}$. Particularly interesting is the comparison between bias and EMD, since for most setups and stations the values of EMD and bias are similar, especially when the model results are significantly different from the observations. The EMD allows us to identify cases where the change in model setup improves only the mean value of distribution, as opposed to the similarities of the whole distribution. For instance, if only the biases were analysed, it could be argued that MYJ-MO is better than the BASE in FINO3 and IJmuiden, while the EMD show that there is only a modest improvement over the BASE at IJmuiden and a worse result at FINO3. Similar conclusions can be drawn for the SLAB-MYJ-MO and YSU-MM5-RUC simulations. As with the bias, no simulation improves the EMD for all sites, and very few simulations improve the EMD at all, especially for the land sites. The SLAB simulation improves the EMD at four sites (SS > 0.1) while the MO simulation improves the EMD at three of the eight sites.

     Two other metrics are presented in Fig. 10: the RMSE for the wind speed (Fig. 10e) and the circular EMD (CEMD) for the wind direction (Fig. 10g) and their respective SS from the BASE simulation (Figs. 10f and 10h). As with the bias and EMD, the RMSE is primarily site dependent, with small differences between the simulations. The RMSE is lowest at Cabauw (1.6–2.1 $\mathrm{m\,s^{-1}}$) and highest at FINO2 ($2.4–2.7\,\mathrm{m\,s^{-1}}$). No simulation significantly improves the RMSE, but it remains unchanged at all sites in the MO, SLAB, NOAHMP-WRF and the all the varied roughness ensembles. The CEMD of the wind direction is variable among sites and simulations, but all the values are small (1.9–7.1°). For this metric, the MO and SLAB show similar values ~~than~~ to the BASE simulation.
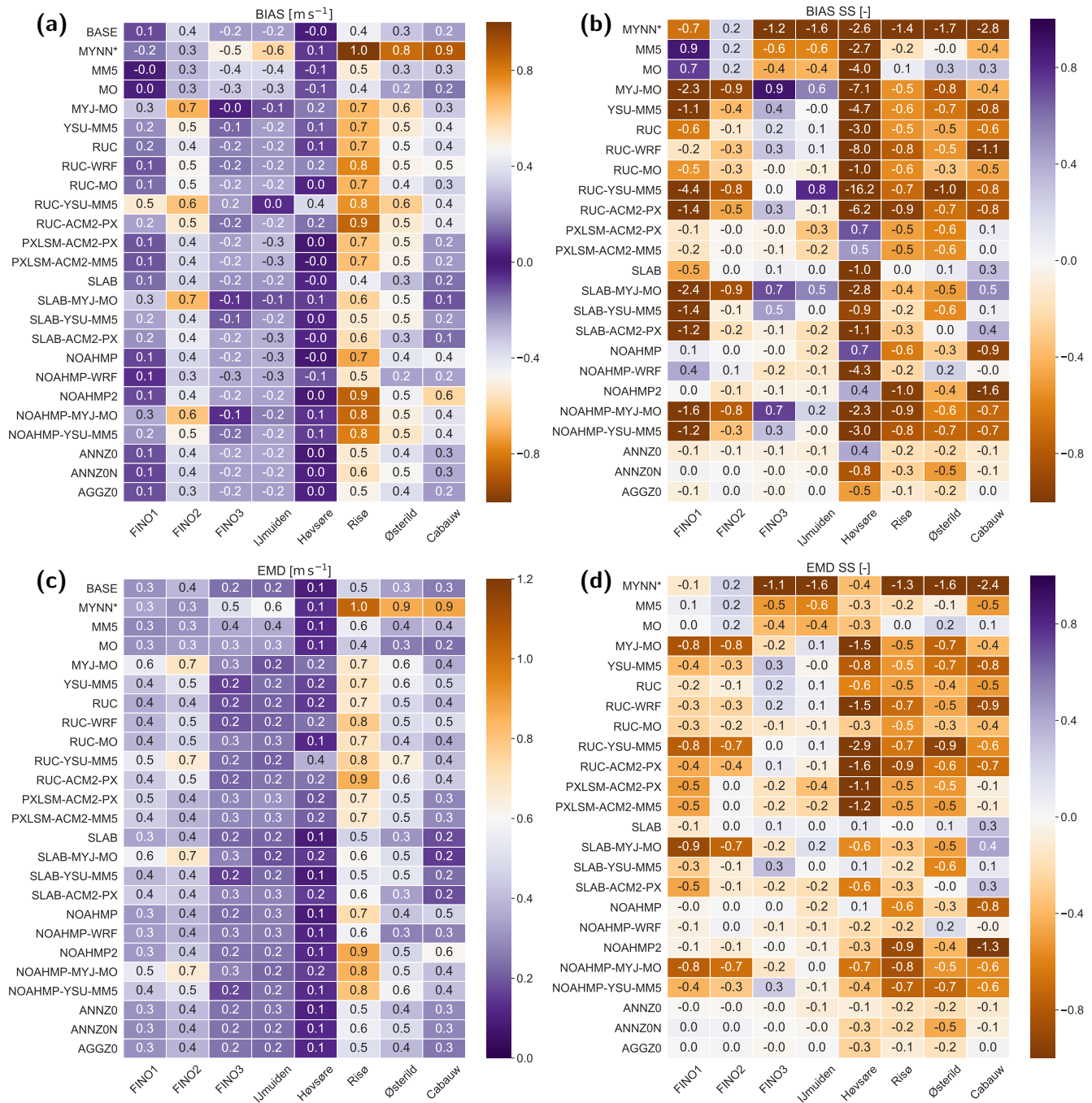
Figure 10. Evaluation metrics: (a) bias $[\mathrm{m\,s^{-1}}]$, (b) bias SS [-], (c) EMD $[\mathrm{m\,s^{-1}}]$, (d) EMD SS [-], (e) RMSE $[\mathrm{m\,s^{-1}}]$, (f) RMSE SS [-], (g) CEMD [°], and (h) CEMD SS [-] between the observed and simulated wind speed and wind direction at the eight sites and underlined heights in Table 1 and the various sensitivity studies in the LSM/SL/PBL ensemble (Table 3). All SS are relative to the BASE simulation.
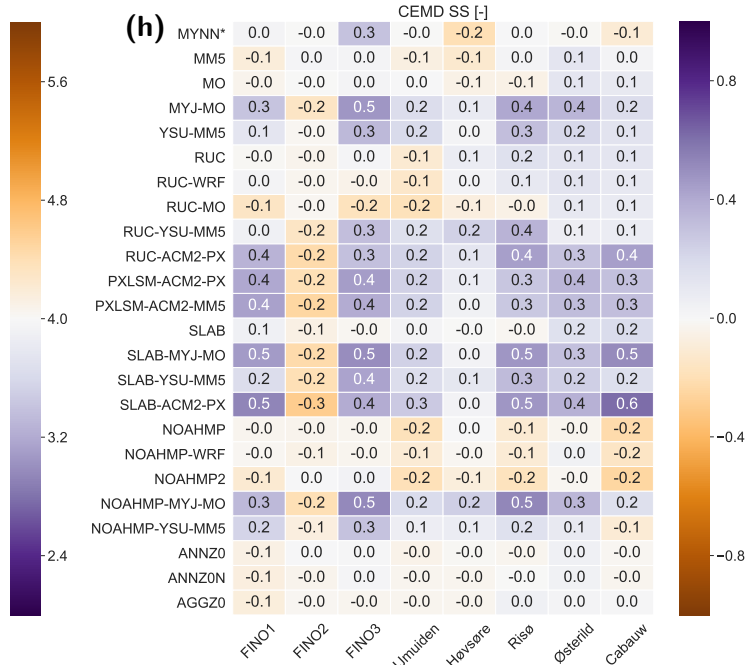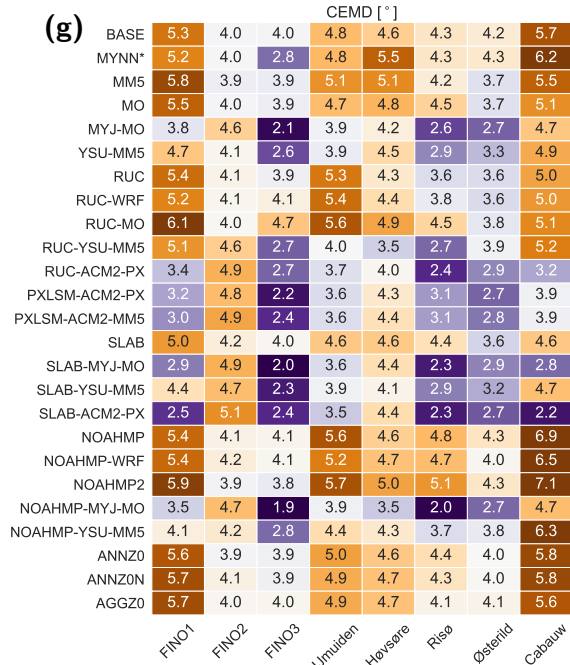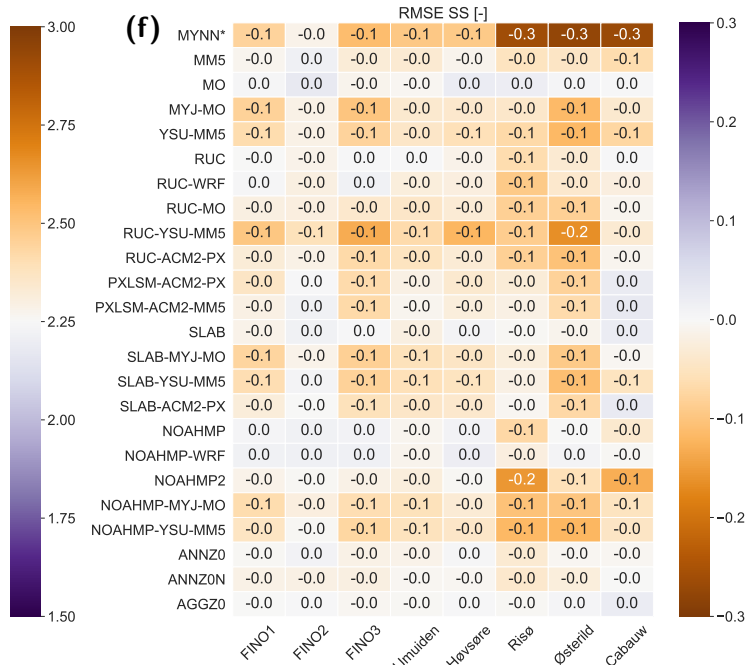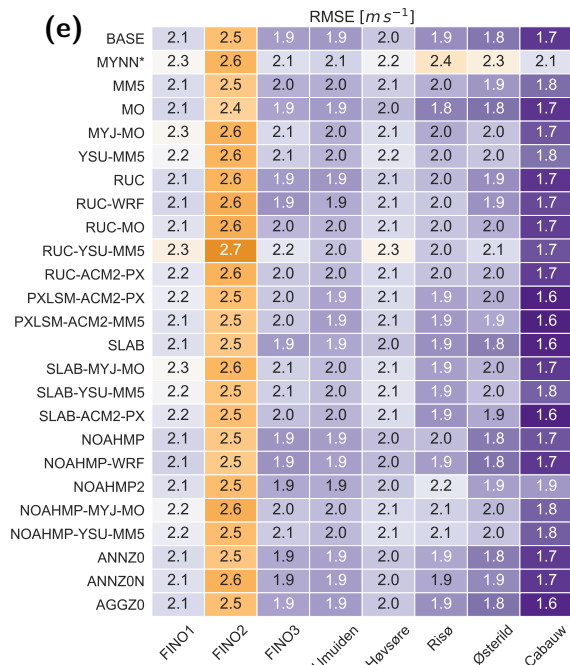
Figure 10. (Continued)

