Final author response

Bjarke Tobias Olsen, Andrea N. Hahmann, Anna Maria Sempreviva, Jake Badger, Hans E. Jørgensen

The comments from the reviewers highlighted important issues that we have adressed. We have made a number of changes to the manuscript. The main change is to section 3. We have removed section 3.4, and replaced it with a new section that go into details with the performance of the models related to three specific model options: the PBL scheme, the grid spacing, and the simulation time. We have also divided section 3 into just three subsections: 3.1 "Mean quantities and distributions", 3.2 "Relating performance to model setup", and 3.3 "Wind energy application".

We have omitted the dataless sites from the paper, since all three reviewers made this suggestion.

NB! Due to the many changes made, and because some sections have been moved around, the mark-up of changes is not accurate in some areas, especially in the last half of the paper.

List of changes

Sections removed: 3.4, 3.4.1
Figures removed: 4, 10, 11, 12
Tables removed: A2, A3

New figure: (fig 6) A figure with two subplots: one that shows the distributions of modelled and observed mean wind speed for each month of the year at the three sites, and one that shows the corresponding distributions of Mean Absolute Error (MAE) for wind speed, for the models, for each month of the year, at the three sites.
New figure: (fig 7) Figure that shows the distributions of MAE for wind speed, for the models, for five classes of atmospheric stabilities.
New figure: (fig 11) Figure that shows the Root Mean Squared Error (RMSE) for the shear exponent vs the normalized RMSE (NRMSE) for wind speed for the modes at the three sites. This is used in the new section "Relating performance to model setup".
Three new tables: (table A5, A6, and A7) Aggregate statistics (mean, median, std, min, max) of RMSE for the shear exponent, and NRMSE for wind speed, grouped according to PBL scheme used, the grid spacing, and the simulation time. This is used in the new section "Relating performance to model setup".

Other changes

1. The introduction has been shortened and improved, as per reviewer comments.
2. The language has been improved and typos removed, as per reviewer comments.
3. Dataless sites has been omitted.
4. Added results and discussion of the ability of the models to capture the annual cycle.
5. Added new results and discussion of the model errors in different atmospheric stability regimes.
6. Combined section 2.2 and 2.3.
7. Shortened, improved, and combined section 4 and 5, as per reviewer comments.

Comments to reviewers

Reviewer comments in **bold text**. Author comments in plain text.

Comments to anonymous reviewer #1

The manuscript provides a valuable comparison of NWP models against wind observations from tall towers. The article is well written and it should deserve publication.
Thank you for the feedback.

One aspect that the authors should consider is the inclusion of dataless sites. The comparisons at these sites do not provide much information and could be removed from the manuscript.
Your comment, and that of the other reviewers, suggests that the comparisons at the dataless sites add more noise than value to the manuscript. We agree with your suggestion and propose to remove them from the future manuscript.

A more important aspect is the relative little attention that the authors pay to the effects of atmospheric stability. According to Table A3 Ri and L are provided by the different teams so there is not a clear reason for not analyzing in more detail the important effects of atmospheric stratification. The behavior of the models could be very different under stable/unstable situations.
We agree, and have further analyzed the data with respect to stability. We propose to add a new section on this topic to the manuscript.

Another relevant aspect for wind energy is how well the models represent the annual evolution and the diurnal cycle. More specific comments are provided below.

With respect to the annual cycle we agree, and propose to add a section about that. Regarding the diurnal cycle, the effects related to changes in the atmospheric stratification, which occurs during the diurnal cycle, are represented well by the new results related to stratification. We propose to add a statement in the new manucscript about this, without adding additional figures.

SPECIFIC COMMENTS

1. **Page 1, Line 10. Clarify what is "average wind speed distribution".**
   We agree that clarification is needed, and propose to reprase the abstract. To be clear, what we ment was the mean of all the modelled wind speed distributions.

2. **Page 2, Line 13. Can you quantify instead of saying "does a much better job"?**
   We suggest changing it to "provides a better representation"

3. **Page 2, line 15-16. The open statement of the paragraph says "many different climates and terrains" but all the examples are for northern Europe. It is better to change the opening sentence or enlarge the number of examples.**
   We propose to rephrase the opening sentence.

4. **Page 2, line 32. Clarify what do you mean by "the observed mean wind speed". Do you mean simulated wind speed?**
   Yes, we agree that this should be corrected.

5. **Page 3, line 8. An important conclusion of Gomez-Navarro et al. is to account for the effects of unresolved topography in the WRF model.**
   We agree that this should be clarified.

6. **Page 3, lines 32-34. Clarify what do you mean by "little knowledge has been derived from assessing the operational NWP models run by the community".**
   We agree that it needs clarification, and propose to rephrase the opening of the paragraph to: "Community-driven model intercomparison projects provide an opportunity to study both model uncertainties, and sensitivities to model components."

7. **Page 7, line 30. What is the distribution of the vertical levels near the surface?**
   Approximately 10, 34, 69, 118, 187 and 275 m. we agree that detail should be added to the manuscript.

8. **Page 8, line 20. Why do you want to remove outliers? In the case of observations you may question the validity of the data but in the case of the simulations you do not question this so you should not remove them.**
   We would like to present the general performance of the models with aggregated statistics. We chose the intermodel mean and standard deviation for

this. In some cases, the output from one or two model(s) is very different from the other models (> 3.5 intermodel standard deviations away from the intermodel mean), which would heavily skew the intermodel mean and standard deviation if included. Since it is so few models we are talking about, we decied to leave them out of the aggregate. The models that are left out are still shown, and the methods we use to calculate the intermodel mean and standard deviation are clearly defined, which makes it completely transparent for the reader.

9. **Page 11, line 7. Jimenez et al. (2016) compared 10 years of observations and WRF simulations at Cabauw. They already pointed out the reduction of the bias with height at this site. You should probably mention this previous work to construct on its findings.**
   Thank you for mentioning this paper, we agree that a reference in the manuscript is appropriate.

10. **Page 16, line 2. Do you think the temporal interpolation is also responsible for the poor results?**
    That is an excellent point. The poor results are, as you say, to a large degree a result of the vertical and temporal interpolation. This should be stressed in the new version of the manuscript.

11. **Fig. 10: Is it correct that some models have a bias of about 20 m/s at Cabauw? That's a very large bias, something looks wrong with that model(s).**
    Thank you for catching this. The unit was wrong, and should have been % not m/s. However, we suggest removing this section from the manuscript, as per the reviewer responses.

12. **Page 21, line 1. Two consecutive "used".**
    Thank you.

13. **Page 22, line 7. Tow consecutive "submitted".**
    Thank you.

14. **Table A.5. The fifth row should be the third one according to the horizontal grid spacing.**
    Thanks. Fixed.

References:
Jimenez, PA, J Vila-Guerau de Arellano, J. Dudhia, F. Bosveld, 2016: Role of synopticand meso-scales on the evolution of the boundary-layer wind profile over a coastal region: the near-coast diurnal acceleration. Meteorol. Atmos. Phys., 128, 39-56.

Comments to anonymous reviewer #2

The manuscript provides a comparison of 25 atmospheric forecasts with mast observations for three different locations with a focus on wind energy related parameters. While the undertaking itself is very important for the wind energy community given the collected data especially from multiple commercial sources. In my opinion, however, there are several issues that need to be addressed before publication.

In general, the language of the manuscript needs some improvement. I gave several corrections in my detailed comments but I suggest a native speaker or a professional editing service to correct all of the numerous (small) errors. Further, I recommend to use present tense instead of past tense for most of the manuscript.
Thank you. We are reviewing the paper accordingly.

The section "Introduction" is too long and needs to be much more concise. Often, the authors do not only cite the essence of a referenced paper, but also provide additional detail about it which does not add value to the actual message. An example for this can be found on page 2 line 28ff: The authors cite Hahmann et al. (2014b) with an explanation on what was done in the study before adding the sentence "A year long wind climatology simulation was used as the test variable". This information is too detailed and can easily be omitted without lessening the message itself. Further, the introduction contains a lot of abbreviations. Some of these are even not used later in the manuscript, e.g. LCOE.
We agree with both points made, and propose to improve the text in both respects.

The use of the three comparison sites without measurements seems to be unneces- sary. First, I would disagree that the data-less sites resemble the mast sites from a climatological perspective (e.g., wind climatology). Second, at horizontal resolutions down to 1km, comparable sites with a focus on near-surface PBL will be very hard to find. Third, the authors themselves do not provide much detail about the comparison. I suggest to omit this part of the comparison.
Agreed. We propose to remove this from the revised manuscript.

Most of my concerns with the manuscript are with the section "Individual model performance" which provides the results for the major objective of model intercomparison: The authors show that the models differ, but they fail to show why. In my opinion, in a comparison study of model simulations, the attribution of differences among the data sets with respect to the representation of the simulated parameters to the characteristics of the simulation systems is most important. While the authors list multiple such characteristics as potentially crucial to the quality of the simulations, e.g., model, physical process schemes, they fail to show a dependence of the single model results to these characteristics with the exception of showing the dependence

of wind speed error to grid spacing in a very simplistic way. I think the reader as well as the quality of the manuscript would profit from more details, e.g., how do longer forecast lead times or smaller grid spacing reflect on the performance of the models presented in a plot similar to Figure 3.
We tend to agree, and suggest a revision of the section. We propose to remove much of the old content, and to add new results to the revised manuscript that provide an analysis of the model results related to three specific model options: PBL scheme, grid spacing, and simulation time.

I suggest to merge sections 4 and 5 into a "Conclusions"-Section which can contain a summary.
We agree with this.

Detailed comments:
Thank you for catching all these!

1. **Page 2 Line 4: "... as ensemble members ..."**
   Thanks!
2. **Page 3 Line 9: "... sensitivities of the WRF"**
   Thanks!
3. **Page 3 Line 23: "... assessment exist."**
   Thanks!
4. **Page 3 Line 25: "... near surface winds were ..."**
   Thanks!
5. **Page 3 Line 26: "... the WRF model was in better ..."**
   Thanks!
6. **Page 3 Line 32: "... to initial conditions, ..."**
   Thanks!
7. **Page 3 Line 34: What community?**
   We agree, it should be clarified that it is the wind energy community.
8. **Page 4 Line 8: "for a number of reasons: ..."**
   Thanks!
9. **Page 4 Line 10: "... who rely ..."**
   Thanks!
10. **Page 4 Line 17: "... of the simplest terrains ..."**
    Thanks!
11. **Page 6 Line 4f: Can the authors provide a reference for this approach. Why not use the data at 50 and 70 meters?**
    Comparison of the (single anemometer) measurements at 40 and 60 meters to the extrapolated/interpolated measurements indicated that the errors due to flow distortion were much larger than the errors from extrapolation/interpolation. Peña et al. (2016) and Fabre et al. (2014) show that

the impact from flow distortion due to the mast can be large. Pena et al. shows a discrepency of more than 10% between two anemometors at the same height, in the case where one is located upstream and one is downstream from the mast, at Høvsøre.

12. **Page 7 Line 16: Nudging is an assimilation method.**
    We agree, it is redundant.

13. **Page 8 Line 2: "This study is . . . "**
    Thanks!

14. **Page 8 Line 22: Tilde is shifted**
    Thanks!

15. **Page 9 Line 5: "... between two levels . . . "**
    Thanks!

16. **Page 9 Line 12: "... the model output data were . . . "**
    Thanks!

17. **Page 10 Line 7: The variance is given in % but there is no reference to what the numbers refer.**
    Thanks, we should clarify that it is % deviation (relative to the observation).

18. **Page 11 Line 2: "... and the intermodel variance is . . . "**
    Thanks!

19. **Page 11 Line 9: "... mesoscale datasets and ERA-Interim show a significant . . . "**
    Thanks!

20. **Page 11 Line 11: "... varies between . . . "**
    Thanks!

21. **Page 11 Line 13: "... clear that the correlation . . . "**
    Thanks!

22. **Page 11 Line 16: "by at"?**
    Thanks!

23. **Page 12 Line 6: "... instead shows an . . . "**
    Thanks!

24. **Page 13 Line 2: "... dataset does not . . . "**
    Thanks!

25. **Page 13 Line 3: "... and tends to . . . "**
    Thanks!

26. **Page 14 Line 9: "... dataset captures the ..., but shows a . . . "**
    Thanks!

27. **Page 14 Line 14: "... dataset, however, does not . . . "**
    Thanks!

28. **Page 14 Line 18: "... roughness varies a lot . . . "**
Thanks!

29. **Page 16 Line 6: "... Fig. 5), two of the sites are investigated."**
Thanks!

30. **Page 16 Line 7: "... with a strong dependency of surface roughness on the wind direction."**
Thanks!

31. **Page 16 Line 8: "... variation were/are binned . . . "**
Thanks!

32. **Page 18 Line 1: "The hypothesis of this study is that . . . "**
Thanks!

33. **Page 18 Line 3: "... factors are expected to . . . "**
Thanks!

34. **Page 18 Line 4: Please provide more detail: What is meant by "source of orography"?**
Elevation data set. We agree, it should be clarified.

35. **Page 18 Line 7: I would expect that the model itself, initial boundary layer conditions and simulation time aka forecast lead time have a large impact on the model estimates. I wonder why the authors hypothesise that the impact of these factors will be of a lesser degree.**
Initial results did not show any significant impact of these factors. However, we suggest adding new results to the paper, which looks at the model performances related to the PBL scheme, grid spacing, and simulation time.

36. **Page 18 Line 10: "... significant correlations were . . . "**
Thanks!

37. **Page 18 Line 18ff: When calculating correlations for wind speed over such distances (up to 500km), large correlation coefficients are to be expected given the data set used. A better approach would be to filter-out low frequency (e.g. days, weeks, months) variations in the time series in order to retrieve the intra-day wind speed variations. Then these can be used in an analysis to remove the obvious correlations between the mast sites.**
We would like to omit this part of the manuscript completely. We do not believe that it adds enough value to the study. We propose revising this part of the paper.

38. **Page 20 Line ": "... by an underestimation . . . "**
Thanks!

39. **Page 21 Line 1: "... schemes used in . . . "**
Thanks!

40. **Page 21 Line 11: "... largest biases are/were observed ..."**
    Thanks!

41. **Page 21 Line 23: "... to accurately estimate ..."**
    Thanks!

42. **Figures 3 to 9: Why is the MM variance plotted when every single MMi is shown in the diagram?**
    It can be tricky to estimate the spread of 20+ lines that are near eachother, and the standard deviation adds a simple metric to show the spread, while not hiding the lines for each model.

43. **Figure 10: The dashed diagonal is misleading as it suggests that there is meaning to itwhich is not as far as I understand (Model resolution in km against wind speed bias in m/s). Please correct me if I am wrong.**
    We suggest omitting the figure completely, as it does not add enough value to the study. However, we propose to revise this part of the study, and add new results that goes into details about the impact of the grid spacing in the modeling results.

References:

Peña, Alfredo et al. "Ten Years of Boundary-Layer and Wind-Power Meteorology at Høvsøre, Denmark", Boundary-Layer Meteorol (2016) 158: 1. doi:10.1007/s10546-015-0079-8

Fabre, Sylvie, et al. "Measurement and simulation of the flow field around the FINO 3 triangular lattice meteorological mast." Journal of Wind Engineering and Industrial Aerodynamics 130 (2014): 99-107.

Comments to anonymous reviewer #3

This paper presents an interesting comparison of mesoscale models at sites with flat orography. While the study is of relevance for the community, I believe it lacks in some aspects which could be easily fixed. First of all, the introduction seems too long. It could benefit of a condensation of some of the informations reported.
We agree with the comment about the length of the introduction. We propose to shorten and improve it.

Also, one could argue about the need of including the dataless sites in the comparison, since they don't add much value to the study. I would consider of removing them.
We agree, and so did the other reviewers. We propose to entirely omit the dataless sites from the paper.

As the authors state in the conclusions, "While it was a key objective of this study to determine the model setup choices that have a large impact on the models ability to estimate the wind climate accurately in the lowest part of the PBL, only weak indications were found.". I suggest putting more emphasis in trying to describe the differences and advantages/disadvantages of using different model configurations.
We tend to agree, and suggest a revision of the section, leaving out much of the old content, and adding new results that go into more detail with three specific model options: PBL scheme, grid spacing, and simulation time.

Typos:
Thanks alot for finding these.

1. **-page 1 line 3: replace "a" with "an"**
   Thanks!
2. **-page 1 line 15: unnecessary "-"**
   Thanks!
3. **-page 2 line 27: replace "Meller" with "Mellor"**
   Thanks!
4. **-page 3 line 11: replace "spacial" with "spatial"**
   Thanks!
5. **-page 4 line 28: replace "is shown" with "as shown"**
   Thanks!
6. **-page 6 line 1: replace "Cabuaw" with "Cabauw"**
   Thanks!
7. **-page 13 line 24: replace "srpead" with "spread"**
   Thanks!
8. **-page 16 line 6: "exists" is repeated**
   Thanks!

9. **-page 16 line 7: replace "represeting" with "representing"**
   Thanks!

10. **-page 21 line 16: replace "used assess" with "used to assess"**
    Thanks!

