



NEWA Report on uncertainty quantification Deliverable D4.4

González-Rouco, J. Fidel ; Bustamante, Elena Garcia ; Hahmann, Andrea N.; Karagili, Ioanna ; Navarro, Jorge ; Olsen, Bjarke Tobias; Sile, Tija; Witha, Björn

Link to article, DOI:
[10.5281/zenodo.3382572](https://doi.org/10.5281/zenodo.3382572)

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
González-Rouco, J. F., Bustamante, E. G., Hahmann, A. N., Karagili, I., Navarro, J., Olsen, B. T., Sile, T., & Witha, B. (2019). *NEWA Report on uncertainty quantification Deliverable D4.4*. NEWA - New European Wind Atlas. <https://doi.org/10.5281/zenodo.3382572>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Report on uncertainty quantification Deliverable D4.4

J. Fidel González-Rouco (1), Elena García-Bustamante (2), Andrea N. Hahmann (3), Ioanna Karagili (3), Jorge Navarro (2), Bjarke Tobias Olsen (3), Tija Sile (4), Björn Witha (5)



Work Package 4,
Mesoscale Group

August 2019

Author: J. Fidel González-Rouco (1), Elena García-Bustamante (2), Andrea N. Hahmann (3), Ioanna Karagili (3), Jorge Navarro (2), Bjarke Tobias Olsen (3), Tija Sīle (4), Björn Witha (5)

Title: Report on uncertainty quantification
Deliverable D4.4

Department: WP 4, Mesoscale Group

NEWA report
August 9, 2019

ISSN:
ISBN:
978-87-93278-88-2

Abstract (max. 2000 char)

This report explores the quantification of uncertainty in the NEWA production run and probabilistic ensemble. Uncertainty quantification relies on the availability of model simulations and observations. In areas where observations are not available, uncertainty addresses the quantification of the variability resulting from the sensitivity of the model chain tool resulting from the sampling of all the available model configurations tested. Uncertainty can be here regarded as the quantification of model sensitivity or model spread. In areas with availability of observations, the realism of the model ensemble is evaluated by establishing metrics of model-data comparison. Therefore, uncertainty is understood here as the result of the contributions of model sensitivity to different model setups, and of model errors in a model-data comparison framework.

The first part of this report explores the uncertainty derived from model sensitivity subjected to the decisions taken regarding the use of different models setups and how these produce variability in model output. The range of this variability has been regarded as spread in model output and has been quantified in various manners. The second part of this report addresses how model performance can be characterised with the data at hand and whether decisions regarding selection of a given model setup for a production run can be taken on the basis of model performance in a variety of situations, using different variables and datasets as observational targets: wind data from tall masts in the Vestas database; wind speed profiles from tall masts and offshore lidars; surface (10-meter) wind data; satellite data and reanalysis outputs.

Contract no:
N/A

Project no:
N/A

Sponsorship:
EU and EUDP

Pages: 106
Tables: 15
Figures: 79
References: 115

Affiliations:
(1) fidelgr@ucm.es
Univ. Complutense Madrid.
(2) CIEMAT Spain.
(3) DTU. Denmark.
(4) University of Latvia
(5) ForWind, Oldenburg University, Germany

Contents

	Page
Executive summary	5
1 Introduction	8
2 Ensemble spread	11
2.1 Analysis of the NW European domain sub-ensemble	11
2.1.1 The role of the mesoscale ensemble	11
2.1.2 Uncertainty in the numerical setup of the model	12
2.1.3 Methodology	13
2.1.4 Results for wind speed	15
2.1.5 Results for wind direction and stability	18
2.2 Design of the NEWA probabilistic ensemble	20
2.2.1 Introduction	20
2.2.2 Selection of a representative year	20
2.2.3 Context for probabilistic ensemble design	20
2.2.4 Member selection for the reduced ensemble	21
2.2.5 Interpretation of the model spread	24
2.2.6 Sub-setting the spread into categories	25
2.3 Ensemble spread: a regional perspective	28
2.3.1 Data and Methodology	29
2.3.2 Regionalisation and model sensitivity	32
3 Model evaluation	37
3.1 Model evaluation using Vestas wind tower data	37
3.1.1 Measurements	37
3.1.2 Models	39
3.1.3 Site characterisation metrics	40
3.1.4 Validation results	43
3.2 Model evaluation based on tall mast data	54
3.3 Model evaluation based on wind profiles (mostly) offshore	62
3.4 Model evaluation using surface wind data	75
3.4.1 Surface wind data	75
3.4.2 Regional model validation over the SW domain: Alaiz region	76
3.4.3 WRF validation over the European domain	81
3.4.4 Conclusions	82
3.5 Model evaluation using satellite wind data	85
3.5.1 Satellite Winds	86
3.5.2 WRF simulations	86
3.5.3 Results	87
3.6 Model comparison with ERA5 reanalysis data and Global Wind Atlas data	90
4 Conclusions	93
References	99

Executive summary

This report explores the quantification of uncertainty in the NEWA project. Uncertainty is understood here as the result of the contributions of model sensitivity to different model setups, and of model errors in a model-data comparison framework. The first part (Section 2 of this report explores the uncertainty derived from model sensitivity subjected to the decisions taken regarding the use of different models setups and how these produce variability in model output. The range of this variability has been regarded as spread in model output and has been quantified in various manners. The second part of this report (Section 3 addresses how model performance can be characterised with the data at hand and whether decisions regarding selection of a given model setup for a production run can be taken on the basis of model performance in a variety of situations, using different variables and datasets as observational targets: wind farm data from Vestas; tall masts and wind profiles; surface wind data; satellite data and reanalysis outputs.

Model spread

Section 2.1 *For uncertainty estimation, the results from an ensemble of simulations was analysed. The dataset from the sensitivity studies was used (Deliverable 4.3, Witha et al., 2019). Analysis for wind speed at 100 m was carried out, with some analysis for wind direction at 100 m and atmospheric stability.*

Results showed that some of the members were not significantly different from the base run and therefore were excluded from further analysis. All the runs with different Land Surface Models or PBL schemes were sufficiently different from the base run, and were chosen for further analysis, together with some runs with changes in forcing (re-analysis) or in boundary conditions.

Section 2.2 *The final uncertainty estimation should cover the whole NEWA domain, while the large ensemble of runs was carried out for a single domain. For computational reasons only, a small number of members could be run for the whole domain, therefore the task was to select the optimal members for these calculations.*

Preliminary results showed that ensemble properties depend on the geographical region, therefore additional runs were carried out for a second domain. Then the ensemble spread was defined in cumulative distribution space, and the members providing most spread with respect to the production run were selected. Two selected ensemble members were run over whole NEWA domain and the results were analysed.

Section 2.3 *The spread provided by a collection of the WRF regional model simulations with different configuration choices has been explored using eigenvector rationalisation approaches. This provides a frame for the quantification of the uncertainty associated with the methodological variance of the NEWA probabilistic ensemble.*

Results show a reduced number of main modes of wind circulation that provides evidence for common variations of the wind field in selected regions. The differences among the set of sensitivity experiments over each region are expressed to facilitate an spatial interpretation of model spread. This approach demonstrates the time dependence of spread on large scale circulation patterns.

The impacts during specific synoptic transients are not negligible when for instance, different LSMs are used. In general, spread tends to increase during periods with larger wind

variability. The YSU PBL scheme, and the results from complex land surface models, like the CLM or the Noah-MP schemes tend to show larger variance comparatively with the rest of simulations, therefore increasing the model spread.

The model integration strategy, can play an important role influencing spread. The initialisation strategy resets the initial conditions every week, thus hampering the memory of the subsurface processes (temperature and humidity) and their ability to develop at longer timescales and therefore hinders the real potential of the physics of land surface models to impact the results.

Model evaluation

Section 3.1 *The NEWA model-chain was evaluated with 291 tall meteorological masts located all over Europe and Turkey.*

The results of the WRF-WAsP model chain, which is used to create the high-resolution wind atlas has a mean wind speed bias of $0.28 \pm 0.76 \text{ m s}^{-1}$, while the raw wind speeds simulated by WRF and ERA5 have mean biases of 0.02 ± 0.78 and $-1.50 \pm 1.30 \text{ m s}^{-1}$, respectively. Using the NREL 5MW reference wind turbine power curve, the NEWA-simulated wind climatologies results in an overestimation of the mean power production by $13.28 \pm 27.36\%$ and $6.20 \pm 25.16\%$ for WRF-WAsP and WRF respectively. ERA5 underestimates the mean power production by $-40.21 \pm 32.69\%$ on average.

For sites in low to moderately complex orography, WRF-WAsP has a lower mean wind speed bias for all sites (0.06 and 0.21 m s^{-1}) and a lower standard deviation (0.49 and 0.76 m s^{-1}) than WRF (0.54 and 0.90 m s^{-1}) and ERA5 (0.69 and 1.21 m s^{-1}). In terrain with many steep hills, WRF-WAsP overestimates the mean wind speed by 0.62 ± 0.91 , compared to $-0.25 \pm 0.83 \text{ m s}^{-1}$ for WRF and $-2.64 \pm 1.17 \text{ m s}^{-1}$ for ERA5.

The validation study showed that uncertainties remain in the surface characterisation, especially of the surface roughness, and in the effective response of the WRF model to orographic and surface roughness variations, and the validity of the WAsP model assumptions used to downscale the WRF output.

Section 3.2 *The NEWA production run has been compared to data of 14 tall masts over Central Europe.*

While at offshore locations the 100 m wind speeds are well predicted by the NEWA production run, the onshore wind speeds are generally overestimated, especially over more complex terrain (here up to 20%). The correlations are high offshore and near the coast (around 0.9) and lower onshore over complex terrain (down to 0.7). For the onshore locations clear annual and diurnal cycles of the correlation have been found.

Section 3.3 *Further validation of the NEWA production run was performed with mostly offshore wind profiles obtained by lidar and met masts.*

At turbine-operating levels offshore and over simple terrain in Northern Europe and the North and Baltic Seas, the NEWA production run slightly underestimates the mean observed wind speed. The overall absolute biases for this sites and heights are $<3\%$ and the correlations >0.86 .

For taller measurements heights ($>150 \text{ m}$), a systematic under-prediction of the mean wind speed by the NEWA simulations is seen in the evaluation, but it could be partially due to lidar tendency to sample more often at higher wind speeds, which are common at these heights.

Section 3.4 *Since wind speed and direction observations at the hub height are scarce, the evaluation of the mesoscale WRF model skill to reproduce the observed wind requires additional comparisons with available wind measurements. Since there are more surface wind datasets*

accessible for research, they can provide an adequate spatio-temporal coverage to explore the model skill.

Thus, a surface wind database over the whole European domain was specifically compiled for the purposes of the NEWA project, the WiSEd dataset, with a dense network of ca. 4000 stations with wind speed and direction records over the whole European domain.

Analysis focused on the northeastern Iberian region showed that the deviations in the surface wind speed of the WRF simulations with respect to observations are generally larger than the intermodel differences, i.e. model spread is usually smaller than model errors.

Analysis over the northeastern Iberian region, shows that the simulated wind speed tends to underestimate the variance of the observed records. Nevertheless, correlations between simulations and observations are higher, as expected, with height than at the surface. In general, despite difficulties of the model at some sites at the surface to represent the correct levels of variability of the wind, specifically related to the complex orography at some areas, the regional simulation is to a large extent skilful in simulating the observed wind over the region.

The observed WiSEd data as well as the wind mast records have been compared to the NEWA production run over the whole European domain to explore ability of the NEWA WRF model configuration to reproduce the surface observed wind.

In general, it was observed that larger winds over northern Europe tend to correlate better with observations but also tend to show larger errors related to the levels of variance represented by the model simulation. The simulated wind at the masts seemingly outperforms that at the surface but does not necessarily improve as the height increases, rather statistics are homogeneous at the different heights.

Section 3.5 *Satellite winds representative at 10 m above the ocean surface were used to evaluate the offshore areas of the NEWA production run.*

An overall model underestimation was especially evident in the Mediterranean Sea region. In the northern European offshore regions, biases between satellite winds and NEWA were near zero.

The NEWA production run underestimation of the mean winds at 100 m was particularly intense in coastal areas, especially offshore from regions of complex topography, i.e. the Aegean Sea, Gulf of Lyon and Adriatic Sea. However, the period of comparison is not totally compatible and the lifting of the wind speeds introduces added uncertainties of the order of $\pm 0.5 \text{ m s}^{-1}$.

ASCAT winds at 10 m above the ocean surface were used as "ground truth" to assess differences between two WRF ensemble members for 2015 and for the north-west domain covering the North Sea. Higher spatial wind variability was revealed in ASCAT compared to the ensemble members although differences between the latter were also evident.

Section 3.6 *The long-term wind climate of the NEWA production run has been compared to that obtained with the ERA5 reanalysis data.*

NEWA indicates significantly higher wind speeds over land, especially in mountainous regions that are not resolved in such detail in ERA5. Over sea both datasets show very similar results except some deviations near islands or mountainous coasts. Generally, NEWA features much higher standard deviations of wind speed and also higher wind speed maxima.

A preliminary visual comparison to the downscaled Global Wind Atlas (GWA2) results indicates that at turbine-operating heights NEWA predicts higher wind speeds not only over land but also over sea.

1 Introduction

The NEWA project was given the challenging task of developing a high resolution regional simulation of the wind field over the broader European domain that would serve as a scientific state-of-the-art reference for applications relevant in the context of wind energy resources (REN21, 2017; Murthy and Rahi, 2017; Tobin et al., 2016; Palutikov et al., 1987), and potentially also for other end users and scientific communities. For instance those concerned with: wind extremes (Cheng et al., 2014; Smits et al., 2005), relevant both for wind energy (Pryor et al., 2005) and for risk of damage on facilities (Khanduri and Morrow, 2003), hydrology (McVicar et al., 2012) or agriculture (Farquhar and Roderick, 2005; Cleugh et al., 1998); all of them demanding information of wind speeds within the planetary boundary layer.

The wind atlas is produced using a chain of models, starting from the global scale ERA5 re-analysis (Copernicus Climate Change Service (C3S), 2019), used for driving a very high 3-km resolution dynamical downscaling with the WRF mesoscale model (Skamarock et al., 2005) and further linking this model chain down to the final resolution of 50 m of the microscale model (Hahmann et al., 2019). The very high resolutions targeted in the model chain, as well as the 30-year long period of simulation, places a milestone in terms of computational requirements. The computational challenge has been additionally burdened by providing an assessment of uncertainties that adds the need for producing an ensemble of model simulations. Indeed, one key element of NEWA is being able to provide an *Uncertainty Quantification* for the wind field estimates, i.e. a probability density function, confidence interval or similar that accounts for the very likely range where observed wind values would be expected to be constrained. The purpose of this report is to describe the rationale for *Uncertainty Quantification* within NEWA and provide some examples of the variety of studies aiming at addressing this problem.

Uncertainty in producing a wind atlas or, for that matter, in generating an atlas for any atmospheric variable within climatological timescales as the ones dealt with in NEWA, stems from three main aspects: the real experiment cannot be controlled, observations representing reality are often not available or are of low quality, and any modelling approach that is taken will be dependent on the decisions regarding model parameters and may include errors and/or missing processes at different spatial and temporal scales. Therefore, a complete evaluation of uncertainty would require to account for uncertainties of experimental and modelling nature, as well as for the assumptions adopted in how model and data are compared.

Experimental uncertainties do include observational errors, but not only these. Additionally, methodological variability in experimental setups, and data management errors (Jiménez et al., 2010b; Lucio-Eceiza et al., 2018a,b) contribute to observational uncertainty. For gridded data sets, interpolation methods and sampling distribution are also an issue (Brinckmann et al., 2015). Finally, non-determinism in the variables associated to the chaotic nature of internal and forced variability in the system adds to observational uncertainty (e.g. Peixoto and Oort, 1984). The use of downscaling approaches to obtain high resolution mesoscale estimates of the wind field is hampered by the general fact that models are a representation of reality and, ultimately, they bear approximations and even sometimes lack of knowledge of the underlying processes. This affects both statistical (Lucio-Eceiza et al., 2019; García-Bustamante et al., 2011, 2012) and dynamical (e.g. Jiménez et al., 2013; Sotillo et al., 2005) downscaling approaches as the ones developed within NEWA. The most obvious simplification regards spatial and temporal resolutions. The increase in resolution of dynamical downscaling approaches does sometimes improve model performance, but not always, due to errors that take place at the large scale and/or to representation errors at the small scale (Jiménez et al., 2010a). Additionally, model simulations are invariably the result of a set of decisions on selecting among a variety of model possible constituent configuration designs and among available physical parameterisations that account for subgrid scale processes (von Storch, 1995). Systematic sampling over these decisions produces variability, i.e.

uncertainty, that reflects sensitivity of the output of the downscaling tool (e.g. García-Bustamante et al., 2012; Jiménez et al., 2012). Thus, uncertainty stems from all the methodological tools and decisions taken during model configuration down to the local microscale that embeds the domain of interest of the target variable. Even at the level of large scale dynamics, deciding upon considering one reanalysis (Fujiwara et al., 2017) or another influences estimations of wind at the local scale (Lucio-Eceiza et al., 2019).

During the last decades, a wealth of studies have helped to widen our understanding of climate variability at regional and continental scales. This has been achieved by developing progress in understanding uncertainty in both observations and also in model simulations (IPCC, 2013). Recent developments in observational data products, addressing a variety of methods and sources help to better understand observational uncertainties and deficiencies in our observation of the system (Hartmann et al., 2013). In turn, computational resources have allowed for taking very large steps in global climate modelling, including a large variety of reanalysis products and the new atmosphere-ocean reanalysis used in NEWA (Copernicus Climate Change Service (C3S), 2019) (Hartmann et al., 2013; Flato et al., 2013). This progress has been fostered by recent generations of General Circulation Models (GCMs) and coordinated evaluation frameworks (Coupled Model Intercomparison Project Phase 5, CMIP5 Taylor et al., 2012), recently expanded for the next Intergovernmental Panel on Climate Change assessment (IPCC) under CMIP6 (Eyring et al., 2016).

Since the last European Wind Atlas (Troen and Petersen, 1989), large progress has been made in mesoscale modelling. The approach taken in the NEWA project has accommodated the previously described context to the computationally demanding objective of producing a Wind Atlas that results from state-of-the-art modelling and knowledge in model data comparison. The mesoscale model is coupled to the microscale component through a statistical downscaling approach (Sanz Rodrigo et al., 2019). The simulation that provides the mesoscale fields has been decided upon a careful sensitivity analysis of potential model configurations (Witha et al., 2019). In order to accomplish model resolutions of 3 km over Europe, the broader European domain is partitioned into a set of 10 partially overlapping tiles that are simulated with the WRF model and blended into a single domain. A large number of simulations is produced for some of the domains in order to decide on the configuration of the final production simulation and quantify uncertainty. The ensemble of simulations incorporates model variants to sample the impact of changes in the boundary conditions (reanalysis data set), sea surface temperature fields, model configuration considering different nudging and spin up options or different physical parameterisations (e.g. land surface model; see NEWA Deliverable 4.3, Witha et al., 2019).

Uncertainty quantification relies on the availability of model simulations and observations within each domain. In areas where observations are not available, uncertainty addresses the quantification of the variability resulting from the sensitivity of the model chain tool resulting from the sampling of all the available model configurations tested. Uncertainty can be here regarded as the quantification of model sensitivity or model spread. Section 2 of this report is devoted to a variety of analyses that target understanding and estimation of model uncertainties characterised as ensemble spread. The text focuses first (Section 2.1) on the analysis of ensemble spread over a large ensemble of WRF simulations centred in one of the model tiles over the NW of Europe. This allows for assessing the influence of different model physics and surface characteristics, and how these impact model biases. New metrics of inter model comparison are presented in Section 2.2 that allow to characterise differences in the probabilistic distribution of wind speed and to derive a means of designing the NEWA probabilistic ensemble. Finally, Section 2.3 closes this part by introducing a complementary approach to characterising changes in model sensitivity with time based on projection onto the Empirical Orthogonal Functions of the system.

In areas with availability of observations, the realism of the model ensemble is evaluated by establishing metrics of model-data comparison (Section 3). This section targets a variety of observational products: from wind tower data distributed over the European domain (Section 3.1); tall mast data (Section 3.2); wind profiles (Section 3.3); to surface wind data (Section 3.4). The section is closed with two model-data comparison sections addressing different types of exist-

ing global products: satellite data (Section 3.5); and Reanalysis and Global Wind Atlas products (Section 3.6).

2 Ensemble spread

Model uncertainty is addressed here as the results of unknowns in relation to the optimal decisions regarding model representation of dynamics and physics; in this case focused on the simulation of wind. Model uncertainty contributes to model performance in terms of reproducing observations but it should not be confused with uncertainty in reproducing observations (see Section 3). Model uncertainty represents the sensitivity of our simulation tools to all decisions in its configuration, including boundary information and surface fields that may impact model output. Sampling the space of possible model configurations and factors that may affect the simulation of wind leads to an ensemble of results that represent the variability of the tool in simulating wind or wind energy at a given time and location. The quantification of the variability in the resulting simulation is referred to here as model spread.

Section 2.1 discusses and presents results of the analysis of a large ensemble of simulations over one of the WRF sub-domains. It introduces the Earth Moving Distance metric and assesses the impact on the simulated wind fields over this sub-domain of using a variety of model setups, specifically different PBL and land model schemes, also having insights into the effects of different land use and stability classes. This analysis has provided the understanding that not all model setups contribute equally to spread. Some model setups contribute very little while the influence of others is dominant. For the development of the probabilistic wind atlas, a limited number of ensemble members could be developed due to computational constraints. The procedure and decision on which ensemble member setups were selected to compute for all the European sub-domains is described in Section 2.2. Finally, the analysis of model uncertainties carried out in Sections 2.1 and 2.2 addresses biases related to model setups that affect the distribution of wind. However, uncertainties may be variable with time and may be dependent on how different model setups manage changes associated to different synoptic configurations and in general to the variability of wind speed related to changes in large scale circulation. Section 2.3 addresses this topic by showing case examples that illustrate how the spread may change with time and can be region dependent.

2.1 Analysis of the NW European domain sub-ensemble

The large number of different WRF setups created in this project serve a dual purpose. First, they were used to determine the best configuration for the production run, but they also could be used to provide an uncertainty estimate. To achieve the second goal, it is necessary to define what is the role of the multi-physics ensemble and what are the principal sources of uncertainty that can be expected to be represented in this ensemble on theoretical grounds. The next step is to quantify the differences between ensemble members and estimate the relationships between the members. Although the wind speed distribution at 100 m is the most important parameter for energy applications, additional parameters such as wind direction and atmospheric stability should be also analysed. As the ensemble was created with dual purpose, then some of the ensemble members are not that different from the base run and therefore cannot add any useful information. Such members should be identified and then excluded from further analysis.

2.1.1 The role of the mesoscale ensemble

The creation and analysis of the mesoscale ensemble required many decisions during the process. To put these choices in their context, the role and the goals of the mesoscale ensemble are introduced first.

The mesoscale ensemble used in NEWA is the so-called “multi-physics” ensemble (e.g. Lee

et al., 2012) as opposed to the initial conditions ensemble, e.g. what is done in the ECMWF ensemble (e.g. Leutbecher and Palmer, 2008). A common (initial condition) NWP ensemble is based on the idea that the behaviour of the atmosphere is chaotic, and therefore some time after integration starts from very similar initial conditions, a set of vastly different, but equally valid and probable states evolves in the simulation. In the case of a wind atlas, when nudging is introduced, the basic premise of the initial conditions ensemble loses its founding as the model is nudged towards the state of the atmosphere in the reanalysis at regular intervals. Also, in a wind atlas one is not interested in the state of the atmosphere at a particular time, but in the mean climatological behaviour of the atmosphere.

In some applications, an ensemble is used to enhance the precision of the predicted values, when the mean of the forecast ensemble is expected to outperform any single member. This is not the case in this work. It is the expectation of the authors that the production run setup, described in Deliverable 4.3 (Witha et al., 2019), will outperform any ensemble-derived values in predicting the wind climate over Europe. The role of the ensemble is only to provide the uncertainty estimation.

Mesoscale models are incapable of resolving microscale processes directly because of the grid resolution they use — 3 km in this case. The description of the sub-grid scale processes is the task of the parameterisation schemes, especially Planetary Boundary Layer (PBL) scheme, together with the Surface Layer (SL) scheme, which is in turn influenced by Land Surface Model (LSM). Therefore, one can argue that by analysing different PBL schemes, the uncertainty in average representation of micro-scale processes is analysed.

However, it is important to distinguish between the description of physical process, e.g. transfer of momentum, and the description of surface properties that quantitatively influence the result. In each grid cell the surface properties are characterised with a set of parameters that describe the properties of small-scale processes in a simplified manner. A parameter most relevant to the wind energy is the surface roughness coefficient z_0 . For most cases each grid-cell has a single surface roughness value for whole 3×3 km area. Such representation cannot account for small scale obstacles and features, and therefore small-scale processes caused by geometrical features cannot be resolved without increasing resolution. However in the context of NEWA project, instead of increasing the mesoscale model resolution, a meso-micro model chain is employed, where the resolution of smaller scale wind field is the task for the micro-scale model such as WAsP.

2.1.2 Uncertainty in the numerical setup of the model

Earlier in this chapter it was asserted that the equations solved in the dynamical core of the model contain no uncertainty (with the caveat about chaos). Strictly speaking, the object solved is not the continuous equations of fluid dynamics, but their numerical representations, that in turn have no claims of being inerrant.

In fact, during the study, some of the numerous options were tested and found to have very little impact, which is a good thing. For completeness they are described here. Similarly, some of the other parameterisation schemes (radiation, microphysics) were tested just to rule out their importance.

There is one exception in this group of ensemble members, and that is the choice of the type of nudging (grid nudging versus spectral nudging and nudging in the outer domain only, or in all domains). Although there are some arguments why spectral nudging should perform better than grid nudging, during the verification phase of this work (Deliverable 4.3., Witha et al., 2019) the grid nudging in all domains showed significantly better results for the RMSE metric than the baseline outer-domain only spectral nudging.

To summarise, the main goal of the mesoscale uncertainty ensemble is to capture:

- The uncertainty in mesoscale processes that are influenced by a choice of PBL scheme — thermally and orographically forced.
- The uncertainty caused by different choice of boundary conditions, especially SST datasets.

Additional ensemble members are analysed, such as with different roughness coefficients, re-analysis forcing or with different numerical parameters, mostly, to confirm the preliminary results suggesting that they are not that important.

In some applications the agreement between different ensemble members can be interpreted as increased confidence in the parameter values provided by such members. In this work such conclusions can be made only with certain caveats. Some of the ensemble members were explicitly designed to test whether a change in some aspect of setup (parameterisation scheme or numerical parameter) even has influence on the wind climate, and therefore an agreement between such ensemble members cannot be used to make any conclusions about the uncertainty in the results.

On the other hand, if the members that use different reanalysis agree among themselves or if the members using different PBL schemes agree in some regions but disagree in others - then one could argue that such comparison contributes to understanding about the sources and quantity of uncertainty in mesoscale ensembles.

There are two main goals of this work — first is to give an overview about the properties of the ensemble spread and second is to create a reduced ensemble that would cover the whole NEWA domain, that is, most of Europe.

Due to the heterogeneity of the ensemble and sensitivity-testing aspects of creation of some of its members, one should not approach all ensemble members as equal and mechanically calculate ensemble statistics. First, the members that significantly differ from the base run must be identified, and the rest should not be used in further analysis. Therefore the first part of the analysis consists of analysing all the members versus the baseline. Then, in order to get a better understanding of the mechanisms and processes behind the ensemble spread, groups of members with similar properties, i.e. members with different sources of boundary and initial conditions are analysed together.

The selection of a sub-ensemble builds on the work and understanding derived in this analysis but lends itself to a more technical approach. It is described in Section 2.2 of this report.

2.1.3 Methodology

Each ensemble member is created by running a specific WRF setup for a time period of one year. The list of ensemble members and the creation of the dataset is described in Deliverable 4.3 (Witha et al., 2019).

The basic object that is used for the analysis is the probability distribution of wind speed at 100 m height in each grid point for each ensemble member. The distribution was calculated using bins of 1 m s^{-1} , 30 bins in total - from 0 to 30 m s^{-1} . Mostly the cumulative distribution function (*scdf*) is used. Two additional variables are analysed with secondary priority - the distribution of wind direction at 100 m, and atmospheric stability near the surface.

Typically, when an ensemble approach is used, a single variable of interest is defined and then all the values provided by the ensemble for this parameter are analysed. Here, a different approach is used. Why did we not choose to analyse the ensemble of average wind speeds, with each average wind speed value provided by a single ensemble member?

Wind power is proportional to the third order of wind speed. Therefore, for a correct description of potential wind power it is necessary for models to correctly replicate the whole wind speed distribution. Preliminary results showed that different ensemble members could have different distribution shapes. The mean value is a bad metric to distinguish distributions from each other, as two distributions can have the same mean value while having vastly different shapes. In addition, for circular variables, such as wind direction, the notion of "mean wind direction" is meaningless.

Therefore, a decision to employ the EMD (Earth Mover's Distance) was made. EMD is a metric that is widely used in image processing. If one imagines two distributions as two piles of dirt, then EMD can be described as the work (in the physical sense) necessary to move the pile from one shape to another. Equivalently, it can be described as the area between the two cumulative distribution functions of the distributions. This interpretation will be later used to select the ensemble members for the reduced ensemble. The unit of EMD is the same unit that is

used for the underlying variable, i.e., if a *cdf* of wind speed in m s^{-1} is analysed, then the unit of the EMD will also be wind speed in m s^{-1} .

Difference in mean wind speed, often called *bias*, is a widely used metric and it has the benefit that people working in wind energy have established an intuition of what constitutes significant differences. For instances, most experts in the field would agree that the difference of average wind speed between two ensemble members of 0.05 m s^{-1} is not significant, while 1 m s^{-1} is quite important. In order to build on this intuition, one can use the property of EMD that if two distributions have the same shape but different means, then the EMD is equal to the difference of means. Namely, the meaning of EMD for a more general audience can be explained as "roughly the same as difference in means, but only accounting for different shape of distribution". If EMD and difference in means are calculated for the same distributions, then EMD always is bigger or equal than the difference in means.

In Figure 1 an example illustrates the difference between EMD and difference of the means of two distributions. For two ensemble members in each grid point both the EMD and the difference in means is calculated. Although there are many points where the differences between these two metrics are negligible, there are points where the both distributions have the same means, while having EMD values of up to 0.3 m s^{-1} .

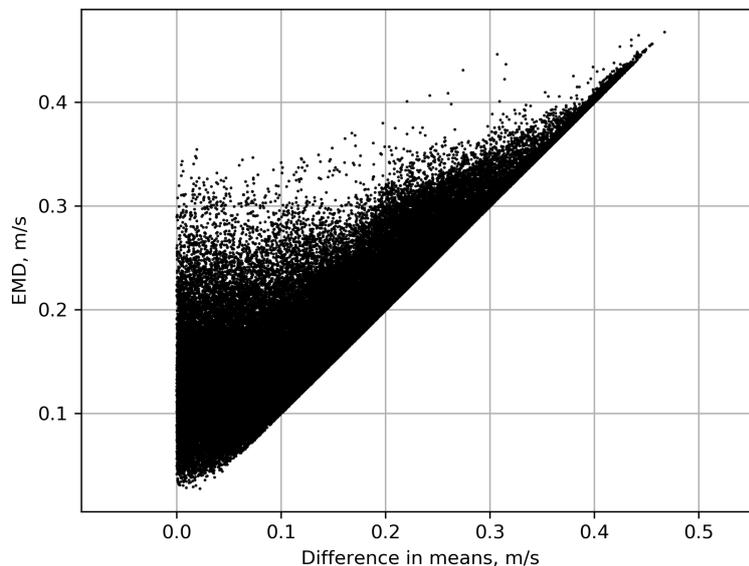


Figure 1: Comparison between EMD (y-axis) and distribution means (x-axis) between two ensemble members (MYNN-MYNN and MYJ-MO) in the NW domain. Each point corresponds to EMD and means in a single grid point.

The EMD is calculated in Python using the PyEMD package (Pele and Werman (2008)). This package allows user-defined distances between histogram points, allowing to apply it to circular variables, such as wind direction. To analyse the stability, 7 stability classes were used according to Gryning et al. (2007) and it was assumed that all the neighbouring stability classes are equidistant from each other, and the numerical value of the between the centres of the classes is 1. Therefore the EMD for stability is expressed in arbitrary units. The EMD for wind direction is expressed in degrees.

If the difference in mean values between three distributions, let's call them A,B and C is analysed, then knowing the difference between A and B on one hand and B and C on other hand it is possible to unambiguously determine the distance between A and C. That is not the case for EMD metric.

There are two types of ensemble members - sensitivity and multi-physics. While multi-physics

members were designed to cover a large spectrum of LSM/PBL/SL combinations, all sensitivity members are small variations of a single base run, which uses the MYNN PBL scheme together with the MYNN Surface Layer scheme. After the verification phase, a different member was chosen for the production run (MYNN PBL scheme with MO Surface Layer scheme). Therefore in the first part of the analysis all members are analysed with respect to old base run (MYNN-MYNN) while in the second part of the work (Section 2.2) the members are mostly compared with the production run (“new base”, MYNN-MO).

2.1.4 Results for wind speed

To reduce the complexity of analysis, the first task was to exclude ensemble members that are not significantly different from the base run and therefore do not provide any additional variability. Two metrics: (a) the difference in the annual averages and the (b) EMD between each ensemble member and the base run were calculated in each grid point. Then the maximum value over all grid points was found. The 20 grid points closest to the edges of the domain were excluded from the analysis.

Table 1 lists all simulated ensemble members and provides information on their difference towards the base run (maximum EMD and difference between means in the domain). A threshold of 0.2 m s^{-1} for difference in means and EMD was defined to consider an ensemble member as sensitive to the changes from the baseline. All sensitive members are marked with "Y" in the column "sens".

In some cases due to technical problems it was impossible to calculate the EMD. In those cases the value in the table is shown as "F". Members 12–15, where the surface roughness length was changed, were excluded from further analysis, according to arguments set out in establishing the framework. Similarly, the RUC-YSU-MM5 member was excluded from further analysis because of suspicions that the results over the water are not correct.

It was important for the ensemble members to represent variability according to the mesoscale framework. The ensemble members were set up with great care, to ensure that the surface representation, such as roughness coefficients, are consistent across all ensemble members. The results showed that even using the same roughness parameter tables for all ensemble members does not necessarily translate to the same roughness coefficients actually being used. That led to the necessity to actually modify the code for some of the ensemble members. These ensemble members are identified in the table by adding "mod" to the ensemble name.

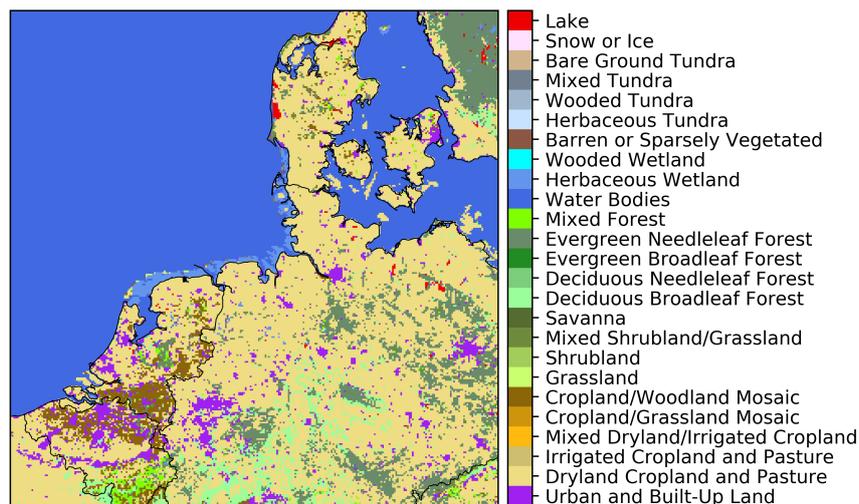


Figure 2: Land use classes for the NW domain

There were two types of modifications. One concerned ensemble members RUC-VEG (24), PXLISM-ACM2 (28-29), where the coastal zone in the North Sea showed significant differences

	short_name	long_name	means	EMD	sensitive	ens
1	xw_36_yw_36	EES81_2551040004_A	0.089	0.096	N	N
2	large_relax	EES81_2551040004_B	0.065	0.072	N	N
3	2-way-nest	EES81_2551040004_C	0.085	0.096	N	N
4	grid-nudging-D3	EES81_2551040004_H	0.258	0.273	Y	Y
5	grid-nudging-D1	EES81_2551040004_I	0.122	0.125	N	N
6	spec-nudging-D3	EES81_2551040004_J	0.194	0.192	N	N
7	ERA1	IIS81_2551040004	0.171	0.177	N	N
8	MERRA2	MMS81_2551040004	0.223	0.222	Y	Y
9	FNL	FFS81_2551040004	0.154	0.158	N	N
10	Vers-361	EES61_2550040004	0.426	F	Y	N
11	MYNN-unmod	EES81_2550040004	1.469	F	Y	N
12	Z0-cycle	EES82_2551040004	0.226	0.230	Y	N
13	Z0-cycle-NCAR	EES82_2551040004_A	0.479	0.483	Y	N
14	Z0-aggr	EES83_2551040004	0.427	0.432	Y	N
15	Z0-aggr-COR	EES84_2551040004	0.328	0.329	Y	N
16	HRSST	EEH81_2551040004	0.087	0.090	N	N
17	SST-ERA5	EEE81_2551040004	0.046	0.049	N	N
18	OISST	EEO81_2551040004	0.519	0.528	Y	Y
19	MYNN-MM5	EES81_2511040004	0.344	0.347	Y	Y
20	MYNN-MO	EES81_2521040004	0.554	0.554	Y	Y
21	MYJ-MO	EES81_2220040004	0.461	0.467	Y	Y
22	YSU-MM5	EES81_2110040004	0.591	0.594	Y	Y
23	RUC	EES81_3551040004	0.395	0.394	Y	Y
24	RUC-VEG mod	EES81_3551040004_A_mod	0.525	0.518	Y	Y
25	RUC-MYNN-MO	EES81_3521040004	0.566	0.572	Y	Y
26	RUC-YSU-MM5	EES81_3110040004	0.510	0.512	Y	N
27	RUC-ACM2-PX	EES81_3770040004	0.590	0.595	Y	Y
28	PXLISM-ACM2-PX mod	EES81_7770040004_mod	F	F	Y	N
29	PXLISM-ACM2-MM5 mod	EES81_7710040004_mod	0.738	0.741	Y	Y
30	SLAB	EES81_1551040004	0.339	0.341	Y	Y
31	SLAB-MYJ-MO	EES81_1220040004	0.507	0.515	Y	Y
32	SLAB-YSU-MM5	EES81_1110040004	0.633	0.632	Y	Y
33	SLAB-ACM2-PX	EES81_1770040004	0.519	0.523	Y	Y
34	NOAHMP mod	EES81_4550040004_mod	0.206	0.221	Y	Y
35	NOAHMP-MYNN-origMP mod	EES81_4550040004_A_mod	1.177	1.180	Y	Y
36	NOAHMP-MYNN-optsfc2 mod	EES81_4550040004_B_mod	0.725	0.728	Y	Y
37	NOAHMP-MYJ-MO mod	EES81_4220040004_mod	0.547	0.547	Y	Y
39	NOAHMP-YSU-MM5 mod	EES81_4110040004_mod	F	F	Y	N
40	grell-freitas	EES81_2551040304	0.071	0.088	N	N
41	rrtmg	EES81_2551040024	0.091	0.094	N	N
42	cam	EES81_2551040003	0.324	0.323	Y	Y

Table 1: Overview of the ensemble members and their difference from the base run in terms of difference of means and EMD (maximum values in the domain). The column "sensitive" indicates if the member yields significantly different results than the base runs or not. The column "ens" indicates if the member was later included in the sub-ensemble (see Section 2.2.4).

in the grid points belonging to the land-use class "Herbaceous wetlands" or number "15", as shown in the land use classes for the domain in Fig. 2. For the members with NOAHMP LSM there were changes associated with LU class "Urban and Built-up land" or number "1".