Table 5. Overview of other sensitivity experiments carried out. The meaning of the simulation code is explained in Table A2 in the appendix.

run name	simulation code	changes to BASE run
Initial, boundary conditions and SST		
ERAI	IIS81_2551040004	ERA-Interim (Dee et al., 2011) forcing
MERRA2	MMS81_2551040004	MERRA2 (Gelaro et al., 2017) forcing
FNL	FFS81_2551040004	FNL (NCAR, 2000) forcing
ERA5SST	EEE81_2551040004	ERA5 SST
HRSST	EEH81_2551040004	NOAA HRSST (Gemmill et al., 2007)
OISST	EEO81_2551040004	NOAA RTGSST (Reynolds et al., 2010)
Model dynamics		
WRFV361	EES61_2550040004	WRF version 3.6.1
NUDPAR	EES81_2551040004_A	lower wavelength in nudging
LRELAX	EES81_2551040004_B	larger relaxation zone
TWOWAY	EES81_2551040004_C	two-way nesting
NUDD3	EES81_2551040004_J	spectral nudging D1, D2, D3
GNUD1	EES81_2551040004_I	grid nudging D1
GNUD3	EES81_2551040004_H	grid nudging D1, D2, D3
Other physics		
RAD3S	EES81_2551040004_D	radiation $\Delta t = 3\,\text{s}$ in all domains
RAD12S	EES81_2551040004_E	radiation $\Delta t = 12\,\text{s}$ in all domains
FASTRA	EES81_2551040024	fast RRTMG code
CAMRAD	EES81_2551040003	CAM radiation (Collins et al., 2004)
CUG-F	EES81_2551040304	Grell-Freitas (Grell and Freitas, 2014) CU scheme
THOMP	EES81_2551080004	Thompson cloud physics (Thompson et al., 2012) + icing