An intercomparison of mesoscale models at simple sites for wind energy applications

Bjarke T. Olsen, Andrea N. Hahmann, Anna Maria Sempreviva, Jake Badger, and Hans E. Jørgensen
DTU Wind Energy, Frederiksborgvej 399, 4000 Roskilde, Denmark
Correspondence to: Bjarke Tobias Olsen (btol@dtu.dk)

Abstract. ~~An intercomparison of model results from 25 different~~

Understanding uncertainties in wind resource assessment associated with the use of the output from Numerical Weather Prediction (NWP) models is ~~presented for the year 2011 at six~~ important for wind energy applications. A better understanding of the sources of error reduces risk and lowers costs. Here, an intercomparison of the output from 25 NWP models is presented for three sites in Northern Europe characterized by simple terrain. The ~~model results and a detailed description of each model was submitted by 18 different modeling groups to a open call for data, and serves as a rare quantitative overview of the model uncertainties associated with state-of-the-art mesoscale models used for wind energy applications today. At three of the sites the model intercomparison was~~ models are evaluated using a number of statistical properties relevant to wind energy and verified with observations~~from nearby meteorological masts. The intercomparison was based on statistical properties of the wind for a number of heights at each site.~~

~~The results show better performance of the models and a smaller inter-model spread~~. On average the models have small wind speed biases offshore and aloft (~~$2-4\%$ mean wind speed bias above 40 meters~~$<4\%$), and ~~greater errors and more spread for inland sites and~~ larger biases closer to the surface ~~(up to $7-9\%$ wind speed bias). For the distributions of wind speed, wind direction, and wind shear only small deviations exist between the observations and the average~~ over land ($>7\%$). A similar pattern is detected for the inter-model spread. Strongly stable and strongly unstable atmospheric stability conditions are associated with larger wind speed errors. Strong indications are found that using a grid spacing larger than 3 km decreases the accuracy of the models, but ~~a small shift of the average wind speed distribution towards high wind speeds at Cabauw, and an underrepresentation of strong shear cases was observed. Although the model setup options were studied to determine a 'best practice', no significant indicator was found.~~ we found no evidence that using a grid spacing smaller than 3 km is necessary for these simple sites. Applying the models to a simple wind energy offshore wind farm highlights the importance of capturing the correct distributions of wind speed and direction.

1 Introduction

Numerical Weather Prediction (NWP) models are increasingly being used ~~for wind energy related applications, ranging from~~ in wind energy applications, e.g. wind power resource mapping and site assessment, ~~development and planning of~~ for planning and developing wind farms, ~~to power-forecasting, maintenance~~ power forecasting, for electricity scheduling, maintenance of wind farms, and energy trading on ~~the~~ electricity markets. ~~For~~ In site assessment, NWP models are commonly part of the model chain used ~~for estimation of~~ to estimate the Annual Energy Production (AEP) ~~. However,~~ and are responsible for a large part of the uncertainty ~~in the estimate is contributed by the uncertainties of the NWP model. This, combined with the~~ of this estimate.

The extensive use of ~~these~~ NWP models, and the ~~fact that each modeltypically has multiple options and parameters available for each sub-component~~vast customization-space of each model, means that a strong demand exists for quantification of a) the ~~sensitivity~~ overall model uncertainties, and b) the sensitivity of the uncertainties to the choice of sub-components and parameters~~, and b) the overall model uncertainties. Having a better understanding of~~. Understanding the sensitivities and uncertainties of ~~NWP models can help lower the~~ the NWP model output can reduce their associated risks, and improve decision making~~, which in turn will lower the Levelized Cost of Energy (LCOE). For model users , having access to sensitivity quantification~~. Model users aware of the sensitivity of individual model components ~~enables optimization of~~ will be able to optimize the model setup for specific applications.

In the following, the NWP models ~~in the study will sometimes~~ will be referred to as ~~'mesoscale'~~ "mesoscale" models, signifying that they partly resolve ~~weather~~ atmospheric phenomena in the mesoscale range, defined as the range of horizontal length scales from about one ~~kilometer, up~~ to several hundreds of kilometers ~~, i.e.~~ (Orlanski, 1975).

A common way to assess NWP model uncertainties is to use an ensemble approach, where a number of parallel model runs, referred to as ~~a~~ ensemble members, are run with slightly perturbed initial conditions ~~for each ensemble member, see e.g. Warner (2004) for details~~(Warner, 2004). The magnitude of the perturbations ~~are~~ is typically limited by the uncertainty associated with the particular perturbed variable, in the ~~hope~~ expectation that the ensemble of solutions will cover the solution-space arising from the uncertainties of the input parameters. Ensemble-based techniques are used for many meteorological application, including: precipitation forecasting (Gebhardt et al., 2011; Bowler et al., 2006), wind power ~~generation forecasting (Constantinescu et al., 2011), and resource assessment (Al-Yahyai et al., 2012)~~production forecasting (Constantinescu et al., 2011). However, one would not expect that the ensembles of ~~one~~ any particular modeling system fully represent the uncertainties of another modeling system. This was ~~shown~~ also demonstrated in the DEMETER project (Development of a European ~~Multimodel~~ multi-model Ensemble for seasonal to inTERannual climate prediction) (Palmer et al., 2004), ~~which also demonstrated that a multimodel ensemble approach~~where a multi-model ensemble approach, consisting of a number of different modeling systems, each split into a number of ensembles, ~~does a much better job at representing~~ provided a better representation of the overall uncertainties than any single model ensemble.

~~NWP model sensitivities related to individual model components~~Mesoscale model uncertainties in wind speed near the ground are particularly sensitive to some model components, e.g. the choice of Planetary Boundary Layer (PBL) scheme, the spin up and simulation time, and the grid spacing. In the last couple of decades these sensitivities have been studied in great detail~~in the last couple of decades, in many different climates and terrains. In Northern Europe~~. Vincent and Hahmann (2015), Draxl et al. (2014), and Hahmann et al. (2014) studied the sensitivities of the Weather Research and Forecasting (WRF) model ~~(Skamarock et al., 2008a) have been studied for coastal and offshore locations in several recent papers:~~ (Skamarock et al., 2008b) in offshore and coastal areas in Northern Europe. Vincent and Hahmann (2015) studied the effect of grid nudging, spin-up time, and simulation time, on near-surface and upper PBL wind speed variance~~for the model in the North Sea and the Baltic Sea.~~

~~One 'short' setup with 36 hours simulation time, including 12 hours of spin-up, and no grid nudging, was compared to two 'long' simulations of 11 days, with a one day spin-up, one using grid nudging above the Planetary Boundary Layer (PBL), and one using spectral nudging. The study showed thatwhile~~ . They showed that: 1) spatial smoothing is observed when nudging is used, ~~it has little impact~~ but the impact is small in the lower part of the atmosphere, and ~~it is concluded that the setups using longer simulation time are appropriate for wind energy applications. Draxl et al. (2014) ran~~ 2) nudged longer simulation times (11 days) only have slightly lower variance than short simulations (36 hours), which makes longer simulations appropriate for climatological wind energy studies. Draxl et al. (2014) studied the ability of the WRF model ~~at a coastal site in Denmark with seven different PBL schemes, in order to study how well the model represents profiles of~~ to represent the wind speed and wind shear ~~at heights from 10 to 160 meters, using the different schemes. The study shows~~ profiles at a Coastal site in Denmark using seven different PBL schemes. They showed that the Yonsei University (YSU) (Hong et al., 2006) ~~PBL scheme represented the wind climate~~ scheme represents the profiles best for unstable atmospheric stability conditions, while the Asymmetric Convective Model version 2 (ACM2) (Pleim, 2007b), and the ~~Meller-Yamada-Janjic~~ Mellor-Yamada-Janjic (MYJ) (Janjić, 1994) PBL schemes ~~worked better~~ had more realistic profiles for neutral and stable conditions respectively. ~~In Hahmann et al. (2014)~~ Using the WRF model ~~sensitivities over the Baltic and North Seas were studied for the following model components: the dataset used for initial and boundary condition, the PBL scheme, the number of vertical grid levels, and the source of Sea Surface Temperature (SST) data. A year long wind climatology simulation was used as the test variable. The study~~ for wind resource assessment, Hahmann et al. (2014) showed that the choice of PBL scheme and ~~spin-up time had the strongest~~ spin up time has the greatest impact on the ~~observed~~ simulated mean wind speed for a number of offshore sites, while the ~~other components played a lesser role~~number of vertical levels, and the source of initial conditions had a smaller impact.

~~The~~ Several studies have investigated the WRF model sensitivities ~~have also been studied~~ in regions of complex terrain. Carvalho et al. (2012) studied the sensitivities related to the choice of ~~restart~~ initialization frequency, grid nudging, and suite of Surface Layer (SL) scheme, PBL scheme, and Land Surface Model. They observe that using grid nudging and frequent ~~restarts (every 2 days~~starts (every second day) gives the best agreement for wind speed with several masts located in complex terrain ~~, and that~~ in Portugal. Carvalho et al. (2012) and García-Díez et al. (2013) found a seasonal dependency of the optimal suite of SL-PBL-LSM ~~exists. A seasonal dependency was found by García-Díez et al. (2013) who looked at systematic biases for the year 2011 for all of Europe, by comparing gridded observations, upper-air data, and high-frequency station observations, with three WRF runs using different PBL schemes. They also warn that sensitivity studies based on a limited period may not be representative for the whole year. In Carvalho et al. (2014b) the sensitivities to the choice of~~ for simulating PBL winds and temperature. Carvalho et al. (2014b) investigated the sensitivities related to the SL and PBL scheme ~~is also studied by comparing the model runs to 13 land-based masts and five offshore buoys. The study shows~~ in WRF model at both land and offshore sites in and near Portugal. They showed that the PX SL scheme (Pleim, 2006) ~~and~~ combined with the ACM2 PBL scheme (Pleim, 2007b) gave the smallest errors for wind ~~energy related metrics. However, for offshore only sites the smallest errors were given by the model using~~ speed, and wind energy production estimates, across the sites, while the QNSE-QNSE ~~PBL scheme and SL scheme (Sukoriansky et al., 2005) . A similar study by~~ (SL-PBL) scheme (Sukoriansky et al., 2005) gave smaller errors for offshore sites. In a similar study Gómez-Navarro et al. (2015) analysed the sensitivities of the WRF model to the choice of PBL scheme, ~~grid spacing and setup~~ and grid spacing, in complex terrain in Switzerland. ~~The evaluation metric was the near surface wind at a number of masts all over the country. The study showed, for their suite of setups, that using~~ They found that using a modified version of the YSU PBL scheme, ~~the highest spacial resolution~~ that account for effects of unresolved topography (Jiménez and Dudhia, 2012), in combination with the smallest grid spacing (2 km), and Analysis Nudging, gave the best agreements with measurements during a number of wind storms. Carvalho et al. (2014a) studied the sensitivities of simulating the local wind resource with the WRF model at several masts in Portugal, to the choice of ~~dataset~~ data set used for initial and boundary conditions~~, and showed~~. They show that using the ~~ECMWF (European Center for Medium-range Weather Forecast)~~ ERA-Interim reanalysis ~~dataset~~ data set (Simmons et al., 2007) gave the ~~best agreement with wind measurements at several sites of high wind resource, when compared to the~~ smallest errors, compared to NCEP (National Centers for Environmental Prediction) R2 (Kanamitsu et al., 2002), CFSR (Saha et al., 2010), FNL, and GFS ~~datasets~~data sets, as well as the NASA (National ~~National~~ Aeronautics and Space Administration) MERRA ~~dataset (Rienecker et al., 2011). Schicker et al. (2016) studied the temperature sensitivities~~ data set (Rienecker et al., 2011).

Sensitivities to the choice of ~~land use dataset in two regions in Austria characterized by complex terrain, such sensitivites can be relevant~~ modeling system have also been studied for wind energy ~~if they influence the wind by misrepresenting the atmospheric stability characteristics. Their study showed that, in general, when compared to surface, upper air, and satellite observations of temperature and wind speed, the CORINE CLC06 dataset (Büttner et al., 2004) gave better model agreements than the USGS (Garbarino et al., 2002), and MODIS (Friedl et al., 2010) datasets.~~

~~Several intercomparison studies of NWP models for near-surface wind and resource assessment exists. In Horvath et al. (2012)~~ applicatio Horvath et al. (2012) compared the MM5 (Grell et al., 1994) and WRF models ~~were compared~~ for a site in west-central Nevada characterized by complex terrain. Both models were run in a grid nesting setup from 27 kilometers to 333 meters grid spacing, and the ~~near-surface wind were~~ near surface wind was compared to wind observations from several ~~50 meter~~ 50-meter tall towers. The study showed that ~~WRF model gave~~ the WRF-derived winds were in better agreement with mean wind speed observations, but ~~it suffered from an overestimation of~~ thermally-driven flows were overestimated in both intensity and frequency~~of thermally driven flow. In Hahmann et al. (2015)~~. Hahmann et al. (2015) compared two downscaling methodologies: the KAMM-WAsP (Badger et al., 2014) and WRF Wind Atlas ~~(?)~~(Hahmann et al., 2014) methods, both based on a ~~coupling~~ model chain approach between a NWP model and a ~~simple microscale model using a linearized flow model, were intercompared~~ linearized flow microscale model, for a number of mast sites in South Africa. The study showed that the WRF-based method gave smaller biases than the KAMM-based approach, which ~~were shown to~~ underestimated the wind speeds.

~~So, while extensive work has been put in assessing model sensitivities to initial initial conditions, by e.g. many-model ensemble studies, and into studying the sensitivities to choice of model components, by e. g. case-control studies, little knowledge has been derived from assessing the operational NWP models run by the community. However, in~~ Community-driven model intercomparison projects provide an opportunity to study both model uncertainties, and sensitivities to model components. In the last decade, several intercomparison ~~studies have been succesfully carried out for other types of models~~ projects have been successfully carried out based on model output submitted by modelers from the wind energy community~~, including the Bolund experiment and the CREYAP exercise~~. The Bolund experiment (Bechmann et al., 2011) was an intercomparison of flow models, from simple linearized flow models to Computational Fluid Dynamics (CFD) models. The ~~modelers were asked to model the flow~~ models were compared to measurements around the small island ~~of~~ Bolund in Denmark~~, and the model results were verified using observations from a previous measurement campaign (Berg et al., 2011)~~. The Comparison of Resource and Energy Yield Assessment Procedures ~~(CREYAP) (Mortensen et al., 2015)~~ (CREYAP; Mortensen et al., 2015) was an intercomparison ~~study~~ of energy yield assessment procedures based on four case-studies. The study revealed a large spread ~~among~~ amongst the different procedures, and highlighted the need for further studies into the uncertainties associated with the models ~~. Carrying out a similar type of intercomparison study for NWP models can be~~ themselves. A similar intercomparison of NWP models is attractive for a number of reasons~~, including: 1) It~~. First, it offers an opportunity ~~, not just~~ for model developers, model users, and stake-holders, to get a better understanding of the model uncertainties~~, but also for users and stake-holders in~~. Secondly, a collaborative intercomparison project, which utilizes model data crowd sourced from the wind energy community~~whom rely on these models. 2) It reduces the workload required to carry out comparative studies because the data, in most cases, already exists, and it increases scalability because including additional model results require very little effort~~, increases the scalability of the study compared to traditional ~~studies. 3) Depending on the level of~~ sensitivity studies, by distributing the workload and computational cost among participants. Finally, if sufficient meta-data is collected, it offers a unique insight into ~~what the community considers 'best-practice' when it comes to NWP modeling for wind energy applications~~the "common-practices" in mesoscale modeling within the wind energy community.

In ~~the present study~~ this paper, a blind intercomparison of the output from 25 different NWP ~~models~~ simulations is presented for ~~six simple sites in northern Europe. It is~~ three locations in Northern Europe. The study is based on model output submitted by the modeling community to an open call for model data for a benchmarking exercise co-organized by the European Wind Energy Association (EWEA, now WindEurope) and the European Energy Research Alliance, Joint Programme Wind Energy (EERA JP WIND)~~, and based on model output submitted from the modeling community to an open call for data. The six~~. The three chosen sites represent some of the simplest ~~terrain for the models. They are all located offshoreor in mostly flat and homogeneous~~ terrains: offshore, inland near the coast and inland in flat terrain, where the ~~subgrid-scale parameterizations of the models are expected to work well. Three sites have a tall meteorological mast with observations at many heightsavailable for verification of the model results. The two main aims of the~~ smoothing of the terrain representation is not an issue. The three sites have quality observations from tall meteorological masts with many heights. The main objectives of this study are: 1) To highlight and quantify the uncertainties of the models ~~, and to~~ and serve as motivation ~~and indicator~~ for future analysis of model uncertainties. 2) To identify model setup decisions that have an impact on the model performance. The models are ~~intercompared~~ evaluated using simple metrics relevant ~~for~~ to wind energy applications.

The structure of the paper is as follows~~:~~. In sect. 2 we present a detailed description of the methodology used~~in this study is presented~~, including a description of the ~~six~~ three study sites and the models used by the participants. ~~In sect~~Sect. 3 presents the intercomparison results~~are shown. A discussion of the results is given in sect. 4~~, and finally ~~in sect. 5 the~~ sect. 4 contains the summary and conclusions of the study~~are presented~~.

2 Methodology

2.1 ~~Mast sites~~ Sites and observations

~~The six locations chosen for the intercomparison are labeled~~ Three sites with quality measurements from tall meteorological masts with different terrain characteristics were chosen for this study: (1) FINO3, ~~Høvsøre, Cabauw, and Dataless1-3, is~~ an offshore mast in the North Sea, (2) Høvsøre, a land mast near the Danish west coast, and (3) Cabauw, a land mast in the Netherlands. The mast locations are shown in Fig. 1, and the ~~location and type of the sites are shown in TableA1. Three of the sites correspond to the locations of tall meteorological masts, located in terrain of different characteristics: land, coastal, and offshore. The mast sites have long-term records of observations and serve as the main intercomparison sites. The year~~ coordinates and characteristics of each site are provided in Table A1. Long-term measurements are available from each of the masts, but a single year (2011) was selected as the ~~case-study due to excellent availability of observations. The three observation-less sites were selected because they each resemble one of the mast-sites, and serve to identify whether a consistent pattern of intermodel variance exists between similar sites~~study period due to its excellent data availability.