Examples for EMD between the base run and members RUC_VEG (unmodified), RUC_VEG (modified), as well as members NOAHMP (unmodified) and NOAHMP (modified) are shown

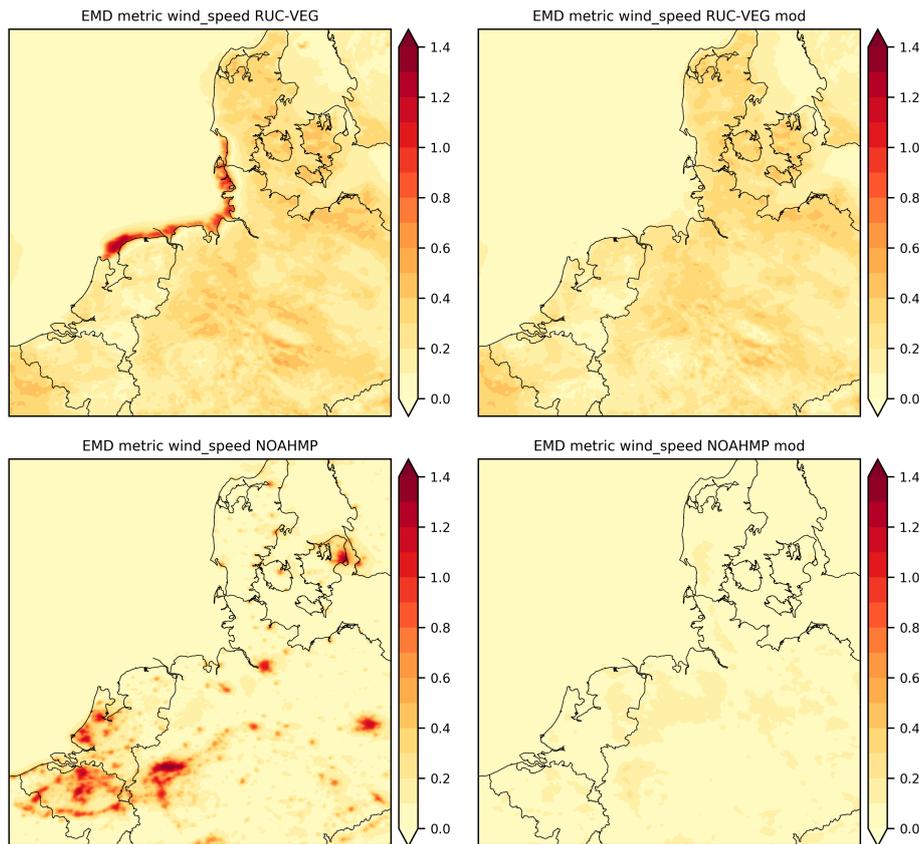


Figure 3: EMD between base run and members RUC_VEG and NOAHMP before and after modifications to remove influence from land-use classes.

in Fig. 3. The values of EMD and the difference in means are very close in Table 1. That is because the maximum value over domain is shown in the table, and as seen in Fig. 1, if the two distributions are significantly shifted with respect to each other than the EMD tends to be close to the difference in means. This underscores the role of the EMD as a metric that can distinguish between two distributions with different shapes that have the same means.

The members that are not sensitive have common properties. Let us analyse those non-sensitive members in more detail. First, members where just the numerical options were tested (1-3) are very similar to the base run, and that is a positive signal that the setup is robust from a numerical perspective.

Interestingly, only one of the members with different nudging options (4–6) has a significant difference and that is the member where grid nudging in all domains is used.

The next set of members concerns different reanalyses used as lateral boundary conditions, initial conditions and forcing for assimilation (7–9) and different Sea Surface Temperature datasets as boundary conditions over the water (15–18). Of all 3 different reanalyses tested, only member using MERRA2 shows some sensitivity although the differences are very small.

Of the SST datasets (16–18) only OISST shows some significant changes in the wind field (Fig. 4). Although the largest differences are over the North Sea, non-negligible differences can also be found onshore, far from the coast.

Similarly, changes in the parameterisation schemes that do not describe surface or near surface parameters, such as cumulus parameterisation (40) and radiation parameterisation (41–42), do not necessarily change the wind direction distribution significantly, with the exception of changing radiation scheme to CAM, that seems to have some effect.

The unmodified version of MYNN-MYNN scheme was excluded from further analysis because there are reasons to believe that the changes introduced in 3.8.1 significantly deteriorate the quality of the results (see also the discussion in Witha et al. (2019)).

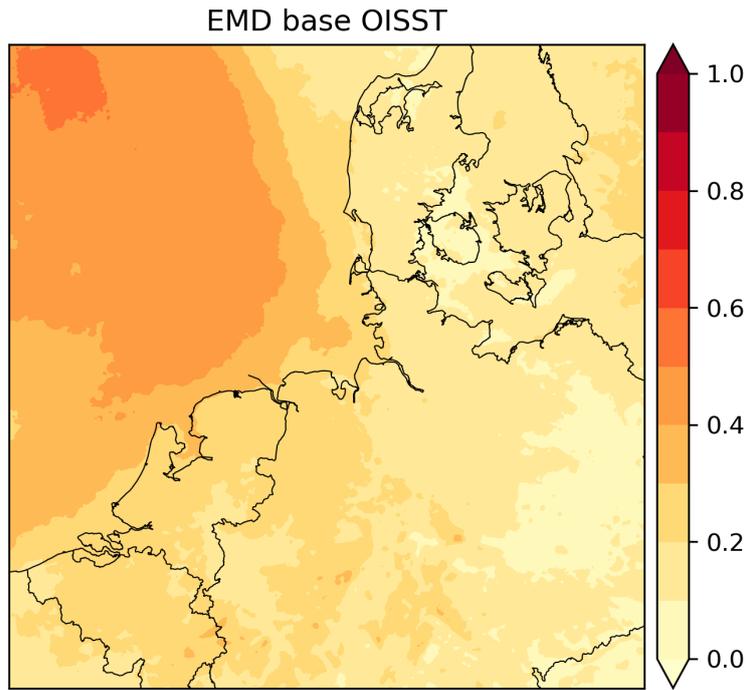


Figure 4: EMD between wind speed distributions of ensemble members "base run" and OISST (18). 2015, NW domain, 100 m height, m s^{-1} .

2.1.5 Results for wind direction and stability

Wind direction is a circular variable, therefore using the EMD metric is preferable over using the difference in distribution means or medians because the latter two have no clear interpretation.

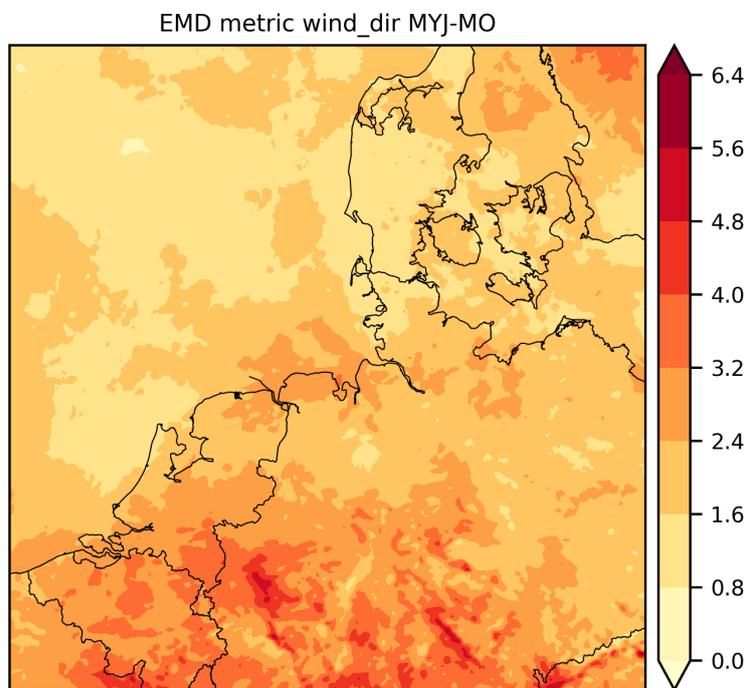


Figure 5: EMD between wind direction distributions of ensemble members "base run" and MYJ-MO (21). 2015, NW domain, 100 m height, degrees.

Similarly, EMD can help identify changes in the stability distribution, however in this case the result is expressed in arbitrary units. An example of EMD for wind direction is shown in Fig. 5. The changes in wind direction are expressed in degrees, and the values are moderate. Interestingly, the largest differences seem to be associated with river valleys in south-west Germany.

The difference in stability distributions (example shown in Fig.6) is closely related to the land-use class, as evidenced by the fact that the locations of major cities can be easily identified in Fig.6).

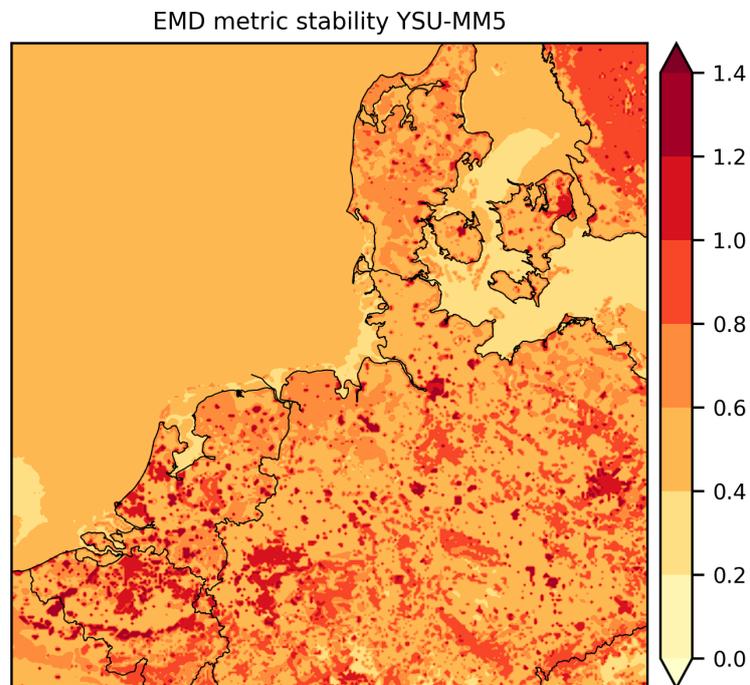


Figure 6: EMD between stability distributions of ensemble members "base run" and YSU-MM5 (22). 2015, NW domain, 100 m height, arbitrary units.

2.2 Design of the NEWA probabilistic ensemble

2.2.1 Introduction

Due to the computational constraints large numbers of test runs can be carried out only for a limited area domain, not for the complete NEWA domain. Due to the dual scope of the ensemble, namely, the necessity of choosing the best member for the production run, the domain had to be located in a way that it would cover the high quality measurement stations of the North Sea and southern Baltic Sea.

In contrast, the final wind atlas is expected to cover most of Europe, and therefore faces even greater computational constraints. Therefore it was necessary to select a small number of those ensemble members that would best represent the properties of the whole ensemble in a meaningful way.

However, it is possible that the relationships between ensemble members in a specific domain depend strongly on geographical parameters, such as the typical land-use class (e.g. Scandinavian domains include a lot of forests). Similarly, typical weather patterns could lead to differences in ensemble spread between northern and southern Europe. Therefore, the selection of a reduced ensemble is not a straightforward task.

2.2.2 Selection of a representative year

The previous ensemble simulations were done for the year 2015 as this year has been focused on in previous studies and analyses. The final reduced NEWA Ensemble should however be run for a year that is as much as possible a representative year with average wind conditions. Furthermore, it should be a year for which ample validation data (Vestas wind tower data, cf. section 3.1) is available. The annual wind conditions have been determined by taking the NEWA production run for SB (Northern Europe) and GR (South-Eastern Europe) domains and comparing the annual mean wind speed maps against the 15 year long-term mean (at that time only 15 years of production run data from 2003-2017 were available). Furthermore wind speed distributions and wind roses have been compared at selected points in the two domains. An example for the location of FINO 3 in the SB domain is shown in Figures 7 and 8. Finally, the year 2012 was chosen as a relatively normal wind year with ample observations available. For the tests described in the following section a number of promising ensemble members was re-computed for the year 2012 and the NW and GR domains.

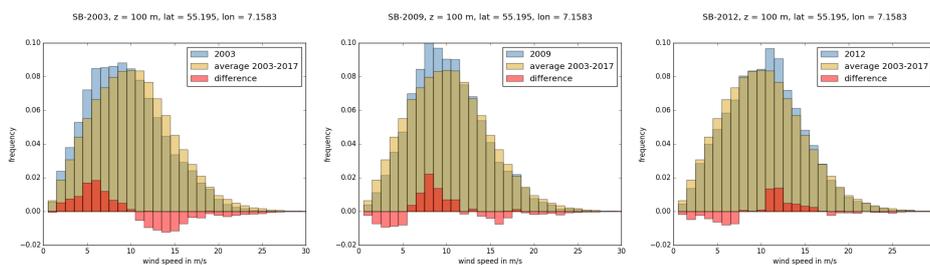


Figure 7: Wind speed distributions for years 2003, 2009 and 2012 compared to the long-term (2003-2017) mean at the location of FINO3 (55.1950°N, 7.1583°E).

2.2.3 Context for probabilistic ensemble design

In the previous sections the ensemble was defined using the cumulative distribution function (*cdf*) of each ensemble member in each grid point and the EMD metric was used to compare each two ensemble members. All ensemble members were verified against the available observa-

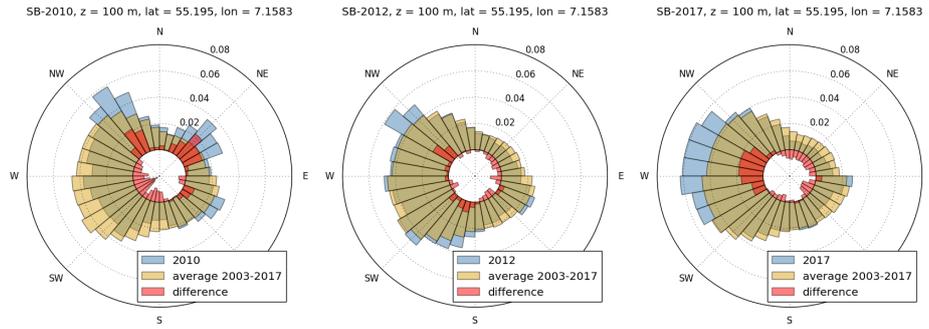


Figure 8: Wind roses for years 2010, 2012 and 2017 compared to the long-term (2003-2017) mean at the location of FINO3 (55.1950°N, 7.1583°E).

tions in 8 coastal stations as described in Witha et al. (2019). All of the members, while differing between themselves, were grouped closely together, and quite often the difference between ensemble members were smaller than the difference between ensemble members and observations. The analysis of the ensemble members described in the previous sections showed that the differences between the ensemble members are limited and most are introduced by the parameterisation schemes. Therefore, a conclusion can be made that every ensemble member provides a reasonable (although not necessarily most correct) estimation of the wind speed at a certain point. In the context of the previous sections, therefore, all the variability captured in the ensemble is to some extent valid and useful, if it is used to characterise the uncertainty of the WRF model as a tool and under the assumptions and caveats described earlier. If infinite computational resources were available, then theoretically, all the members could be calculated for the whole of Europe. However, because of the limitations of computing power it was necessary to choose a few members that capture most of the uncertainty present in the full ensemble.

2.2.4 Member selection for the reduced ensemble

Let us start by defining how the uncertainty captured by the ensemble is quantified using cumulative distribution functions, and let us define the gain in spread that is achieved by adding a member to a reduced ensemble.

As previously stated, the basic object of the analysis is the *cdf* of a single ensemble member wind speed distribution in a single grid point. Now let us define the basic object of the ensemble spread as the collection of *cdfs* of all the ensemble members in a single grid point. The area between each two *cdfs* is the value of EMD metric between those two distribution. However, now a new metric that captures the spread between 3 or more ensemble members is necessary. We propose to use an extension of the definition of EMD and define the *total spread* as the area of the envelope of all *cdfs*. Hence, for each wind-speed bin we calculate the difference between the highest and lowest value of cumulative probability and then sum the differences over all wind-speed bins.

With such a definition it is straightforward to define the amount of spread gained by adding a member to the ensemble. It is the difference in the total spread before and after the addition of a new ensemble member.

We already have the first ensemble member, the production run. The next member can be selected by calculating and ranking the rest of the members according to the gain in spread each of them could provide. After the first two members are selected, it is important to recompute the gain that the rest of the ensemble members can provide to the ensemble spread that is already covered by the two members of the reduced ensemble. Then the third ensemble member can be chosen, and so on.

An example of this process is shown in Figure 9. This approach is straightforward if a single grid point is analysed, however initial analysis suggested that a member can add spread in some geographical regions while completely agreeing with other members elsewhere. One approach

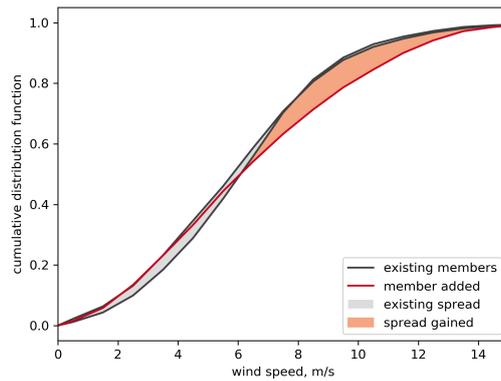


Figure 9: Illustration of the calculation of the increase in ensemble spread after adding another member to the ensemble.

would be to treat all grid points as having equal weight and simply sum or average over all points in the domain. However, the success of such approach depends on whether the test domain is representative of the wider region that needs to be covered by the final ensemble.

The initial test domain (NW) covering Denmark, the Netherlands and most of Germany was chosen to contain as many observational sites as possible. During the project it was decided that a single member should be run for the whole NEWA domain for testing purposes and the member with YSU-MM5 PBL/SL scheme was chosen. A simple preliminary analysis, comparing the difference in means for the whole NEWA domain (Fig. 10) suggested that the differences are strongly influenced by orography, especially in the southern part of Europe. The NW test domain contains no significant mountain ranges. Therefore a decision was made to run the ensemble members that were proven to be sensitive and significantly different for another domain, that would contain more complicated orography, and in other ways could represent different climatological conditions. The production domain covering Greece ("GR") was chosen for this task, and the respective ensemble members were run for the year 2012.

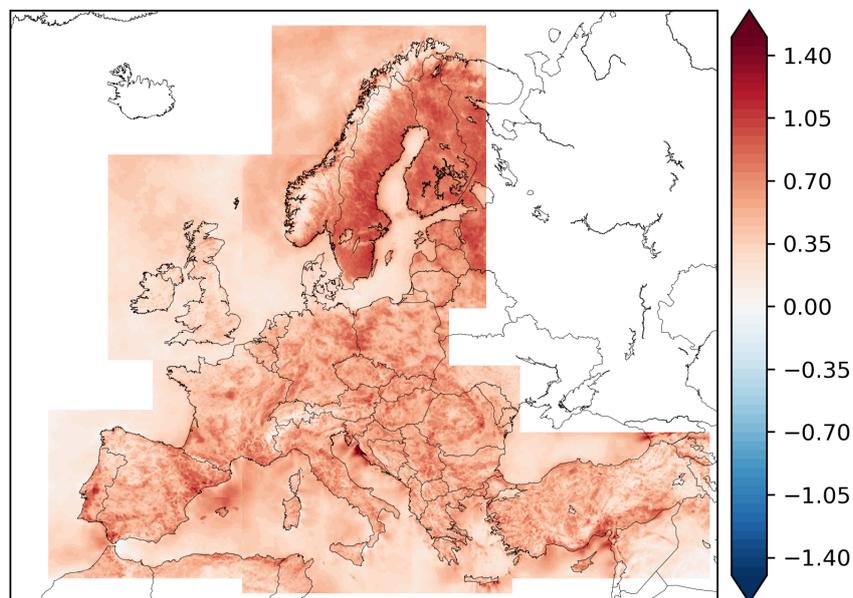


Figure 10: Difference in wind speed annual means, YSU-MM5 minus production run, year 2012 [m s^{-1}] 100m height

As the NW domain and GR domain have different sizes, the average gain in spread over both

of the domains was calculated. The calculation of spread gain against the base run is summarised in Table 2. As there is only one model in the ensemble at this point, the calculation of the gain is equivalent with choosing the ensemble member that has the highest EMD against the base run, averaged over all domain grid points.

member name	average spread gained [m s^{-1}]
YSU-MM5	0.37
NOAHMP-YSU-MM5	0.36
MYJ-MO	0.35
RUC-ACM2-PX	0.33
PX-LSM-ACM2-MM5	0.32
NOAHMP-MYJ-MO	0.32
SLAB-MYJ-MO	0.27
SLAB-YSU-MM5	0.24
SLAB-ACM2-PX	0.21
RUC-VEG	0.21

Table 2: Average spread gain over NW (year 2015) and GR (year 2012) domains versus base run. Only the 10 members with most gain are shown.

Table 2 shows that the second member of the ensemble should be the one with YSU-MM5 parameterisation scheme. The gain from this member is the largest. However, the gain from some other members is not much smaller, e.g., NOAHMP-YSU-MM5 where both the PBL scheme and Land Surface Model is changed.

It is important to note that there are some differences between the NW and GR domains (Table 3). In both domains, members with YSU-MM5 or MYJ-MO PBL scheme seem to provide the most spread. However, in the NW domain YSU-MM5 differs more from the production run, whereas in the GR domain MYJ-MO differs more.

These results could be influenced by the percentage of domain that is sea, as MYJ-MO provides most spread over the water points.

NW		GR	
member name	average spread gained [m s^{-1}]	member name	average spread gained [m s^{-1}]
YSU-MM5	0.33	MYJ-MO	0.44
NOAHMP-YSU-MM5	0.31	NOAHMP-MYJ-MO	0.43
RUC-ACM2-PX	0.27	NOAHMP-YSU-MM5	0.41
MYJ-MO	0.26	YSU-MM5	0.41
RUC-VEG	0.26	PXLSM-ACM2-MM5	0.40

Table 3: Same as in Table 2, but calculated for NW (year 2015) and GR (year 2012) domains separately.

After adding the YSU-MM5 member to the ensemble, let us find out which member should be added to the ensemble as third member. The spread gained, using the methodology described earlier, is summarised in Table 4. Two things should be noted. First, the spread that could be gained by adding a third member is significantly smaller than the spread that was gained by adding a second member to the base run. This can be explained by the fact that the production run has shown one of the lowest values of wind-speeds of all the ensemble members. Therefore, adding a member with higher wind-speeds to a large extent covered the spread seen in the full ensemble.

It should also be noted that although the ensemble member NOAHMP-YSU-MM5 was shown to provide a lot of spread against production run in Table 2, it seems not to provide any additional spread now that YSU-MM5 is included.

member name	average spread gained [m s^{-1}]
MYJ-MO	0.10
NOAHMP-MYJ-MO	0.10
SLAB-MYJ-MO	0.09
OISST	0.08
PXLSM-ACM2-MM5	0.07

Table 4: Same as in 2, but assuming that production run and YSU-MM5 already are in the ensemble. The results are shown only for 5 members with largest gain.

member name	average spread gained [m s^{-1}]
OISST	0.08
SLAB	0.06
PXLSM-ACM2-MM5	0.05
NOAHMP-MYNN-OrigMP	0.05
grid-nudging-D3	0.05

Table 5: Same as in 2 and 4, but assuming that production run, YSU-MM5 and MYJ-MO already are in the ensemble.

After adding the third member (MYJ-MO) it was decided that the ensemble spread gained from adding the fourth ensemble member does not justify the computational expense necessary. Adding OISST would add 0.08 m s^{-1} of spread for every grid point, which is comparable to the uncertainty in the model data associated with the technical implementation of model, such as the HPC system used.

2.2.5 Interpretation of the model spread

As previously described, three ensemble members were calculated. Using a wind speed distribution *cdf* as a basic element of the ensemble has its drawbacks and advantages. The drawback is that expressing that in a single map is not straightforward. The advantage is that more information about the ensemble spread is retained and can be expressed in way that is interesting for the wind atlas users.

To illustrate these properties of the ensemble, the ensemble spread at certain probabilities of the wind speed *cdf* is shown for a single grid point in Fig. 11 (example for a larger ensemble). The spread at 0.1 cumulative probability here is defined as the difference between highest and lowest wind speeds [$w_{0.1}^{max} - w_{0.1}^{min}$] that different ensemble members provide as the possible values for cumulative probability at 0.1.

Similarly, spread at percentiles 0.5 (spread of ensemble medians) and spread at 0.9 percentile is calculated. It is important to stress that each of these three values of spread can be different. Spread at 0.9 percentile can be used to estimate uncertainty associated with extremely high winds, while spread at 0.1 percentile can inform about uncertainty in low-wind speed region.

One could make similar plots but choose to measure the ensemble difference in cumulative probabilities of wind speed exceeding e.g. 20 m s^{-1} . However, that would create problems when comparing different regions, namely, a spread of probabilities for 20 m s^{-1} could be interpreted either as a very high confidence in the model results for this range of wind-speeds or as the fact that the wind speed is rarely in this range, and it would not be possible to separate those two cases.

To aid the illustration of the spread at percentiles, the respective wind speeds for the percentiles from the base run are plotted in Figure 12.

The ensemble spread at different percentiles is shown in Figure 13. It is important to note that the colour scales are different and in general the spread at higher percentiles is larger. However, and interestingly, the spatial distribution of spread at the 0.1 percentile is completely different from the spread at 0.9 percentile, and the differences cannot be solely explained by the differences

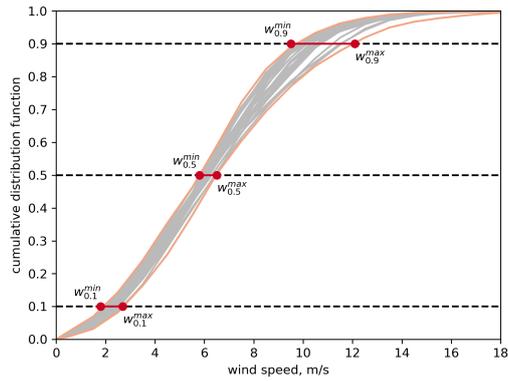


Figure 11: Illustration of quantification of ensemble spread at different probabilities of cumulative distribution function. See text for details. For illustration, an ensemble with more members is used.

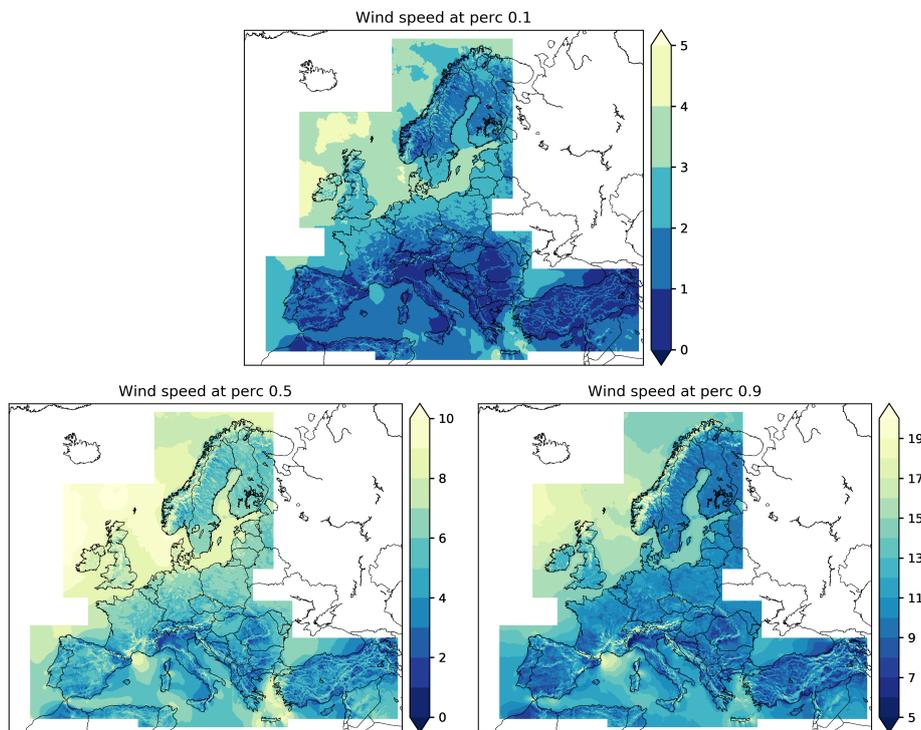


Figure 12: Base run wind speed corresponding to the 0.1 percentile, the median of the distribution and the 0.9 percentile [m s^{-1}] for the year 2012. Please note that colour scales are different.

in the wind-speed itself. For instance, the highest wind-speeds in the whole study region can be seen over the Atlantic ocean, west of the British isles, but the spread in that region is moderate.

2.2.6 Sub-setting the spread into categories

When the results of the project are communicated to the end-users of the atlas, it is important that the results are presented in a way that reflects the uncertainty in the data. Taking into account all the caveats of creating the ensemble, described in Section 2.1, it was decided that the best way how to communicate the ensemble spread is to assign each grid point a category that would reflect the uncertainty.

As the EMD metric forms the basis of the ensemble methodology in this case, and as the envelope of *cdfs* is the measure of the total spread in the ensemble as described in Section 2.2.4,

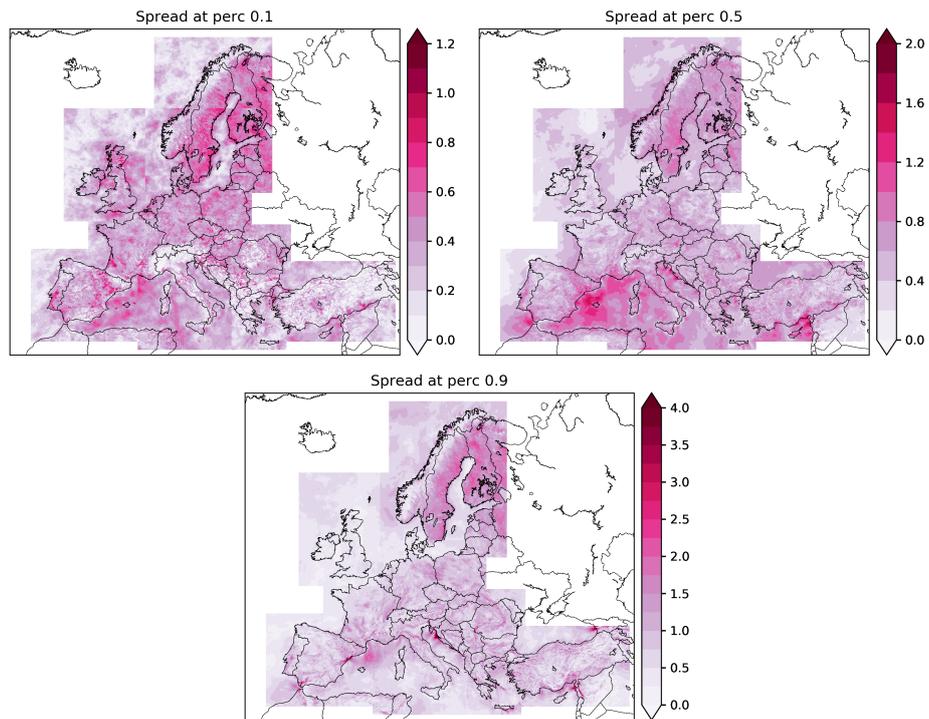


Figure 13: Ensemble spread corresponding to the 0.1 percentile, the spread of the medians and the spread corresponding to 0.9 percentile [m s^{-1}] for the year 2012. Please note that colour scales are different.

then it would be logical to use it as the basis for categorising the grid points. For all domains it is shown in Figure 14.

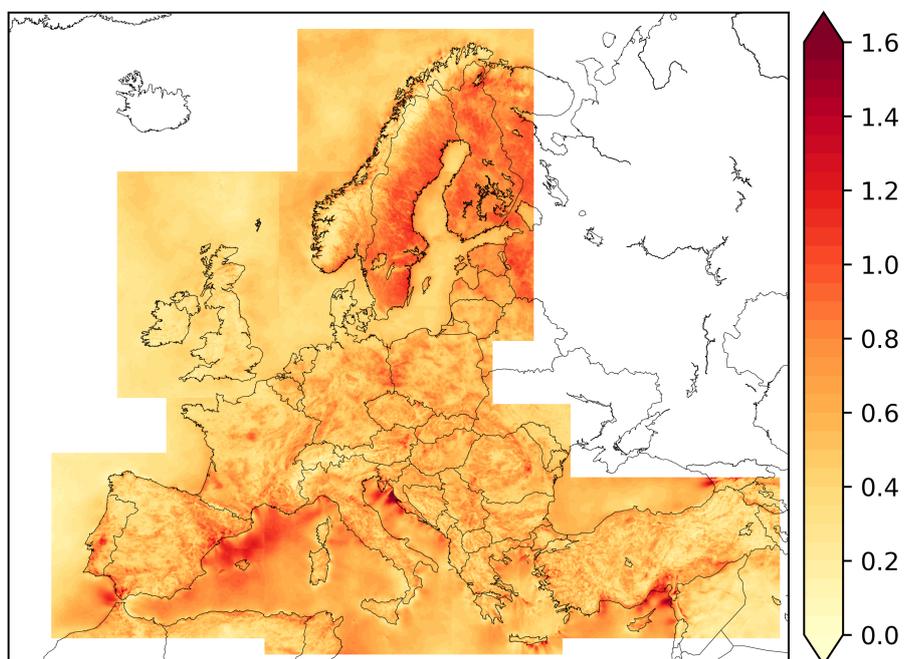


Figure 14: Total ensemble spread for the reduced ensemble. See text for details.

The distribution of total spread values over all the grid points is shown in Figure 15. It shows that the values of the total spread are mostly lower than 1 m s^{-1} and the distribution is reasonably

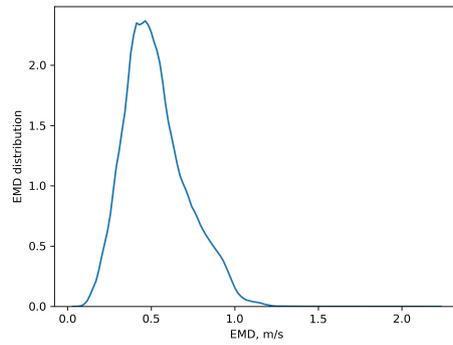


Figure 15: The distribution of total spread values shown in Fig. 14

close to normal with most values grouping around 0.5 m s^{-1} . Therefore, we propose to divide the grid points into five categories from very low uncertainty (total spread less than 0.25 m s^{-1}) to very high uncertainty (total spread more than 1.0 m s^{-1}) as listed in Table 6. The final map with categories assigned to each grid point is shown in Figure 16.