5.5 Other sensitivity experiments

A second set of sensitivity experiments was carried out to identify other factors that could potentially be important for the simulation of wind speed and direction within the WRF model. These experiments are listed in Table 5 and can be grouped into three main categories. First, we tested the impact of various initial and boundary conditions, by using ERA-Interim, MERRA2 and FNL fields as forcing. The effect of various sources of sea surface temperature (SST) was also tested. In a second set, we tested other model dynamics including the effect of spectral versus grid nudging, enlarging the lateral boundary zone, changing the wavelength of the minimum spectral nudging length and enabling 2-way nesting. The third set of experiments tested other model physics not related to the surface and PBL, that is radiation, cumulus convection and explicit moisture schemes.
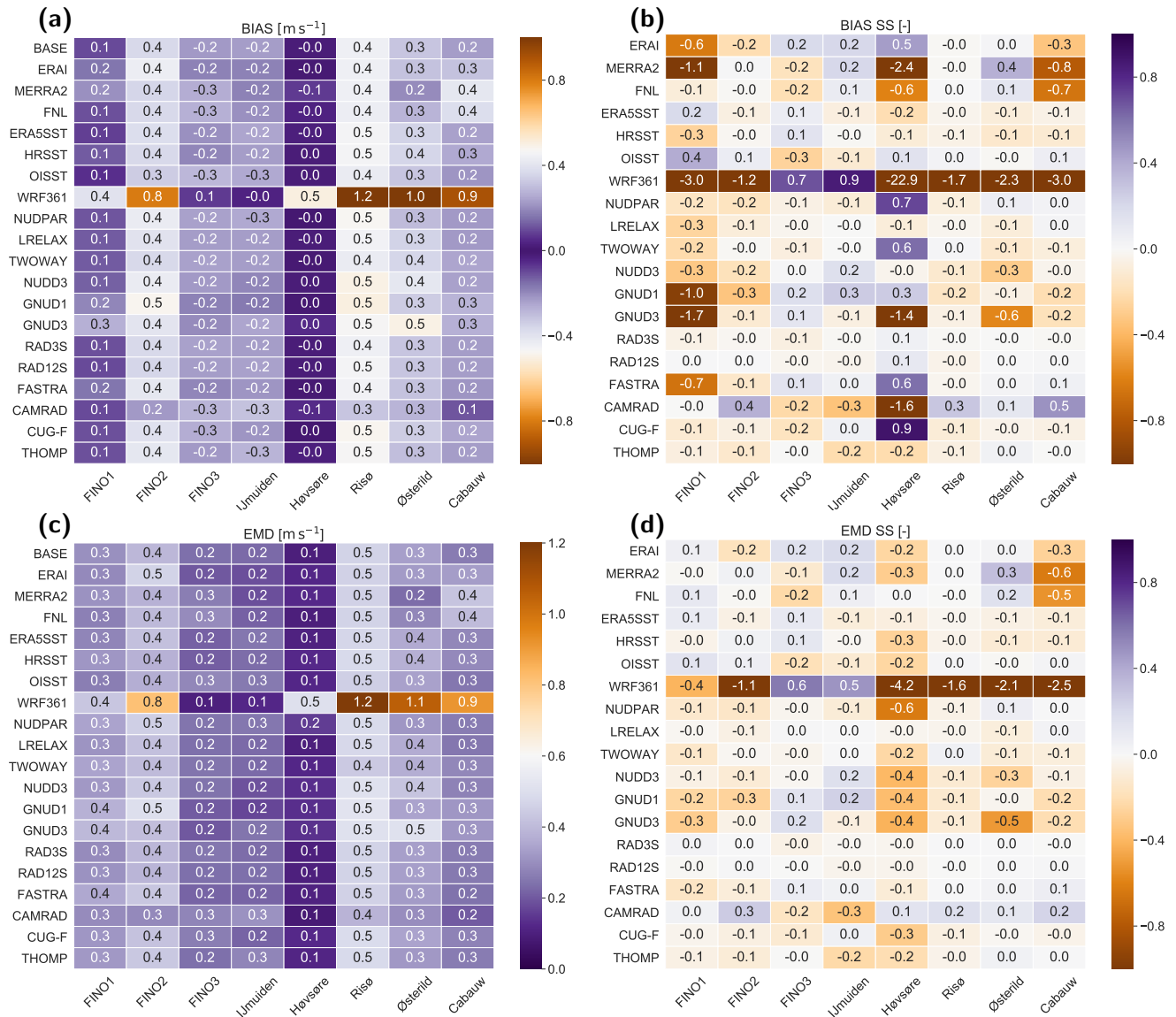
Figure 11. Evaluation metrics: (a) bias [m s^{-1}], (b) bias SS [-], (c) EMD [m s^{-1}], (d) EMD SS [-], (e) RMSE [m s^{-1}], (f) RMSE SS [-], (g) CEMD [°], and (h) CEMD SS [-] between the observed and simulated wind speed at the eight sites and underlined heights in Table 1 and the various sensitivity studies in the other ensemble (Table 5). All SS are relative to the BASE simulation.

Figure 11 shows all the evaluation metrics of the wind speed and wind direction for each simulation for the other set of sensitivity experiments. In contrast to the LSM/PBL/SL ensemble members, for the bias (Figure 11a) these simulations provide results that are very similar to the BASE simulation, except for the WRF361, GNUD3 and CAMRAD simulations. Switching the source of initial and boundary conditions to ERA-Interim, MERRA2 or FNL has very small impact to the bias. Changing
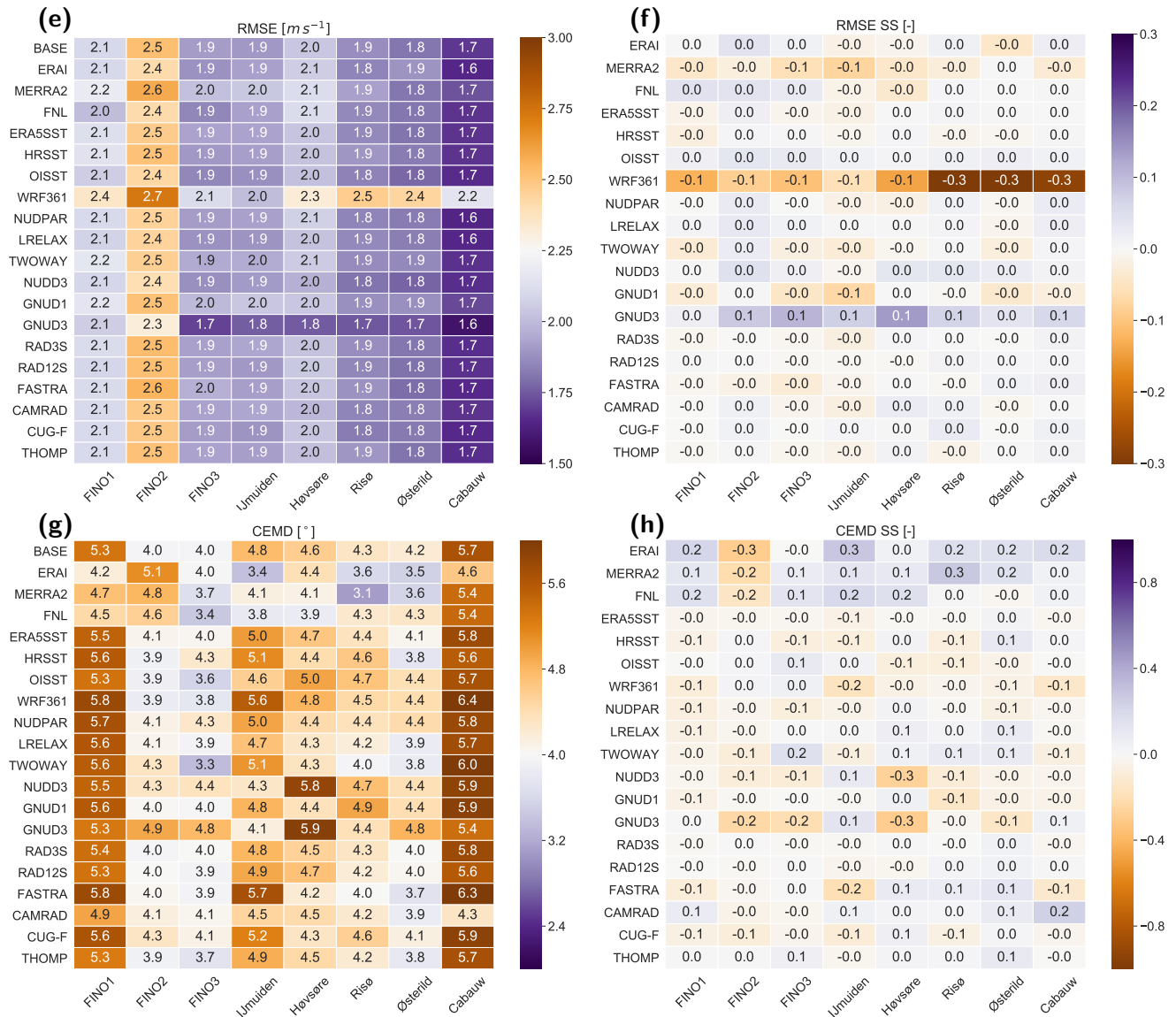
Figure 11. (Continued)

the source of SST has negligible effect for all of the offshore masts. Most of the changes to the dynamic settings have very small consequences to the bias. The only larger change is the use of grid nudging in all three WRF domains, simulation GNUD3. The biases increase (SS ≥ 0.1) in seven of the eight sites when this setting is activated, probably because of the slow down of the winds in the ERA5 reanalysis over land (see Fig. 9 of Dörenkämper et al., 2020). Interestingly, using this setting decreases the

5 RMSE between the simulated and observed time series (Fig. 11f). Since the intent of the NEWA atlas is to provide an accurate

24

description of the wind climatology, distribution metrics (e.g. EMD) are considered more important than time-dependent ones (e.g. RMSE).