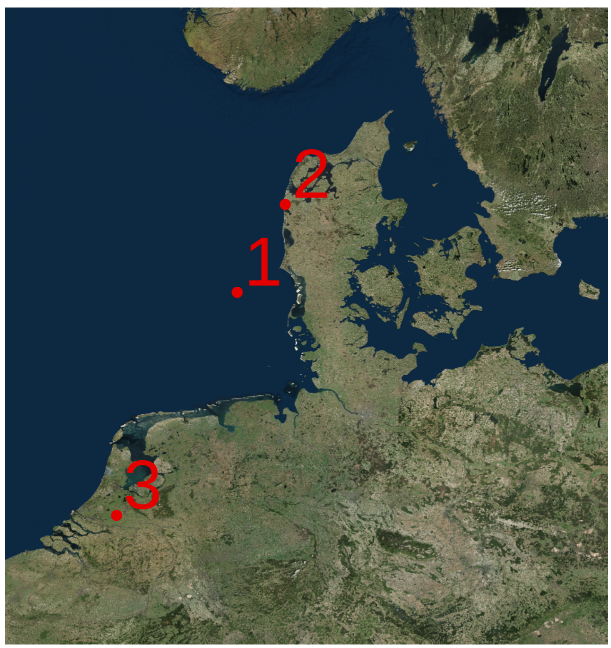


Figure 1. ~~The six~~ Map of Northern Europe with the three site locations used in the model intercomparison: (1) FINO3, in the North Sea. (2) Høvsøre, Denmark. (3) Cabauw, The Netherlands.~~(4) Dataless1, Skagerrak Sea. (5) Dataless2, Bay of Aarhus. (6) Dataless3, Germany.~~

~~The~~ FINO3 ~~site~~ (Fabre et al., 2014) is a ~~platform~~ marine platform located in the North Sea 80 kilometers off the coast of Denmark, with a meteorological mast ~~, reaching an elevation of~~ reaching 120 ~~meters.~~ m above mean sea level (AMSL). We used measurements at 40, 60 and 90 m AMSL in this study. The Høvsøre (Peña et al., 2014) ~~is a mast located approximately~~ mast is located about 2 ~~kilometers from~~ km east of the coastline in western Jutland, Denmark. Apart from the sharp ~~roughness-change represented by the coastline~~surface roughness change at the coastline, and the presence of a small ~~enscarpment, the terrain can be characterized as~~ coastal escarpment, the surrounding terrain is homogeneous and flat. We used measurements at 10, 40, 60, 80, 100 m at this site. The Cabauw mast (Ulden and Wieringa, 1996) is located ~~inland (~~40 ~~kilometers to the coast)~~ km inland near the small towns of Cabauw and Lopik in the Netherlands. The surroundings are flat and characterized by fairly homogeneous agricultural fields, although ~~some~~ with patches of forest and buildings~~are located in the surroundings. The Dataless1 site is located in the Skagerrak Sea approximately 50 kilometers off the northern coast of Denmark. Dataless2 is a site 5 kilometers offshore the east coast of Denmark, near the town of Aarhus. Besides the town and a small enscarpment the landscape is relatively flat and homogenous. Dataless3 is site north-east of the city of Bergen in northern Germany. The surrounding area is relatively flat and homogenous, however some small patches of forest are located nearby~~. Here we used measurements at 10, 20, 40, 80, 140, and 200 m.

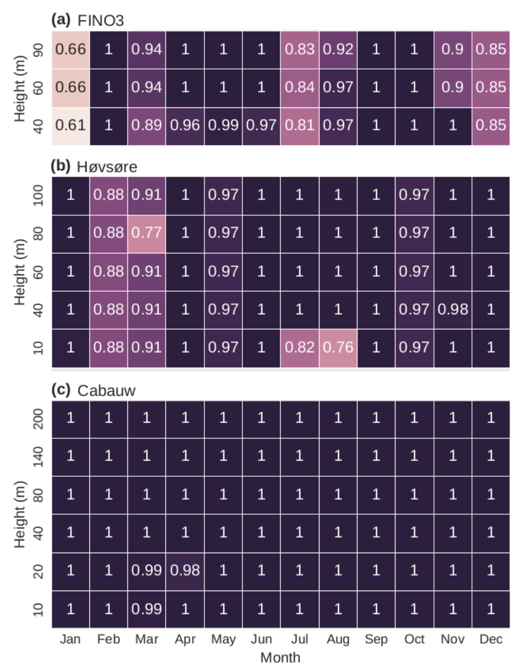


Figure 2. Availability of wind speed and direction ~~data~~ observations for (a) FINO3, (b) Høvsøre, and (c) Cabauw given as ~~a~~ the fraction of completeness for each month of the year 2011 for each ~~comparison~~ height.

~~The wind speed data availability~~ Figure 2 shows availability of wind speed observations for 2011 at the three meteorological masts~~is shown in Fig. 2. At Cabuaw the gaps were filled via interpolation due to the low abundance and magnitude of gaps~~. At Cabauw, the data was gap-filled by simple interpolation as the missing values were few (less than 2% missing data ~~for any given month) . This means that the gap-filled availability was 100%. The observations~~ per month) and the gaps short. The time series from the two other sites were not gap-filled. ~~The intercomparison heights chosen for each site are shown for each site in Table ??, largely chosen because of the placements of instruments on the masts.~~

At FINO3, the wind speed measurements at three of the ~~measurement heights:~~ heights, 50, 70, ~~and~~ 90 ~~meters, comes from three separate booms with cup anemometers separated by a~~ m, are a combination of the measurements from three anemometers at three separate booms 120° ~~angle. This is done to reduce effects of flow distortion from the tower on the wind speed measurements, by combining data from the three anemometors. However, two of the heights used in the intercomparison in this study~~apart. This procedure minimizes the effects of the mast flow distortion. At the other two heights, 40 and 60 ~~meters, had only one cup anemometor available, so~~ m, only one anemometer is available, and the wind measurements are therefore susceptible to flow distortion. Thus, instead of using the ~~single-anemometor~~ single-anemometer data from 40 and 60 ~~meters, the~~ m, the measurements from 50 and 70 ~~meter data was vertically extrapolated down~~ m were vertically interpolated in log height to 40 and 60 ~~meters respectively, using log-law extrapolation in height. This was done under the assumption~~ m. This assumes that the errors due to interpolation and extrapolation are much smaller than those caused by mast flow distortion.

2.2 Submission procedure and models

~~The modelled time-series evaluated in this study were submitted by the participants to the~~ EWEA issued an open call for data ~~issued by the European Wind Energy Association (EWEA). The~~ and the submission procedure consisted of a template spreadsheet and a questionnaire downloadable from the EWEA ~~webpage, which included a questionnaire. The spreadsheet was filled with time-series~~ website. The participants filled the spreadsheet with the time series of the required variables at each location and height. The questionnaire contained ~~queries~~ details about the setup of the modeling system used. The participants ~~then~~ returned the spreadsheet to EWEA, whom passed it on to the authors in an anonymized version.

~~Table ?? shows the~~ The requested model variables were hourly wind speed and direction, air temperature, and atmospheric stability. The questionnaire ~~contained questions detailing~~ asked about the modeling setup, ~~including information about the following:~~ i.e. the model code and version, ~~Surface Layer (SL) scheme, Planetary Boundary Layer (PBL) scheme,~~ the surface and planetary boundary layer schemes, the Land Surface Model (LSM), the grid nests size(s) and spacing(s), the vertical levels, ~~landuse data,~~ the land use data, the length of the simulation and spin-up time, as well as the source of the initial and boundary conditions. ~~Furthermore the participants were~~ The participants were also asked to comment on any additional modifications made to the model, ~~as well as details on what assimilation, nudging, and ensemble~~ including assimilation, ensemble or other methods used.

2.3 ~~Models~~

~~The modeling groups participating with model data are listed in Table A3. There are representatives from~~ Table A3 lists the various groups participating in the exercise. It includes representatives from private companies, universities, research centres, and meteorological institutes. ~~The represented models , including~~ Table A4 summarizes the models and the different model setup options used~~, are shown in Table A4.~~

~~It is clear from Table A4 that the~~ . The WRF model is by far the most commonly used model in the ~~most used in this~~ study, with 18 out of 25 ~~groups using it. It is also clear that the Noah~~ models (Table A4). The Noah LSM was the most ~~popular LSM~~ common LSM used, and the Era-Interim Reanalysis the most ~~popular~~ common source of boundary and initial conditions. The ~~choice of PBL scheme and source of landcover data~~ PBL scheme used and the source of land cover data were more varied amongst the participants. ~~The simulation length , including spinup time, of individual sub-simulation for most of the setups used was~~ Most models used a maximum simulation length of less than 100 hours, including the spin-up time (most typically 12 hours ~~of~~ spin-up and 36 hours of total simulation~~. However, it did vary~~). The simulation and spin-up length ranged from 1 hour ~~spin-up~~ and 7 simulation ~~up~~ to continuously running for the full year.

~~As a source of reference,~~ For reference, wind time series from the ERA-Interim reanalysis (Dee et al., 2011) ~~was~~ were included in the comparisons whenever possible. The ERA-Interim reanalysis ~~dataset~~ data set is a global ~~dataset~~ data set based on extensive assimilation of surface and upper-air observations~~to the IFS global model using 4D-Var (Courtier et al., 1994). The spatial resolution of the dataset is approximately~~ . The data is available on a grid spacing of about 80 km in the horizontal with 60 vertical levels, ~~so bilinear interpolation was used for interpolation to the site locations. To get data on appropriate height levels~~ with values at approximately 10, 34, 69, 118, 187 and 275 m above the model surface. We used bilinear interpolation to interpolate to the sites coordinates, and linear interpolation in ~~height was used. The dataset comes~~ the vertical. The data set is available in 6 hour intervals~~at 0, 6~~, ~~12, and 18 Coordinated Universal Time (UTC), so~~ thus linear interpolation in time was used to ~~fill the sampling gaps~~obtain hourly samples.

2.3 Statistical methods

This study ~~was~~ is based on direct comparison between the observations and model output at collocated positions, as well as intercomparison ~~between the output from the models~~of the modelled output. The sampling frequency for the study was chosen to be one hour. For the observation data this means ~~hour mean values, while~~ hourly mean values; for the mesoscale models ~~instantaneous values are used. This was done based on the assumption that the intra-hourly variance is low for the models, such that instantaneous values are very similar to the hourly means~~the inter-hourly variation is small, so instantaneous values were used. To ensure temporal ~~collocation, missing observations were used as a mask for~~ consistency between observations and modelled output, instances of missing data from the observations were removed from the modeled output. Furthermore, to get ~~vertically consistent~~ consistent vertical profiles, only ~~observations~~ instances where all heights for a particular mast had available data were used. The model ~~outputs submitted by the participants~~ output submitted were assumed to be quality checked by the submitter, but it was also checked by the authors for obvious ~~nonphysical~~ non-physical or inconsistent behavior, and ~~removed~~ not used in that case.

~~From the variables presented in Table ??,~~

Inter-model mean and inter-model variations

The emphasis of this study is on the wind speed, u, and wind direction, as they are the most important ~~ones~~ variables for wind energy applications~~and was emphasized the most~~. In the following, a subscript ~~m will signify~~ m signifies the temporal mean of

~~that~~ a variable, i.e. u_m is the temporal mean wind speed. This is not to be confused with the mean value of the ~~models, which we denote~~ model-ensemble, also referred to as the inter-model mean, which is denoted with a tilde~~, i.e.~~. For example, the mean of the ~~model mean of temporal means~~ model-ensemble for the temporal mean wind speed is denoted $\tilde{u}_m$, and calculated as:

$$\tilde{u}_m = \frac{1}{N}\sum_{i}^{N} u_{m,i} \tag{1}$$

Here ~~the index~~ i is ~~a reference to a specific model submission~~the model index, and N is the total number of models. Likewise, it is useful to define ~~the variation:~~ its standard deviation:

$$\tilde{\sigma}_{u_m} = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(u_{m,i} - \tilde{u}_m\right)^2}, \tag{2}$$

~~Here σ~~ which is the standard deviation ~~, in this case of the intermodel~~ of the inter-model variation between the temporal model means. Since $\tilde{u}_m$ and σ_{u_m} are both sensitive to outliers, ~~so the procedure applied in this study was~~we used the following procedure:

1. Calculate $\tilde{u}_m$ and $\tilde{\sigma}_{u_m}$
2. Remove models ~~where $|u_{m,i} - \tilde{u}_m|$ is greater than $3.5\tilde{\sigma}_{u_m}$~~ whose mean $|u_{m,i} - \tilde{u}_m| > 3.5\,\tilde{\sigma}_{u_m}$
3. Recalculate ~~$\tilde{u_m}$~~ $\tilde{u}_m$ and $\tilde{\sigma}_{u_m}$ with the new subset of models

~~This was done in an effort to reduce the sensitivity to the outliers.~~ The value of ~~3.5 standard deviations~~ $3.5\,\tilde{\sigma}_{u_m}$ was chosen somewhat arbitrarily to ensure that only ~~somewhat ‘extreme’ outliers is removed. Only submission with data~~ "extreme" outliers were removed. The procedure included only models with output available at all ~~heights was included in the calculation of the inter model mean and variation. This was done to get vertically consistent values~~the heights, to ensure a vertically consistent profile of the mean and its variation. Typically, only one or two models were removed by this criteria.

2.3.1 ~~Coefficient of variation~~

Coefficient of variation

~~At the six sites used in this study the variation of wind speed σ_u scales~~ Variations in wind speed often scale with the mean wind speed~~u_m, so~~. Thus, to allow for intercomparison of wind speed variation intensity across vertical levels we define the coefficient of variation, $C_{v,u}$ ~~was used~~. It is defined as the ~~variation over the mean~~ratio of the standard deviation and the mean, σ_u/u_m, and is a unit-less measure of the relative variation at the sampling time scale. At timescales of seconds it is known as ~~turbulent~~ the turbulence intensity, but in this case, with a sampling frequency of one hour, it represents the intensity of variations of synoptic- and mesoscale weather phenomena.

2.3.1 ~~Wind speed shear exponent~~

Wind speed shear exponent

~~The shear exponent(~~To diagnose the wind sheer in the boundary layer, we use the wind sheer exponent, α~~) given by eq. (??) provides a measure of the relative change of wind speed with height between to levels~~ , which uses the wind speed u_1 and u_2 at two heights z_1 and z_2~~.~~ , given by the expression:

$$\cancel{\alpha} u_2 = \cancel{\frac{\ln\left(\frac{u_1}{u_2}\right)}{\ln\left(\frac{z_1}{z_2}\right)}} u_1 \left(\frac{z_2}{z_1}\right)^{\alpha} \tag{3}$$

In the surface layer α is strongly influenced by the surface roughness and the atmospheric stability. ~~It is important that the mesoscale models capture the distributions of~~ By comparing the modelled to the measured α ~~well, because it is an indirect measure of how well the mesoscale models capture the local effectsof roughness and stability~~it is thus possible gain insights into how the model captures these effects.

Error metrics

The Root Mean Squared Error (RMSE) and the Normalized RMSE (NRMSE) were used as error metrics to obtain single value measures of the error across heights at a site. The RMSE and NRMSE are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(x_j^M - x_j^O\right)^2}, \tag{4}$$

$$NRMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^M - x_j^O}{x_j^O}\right)^2}, \tag{5}$$

for a set of n modelled values x_j^M and observed values x_j^O. The RMSE was used for variables that do not scale with height in the surface layer, e.g. wind speed shear exponent; the NRMSE was used for variables that do scale with height, e.g. wind speed.

2.4 Wind energy application

To investigate the errors ~~and spread of the models for simple applied~~ associated with the use of each model in wind energy applications, ~~the models output were used for a wind resource assesment exercise~~we performed a simple wind resource assessment exercise, using both measurements and modelled time series at FINO3.

A typical approach to resource assessment ~~in the wind energy sector~~ is to run a mesoscale model ~~in the area of interest~~ for a number of years, followed by a downscaling process where the ~~wind climate~~ wind-climate statistics obtained from the mesoscale model ~~is~~ are used as input ~~for~~ to a microscale model (Badger et al., 2014; Hahmann et al., 2015). In simple terrain, the microscale model usually consists of a ~~simple flow model , similar to~~ flow model like the one used by the Wind Applications and Analysis Program (WAsP). WAsP uses a linearised flow model based on ~~the principles of Jackson and Hunt (1975), and consists~~ Jackson and Hunt (1975). The procedure in WAsP consists first of an upscaling~~procedure~~ , where local effects from variations in orography, surface roughness, and ~~objects~~obstacles, are removed from the ~~wind climate~~ wind-climate statistics. This is referred to as ~~‘generalisation’~~ "generalisation" of the wind climate, ~~and the generalized statistics are~~ which makes it representative for a larger ~~surrounding~~ area than the site specific wind climate. The size of ~~the area that it represents~~ this area depends on the complexity of the surface roughness, and orographic variations in that area. ~~The generalised wind climate can then be downscaled to a specific site of interest by ‘reversing’~~ To obtain a site-specific wind climate at a new site in this area, the generalised wind climate is downscaled by "reversing" the generalization process, i.e. ~~putting back in the site specific~~ by introducing the site-specific effects of orography, surface roughness, and ~~objects~~obstacles of the new site.

Given ~~a downscaled wind climate and a turbine-specific~~ the wind-climate and the turbine power curve, the expected power output can ~~then be calculated~~ be calculated for any site. Since the participants in this intercomparison ~~did not submit their~~ were not requested to submit the model-specific orography and roughness maps near each site, it is not possible to go through the generalization procedure, and subsequent downscaling process at the inland sites. However, for the offshore site FINO3 there are no effects of orography, and ~~if differences in surface roughness is assumed to negligible between~~ the differences in roughness between the models can be assumed to be negligible. Therefore, we can use the raw model output at this site to estimate the wind resources estimated by each of the models, ~~then the downscaling process can be applied without the generalization. This was done for the FINO3 site~~ without the generalisation procedure.

We performed the wind resource exercise at 90 ~~meters, assuming~~ m at FINO3, assuming first a single Vestas V80 turbine ~~(s)~~ at the site, and then repeated for the wind farm of Horns Rev, which is a 80 turbine ~~large~~ wind farm located near FINO3. ~~This was done using~~ The resource estimations for the wind farm includes the simple wake ~~parameterization~~ parametrization present in the WAsP model~~to study how wake effects can alter the results~~, which was used to estimate the power losses.