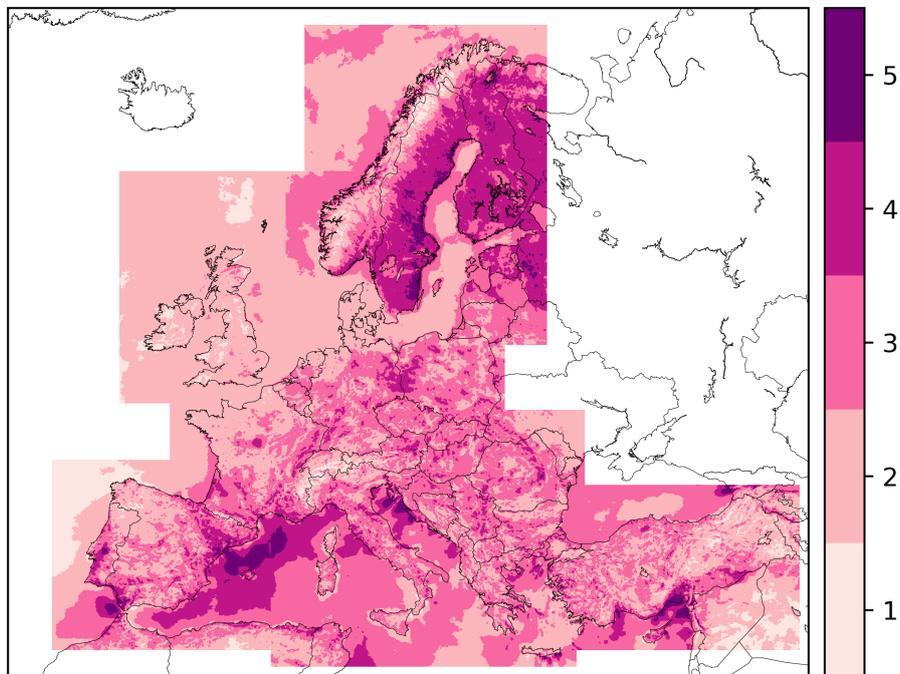


Figure 16: Ensemble spread categories according to Table 6

Category	Total spread [m s^{-1}]
1 Very low uncertainty	less than 0.25
2 Low uncertainty	0.25 ... 0.5
3 Medium uncertainty	0.5 ... 0.75
4 High uncertainty	0.75 ... 1.0
5 Very high uncertainty	more than 1.0

Table 6: Total spread values used for assigning the uncertainty categories at each grid point

2.3 Ensemble spread: a regional perspective

The selection of an optimal model configuration implies examining multiple choices with different parameterisation schemes or a variety of model setup options. Even if a preferred configuration emerges from such an analysis, the optimal setup might be domain, simulated period or time-scale dependent. Therefore little can be said about what would be a potentially best suited setup in the case of regional simulations covering long periods or very large areas as is the case of the European domain within the NEWA experiments.

Nonetheless, an analysis of how much the simulations resulting from a variety of configuration experiments depart from each other is still interesting to address the question how robust is the model to changes in its configuration or how much a specific variable, such as the wind field, depends on certain aspects of the model setup.

The analysis of differences among a pool of simulations that target the most realistic representation of a specific climatic variable by the regional model can provide as well a measure of the uncertainty associated with the methodological variance. It is in this sense in which the term spread is used within this document. Therefore, evaluating the regional spread given by a set of simulations with different setup choices provides a frame for the quantification of the wind atlas uncertainty.

However, comparing multiple simulations together and providing a measure of such an uncertainty in a spatio-temporal frame is not a straight-forward issue. Thus, this section is devoted to introduce a methodology that involves an estimation of how much the different simulations tested depart from each other and thus, an analysis of how robust is the model to changes in its setup from a spatial as well as time-dependent perspective. The latter is accomplished by identifying subregions with homogeneous behaviour of the wind. The first part of this section aims at describing the methodology for the regionalisation of the wind field over a region of complex terrain as well as the collection of WRF simulations performed in this part of the analysis. This section is focused on the sensitivity experiments carried out for the Iberian Peninsula (IB) sub-domain (see domains in Figure 17 as defined for the NEWA sensitivity experiments). The second part of the section illustrates the procedure, shows the regions found and evaluates the ensemble regional spread over the IB domain.

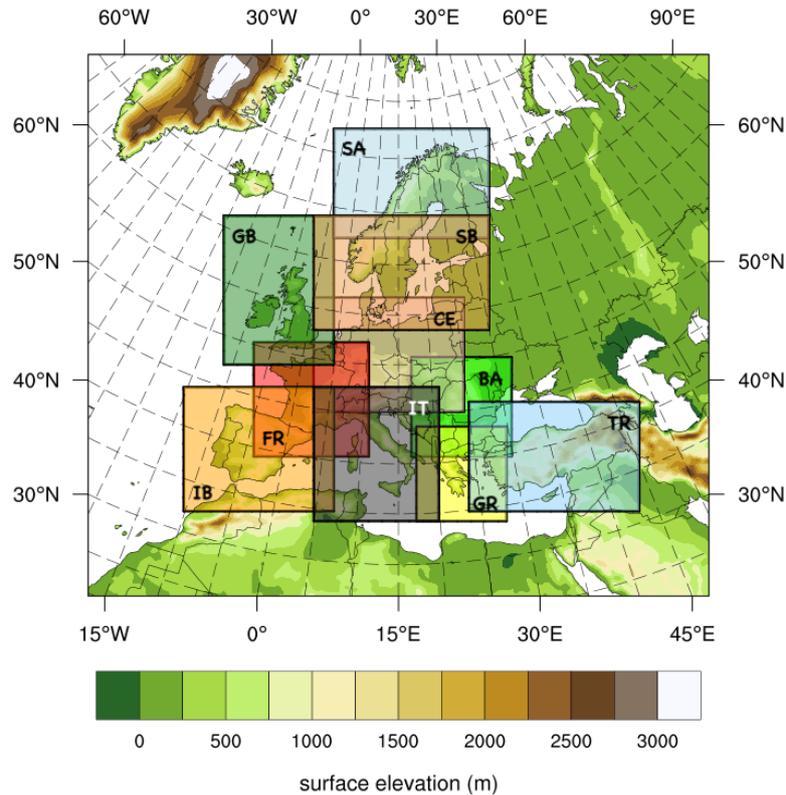


Figure 17: WRF model sub-domains used in NEWA mesoscale sensitivity experiments.

2.3.1 Data and Methodology

Thirteen high resolution simulations over the IB domain with alternative physical parameterisations and setup choices were used to investigate the model spread. Two discontinuous years (2011 and 2015) were simulated using the version 3.6.1 of the WRF model (Skamarock et al., 2005) with a configuration of three one-way nested domains increasing the resolution progressively up to 3 x 3 km horizontal resolution in the innermost domain. Boundary and initial conditions were provided by ERA-Interim v 2.0 (Dee et al., 2011), the land use data is CORINE 100 m update 2012 (Copernicus Land Monitoring Service, 2019) and the SST is provided by OSTIA (Donlon et al., 2012). More details about the common setup of all thirteen simulations can be found in Table 7. The options involve two Planetary Boundary Layer (PBL) parameterisations, *i.e.*, the YSU (Hu et al., 2013) and MYNN (Nakanishi and Niino, 2006) schemes, four variants for the Land Surface Model (LSM, namely NOAH, RUC, NOAH-MP and CLM; Chen et al., 1996; Benjamin et al., 2004; Niu et al., 2011; Bonan et al., 2002, respectively) and three alternate integration methods including daily initialisation with no spectral nudging (S1), weekly initialisation with spectral nudging only in the external domain (W1), weekly initialisation with spectral nudging in all three domains (W3). In the W1 and W3 cases, initialised at 12:00 GMT, the first 24 hours were disregarded. Additionally, an experiment with 91 vertical levels instead of the usual 61 was performed.

Experiment	PBL scheme	LSM	Initialisation	Nudging	Vertical levels
NH0M20W3	MYNN	NOAH	8 days	spectral nudging D1–D3	61
NH0M20W1	MYNN	NOAH	8 days	spectral nudging D1	61
RC0M20W1	MYNN	RUC	8 days	spectral nudging D1	61
NP0M20W1	MYNN	NOAH-MP	8 days	spectral nudging D1	61
NP1M20W1	MYNN	NOAH-MP	8 days	spectral nudging D1	61
CLMM20W1	MYNN	CLM	8 days	spectral nudging D1	61
NHYSU0W1	YSU	NOAH	8 days	spectral nudging D1	61
NHYSU1W1	YSU	NOAH	8 days	spectral nudging D1	61
NHYSU0W3	YSU	NOAH	8 days	spectral nudging D1–D3	61
NHYSU1W3	YSU	NOAH	8 days	spectral nudging D1–D3	61
NHYSU0S0	YSU	NOAH	36 hours	no nudging	61
NHYSU1S0	YSU	NOAH	36 hours	no nudging	61
M20W1L91	MYNN	NOAH	8 days	spectral nudging D1	91

Table 7: WRF model setup common to all thirteen simulations participating in the regional spread analysis. The 1st column indicates the acronym of the respective experiment, the 2nd column refers to the PBL scheme used, the 3rd column specifies the LSM used, the 4th column represents the length of the initialisation period while the 5th column indicates if spectral nudging is applied and what domains affected, finally the last column accounts for the number of vertical levels. The *reference configuration* is emphasised in bold.

The rest of configuration options are shared among the different simulations and they are enumerated in Table 8. The simulation identified as NH0M20W1 (bold in Table 7) is designated as the *reference configuration* and will serve as the baseline with which the rest of simulations are compared in the analyses.

Longwave Radiation	RRTMG scheme (Iacono et al., 2008)
Microphysics	WRF Single-Moment 5-class scheme (Hong et al., 2004)
Shortwave Radiation	RRTMG shortwave (Iacono et al., 2008)
Cumulus Parameterisation	Kain-Fritsch scheme (Kain, 2004) on D1 and D2
Diffusion	Simple diffusion 2D deformation 6th order positive definite numerical diffusion rates of 0.06, 0.08, and 0.1 for D1, D2, and D3 vertical damping.
Advection	Positive definite advection of moisture and scalars.
Grid relaxation zone	5 points
Nudging	Spectral nudging (Miguez-Macho et al., 2004) nudging coefficients: 0.0003 s^{-1} nudging above the PBL (u and v) and above level 20 wave numbers: 15 (x) and 11 (y) in D1, variable in D2 and D3.

Table 8: Configuration of the WRF model common to all simulations

In order to provide a compressed view of the model spread at the regional scale that accounts for its temporal as well as its spatial variability we follow a regionalisation strategy, searching for domains that evidence shared wind variability and we explore the spatial patterns found as well as the representative series. To this aim we first apply an Empirical Orthogonal Function (EOF) analysis (von Storch and Zwiers, 1999). In climate as well as in other disciplines, the EOF analysis is applied to obtain models or patterns of variability. Their variation in time is provided by the Principal Components (PCs). The technique consists in partitioning the original field into orthogonal or independent modes of variability. They are obtained by calculating the eigenvalues and eigenvectors from the covariance (or correlation) matrix of the original field. The eigenvalues imply an estimation of the amount of explained variance by each mode. These patterns are usually

ordered according to the amount of variance they retain from the original field. As a result, the EOFs and PCs provide a condensed representation of the variations of the field by making use of a reduced number of main patterns or modes of circulation (von Storch and Zwiers, 1999). By definition, EOF modes and PC series are uncorrelated. The orthogonality condition may impose however some constraints to the physical interpretation of the patterns. These patterns can be as well dependent upon the domain selected for the analysis, therefore the appearance of a certain mode of variation could be an artefact due to the setup of the experiment.

In order to avoid such limitations in the classical EOF analysis, the Rotated EOFs (REOFs) can be calculated. The REOF approach provides more regional or localised patterns or structures at the expense of compromising for instance the orthogonality (Hannachi et al., 2007). In such a way, patterns obtained from REOF are simpler and more easily interpreted from a physical perspective. The Varimax rotation technique is frequently used in this context (Hannachi et al., 2006). The rotation is a readjustment that consists in maximising the variance of those modes with larger loads and identically, minimising the variance of the smaller loads. This procedure simplifies the structures and therefore the physical interpretation of the modes found. A particularity of this regionalisation method is that it preserves the orthogonality of the eigenvectors (PCs) but not of the eigenvalues (EOFs) after rotation. Also, compared to other methodologies, such as cluster analysis, it may be the case that a certain area or site pertains to various subregions since the regions obtained from rotated PCA (RPCA) modes can overlap to some extent. Nevertheless, this can easily be interpreted from the perspective of transition zones in the borderline between regions.

Therefore, a standard EOF analysis is performed over the wind field anomalies (annual cycle removed) and a set of EOFs is retained in the first place. The Varimax rotation is applied then to this subset of EOFs. However the rotation criterion is somehow subjective and different options are available in the literature (Jolliffe, 2011). The aim of Varimax rotation is to maximise the variance and this is usually performed by increasing the squared correlation of factors with larger loads and decreasing the correlation of those with generally smaller loads. It cannot be ruled out that REOFs are still domain dependent to a certain extent. Also the number of initially retained EOFs is to some degree subjective.

Despite method limitations, it still is a useful tool to understand the structures of auto-covariance of a certain climatic field. To our knowledge its use in the frame of sensitivity analysis is not very extended so far in the literature although it has been applied in studies that attempted a regionalisation of different climate fields to identify common patterns of variance (White et al., 1991; Mearns et al., 2003; Argüeso et al., 2011). Few references to the case of the wind field can be found in the scientific literature (Burlando et al., 2008; Jiménez et al., 2010a; Lorente-Plazas et al., 2015). It allows to synthesise the information provided by multiple experiments providing at a time an idea of the spatio-temporal methodological variance. Therefore, to summarise the model spread over the IB domain in time and space the process involves the projection of the wind field from the different sensitivity WRF simulations onto the EOFs and the Varimax REOFs. The latter would allow to identify regions with shared temporal variability and therefore we would be able to characterise the model spread over the regions with homogeneous wind behaviour. Such an approach would contribute to a more fair comparison avoiding that differences among the various simulations are related to large differences in the wind variability across the regions.

As commented above, the aim of the regionalisation technique based on the rotation of the selected principal modes is to obtain wind regions in a much simpler structure than that provided by the EOFs. In such a structure the variables are as close as possible to a hyperplane of at least one principal mode (Richman, 1986). Ideally, each observational site or model grid point should obtain a high load in just one rotated loading map and null loads in the rest (perfect simple structure). However, in reality, loads are not null which makes it necessary to define a threshold value to help define the subregions, that is, only locations with loads higher than the critical value belong to a specific subregion. In our case the critical value is defined based on the value of the wind velocity. The critical value decided defines an isoline that separates one region from the others. The threshold is decided by analysing the graphical representation of the REOFs. Also the correlations of the representative wind time series for each region (the mean

wind calculated over all grid points within the specific region) with the original wind time series were used to help decide the threshold value. The critical value is the response to a balance decision between not allowing too much overlapping among the different resulting regions and the resemblance of the series with the original data (correlations). Here the critical value was 0.0125 before renormalising the loads with the variance of the original wind field. Finally this value determined the creation of six subregions.

2.3.2 Regionalisation and model sensitivity

The method involves as a first step the calculation of the EOFs from the simulated zonal and meridional wind components EOFs from the *reference simulation*. As commented, the number of EOFs to be kept for the rotation is to some extent subjective. In our case, six EOFs were initially retained from the original simulated fields. They accounted for an 83% of explained variance of the original wind components. Figure 18 shows the first two EOFs as well as the corresponding PCs. The first EOF (45% explained variance) represents a dominant NW-SE wind pattern with stronger wind intensities along the Ebro valley that channels circulations in this area generating strong and cold winds over the region known as *Cierzo* winds. The Cierzo wind is typically more intense during the cold months, from October to March, as is also illustrated by the first PC, that shows large scores during this period of the year, both for 2011 and 2015. A noticeable decay, however, of the strength of the pattern identified by the first EOF is visible in December 2011 and November 2015. The second EOF (25%) shows a more zonal wind circulation over the area evidencing the role of the topography in modulating the wind behaviour. The associated PC series reflects somehow larger levels of variance during 2015 compared to 2011, illustrating therefore the natural variability of the wind field year by year. EOF3 and EOF4 (not shown) are as well strongly influenced by the terrain. Therefore they present much more local detail of the wind field variability.

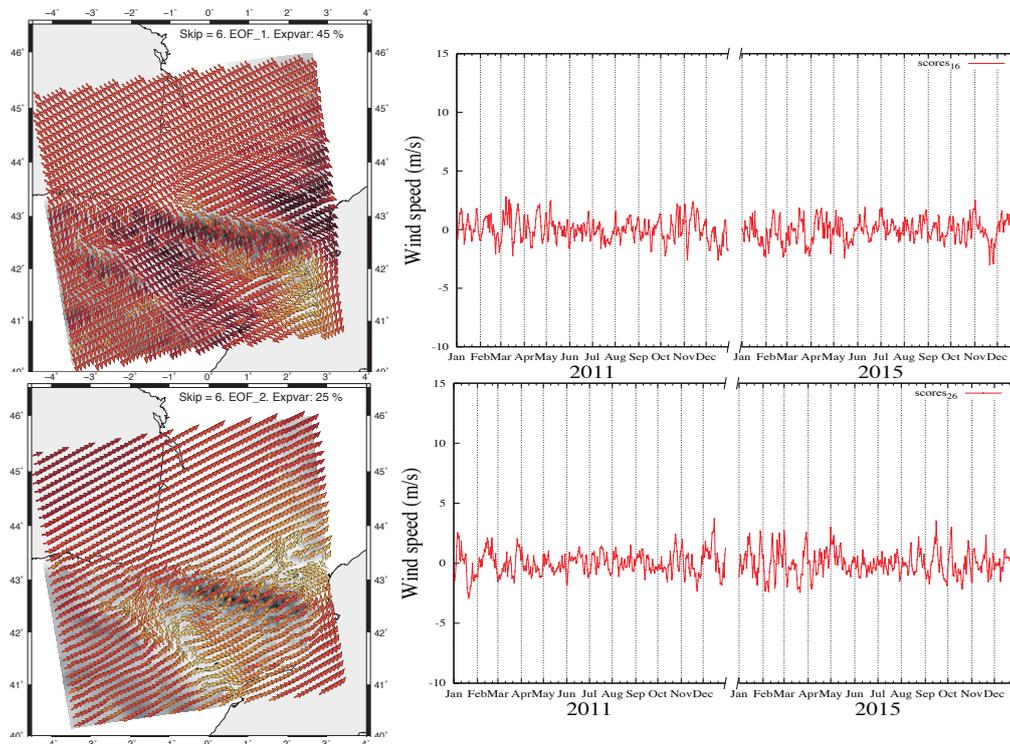


Figure 18: Two first EOFs (left) and the corresponding PCs (right) calculated for the wind components of the *reference simulation* for the two years simulated, 2011 and 2015. The colour of the arrows is indicative of the wind velocity at each grid point.

After calculating the EOFs we applied a Varimax rotation to the first six EOFs to obtain the rotated EOFs or REOFs. We used the first four REOFs that are no longer uncorrelated and they account for more than 100% of the retained first six EOFs variance. Out of them we represent the first two REOFs in Figure 19 with their associated PCs. The first REOF (Figure 19 left) evidences further regional detail of the wind circulation comparatively to the first EOF mode represented in Figure 18. It can be appreciated the characteristic anticyclonic circulation over the Gulf of Biscay that is deflected to overcome the natural obstacle imposed by the large mountain range and then is channelled along the Ebro Valley.

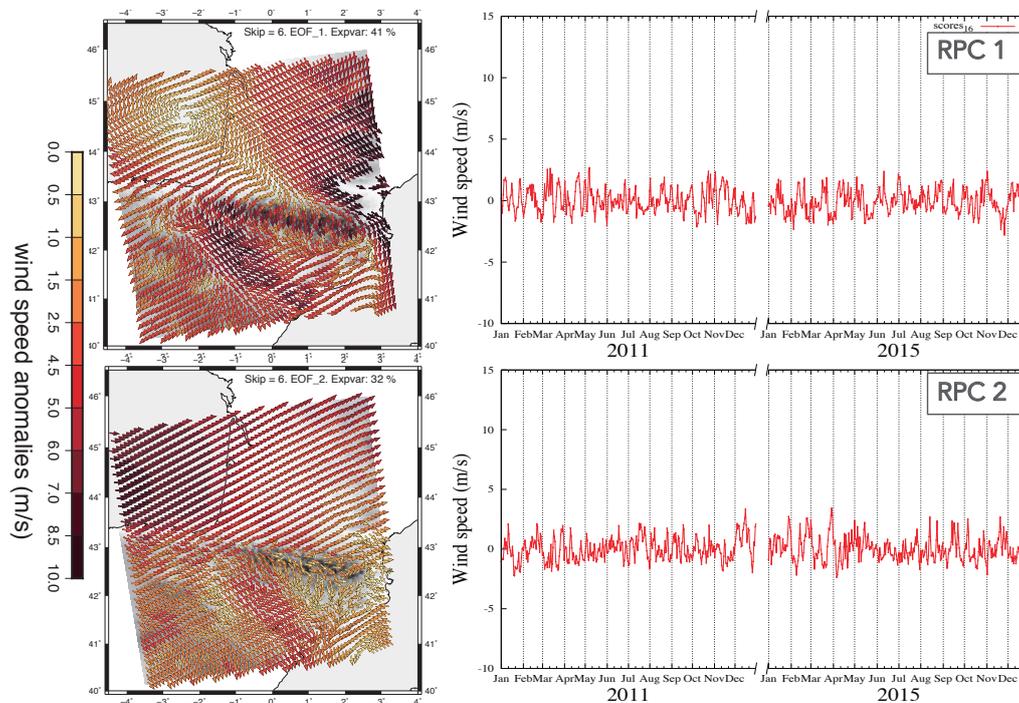


Figure 19: As in Figure 18 but for the first two rotated EOFs (REOFs, left) and corresponding PCs (right).

The second REOF shows two clear modes of wind circulation. Over the northeastern part of the region under study the flow has a preferred westward zonal orientation while to the south on the mountain ranges, the wind is again more meridionally oriented. Therefore it is apparent how the rotation has helped identifying the dominant modes of circulation over the region comparatively to the case where only the EOFs are calculated (Figure 18). The RPC1 and RPC2 (Figure 19, right) are nevertheless to a great extent similar to the corresponding original PCs in Figure 18.

We represent the renormalised RPCs using the variance of the original wind field, so that the time series have now physical units in Figure 20 (left) for all thirteen different configurations of the WRF model that take part in this part of the experiment.

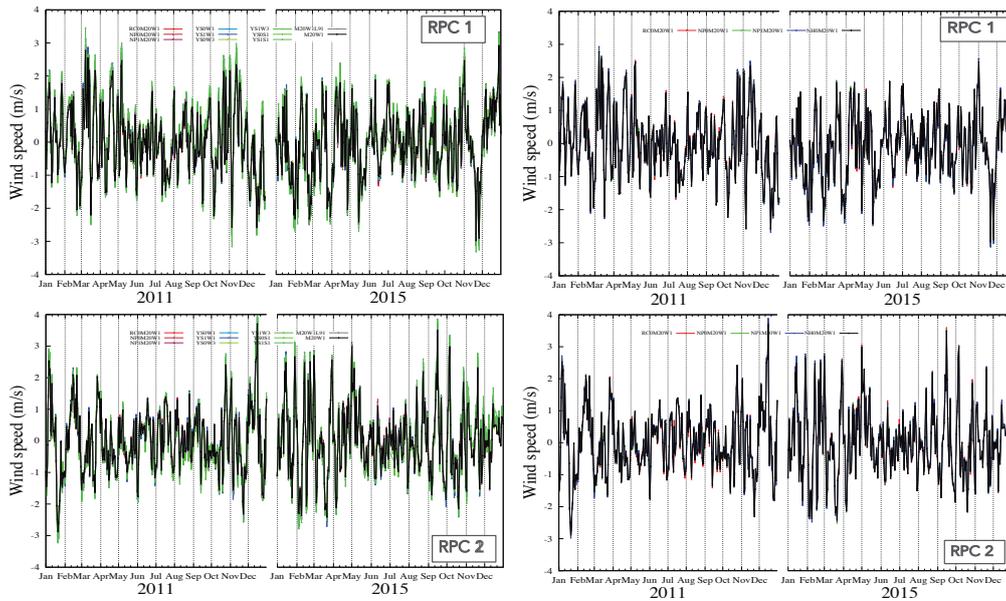


Figure 20: Renormalised RPC1 and RPC2 of all thirteen simulations (left) and only for those cases with different LSMs (right).

All time series evidence clear common variability throughout the two years simulated, 2011 and 2015, especially for the case of the RPC1, where it can be noticed only slight differences in the amplitude of changes, being these to some extent larger for those simulations that make use of the YSU PBL parameterisation compared to the rest. In the case of RPC2 slightly larger differences among the simulations can be detected, in particular towards the end of both years 2011 and 2015. It is reasonable that, as the explained variance decreases, the differences among the series become more visible since the behaviour of lower rank RPCs becomes progressively less dominant (the smaller the explained variance, the less relevant the mode).

If we focus only on those simulations that differ from each other in the LSM parameterisation used (Figure 20, right) we observe that, even though some differences are present, especially in the case of the RUC LSM compared to the rest, the series are very similar to each other in these timescales. We only show the first two RPCs for the sake of brevity. Similar conclusions can be achieved based on the rest of RPCs. The fact that simulations based on the use of different LSMs do not depart largely from each other is not surprising and can be related, among perhaps other causes, to the strategy of simulation that reset all initial conditions every few days not allowing thus the memory of longer lifetime processes within the subsurface to develop and hampering therefore the realism of the land-surface interactions within the simulations. Indeed, if differences between LSMs are designed to reflect the effect of alternate physics affecting the thermodynamical processes taking place in the subsurface, the circumstance of resetting the simulation every few days will not allow these differences to be expressed in the simulations. A meaningful experiment would be to test the impact of making use of different LSMs in a configuration where simulations are performed in a "continuous" way, without initialising the conditions every few days and allowing the full memory of the soil to develop. This type of experiment will be accomplished somewhere else.

As commented in Section 2.3.1, the model sensitivity or model spread, as a measure of the uncertainty associated to the change of physical parameterisations and model configuration can be analysed based on segregating those regions with a homogeneous behaviour of the wind field. In such a way, differences among simulations (or with observations) are not due to the particular variability of the wind field, that changes from one subregion to another, but they depend specifically upon the different configuration schemes used. We have therefore divided the area under study into subregions with analogous temporal wind components variability. The latter is illustrated in Figure 21 left (subregions in colours) based on the technique of the Varimax ro-

tation. The representative wind time series at each subregion (mean zonal and meridional wind components) calculated as the projection of the original wind field at each region onto the corresponding subregion mode have also been represented in Figure 21, right. The top panel shows the representative 31-years running mean wind of the region marked with black dots for all thirteen simulations performed within our sensitivity experiment at the regional scale over the IB domain, while the middle (bottom) panel wind series corresponds to the region assigned with red (green) dots in the left panel.

Although the subregions are clearly segregated as can be noticed in the left panel of Figure 21, still the variability described by the representative wind series is to some extent comparable among the different subregions. The zonal wind component evidences larger amplitude of changes during the first of the two simulated years (2011) with a noticeable enhanced variability by the end of the year and also a clear medium-term trend towards negative values during the second year (2015), which is much more pronounced over the region of the Ebro Valley (green subregion). The meridional component seems to have less variability compared to the zonal one although this is an effect of the axis scales (note the larger range of variations in the case of the right axis for the meridional component) and also evidences an opposition-in-phase sequence of changes compared to the zonal wind variations that can be related to the conservation of momentum (when one component increases the other one decreases in favour of the first). The combination of positive meridional and negative zonal trends throughout 2015 indicates a strength of the northwest-southeast circulation over the region, enhancing the classical flow over the region that tends to channel strong winds along the Ebro Valley in this area.

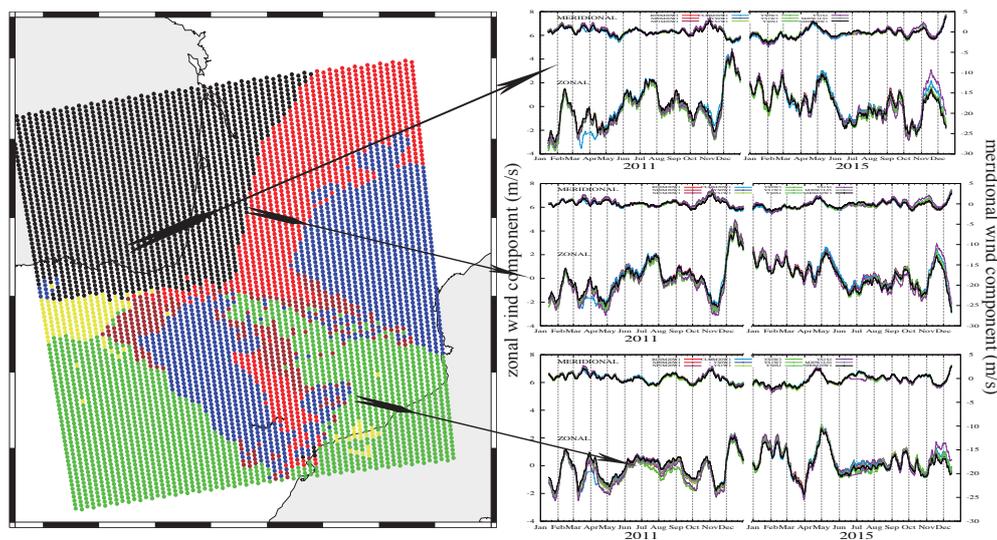


Figure 21: Spread of all thirteen simulations according to the different regions identified. Left: regions classified through the regionalisation technique applied (see Section 2.3.1). Right: Mean zonal and meridional wind components calculated over all model grid points belonging to each region (top belongs to the region identified with black dots and middle (bottom) panel represents the mean wind over the red (green) dots region). Series are anomalies with respect to the annual cycle and 31-year moving average filter outputs.

The regional spread over the IB region is characterised by large similarities among the different simulations within the sensitivity experiment, therefore it can be said that a limited model spread is detected over the region, indicating that most of the simulations share analogue variability over the region. The larger differences among simulations can be noticed during those periods with larger variance or larger amplitudes of change. Also the simulations based on the CLM and the Noah-multiphysics LSMs seems to present larger differences relative to the pool of simulations. Whether this is an effect of the much more complex physics and the added realism of these models with respect to the others would need further investigation.

In general, it can be said that a sound strategy to characterise the model sensitivity or model spread based on the segregation of the domain under study into subregions with shared variability is a powerful and straightforward tool that allows an evaluation of the model spread from a temporal variation as well as spatial variability perspective. The application of this strategy has shown that no large differences can be found among the set of simulations planned in the sensitivity analysis, although, at least in the case of the alternative LSMs explored, it cannot be ruled out that these differences are not fully expressed due to the running strategy used usually within the community and herein, which implies the reset of the initial conditions every few days. The latter calls for further experiments with an alternate 'continuous' running strategy that allows the soil conditions to develop full memory.

A question that holds is how the model spread compares to the model validation with observations. Whether the differences among the pool of simulations is larger/smaller than those between simulation and observations is addressed in Section 3.4 of this document.

3 Model evaluation

Uncertainty quantification may include observational errors in areas where observations are available. Differences between model output and observations depend then on the model setup, discussed in Section 2, and model errors. Observational uncertainties are also a reality and have a role in this context; however, they are not considered herein except for the application of dataset specific quality control strategies that try to reduce them.

The results of comparing model and observations can vary in space and time. They may depend on the specific location being addressed with some model setups being more adequate for some sites or regions and others performing better elsewhere. They also may depend on time with some model configurations performing better in some specific season or synoptic configuration. Also, results may vary according with the type of variable or dataset being analysed; e. g. one model setup may perform better for wind tower data and others may produce a more realistic output in comparison with wind profile data or in some specific wind farms. Case examples of such situations will be unfolded through this section. Therefore, it is arguable that there is no model setup that can be defined as universally valid or most valid for all instances and cases and it is often the case that some model setups can be more appropriate to simulate the physics over certain areas, seasons and variables of interest (Jiménez et al., 2011; Jerez et al., 2010; Kotlarski et al., 2014). From this perspective it is meaningless to render one specific model configuration as valid or invalid as different setups can be regarded as invalid or valid depending on their performance when reproducing different variables of interest at different sites and time intervals (Oreskes et al., 1994; Oreskes, 1998; von Storch, 2010). The question addressed in this report is how model performance can be characterised with the data at hand and whether decisions regarding selection of a given model setup for a production run can be taken on the basis of model performance in a variety of situations, using different variables and datasets as observational targets. We therefore move from the concept of model validation to that of model evaluation, following Flato et al. (2013), as a procedure to assess the performance of different model setups at different instances with the purpose of selecting the most appropriate model configurations for the purposes of NEWA.

Section 3.1 targets model assessment at relevant sites for wind energy production. Sections 3.2 and 3.3 address the simulation of wind at height by comparing simulations with tall mast and wind profile data, the latter mostly focussed on offshore conditions. Section 3.4 evaluates model performance using a combination of wind tower and mostly wind surface data. The section is closed with two model-data comparison sections addressing different types of existing global products: satellite data (Section 3.5); and Reanalysis and Global Wind Atlas products (Section 3.6).

3.1 Model evaluation using Vestas wind tower data

The NEWA model-chain was validated using measurements from 291 tall masts located all over Europe. The measurements were made available by Vestas under a non-disclosure agreement, requiring anonymisation of both the raw measurements and the mast meta-data. The following validation thus focuses on aggregate statistics, rather than the details of individual sites.

3.1.1 Measurements

The masts are distributed across Europe, including Turkey. Figure 22 shows the number of masts located in each of the largest European countries. Parts of Europe, including Poland, Great Britain, France, Italy, and Turkey, are well populated by masts, while other countries such as Spain, Germany and several other central European countries are only sparsely populated. Fortunately, the masts are not concentrated in one region of Europe but cover most edges and the

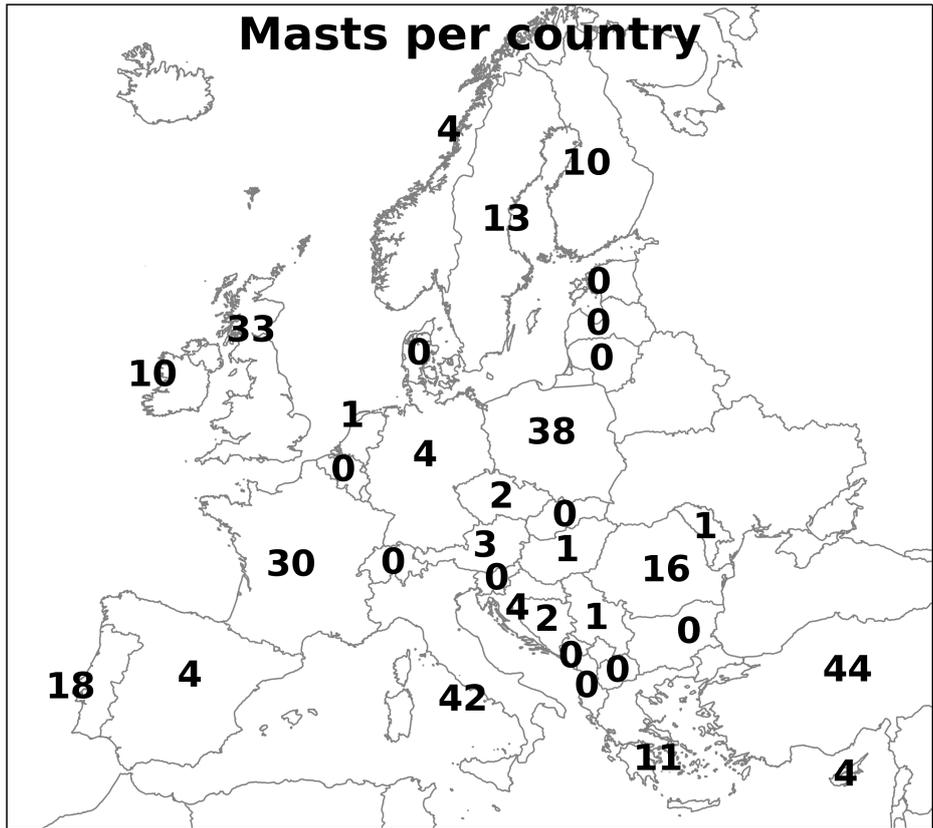


Figure 22: The number of masts used for the validation located in each of the largest European Countries. Several smaller countries contain no masts and are omitted for readability.

centre of the continent.

Most of the measurement campaigns from which the data comes covered periods of 1–2 years. Therefore, it was decided to use only one year of continuous measurements, but with the requirement that at least 80% data availability per month was fulfilled for each mast. Using one year of measurements of high availability ensures a minimal influence of seasonal biases, and keeping the period duration the same for all masts ensures that emergent differences in modelling errors between masts are not due to the measurement duration.

Only one set of measurements (wind speed and direction) was used from each of the masts, even if measurements were available from many heights. The data from different heights were ranked according to their height above the ground and data-availability. Measurements higher up and with more available data were prioritised. Lower level measurements were only considered if no period of 80% monthly data-availability was available for the top measurements.

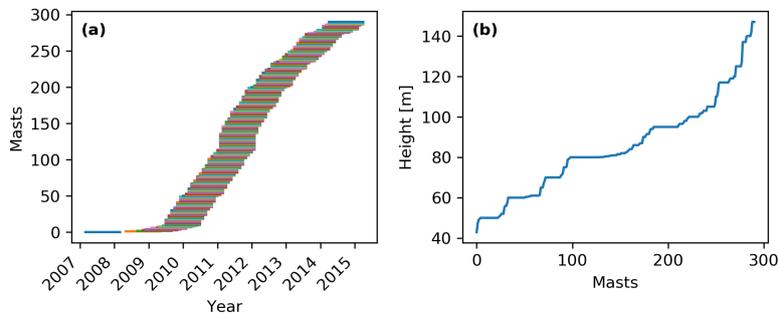


Figure 23: The distribution of measurement periods (a) and wind speed measurement heights (b) for the 291 masts.