The EMD and EMD SS for this set of sensitivity experiments was calculated and the results are shown in Figure 11c and 11d. The conclusions about the usefulness of the EMD apply here as well. For instance, at FINO1 the ERAI and MERRA2 runs show an increase in bias compared to the BASE, while the EMD values show that these runs actually have more similar wind speed distribution to observations than the BASE. Otherwise, the EMD metric confirms the conclusions described earlier about the small impact of all of these changes and the relative decrease in quality when using grid-nudging (GNUD3).

Other dynamical and physics options such as the radiation time step (RAD3S or RAD12S), the change of nudging constants (NUDPAR, LRELAX) or nesting (TWOWAY) show very small influence to nearly all the evaluation metrics. As with the LSM/PBL/SL, the simulations show very small differences in RMSE and CEMD with respect to the BASE, except for GNUD3 as mentioned above, which shows improved RMSE. It is also interesting to note that the ERAI, MERRA2 and FNL simulations show small improvements in the CEMD at many sites compared to the BASE simulation. The reason for this behaviour is not well understood.

In conclusion, many other changes to the WRF model settings have inconsequential effects to the simulation of the wind speed and direction at wind turbine hub height. The use of grid nudging in all domains (simulations GNUD3) improves the RMSE at most sites, but has a negative effect to the BIAS, EMD and CEMD. The change in radiation parameterisation has a small effect in relation to the BASE simulation. Because the effect is small and the CAM radiation parameterisation (Collins et al., 2004) is more expensive in terms of computational resources, it was ultimately not used in the NEWA production run.

5.6 Sensitivity to domain size

An additional decision to be made regarding the NEWA mesoscale simulations was the domain configuration; that is using a single large domain or several small domains to cover Europe. From a pure computational perspective, one single domain is more efficient, because the WRF model code scales better with larger domains (Kruse et al., 2013) and there is only data from one domain to post-process. However, the output files are very large and the simulation needs to be completed before post-processing can begin. The limiting factor here is the scratch space available at modern HPC systems that is typically not more than 100 TB. Furthermore, large areas outside of the region of interest (the NEWA domain, see Section 1) would be simulated, that is parts of the Atlantic Ocean, the Norwegian Sea and non-EU countries in Eastern Europe, which were outside of the planned New European Wind Atlas. Apart from these technical questions, it was unknown how the domain size influences the quality of the simulated fields.

To study the sensitivity of the simulated wind speed, we carried out simulations for three differently sized domains over the North Sea using the same setup and resolution as in Section 5.1, but for 2016, when data availability was larger. The number of grid points in the inner domain in these simulations are: small (SM) 121×121, medium (MD) 241×241 and large (LG) 481×481, which correspond to square domains with edge lengths on the WRF model projection equivalent to 360 km, 720 km, and 1440 km, respectively. The three domains are centred at the same coordinates and only differ by the number of grid points. The size of the boundary zone, in grid cells, between D1 and D2 and D2 and D3 is kept the same. Two sets of WRF model

25
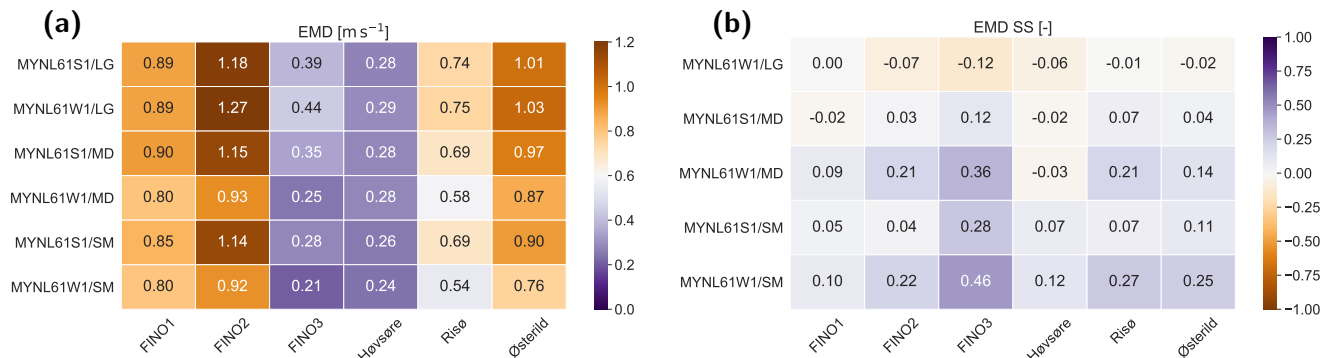
Figure 12. (a) EMD [m s^{-1}] and (b) EMD SS [-] between the observed and simulated wind speed for the six inner sites and underlined heights in Table 1 for the six domain size experiments. The EMD SS is compared to that of the EMD of the MYNL61W1/SM simulation.

simulations were done for each of the three domains, MYNL61S1 and MYNL61W1 in Table 2. For evaluating the results of the simulations we use the same data as in Section 5.1, but only six of the masts are contained within the SM domain.

Figure 12 shows the EMD between the various WRF model simulations and the observations for the six sites. It should be noted that these EMD values are larger than those previously shown for the LSM/SL/PBL and other ensemble. The reason for this is that the simulations were carried out with WRF V3.6.1 (see brown line in Fig. 11c) and the measurements at FINO1 and FINO2 are more affected by the presence of neighbouring wind farms in 2016 than what they were in 2015. The EMD from all simulations are summarised as follows: for five out of six sites the MYNL61W1/LG simulation have the largest EMD, and for all sites the MYNL61W1/SM simulation has the smallest EMD. Similar results (not shown) emerge for the bias and RMSE. In conclusion, for this region, biases in mean wind speed are influenced by the size (and possibly also location) of the domain, smaller domains have generally lower wind speeds and thus lower biases. This effect is most pronounced in the week-long nudged simulations. The RMSE increases with increasing domain size and integration time.

The results from these experiments guided the design of the NEWA domains for the production run. Instead of a single, or a few, very large domains, we chose to conduct the simulations in a rather large number of medium-sized domains. While generating different time series, overlapping areas in simulations generally show similar wind climates (Witha et al., 2019, Section 2.1.3). We decided against very small domains, which would add to the needed computational resources and the extra effort of dealing with hundreds of model domains and because most countries would be covered by multiple domains It was desired that each country should be covered by only one domain to avoid edge differences in overlapping domains. The final domain configuration, which is presented in this paper's companion (Dörenkämper et al., 2020), fulfils this requirement for all countries except Norway, Sweden and Finland which are so elongated that a correspondingly large domain would be detrimental to the accuracy of the results.

26

5.7 Summary of the sensitivity experiments

A long list of sensitivity studies were carried out to identify a suitable configuration for the NEWA production run. Here is a summary of the findings:

1. In the initial sensitivity experiments, the largest differences in annual mean wind speed at 100 m in Fig. 4 and AGL/AMSL in Table 2 are between simulations using two PBL schemes (MYNN and YSU) and coincide with regions of high surface roughness in all domains over Europe. Over the sea, the differences could be traced to differences in atmospheric stability, but were modest.

2. The weekly simulations using spectral nudging tend to perform better (lower bias, EMD and RMSE) for the eight sites in northern Europe. The simulation using the MYNN scheme in the WRF model version 3.6.1 outperformed the simulations using the YSU scheme in this region.