3 Results

3.1 Mean ~~quantatives~~quantities and distributions

3.1.1 ~~Mast sites~~

The following subsection is dedicated to the general performance of the models, and their ability to capture the mean and the distributions of a number of wind-related quantities. As previously stated, the goal is to highlight the weaknesses of the models to encourage further analysis of model sensitivities.

~~From~~

3.1.1 Mean wind speed

Figure 3 shows the vertical profiles of mean wind speed (u_m)~~, presented in Fig.3, it is clear that at the offshore site FINO3 most mesoscale models (MM~~at the three sites. At FINO3 (Fig. 3a), most Mesoscale Models (MMs) underpredict u_m at ~~the three~~ all heights. However, the bias on average is less than ~~$0.27\,\mathrm{m/s}$ on average, corresponding to about 2.8%, which~~ $0.27\,\mathrm{m\,s^{-1}}$ ($\sim 2.8\%$). This is a small ~~number, especially compared to~~ bias compared to that of the ERA-Interim data, which shows a larger ~~negative~~ bias than all ~~of~~ the mesoscale models. The ~~intermodel~~ inter-model variance $\tilde{\sigma}_{u_m}$ at FINO3 is ~~$2.7-3.1\%$, decreasing with height, which~~ 2.7–3.1% of the inter-model mean, and decreases with height. That is the lowest combined inter-model variance of any of the ~~six~~ three sites.

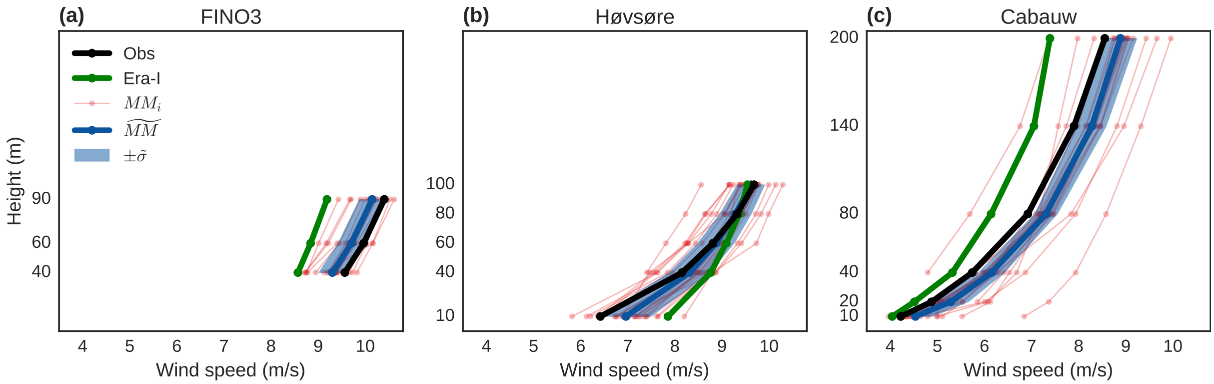


Figure 3. Vertical profiles of mean wind speed (u_m) ~~, for all of 2011,~~ at the ~~six~~ three sites ~~,~~ for: the observations (black), the ERA-Interim ~~dataset~~ data set (green), the Mesoscale Models MM_i (red), and the ~~mesocale models~~ inter-model mean $\widetilde{MM}$ (blue line) and ~~intermodel variance $\widetilde{MM} \pm \tilde{\sigma}$~~ its standard deviation $\pm\tilde{\sigma}$ (blue shade).

At Høvsøre ~~the mesoscale models and ERA-Interim~~ (Fig. 3b), the MMs generally have small wind speed ~~biasses~~ biases above 10 ~~meters, showing a bias of the mesoscale model mean $\tilde{u}_{\mathrm{m}}$ that~~ m. The error of the inter-model mean of the models is smaller than ~~$\pm 0.16\,\mathrm{m/s}$ ($< \pm 1.9\%$~~$\pm 0.16\,\mathrm{m\,s^{-1}}$ ($\sim 1.9\%$), and ~~a intermodel variance is $3.0-5.2\%$~~the inter-model variance is 3.0–5.2%, decreasing with height, which is low compared to the ~~other sites~~biases at the other site on land (Fig. 3c). At 10 ~~meters the mesoscale models generally~~ m, most MMs overpredict the mean wind speed~~and the model mean wind speed $\tilde{u}_{\mathrm{m}}$~~. The inter-model mean has a positive bias of ~~$0.54\,\mathrm{m/s}$ ($\sim 8.41\%$).~~ $0.54\,\mathrm{m\,s^{-1}}$ ($\sim 8.4\%$). The largest inter-model variance is also seen at 10 m (7.8%). The ERA-Interim also overpredicts the mean wind speed at 10 ~~meters, with an even~~ m, with a larger bias than $\tilde{u}_{\mathrm{m}}$~~, and the largest intermodel variance is also observed there (7.8%)~~. Above 10 m, ERA-Interim has smaller errors, but the shape of the profile is not well captured. Signs of a "kink" in both the observed and modelled profiles are present, which could indicate the transition from the low surface roughness of the sea to the higher surface roughness inland.

At Cabauw (Fig. 3c), most of the ~~mesoscale models~~ MMs overpredict u_m. Only one of the ~~mesoscale and the reanalysis datasets,~~ models and the ERA-Interim shows a significant underprediction, and in the case of the reanalysis, this underestimation increases with height. The overprediction by the rest of the ~~mesoscale models~~ MMs varies in magnitude, but the average of the models, excluding the outliers, ~~are~~ is in the range ~~$4-9\%$ for~~4–9% across the different heights~~, with the~~. The largest relative errors ~~near the surface. The inter model~~ are at the lowest levels. The inter-model variance ($\tilde{\sigma}_{u_m}$) at Cabauw ~~vary between $3.3-8.1\%$ at~~ varies between 3.3–8.1% across the different heights, and is ~~highest~~ largest at the lowest levels. The decrease of wind speed bias with height was also observed by Jiménez et al. (2016), whom associated this with excessive turbulent mixing, which may be caused by a misrepresentation of the surface roughness length.

~~Vertical profiles of correlation coefficient between model and observation for wind speed for all of 2011 at the three mast sites, for: ERA-Interim (green), the mesoscale models~~ ~~MM_i~~ ~~(red), and the mesoscale models mean and intermodel variance~~ ~~$\widetilde{MM} \pm \tilde{\sigma}$~~ ~~(blue).~~

~~Wind speed correlation coefficients between the models and the observations are shown in Fig. ??. The figure reveals thatthe correlation between the mesoscale models and the observations, on an hourly time scale is~~ ~~> 0.8~~~~. It is also clear that correlation is generally higher at FINO3, and slightly lower~~

3.1.2 Frequency distribution of wind speed

Figure 4 shows that, on average, the MMs capture the wind speed distributions well compared to the observations. The only exception is a slight shift towards higher wind speeds at Cabauw, corresponding to the positive bias in mean wind speed observed in Fig. 3. The ERA-Interim data set captures the distribution well at Høvsøre~~and Cabauw. At Høvsøre and Cabauw the correlation decrease with height. The correlation of~~ , but it has distributions that are shifted towards lower wind speeds at FINO3 and Cabauw, corresponding to the bias in Fig. 3.

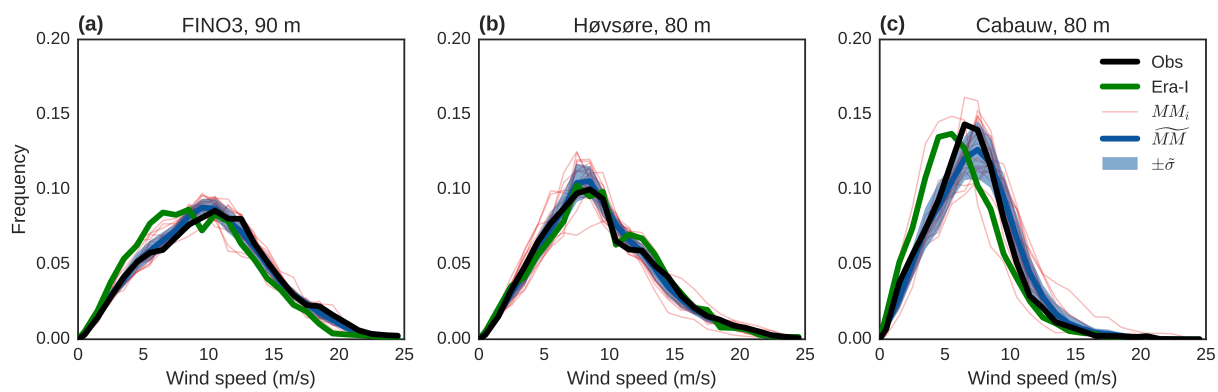


Figure 4. Wind speed distributions at the three sites (FINO3 at 90 m, Høvsøre at 80 m and Cabauw at 80 m), for: the observations (black), the ERA-Interim data set (green), the Mesoscale Models MM_i (red), and the inter-model mean (blue line) and its standard deviation $\widetilde{MM} \pm \tilde{\sigma}$ (blue shade).

3.1.3 Distribution of wind direction

Figure 3 shows that the MMs generally capture the mean wind speed well, this is also true for the wind direction distributions, commonly called "wind roses". The distributions are split into 15° sectors at heights of either 80 or 90 meters. Figure 5 also shows that the models are in good agreement. In all three sites the MMs capture the distribution better than the reanalysis data~~is similar to the mesoscale~~. At all sites, but most markedly at Cabauw, the ERA-Interim distribution is rotated clockwise relative to the distribution from the observations and MMs. This rotation might result in a different wind farm layout if its power is optimized according to the wind roses from MMs or the ERA interim.

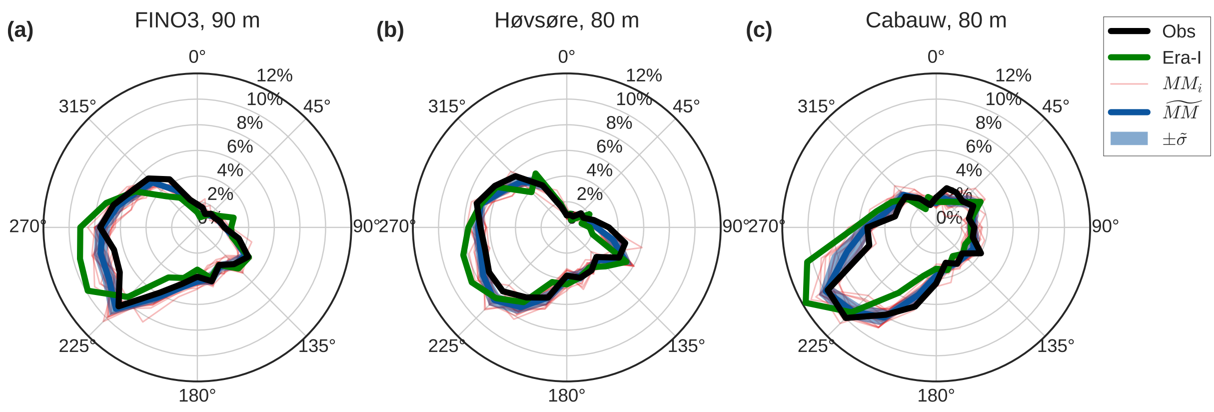


Figure 5. Wind direction distributions at the three sites (FINO3 at 90 m, Høvsøre at 80 m and Cabauw at 80 m), based on 24 sectors, for: the observations (black), the ERA-Interim data set (green), the Mesoscale Models MM_i (red), and the inter-model mean $\widetilde{MM}$ (blue line) and its standard deviation $\pm\tilde{\sigma}$ (blue shade).

3.1.4 Annual wind speed cycle

Figure 6a shows the monthly distribution of the mean wind speed for the MMs, and the measurements. Apart from a few models outside the 3x quartile range, most models capture the diurnal cycle well. Interesting, the figure also reveals that both the overestimation by the models at Cabauw and ~~aloft at~~ the underestimation at FINO3, seen in Fig. 3, is evenly distributed throughout the year. At Høvsøre, ~~but slightly lower at FINO3 and at the lowest levels at Høvsøre. The intermodel variance is $\approx 2.5\%$ everywhere, by at Cabauw, especially at the lowest levels, it is slightly higher ($\approx 3.4\%$).~~ a mix of under- and overestimations are observed.

~~The mean coefficient of variation ($C_{v,u}$)~~

Figure 6b shows the monthly distribution of the Mean Absolute Error (MAE) for wind speed for the MMs. Summer and spring are generally associated with larger deviations between the modeled and observed wind speeds. It is well established that fall and winter weather in Northern Europe is governed by large-scale planetary and synoptic weather phenomena, that is well captured by mesoscale models. During spring and summer, meso- and thermally induced phenomena (e.g. sea breezes and convection) have a larger impact on the flow, which is more difficult for the models to correctly capture. The lowest MAE is observed at FINO3 in February, October and November with most MAE values near 10%. The largest MAE are in November at Cabauw (values in the range 30–45%). For June and July FINO3 shows

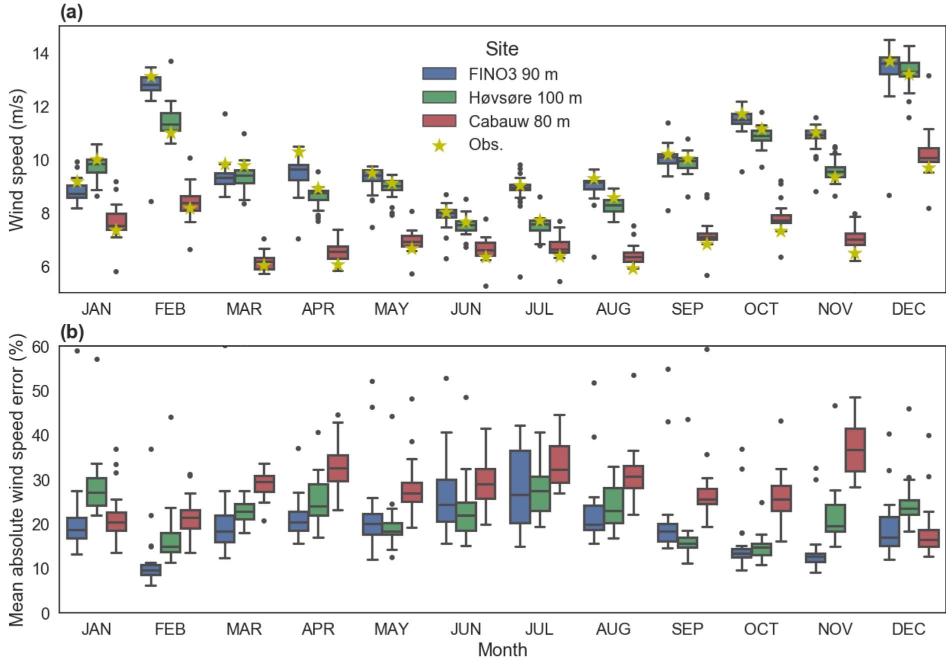


Figure 6. (a) Monthly distributions of mean wind speed for the MMs (boxplots) and observations (star), at each location (colors). (b) Monthly distributions of the models for the Mean Absolute Error (MAE) for wind speed at each location (colors).

3.1.5 Effect of atmospheric stability

It is generally acknowledged that non-neutral atmospheric stability conditions pose one of the greatest challenges for MMs. To study the performance of the models in different stability regimes, the stability parameters supplied for each model (inverse Obukhov length or bulk Richardson number) were used to group the hourly samples into five stability classes, shown in Table A2. Because the models represent atmospheric stability in different ways, the number of samples in each stability group varies for the different models. However, the number of samples in each group was never below 150 hours (out of 8760 hours), and it was more than 400 in most cases. The MAE for wind speed ~~is~~ was calculated for each of groups and for all models. The results are shown in Fig.~~8 for all six sites. It shows that at~~ 7.

At all three sites, the smallest deviations between modelled and measured wind speeds are found when the models perceive the surface layer stability from unstable (U) to stable (S). The MAE in these cases typically range from 10% to 35%, with just a few models outside of 3x quartile range. The largest deviations are found when the models estimate very stable conditions (VS) or very unstable conditions (VU) (typical values in the range 15–45% MAE). The site where the largest errors are found is Cabauw, and the smallest is FINO3. This is in agreement with the results in section 3.1.

3.1.6 Coefficient of variation of wind speed

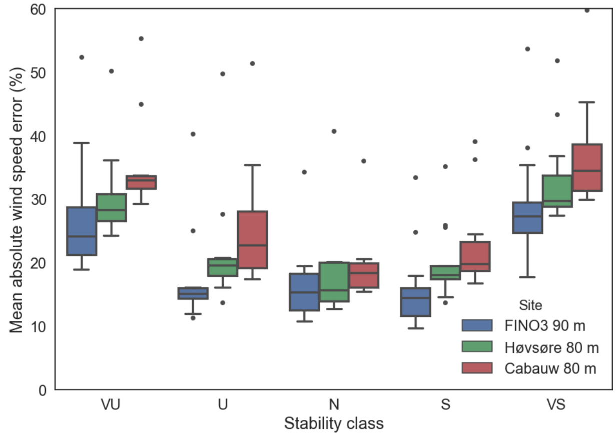


Figure 7. Distribution of Mean Absolute Error (MAE) for wind speed at the three sites for five stability classes: Very Unstable (VU), Unstable (U), Neutral (N), Stable (S), Very Stable (VS). See definitions in Table A2.