The measurements were quality controlled by the data provider, so no rigorous quality-control was made in addition for this study. Only a light screening of the data was performed to catch obvious measurement errors, artefacts, icing, or nonphysical signals. No such obvious problems were detected.

Only measurements taken at 40 m above the ground or higher were used in the validation. This decision was made to avoid the larger uncertainties generally associated with measurements taken closer to the surface, and to ensure that measurements were representing turbine hub-height level winds as best as possible. Figure 23 shows the distributions of periods and heights of the measurements. Most of the measurements were taken between 2010 and 2015. The heights shown in the figure correspond to the height of the wind speed measurements (cup or sonic anemometers). The wind directions were obtained from either the sonic anemometers or from the nearest wind vane, typically located 0 – 40 m below the wind speed measurements. The measurement heights range from 40 to \approx 150 m, but the bulk are located between 60 and 120 m above ground.

3.1.2 Models

Each model in the NEWA model-chain: ERA5, WRF, and WRF-WAsP was validated to show the added value of each component.

ERA5 and WRF

The ERA5 Reanalysis dataset and the NEWA WRF 3km (d03) dataset were used in the validation. The ERA5 Reanalysis data used for the validation were the 10 m and 100 m single level velocity components (U and V) and pressure level velocity components available from . 10 and 100 m velocity components were interpolated to 25, 50, and 75 m by shear extrapolation, and the single level velocity components at 100 m and the pressure level velocity components were linearly interpolated to 150, 200 and 250 m.

The WRF model output was post-processed before the validation. Only the velocity components at 50, 75, 100, 150, 200, 250 and 500 m height above ground were kept.

Time-series of wind speed and direction were extracted from both the ERA5 Reanalysis and the NEWA WRF datasets. The data was extracted at each mast location at the height of the wind speed measurements. The time-series were extracted at the available frequencies, 1 hour for ERA5 and 30 minutes for WRF, and then interpolated in time to a frequency of 10 minutes corresponding to the frequency of the measurements. Linear interpolation in the wind velocity components was used for both the spatial and temporal interpolation.

WRF-WAsP

To obtain the final high-resolution wind climate estimates, the WRF-WAsP downscaling methodology (Hahmann et al., 2019, 2014; Badger et al., 2014) was used. It is a statistical downscaling methodology, meaning that it works as a correction to the multivariate wind climate probability density function (typically a histogram of wind speeds and directions) obtained from long-term WRF simulations. The assumption of the methodology is that local microscale effects at a given site, be it orographic, internal boundary layer, or effective upstream roughness effects, are under-resolved in the coarse mesoscale model due to the finite effective resolution of the model. In contrast, state-of-the-art high-resolution maps of elevation and land-use, like SRTMV3 and CORINE can be used by WAsP to estimate the linearised microscale effects at fine resolution. Thus, to downscale the mesoscale wind climates the wind climates are first adjusted according to the linearised microscale effects resolved by the mesoscale model, corresponding to the removal of under-resolved microscale effects. Following this, the wind climates are adjusted again according to the linearised microscale scale effects derived by WAsP from the high-resolution maps, corresponding to enrichment of the wind climates with high-resolution microscale effects.

Within the WRF-WAsP framework Hahmann et al. (2019) outlines two methods of downscaling: 1) generalisation and downscaling, and 2) direct downscaling.

In method 1) an intermediate wind climate, the so-called *generalised* wind climate, is calculated by applying the mesoscale site corrections and then fitted to Weibull distributions by sector and corrected for atmospheric stability effects. Finally, the generalised wind climate is down-scaled by applying the microscale site corrections to the Weibull distributions.

In method 2) no intermediate generalised wind climate, no Weibull fit, and no stability correction is made, instead both the mesoscale site corrections and the microscale site corrections are applied to the wind climate, resulting in a new corrected histogram instead of corrected Weibull parameters.

The long-term NEWA microscale wind atlas uses method 1. However, since errors resulting from fitting the histograms to a Weibull distribution are larger for shorter sampling periods, it was decided to use method 2 for the present validation.

In the present validation study the microscale correction factors of the WRF model were calculated in each grid-cell centre at predefined heights: 50, 75, 100, 150 and 200 m. The correction factors cannot be linearly combined, so no spatial interpolation to the mast locations was done. Instead, the nearest grid cell and the nearest height (from the list of heights above) was used as the starting point for the downscaling. The downscaling includes vertical extrapolation to the measurement height using the geostrophic drag-law and the effective surface roughnesses in the mesoscale model to estimate the geostrophic wind, and then in WAsP to estimate the wind at the measurement height from the estimated geostrophic wind.

The 3'' (≈ 90 m) Shuttle Radar Topography Mission Global Coverage Version 3 (SRTMV3) was used in WAsP as surface elevation to estimate the high-resolution microscale effects used for downscaling. For surface roughnesses the 100 m CORINE land cover (CLC) was used, by converting each land-use class to a constant roughness using a conversion table.

No objective or thoroughly validated roughness conversion table for CORINE land-use classes exists. Silva et al. (2007) suggested roughnesses for each class, but the values were only validated at a limited number of sites. Another set of roughness values has been estimated at DTU Wind Energy¹, which has similarly not been comprehensively validated. Nevertheless, the latter table was used as the starting point for the validation and is referred to in the following as roughness conversion table A (RCT-A). The values of RCT-A are shown in table 9. Some roughness values in RCT-A, like those for *Non-irrigated arable* (0.05 m) and *Pastures* (0.03 m), which respectively constitute approx. 1/4 and 1/13 of the land near the validation sites (on average), are lower than in e.g. Silva et al. (2007) and lower than the equivalent values used inside the NEWA-configured WRF model (table 9 column " z_0 WRF"). Due to this, and due to the general uncertainties associated with the WRF-WAsP methodology, and further motivated by the initial validation results in this study using RCT-A, a second conversion table B (RCT-B) was created (Table 9; z_0 B) that have higher roughnesses for *Non-irrigated arable* and *Pastures*, and some of the other land-use classes typically associated with low surface roughness. Validation results for the two tables are compared in Section 3.1.4.

3.1.3 Site characterisation metrics

A number of descriptive metrics were calculated for each validation site to characterise important attributes of the site: the orographic complexity of the sites was characterised by the Ruggedness Index (RIX), the roughness complexity of the sites was characterised by a Roughness Complexity Index (RCI) discussed below, and the distance to the nearest coastline was estimated to capture coastal effects.

The Ruggedness Index (RIX) (Mortensen et al., 2008) characterises the complexity of the orography at each site. It is defined as the fraction of the terrain within a given radius that slopes more than a certain threshold slope, in the present study this threshold was the default WAsP threshold: radius = 3500 m and critical slope = 0.3 (16.7°). RIX is used as a crude estimate of

¹Pers. comm. Rogier Floors, Niels Gylling Mortensen and Andrea N. Hahmann, DTU Wind Energy, March 2019.

Code	Category	Roughness length z_0 [m]		
		A	B	WRF
1	Continuous urban fabric	0.5	1.0	1.0
2	Discontinuous urban fabric	0.5	1.0	1.0
3	Industrial or commercial units	0.7	0.7	0.5
4	Road and rail networks and assoc. land	0.1	0.2	0.5
5	Port areas	0.5	0.5	0.5
6	Airports	0.01	0.1	0.5
7	Mineral extraction sites	0.05	0.15	0.5
8	Dump sites	0.05	0.15	0.5
9	Construction sites	0.3	0.3	0.5
10	Green urban areas	0.8	0.8	0.5
11	Sport and leisure facilities	0.2	0.3	0.5
12	Non-irrigated arable land	0.05	0.1	0.1
13	Permanently irrigated land	0.03	0.1	0.1
14	Rice fields	0.03	0.1	0.1
15	Vineyards	0.3	0.3	0.2
16	Fruit trees and berry plantations	0.4	0.4	0.2
17	Olive groves	0.4	0.4	0.2
18	Pastures	0.03	0.1	0.1
19	Annual crops assoc. with perm. crops	0.1	0.2	0.2
20	Complex cultivation patterns	0.15	0.2	0.2
21	Agriculture with sig. areas of nat. veg.	0.15	0.2	0.2
22	Agro-forestry areas	0.5	0.5	0.2
23	Broad-leaved forest	1.0	1.0	0.9
24	Coniferous forest	1.2	1.2	0.9
25	Mixed forest	1.1	1.1	0.5
26	Natural grasslands	0.03	0.1	0.1
27	Moors and heathland	0.05	0.12	0.12
28	Sclerophyllous vegetation	0.07	0.12	0.12
29	Transitional woodland-shrub	0.4	0.4	0.12
30	Beaches - dunes - sands	0.003	0.01	0.01
31	Bare rocks	0.05	0.05	0.01
32	Sparsely vegetated areas	0.03	0.03	0.01
33	Burnt areas	0.2	0.2	0.01
34	Glaciers and perpetual snow	0.005	0.005	0.001
35	Inland marshes	0.05	0.05	0.001
36	Peat bogs	0.03	0.03	0.001
37	Salt marshes	0.02	0.02	0.001
38	Salines	0.005	0.005	0.001
39	Intertidal flats	0.0002	0.001	0.001
40	Water courses	0.0002	0.0002	0.0001
41	Water bodies	0.0002	0.0002	0.0001
42	Coastal lagoons	0.0002	0.0002	0.0001
43	Estuaries	0.0002	0.0002	0.0001
44	Sea and ocean	0.0002	0.0002	0.0001

Table 9: Roughness values for each land-use category.

the degree to which flow separation occurs. The WASP assumption is that only gentle terrain variations are present and that the flow stays attached, so RIX indicates how much the assumptions of the linearised flow model in WASP (IBZ model) are violated for a particular site. RIX values are calculated in radial line segments along a chosen number of angles from the origin. In this study the angles of the 12 sector centres were used and only the sector-average RIX value was considered.

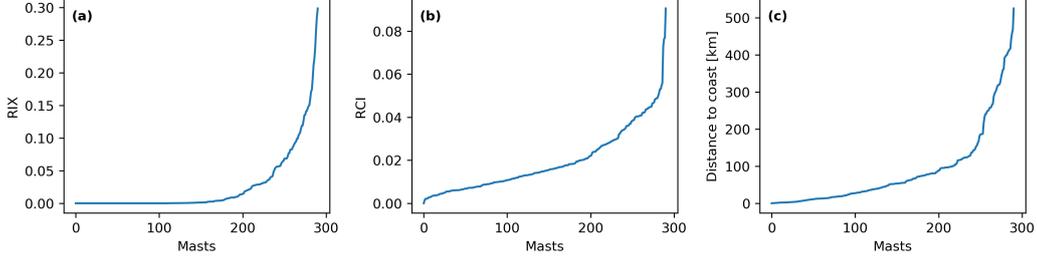


Figure 24: Distribution of RIX, RCI, and distance to the coast for the 291 masts.

Figure 24 (a) shows the distribution of RIX values for the mast locations. The orography at a third of the sites do not slope more than the critical slope and thus have RIX = 0, while RIX for the remaining sites are distributed between 0 and 30% (near-exponentially increasing).

The Roughness Complexity Index (RCI) characterises the complexity of the surface roughness surrounding each site. It is the average absolute speed-up at the point due to internal boundary layer effects as given by Eq. 1,

$$RCI = \frac{1}{N} \sum_{i=1}^N |\delta A_{rou,i}| \quad (1)$$

where δA_{rou} is the speed-up calculated by the WASP internal boundary layer (IBL) model (Sempreviva et al., 1990) and i denotes the sector index. It reflects the combined influence of upstream roughness changes weighted by a decaying exponential function of the distance to each roughness change. RCI = 0 corresponds to zero influence from upstream roughness changes. The largest RCI value of all the sites in this validation is ≈ 0.1 , but the typical value is in the range 0.01–0.02, see Figure 24 (b).

A Roughness Magnitude Index (RMI) based on the geometric average effective surface roughness across the 12 wind sectors calculated by the WASP model from the high-resolution roughness map $\overline{z_{0c}}_{WASP}$ was used to distinguish low and high surface roughness sites, e.g. grassland or pastures, from sites with abundant forests, or urban areas. RMI differs from RCI in that it does not measure variation, only average magnitude.

Metric	Category		
	<i>Low</i>	<i>Medium</i>	<i>High</i>
RIX	0 %	0 – 2 %	2 – 100 %
RCI	0 – 0.01	0.01 – 0.02	0.02 – ∞
RMI	0 – 0.12	0.12 – 0.2	0.2 – ∞

Table 10: Categories of the ruggedness index (RIX), the roughness complexity index (RCI), and the roughness magnitude index (RMI). The thresholds are right inclusive and left exclusive, so RIX 0 is *low* and RIX 0.05 is *medium*.

For statistical grouping purposes RIX, RCI, and RMI were each categorised into three groups: *low*, *medium*, and *high* based on the thresholds shown in table 10.

The distance to the (nearest) coastline was calculated for each mast using the high-resolution coastline dataset from the European Environment Agency (EEA). Figure 24 (c) shows the distribution of distances to the coast for the masts. ≈ 15 of the masts are located less than 10 km from the coast, ≈ 50 less than 50 km, and ≈ 75 less than 100 km.

3.1.4 Validation results

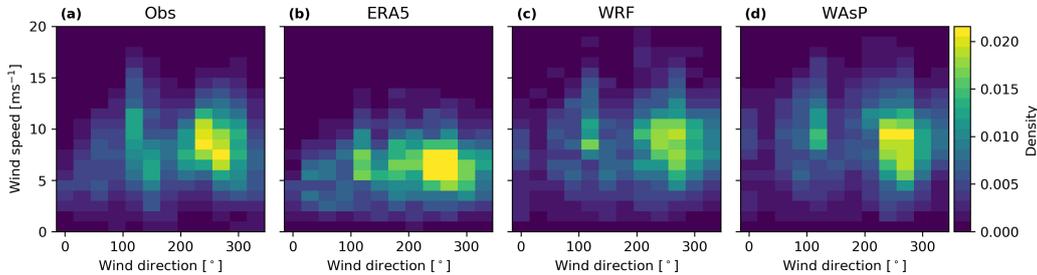


Figure 25: Observed (a) and modelled (b: ERA5; c: WRF, d: WRF-WAsP) wind climates at one example site. The wind climate histogram consists of 30 wind speed bins and 12 wind direction bins spanning intervals of 1 m s^{-1} and 30° respectively.

The wind climate is defined as the discrete multivariate probability density function (PDF) of wind speed and direction as shown for one example site in Figure 25. The wind speed and wind direction data from the models are binned into 30 bins of 1 m s^{-1} wind speed from 0 to 30 m s^{-1} and 12 bins of 30° wind sectors counterclockwise from north (0°).

The short form of *WAsP* is used interchangeably with *WRF-WAsP* in text, tables, and figures. Unless explicitly stated, WRF-WAsP results reported below are based on RCT-B.

Wind speed biases using roughness conversion table A vs. B

To study the sensitivity of the roughness conversion table used for WRF-WAsP an initial analysis was made of wind speed biases ($\bar{U}_{Mod} - \bar{U}_{Obs}$) for WRF-WAsP using RCT-A or RCT-B (table 9).

Figure 26 shows the modelled vs. observed mean wind speeds for WAsP using RCT-A and RCT-B and Table 11 shows the mean and standard deviations of wind speed biases for the masts in the three RIX categories for ERA5, WRF and WRF-WAsP using RCT-A and RCT-B.

RIX Cat.	n	ERA5 $\mu \pm \sigma$	WRF $\mu \pm \sigma$	WAsP $\mu \pm \sigma$	
				RCT-A	RCT-B
All	291	-1.50 ± 1.30	$\mathbf{0.02} \pm 0.78$	0.46 ± 0.78	$0.28 \pm \mathbf{0.76}$
Low	110	-0.72 ± 0.69	0.21 ± 0.54	0.28 ± 0.53	$\mathbf{0.06} \pm \mathbf{0.49}$
Medium	96	-1.39 ± 1.21	$\mathbf{0.03} \pm 0.90$	0.43 ± 0.81	$0.23 \pm \mathbf{0.76}$
High	85	-2.64 ± 1.17	$\mathbf{-0.25} \pm \mathbf{0.83}$	0.75 ± 0.93	0.62 ± 0.91

Table 11: Mean and standard deviation of $\bar{U}_{bias} = \bar{U}_{Model} - \bar{U}_{Obs}$ ($\mu \pm \sigma$) for ERA5, WRF and WRF-WAsP using RCT-A and RCT-B for all the masts and grouped by the RIX categories separately. n shows the number of masts in each group aggregated.

The scatter-plot (Fig. 26) shows that using RCT-A results in greater mean wind speeds for most masts compared to RCT-B. This shows that the increase of roughness values for the most frequent land-use classes in the CORINE dataset has large impact on the results. The larger mean wind speeds for RCT-A correspond to greater biases, relative to RCT-A, in most cases. The greatest biases occur for the *high* RIX sites for both roughness conversion tables (squares).

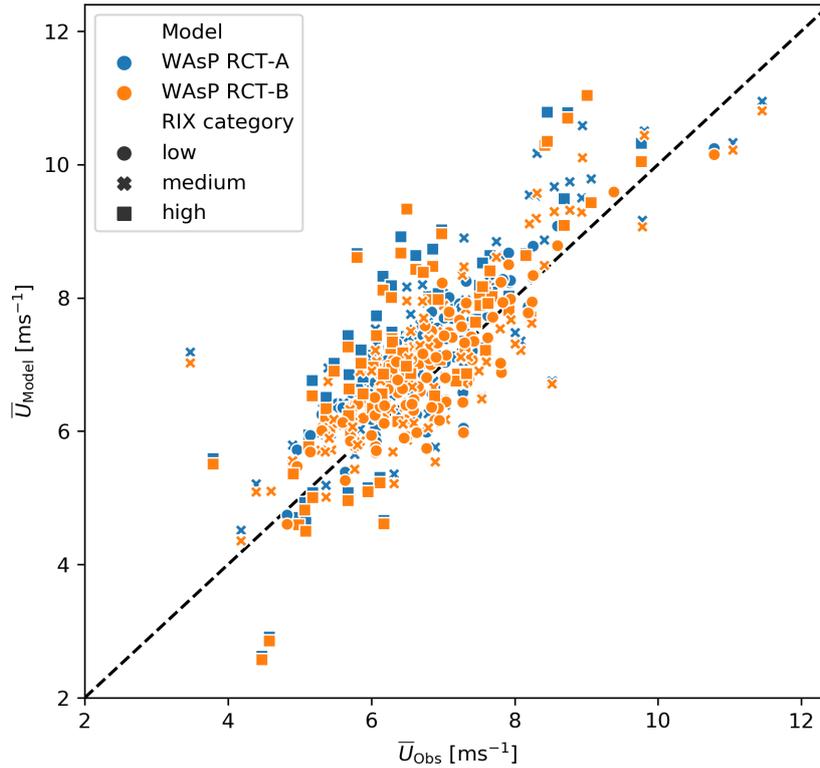


Figure 26: Observed vs. modelled mean wind speeds for WASP using RCT-A (blue) and WASP using RCT-B (orange). The Symbols indicate the orographic complexity category at each site: low (circle), medium (cross), and high (square) as defined in Table 10.

Aggregating the wind speed biases (\bar{U}_{bias}) by the mean (μ) and standard deviation (σ) for all the sites, and grouped by RIX category (table 11), shows clearly the overestimation of \bar{U}_{bias} seen for RCT-A in Fig. 26, on average by 0.46 m s^{-1} for all the sites. For all four groups (all-low-medium-high) μ and σ for \bar{U}_{bias} is smaller for RCT-B than RCT-A as well. WRF-WAsP using RCT-A results in larger μ s than WRF while σ is on par for all the masts and for the *low* RIX group. For *medium* and *high* RIX respectively WRF-WAsP RCT-A results in smaller (0.81 for WRF-WAsP RCT-A against 0.90 for WRF) and larger (0.93 against 0.83 m s^{-1}) spreads of \bar{U}_{bias} than WRF. WRF-WAsP RCT-B results in smaller σ of \bar{U}_{bias} than WRF for all the masts combined and for both the *low* and *medium* RIX sites while σ is larger than WRF for *high* RIX sites. WRF-WAsP RCT-B overestimates the mean wind speeds on average while WRF show near zero bias for all the mast combined. Only for the *low* RIX sites does WRF-WAsP RCT-B have a smaller mean bias of \bar{U}_{bias} than WRF. ERA5 generally underestimates the mean wind speeds significantly and has the largest spreads of \bar{U}_{bias} of any of the models.

Mean wind speed bias

In the following, the mean wind speed biases (\bar{U}_{bias}) for ERA5, WRF, and WRF-WAsP based on RCT-B are studied in more detail. Figure 27 shows the scatter-plots of observed vs modelled mean wind speed for the three models. ERA5 clearly underestimate the mean wind speed for most masts, especially for the *medium* and *high* RIX sites. The scatter for WRF and WASP falls more evenly on both sides of the 1-to-1 line and seem quite similar, apart from the *high* RIX sites for WASP that show greater wind speeds for some of the masts.

The relationship between four different predictors and \bar{U}_{bias} are shown in Fig. 28: (1) RIX, (2) RCI, (3) geometric mean effective surface roughness $\bar{z}_{0c,WAsP}$ and (4) measurement (wind speed) height above ground z . To achieve near-normal distribution of the predictors, the RIX values were

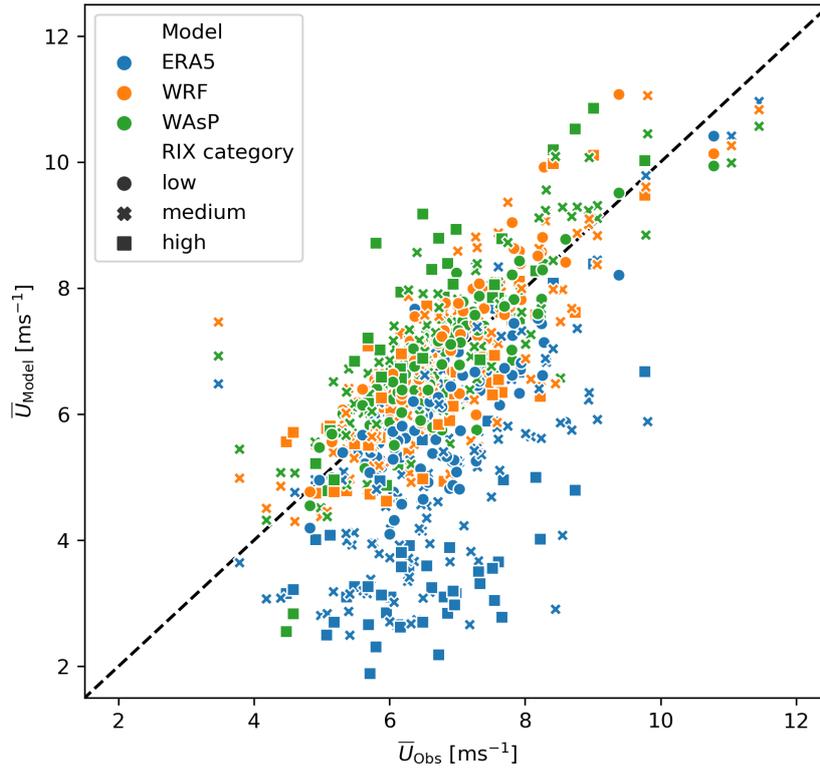


Figure 27: Scatter-plot of \bar{U}_{Obs} vs. \bar{U}_{Model} for ERA5 (blue), WRF (orange), and WASP (green). The symbols indicate the RIX category: *low* (closed circle), *medium* (cross), and *high* (square).

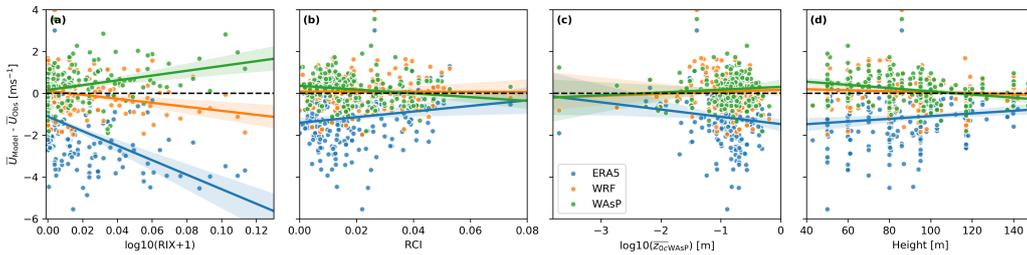


Figure 28: Scatter-plots of $\bar{U}_{bias} = \bar{U}_{Model} - \bar{U}_{Obs}$ for ERA5, WRF, and WASP vs. four different predictors: **(a)** RIX, **(b)** RCI, **(c)** geometric mean effective surface roughness $\bar{z}_{0c,WASP}$, and **(d)** measurement height z . The regression line-fit is based on iterative re-weighted least squares estimates using Huber's T-norm (Huber, 1973) (robust statistics), and the error bands are the 95th confidence interval estimated via bootstrapping using 200 iterations.

$\log(x + 1)$ -transformed and the $\bar{z}_{0c,WASP}$ values were $\log(x)$ -transformed. The regression line-fit in the figure uses robust statistics with iterative re-weighted least squares using Huber's T-norm (Huber, 1973).

In Fig. 28 **(a)** $\log_{10}(RIX+1)$ shows a clear correlation with \bar{U}_{bias} . For ERA5 and WRF this relationship results in a negative slope with RIX while for WASP it results in a positive slope with RIX. The WASP IBZ model assumes attached flows, so in steep terrain, where detachment is likely to occur, over-estimations of the wind speed at higher RIX sites are not surprising. On the other hand, the negative slope for ERA5 and WRF may be explained by under-resolved orographic speed-ups, since masts in high-RIX regions are more likely to be placed on hill-tops where orographic speed-ups occur.

No strong connections between RCI and \bar{U}_{bias} or $\bar{z}_{0c,WASP}$ and \bar{U}_{bias} are seen for WRF and WASP, while a gentle slope is seen for ERA5 (negative for RCI and positive for $\bar{z}_{0c,WASP}$).

A negative slope with height between the measurement height and \bar{U}_{bias} is seen for WASP, i.e.

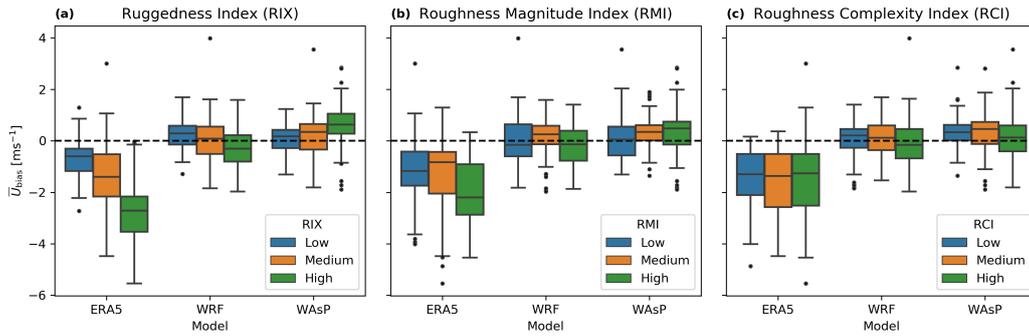


Figure 29: Distribution of mean wind speed biases for the three models: ERA5, WRF, and WAsP, for the three categories (*low*, *medium*, and *high*) of the Ruggedness Index (RIX), Roughness Magnitude Index (RMI), and the Roughness Complexity Index (RCI). The boxes are the 2nd and 3rd distribution quarterlies, whiskers extends to 1.5x of the interquartile range (distance from start of 2nd quartile to end of 3rd quartile) or to the outermost data point. Dots indicate outliers outside the 1.5x interquartile range.

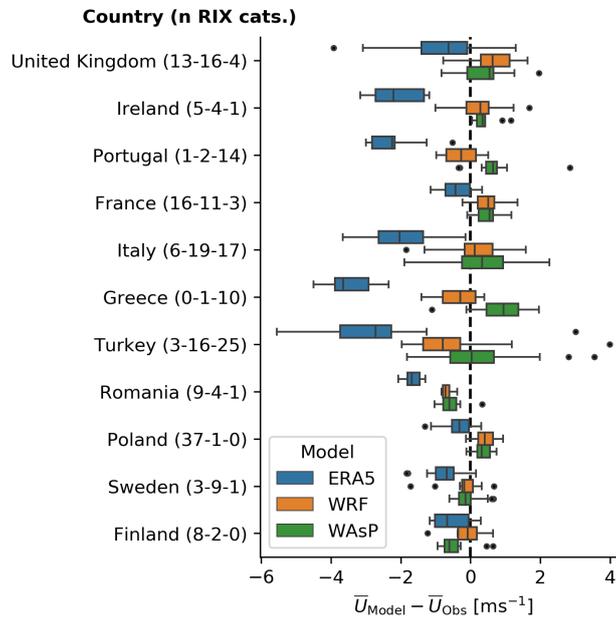


Figure 30: Box-plots of \bar{U}_{bias} 's for ERA5, WRF, and WAsP for the 11 European countries containing most masts. The number of masts in each RIX category (*low-medium-high*) is shown in the parenthesis, i.e. United Kingdom has 33 masts, 13 in *low* RIX terrain, 16 in *medium* and 4 in *high*. Boxes are the 2nd and 3rd quartiles, whiskers extends to the 1.5x the interquartile range (distance from start of 2nd quartile to end of 3rd quartile) or to the outermost data point. Dots indicate outliers outside 1.5x the interquartile range.

overestimation closer to the surface and underestimation further aloft. This is in contrast to WRF, which have a much smaller, but likewise negative, slope. The increased slope for WAsP may be an indication of inaccurate effective surface roughnesses used in the drag-law in the WRF-WAsP methodology. However, it could also be related to a number of other factors in the methodology related to the orographic or internal boundary layer models. For ERA5 a positive correlation is seen between \bar{U}_{bias} and z , which indicates that absolute \bar{U}_{bias} decreases with height (coming from a negative bias). This could in part be explained by the uncertainties related to the flow closer to the surface, that play a smaller role aloft. However, the coarse resolution of the ERA5 data used for this validation could also have an influence.

Interaction effects between predictors and conditional correlations were not studied in any

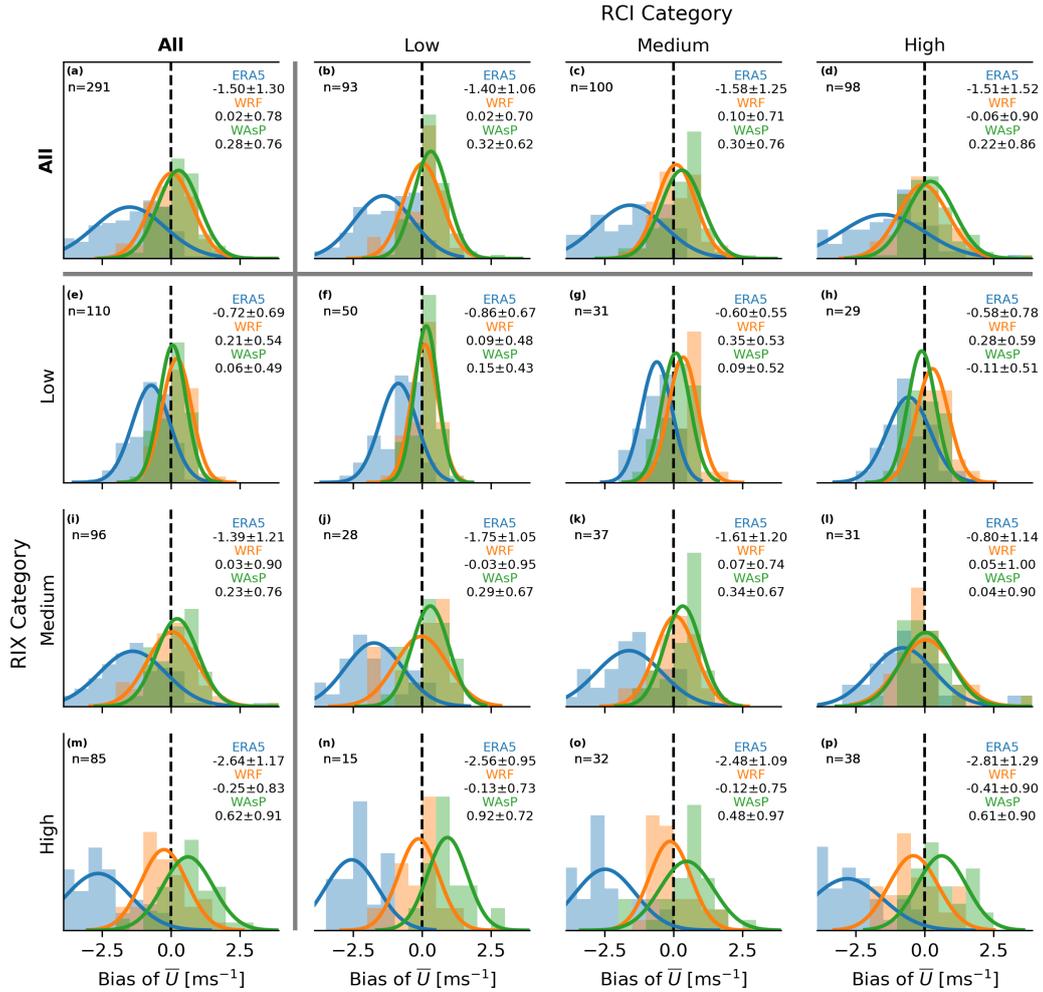


Figure 31: Histograms and estimated normal distributions of $\bar{U}_{\text{bias}} = \bar{U}_{\text{model}} - \bar{U}_{\text{obs}}$ for ERA5 (blue), WRF (orange) and WASP (green). The panels constitute the matrix of conditional results for four groups of RIX and RCI categories (*all-low-medium-high*). As the name suggests the *all* category contains all the masts. The values of the mean and standard deviation of the \bar{U}_{bias} are shown for each model in the upper right corner of each panel, corresponding to the normal-distribution curve (line) for that same model.

great detail, so further analysis of this could reveal further patterns.

Figure 29 shows the distribution of wind speed biases for the models divided into the category groups listed in Table 10. The RIX categories (Fig. 29 a) show the clearest relationship to the wind speed biases: negative biases tend to increase with larger complexity for ERA5 and positive biases tend to increase for WASP. For WRF wind speed biases tend to decrease with larger complexity.

RIX and surface roughness have a relative large correlation ($r \approx 0.4$) since the hilly or mountainous sites have many trees as well, therefore some of the signals seen for RIX are the same for RMI (Fig. 29 b) as well. The spread of wind speed biases is smallest for the *medium* RMI category, which could mean that these sites have smaller differences between the mesoscale and microscale surface roughnesses, which decreases the likelihood of large errors induced by large surface roughness errors caused by mischaracterisation of the land-use in the downscaling.

For RCI (Fig. 29 c), the spread of wind speed biases for the models tends to increase with larger complexities. Generally, no clear trend is seen for the distribution median for larger roughness complexities for ERA5 and WASP, however a weak decreasing trend is seen for WRF.

To investigate regional differences in the simulated wind climates the distribution of \bar{U}_{bias} for

the 11 countries containing most masts were computed, see Fig. 30. The figure shows the same pattern as Fig. 27, that ERA5 underestimates \bar{U} , especially for *high* RIX sites. All the three models show greater spread of \bar{U}_{bias} for countries with a significant number of masts in *high* RIX terrain and smaller spread in predominantly *low* RIX terrain. WRF shows a pattern of slightly overestimating \bar{U} in north-western Europe, including the British isles, France and Poland, while for countries in south-eastern Europe: Romania, Greece, and Turkey, WRF shows an underestimation of \bar{U} . The WAsP results are highly linked to the WRF results, and a similar pattern is seen for WRF, but with deviations between WAsP and WRF especially for countries with many masts in *high* RIX terrain.

The aggregate \bar{U}_{bias} results for the models are presented in Fig. 31. Across all the masts WRF has the lowest mean bias of 0.02 m s^{-1} , while WAsP has small positive mean bias of 0.28 m s^{-1} and ERA5 significantly underestimates the mean wind speed by -1.50 m s^{-1} on average. The overall σ of \bar{U}_{bias} for WRF and WAsP is similar (0.78 vs 0.76 m s^{-1}), while ERA5 shows a significantly larger σ (1.30 m s^{-1}). Inspecting the different sub-groups of RIX and RCI categories reveals that RIX is the strongest predictor of μ and σ (left-most column), especially for the WAsP results that suffer from increased overestimation and spread for sites of greater orographic complexity where flow attachment is a bad assumption. However, all three models also show larger σ for higher RCI complexity (top row). As expected, the most robust results (smallest σ) for WRF and WAsP comes from the masts in *low* RIX and *low* RCI complexity locations. The mean \bar{U}_{bias} for WRF and WAsP for the same masts (in *low* RIX complexity) are also relatively low (WRF: 0.09 m s^{-1} ; WAsP: 0.15 m s^{-1}).

Generated power

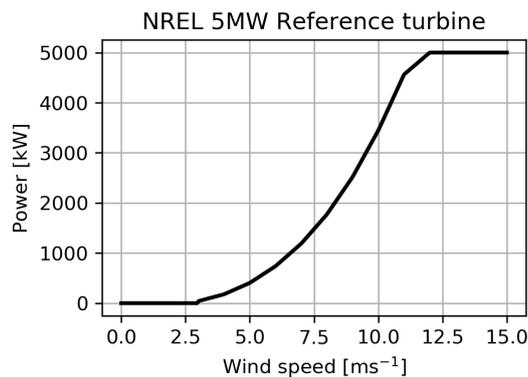


Figure 32: Power curve for the NREL 5MW reference wind turbine used in the validation to validate the accuracy of the modelled estimate of power generation (Jonkman et al., 2009).