3. The use of the WRF model V3.8.1 and the MYNN scheme increased the biases compared to observations at nearly all sites and most levels. A couple of settings, mynn_mixlength=0 and COARE_OPT=3.0, turn the MYNN scheme nearly back to the conditions in the WRF model version 3.6.1. However, above 100 m a.g.l. the modified MYNN scheme gives lower wind speeds than the one in WRF V3.6.1 at all sites.

4. A series of 25 experiments varying the land, PBL and surface layer scheme (LSM/PBL/SL) shows a spread in the mean wind speed of about $1 \, \mathrm{m \, s^{-1}}$. When comparing to wind speed observations, most LSM/PBL/SL ensembles show negative biases over the ocean (except for FINO2) and positive over land, which are more consistent between sites than LSM/PBL/SL combinations. This likely reflects misrepresentation of the land surface around each site than deficiencies in the LSM/PBL/SL schemes themselves.

5. Changes to the WRF model lookup table for surface roughness length have large consequences for the simulated wind speed, but is nearly invisible to the evaluation against the tall masts because these are located in areas away from those impacted by the changes.

6. The use of the EMD metric and the skill score helps clarify the comparison of the improvements between the various LSM/PBL/SL and the BASE simulation, especially if the bias is small. No simulation improves the EMD for all sites, and very few simulations improve the EMD metric at all, especially for the land sites. However, the MO simulation, which uses the NOAH, MYNN and MO surface layer schemes, improves the results, SS > 0.1, at four of the eight sites and was finally chosen as the physical model configuration for the NEWA production run.

7. A set of additional sensitivity experiments changing the source of forcing data, SST, dynamic options and other physical parameterisations, shows smaller changes from the baseline simulation than the various LSM/PBL/SL experiments. Nearly all the changes have inconsequential effects to the simulation of the wind speed at wind turbine hub height. Only the simulation using the CAM radiation scheme showed improvements over the RRTMG scheme used in the baseline

experiment. However, it was concluded that the modest improvements were not worth the additional expense of running this scheme in the production run.

8. A final set of experiments testing the effect of the size of the domain on the simulated wind speed error statistics showed that for a domain centred over Denmark, the simulations using the smaller domain have lower wind speeds which compare better to measurements and time correlations decrease with domain size. It is however unclear if this is a consequence of the domain size itself or the location of the main inflow boundary to the domain in the simulations.

6 Discussion and outlook

In the companion paper (Dörenkämper et al., 2020), we document the final model configuration and how we computed the final wind atlas, including a detailed description of the technical and practical aspects that went in to running the WRF simulations and the downscaling using the linearised microscale model WAsP (Troen and Petersen, 1989). This second paper also shows a comprehensive evaluation of each component of the NEWA model-chain using observations from a large set ($n = 291$) of tall masts located all over Europe. That work concludes that the NEWA wind climate estimated by the WRF model simulations using the model configuration selected here is significantly more accurate than using the raw ERA5 wind speed and direction which was used to force these simulations.

From the results of the sensitivity analysis, it is clear that some model setups have better scores for one validation metric (e.g. EMD) and some setups have better scores in other metrics (e.g. RMSE). Also, the performance of the setup varies considerably from station to station. In addition, one can clearly see that the performance at each station is systematic. At some sites all setups tend to perform better and at other sites worse. Thus, it was decided not to base configuration decisions on average metrics across the sites. From the results of the sensitivity analysis, several model setups could have been chosen (e.g. MO, SLAB or CAMRAD). The MO setup, which was ultimately chosen for the production runs (Dörenkämper et al., 2020), is one of the best performing, according to the verification results. It performs well both in terms of its wind speed distribution (EMD) and relatively well in a time-series sense (RMSE) compared to other model setups. It has the additional benefit of runs being numerically stable when compared to BASE, i.e., the runs failed less (this aspect was important, because of the necessity for a person to monitor and re-submit the runs), and MO also had a favourably small computational time when compared to other setups.

Many other details of the model setup have not been tested. For example, the sensitivity of the simulated wind climate to the number and location of the vertical levels in the PBL was not systematically tested. Previous studies (Hahmann et al., 2015) and earlier simulations in the NEWA project showed small impact, but these were conducted with a single PBL scheme. The dependence of the simulated wind climate on the grid spacing is also not explored here. Previous publications offshore (e.g. Floors et al., 2018b) and onshore (Rife and Davis, 2005; Gómez-Navarro et al., 2015; Smith et al., 2018) have done so, but the investigation on what is the ideal grid spacing of the mesoscale simulations when further downscaling is done also remains unknown. In a similar theme, the dependence of the WRF model simulation of the wind climate on the size and location of the computational domain also remains unresolved. Smaller domains in the WRF simulation tend to have smaller wind speed

biases and higher RMSE compared to tall mast observations, but it was unclear if this was really a result of the size of the domain or rather the location of the boundaries in relation to the large-scale flow. More numerical experiments should be carried out to identify all these potential interactions.

As with any modelling study, some questions remain unresolved simply because of the expensive nature of the numerical
5 experiments. For convenience and simplicity, we separated the sensitivity experiments dealing with LSM/PBL/SL and the other parameterisations changing only one scheme or parameter at a time. However, the experiment using the CAM radiation scheme had better verification statistics than the other simulations. The use of the CAM scheme was not tested using the preferred LSM/PBL/SL combination. Thus, we suggest that a better way to go in this process of model selection is to go through the changes and evaluation in a sequential way. But the number of ensemble members can rapidly become unmanageable.
10 Algorithms in this direction are currently being applied for tuning Earth System Models (Li et al., 2019) and could perhaps be evaluated to best determine the best suited WRF setups for different applications, not just wind resource assessment.

Finally, we have not evaluated the results of the ensemble simulations with the large observational dataset used in the companion paper (Dörenkämper et al., 2020). We have also not evaluated the simulated wind climatologies when we use the full downscaling model chain. Nevertheless, the evaluation of the production run with the data included there, support the
15 performance of the configuration selected herein.

Code availability. The WRF model code is open source code and can be obtained from the WRF Model User's Page (http://www2.mmm. ucar.edu/wrf/users/, doi:10.5065/D6MK6B4K). For the NEWA production run we used WRF version 3.8.1. The code modifications as well as namelists, tables and domain files we used are available from the NEWA GitHub repository: https://github.com/newa-wind/Mesoscale and permanently indexed in Zenodo (Hahmann et al., 2020). The WRF namelists and tables for all the ensemble members are also available in
20 the repositories. The code used in the calculation of EMD metric is available from: https://pypi.org/project/pyemd/

Data availability. The NEWA data from the production run is available from https://map.neweuropeanwindatlas.eu/.
All namelists and tables needed to carry out the ensemble simulations are available from https://doi.org/10.5281/zenodo.3709088
The summary statistics for all the ensemble simulations are available from https://doi.org/10.5281/zenodo.4002351
The forcing data for the mesoscale simulations are publicly available:
25 ERA5 - https://climate.copernicus.eu/climate-reanalysis,
OSTIA - http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SST_GLO_SST_
L4_NRT_OBSERVATIONS_010_001,
CORINE - https://land.copernicus.eu/pan-european/corine-land-cover,
ESA-CCI - http://cci.esa.int/data.
30 Some of the tall mast data used for the evaluation of the wind atlas is confidential and thus not publicly available.