Figure 8 shows the mean coefficient of variation ($C_{v,u}$) for wind speed, at the three sites. At FINO3, the average of the ~~mesoscale models~~ MMs $\tilde{C}_{v,u}$ is similar to the ~~the observations, the difference is~~ observations, with a bias of less than 1% at all three heights. Ignoring one ~~outlier, the intermodel variance range between 3.0 – 3.5%~~ "outlier", the inter-model variance ranges between 3.0% and 3.5% at the three heights. The ~~outlier with~~ "outlier", which shows much lower values~~is due to lower~~ , is a consequence of the low variance for that ~~particular model not due to any significant difference in the mean wind speed~~ model compared to the other models. It was removed by the filtering method described in sect.~~2.3 in the calculation~~ 2.3 when calculating the mean of the models ~~mean~~ ($\tilde{C}_{v,u}$) and the inter-model variance ($\tilde{\sigma}_{C_{v,u}}$). The ERA-Interim ~~dataset also captured~~ data set also captures the magnitude of $C_{v,u}$ well~~at FINO3, showing values similar to the observations and the mesoscale models~~.

At Høvsøre, $C_{v,u}$ decreases with height for both the observations ~~,~~ and most of the ~~mesoscale models. The model average~~ MMs. The inter-model mean of the models ($\tilde{C}_{v,u}$) agrees well with the observations, but underestimates it by about ~~2%~~2%. The ERA-Interim ~~dataset~~ data set does not capture ~~the~~ this behavior, and instead ~~show~~ shows an increase with height~~, but at~~. At the highest levels~~it does agree quite well with the model average and observation~~ , however, it reaches the average of the models and the observed values. The spread of the ~~mesoscale models~~ MMs ($\sigma_{C_{v,u}}$) is slightly higher than at FINO3 (~~3.6 – 4.4%),~~ 3.6–4.4%), and is highest at the ~~surface~~lowest levels.

~~The magnitude of~~ At Cabauw, $C_{v,u}$ ~~is largest~~ at 10 m is the largest value found across all sites~~at 10 meters at Cabauw~~. Above 10 ~~meters~~ m a sharp drop-off ~~with height~~ is found up to 80 ~~meters is observed, followed by a small~~ m, where is starts to slowly increase up to 200 ~~meters~~m. Most of the ~~mesoscale models are able to~~ MMs capture this behavior, which is reflected in the mean of the models ($\tilde{C}_{v,u}$). However, the models underestimate the magnitude and the drop-off of $C_{v,u}$ at the lowest levels ~~,~~ with a bias up to 12% at 10 and 20 ~~meters the bias of the average of the mesoscale models is ≈ 12%~~ m. Above 80 ~~meters the mean of the models and observations agree quite well~~m the models agree with the observations. The ERA-Interim ~~dataset do~~

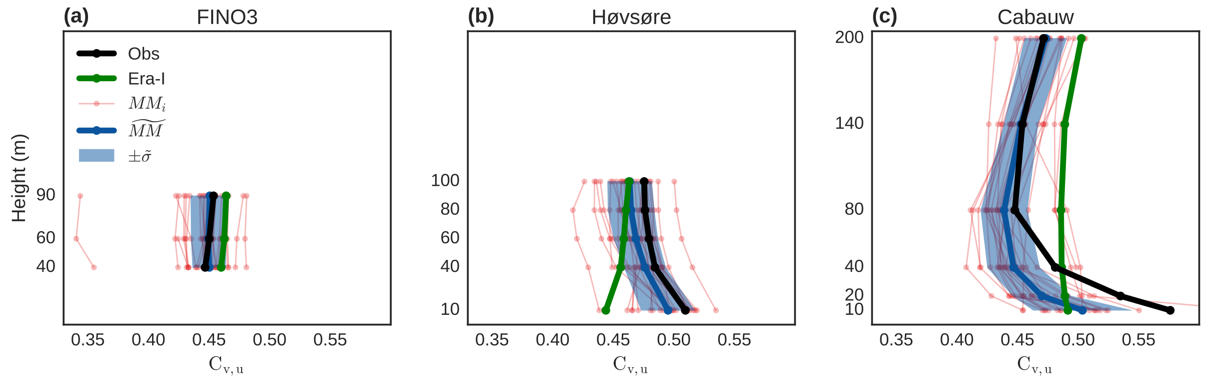


Figure 8. Vertical profiles of the coefficient of variation for wind speed $C_{v,u}$ ~~for all of 2011~~ at the ~~six~~ three sites, for: observations (black), ERA-Interim (green), the mesoscale models MM_i (red), and the mesoscale models mean and ~~intermodel~~ inter-model variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).

~~not show much change~~ data set is nearly constant with height, and ~~tend to overestimate~~ underestimates $C_{v,u}$ ~~above 40 meters, and underestimate it~~ below 40 m, and overestimates it above. The inter-model variance ($\tilde{\sigma}_{C_{v,u}}$) of the ~~mesoscale models~~ MMs is largest at the lowest levels, 8.0% at 10 ~~meters~~m, and gradually decreases to less than ~~4%~~ 4% at 200 ~~meters~~m.

3.1.7 ~~Dataless sites~~

~~The mesoscale models mean wind speed (u_m) at the three Dataless sites (Fig. 3 d, e, and f) reveals that at~~

Effect of upstream conditions on the variation of wind speed

The coastal site Høvsøre and the offshore site ~~Dataless1 the mesoscale models have a larger mean wind speed than the reanalysis, similarly to the~~ FINO3 ~~offshore site.The intermodel variance $\tilde{\sigma}_{u_m}$ is 2.3 − 4.3% with the largest variance found at the lowest levels.~~ is used to investigate whether there is a dependency o the coefficient of variation for wind speed (shown in Fig. 8) on upstream surface conditions. With a nearby coastline aligned north-south, Høvsøre represents the case with anisotropic surface roughness conditions: westerly winds comes from the sea (onshore flow), and easterly winds from land (offshore flow). In contract, the offshore site FINO3 has isotropic upstream surface roughness. To study the differences, the coefficients of variation were binned according to four wind direction sectors, each spanning 90 degrees: north, east, south, and west. The values for the east and west sectors were then extracted and analyzed. Figure 9 shows the profiles of $C_{v,u}$ for the two wind directions at FINO3 and Høvsøre.

~~At the Dataless2 site, offshore the coast in Denmark, the mesoscale models and the reanalysis data agree quite well, but the mean wind speed of the reanalysis data are slightly lower than the average of the mesoscale models. The inter model variance ($\tilde{\sigma}_{u_m}$)at the site is 5.7% at 10 meters, and gradually decrease to 4.0% at 120 meters.~~

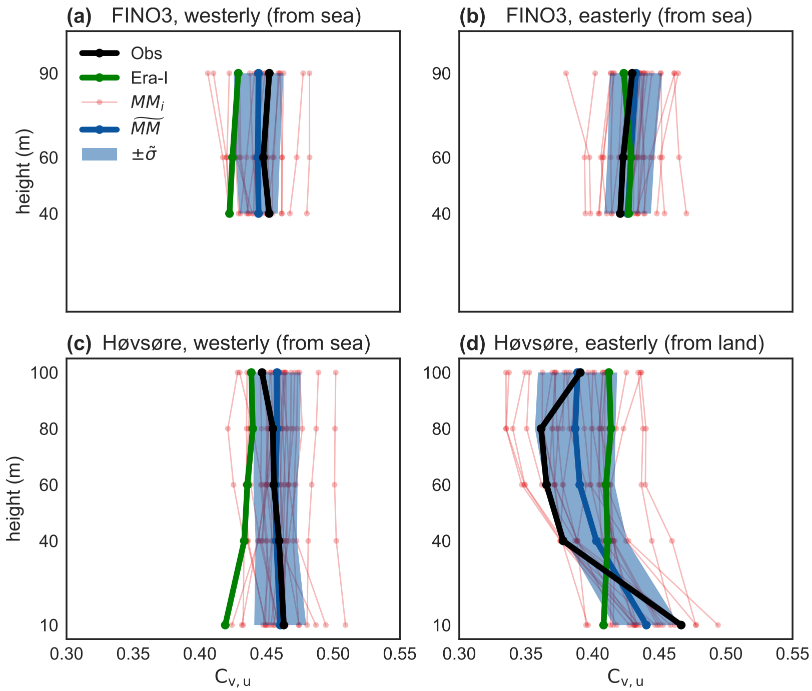


Figure 9. Coefficient of variation for wind speed $C_{v,u}$ for easterly (top) and westerly (bottom) winds at FINO3 (left) and Høvsøre (right), for: the observations (black), the ERA-Interim data set (green), the MMs MM_i (red), and the mesoscale models mean and inter-model variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).

~~At the Dataless3 site the mean wind speed of the reanalysis data is generally lower than the mean wind speed from the mesoscale models. The magnitude of this difference is larger aloft. The site also shows the largest inter-model variance of any of the six sites: $10.1 - 13.6\%$ at 10 - 40 meters, and $6.7 - 6.8\%$ at 80 - 120 meters. One mesoscale model show considerably larger mean wind speed than the rest, but was not included in the calculation of the model mean and variance .~~

~~The coefficient of variation at the three Dataless sites is presented in Fig. 8 (d), (e), and (f). It shows that relative variations are larger at Dataless1 than at~~ At FINO3, ~~but there is a good agreement between the mesoscale models ($\tilde{\sigma}_{C_{v,u}}$ is $< 2.5\%$). The ERA-Interim dataset show a large coefficient of variation compared to most mesoscale models, but not significantly.~~

~~The~~ the coefficient of ~~variation at Dataless2 is slightly lower than for Dataless1 and decrease with height for most mesoscale models, but not for the ERA-Interim dataset. The~~ variance is almost constant with height and slightly lower for easterly winds than for westerly flow. This is true for both models and observations. The sample size for easterly winds is smaller, about half, than for westerly flow. However, both sample sizes are large (N > 1000), so the influence from sample sizes is expected to be small. The average of the MMs captures the observed behavior well for both westerly and easterly winds, and the inter-model variance is ~~comparable to that at Høvsøre ($2.3 - 5.0\%$) and is also slightly higher at the lowest levels~~similar for the two sectors. The ERA-Interim agrees better with the observations during easterly flow at FINO3.

At ~~Dataless3 large values of~~ Høvsøre, the coefficient of variation ~~for wind speed is observed at the lowest levels, similar to Cabauw, and a sharp drop-off~~ is larger for westerly than for easterly winds. Easterly winds show larger coefficients of variation at 10 m than higher up. The reduction of $C_{v,u}$ with height up to 40 ~~meters is also seen. Just like at Cabauw,~~ m for easterly flow is underestimated by most of the mesoscale models, and completely missed by the ERA-Interim ~~do not show the drop-off with height. The mesoscale modelshave a stronger agreement at Dataless3 than at Cabauw ($2.9-4.8\%$), but show the largest srpead at 10 meters, exactly like for Cabauw.~~

~~In general the observed behavior at the Dataless sites for both mean wind speed and coefficient of variation for wind speed is very similar to the behavior observed at each of counterpart-sites with observations, apart from a slightly better agreement between the mesoscale models at low levels at the Dataless2 than at Høvsøre, and less agreement amongst the models at the Dataless3 site compared to Cabauw, but it is probably a reflection of differences in terrain-complexity of the sites.~~

3.2 ~~Distributions~~

~~Figure 3 shows that the mesoscale models generally capture the mean wind speed well. In the following the ability of the models to reproduce distributions of wind direction (Fig. 5), wind speed (Fig. 4), and shear exponent (Fig. 10) is demonstrated. The distributions are presented at heights relevant to wind energy applications, between 80 and 90 meters, and in the case of wind shear exponent, between 40 meters and 80-90 meters. In the following only the sites with observations have been included, because the results at each of the three data-less sites are similar to the results from the corresponding site with observations (offshore, coastal, inland).~~

~~The wind roses (distribution of~~ data set. For westerly winds, the mean of the ~~wind direction), shown in Fig. 5 in 15° sectors at heights of either 80 or 90 meters, reveal that the models capture the observed distributions well, and the intermodel variance is low. In all three cases the mesoscale models have captured~~ models and the ~~distribution better than the reanalysis data does.~~

~~Distributions of wind direction at the six sites based on 24 directions, each representing 15°, for: the observations (black), the ERA-Interim dataset (green), the Mesoscale Models MM_i (red), and the mesocale models mean and intermodel variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).~~observations agree, but is underestimated by ERA-interim.

~~Wind speed distributions (Fig. 4) show that on average the mesoscale models capture the distributions well, apart from a slight shift towards the higher wind speeds at Cabauw, resulting in the positive bias in mean wind speed observed in Fig. 3~~The dependence on height of $C_{v,u}$ is only present at Høvsøre for easterly winds, and points to the influence of upstream surface conditions on the variation. The observed pattern is captured by the MMs, but the models show a more "smooth" vertical transition than do the observations. The ERA-Interim ~~dataset capture the distribution well at Høvsøre, but show a distribution shifted towards lower windspeeds at FINO3 and Cabauw~~does not capture the pattern.

~~Distributions of wind speed at three mast sites, for: the observations (black), the ERA-Interim dataset (green), the Mesoscale Models MM_i (red), and the mesocale models mean and intermodel variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).~~

~~Frequency of occurance of the shear exponent (α) at the three mast sites, for: the observations (black), the ERA-Interim dataset (green), the Mesoscale Models MM_i (red), and the mesocale models mean and intermodel variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).~~

3.1.1 Distribution of wind speed shear exponent

~~The distributions of the~~ Figure 10 shows the distributions of wind speed shear exponent (α) for ~~the three mast sites are presented in Fig. 10~~each of the three sites calculated between 40 and 80 or 40 and 90 m. Under neutral ~~conditions and a uniform surface roughness(for all wind directions)~~ atmospheric stability conditions and isotropic surface roughness, a sharp distribution centered around a single ~~α value should be observed, so for offshore sites like~~ value is expected. This means that for an offshore sites such as FINO3, the spread in shear exponent comes primarily from variations in ~~the atmospheric stability, the Fig. reveals that in that particular case the models capture~~ atmospheric stability~~well on average~~. With this in mind, the distributions show that most MMs capture the stability well at the site. The ERA-Interim ~~dataset however, do not seem to~~ data set does not capture the strongest shear situations ~~.~~ well. This can be easily explained by the low data frequency (6 hours).

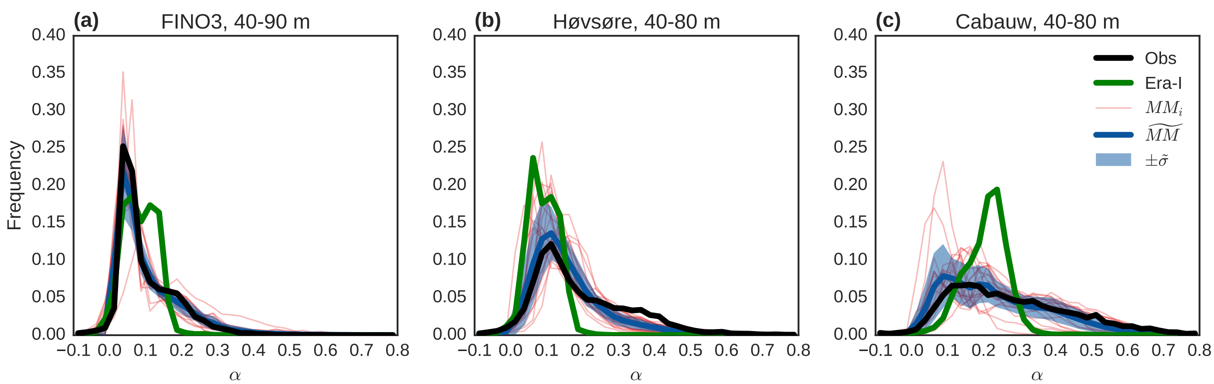


Figure 10. Frequency of occurrence of the shear exponent (α) at the three sites, for: the observations (black), the ERA-Interim data set (green), the Mesoscale Models MM_i (red), and the inter-model mean (blue line) and standard deviation $\widetilde{MM} \pm \tilde{\sigma}$ (blue shade).

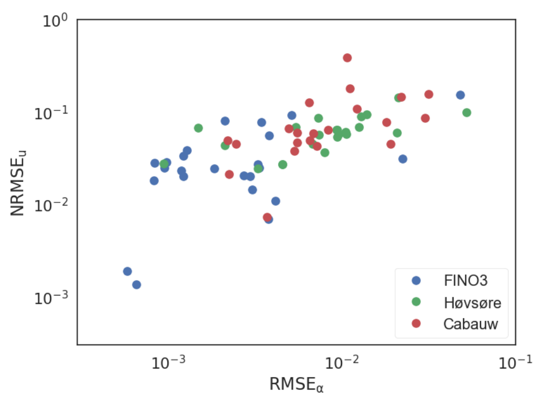
At Høvsøre and Cabauw~~the distribution~~, the distributions of α ~~is a~~ reflect the combined effect of both the non-homogenous upstream surface roughness, ~~which cannot be expected to be uniform (it varies with direction), and the varying~~ and the variations in atmospheric stability. ~~For example, one would expect that at~~ At the coastal site~~Høvsøre the upstream roughness vary a lot~~, the wind speed profile changes depending on whether the ~~wind is coming~~ fetch is from land or ~~sea, something that was also observed by Hahmann et al. (2014). The figure~~ from the sea, which is also reflected in the distribution of α (Hahmann et al., 2014). Figure 10 also shows that while the shear distributions are generally ~~well captured~~also well captured at Høvsøre and Cabauw, a slight shift towards lower ~~shear~~ values is observed at both sites. ~~The~~ This points to an underestimation of the surface roughness, a misrepresentation of the atmospheric stability, or a combination of the two. Just like at FINO3, the ERA-Interim ~~dataset~~ data set does not capture the weak and strong shear cases at Høvsøre ~~, and at Cabauwneither the weak or strong shear cases are sufficiently represented. Part of the reason for the poor results for ERA-Interim is the linear interpolation from model levels to fixed height levels used in the data extraction. However, at least three model levels were used in the interpolation to the two height levels (40 and 80 meters), so even though some dampening~~ and Cabauw.