To study how well estimated power generation from the modelled wind climates approximate the actual production for a hypothetical wind turbine at the site, the mean mean power generation was estimated by using the observed and modelled wind climate PDFs and the NREL 5MW reference wind turbine power curve (Jonkman et al. (2009), see Fig. 32). Given the shape of the the power curve, with power generation changing only in the wind speed range between 3 and 12 m s^{-1} , this is the most important range to accurately estimate the wind climate PDF. Deviations in the PDF below 3 or above 12 m s^{-1} do not change the mean power production.

In figure 33 the modelled vs. observed generated power is plotted for all three models. The RIX complexity at each mast is shown by the marker of each point. The plot is similar to the scatter plot for \bar{U} , but the relative errors are enhanced given the non-linear relation between wind speed and power. The aggregate results of \bar{P}_{bias} in percent is shown in Fig. 34. Both WRF and WAsP overestimate \bar{P} in *low* and *medium* RIX terrain, more so WRF in *low* RIX terrain ($\mu = 11.32\%$)

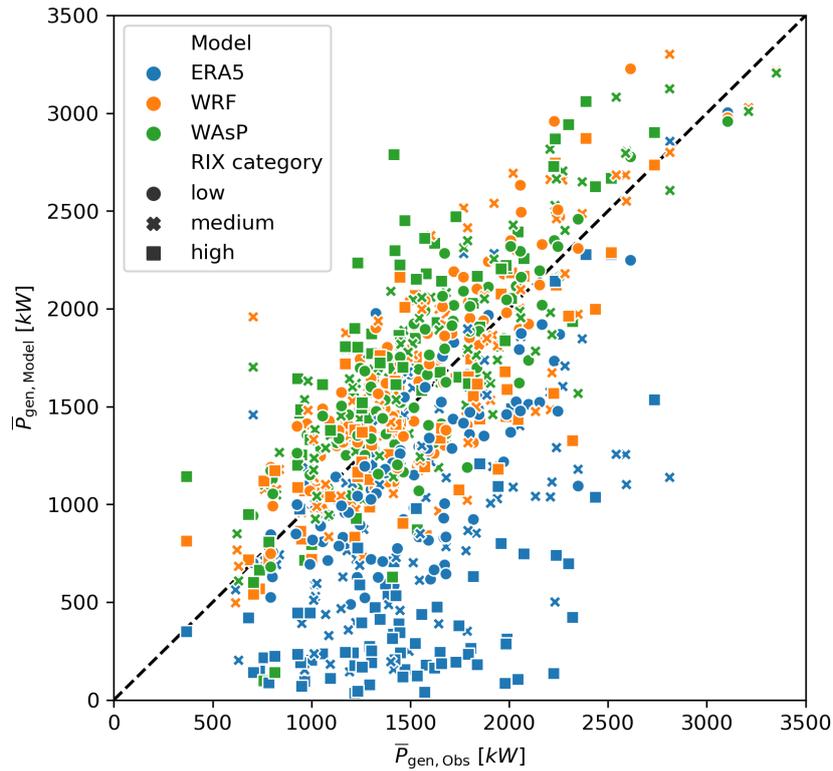


Figure 33: Scatter-plot of the observed vs. modelled mean generated power $\bar{P}_{gen,obs}$ vs. $\bar{P}_{gen,model}$ for ERA5 (blue), WRF (orange), and WASP (green). The symbols indicate the RIX category: *low* (closed circle), *medium* (cross), and *high* (square).

and WASP in *medium* RIX terrain ($\mu = 13.13\%$). Similarly to \bar{U} the positive bias and the spread of \bar{P}_{bias} for WASP increases for increasing orographic complexity. From $7.33 \pm 17.56\%$ for *low* RIX to $21 \pm 36.66\%$ for *high* RIX. The mean of the \bar{P}_{bias} for WRF decreases with increasing orographic complexity, which seems counterintuitive. One possible explanation is that it is caused by a compensation of errors where the overestimation of the synoptic or mesoscale scale wind speed is compensated by an underestimation of the sub-grid-scale orographic speed-up. ERA5 significantly underestimates the power production. From 20% on average in low orographic complexity to 70% in *high* complexity.

Wind direction

The wind direction is another important characteristic to capture accurately in the wind climate. The Earth Movers Distance (EMD, see Section 2.1.3 for an explanation) was used to measure statistical distance between the observed and modelled wind direction PDFs and denoted EMD_{WD} in the following. Figure 35 shows distributions of $\log_{10}(EMD_{WD})$ for a matrix of different conditions. The $\log_{10}(x)$ transformation was done to achieve near-normality of the data. Smaller μ means more similar wind direction PDFs. The figure shows that WRF and WASP share near-identical accuracy of the wind direction PDFs for *low* RIX and RCI, while greater difference between WRF and WASP are seen for greater complexities. In *low* RIX complexity ERA5 has the largest spread of the models ($\sigma = 0.28$ overall) but also the lowest mean $\log_{10}(EMD_{WD})$ ($\mu = -0.65$). In more complex terrain more high-resolution models (WRF and WASP) have lower statistical distance than ERA5 on average.

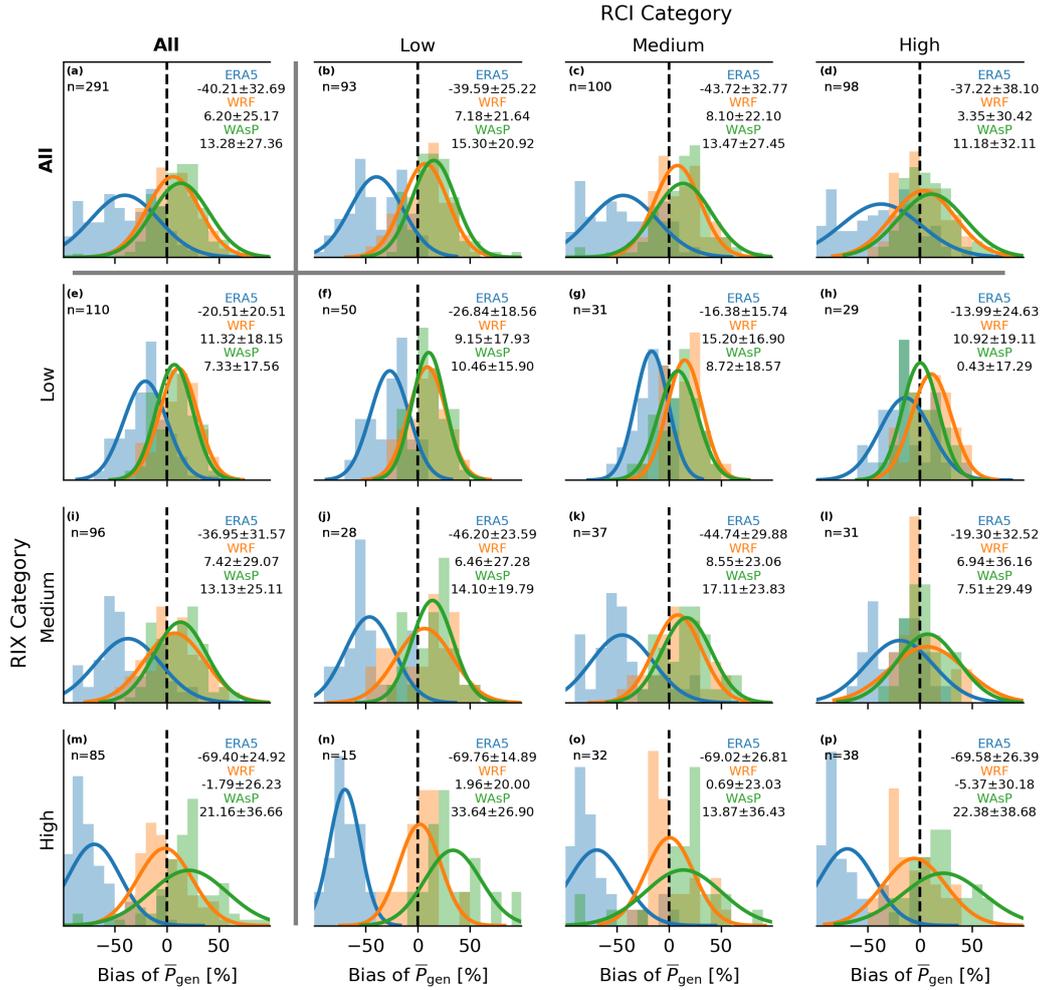


Figure 34: Histograms and estimated normal distributions of $\bar{P}_{bias} = \bar{P}_{model} - \bar{P}_{obs}$ in percent for ERA5 (blue), WRF (orange), and WAsP (green). The panels constitute the matrix of conditional results for four groups of RIX and RCI categories (*all-low-medium-high*). As the name suggests the *all* category contains all the masts. The values of the mean and standard deviation of the \bar{P}_{bias} are shown for each model in the upper right corner of each panel, corresponding to the normal-distribution curve (line) for that same model.

Sensitivity to surface roughness

Given the large uncertainties surrounding the roughness conversion table for CORINE land-use classes (or any other dataset for that matter), this final analysis in this validation study investigates the sensitivity of the aggregate results for \bar{U}_{bias} , \bar{P}_{bias} , and $\log_{10}(\text{EMD}_{WD})$ when scaling the conversion values in RCT-B (table 9). The uncertainty of individual roughnesses assigned to each class is unknown, but it could be as large as a factor of 2 or 3 (Kelly and Jørgensen, 2017). To study the effect of scaling the roughnesses by this magnitude, the z_0 in RCT-B was halved and doubled, and the results computed for these new scaled values. In table 12 the resulting statistics are shown.

Across the different metrics, the most accurate overall results are obtained by doubling the roughnesses in RCT-B. This is in part due to the underestimation for *low* RIX sites balancing the overestimation at *high* RIX sites. However, the absolute mean biases are also smaller for a doubled z_0 than for the default values. To no surprise, halving the roughnesses results in increased overestimation and larger spread of results. The results indicate that inaccuracies are present either in the WRF-WAsP model chain or the measurements because a doubling of the z_0 values in RCT-B are much larger than the typical values used in the scientific community. However,

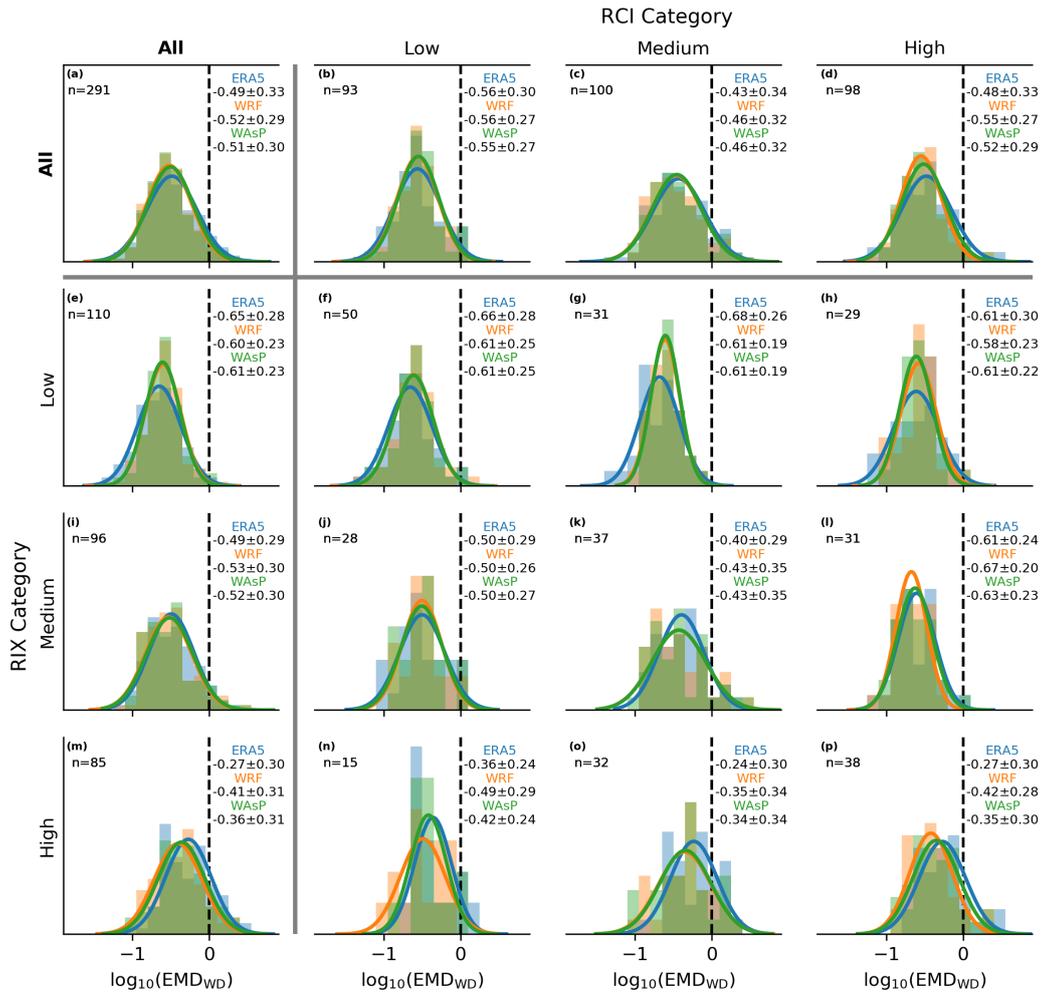


Figure 35: Histograms and estimated normal distributions of EMD_{WD} (smaller EMD is better, it means more similar PDFs) for ERA5 (blue), WRF (orange), and WAsP (green). The panels constitute the matrix of conditional results for four groups of RIX and RCI categories (*all-low-medium-high*). As the name suggests the *all* category contains all the masts. The values of the mean and standard deviation of the EMD_{WD} are shown for each model in the upper right corner of each panel, corresponding to the normal-distribution curve (line) for that same model.

from the results presented here, it is not possible to determine the main sources of error between the observations and WAsP-modelled quantities.

Measurement errors cannot be ruled out, but all measurements used were quality controlled. Another potential source of error is related to the WRF-WAsP methodology and how it was used in this validation study. The methodology assumes that the effective response of WRF to variations in orography and surface roughness can be captured by the linearised flow and the internal boundary layer model of WAsP. However, this has not been thoroughly validated, and may result in inaccurate generalised winds that are used for the downscaling with microscale effects derived from high-resolution terrain datasets. Finally, the surface characteristics, and especially the surface roughness characterisation, lacks accurate, robust, and validated datasets and thus remains a large potential source of errors.

Conclusions

291 tall meteorological masts located all over Europe were used to validate the ERA5, WRF, and WRF-WAsP models, which constitutes the NEWA model-chain. One contiguous year of measurements from 40 – 150 m above ground were used from each of the masts between 2007

Metric	RIX Cat.	WRF-WAsP RCT-B $\mu \pm \sigma$			Units
		$1/2z_0$	z_0	$2z_0$	
\bar{U}_{bias}	All (291)	0.61 ± 0.78	0.28 ± 0.76	-0.11 ± 0.75	$[\text{m s}^{-1}]$
\bar{U}_{bias}	Low (110)	0.37 ± 0.50	0.06 ± 0.49	-0.30 ± 0.49	$[\text{m s}^{-1}]$
\bar{U}_{bias}	Medium (96)	0.57 ± 0.78	0.23 ± 0.76	-0.18 ± 0.75	$[\text{m s}^{-1}]$
\bar{U}_{bias}	High (85)	0.98 ± 0.92	0.62 ± 0.91	0.21 ± 0.91	$[\text{m s}^{-1}]$
\bar{P}_{bias}	All (291)	23.93 ± 29.10	13.28 ± 27.36	0.88 ± 25.38	$[\%]$
\bar{P}_{bias}	Low (110)	18.21 ± 19.83	7.33 ± 17.56	-5.15 ± 15.76	$[\%]$
\bar{P}_{bias}	Medium (96)	24.15 ± 27.76	13.13 ± 25.11	0.11 ± 22.81	$[\%]$
\bar{P}_{bias}	High (85)	31.08 ± 37.78	21.16 ± 36.66	9.56 ± 34.24	$[\%]$
$\log_{10}(\text{EMD}_{\text{WD}})$	All (291)	-0.50 ± 0.28	-0.51 ± 0.30	-0.51 ± 0.30	$[-]$
$\log_{10}(\text{EMD}_{\text{WD}})$	Low (110)	-0.59 ± 0.22	-0.61 ± 0.23	-0.62 ± 0.24	$[-]$
$\log_{10}(\text{EMD}_{\text{WD}})$	Medium (96)	-0.51 ± 0.29	-0.52 ± 0.30	-0.53 ± 0.31	$[-]$
$\log_{10}(\text{EMD}_{\text{WD}})$	High (85)	-0.37 ± 0.30	-0.36 ± 0.31	-0.36 ± 0.31	$[-]$

Table 12: Mean (μ) and standard deviation (σ) of the mean wind speed bias (\bar{U}_{bias}), the mean power generation bias (\bar{P}_{bias}), and the statistical distance between observed and modelled wind direction PDF's ($\log_{10}(\text{EMD}_{\text{WD}})$) for WRF-WAsP RCT-B, with roughness values for each CORINE land-use class halved ($1/2z_0$), default (z_0), and double ($2z_0$). The statistics are presented for all masts and for each group of RIX complexities. The number of masts in each group is shown in the RIX cat. column.

and 2016.

The terrain around each mast was categorised using the ruggedness index (RIX) into three groups (*low-medium-high*) based on the orographic complexity and three groups (same labels) using the WAsP calculated flow speed-up due to internal boundary layer effects to characterise the complexity of surface roughness variations.

The main findings were:

- The average wind speed bias for the 291 masts is $0.28 \pm 0.76 \text{ m s}^{-1}$ for WRF-WAsP compared to $-1.50 \pm 1.30 \text{ m s}^{-1}$ and $0.02 \pm 0.78 \text{ m s}^{-1}$ for ERA5 and WRF, see Fig. 31.
- In simple terrain (*low* RIX) and terrain with mostly gently sloping hills (*medium* RIX), the wind speed bias \bar{U}_{bias} estimated by WRF-WAsP has the smallest spread of any of the models ($\sigma=0.43 \text{ m s}^{-1}$ in *low* RIX and $\sigma=0.76 \text{ m s}^{-1}$ in *medium* RIX), and the average wind speed bias is low in simple terrain (*low* RIX) $\mu=0.06 \text{ m s}^{-1}$, compared to $\mu=0.21 \text{ m s}^{-1}$ for WRF.
- In *medium* RIX terrain WRF-WAsP overestimates the mean wind speed by 0.23 m s^{-1} on average, while the mean bias of WRF is 0.03 m s^{-1} . In terrain with the greatest occurrence of steep hills (*high* RIX) WRF-WAsP overestimates the mean wind speed by 0.62 m s^{-1} . The WAsP flow model assumes attached flow, which is an inaccurate assumption in steep terrain, which may explain the overestimation of mean wind speed by WRF-WAsP. A similar trend, greater complexity means larger spread, was seen for roughness complexity. The smallest spread for WRF-WAsP was seen for the lowest roughness complexity.
- The observed and modelled wind climates were used to estimate mean power production for a hypothetical wind turbine at each mast location using the power curve of the NREL 5MW reference turbine. The study shows that WRF and WRF-WAsP overestimate the mean power production for the masts by on average $6.20 \pm 25.16\%$ and $13.28 \pm 27.36\%$, respectively, while ERA5 underestimates the mean power production for the same masts by $-40.21 \pm 32.69\%$ on average.
- In *low* and *medium* RIX terrain the smallest mean bias and spread is seen for WRF-WAsP, $7.33 \pm 17.56\%$ in *low* RIX and $13.13 \pm 25.11\%$ in *medium* RIX, while the values for WRF

are $11.32 \pm 18.15\%$ and $7.42 \pm 29.07\%$ respectively for the same masts. In *high* RIX terrain the mean bias and spread is much larger for WRF-WAsP ($21.16 \pm 36.66\%$) than for WRF ($-1.79 \pm 26.23\%$), possible due to an overestimation of the orographic speed-ups.

- The Earth Movers Distance (EMD) was used to measure the similarity between the observed and modelled wind direction PDFs (wind roses). On average, the WRF and WRF-WAsP wind directions are very similar, with a similar level of accuracy, and ERA5 has slight less accurate wind direction PDFs. However, when only masts in the *low* RIX locations are considered, the wind direction PDFs for ERA5 are slightly more accurate than WRF and WRF-WAsP. For *high* RIX locations the opposite is true (WRF and WRF-WAsP are slightly more accurate).

The WRF-WAsP methodology relies heavily on accurate maps of elevation and surface roughness. Especially accurate surface roughness maps remain a topic in need of further advancement. In this validation study, and in the computation of the actual NEWA atlas itself, the CORINE land-use dataset has been used to obtain a roughness map for Europe by converting land-use classes to fixed surface roughness values. However, no thoroughly validated roughness conversion table exists and suggested roughness values for each class differ substantially. Further, the accuracy of characterising roughness into a small number of predefined classes will inevitably be limited. A doubling of the roughness values in the roughness conversion table revealed better estimates of the mean wind speed, mean power generation, and direction PDF, which shows that inaccuracies still remain in the WRF-WAsP methodology or measurements, calling for a better understanding of the effective response in the WRF model to terrain variations and for more accurate roughness maps.

3.2 Model evaluation based on tall mast data

Data from 17 tall masts over Central Europe have been collected and processed. The locations of the masts are shown in Figure 36. The data of these research met masts was either publicly available, available within the NEWA consortium or was provided upon request from the owners for use within NEWA (see the acknowledgements at the end of the report). 14 out of the 17 masts were finally used for the model evaluation. Three masts had to be excluded as the data quality was not sufficient. Primary reasons for low data quality were massive towers affecting the wind speed measurements or too many gaps in the data. The masts are listed in Table 13. The data availability provided there is the percentage of valid measurements during the indicated time period for the wind speed at or close to 100 m height which was mostly used for the results shown later. For other heights and sensors the availability is of course different.

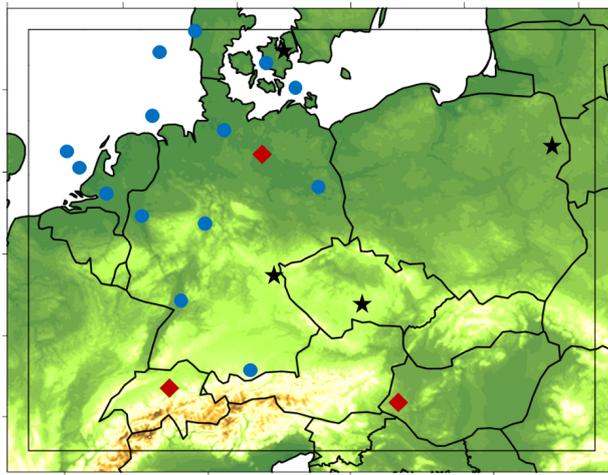


Figure 36: Locations of the tall masts collected for the model evaluation. Blue circles: masts used in the evaluation; red diamonds: masts not used due to bad data quality; black stars: prospective sites whose data could not be acquired in time.

mast name	full years covered	data availability [%]	main evaluation height
Cabauw	2001–2017	99.46	80 m
FINO1	2004–2016	95.36	100 m
FINO2	2008–2015	93.58	100 m
FINO3	2011–2017	82.94	90 m
Hamburg	2015–2017	99.36	110 m
Hohenpeißenberg	2016–2017	99.91	93 m
Høvsøre	2005–2017	96.79	100 m
Ijmuiden	2012–2015	99.04	89 m
Jülich	2005–2017	99.79	120 m
Karlsruhe	2008–2016	98.76	100 m
Kassel	2013/14, 2016/17	99.89, 95.94	135 m, 140 m
Lindenberg	2009–2017	99.43	98 m
OWEZ	2009–2010	91.45	116 m
Tystofte	2000–2013	88.44	39 m

Table 13: List of tall masts. Data availability for the wind speed sensor at the indicated height (which is the closest height to 100 m). The Kassel data contains two full years of data but not a full calendar year.

The mast data have been synchronised with the NEWA production run data, i.e. time stamps where one dataset (usually the mast dataset) has NaN values were discarded from the comparison. The observations have been sampled to 30 min which is the frequency of the modelled data. The simulated data have been interpolated to the measured heights. The comparison has been done for all measurement heights with a sufficient data availability. However, to keep clarity we are mostly showing results for one measurement height at or close to 100 m in this report. For simplicity this height is referred to as 100 m, although for several stations the actual height is not exactly 100 m (cf. Table 13).

First, an overview of error statistics for 12 stations with full year data will be given. For convenience we plotted the metrics in a matrix so that the performance of all the stations and all the years can be assessed at a glance. Tystofte was excluded from these tables as it has only data for 39 m height which is too far off from 100 m.

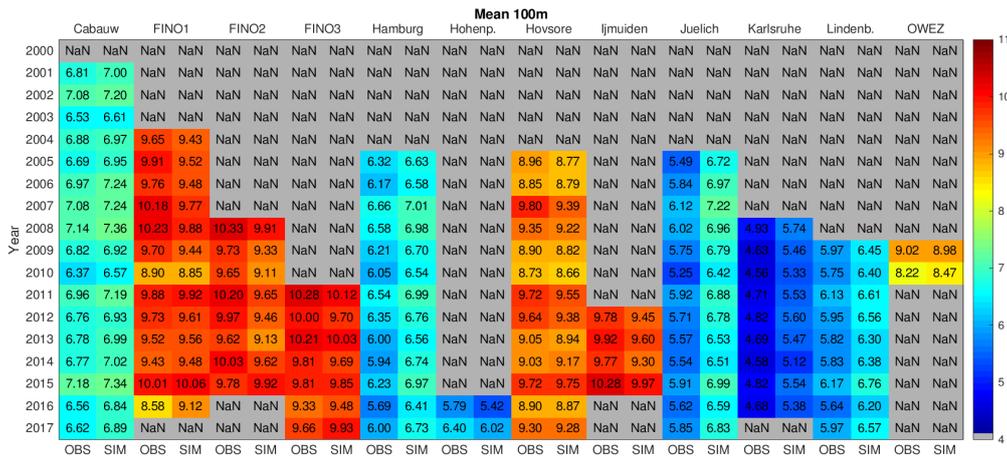


Figure 37: Annual mean wind speed for the NEWA production run and observed data at 12 mast locations at or close to 100 m height.

Figure 37 shows the observed and simulated annual mean wind speed at 100 m for the 13 mast locations. For all the onshore locations the simulated mean wind speed is always higher than the observed mean wind speed. At the offshore locations often but not always the simulated wind speed is lower than the observed one, though the difference is less than onshore.

This becomes more clear when looking at Figure 38 that shows the normalised bias for each location and year. A positive bias is evident for all onshore locations and all years. The bias is smallest for Cabauw with only 1–4 % while most of the other onshore locations have a bias between 10 and 20 %. The offshore locations have mostly a small negative bias between 0 and 5 %, sometimes also a small positive bias (up to 6 %).

The RMSE values shown in Figure 39 are mostly ranging around 2 m s^{-1} . The highest values of around 2.7 m s^{-1} are reached at the Hohenpeißenberg site which is located on top of a large hill close to the Alps. Also Jülich has quite high values between 2.5 and 2.6 m s^{-1} . The lowest RMSE is found for the Cabauw mast (around 1.8 m s^{-1}). The (Pearson) correlation between observed and simulated data is shown in Figure 40. At the offshore locations FINO1-3, Ijmuiden and OWEZ and also at the coastal location Høvsøre the correlation reaches high values of close to 0.9. Also Cabauw (very flat terrain, not far from the coast) has rather high values between 0.8 and 0.85. The lowest correlations are found for the onshore locations Lindenberg, Jülich and Karlsruhe (far from the coast, forested or rather complex terrain) and range around 0.7–0.75.

All these annual metrics show that the inter-annual variation not only of wind speeds but also of model performance can be quite high. The mean wind speeds (see Figure 37) can vary by as much as 1 m s^{-1} per year with a standard deviation of around 3 %. The standard deviation of the annual correlations is typically around 0.02 (2–3 %), the maximum spread is between 0.03 and 0.08 (cf. Figure 40).

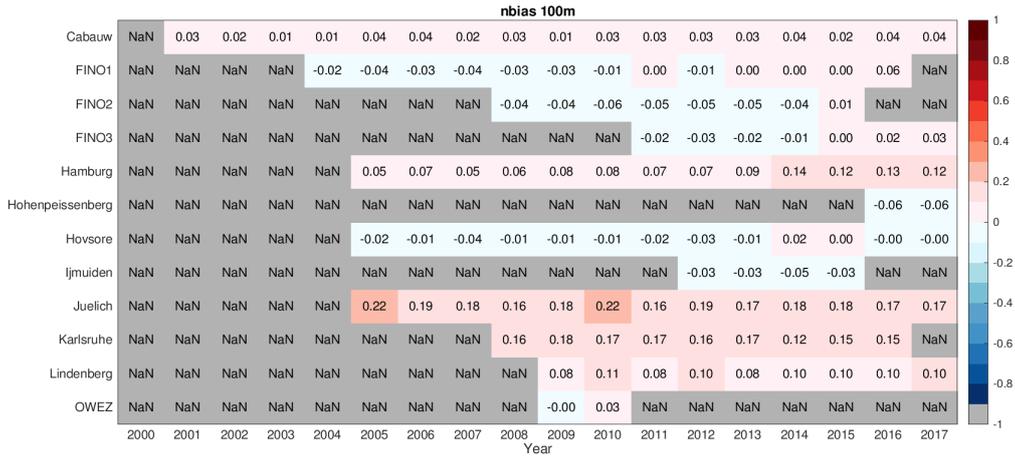


Figure 38: Annual normalised bias of the NEWA production run data compared to observed data at 12 mast locations.

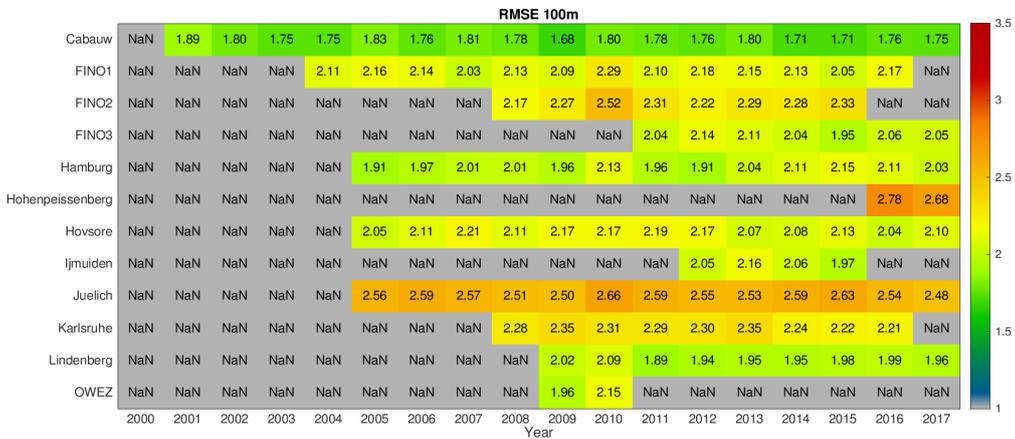


Figure 39: Annual RMSE of the NEWA production run data compared to observed data at 12 mast locations.



Figure 40: Yearly correlation (Pearson) of the NEWA production run data with observed data at 12 mast locations.

To investigate if the model performance changes as function of the season or the time of the day, we plotted similar matrices also for the month of the year and the hour of the day. As the correlations for the month of the year in Figure 41 show, a clear annual cycle becomes visible. For the onshore locations it is obviously more pronounced than for the offshore locations. A maximum of the correlation is reached in winter, a minimum in summer. At the coastal site Høvsøre the highest correlation is 0.90 in December and January, the lowest correlation 0.81 in July. The onshore site Hamburg experiences much stronger variations between 0.86 in January and 0.66 in August. For these monthly comparisons the length of the dataset has to be considered. Naturally, a site will generate different monthly statistics when only two years are considered compared to a site with 15 years of data. Also a diurnal cycle can be seen in the data (see Figure 42) but, not surprisingly, only for the onshore locations. The correlation reaches a minimum during the evening hours and first half of the night and a maximum in the late morning close to noon. The variation between day and night is, however, smaller than between summer and winter.

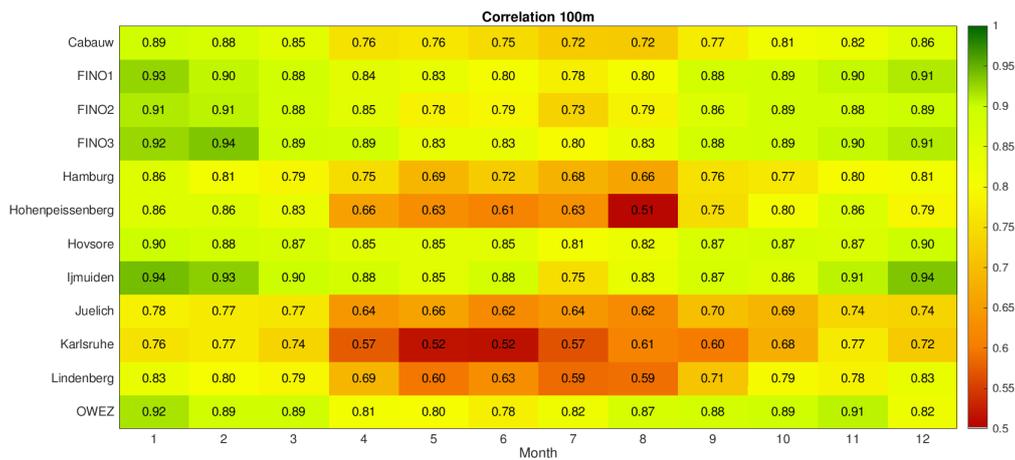


Figure 41: Monthly correlation (Pearson) of the NEWA production run data with observed data at 12 mast locations, aggregated over all available full years of data.

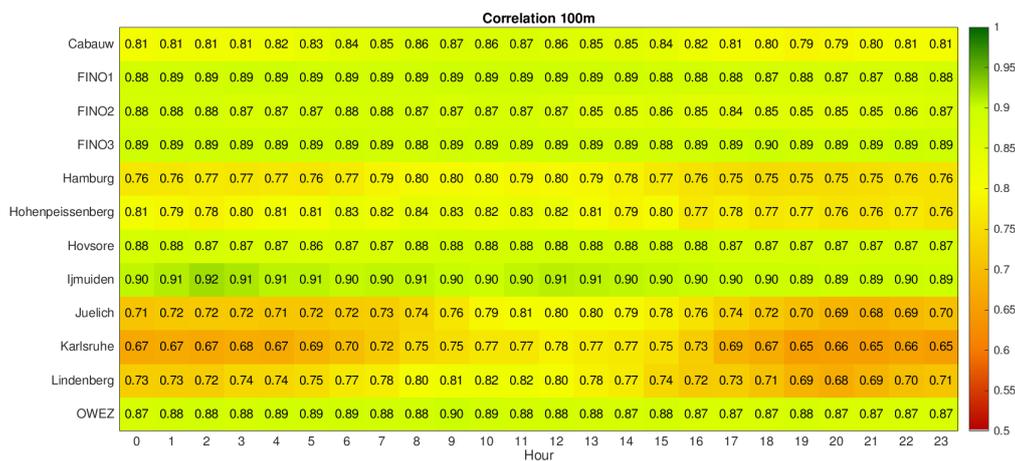


Figure 42: Hourly correlation (Pearson) of the NEWA production run data with observed data at 12 mast locations, aggregated over all available full years of data.

Next, we are presenting site-specific statistics — exemplarily for three sites: FINO1 (offshore), Cabauw (onshore, flat terrain, close to the coast) and Karlsruhe (onshore, rather complex forested terrain, far from the coast). Similar statistics and figures have been produced for all other sites, as well.

At the offshore site FINO1 the observed wind speed histograms are well reproduced by the NEWA production run (Figure 43). Both onshore sites show some differences in the distributions which are more pronounced at the more complex site Karlsruhe. The observations contain more frequently lower speeds in the range 2–7 m s⁻¹ while the simulations contain more often higher wind speeds above 8 m s⁻¹. A similar shift towards higher simulated wind speeds can be observed for all of the onshore sites and especially in lower heights.

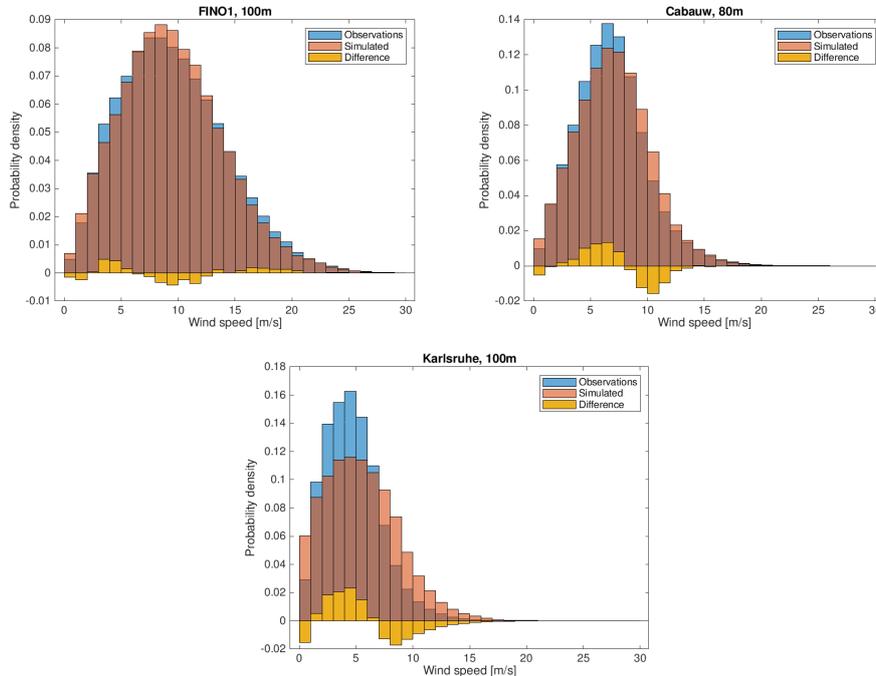


Figure 43: Wind speed histograms for three sites: observed wind speed (blue), simulated wind speed (red) and the difference between observed and simulated wind speed (yellow).