Table A1. The WRF model setup common to all simulations and to the BASE simulation.

option	setting
Common setup:	
Model grid	D1/D2/D3 with 27 km / 9 km / 3 km horizontal grid spacing
	Lambert conformal grid projection
Terrain data	Global Multi-resolution Terrain Elevation Data 2010 at 30" (Danielson and Gesch, 2011)
Land use	CORINE land-cover classification (Copernicus Land Monitoring Service, 2019)
	ESA-CCI land-cover (Poulter et al., 2015) outside the CORINE domain
Vertical discretisation	61 vertical levels with model top at 50 hPa.
Model levels	20 model levels below 1 km
Diffusion	Simple diffusion (option 1), 2D deformation (option 4)
	6th order positive definite numerical diffusion (option 2)
	No vertical damping
	Positive definite advection of moisture and scalars
BASE setup:	
Forcing data	ERA5 (Hersbach et al., 2020) reanalysis at $0.3°$ on pressure levels
Sea surface temperature	Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012)
	fractional sea-ice activated
	Lake temperatures from time-averaged ERA5 ground temperatures
Cloud micro-physics	WRF Single-Moment 5-class scheme (Hong et al., 2004)
Cumulus convection	Kain-Fritsch Scheme (Kain, 2004); D1 and D2
PBL scheme	MYNN level 2.5 (Nakanishi and Niino, 2009)
Surface layer scheme	MYNN (Nakanishi and Niino, 2009) with mods (see text)
Land surface model	Unified Noah Land Surface Model (Tewari et al., 2004)
Shortwave and longwave radiation	RRTMG (Iacono et al., 2008) at 12 minute interval
Nesting	one way nesting with smooth (option 2)
Nudging	spectral nudging U, V, T and q on D1
	above level 20, no PBL nudging
Nudging constant	$0.0003\,\mathrm{s}^{-1}$
Nudging wavelength	14 (x) and 10 (y) equivalent to $6 \times \Delta x$ of ERA-Interim reanalysis grid spacing

Table A2. Explanation for the various digits of the sensitivity experiments that refers to the code available in GitHub.

digit	option	convention
1	IC/BC data	E: ERA5, I: ERA-Interim, M: MERRA2, C: CFSR2, F: FNL
2	Land IC	same as digit 1 (E, I, M, C, F), G: GLDAS
3	SST data	S: OSTIA, H: HRSST, O: OISST, or same as digit 1 (E, I, M, C, F)
4	WRF version	6: WRFV3.6.1, 8: WRFV3.8.1
5	Roughness option	1: constant, 2: annual cycle, 3: aggregated
6	Separator underscore	
7	Land Surface Model	Code as in WRF: Thermal diffusion=1, NOAH=2, RUC=3, CLM4=4, PLX=7
8	PBL scheme	Code as in WRF: YSU=1, MYJ=2, QNSE=4, MYNN2=5, MYNN3=6, ACM2=7
9	Surface layer	Code as in WRF: Revised MM5=1, M-O=2, QMSE=4, MYNN=5, P-X=7
10	Modified PBL and surface layer?	no=0, yes=1
14–15	Cloud Microphysics	Code as in WRF: WSM5=04, Thompson=08, Thompson+aerosol=28
16–17	Convective scheme (D1,D2)	Code as in WRF: No=00, K-F=01, B-M=02, Grell-Devenyi=93
18–19	SW/LW radiation	Code as in WRF: CAM=03, RRTMG=04, Fast RRTMG=24
20–21	Separator (if need) + extra option	A, B, C, e.g. two-way nesting

References

Anderson, J. R., Hardy, E. E., Roach, J. T., , and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, Tech. rep., United States Geological Service, https://pubs.usgs.gov/pp/0964/report.pdf, 1976.

Badger, J., Frank, H., Hahmann, A. N., and Giebel, G.: Wind-climate estimation based on mesoscale and microscale modeling: Statistical-dynamical downscaling for wind energy applications, J. Appl. Meteorol. Clim., 53, 1901–1919, https://doi.org/10.1175/JAMC-D-13-0147.1, 2014.

Benjamin, S. G., Grell, G. A., Brown, J. M., and Smirnova, T. G.: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model., Mon. Wea. Rev., 132, 473–494, https://doi.org/10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2, 2004.

Bosveld, F. C.: Cabauw In-situ Observational Program 2000 – Now: Instruments, Calibrations and Set-up, Tech. rep., KNMI, http://projects.knmi.nl/cabauw/insitu/observations/documentation/Cabauw_TR/Cabauw_TR.pdf, 2019.

Chávez-Arroyo, R., Lozano-Galiana, S., Sanz-Rodrigo, J., and Probst, O.: Statistical-dynamical downscaling of wind fields using self-organizing maps, Appl. Therm. Eng., 75, 1201–1209, https://doi.org/10.1016/j.applthermaleng.2014.03.002, 2015.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., and Briegleb, B.: Description of the NCAR Community Atmosphere Model (CAM 3.0), Tech. Rep. NCAR/TN-464+STR, Mesoscale & Microscale Meteorology Division, NCAR, USA, 2004.

Copernicus Land Monitoring Service: CORINE Land Cover, https://land.copernicus.eu/pan-european/corine-land-cover, accessed: 2019-04-15, 2019.

Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), Tech. Rep. 2011-1073, U.S. Geological Survey Open-File Report, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorolog. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dellwik, E., Arnqvist, J., Bergström, H., Mohr, M., Söderberg, S., and Hahmann, A.: Meso-scale modeling of a forested landscape, J. Phys. Conf. Ser., 524, https://doi.org/10.1088/1742-6596/524/1/012121, 2014.

Donlon, C. J., Martin, M., Stark, J. D., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice analysis (OSTIA)., Remote Sens. Environ., 116, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Marugán, M. S., Sīle, T., Trei, W., and Žagar, M.: The Production of the New European Wind Atlas, Part 2: Production and Validation, Geosci. Model Dev. Discuss., in review, https://doi.org/10.5194/gmd-2020-23, 2020.

Draxl, C., Hahmann, A. N., Peña, A., and Giebel, G.: Evaluating winds and vertical wind shear from WRF model forecasts using seven PBL schemes, Wind Energy, 17, 39–55, https://doi.org/10.1002/we.1555, 2014.

Draxl, C., Clifton, A., Hodge, B.-M., and McCaa, J.: The Wind Integration National Dataset (WIND) Toolkit, Appl. Energ., 151, 355–366, https://doi.org/10.1016/j.apenergy.2015.03.121, 2015.

Dudhia, J.: A multi-layer soil temperature model for MM5., in: The Sixth PSU/NCAR Mesoscale Model Users' Workshop, Boulder, Colorado, USA, 1996.

Edson, J., Jampana, V., Weller, R., Bigorre, S., Plueddemann, A., Fairall, C., D. Miller, S., Mahrt, L., Vickers, D., and Hersbach, H.: On the exchange of momentum over the open ocean, J. Phys. Oceanogr., 43, 1589–1610, https://doi.org/10.1175/JPO-D-12-0173.1, 2013.

5   Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., and Edson, J. B. a.: Bulk parameterization of air-sea Fluxes: Updates and verification for the COARE algorithm, J. Climate, 16, 571–591, https://doi.org/10.1175/1520-0442(2003)016<0571:bpoasf>2.0.co;2, 2003.

Fernández-González, S., Martín, M. L., Merino, A., Sánchez, J. L., and Valero, F.: Uncertainty quantification and predictability of wind speed over the Iberian Peninsula, J. Geophys. Res., 122, 3877–3890, https://doi.org/10.1002/2017JD026533, 2017.

Fernández-González, S., Sastre, M., Valero, F., Merino, A., García-Ortega, E., Luis Sánchez, J., Lorenzana, J., and Martín, M. L.: Characterization of spread in a mesoscale Ensemble prediction system: Multiphysics versus Initial Conditions, Meteorol. Zeitschrift, 28, 59–67, https://doi.org/10.1127/metz/2018/0918, 2018.

Floors, R., Enevoldsen, P., Davis, N., Arnqvist, J., and Dellwik, E.: From lidar scans to roughness maps for wind resource modelling in forested areas, Wind Energy Science, 3, 353–370, https://doi.org/10.5194/wes-3-353-2018, 2018a.