3.2 Relating performance to model setup

To identify what model setup choices lead to better model performance, the statistics of each model across all heights are reduced to just two values at each site: NRMSE for wind speed (NRMSE$_u$) and RMSE for wind speed shear exponent (RMSE$_\alpha$). The shear exponent was calculated between pairs of nearby levels, e.g. at FINO3 two values were calculated, one between 40 and 70 m, and one between 70 and 90 m. The RMSE$_\alpha$ was then calculated as described in section 4 between modelled and observed values of the shear ~~coefficient is expected from the interpolation method, it does not fully explain the poor results~~exponent across all height-pairs.

3.3 ~~Effect of upstream conditions on variation of wind speed~~

~~In order to investigate whether a dependency on upstream conditions exists for~~ Figure (11) shows NRMSE$_u$ and RMSE$_\alpha$ for all MMs at all three sites. It shows, similarly to section 3.1, that the models generally have smaller mean wind speed and mean shear exponent errors at the offshore site FINO3. But, as previously shown, errors are larger near the surface, and the ~~coefficient of variation for wind speed (shown in Fig. 8) exists, two of the sites were studied. One offshore (~~three levels used at FINO3 ~~), representing the isotropic upstream surfaceroughness case, and one with a nearby coastline going from south to north (Høvsøre), represeting the case with a strong surface roughness dependence on direction. The coefficients of variation was binned according to four wind directions representing 90 degree sectors: north, east, south, and west. The values for the east and west sectors at the two sites was then extracted and analyzed.~~ is at 40 m and above, unlike Høvsøre and Cabuaw where levels below 40 m are included.



~~Coefficient of variation for wind speed $C_{v,u}$ for split into westerly and easterly flow at FINO3 (a,b) and Høvsøre (c,d), for: the observations (black), the ERA-Interim dataset (green), the Mesoscale Models MM_i (red), and the mesocale models mean and intermodel variance $\widetilde{MM} \pm \tilde{\sigma}$ (blue).~~

Figure 11. RMSE for wind speed shear exponent (RMSE$_\alpha$) versus Normalized RMSE for wind speed (NRMSE$_u$) at the three sites.

The models were then grouped according to specific model components. Given the range of setup choices that influence the model performance, large groups were needed to obtain useful statistics. With this in mind, three setup options were chosen for analysis: PBL scheme, grid spacing, and simulation lead-time, and statistics of NRMSE$_u$ and RMSE$_\alpha$ were computed for each group. The choice of groupings was based mainly on two criteria: 1) it was possible to form groups with at least six members in each group. 2) each of the options were highlighted in the literature as being important for model performance (Hahmann et al., 2014; Gómez-Navarro et al., 2015; Carvalho et al., 2012; Draxl et al., 2014). Several other setup options were considered: MM, LSM, land cover, spin-up time, and data set used for initial and boundary conditions, but either it was not possible to group them in a meaningful way, or they were deemed of too little importance based on previous studies. Models missing information about particular setup options, or missing output at some heights, were excluded from this analysis.

~~Figure 9 shows the two cases for FINO3 and Høvsøre. At FINO3 the coefficient is slightly lower for easterly flow for both the models and observations, and no dependence on height is observed. The sample size is smaller for easterly directions, about half that of westerly directions, but both sample size are large (N > 1000) so the difference in $C_{v,u}$ due to differences in sample sizes is expected to be small. For both wind directions the average of the mesoscale models, and ERA-Interim, captures the observed behavior well, and the inter-model variance for the mesoscale models is similar for both directions. At Høvsøre westerly flow comes from the sea, and easterly from land. The figure shows that wind sectors coming from sea at Høvsøre have a higher coefficient of variation compared to the land sector. Easterly winds (coming from land)show larger coefficients of variation at the lowest levels, which decreases with height. This reduction with height is underestimated by most of~~

3.2.1 PBL scheme

The PBL scheme in a MM ensures an accurate representation of thermodynamic and kinematic structures of the lower troposphere (Cohen et al., 2015). To study the influence of the PBL schemes used, the ~~mesoscale models , and completely missed by the ERA-Interim dataset. For westerly winds (coming from the sea) at Høvsøre the models and observations agree almost exactly~~MMs were split into three groups: YSU, MYJ, and Other. The statistics of NRMSE$_u$ and RMSE$_\alpha$ for these groups are shown in Table A5. The YSU group consists of six models that used the YSU PBL scheme (Hong et al., 2006). The models in this group span a range of grid spacings and lead-times, but models with larger than average grid spacing and longer than average lead-times dominate the group. The MYJ group contains six models that used the MYJ PBL scheme (Janjić, 1994), most of them use a short lead-time limit, and a grid spacing that is close to the average for the MMs in this study. The last group labeled 'Other' contains nine models that used a mix of different PBL schemes (see Table A4). These models have a wide representation of different grid spacings and lead-times.

At FINO3, the group consisting of models not using either the YSU or MYJ PBL schemes generally have smaller wind speed errors; even though the group also contains the model with the largest NRMSE$_u$. The models using the MYJ PBL scheme have smaller wind shear exponent errors, and on average also smaller wind speed errors than YSU. But the median model in the YSU and MYJ groups have similar wind speed errors.

~~The westerly wind sector at Høvsøre, where the wind comes from the sea, show a very similar pattern to the two sectors at FINO, that is: no change with height. This points to the fact that the pattern observed for the easterly wind directions at Høvsøre~~

~~, where a clear dependence on height is observed, is due to influence from the upstream surface conditions. These influences do not appear to be well captured by the ERA-Interim dataset~~At Høvsøre, the three groups have very similar mean wind speed error-statistics, with YSU showing only slightly smaller errors. However, for wind shear exponent the models in the YSU group have the smallest errors, both on average and for the median model. Draxl et al. (2014) studied similar error-statistics at Høvsøre for the WRF model run with a number of different PBL schemes during October 2009. They, unlike this study, found that MYJ gave slightly smaller errors than YSU. However, Draxl et al. (2014) used a version of the YSU scheme with a bug that was corrected in WRF version 3.4.1 (Hahmann et al., 2014).

At Cabauw, the YSU group has smaller errors than the other groups for both wind speed and wind shear exponent, but the ~~mesoscale models are , to some degree, able to capture them~~errors for the median model in the YSU and MYJ groups are quite similar. The single most accurate model is found in the 'Other' group, but that group as a whole has larger errors.

3.3 ~~Individual model performance~~

~~The previous sections have been focused on distributions, mean quantities, and the variance of the mesoscale models. In this subsection the attention is on identifying the things that set the models apart, be it due to the chosen parametrization, the dataset used for initial and boundary conditions, the setup~~

3.2.1 Grid spacing

A mesoscale model should be able to explicitly resolve smaller and smaller phenomena as the grid spacing is decreased. Skamarock (2004) illustrated that the effective resolution of the WRF model is approximately seven times the grid spacing used. However, mesoscale models, as the name suggests, have been developed to simulate the 'meso'-scale, they are often not capable of simulating weather at scales that lie between the micro- and mesoscale, i.e. between approximately 100 and 2000 m. To study the importance of the grid ~~, the time integration choice, or otherwise.~~

~~Absolute mean wind speed bias at 40 meters vs. model resolution for three sites with a mast. R^2 is the correlation coefficient, and a is the slope of a least-squares fit to the data points.~~

~~The hypothesis of this study was that, given enough samples (models) in the study, it would be possible to identify the statistical effects of choosing a particular model setup, compared to another.~~

~~The following factors were hypothesised to influence the surface layer winds: The model grid spacing and source of orography and land use data, due to the importance of accurately representing the orography and the upstream surface roughness. The choice of SL and PBL scheme. To a lesser degree: the model, the boundary and initial conditions, the spin-up and simulation time,~~ spacing, the models were ranked by grid spacing, similar to table A4. The models were then split into three groups: Fine, Moderate, and Coarse. The Fine group consists of seven models that all have a grid spacing below 3 km. The Moderate group consists of eight models at exactly 3 km, and the ~~domain placement.~~

~~Chance of mesoscale models to have a mean wind speed error (E_c) larger or smaller than the median error if they have smaller or larger grid spacing than the median (Δ_c) for the three mast sites at 40 meters.~~ Coarse group consists of six models above 3 km. The Fine group contains models that are well distributed in terms of PBL schemes and simulation lead-time. The

Moderate models also has a good representation of different PBL schemes and lead-time limits, but the MYJ PBL scheme and short lead-times are most common. The Coarse group contains no models using the MYJ PBL scheme, and half of the models use a short lead-time.

~~The hypothesized correlations were investigated using a number of approaches, and although some weak indications of correlation between the model grid spacing and~~ Table A6 shows the statistics for NRMSE_u and RMSE_α. At FINO3, the ~~absolute wind speed bias (Fig. ??), no significant correlation were found between the different performance metricstested for~~ Fine group has the smallest wind speed errors. For the wind shear exponent, the ~~surface layer winds and the factors listed above~~smallest error is found in the Coarse group, but, on average, the Fine and Moderate groups have smaller errors. At Høvsøre, the Fine and Moderate groups have similar errors for both wind speed and shear exponent. However, the model with the smallest shear exponent error is found in the Coarse group. At Cabauw, the Moderate group shows the smallest errors for both metrics, followed by the Fine group. But, just as for Høvsøre, the model with the smallest RMSE_α is found in the Coarse group.

~~Figure ?? shows the chance, for a random model , of having a mean wind speed error (E_c) that is larger or smaller than the median error, depending on whether the model grid spacing is larger or smaller than the median model grid spacing (Δ_c), for the three mast sites at 40 meters. At FINO3 and Høvsøre the chance of a lower than median error is the same, whether the model grid spacing is larger or smaller than the median spacing. At Cabauw a higher chance of a smaller mean wind speed error is observed for smaller grid spacing,~~

3.2.2 Simulation time

As the solution in mesoscale models is integrated forward in time, the uncertainties associated with the errors in the initial conditions increase (Yoden, 2007). This can cause the model solution to drift away from the true solution. Furthermore, amplification errors can reduce the variance, which reduces the accuracy of the model in a statistical sense. To study the influence of the simulation time on the model performance, the models were ranked and split into three groups: Short, Medium, and Long. The Short group consists of nine models with a lead-time below 48 hours. Four models in the group use the MYJ scheme, and one the YSU scheme. The Short group has a good representation of models with different grid spacings. The Medium group includes eight models with a lead-time between 48 and ~~a higher chance for larger errors is seen for larger~~ 335 hours. The group has a good representation of different PBL schemes and grid spacing. The Long group consists of seven models with a lead-time limit above 335 hours. Five of the models use the YSU PBL scheme, and most of the models use a larger than average grid spacing.

3.2.3 ~~Performance consistency across sites~~

~~To investigate whether the models that perform well at one site also perform well at the other sites, the mean wind speed error, as well as the wind speed correlation with observations, was compared for site-wise pairs, see Fig. ??. The figure reveals that for wind speed correlation (Fig. ?? b) a clear connection is seen between the correlation at one site and the correlation at another, for all site pairs, resulting in a coefficient of determination R^2 of 0.57 or above for the~~ Table A7 shows the errors statistics for the three ~~site-pairs. For mean wind speed error (Fig. ?? a) the correlation between FINO3 and Høvsøre ($R^2 = 0.63$), and FINO3 and Cabauw($R^2 = 0.65$) is quite high, while it is smaller for Høvsøre and Cabauw ($R^2 = 0.44$). Even though a correlation is observed between the performance at one site, and another, for these two metrics, two things should be kept in mind: 1) It is not always the same models that perform well for mean wind and for wind speed correlation 2) the temporal correlation between the sites is quite high, Mehrens et al. (2016) showed that for sites in the North Sea area the temporal correlation of 10 minute average wind speed measurements is $R^2 \approx 0.5$ at distances up to 500 km, which makes it difficult to quantify and separate the effects from the temporal correlation and from and~~ simulation-time groups. At FINO3, the median model from the Short group has the lowest NRMSE$_u$ and RMSE$_\alpha$, but because one model has large errors, the lowest mean errors are found in the Medium group. The Medium group has smaller errors across all metrics compared to the Long group.

At Høvsøre, the Short and Long groups have similar error statistics for wind speed, and both measures are lower than those for the Medium group. For RMSE$_\alpha$ the median model from the Short group has the smallest error, while, on average, the errors are smallest in the Medium group.

At Cabauw, the smallest errors for both wind speed and shear exponent are, on average, found in the Long group, while the median model with the smallest errors are in the Short group. It is worth noting that five of the seven models in the Long group use the YSU PBL scheme, and in section 3.2.1 the models using the YSU PBL scheme were shown to have smaller errors at Cabauw, so it cannot be ruled out that the small errors in the Long group at Cabuaw is related to the ~~model skill. The sites are between 150 and 550 kilometers from each other, so some temporal correlation should be expected~~over representation of the YSU scheme and not the simulation length.

~~a) Errors of u_m at one site for the models plotted against the error of u_m at another site for the same model. b) Correlation of u at one site plotted against the correlation of u at another site. The coefficient of determination R^2 is shown for the different site-pairs in the legends. In both a) and b) the height is 80 meters for Høvsøre and Cabauw, and 90 meters for FINO3.~~

3.3 Wind energy application

~~The mesocale model timeseries~~ As described in sect. 2.4, the output from the mesoscale models was applied to a simple wind energy ~~resource assessment application using the 90 meter data at the offshore site~~ exercise. The 90-m wind resource of a Horns Rev wind farm was estimated using the output from the various MMs at FINO3~~, as described in section 2.3. In figure 12 the error of the calculated~~ . Figure 12 shows the errors for four metrics: 1) error in mean wind speed u_m, 2) error in mean power density P_m, 3) error in mean power density ~~with an assumed~~ using a single power curve $P_{m,pc}$, and 4) error in the mean power density of ~~80 turbines in~~ a wind farm of 80 turbines $P_{m,wf}$, including wake effects~~$P_{m,wf}$~~.

~~A~~ Figure 12 shows that the majority of the models ~~shown in Fig. 12~~ have less than ±5% error ~~for~~ in mean wind speed. The errors are mostly ~~due to an underestimation of the mean wind speed, andfor some of them severely (< 10%~~under-estimations, and, in a few cases, severe underestimation of more than 10% (outside the scale of the Figure). For the mean power density, the spread of the models ~~are~~is, as expected, much larger due to the ~~third power~~ "third power" dependence on the wind speed. However, when the power density is calculated using a turbine power-curve, where the highest wind speeds (~~> 14 m/s~~>14 m s^{-1}) are less important~~for the power than it is for power density, the~~ , the inter-model variance is comparable ~~to the variance~~ that

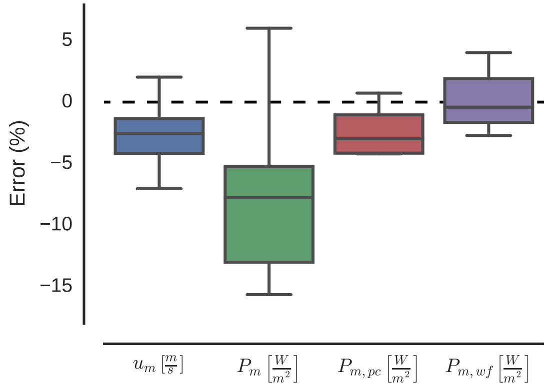


Figure 12. Distributions of errors from the ~~models~~ model's output at 90 ~~meters~~ m at FINO3 for the following ~~statistical quantaties~~errors: 1) the mean wind speed u_m (blue), 2) the power density P_m (green), 3) the power density with an implied power curve $P_{m,pc}$ (red), and 4) ~~average~~ the averaged power density of a wind farm including the same implied power curve as 3) and the wake effects (purple). Outliers are not shown~~, but~~ ; the most extreme ones ~~go to −60%~~ are -25% for ~~P_m~~u_m, ~~−37%~~ -60% for ~~$P_{m,wf}$~~P_m, ~~−35%~~ -37% for ~~$P_{m,pc}$~~$P_{m,wf}$, and ~~−25%~~ -35% for ~~u_m.~~ $P_{m,pc}$

for mean wind speed. For the wind farm ~~situation~~case, where the power density ~~is dependent~~ depends on the wind direction distribution, ~~and the losses due to wake effects~~because of the wake losses, the variance is comparable in size to that of the mean wind speed and $P_{m,pc}$, and most models have errors smaller than ~~±2%~~$\pm 2\%$. The improvement seen for $P_{m,wf}$ is caused by ~~a underestimation~~ the under estimation of the wake effects ~~for~~ by most models, leading to a relative increase in mean power density, ~~offsetting~~ off-setting the underprediction from the modelled wind speed distribution. However, the relative effect of over- or underprediction the wake effects may just as well enhance the total power density errors, given slightly different wind direction distributions.

4 ~~Discussion~~

~~The increase of both model errors and inter-model variance, observed with decreasing height at the inland sites, indicates an influence of misrepresentation of surface characteristics, such as surface roughness and orography of the models, and it highlights the need for downscaling of the mesoscale results, to include high resolution information and microscale effects, especially if the models are used for siting and resource assessment.~~

4 Summary and conclusions

~~While the current study offered great insight, especially into the inter-model variance of mesoscale models, several things could be improved in future studies, including: 1)The model representation of surface roughness for the nine grid cells closest~~

~~to the site of interest, was submitted by less than half of the participants. In future studies having these details for all models may help detect misrepresentation of~~ The mesoscale models in this study are able to reproduce well the observed mean wind speed profiles, and the distributions of wind speed. At FINO3 and above 10 meters at Høvsøre, the average of the models has a bias of 3% or less. The largest mean wind speed biases (7–9%) are found at the lowest levels at Høvsøre and Cabauw. Similarly, the MMs were able to reproduce the relative variations of wind speed well in most cases (Fig. 8), but underestimated the relative variations at the lowest levels at Cabauw. A simple analysis of the impact of upstream surface roughness conditions on the relative wind speed variations, suggested that the models may be misrepresenting the surface characteristics ~~. 2)A larger sample size would improve the robustness of the statistics, and allow formore grouping of the model output, which may allow for a formulation of best-practice principles for NWP modeling for wind energy. 3) This study showed that for offshore sites the model errors and inter-model variance is quite low, so future studies would probably benefit most from focusing on inland sites of~~ (Fig. 9), which could be a misrepresentation of either the landuse classification, the conversion of landuse classes into surface roughness lengths, or in the PBL scheme. This problem highlights the need for: 1) further analysis of the representativeness of the surface characteristics in mesoscale models, and 2) downscaling the mesoscale results using a coupled microscale model to capture subgrid-scale influence from variations in orography and surface roughness. The modeled distributions of the wind direction showed only minor differences compared to the observed ones.