Although the general shape of the wind rose is well reproduced, all three sites feature a shift of wind directions in the simulated data towards the right (towards higher degrees). This is also true for other heights we compared.

Figure 45 presents all data points in a scatter plot together with a linear fit and the bin averages. Again it becomes evident that offshore the simulated and observed data match very well, the linear fit is almost coinciding with the 1:1 line. At Cabauw the scatter is slightly higher and the linear fit has a slightly smaller slope compared to the 1:1 line. At Karlsruhe the slope is much smaller indicating a strong overprediction of high wind speeds and underprediction of low wind speeds.

Figure 46 presents again the annual correlations but for each site separately in a xy-diagram and furthermore for all the evaluated heights. In general the correlation increases with height. This increase is pronounced for the onshore stations and almost negligible offshore (except the lowest height). As already discussed above, the inter-annual variations are considerable, and some characteristic years are emerging. One example is the year 2010 which has low correlations at both FINO1 and Cabauw as well as most of the other sites (cf. also Figure 40). In fact, Karlsruhe is the only exception. This could be related to large-scale circulation patterns, as it is the most southern location with data in 2010. Another example is the year 2015 that features high correlations at all three sites shown here (and most of the other evaluated sites).

Plotting the correlation by month as shown in Figure 47 reveals a clear seasonal cycle at all locations with high correlations in winter and much lower correlations in summer. The decrease of correlations towards lower heights seems to be less pronounced in winter than in summer.

Plotting in the same way the correlation for each hour of the day, a pronounced daily cycle becomes evident (Figure 48). Interestingly, the correlation does not continuously increase with

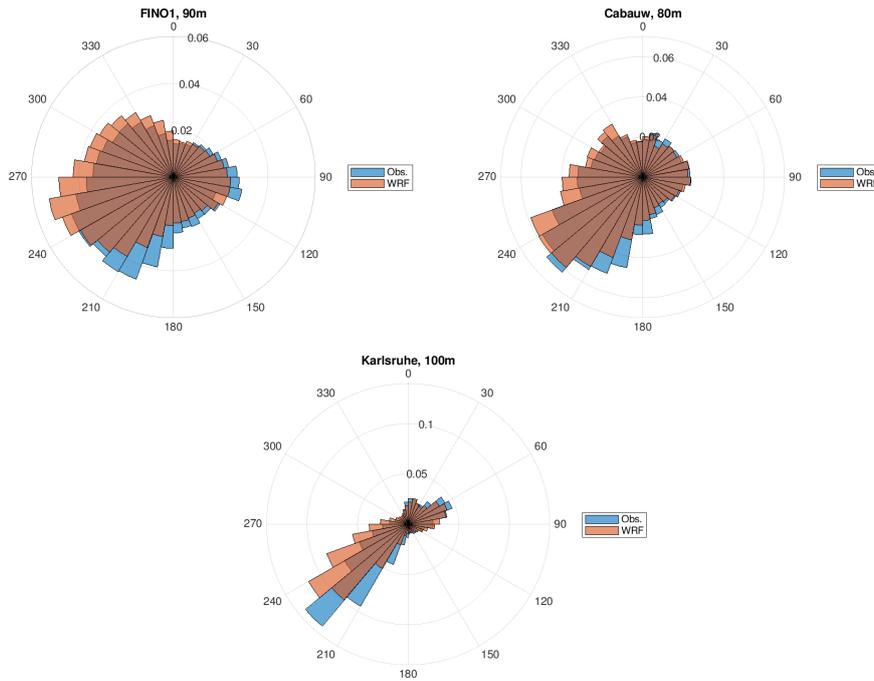


Figure 44: Wind roses for three sites: observed wind directions (blue) and simulated wind directions (red).

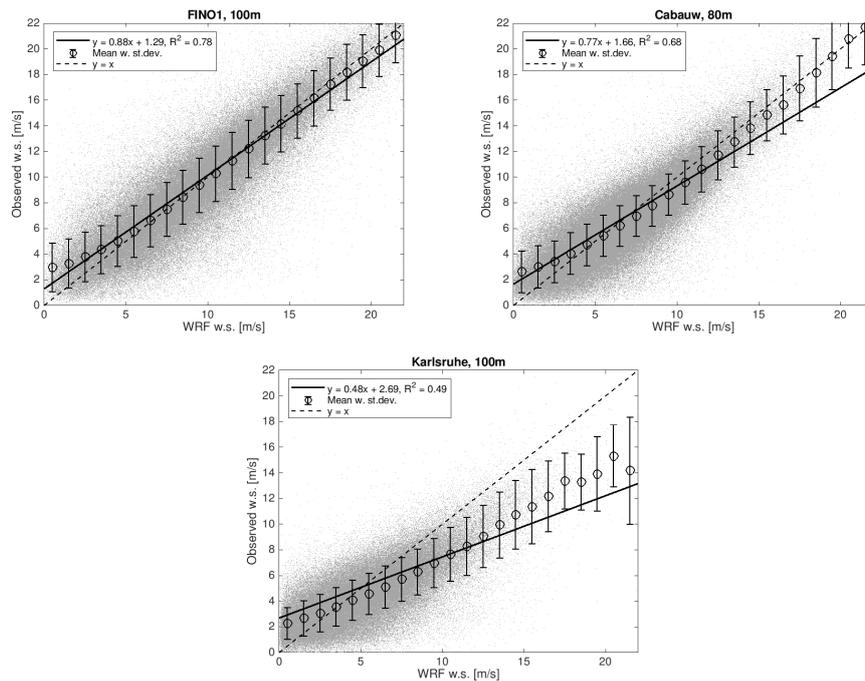


Figure 45: Observed (vertical axis) versus simulated wind speed (horizontal axis) for three sites with linear fit (solid line), 1:1 line (dashed) and averages for each simulated wind speed bin (circles with error bars).

height during the night at the onshore sites. Here, the lower levels have a higher correlation than the levels above, except the very high levels (e.g. 200 m) where correlation increases again. Apparently, the model has problems with the shallow stable nighttime boundary layer. The lowest level at Cabauw (10 m) has two clear minima of correlation around (the average) sunrise and

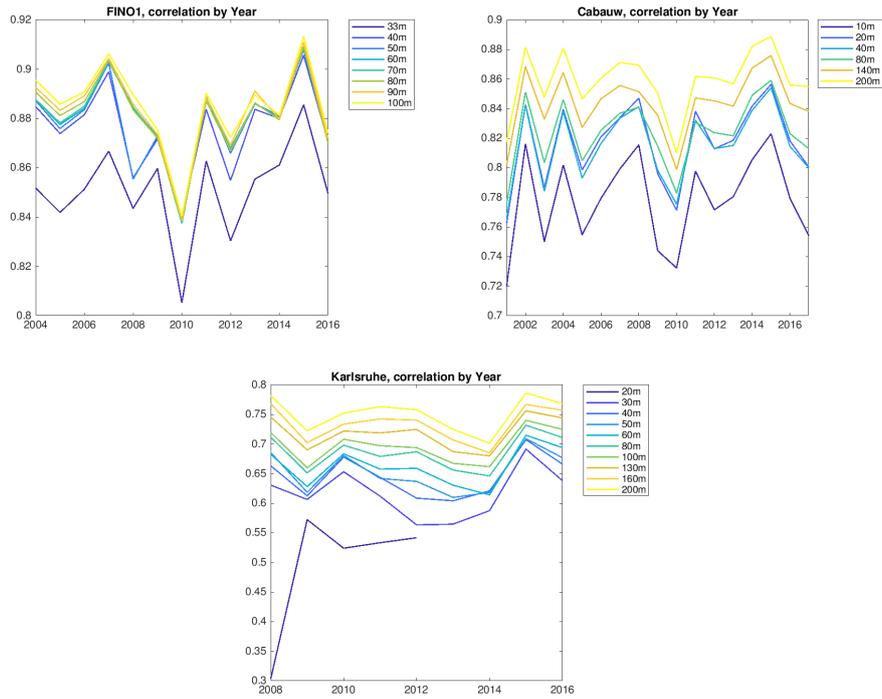


Figure 46: Correlation by year for three sites and different heights.

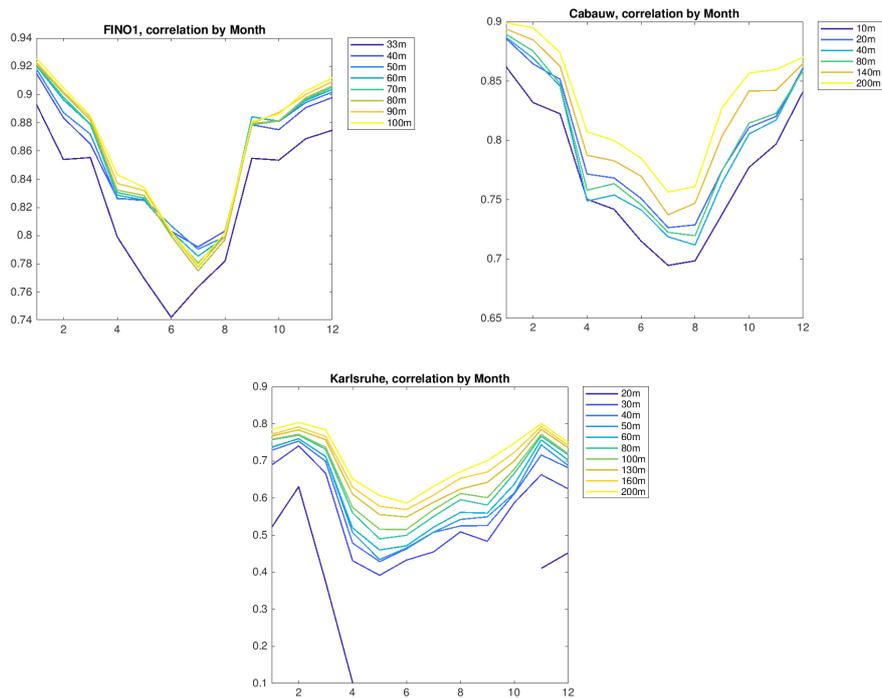


Figure 47: Correlation by month of the year for three sites and different heights.

sunset pointing to difficulties of the model to simulate the morning and evening transition of the surface layer.

Finally, we analysed the correlation for different wind directions (Figure 49). FINO1 and Cabauw show the same trend: higher correlations for the sector from south to west and low correlations for northerly wind directions. This is somehow correlated with the frequency of oc-

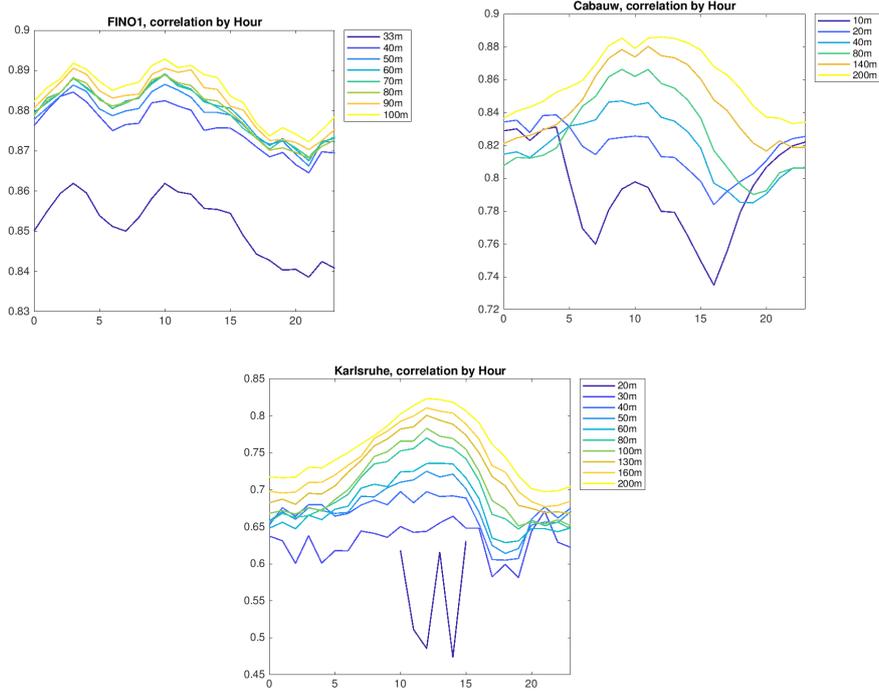


Figure 48: Correlation by hour of the day for three sites and different heights.

currence (compare with Figure 44): high correlations for the main wind direction, low correlation for the sector which has the lowest frequency of occurrence. A similar conclusion can be drawn for Karlsruhe which features a clear bimodal wind direction distribution (see Figure 44 bottom). The two main wind directions southwest and northeast have also the highest correlations while the lowest correlations are found for southeasterly and northwesterly wind that are very rarely occurring.

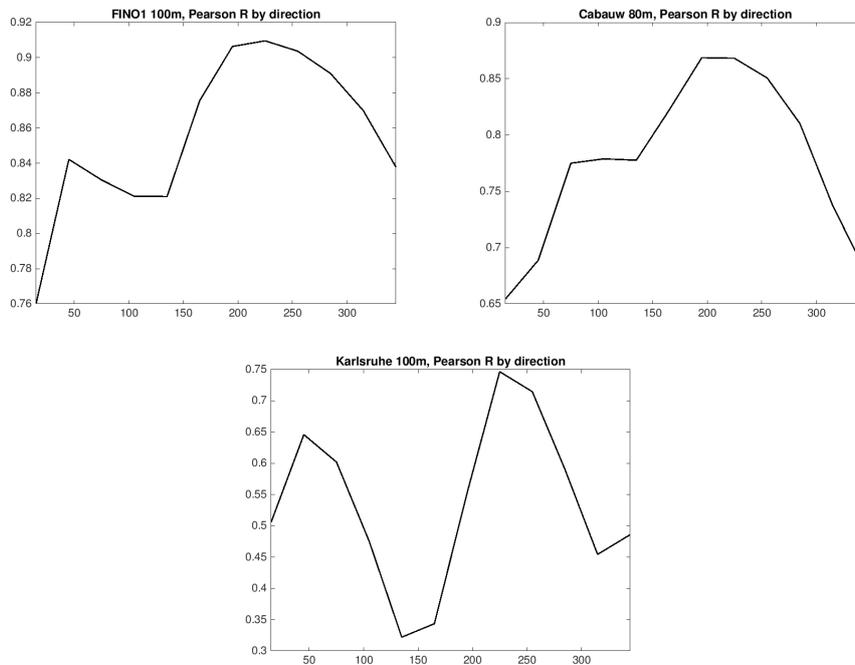


Figure 49: Correlation by wind direction for three sites.

3.3 Model evaluation based on wind profiles (mostly) offshore

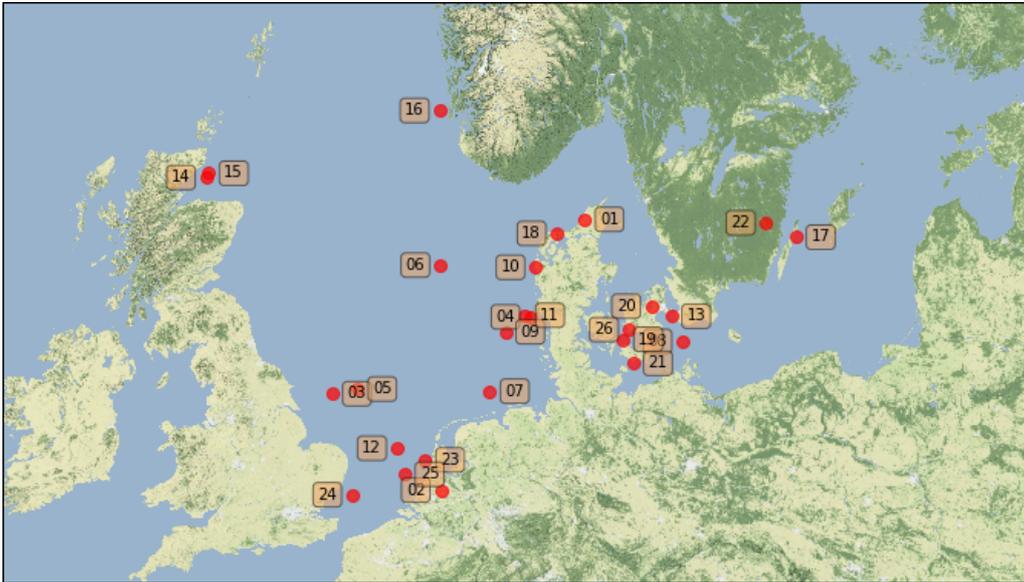


Figure 50: Location of the sites used in the validation of the NEWA mesoscale simulations. The numbers correspond to those sites described in Table 14.

This section presents the validation of the wind speeds simulated in the NEWA production run with mostly lidar and mast wind data from the North and Baltic Seas and some land-based sites in Denmark, Sweden and Holland (see Figure 50). The observations from these sites cover the period from 2000–2018 with very varied lengths of data availability (see Figure 51). The name of the sites, height range at which wind speeds are measured, type of site and measurement device and notes regarding the location and quality of the data are presented in Table 14. The anonymised stations are grouped as Northern North Sea (NNS), Central North Sea (CNS) and South North Sea (SNS). These stations were part of the NORSEWiND project (Hasager et al., 2013).

The mast data has been quality controlled and a rough attempt has been made to minimise the effect of the mast flow distortion on the wind speed measurement. At FINO1, FINO2, Risø, Høvsøre and Østerild, where wind speed measurements are available from only one boom, winds originating $\pm 10^\circ$ of the boom direction are filtered. At other sites, where wind speeds are measured at more than one boom direction (e.g. FINO3), the wind speed measurements are combined according to the wind direction to minimise the mast flow distortion. At IJmuiden, the data was processed as discussed in Kalverlaa et al. (2017). At Cabauw, the data was processed and gap filled by KNMI and used as is. Figure 51 shows the data availability per month for all the sites for 2000–2018 after the quality control and filtering has been carried out. Some mast sites over land, e.g. Cabauw and Tystofte, show nearly complete time series, while other lidar sites, e.g. CNS4 and Site SNS2, show poor data availability ($< 30\%$ in most months) and relatively short time series (< 2 years).

The data used is that described in Table 14, the site locations are shown in Figure 50. The table also points out some of the issues, e.g. mast flow distortion or presence of wind farms, that limit the quality of the wind measurements.

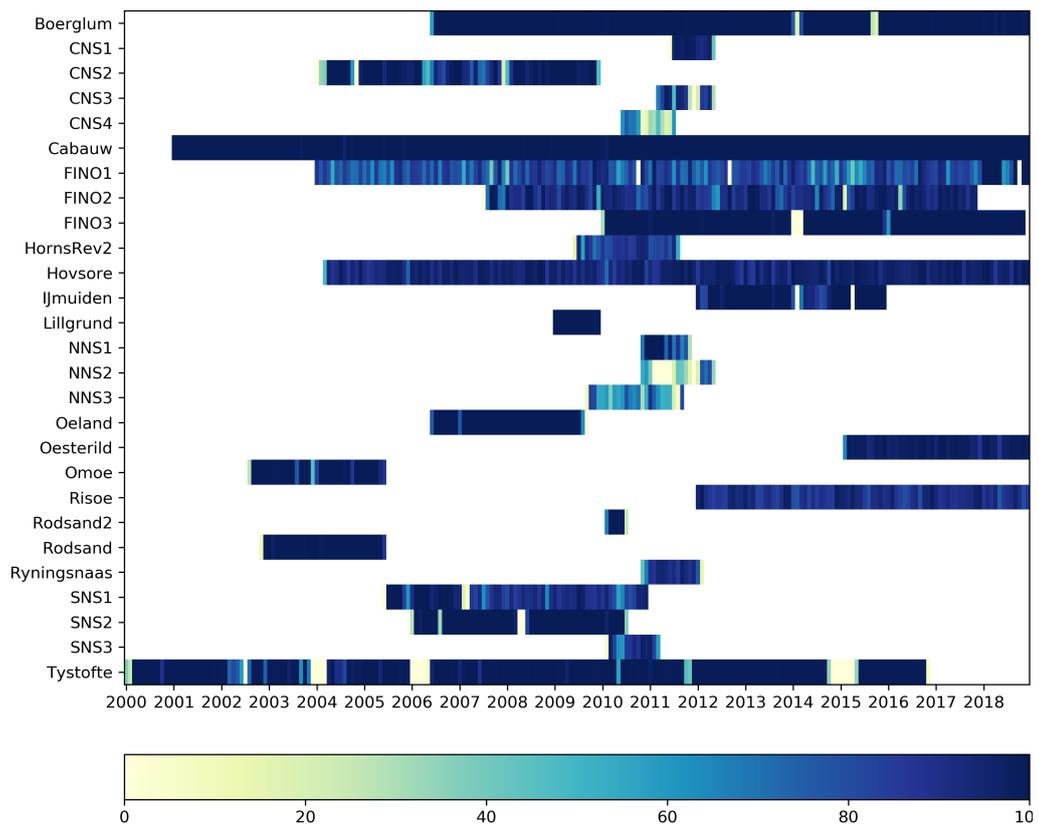


Figure 51: Monthly data availability (%) as a function of time for the wind speed time sites in Figure 50 and Table 14 after filtering.

Table 14: Data description at the various validation sites. The domain label represents the original WRF domain used. The underlined height represents the level used in later validation in the scatter diagrams (Figs. 56–58) and the error histograms (Figs. 59–61). The various rows are labelled according to location: offshore (light blue), coastal (light green) and land (beige).

N	site	NEWA domain	height range [m AGL/AMSL]	type ² location ³	Notes
1	Børglum	SB	10, 20.5, <u>31.5</u>	M/L	
2	Cabauw	CE	10–200, <u>140</u>	M/L	
3	CNS1	GB	92–252, <u>108</u>	L/O	
4	CNS2	SB	20–70, <u>70</u>	M/O	DI, WF
5	CNS3	GB	92–182, <u>102</u>	L/O	
6	CNS4	SB	85–295, <u>105</u>	L/O	LD
7	FINO1	CE	34–104.5, <u>91.5</u>	M/O	DI ⁴ , WF ⁵
8	FINO2	CE	32.4–102.5, <u>92.4</u>	M/O	DI
9	FINO3	CE	50–90, <u>90</u> , 100	M/O	WF
10	Høvsøre	SB	10–116, <u>100</u>	M/C	DI
11	Horns Rev 2	SB	66–286, <u>106</u>	L/O	WF
12	IJmuiden	CE	27–290, <u>115</u>	M+L/O	
13	Lillgrund	SB	20–65, <u>65</u>	M/O	WF
14	NNS1	GB	76–106, <u>106</u>	L/O	
15	NNS2	GB	60–270, <u>100</u>	L/O	LD ⁶
16	NNS3	SB	67–300, <u>100</u>	L/O	
17	Öland	SB	10–50, <u>50</u>	M/C	
18	Østerild	SB	40–244, <u>106</u>	M/L	
19	Omø	SB	9.6, 29.6, <u>50.6</u>	M/O	DI
20	Risø	SB	44.2–125.2, <u>125.2</u>	M/L	DI
21	Rødsand II	SB	57–68, <u>68</u>	M/O	SH
22	Ryningsnäs	SB	40–138, <u>138</u>	M/F	
23	SNS1	CE	21–106 <u>116</u>	M/O	
24	SNS2	GB	52.5–72.5, <u>72.5</u>	M/O	
25	SNS3	CE	70–250, <u>110</u>	L/O	
26	Tystofte	SB	10–39, <u>39</u>	M/L	

²Type: L: lidar, M: mast

³Location: O: offshore, C: coastal, L: land, F: forest

⁴DI: flow distortion, affected sectors are filtered from long-term mean

⁵WF: flow maybe affected by wind farm in some sectors

⁶LD: Limited data

Figure 52 shows a good match between the model-simulated wind speeds and those measured at the lidar sites offshore. The biases in mean wind speed at around 100 m AGL are in general < 3%. The comparison at Horns Rev 2 is influenced by the presence of the wind farm at low levels and the comparison at NNS2 is limited by low data availability. The taller measurements sites also show underestimation by WRF above 150 m AGL. However, a fraction of this underestimation can be explained by higher recovery rates favouring higher wind speeds as outlined in Floors et al. (2018). Especially because the underestimation is not so apparent in the tall mast measurements at SNS1 and Østerild.

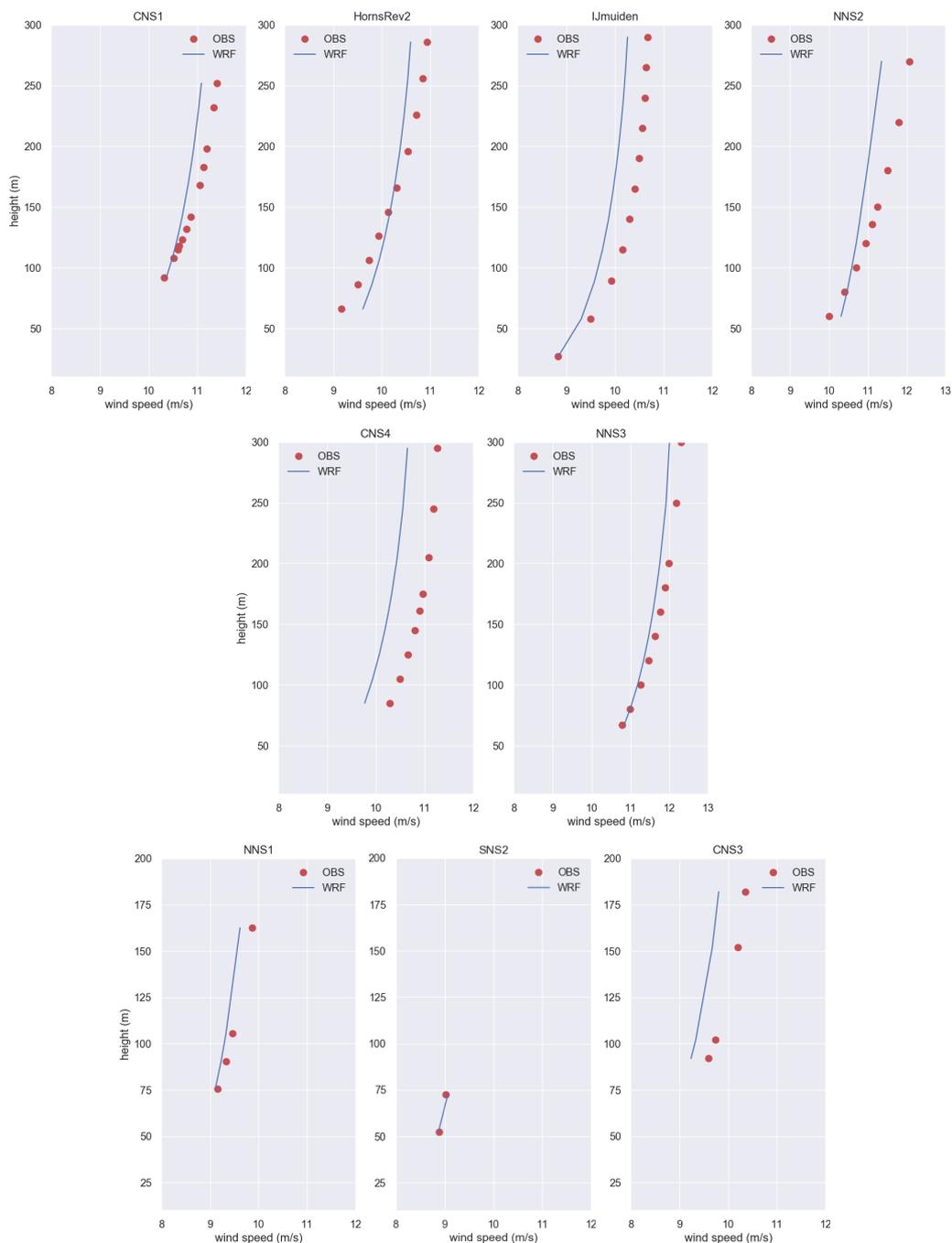


Figure 52: Comparison of long-term averaged wind speed (m s^{-1}) as a function of height for the offshore lidar measurements and the NEWA model simulations. Note that the axis of the plots might vary from one row to the others.

At the tall masts offshore, Figure 53, there is an excellent correspondence between the observed and simulated wind speed at the two German masts FINO1 and FINO2 after filtering of the mast-distorted measurements. The comparison is also excellent at the near-coastal sites in Figure 54, except for the measurements at Lillgrund, which is near the wind farm at the site. At the land-based sites in Figure 55 the comparison is also good at higher levels. Near the ground, the measurements are influenced by the local terrain conditions, which are often not well represented by WRF, e.g. at Risø and Østerild.

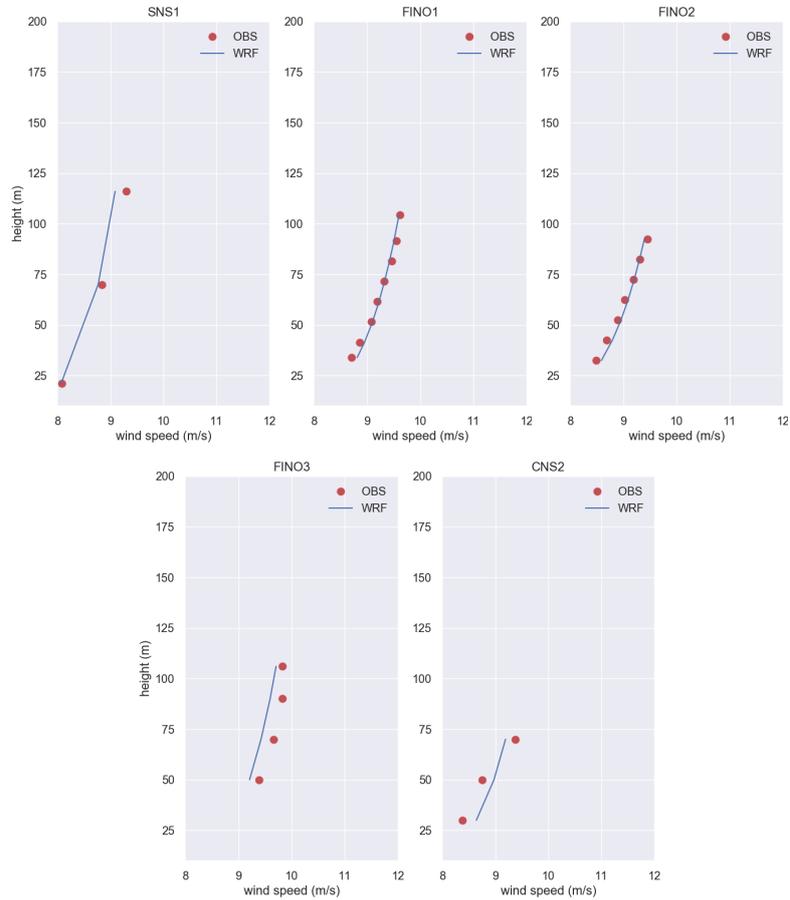


Figure 53: Comparison of long-term averaged wind speed (m s^{-1}) as a function of height for the offshore mast measurements and the NEWA model simulations.

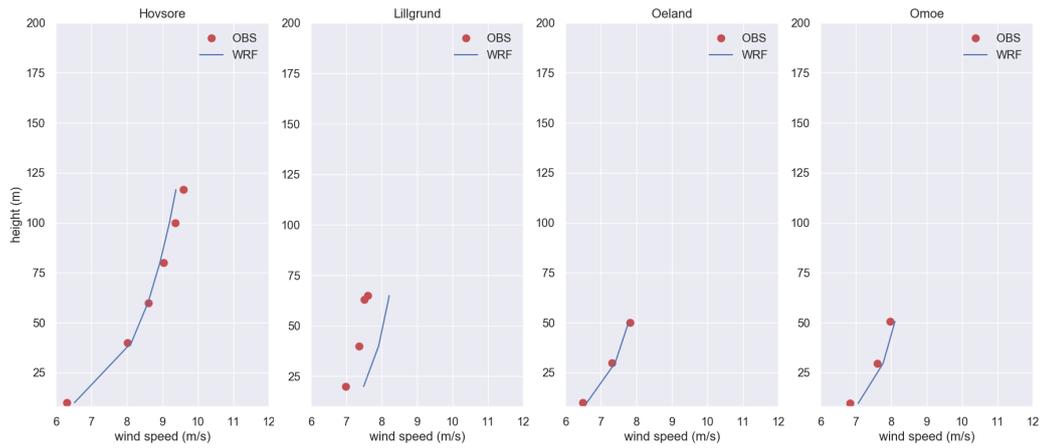


Figure 54: Comparison of long-term averaged wind speed (m s^{-1}) as a function of height for the coastal mast observations and the NEWA model simulations. Note that the axis on each row of graphs might be different.

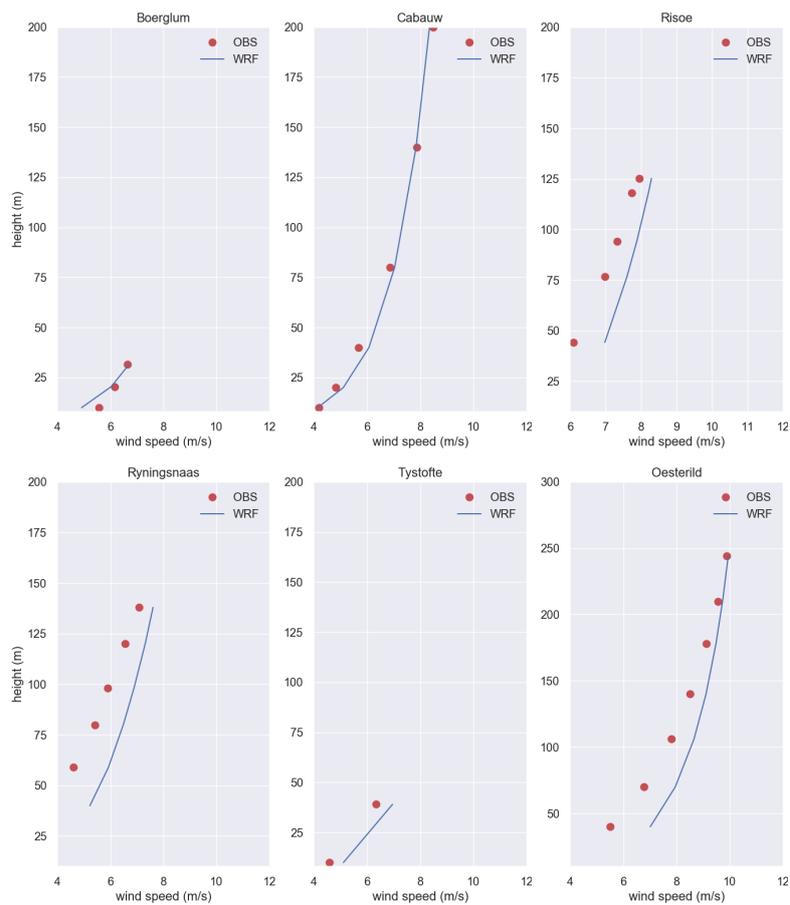


Figure 55: Comparison of long-term averaged wind speed (m s^{-1}) as a function of height for the land-based mast observations and the NEWA model simulations. Note that the axis on each row of graphs might be different.

The next set of figures (Figures 56-58) show the scattergrams between the time series of wind speed observations and WRF-simulated winds. We selected a height near 100 m AGL for each site (underlined in Table 14) for this comparison since this is often near the hub height of current wind farms offshore.

In general, the correspondence between the time series of observed wind speed and that simulated is good, with a tendency for larger spread at lower wind speeds than at high wind speed. The correlations vary from 0.91 at IJmuiden, which has well-sampled and well quality controlled mast and lidar measurements, to 0.68 at NNS2, which has low data availability and a short time series (Figure 51).