Floors, R., Hahmann, A. N., and Peña, A.: Evaluating mesoscale simulations of the coastal flow using lidar measurements, J. Geophys. Res., 123, 2718–2736, https://doi.org/10.1002/2017JD027504, 2018b.

Frank, H. and Landberg, L.: Modelling the wind climate of Ireland, Bound.-Lay Meteorol., 85, 359–378, https://doi.org/10.1023/A:1000552601288, 1997.

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, Q. J. R. Meteorolog. Soc., 139, 501–514, https://doi.org/10.1002/qj.1976, 2013.

20   Gelaro, R., McCarty, W., and et al, M. J. S.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gemmill, W., Katz, B., and Li, X.: Daily Real-Time Global Sea Surface Temperature - High Resolution Analysis at NOAA/NCEP., Office note nr. 260, 39 pp, NOAA/NWS/NCEP/MMAB, 2007.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrizations and nesting techniques: evaluation of surface wind over complex terrain, Geosci Model Dev., 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, 2015.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmos. Chem. Phys., 14, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014, 2014.

Hahmann, A. N., Rostkier-Edelstein, D., Warner, T. T., Vandenberghe, F., Liu, Y., Babarsky, R., and Swerdlin, S. P.: A reanalysis system for the generation of mesoscale climatographies, J. Appl. Meteorol. Clim., 49, 954–972, https://doi.org/10.1175/2009JAMC2351.1, 2010.

30   Hahmann, A. N., Vincent, C. L., Peña, A., Lange, J., and Hasager, C. B.: Wind climate estimation using WRF model output: Method and model sensitivities over the sea, Int. J. Climatol., 35, 3422–3439, https://doi.org/10.1002/joc.4217, 2015.

Hahmann, A. N., Davis, N. N., Dörenkämper, M., Sīle, T., Witha, B., and Trey, W.: WRF configuration files for NEWA mesoscale ensemble and production simulations, https://doi.org/10.5281/zenodo.3709088, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., ías Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Q. J. R. Meteorolog. Soc., 2020, 1–51, https://doi.org/DOI: 10.1002/qj.3803, 2020.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation., Mon. Wea. Rev., 132, 103–120, https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Wea. Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models., J. Geophys. Res., 113, D13 103., https://doi.org/10.1029/2008JD009944, 2008.

Janjic, Z. I. and Zavisa, I.: The Step–Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes., Mon. Wea. Rev., 122, 927–945, 1994.

Jiménez, P., García-Bustamante, E., González-Rouco, J., Valero, F., Montávez, J., and Navarro, J.: Surface Wind Regionalization over Complex Terrain: Evaluation and Analysis of a High-Resolution WRF Simulation, J. Appl. Meteorol. Clim., 49, 268–287, https://doi.org/10.1175/2009JAMC2175.1, 2010.

Jiménez, P., González-Rouco, J., Montávez, J., Navarro, J., García-Bustamante, E., and Dudhia, J.: Analysis of the long-term surface wind variability over complex terrain using a high spatial resolution WRF simulation, Clim. Dyn., 40, 1643–1656, https://doi.org/10.1007/s00382-012-1326-z, 2013.

Jiménez, P. A., Vilà-Guerau de Arellano, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., García-Bustamante, E., and Dudhia, J.: The Effect ofHeat Waves and Drought on Surface Wind Circulations in the Northeast of the Iberian Peninsula during the Summer of 2003, J. Climate, 24, 5416–5422, https://doi.org/10.1175/2011JCLI4061.1, 2011.

Jiménez, P. A., Dudhia, J., Gonzalez-Rouco, J. F., Navarro, J., Montavez, J. P., and Garcia-Bustamante, E.: A Revised Scheme for the WRF Surface Layer Formulation, Mon. Wea. Rev., 140, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1, 2012.

Kain, J. S.: The Kain–Fritsch convective parameterization: An update., J. Appl. Meteorol. Clim., 43, 170–181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Kalverla, P. C., Steeneveld, G. J., Ronda, R. J., and Holtslag, A. A.: An observational climatology of anomalous wind events at offshore meteomast IJmuiden (North Sea), J. Wind Eng. Ind. Aerodyn., 165, 86–99, https://doi.org/10.1016/j.jweia.2017.03.008, http://dx.doi.org/10.1016/j.jweia.2017.03.008, 2017.

Katragkou, E., Garciá-Diéz, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., and Jacob, D.: Regional climate hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble, Geosci Model Dev., 8, 603–618, 2015.

Kruse, C., Vento, D. D., Montuoro, R., Lubin, M., and McMillan, S.: Evaluation of WRF scaling to several thousand cores on the Yellowstone supercomputer, in: Proceedings of the Front Range Consortium for Research Computing Conference, 2013.

Lee, J. A., Kolczynski, W. C., McCandless, T. C., and Haupt, S. E.: An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction, Mon. Wea. Rev., 140, 2270–2286, https://doi.org/10.1175/MWR-D-11-00065.1, 2012.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F., and Smith, J. A.: Development and Evaluation of a Mosaic Approach in the WRF-Noah Framework., J. Geophys. Res., 118, 11,918–11,935, https://doi.org/10.1002/2013JD020657, 2013.

Li, S., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., Wallom, D. C. H., Betts, R. A., and Wettstein, J. J.: Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation, Geosci Model Dev., 12, 3017–3043, https://doi.org/10.5194/gmd-12-3017-2019, 2019.

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., and Beltrami, H.: Quality Control of surface wind observations in North Eastern North America. Part I: Data Management Issues, J. Atmos. Ocean. Tech., 35, 163–182, https://doi.org/10.1175/JTECH-D-16-0204.1, 2018a.

5   Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., Beltrami, H., and Conte, J.: Quality control of surface wind observations in North Eastern North America. Part II: Measurement errors, J. Atmos. Ocean. Tech., 35, 183–205, https://doi.org/10.1175/JTECH-D-16-0205.1, 2018b.

Lupu, N., Selios, L., Warner, Z., et al.: A new measure of congruence: The Earth Mover's Distance, Political Analysis, 25, 95–113, https://doi.org/10.1017/pan.2017.2, 2017.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys. and Space Phys.,

10   20, 851–875, 1982.

Mortensen, N. G., Heathfield, D. N., Rathmann, O., and Nielsen, M.: Wind Atlas Analysis and Application Program: WAsP 10 Help Facility, Tech. rep., DTU Wind Energy, https://orbit.dtu.dk/files/116352660/WAsP_10_Help_Facility.pdf, 2011.

Nakanishi, M. and Niino, H.: An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog., Bound.-Layer Meteor., 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

15   Nakanishi, M. and Niino, H.: Development of an improved turbulence closure model for the atmospheric boundary layer., J. Meteor. Soc. Japan, 87, 895–912, https://doi.org/10.2151/jmsj.87.895, 2009.

Nawri, N., Petersen, G., Bjornsson, H., Hahmann, A., Jónasson, K., Hasager, C., and Clausen, N.-E.: The wind energy potential of Iceland, Renew. Energ., 69, https://doi.org/10.1016/j.renene.2014.03.040, 2014.

NCAR: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, https://doi.org/10.5065/D6M043C6,

20   2000.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., et al.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12 109, https://doi.org/10.1029/2010JD015140, 2011.

Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, Mon. Wea. Rev., 117, 536–549,

25   https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

Olsen, B. T., Hahmann, A. N., Sempreviva, A. M., Badger, J., and Jørgensen, H. E.: An intercomparison of mesoscale models at simple sites for wind energy applications, Wind Energy, 2, 211–228, https://doi.org/10.5194/wes-2-211-2017, 2017.