For future benchmarking exercises, our study shows that the focus should be on the model representation of surface characteristics, such as orography and landuse, and their associated surface roughness. An attempt was made here to include these details, but because only a subset of the participants supplied this information, it was not feasible. Further studies could also benefit from including more land masts with low to moderate complexity~~where effects from the surface characterisation in the model can be studied in greater detail. 4) Future studies could also focus on specific cases, or phenomena, that is importantfor wind energy, for example Low Level Jets (LLJ). Several studies, e.g. **?**, have shown that the PBL schemes used used in most mesoscale models, including YSU, MYJ, and MYNN have difficulties in accurately capturing surface layer winds in stable conditions, including LLJ's. Capturing LLJ's can be of vital importance because they are associated with strong winds, strong shear, and high turbulence levels, which have the potential to either increase power output , or damage the turbine, depending on the strength and location of the jet.~~

, where capturing the surface characteristics is important, but still manageable by mesoscale models.

~~In this paper an intercomparison of results from 25 different NWP models has been presented for six locations characterized by simple terrain in Northern Europe~~The impact of choosing specific model sub-components was studied in some detail. To allow this, the output from the models was reduced to two metrics at each site, one related to the wind speed bias (NRMSE for wind speed), and one related to the shape of the wind speed profile (RMSE for wind speed shear exponent). The models were ~~compared with each other and, at three of the sites, with observations from nearby meteorological masts for the year 2011. The results, including model meta-data, was submitted by modeling groups from the wind energy sector, and the model setups represent the practice used within the industry today.~~

~~The study showed that on average the mean wind speed as estimated by the mesoscale models had low biases ($< 3\%$) at the offshore site FINO3 and above 10 meters at the coastal site of Høvsøre , meanwhile the largest biases ($7-9\%$) was observed~~

~~at the lowest levels at Høvsøre (10 meters) and Cabauw (10 and 20 meters). A similar pattern existed for the model spread, which was greater at the lowest levels and smaller aloft, and largest at inland and coastal sites, and smaller offshore. The same pattern was also present for correlation between modeled and measured wind speed at the three mast sites: weaker correlation inland and closer to the surface.~~

~~The coefficient of variation (σ/u) was used assess how well the models are able to capture the relative wind speed variation. It revealed that the average of the models have biases of less than 3% at FINO3, Høvsøre, and above 40 meters at Cabuaw, but at 10-40 meters at Cabauw the biases was $7-13\%$ due to a large coefficient of wind speed variation at the lowest levels, that decrease with height, that is not well captured by the models.~~

~~The study also showed that for the distribution of wind direction (wind rose) only small deviations between the mesoscale models and the observations are seen. For wind speed, the models also represent the distributions accurately, apart from a slight shift towards the high wind speeds at Cabauw. For distributions of the shear exponent (α), which reflects the ability of the models to accurate estimate the combined effects of surface roughness and atmospheric stability, the models do well at FINO3, but seem unable to capture all cases of strong shear at Høvsøreand Cabauw.~~

~~A detailed study of~~ then separated into large groups according to their model setup for three setup choices: PBL scheme, grid spacing, and simulation lead-time. At FINO3~~and Høvsøre showed that for wind directions coming from sea the coefficient of variation for wind speed is constant with height, and the models capture the magnitude and behavior well, while for wind directions coming from land at Høvsøre the coefficient is dependent on height. On land the observed coefficient of variation was generally lower, and the models were less accurate, compared to wind directions coming from sea.~~

~~Strong evidence was found, for some metrics (mean wind speed, correlation), that the model performance was consistent between site-pairs for the mast sites.~~, the grouping revealed that the models using the MYJ PBL scheme had smaller wind speed and shear exponent errors than those that use the YSU scheme. At Høvsøre and Cabauw, the opposite was true. However, ~~since the mast sites are located within a distance where some spatial correlation is expected, the consistency cannot be exclusively contributed to the model skill~~the differences between the two groups were not significant and the median model from the two groups had similar errors. Grouping the models according to grid spacing showed that the models with 3 kilometer grid spacing or smaller had lower errors than the group with the largest grid spacings. No conclusive evidence was found that reducing the grid spacing below 3 kilometers results in smaller errors. For simulation lead-time, the median model from the group with short lead-times had the smallest errors at all sites, with the exception of the shear exponent error at Høvsøre. However, no significant difference between the mean of the groups were found, which suggests that the PBL scheme and grid spacing may be of greater importance for the performance at these sites. Future studies should include many more runs to provide more robust statistics, which can provide a basis for "best-practice" guidelines for wind energy applications using NWP models.

~~A wind energy case study was made~~ Last, we used the observed and modelled time-series for a classical wind energy application, the estimation of power production at a hypothetical wind farm at FINO3~~, where the observations and the output from the models output was used to estimate the power output for one~~. The power production, including wake losses, was estimated for both a single turbine and for a ~~whole wind farmthat included wake effects. The study~~ wind farm, using a standard power curve. The exercise showed that while a large spread exists between the modeled power density, it is reduced when the power is calculated ~~via~~ using a power-curve. It also showed the importance of accurately estimating the wind ~~speed~~ direction distribution, since ~~the~~ a small deviation in the distributions ~~changed the power distribution strongly.~~

~~While it was a key objective of this study to determine the model setup choices that have a large impact on the models ability to estimate the wind climate accurately in the lowest part of the PBL, only weak indications were found~~might induce large changes in the power production, because of its sensitivity to the wind farm layout.

5 Data availability

The output data from the mesoscale models have been submitted ~~submitted~~ to the European Wind Energy Association (EAWE) for the mesoscale benchmarking study under an agreement that ensures that individual participants are anonymized in the reported results, and that the model output was not publicly shared. The measurements from the meteorological masts FINO3, Høsøre, and Cabauw ~~,~~ are provided by the data owners under an agreement of not sharing the data with any ~~3rd~~ third party.

Competing interests. The authors declare that they have no conflict of interests

Acknowledgements. We would like to thank the three anonymous reviewers for constructive criticism. Their feedback elevated the level of the paper. Funding from the EU and the Danish Energy Agency through the project EUDP 14-II, ERA-NET Plus - "New European Wind Atlas" is greatly appreciated. The authors would also like to thank the European Wind Energy Association (EWEA) for organizing this mesoscale benchmarking study, the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU), and the Project Management Jülich (PTJ) for sharing the FINO3 mast data, and the Cabauw Experimental Site for Atmospheric Research (CESAR) for making the measurements from the Cabauw mast freely available online (www.cesar-labaratory.nl). Furthermore, we would like to thank the Test and Measurement section of DTU for providing the Høvsøre mast data~~, and finally~~ . Finally, we would like to thank ~~the participants for providing their model data for the intercomparison.~~ all the modeling groups that submitted output for this intercomparison. This study would not be possible without their contributions.

References

Al-Yahyai, S., Charabi, Y., Al-Badi, A., and Gastli, A.: Nested ensemble NWP approach for wind energy assessment, Renewable Energy, 37, 150–160, doi:10.1016/j.renene.2011.06.014, http://dx.doi.org/10.1016/j.renene.2011.06.014, 2012.

Arino, O., Bicheron, P., Achard, F., and Latham, J.: The most detailed portrait of Earth, Eur. Space Agency, http://www.esa.int/esapub/bulletin/bulletin136/bul136d_arino.pdf, 2008.

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, Journal of Applied Meteorology and Climatology, 53, 1901–1919, doi:10.1175/JAMC-D-13-0147.1, http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-13-0147.1, 2014.

Bechmann, A., Sørensen, N. N., Berg, J., Mann, J., and Réthoré, P. E.: The Bolund Experiment, Part II: Blind Comparison of Microscale Flow Models, Boundary-Layer Meteorology, 141, 245–271, doi:10.1007/s10546-011-9637-x, 2011.

Berg, J., Mann, J., Bechmann, A., Courtney, M. S., and Jørgensen, H. E.: The Bolund Experiment, Part I: Flow Over a Steep, Three-Dimensional Hill, Boundary-Layer Meteorology, 141, 219–243, doi:10.1007/s10546-011-9636-y, 2011.

Bowler, N. E., Pierce, C. E., and Seed, A. W.: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP, Quarterly Journal of the Royal Meteorological Society, 132, 2127–2155, doi:10.1256/qj.04.100, http://doi.wiley.com/10.1256/qj.04.100, 2006.

Büttner, G., Feranec, J., and Jaffrain, G.: The CORINE land cover 2000 project, EARSeL {. . . }, http://www.eproceedings.org/static/vol03_3/03_3_buttner2.pdfhttp://www.eproceedings.org/static/vol03%7B_%7D3/03%7B_%7D3%7B_%7Dbuttner2.pdf, 2004.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., and Santos, C.: A sensitivity study of the WRF model in wind simulation for an area of high wind energy, Environmental Modelling & Software, 33, 23–34, doi:10.1016/j.envsoft.2012.01.019, http://linkinghub.elsevier.com/retrieve/pii/S1364815212000382, 2012.

Carvalho, D., Rocha, a., Gómez-Gesteira, M., and Silva Santos, C.: WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for Portugal, Applied Energy, 117, 116–126, doi:10.1016/j.apenergy.2013.12.001, http://dx.doi.org/10.1016/j.apenergy.2013.12.001, 2014a.

Carvalho, D., Rocha, a., Gómez-Gesteira, M., and Silva Santos, C.: Sensitivity of the WRF model wind simulation and wind energy production estimates to planetary boundary layer parameterizations for onshore and offshore areas in the Iberian Peninsula, Applied Energy, 135, 234–246, doi:10.1016/j.apenergy.2014.08.082, http://linkinghub.elsevier.com/retrieve/pii/S0306261914008939, 2014b.

Champeaux, J. L., Masson, V., and Chauvin, F.: ECOCLIMAP: a global database of land surface parameters at I km resolution, Meteorological Applications, 12, 29–32, doi:10.1017/S1350482705001519, http://doi.wiley.com/10.1017/S1350482705001519, 2005.

Cohen, A. E., Cavallo, S. M., Coniglio, M. C., Brooks, H. E., Cohen, A. E., Cavallo, S. M., Coniglio, M. C., and Brooks, H. E.: A Review of Planetary Boundary Layer Parameterization Schemes and Their Sensitivity in Simulating Southeastern U.S. Cold Season Severe Weather Environments, Weather and Forecasting, 30, 591–612, doi:10.1175/WAF-D-14-00105.1, http://journals.ametsoc.org/doi/10.1175/WAF-D-14-00105.1, 2015.

Constantinescu, E. M., Zavala, V. M., Rocklin, M., Lee, S., and Anitescu, M.: A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation, IEEE Transactions on Power Systems, 26, 431–441, doi:10.1109/TPWRS.2010.2048133, 2011.

Courtier, P., Thepaut, J. N., and Hollingsworth, a.: A Strategy for Operational Implementation of 4d-Var, Using an Incremental Approach, Quarterly Journal of the Royal Meteorological Society, 120, 1367–1387, doi:10.1002/qj.49712051912, 1994.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Climate Dynamics, 15, 183–203, doi:10.1007/s003820050276, 1999.

Cuxart, J.: A turbulence scheme allowing for mesoscale and large-eddy simulations, Quarterly Journal of the Royal Meteorological Society, http://onlinelibrary.wiley.com/doi/10.1002/qj.49712656202/full, 2000.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Draxl, C., Hahmann, A. N., Peña, A., and Giebel, G.: Evaluating winds and vertical wind shear from Weather Research and Forecasting model forecasts using seven planetary boundary layer schemes, Wind Energy, 17, 39–55, doi:10.1002/we, http://onlinelibrary.wiley.com/doi/10.1002/we.1608/full, 2014.

Fabre, S., Stickland, M., Scanlon, T., and Oldroyd, A.: Measurement and simulation of the flow field around the FINO 3 triangular lattice meteorological mast, Journal of Wind, http://www.sciencedirect.com/science/article/pii/S0167610514000804, 2014.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sensing of Environment, 114, 168–182, doi:10.1016/j.rse.2009.08.016, http://dx.doi.org/10.1016/j.rse.2009.08.016, 2010.

Garbarino, R., Struzeski, T., and Casadevall, T.: US Geological Survey, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.404.5834, 2002.

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, Quarterly Journal of the Royal Meteorological Society, 139, 501–514, doi:10.1002/qj.1976, 2013.

Gebhardt, C., Theis, S., Paulat, M., and Ben Bouallègue, Z.: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries, Atmospheric Research, 100, 168–177, doi:10.1016/j.atmosres.2010.12.008, 2011.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: Evaluation of wind storms over complex terrain, Geoscientific Model Development, 8, 3349–3363, doi:10.5194/gmd-8-3349-2015, 2015.

Grell, G., Dudhia, J., and Stauffer, D.: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5), https://www.researchgate.net/profile/Georg_Grell/publication/230642695_A_description_of_the_fifth-generation_Penn_StateNCAR_Mesoscale_Model_MM5/links/00b7d51b624259ae9f000000.pdf, 1994.

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, International Journal of Climatology, doi:10.1002/joc.4217, 2014.

Hahmann, A. N., Lennard, C., Badger, J., Vincent, C. L., Kelly, M. C., Volker, P. J. H., Refslund, J., Lennard, C., Badger, J., Vincent, C. L., Kelly, M. C., Volker, P. J. H., and Refslund, J.: Mesoscale modeling for the Wind Atlas of South Africa (WASA) project, DTU Wind Energy, No. 0050, 80 pp, doi:10.13140/RG.2.1.3735.6887, 2015.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, Monthly Weather Review, 134, 2318–2341, doi:10.1175/MWR3199.1, http://journals.ametsoc.org/doi/abs/10.1175/MWR3199.1?prevSearch=&searchHistoryKey=, 2006.

Horvath, K., Koracin, D., Vellore, R., Jiang, J., and Belu, R.: Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models, Journal of Geophysical Research Atmospheres, 117, 1–19, doi:10.1029/2012JD017432, 2012.

Jackson, P. S. and Hunt, J. C. R.: Turbulent wind flow over a low hill, Quarterly Journal of the Royal Meteorological Society, 101, 929–955, doi:10.1002/qj.49710143015, http://doi.wiley.com/10.1002/qj.49710143015%5Cnhttp://www.cpom.org/people/jcrh/QJRMS-101.pdf, 1975.

Janjić, Z. I.: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes, Monthly Weather Review, 122, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1994)122%3C0927:TSMECM%3E2.0.CO;2?prevSearch=&searchHistoryKey=, 1994.

Janjić, Z. I.: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model, NCEP office note, http://www.lib.ncep.noaa.gov/ncepofficenotes/files/on437.pdf, 2002.

Jiménez, P. A. and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind in the wrf model, Journal of Applied Meteorology and Climatology, 51, 300–316, doi:10.1175/JAMC-D-11-084.1, 2012.

Jiménez, P. A., de Arellano, J. V. G., Dudhia, J., and Bosveld, F. C.: Role of synoptic- and meso-scales on the evolution of the boundary-layer wind profile over a coastal region: the near-coast diurnal acceleration, Meteorology and Atmospheric Physics, 128, 39–56, doi:10.1007/s00703-015-0400-6, 2016.

Kallberg, P.: The HIRLAM level 1 system, Documentation manual, https://scholar.google.dk/scholar?hl=en&q=the+hirlam+level+1+system&btnG=&as_sdt=1%252C5&as_sdtp=#0, 1989.

Kallos, G., Nickovic, S., and Papadopoulos, A.: The regional weather forecasting system SKIRON: An overview, Proceedings of the symposium on regional weather prediction on parallel computer environments, https://scholar.google.dk/scholar?hl=en&q=The+Regional+weather+forecasting+system+SKIRON&btnG=&as_sdt=1%252C5&as_sdtp=#0, 1997.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP-DOE AMIP-II reanalysis (R-2), Bulletin of the American Meteorological Society, 83, 1631–1643+1559, doi:10.1175/BAMS-83-11-1631, 2002.

Lean, H. and Clark, P.: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom, Monthly Weather Review, http://journals.ametsoc.org/doi/abs/10.1175/2008MWR2332.1, 2008.

Lock, a. P., Brown, a. R., Bush, M. R., Martin, G. M., and Smith, R. N. B.: A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests, Monthly Weather Review, 128, 3187–3199, doi:10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2000)128%7B%25%7D3C3187:ANBLMS%7B%25%7D3E2.0.CO;2, 2000.

Loveland, T. R. and Belward, a. S.: The IGBP-DIS global 1km land cover data set, DISCover: First results, International Journal of Remote Sensing, 18, 3289–3295, doi:10.1080/014311697217099, 1997.

Mehrens, A. R., Hahmann, A. N., Larsén, X. G., and von Bremen, L.: Correlation and coherence of mesoscale wind speeds over the sea, Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.2900, http://doi.wiley.com/10.1002/qj.2900, 2016.

Moigne, P. L. and Boone, A.: SURFEX scientific documentation, http://geosci-model-dev.net/6/929/2013/gmd-6-929-2013-supplement.pdf, 2009.

Mortensen, N. G., Nielsen, M., and Jørgensen, H. E.: Comparison of Resource and Energy Yield Assessment Procedures 2011-2015 : What have we learned and what needs to be done ? 1 Introduction 2 Case study wind farms, pp. 1–10, 2015.

Nakanishi, M. and Niino, H.: An Improved Mellor-Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, Boundary-Layer Meteorology, 119, 397–407, doi:10.1007/s10546-005-9030-8, http://www.springerlink.com/content/j04441r721280776/, 2006.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, D12 109, doi:10.1029/2010JD015139, http://doi.wiley.com/10.1029/2010JD015139, 2011.