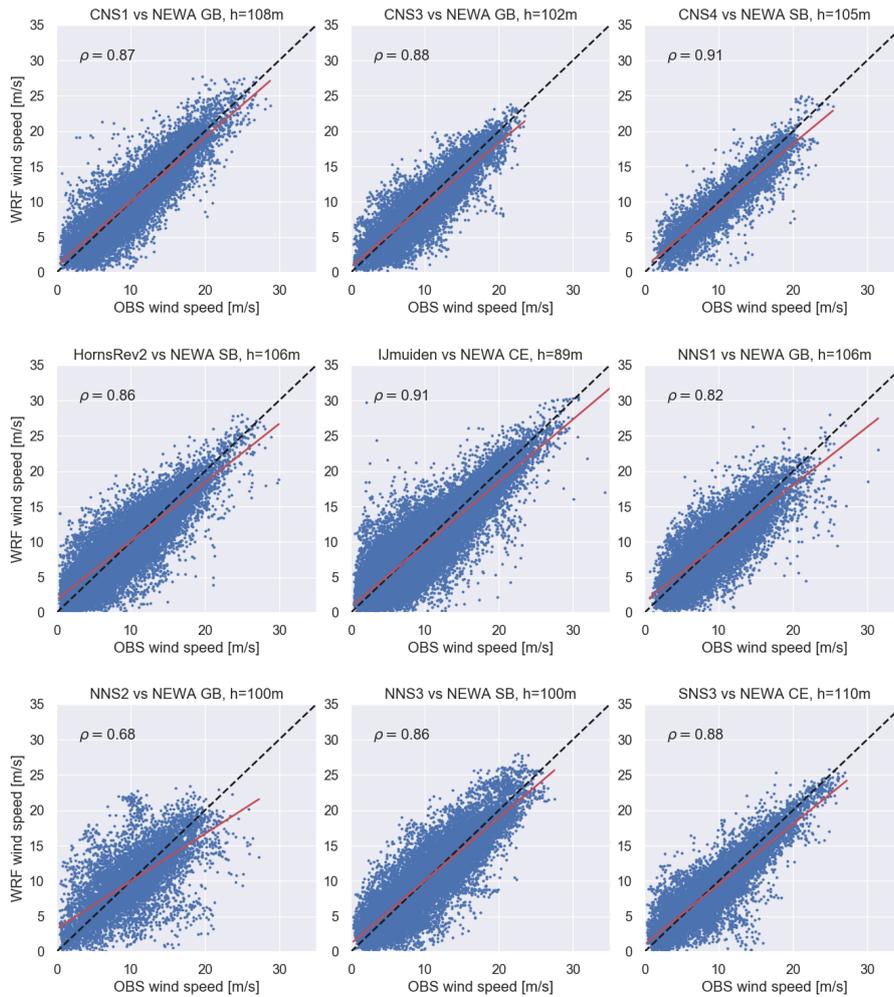


Figure 56: Comparison of observed (OBS) versus simulated (WRF) wind speeds for lidar sites offshore. The correlation between the two time series is shown in the upper right of each plot.

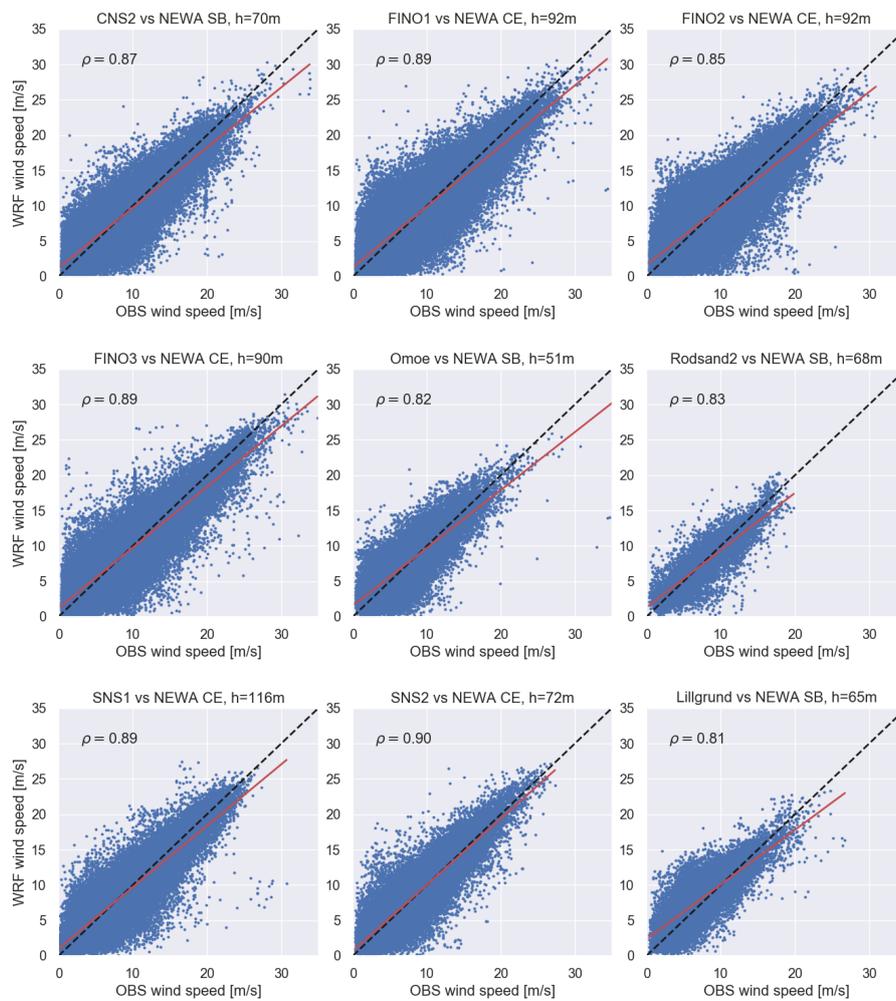


Figure 57: Comparison of observed (OBS) versus simulated (WRF) wind speeds for the mast sites offshore. The correlation between the two time series is shown in the upper right of each plot.

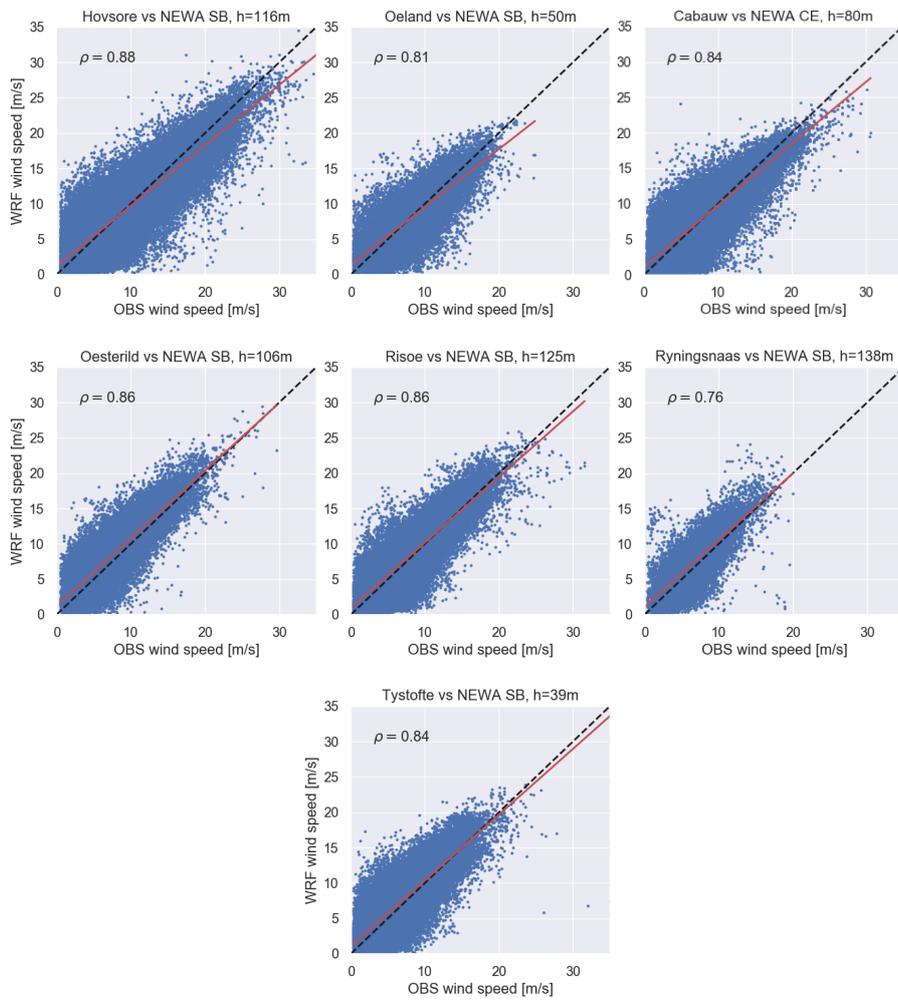


Figure 58: Comparison of observed (OBS) versus simulated (WRF) wind speeds for the coastal and land sites. The correlation between the two time series is shown in the upper right of each plot.

The next set of figures (Figures 59-61) show the distribution of the wind speed biases between the wind speed simulated by WRF and the observations at the various lidar and mast sites. The chosen heights are the same as those used in the scattergrams. Most offshore sites have very symmetric distributions of the errors, with standard deviations of 3.47 (CNS4) to 12.41 (NNS2). Once again, NNS2 has a short time series and poor data availability.

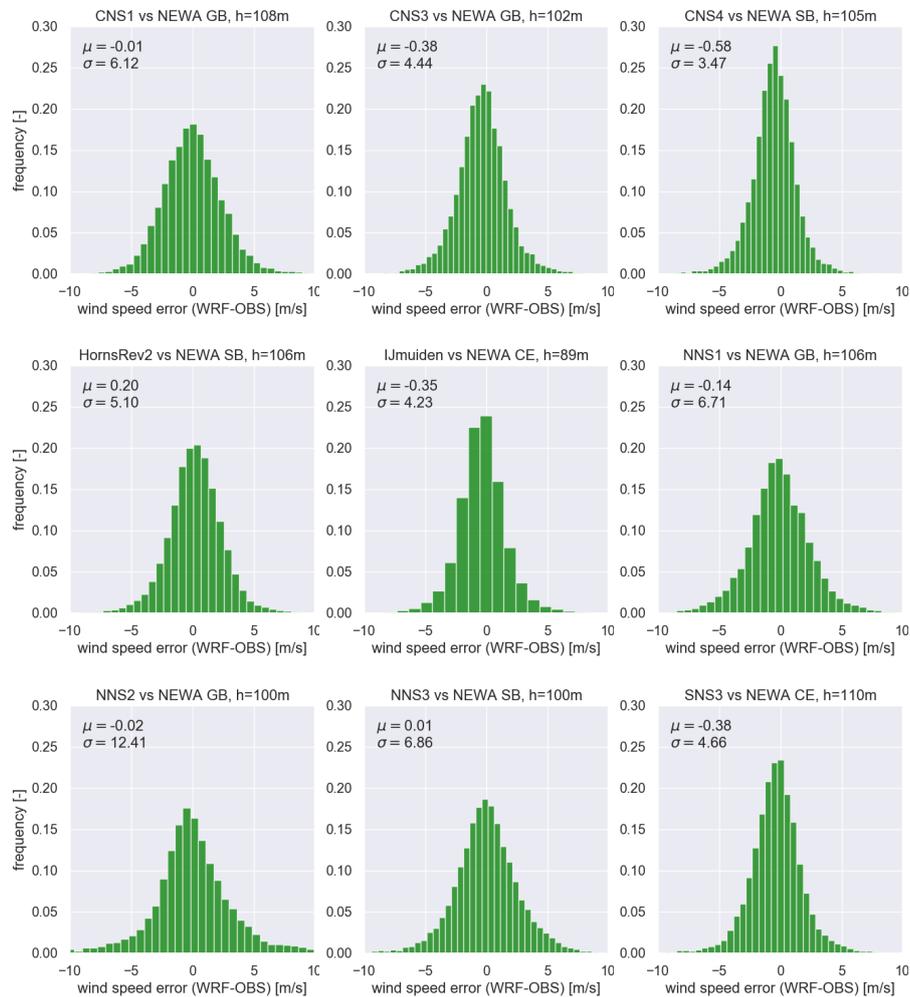


Figure 59: Distributions of the wind speed biases (WRF minus observations) for lidar offshore sites. The mean and standard deviation of the biases is shown in the upper left of each figure.

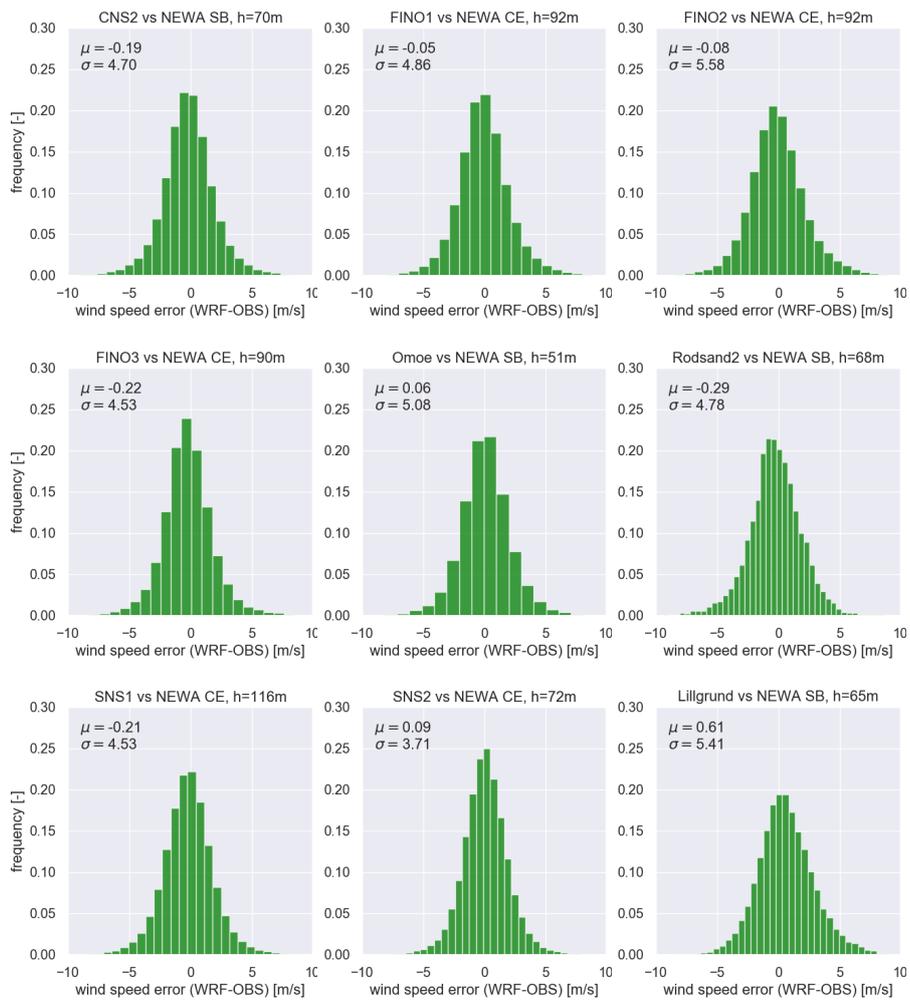


Figure 60: Distributions of the wind speed biases (WRF minus observations) for mast offshore sites. The mean and standard deviation of the biases is shown in the upper left of each figure.

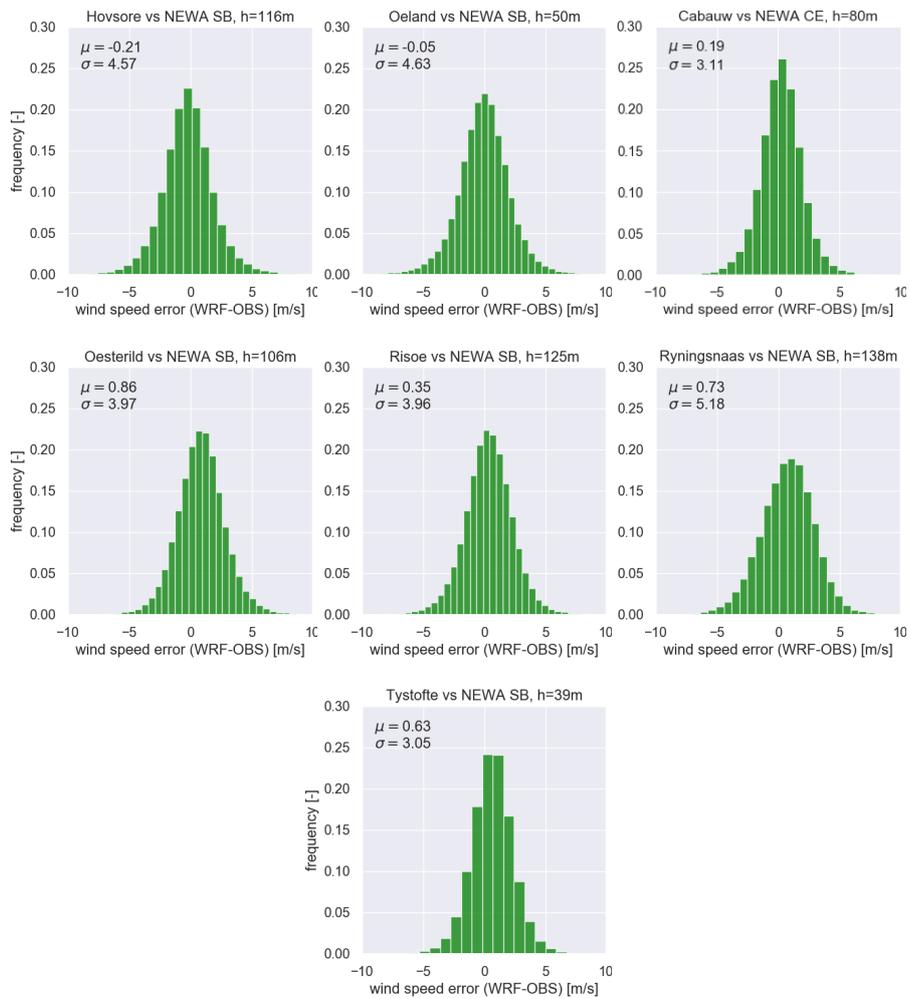


Figure 61: Distributions of the wind speed biases (WRF minus observations) for the coastal and land-based sites. The mean and standard deviation of the biases is shown in the upper left of each figure.

station name	height [m]	N	BIAS [%]	VRAT [-]	CORR [-]	RMSE [m s ⁻¹]	MAE [m s ⁻¹]
Børglum	31.50	215727	0.89	0.86	0.85	1.76	1.34
Cabauw	140.00	315225	-0.23	1.07	0.86	1.95	1.46
CNS1	108.00	14503	-0.12	1.11	0.87	2.47	1.89
CNS2	70.00	91299	-1.95	0.98	0.87	2.18	1.60
CNS3	102.00	14779	-3.96	1.22	0.88	2.14	1.60
CNS4	105.00	7267	-5.53	1.00	0.91	1.95	1.43
FINO1	91.50	204293	-0.55	0.91	0.89	2.20	1.63
FINO2	92.40	146145	-0.89	0.91	0.85	2.36	1.76
FINO3	90.00	144952	-2.25	0.92	0.89	2.14	1.55
HornsRev2	106.00	31610	2.07	0.92	0.86	2.27	1.70
Høvsøre	100.00	246128	-1.75	0.94	0.88	2.12	1.58
Ijmuiden	115.00	66551	-4.05	0.94	0.90	2.18	1.58
Lillgrund	65.00	17514	8.13	0.92	0.81	2.41	1.83
NNS1	105.50	15514	-1.46	1.07	0.82	2.59	1.95
NNS2	100.00	8507	-0.19	1.11	0.68	3.52	2.47
NNS3	100.00	20486	0.07	1.02	0.86	2.62	1.97
Öland	50.00	55550	-0.61	1.03	0.81	2.15	1.61
Østerild	106.00	64191	11.07	1.27	0.86	2.17	1.68
Omø	50.60	48257	0.80	1.00	0.82	2.26	1.57
Risø	125.20	110093	4.36	1.15	0.86	2.02	1.54
Rødsand2	68.00	7109	-3.46	0.95	0.83	2.20	1.66
Ryningsnäs	138.00	19073	10.26	1.49	0.76	2.39	1.86
SNS1	116.00	83915	-2.28	0.98	0.89	2.14	1.59
SNS2	72.50	74146	0.96	1.07	0.90	1.93	1.43
SNS3	110.00	15915	-4.02	0.95	0.88	2.19	1.60
Tystofte	39.00	260687	9.94	1.24	0.84	1.86	1.44

Table 15: Summary statistics for wind speed the comparison of the WRF simulations and the various observations. BIAS: WRF minus OBS, VRAT: Variance ratio WRF/OBS, RMSE: root mean square error, MAE: mean absolute error. The various rows are labelled according to location: offshore (light blue), coastal (light green) and land (beige).

The overall statistics of the validation of the WRF-simulated time series is presented in Table 15. Here we include the bias, ratio of the variances, and the standard RMSE and MAE. The bias is $< \pm 3\%$ at all offshore and coastal sites with good data availability (except for Lillgrund, where there is strong influence of the wind farm.).

The ratio of the variances shows interesting systematic behaviour: nearly 1 or < 1 at most offshore sites, and larger than 1 (1.07–1.49) at all land sites. The RMSE values range from 1.86 at Tystofte to 2.4 at the forest site of Ryningsnäs in all well-sampled sites.

In summary, the NEWA simulations compare quite well against wind speeds from tall masts and lidar offshore, except when the time series are short. At turbine-rotor-level (~ 100 m) the overall absolute biases are of the order of $< 3\%$ and correlations > 0.86 . For taller measurements (> 150 m), a systematic under-prediction of the mean wind speed by the NEWA simulations is seen in the evaluation, but it could be partially due to lidar tendency to sample more often at higher wind speeds, which are common at these heights.

3.4 Model evaluation using surface wind data

An ideal scenario in the evaluation of the NEWA model-chain would be to make use of a large and sufficient amount of tall mast data to understand how realistic simulations are at the hub height of a wind turbine, i.e., around 100 m above the ground. Unfortunately, as reported in previous sections, this type of data is frequently not available since providers are usually companies and wind farm developers that do not normally make the masts wind data publicly available.

Therefore, to assess the ability of the model strategy in reproducing the observed wind it is compulsory to complete comparisons with other available wind data. Usually surface wind datasets are free for research and they are more abundant. Thus, they can provide a good spatio-temporal coverage to explore the model skill in reproducing the observed wind.

Of course, one should be aware that the conclusion met about the validation of a model simulation at the surface or above are not necessarily identical and that the reasons why the simulation does not adequately describe the observed wind behaviour at 100 m height might not be the same than those that distort the ability of the model to reproduce the surface wind. Still, the validation at the surface is useful as it might suggest certain biases or misrepresentations that at some point become also relevant for the simulated wind above.

Therefore, this section's purpose is to illustrate how the regional simulation reproduces the observed wind at the surface. To this aim, a surface wind dataset over the whole European domain was specifically compiled for the purposes of the NEWA project and is presented in Section 3.4.1 with wind speed and direction values over a huge spatio-temporal domain. The validation of specific simulations over the SW (Iberian Peninsula) domain is presented in Section 3.4.2 while the NEWA production run simulated wind is compared to the surface observed wind in Section 3.4.3.

3.4.1 Surface wind data

The WiSED (Wind Surface European Database) data is originally a composite of wind speed and wind direction data from eight different wind datasets accounting for almost 13.000 observational sites, most of which lack an exhaustive quality control. At the time of the writing of this report, a comprehensive automated quality assurance procedure is being applied based on previous works by Lucio-Eceiza et al. (2018a,b).

Figure 62 illustrates the extension of the WiSED database after unifying all eight different sources of data. The panel on the left shows a diagram with the number of sites within the dataset and the period they cover, respectively. Some sites have considerably long records (they go back as far as the beginning of the 20th century) although the network becomes more dense after mid 20th century and the vast majority of sites have records after 1975. The right panel of Figure 62 represents the location of the 4.064 sites after compositing and unifying sites from all eight sources and their temporal coverage in colours. Symbols also refer to the length of the period covered by the observations at each site.

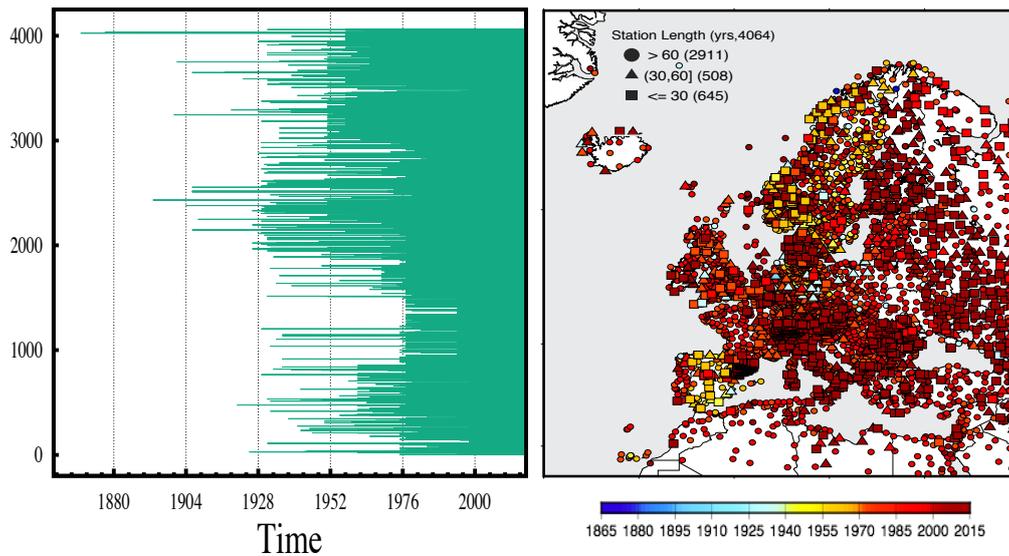


Figure 62: WiSED database. Left: Number of sites with their corresponding temporal length (initial year of records). Right: location of the 4.064 sites after compositing from all eight data sources. Symbols and colours refer to the temporal coverage at each site.

Figure 63 shows the WiSED mean wind calculated from the remaining 4.064 sites after unifying data from all sources. The most windy areas are located over northern Europe: western United Kingdom, Baltic regions and the Scandinavian Peninsula. Some quality issues seem to affect Romania and Bulgaria that evidence very high mean wind speeds. These issues are expected to be corrected after the quality control is completed.

For the validation of simulations over the IB domain a subset of WiSED stations belonging to the area of interest are used. Also, a reduced dataset of four tall mast wind records within the area has been used as well for the validation over the SW domain with heights varying from 10 up to 120 m. These are not represented in the maps presented above but will be part of the validation analysis in Section 3.4.2.

The WRF simulations used in this part of the analysis correspond to those presented in Section 2.3. The *reference simulation* is as therein, the simulation NH0M20W1 (see Table 7). In the present section it is denoted as NH0W1 for the sake of brevity. We also use the simulations RCOM20W1 and CLMM20W1, with different LSMs (RUC and CLM, respectively), for comparison with the reference case. The latter will be denoted as RU0W1 and CLM0W1 herein.

The next section is devoted to illustrate how the WRF simulation reproduces the wind field over the SW Iberian Peninsula domain at the surface provided by the WiSED observational records and at some tall mast series.

3.4.2 Regional model validation over the SW domain: Alaiz region

The reference simulation NH0W1 is compared to observations over the SW domain, in the north-eastern region of the Iberian Peninsula and southwestern France, centred over the wind farm Alaiz, in Navarre, a region of complex orography, surrounded to the south by the deep Ebro Valley that tends to accelerate flows in the NW-SE direction (known as the Cierzo winds), but also in the opposite direction, creating a warmer and more humid wind flow from the Mediterranean to the northwest, known as Bochorno winds, and to the north by the Pyrenees mountain range. Figure 64 (left) represents the mean wind speed bias (simulation minus observation) between the *reference simulation* and the WiSED observations that fall within the domain of Alaiz. This is represented by the coloured circles (squares) for the surface stations (tall mast records). The innermost symbol at each site corresponds to the reference NH0W1 simulation while the outer one illustrates the bias from the RU0W1 case for comparison. While the bias tends to be positive at

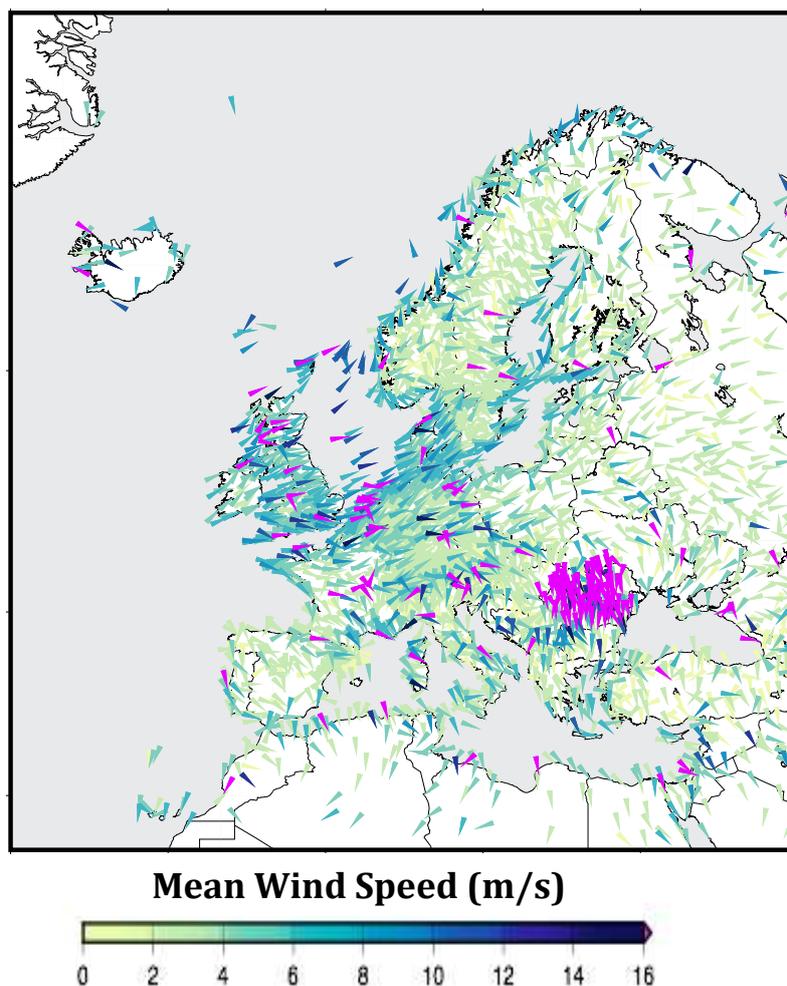


Figure 63: Mean wind speed over the European domain calculated using the 4.064 sites within the WiSED dataset, previous to the quality control procedure.

the surface it seems that the opposite behaviour can be expected when observations in height and the corresponding series from the simulations are compared. With the exception of some sites whose bias seems to be larger compared to the rest (the absolute bias can peak 4 m s^{-1} at some locations), the bias is comprised in the range $[-2, 2] \text{ m s}^{-1}$.

What calls the attention in Figure 64, left is that the colours of the inner and outer symbols tend to be the same for most of the sites. The latter implies that differences between simulations and observations are in general larger than differences between the two simulations compared. Therefore, it can be said that deviations with respect to observations are generally larger than the inter-model differences.

The inter-model differences (RU0W1-NH0W1) can be appreciated over the innermost domain of the two WRF simulations used in this analysis in Figure 64(right). It can be noticed that the differences range between -1.5 m s^{-1} and 1.5 m s^{-1} and they are very much connected with the topography of the terrain. Therefore, each of the LSM used in these two simulations seems to contribute to differences in the simulated wind depending on how each one resolves the thermodynamical processes involved in the land-surface interactions. Nevertheless, as commented before, these differences are sensibly smaller than the bias between the simulations and the observations.

The bias between the two simulations seems to be less important with height, where the influence of the soil-atmosphere interactions diminishes. This can be observed both in left and right panels of Figure 65, where differences between the wind simulated by RU0W1 and NH0W1

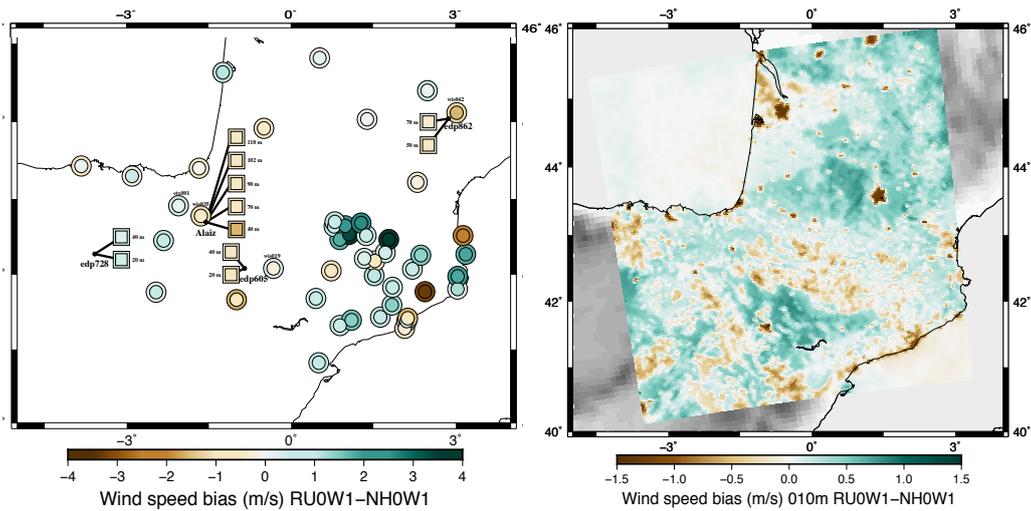


Figure 64: Surface wind speed bias between simulations and observations over the Alai domain in the Iberian Peninsula. Left: Bias between the reference simulation NH0W1 (sensitivity experiment simulation RU0W1) and observed WiSED wind at the surface is represented by the innermost (outer) circles while bias between the two simulations and records from the four tall masts acquired in this region are represented by the coloured (see colour scale) squares. Right: Mean wind difference between NH0W1 and RU0W1 simulations over the innermost WRF domain.

calculated at 120 m height are represented.

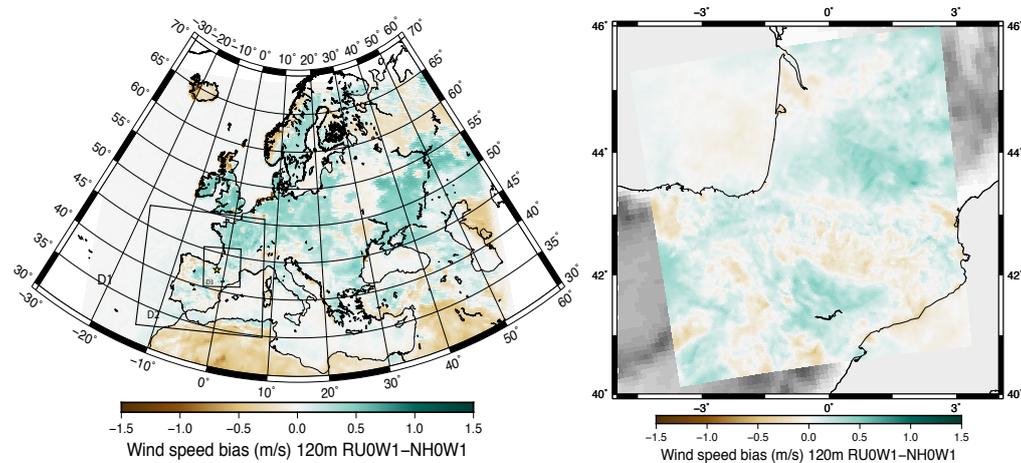


Figure 65: Wind speed bias at 120 m between the NH0W1 and the RU0W1 simulations over the outermost (D1) European domain (left) and the innermost (D3) Alai domain (right).

A closer look at the differences between observations and simulation is provided in Figure 66, where the temporal variability of the daily time series for three sites within the observational database is represented. Figure 66a corresponds to measurements at 40 m height from the mast located at the Alai site (see map of Figure 64, left). We can appreciate that during the period represented differences between the two simulations (the reference one, NH0W1 and the CLM case, CLM0W1) are smaller than those with the observed wind speed in blue. The simulations tend to show some reduced levels of variance compared to observations, possibly as a result of the smoothing effect from the model. This effect is even more noticeable in Figure 66b, where differences among simulations are even smaller if compared with the observations. In this case, surface measurements at the edp862 location (see map of Figure 64, left) are represented.