Olson, J., Kenyon, J., Brown, J., Angevine, W., and Suselj, K.: Updates to the MYNN PBL and surface layer scheme for RAP/HRRR, NOAA Earth System Research Laboratory, Boulder, CO, USA, http://www2.mmm.ucar.edu/wrf/users/workshops/WS2016/oral_presentations/6.

30   6.pdf, 2016.

Pele, O. and Werman, M.: A linear time histogram metric for improved sift matching, in: Computer Vision–ECCV 2008, pp. 495–508, Springer, 2008.

Peña, A.: Østerild: A natural laboratory for atmospheric turbulence, J. Renew. and Sust. Energ., 11, 063 302, https://doi.org/10.1063/1.5121486, 2019.

35   Peña, A., Floors, R., Sathe, A., Gryning, S.-E., Wagner, R., Courtney, M. S., Larsén, X. G., Hahmann, A. N., and Hasager, C. B.: Ten Years of Boundary-Layer and Wind-Power Meteorology at Høvsøre, Denmark, Boundary-Layer Meteorol., 158, 1–26, https://doi.org/10.1007/s10546-015-0079-8, 2015.

Petersen, E. L.: In search of the wind energy potential, J. Renew. and Sust. Energ., 9, 052 301, https://doi.org/10.1063/1.4999514, 2017.

Pinard, J. D. J.-p., Benoit, R., and Yu, W.: A WEST wind climate simulation of the Mountainous Yukon, Atmosphere-Ocean, 43, 259–282, https://doi.org/10.3137/ao.430306, 2005.

Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing., J. Appl. Meteorol. Clim., 46, 1383–1395, 2007.

5   Poulter, B., MacBean, N., Hartley, A., and coauthors: Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative, Geosci. Model Dev., 8, 2315–2328, https://doi.org/10.5194/gmd-8-2315-2015, 2015.

Rabin, J., Delon, J., and Gousseau, Y.: Circular Earth Mover's Distance for the comparison of local features, in: 2008 19th Int. Conf. Pattern Recognit., pp. 1–4, IEEE, https://doi.org/10.1109/ICPR.2008.4761372, 2008.

10   Refslund, J., Dellwik, E., Hahmann, A. N., Barlage, M. J., and Boegh, E.: Development of satellite green vegetation fraction time series for use in mesoscale modeling: application to the European heat wave 2006, Theor. Appl. Climatol., 117, 377–392, https://doi.org/10.1007/s00704-013-1004-z, 2014.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

15   Reynolds, R. W., Gentemann, C. L., and Corlett, G. K.: Evaluation of AATSR and TMI Satellite SST Data, J. Climate, 23, 152–165, https://doi.org/DOI 10.1175/2009JCLI3252.1, 2010.

Rife, D. L. and Davis, C. A.: Verification of temporal variations in mesoscale numerical wind forecasts, Mon. Wea. Rev., 133, 3368–3381, http://journals.ametsoc.org/doi/abs/10.1175/MWR3052.1, 2005.

Rubner, Y., Tomasi, C., and Guibas, L. J.: The Earth Mover's Distance as a Metric for Image Retrieval, International Journal of Computer
20   Vision, 40, 99–121, https://doi.org/10.1023/A:1026543900054, https://doi.org/10.1023/A:1026543900054, 2000.

Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J., and Tovar-Pescador, J.: Influence of land-use misrepresentation on the accuracy of WRF wind estimates: Evaluation of GLCC and CORINE land-use maps in southern Spain, Atmos. Res., 157, 17–28, https://doi.org/10.1016/j.atmosres.2015.01.006, 2015.

Siuta, D., West, G., and Stull, R.: WRF hub-height wind forecast sensitivity to PBL scheme, grid length, and initial condition choice in
25   complex terrain, Wea. Forecasting, 32, 493–509, https://doi.org/10.1175/WAF-D-16-0120.1, 2017.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, Tech. Rep. NCAR/TN-475+STR, National Center for Atmospheric Research, 2008.

Smith, E. N., Gibbs, J. A., Fedorovich, E., and Klein, P. M.: WRF model study of the Great Plains low-level jet: Effects of grid spacing and
30   boundary layer parameterization, J. Appl. Meteorol. Clim., 57, 2375–2397, https://doi.org/10.1175/JAMC-D-17-0361.1, 2018.

Strobach, E. and Bel, G.: Regional Decadal Climate Predictions Using an Ensemble of WRF Parameterizations Driven by the MIROC5 GCM, J. Appl. Meteorol. Clim., 58, 527–549, https://doi.org/10.1175/JAMC-D-18-0051.1, 2019.

Tammelin, B., Vihma, T., Atlaskin, E., Badger, J., Fortelius, C., Gregow, H., Horttanainen, M., Hyvönen, R., Kilpinen, J., Latikka, J., Ljungberg, K., Mortensen, N. G., Niemelä, S., Ruosteenoja, K., Salonen, K., Suomi, I., and Venäläinen, A.: Production of the Finnish
35   Wind Atlas, Wind Energy, 16, 19–35, https://doi.org/10.1002/we.517, 2013.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R. H.: Implementation and verification of the unified Noah land surface model in the WRF model., in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, Seattle, 12-16 January 2004, AMS, 2004.

Thompson, D. R., Horstmann, J., Mouche, A., Winstead, N. S., Sterner, R., and Monaldo, F. M.: Comparison of high-resolution wind fields extracted from TerraSAR-X SAR imagery with predictions from the WRF mesoscale model, J. Geophys. Res., 117, C02 035, https://doi.org/DOI 10.1029/2011JC007526, 2012.

Troen, I. and Petersen, E. L.: European Wind Atlas, Published for the Commission of the European Communities, Directorate-General for Science, Research, and Development, Brussels, Belgium by Risø National Laboratory, https://backend.orbit.dtu.dk/ws/portalfiles/portal/112135732/European_Wind_Atlas.pdf, 1989.

Vincent, C. L. and Hahmann, A. N.: The Impact of Grid and Spectral Nudging on the Variance of the Near-Surface Wind Speed, J. Appl. Meteorol. Clim., 54, 1021–1038, https://doi.org/10.1175/JAMC-D-14-0047.1, 2015.

von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, 1999.

Wang, W., Dudhia, J., and Chen, M.: Application of WRF - How to get better performance, National Center for Atmospheric Research, Boulder, CO, USA, http://www2.mmm.ucar.edu/wrf/users/tutorial/201901/chen_best_practices.pdf, 2019.

Westerhellweg, A., Neumann, T., and Riedel, V.: FINO1 Mast Correction, https://pdfs.semanticscholar.org/cf85/2b7bc731b071162e537edf45f9578f4ec86e.pdf, 2012.

Wijnant, I., van Ulft, B., van Stratum, B., Barkmeijer, J., Onvlee, J., de Valk, C., Knoop, S., Kok, S., Marseille, G., Baltink, H. K., and Stepek, A.: The Dutch Offshore Wind Atlas (DOWA): Description of the dataset, Tech. Rep. TR-380, Royal Netherlands Meteorological Institute (KNMI), https://www.dutchoffshorewindatlas.nl/, 2019.

Witha, B., Hahmann, A., Sile, T., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Leroy, G., and Navarro, J.: WRF model sensitivity studies and specifications for the NEWA mesoscale wind atlas production runs, Tech. rep., Carl von Ossietzky University of Oldenburg, https://doi.org/10.5281/ZENODO.2682604, 2019.

Wohlfart, C., Winkler, K., Wendleder, A., and Roth, A.: TerraSAR-X and Wetlands: A Review, Remote Sensing, 10, https://doi.org/10.3390/rs10060916, 2018.

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the Weather Research and Forecasting model, Bound.-Lay Meteorol., 162, 117–142, https://doi.org/10.1007/s10546-016-0185-2, 2017.