Noilhan, J. and Mahfouf, J. F.: The ISBA land surface parameterisation scheme, Global and Planetary Change, 13, 145–159, doi:10.1016/0921-8181(95)00043-7, 1996.

Orlanski, I.: A rational subdivision of scales for atmospheric processes, Bulletin of the American Meteorological Society, http://www.citeulike.org/group/17501/article/12086670, 1975.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., Déqué, M., Díez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J. F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J. M., and Thomson, M. C.: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), Bulletin of the American Meteorological Society, 85, 853–872, doi:10.1175/BAMS-85-6-853, 2004.

Pan, H. L. and Mahrt, L.: Interaction between soil hydrology and boundary-layer development, Boundary-Layer Meteorology, 38, 185–202, doi:10.1007/BF00121563, 1987.

Peña, A., Floors, R., and Gryning, S. E.: The Høvsøre Tall Wind-Profile Experiment: A Description of Wind Profile Observations in the Atmospheric Boundary Layer, Boundary-Layer Meteorology, 150, 69–89, doi:10.1007/s10546-013-9856-4, 2014.

Pielke, R., Cotton, W., and Walko, R.: A comprehensive meteorological modeling system—RAMS, Meteorology and Atmospheric Physics, http://link.springer.com/article/10.1007/BF01025401, 1992.

Pleim, J.: A Simple, Efficient Solution of Flux–Profile Relationships in the Atmospheric Surface Layer, Journal of Applied Meteorology and Climatology, 45, 341–347, doi:10.1175/JAM2339.1, http://ams.allenpress.com/perlserv/?request=get-abstract%7B&%7Ddoi=10.1175/JAM2339.1%7B%25%7D5Cnpapers2://publication/doi/10.1175/JAM2339.1http://ams.allenpress.com/perlserv/?request=get-abstract&doi=10.1175/JAM2339.1%5Cnpapers2://publication/doi/10.1175, 2006.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing, Journal of Applied Meteorology and Climatology, 46, 1383–1395, doi:10.1175/JAM2539.1, 2007a.

Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part II: Application and Evaluation in a Mesoscale Meteorological Model, Journal of Applied Meteorology and Climatology, 46, 1396–1409, doi:10.1175/JAM2534.1, http://journals.ametsoc.org/doi/abs/10.1175/JAM2534.1, 2007b.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, Journal of Climate, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C. Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Schicker, I., Arnold Arias, D., and Seibert, P.: Influences of updated land-use datasets on WRF simulations for two Austrian regions, Meteorology and Atmospheric Physics, 128, 279–301, doi:10.1007/s00703-015-0416-y, http://link.springer.com/10.1007/s00703-015-0416-y, 2016.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, Monthly Weather Review, 139, 976–991, doi:10.1175/2010MWR3425.1, http://journals.ametsoc.org/doi/abs/10.1175/2010MWR3425.1, 2011.

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-Interim: New ECMWF reanalysis products from 1989 onwards, ECMWF newsletter, 110(110), 25–35., 2007.

Skamarock, W., Klemp, J., Dudhi, J., Gill, D., Barker, D., Duda, M., Huang, X.-Y., Wang, W., and Powers, J.: A Description of the Advanced Research WRF Version 3, doi:10.5065/D6DZ069T, 2008a.

Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Monthly Weather Review, 132, 3019–3032, doi:10.1175/MWR2830.1, 2004.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., Powers, J. G., Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G Duda, X.-Y. Huang, W. Wang, and Powers, J. G.: A Description of the Advanced Research WRF Version 3., Tech. rep., National Center for Atmospheric Research, Boulder, CO, USA, 2008b.

Sukoriansky, S., Galperin, B., and Perov, V.: Application of a New Spectral Theory of Stably Stratified Turbulence to the Atmospheric Boundary Layer over Sea Ice, Boundary-Layer Meteorology, 117, 231–257, doi:10.1007/s10546-004-6848-4, http://link.springer.com/10.1007/s10546-004-6848-4, 2005.

Ulden, A. V. and Wieringa, J.: Atmospheric boundary layer research at Cabauw, Boundary-Layer Meteorology 25th Anniversary, http://link.springer.com/chapter/10.1007/978-94-017-0944-6_3, 1996.

Vincent, C. L. and Hahmann, A. N.: The impact of grid and spectral nudging on the variance of the near-surface wind speed, Journal of Applied Meteorology and Climatology, 54, 1021–1038, doi:10.1175/JAMC-D-14-0047.1, 2015.

Walko, R. and Tremback, C.: ATMET Technical Note 1, Modifications for the Transition from LEAF-2 to LEAF-3, ATMET, LLC, Boulder, Colorado 80308-2195, https://scholar.google.dk/scholar?hl=en&q=ATMET+Technical+Note+1%252C+Modifications+for+the+Transition+from+LEAF-2+to+LEAF-3%252C+AT&btnG=&as_sdt=1%252C5&as_sdtp=#0, 2005.

Warner, T. T.: Numerical Weather and Climate Prediction, 2004.

Yoden, S.: Atmospheric Predictability TL - 85B, Journal of the Meteorological Society of Japan. Ser. II, 85B VN -, 77–102, doi:10.2151/jmsj.85B.77, 2007.

Table A1. Site description, including latitude and longitude coordinates, classification of the site, and the height of the mast z_s as well as the location terrain elevation relative to sea-level z_{asl} and prevailing wind direction.

Nr.	Name	Latitude
1	FINO3	55.195
2	Høvsøre	56.441
3	~~Cabuaw~~ Cabauw	51.970

~~4 Dataless1 57.673 9.024 Offshore - 0 WSW 5 Dataless2 56.011 10.354 Coastal - 0 SE 6 Dataless3 52.830 10.000 Inland - 77 SE~~ height

Table A2. ~~Intercomparison heights for each site~~ Ranges of inverse Obukov length (~~marked by the x~~1/L) and bulk Richardson number (Ri$_b$) used in the stability classification.

~~Height [m] FINO3~~ Stability class	~~Høvsøre~~ Class name	~~Cabauw~~ 1/L interval [m^{-1}]	~~Datal~~
~~200 x 140 x 120 x 100 x x 90 x 80 x x x 60 x x 40 x x x x 20 x x 10 x~~ VU	~~x~~ Very unstable	~~x~~ $1/L < -0.005$	~~x~~
~~Variables submitted by contributers. Variable name Symbol~~ U	~~Units Bulk Richardson number~~ Unstable	~~R_i~~ $-0.005 \le 1/L < -0.002$	~~-~~ $-0.2 \le$
~~(Obukov length)$^{-1}$ $1/L$~~ N	~~m^{-1} Surface temperature~~ Neutral	~~T_s~~ $-0.002 \le 1/L < 0.002$	~~K~~ -0.05
~~Air temperature T~~ S	~~K Wind speed~~ Stable	~~u~~ $0.002 \le 1/L < 0.005$	~~m/s~~ 0.05
~~Wind direction~~ VS	~~° Specific humidity~~ Very stable	~~Q~~ $0.005 \le 1/L$	~~kg/kg~~ 0

Table A3. Participants in the study in alphabetical order.

Participant	Institution	Country
3E	Company	Belgium
Anemos GmbH	Company	Germany
ATM Pro	Company	Belgium
CENER	Research Center	Spain
CIEMAT	Research Center	Spain
DEWI	Company	Germany
DTU Wind Energy	University	Denmark
DX Wind Technologies	Company	China
EMD International	Company	Denmark
ISAC-CNR	Research Center	Italy
KNMI	Meteorological institute	The Netherlands
Met Office	Meteorological institute	United Kingdom
RES ltd.	Company	United Kingdom
Statiol ASA	Company	Norway
University of Oldenburg	University	Germany
Vestas	Company	Denmark
Vortex	Company	Spain

Table A4. Setup description of the 25 model setups ranked by horizontal grid spacing of the finest grid. The columns are: the model name and version (Model), the PBL scheme (PBL), the land surface model (LSM), whether nesting was used (Nest.), the horizontal grid spacing (Δ), the land cover source, Simulation and ~~spinup~~ spin-up time (Sim. time), and initial and boundary condition data (B.C.).

Nr.	Model	PBL	LSM	Nest.	Δ [km]	Landcover	Sim. time [h]	B.C.
1	WRF V3.6.1[a]	Custom	-	yes	1	CORINE[b]	48-24	Era-I[c]
2	MAESTRO V15.01	-	-	no	1	CORINE	-	Era-I
3	WRF V3.6.1	MYJ[d]	Noah[e]	yes	2	USGS[f]	78-6	Era-I
4	WRF V3.3.1	MYJ	-	yes	2	GlobCover[g]	11064-24	Era-I
5	WRF V3.5.1	YSU[h]	Noah	yes	2	CORINE	30-6	Era-I
6	WRF V3.5.1	YSU	Noah	yes	2	-	264-24	Era-I
7	HARMONIE V37h1.1[i]	SURFEX[j]	ISBA[k]	yes	~~1.5~~ 2.5	ECOCLIMAP[l]	7-1	Era-I
8	WRF V3.6	ACM2[m]	Noah	yes	3	USGS-MODIS	84-12	FNL[n]
9	WRF V3.4	MYJ	Noah	no	3	USGS	28-4	Era-I
10	WRF V3.6.1	YSU	Noah	yes	3	CORINE	672-96	CFSR[o]
11	WRF V3.0.1	MYJ	Noah	yes	3	GlobCover	36-6	CFSR
12	WRF V3.6.1	MYNN[p]	Noah	yes	3	USGS	816-72	Era-I
13	WRF V3.0.1	MYJ	Noah	yes	3	GlobCover	36-12	MERRA[q]
14	WRF V3.0.1	MYJ	Noah	yes	3	GlobCover	36-12	Era-I
15	WRF V3.1	MYJ	Noah	yes	3	MODIS[r]	54-6	FNL
16	WRF V3.6.1	YSU	Noah	yes	3	CORINE	336-96	CFSR
17	WRF V3.5.1	MYJ	Noah	yes	4	IGBP-MODIS[s]	264-24	Era-I
18	UM V8.4[t]	Lock[u]	JULES[v]	yes	4	IGBP-MODIS	36-6	Era-I
19	WRF V3.5.1	YSU	Noah	yes	5	USGS	~~8856-24~~ 2424-24	Era-I
20	SKIRON V6.9[w]	MYNN	OSU[x]	no	5	USGS	51-3	GFS[y]
21	WRF V3.5.1	YSU	Noah	yes	5	USGS	~~8856-24~~ 2424-24	Era-I
22	WRF V3.5.1	YSU	Noah	yes	6	IGBP-MODIS	264-24	Era-I
23	HIRLAM V6.4.2[z]	CBR[aa]	ISBA	no	11	USGS	9-3	IFS[ab]
24	RAMS V6.0[ac]	MYNN	LEAF[ad]	no	12	CORINE	36-12	IFS
25	MM5 V3[ae]	YSU	-	no	20	CORINE	744-24	MERRA

[a]Skamarock et al. (2008b) [b]Büttner et al. (2004) [c]Dee et al. (2011) [d]Janjić (2002)
[e]Niu et al. (2011) [f]Garbarino et al. (2002) [g]Arino et al. (2008) [h]Hong et al. (2006)
[i]Seity et al. (2011) [j]Moigne and Boone (2009) [k]Noilhan and Mahfouf (1996) [l]Champeaux et al. (2005)
[m]Pleim (2007a) [n]NCEP Final analysis [o]Saha et al. (2010) [p]Nakanishi and Niino (2006)
[q]Rienecker et al. (2011) [r]Friedl et al. (2010) [s]Loveland and Belward (1997) [t]Lean and Clark (2008)
[u]Lock et al. (2000) [v]Cox et al. (1999) [w]Kallos et al. (1997) [x]Pan and Mahrt (1987)
[y]Global Forecast System [z]Kallberg (1989) [aa]Cuxart (2000) [ab]Integrated Forecasting System
[ac]Pielke et al. (1992) [ad]Walko and Tremback (2005) [ae]Grell et al. (1994)

Table A5. Statistics of NRMSE for wind speed ($NRMSE_u$) and RMSE for wind speed shear exponent ($RMSE_\alpha$) associated with the groups of PBL schemes across all heights at each site. The number of models in each group is: 6 in the "YSU", 6 in the "MYJ", and 9 in the "Other" group. The smallest value for each metric is in **bold**.

FINO3

Metric	PBL	Mean	Median	St.d.	Min	Max
$NRMSE_u$	YSU	0.047	0.029	0.028	0.018	0.091
	MYJ	0.032	0.029	**0.011**	0.020	**0.055**
	Other	**0.028**	**0.014**	0.045	**0.001**	0.154
$RMSE_\alpha$	YSU	0.029	0.019	0.034	0.004	0.116
	MYJ	**0.010**	**0.010**	**0.007**	0.004	**0.025**
	Other	0.057	0.019	0.120	**0.003**	0.396

Høvsøre

Metric	PBL	Mean	Median	St.d.	Min	Max
$NRMSE_u$	YSU	**0.061**	**0.058**	0.037	**0.024**	0.144
	MYJ	0.063	0.064	**0.013**	0.045	**0.090**
	Other	0.062	0.059	0.026	0.027	0.100
$RMSE_\alpha$	YSU	**0.035**	**0.018**	0.029	**0.005**	0.087
	MYJ	0.049	0.044	**0.011**	0.030	**0.061**
	Other	0.086	0.051	0.100	0.027	0.365

Cabauw

Metric	PBL	Mean	Median	St.d.	Min	Max
$NRMSE_u$	YSU	**0.058**	**0.049**	**0.033**	0.021	**0.127**
	MYJ	0.066	0.053	0.037	0.038	0.146
	Other	0.124	0.086	0.106	**0.007**	0.389
$RMSE_\alpha$	YSU	**0.025**	**0.022**	**0.007**	0.018	**0.036**
	MYJ	0.045	0.023	0.036	0.020	0.117
	Other	0.064	0.075	0.036	**0.015**	0.113

Table A6. Statistics of NRMSE for wind speed (NRMSE$_u$) and RMSE for wind speed shear exponent (RMSE$_\alpha$) associated with the group model grid spacing across all heights at each site. The number of models in each group is: 7 in "Fine", 8 in 'Moderate', and 6 in 'Coarse'. The smallest value for each metric is in **bold**.

FINO3

Metric	Grid spacing	Mean	Median	St.d.	Min	Max
NRMSE$_u$	Fine	**0.024**	**0.020**	**0.015**	**0.001**	**0.055**
	Moderate	0.037	0.027	0.025	0.007	0.080
	Coarse	0.044	0.025	0.046	0.002	0.154
RMSE$_\alpha$	Fine	**0.013**	0.013	**0.008**	0.005	**0.025**
	Moderate	0.015	**0.011**	**0.008**	0.004	0.028
	Coarse	0.067	0.019	0.121	**0.003**	0.396

Høvsøre

Metric	Grid spacing	Mean	Median	St.d.	Min	Max
NRMSE$_u$	Fine	0.057	**0.057**	0.026	**0.024**	0.093
	Moderate	**0.054**	**0.057**	**0.012**	0.027	**0.064**
	Coarse	0.075	0.068	0.034	0.028	0.144
RMSE$_\alpha$	Fine	**0.040**	**0.040**	0.021	0.015	0.076
	Moderate	0.047	0.048	**0.010**	0.030	**0.060**
	Coarse	0.088	0.055	0.109	**0.005**	0.365

Cabauw

Metric	Grid spacing	Mean	Median	St.d.	Min	Max
NRMSE$_u$	Fine	0.086	0.064	0.056	**0.007**	0.178
	Moderate	**0.048**	**0.046**	**0.015**	0.021	**0.078**
	Coarse	0.146	0.107	0.115	0.049	0.389
RMSE$_\alpha$	Fine	0.052	0.030	0.036	0.016	0.117
	Moderate	**0.031**	**0.021**	**0.017**	0.020	**0.066**
	Coarse	0.063	0.060	0.041	**0.015**	0.113

Table A7. Statistics of NRMSE for wind speed ($NRMSE_u$) and RMSE for wind speed shear exponent ($RMSE_\alpha$) associated with each group of simulation lead-time across all heights at each site. The number of models in each group is: 9 in the 'Short', 8 in the 'Medium', and 7 in the 'Long'. The smallest value for each metric is in **bold**.

FINO3

Metric	Sim. length	Mean	Median	St.d.	Min	Max
$NRMSE_u$	Short	0.032	**0.020**	0.044	**0.001**	0.154
	Medium	**0.028**	0.025	**0.014**	0.007	**0.055**
	Long	0.051	0.031	0.028	0.025	0.091
$RMSE_\alpha$	Short	0.052	**0.010**	0.122	**0.003**	0.396
	Medium	**0.016**	0.016	**0.006**	**0.003**	**0.025**
	Long	0.029	0.022	0.036	0.004	0.116

Høvsøre

Metric	Sim. length	Mean	Median	St.d.	Min	Max
$NRMSE_u$	Short	**0.058**	0.059	0.023	**0.024**	0.100
	Medium	0.070	0.068	**0.016**	0.044	**0.093**
	Long	0.062	**0.057**	0.039	0.027	0.144
$RMSE_\alpha$	Short	0.081	**0.044**	0.102	0.018	0.365
	Medium	**0.044**	0.056	**0.023**	0.009	**0.076**
	Long	0.046	0.048	0.025	**0.005**	0.087

Cabauw

Metric	Sim. length	Mean	Median	St.d.	Min	Max
$NRMSE_u$	Short	0.088	**0.058**	0.108	**0.007**	0.389
	Medium	0.103	0.097	0.058	0.043	0.178
	Long	**0.068**	0.064	**0.035**	0.021	**0.127**
$RMSE_\alpha$	Short	0.046	**0.021**	0.038	**0.015**	0.113
	Medium	0.058	0.054	0.038	0.018	0.117
	Long	**0.031**	0.025	**0.012**	0.020	**0.052**