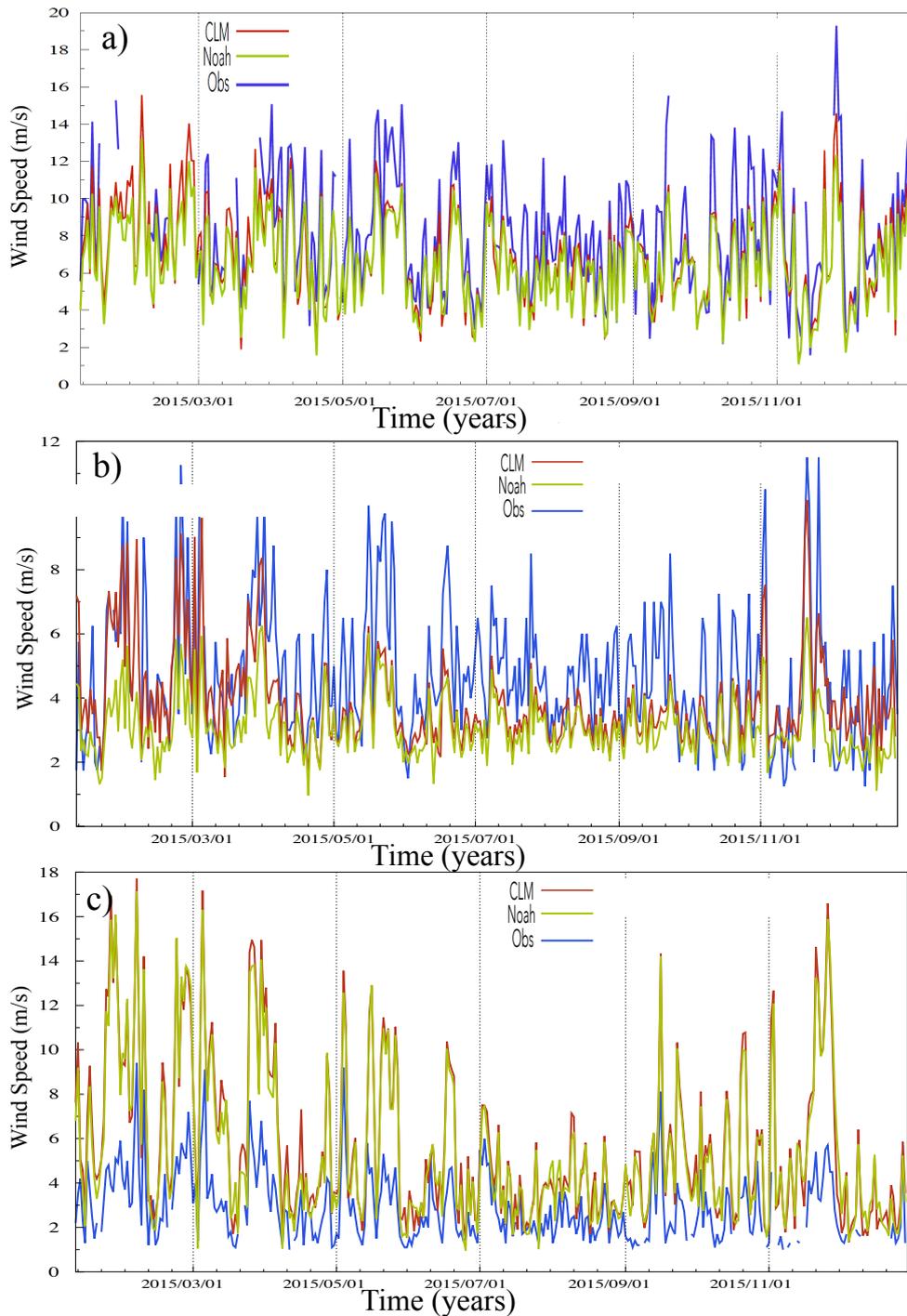


Figure 66: Daily wind series from simulations using two different LSMs (Noah in green and CLM in red) and from observations (blue) at three different sites over the Alaiz domain.

Situations like that in Figure 66c where both simulations overestimate the observed wind are as well possible. We have selected one of the sites with largest bias from the map in Figure 64. The performance of the model might depend therefore on a specific time interval or site where the complexity of the terrain hinders an adequate representation of the observed wind.

The similarity between the regional model simulations and the observations can be quantified as well by using the Brier skill score (BSS) statistics that provides an estimate of the observational variance that the model accounts for (von Storch and Zwiers, 1999). Figure 67 shows BSS values

at all sites within the Alaiz domain. Although values are comprised within the range $[0.5, -0.5]$ m s^{-1} , they are generally small all over the region and those sites with larger BSS estimates are not necessarily the same as those showing larger wind speed bias in Figure 64(left). Negative BSS values indicate that observations tend to show larger variance than the simulations. This is the behaviour we can observe in three of the tall mast BSS estimates.

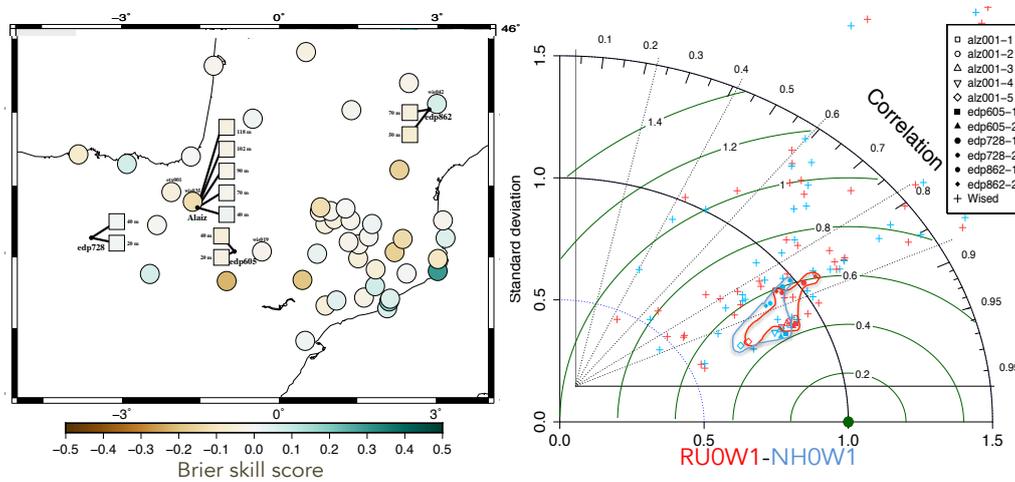


Figure 67: Validation statistics over the Alaiz domain: Brier Skill Score (left) and Taylor diagram (right) of the *reference simulation* (blue). The Taylor diagram includes as well statistics for the simulation with the RUC LSM (red). Crosses in the Taylor diagram represents statistics at WiSED sites while the rest of symbols identify the values obtained at the tall mast locations.

Additionally, in Figure 67(right) a Taylor diagram showing the statistics at each site when comparing simulations (*reference* in red and RUC LSM in blue) with observation is represented. This is a polar diagram where the angle is indicative of the correlation value and the radial coordinate accounts for the standard deviation ratio between estimations and observations at each location. The crosses represents statistics at WiSED sites while the rest of symbols identify the values obtained at the tall mast locations. In the case of the WiSED sites the performance of the two simulations is very similar. Correlation values between daily observations and simulations are well spread between 0.4 and 0.85. In the case of the mast wind, the correlations at the different heights of the mast are higher (correlation values between 0.85 and 0.95), illustrating that the complexity of the terrain over the region does not affect the simulation of the wind speed with height. Also the standard deviation values are much comprised and closer to the 1.0 (perfect estimation) value in the case of the mast measurements compared to the surface WiSED observations. Nonetheless, in this last case, the majority of sites evidence a standard deviation ratio close to 1.0. It can be said that the Taylor diagram shows that the standard deviation ratios are slightly better for RUC (red) land surface model.

Although the WRF model is able to realistically reproduce the wind field over the region, some biases and issues regarding the levels of variance reproduced are also detected at the local scale, that is, depending on the site. It seems in addition that for the model is easier to simulate the correct levels of variability in height compared to sites with wind at the surface. This is reasonable to understand since the complexity of the terrain and the representation of the land-surface processes might be behind the difficulties of the regional model to capture the variability of the wind at the surface in specific sites comparatively with the tower records, where the influence of the soil is certainly diminished. Nevertheless, it can be said that differences between observations and simulation, although larger than inter-model differences, are still in an acceptable range so that the quality of the simulations is not compromised.

The next section provides a hint of how the so-called NEWA production run performs over the whole European region.

3.4.3 WRF validation over the European domain

In this section some general ideas about the ability of the WRF model production run (as described in Witha et al. (2019)) to reproduce the wind field at the surface WiSED stations and at several tall masts are presented. For information about the tall mast wind data that is used in this section the reader is addressed to Section 3.2.

Figure 68 illustrates the mean wind field at the surface from the WiSED database (circles) and the mean wind values at the different heights from the tall masts (squares, triangle and triangle inverted depending on the difference between observation and simulation heights, the observation heights are also indicated). The most populated area is central Europe. As mentioned in Section 3.4.1 the windiest areas are northern coastal regions (United Kingdom and Scandinavia), where mean wind speed values at the surface can be as large as at the tall masts.

These data have been compared to the NEWA production run simulation in order to extrapolate the level of ability of the official WRF model configuration to reproduce the wind at the surface.

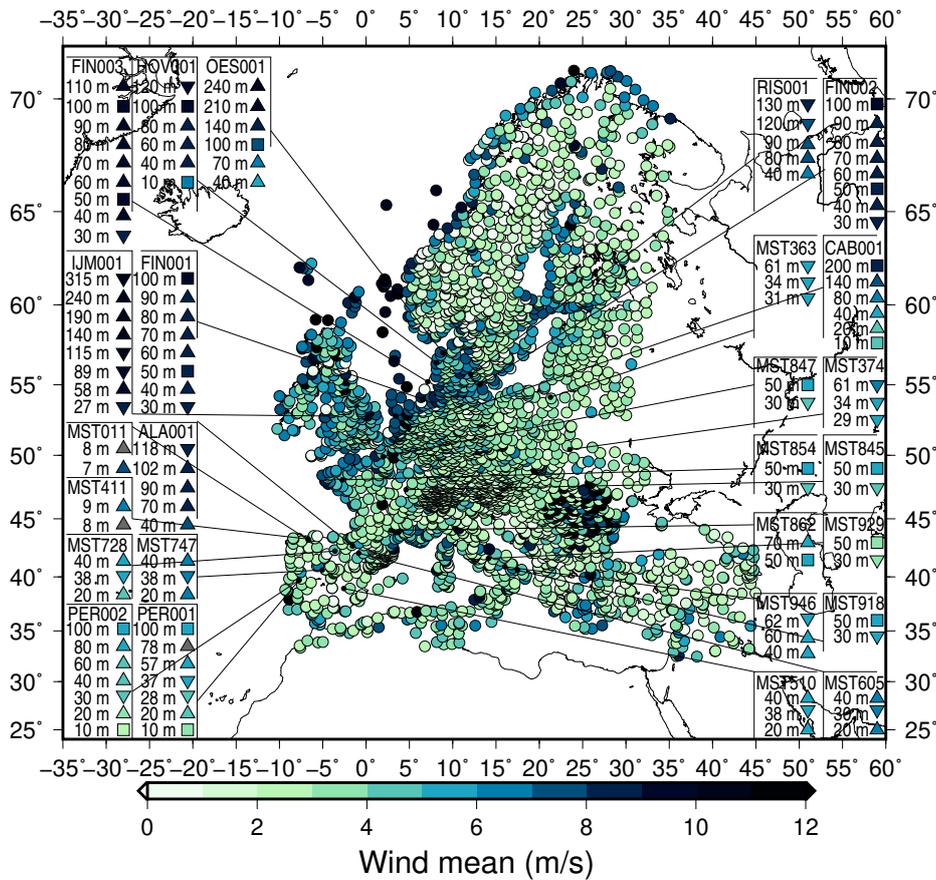


Figure 68: Mean wind speed calculated from the WiSED database at the surface and the tall masts over the whole European domain.

Figure 69 shows the mean wind speed bias calculated as the difference between the simulated and the observed wind at each grid point co-located with the observational sites. Positive values, indicating an overestimation of the observed wind by the simulation, are dominant over the whole domain. This can be partially a quality control issue regarding observed data, although an overestimation from the model cannot be disregarded. The bias at the masts tends to be negative or near to zero in general with height but they show a trend towards positive biases closer to the surface.

The correlation between simulations and observations at each grid point co-located with the observational sites has also been calculated as an estimate of the co-variability between both. Correlation values are represented in Figure 70. The map indicates that generally values are high

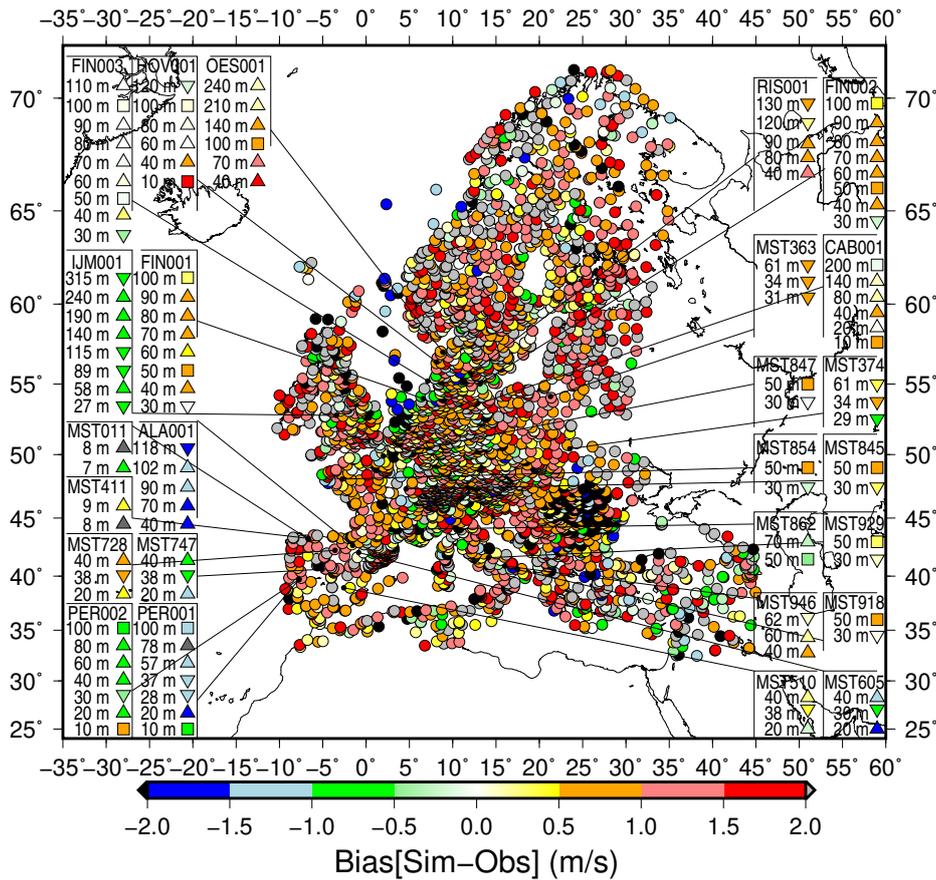


Figure 69: Mean wind bias (simulation minus observation) at the surface WiSED stations (circles) and the tall masts acquired (squares, triangle if the difference between the simulation and observation height is less than 10 m and inverted triangle if larger, height indicated as well) over the European domain. Simulations are obtained for each grid point co-located with the observational sites.

(above 0.5) over the whole European domain with the highest values in the western EU. The tall masts tend to show also high correlation values that do not necessarily change with height.

Finally, the root mean square error (RMSE) between the simulation and the observations is presented in Figure 71. The colour scale has been adapted for the RMSE. We can appreciate larger RMSE values over coastal areas in Europe, coincident to a great extent with the pattern shown by the absolute bias in Figure 69.

Thus, apparently the co-variance, given by the correlation values, does not necessary coincide with the bias and RMSE values, and high correlations proved to be compatible with larger errors from the simulations when representing the levels of variance of the wind comparatively with the observations.

As mentioned before, although some issues observed throughout the statistics presented in this section can be related to the quality of the observations in some regions, it cannot be disregarded that the simulation performs better in terms of correlation over the northern coasts and the Scandinavian Peninsula, where the wind speed has been found also to be larger.

3.4.4 Conclusions

In this section the regional model part of the model-chain has been analysed compared to surface wind data (10 m height) as well as with some tall mast measurements with heights ranging from 10 m to 240 m.

Two *foci* of attention have been discussed in this section. On the one hand, the reference sim-

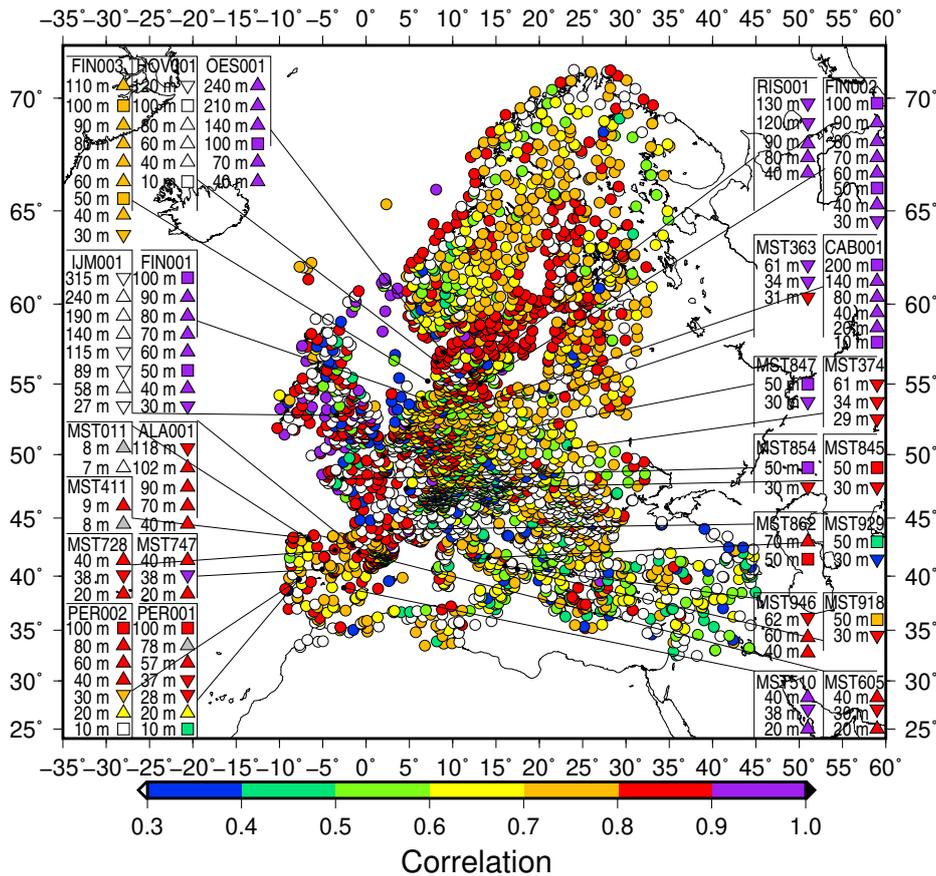


Figure 70: As in Figure 69 but for the correlation between the observation and the simulation.

ulation over the so-called Alaiz domain, corresponding to the northeastern region of the Iberian Peninsula (SW domain, see tiles in Figure 17 from Section 2.3) was evaluated and compared with some alternative simulations that differed in the LSM used in each case. Comparisons between simulation and observations revealed that despite a reasonable ability of the model to realistically reproduce the observed wind variability, the levels of variance simulated are not always as close to the observed ones, especially at the surface sites as expected. Differences between the different simulations at the local scale are slight, although simulated standard deviation ratios seem to be consistently better for the RUC model comparatively with the reference LSM used (Noah). In general, it can be said, that differences between observations and simulations at the local scale are larger than the model spread when using a pool of different simulations to evaluate the sensitivity of the model to changes in its configuration.

On the other hand, an evaluation of the skill of the official NEWA production run configuration has been presented for the whole European domain in this section. The regional model proved ability to represent the observations either at the surface or at the tall mast locations included in this part of the analysis. Statistics at the towers tend to overcome those at the surface, an expected feature since records in height are free of the surface influence and the topography that adds complexity to the simulation of the wind field. The WRF model seems to have larger skill to capture the variability of the wind over the windiest regions (northern coasts, the Scandinavian Peninsula and central Europe) although the levels of variance in these cases are not necessarily fully in agreement with those of the observations. The simulated wind at the masts, as said, seems to outperform that over the surface stations and does not seem to improve as the height increases to the top of the tower, rather statistics are homogeneous across the different heights recorded.

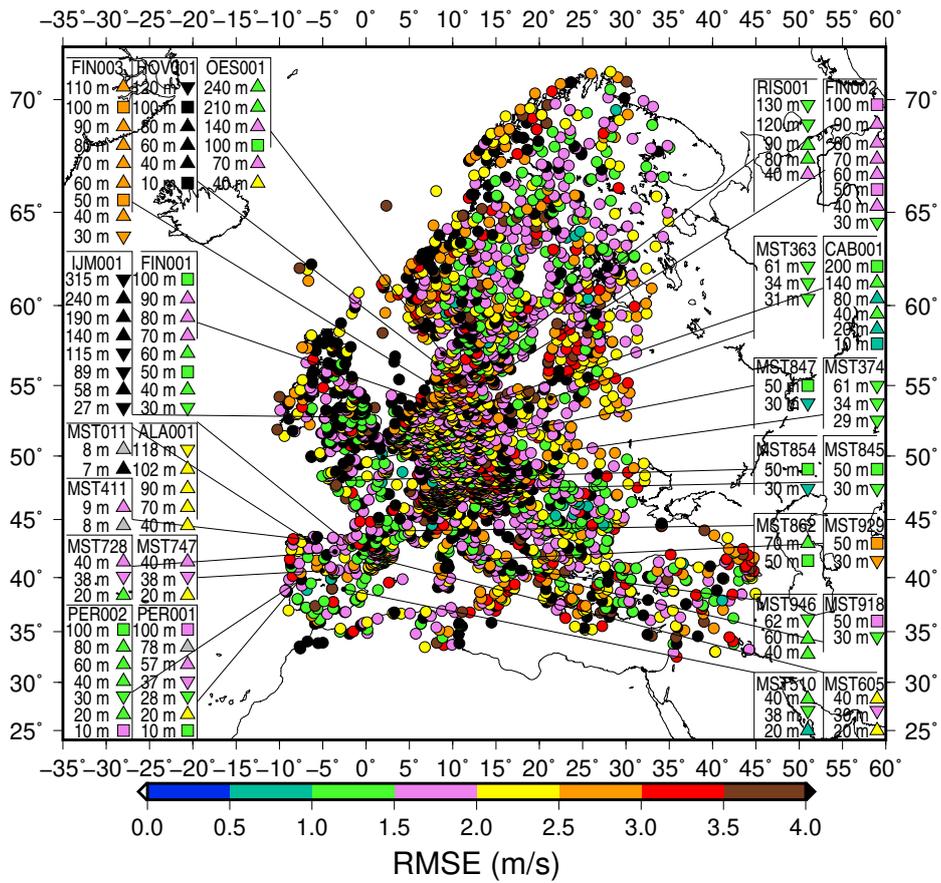


Figure 71: As in Figure 69 but for the RMSE values between the observation and the simulation.

3.5 Model evaluation using satellite wind data

The strength of satellite wind fields lies in their capability to monitor large spatial domains over extensive periods of time although the sampling frequency achieved is poor compared to the sampling frequencies of typical *in-situ* sensors, i.e. 10 minute averages, or numerical models, i.e. hourly. Satellite observations of the wind over the ocean surface can prove valuable as indicators of the resource distribution and provide information about the areas where high-resolution mesoscale model experiments can be performed.

Wind speed at 10 m above the ocean surface is routinely retrieved from space-borne radars, i.e. Synthetic Aperture Radars (SAR) and scatterometers. These radar instruments are sensitive to cm-scale waves generated at the sea surface due to the instantaneous wind stress. The radar pulses are scattered from the ocean surface back to the instruments and the amount of back-scattered signal per unit area, i.e. the normalised radar cross section (NRCS), depends on the size and geometry of roughness elements on the scale of the radar wavelength at the Earth surface. In the case of a smooth surface, the returned NRCS is low due to reflection of the radar pulse away from the instrument. Roughness elements, generated by the surface wind stress as the wind over the ocean increases, increase the signal back-scattered to the instrument. The empirical relationships used for the retrieval of wind speed and direction, i.e. the Geophysical Model Functions (GMFs), are derived from the relation of the NRCS, the wind speed and direction at 10 m and the radar viewing geometry. GMFs are traditionally sensor or mission specific, and so far have been derived using the Equivalent Neutral Wind (ENW) notation (Liu and Tang, 1996) – the wind at 10 m assuming neutral atmospheric stratification. Due to multiple looks over the same area, scatterometers can retrieve the wind speed and direction but their spatial resolution is coarser than SAR winds (Karagali et al., 2013b). The single look mode of SAR requires the *a priori* knowledge of wind direction to retrieve the wind speed. Global atmospheric models can provide wind directions to be utilised for operational SAR wind retrievals.

SAR wind retrievals have been used to identify and study wind farm wakes offshore, e.g. in Hasager et al. (2015c). Wind resource assessment from Earth Observation ocean surface winds has been performed during the past decade, especially focusing on the northern European Seas. Winds from the Quick Scatterometer (QuikSCAT) and the European Remote-Sensing Satellite (ERS) SAR were used in Hasager et al. (2008) to perform an analysis of the wind resources in the North Sea, highlighting the applicability of SAR for local-scale and that of QuikSCAT for basin-scale studies. The amount of required ocean wind field retrievals to achieve accurate resource statistics was investigated in Hasager et al. (2008). The full QuikSCAT archive was used in Horstmann et al. (2004); Karagali et al. (2013a, 2014) to perform validation, resource assessment and long-term characterisation of the surface winds, especially their spatial variability compared to modelled wind fields. The full QuikSCAT and part of the Advanced Scatterometer (ASCAT) archive were used in Hasager et al. (2015b), for validation with *in-situ* stations and to demonstrate the potential for the combined use of scatterometer observations from different platforms. Wind retrievals from the Environmental Satellite (Envisat) Advanced SAR (ASAR) have been used to perform wind resource estimations for various regions including Iceland (Hasager et al., 2015a), China (Chang et al., 2014), the Great Lakes (Doubrawa et al., 2015), and the Baltic Sea (Hasager et al., 2011) amongst others.

The advantages of satellite ocean surface winds over modelled winds are due to the higher effective spatial resolution of the former. Studies have shown deficiencies in kinetic energy of modelled winds when compared to the ERS (Chin et al., 1998; Halpern et al., 1999) and QuikSCAT (Zecchetto and De Biasio, 2003). Recently, Karagali et al. (2013b) examined the spectral characteristics of Envisat ASAR and QuikSCAT wind retrievals and showed the contribution of each sensor to describe spatial scales of different sizes. While the 10 m wind information can be resolved with satisfactory accuracy given the design characteristics of such space-borne sensors, atmospheric levels relevant for wind turbine operation are much higher and the extrapolation from 10 m depends on the atmospheric stability. Studies, e.g. Kara et al. (2008) have shown that

the marine boundary layer over the global ocean is, on average, slightly unstable. Badger et al. (2016) performed an extrapolation of the mean wind from SAR at turbine relevant heights, using a long term stability corrected wind profile derived from mesoscale model simulations.

For the purposes of NEWA, the 10-m wind information from SAR and scatterometers from 2002 to 2018 was used to derive a mean wind "climate" over the European Seas. Since for wind turbine hub-heights offshore levels higher than 10 m are more relevant, extrapolation of the long-term mean winds to higher atmospheric levels was performed according to Badger et al. (2016). Thus, the logarithmic wind profile was used in two versions, i) assuming neutral stability (referred to as ENW) and ii) using a 10-year long stability correction (referred to as SDW), derived from the mesoscale model WRF described in Nuño Martinez et al. (2018).

Ideally, a new stability correction will be derived by the NEWA WRF simulations to be used for extrapolation of the satellite wind atlases, for direct comparisons. Furthermore, comparisons between the satellite derived offshore winds and the WRF model outputs were used to assess the spatial variability of the mean wind speed at different heights and the impact of different WRF tuning options on the modelled outputs.

3.5.1 Satellite Winds

ASCAT The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) operates a series of polar orbiting meteorological satellites (MetOp), two of which are already in operational phase and the third, launched in October 2018, is in demonstration phase. ASCAT is the C-band scatterometer on board the MetOp-A/B/C platforms, measuring in two 500 km-wide swaths. The wind product used in the present study is the newly available 12.5 km Coastal Stress Equivalent Wind, obtained through the Copernicus Marine Environmental Monitoring Service (CMEMS, <http://marine.copernicus.eu/>, De Kloe et al. (2017)).

SAR SAR wind retrievals are routinely performed at DTU Wind Energy, using a processing chain built around the SAR Ocean Products System (SAROPS) by the NOAA Center for Satellite Applications and Research (STAR), US National Ice Center and Johns Hopkins University, Applied Physics Laboratory Monaldo et al. (2014). The SAR wind retrievals are performed using the CMOD5.N GMF Hersbach (2010), thus are representative of the Equivalent Neutral Wind (ENW), at a 600 m resolution to eliminate effects of random noise and of surface inclination due to longer-period ocean waves. These are available from <https://satwinds.windenergy.dtu.dk/>. The derived SAR mean wind resource used in this study has a 2 km resolution.

The European Space Agency (ESA) platform Envisat was launched in 2002 and its data acquisition was terminated in April 2012. It carried, amongst other instruments, an Advanced Synthetic Aperture Radar (ASAR) operating in C-band with several different modes. The majority of Envisat ASAR scenes in this archive have been acquired in Wide Swath mode (WSM) or Global Monitoring mode (GMM) with a swath width of 400 km and variable lengths. The Image mode (IM) and Alternating Polarisation mode (APP) are suitable for ocean wind retrieval over smaller areas due to their swath width of 100 km. This archive holds a limited collection of processed data based on these modes.

Another ESA mission, Sentinel-1 consists of two separate platforms, each with a C-band SAR (5.3 GHz) on board. Sentinel-1A was launched on 3 April 2014 and Sentinel-1B on 25 April 2016. Both sensors are currently operational.

3.5.2 WRF simulations

For comparisons with satellite wind fields, the NEWA test simulations performed 2015 using the method of Hahmann et al. (2015) were used; simulation experiments used here had the same daily initialisation at 00:00 GMT and a difference of the selected Planetary Boundary Layer (PBL) scheme, i.e. MYNN and YSU. These two simulations were named as MYNL61S1 and YSUL61S1 and will be referred to using these code names. Furthermore, the NEWA production simulation for mean wind at 100 m has been used for comparisons with the mean winds from

ASCAT and SAR, extrapolated at 100 m.

3.5.3 Results

For the test experiments of 2015, the mean bias between the WRF simulations and ASCAT (WRF-ASCAT) is shown at the top row of Figure 72. While both WRF simulations provide similar spatial variability of the bias with ASCAT, MYNL61S1 (left) resulted in higher positive differences in most of the North Sea basin and the small part of the Baltic Sea covered by this domain.

Exceptions were found in areas with intensively higher ASCAT winds, such as some areas offshore from the Netherlands and Germany. In these cases, higher ASCAT winds can be partially considered as artefacts due to i) the consistent presence of ships offshore Rotterdam port and ii) the high concentration of wind farms north of Bremerhaven. Both ships and wind turbines, considered as hard targets, can increase the signal back-scattered to the instrument and thus the derived wind speed magnitude. Existing studies, e.g. Badger et al. (2016); Karagali et al. (2013b); Peña Diaz et al. (2011), comparing satellite winds with WRF model outputs have found lack of spatio-temporal variability in the model compared to the wind retrievals.

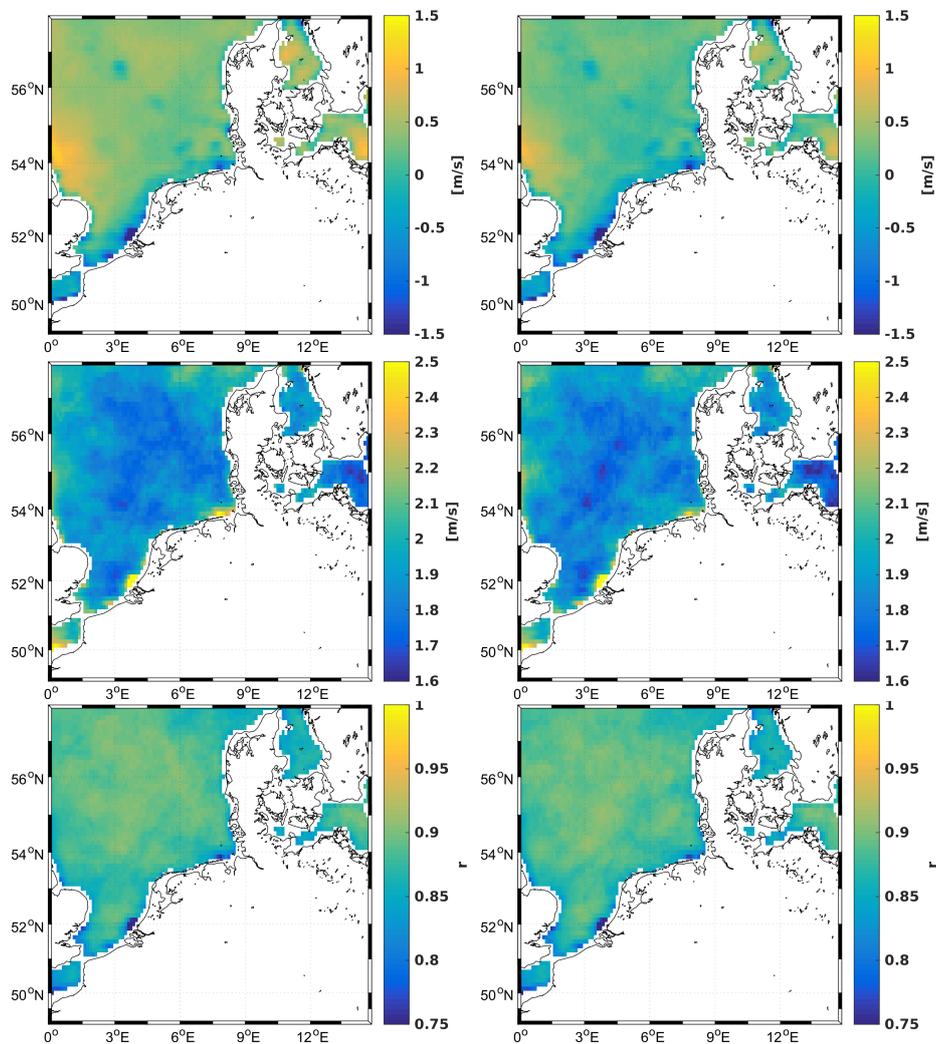


Figure 72: Comparison of the 1 year mean wind speed at 10 m: mean bias (top), standard deviation (middle) and correlation r (bottom) between ASCAT and MYNL61S1 (left), ASCAT and YSUL61S1 (right).

The standard deviation, σ , of the WRF-ASCAT estimates, shown in the middle row Figure

72, indicated lower values in the central part of the North Sea basin for both MYNL61S1 and YSUL61S1 experiments, with a slight tendency of the YSUL61S1 simulation for lower σ values in larger part of the domain. The correlation coefficient, r , shown at the bottom row indicated overall values of 0.9 and higher with exceptions including the coastlines and offshore from Rotterdam and Bremerhaven.

The percentages of unstable conditions were derived from the surface-layer Monin-Obukhov stability parameter, L , an output field from the MYNL61S1 simulation (2015) and are shown in the left panel of Figure 73. Over the water conditions were found frequently unstable, especially closer to the coast lines and in particular off the shores of the Netherlands, Belgium and Germany. As mentioned before, equivalent neutral winds (ENW) from satellites will be higher than naturally occurring winds under unstable conditions.

To attempt an interpretation regarding the frequency of unstable conditions over the North Sea, findings from Karagali and Høyer (2014) were used. The right panel of Figure 73, adapted from Karagali and Høyer (2014), shows an estimate of the maximum diurnal variability signal of the sea surface temperature (SST). This signal was defined as the mean SST during a day (24 h) minus the foundation SST of that day, i.e. the SST under well mixed conditions or during night-time. This provided a daily mean diurnal warming estimate which was averaged monthly and Figure 73 shows the maximum of these values.

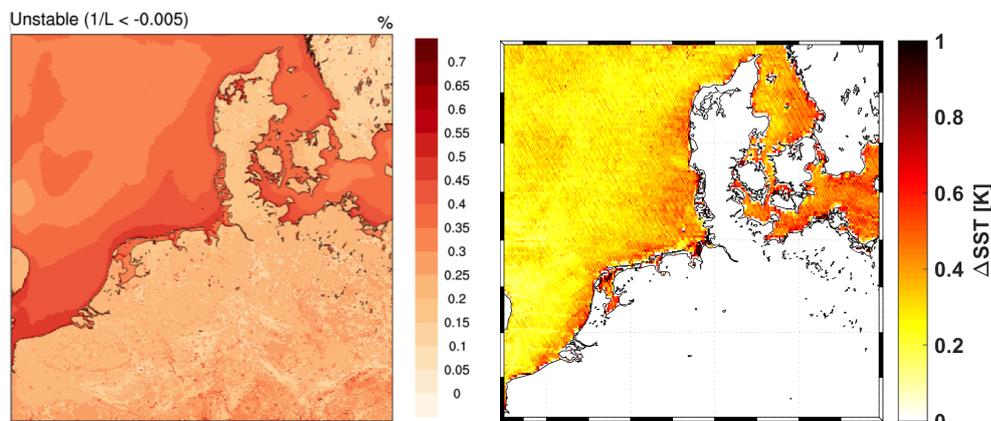


Figure 73: Fraction of unstable conditions over the North Sea during 2015 as simulated by WRF during the sensitivity experiment using MYNL61S1 (left). Mean of the sea surface temperature maximum diurnal variability from 2006 to 2012 (right).

The spatial match of most frequently unstable conditions from WRF corresponds well to areas with highest amplitudes of the diurnal temperature cycle. This can be explained by the fact that large and sudden increases of the surface temperature occurring under clear sky and up to moderate wind conditions of $\sim 6 \text{ m s}^{-1}$ (Karagali et al., 2012), result in unstable conditions. This is a good indicator that a proper and accurate SST boundary condition should be used for the simulations.

Mean winds at 100 m from the WRF production run (1989-2018) were compared with satellite derived winds from ASCAT and SAR (Figure 74). The top left panel shows the ASCAT ENW mean wind at 100 m (extrapolated assuming neutral conditions) minus the mean wind from the production run. Near zero biases are recorded in the central North Sea and the Black Sea while more intense positive biases, indicating higher ASCAT mean winds, are found in the largest part of the Mediterranean and the North Atlantic. More intense negative biases, i.e. higher NEWA mean winds, were recorded in the Baltic Sea, the Mediterranean side of the Strait of Gibraltar and off the east coast of England. These biases are significantly reduced when examining the difference between ASCAT SDW (extrapolated using a stability correction) and NEWA, see Figure 74 top right panel. Note how biases are significantly positive, i.e. higher ASCAT winds, close to coastlines and offshore from extreme topographic features, e.g. Norwegian west coast, the Aegean islands of Greece, the Ligurian coast of Italy, the Gulf of Lions in France.

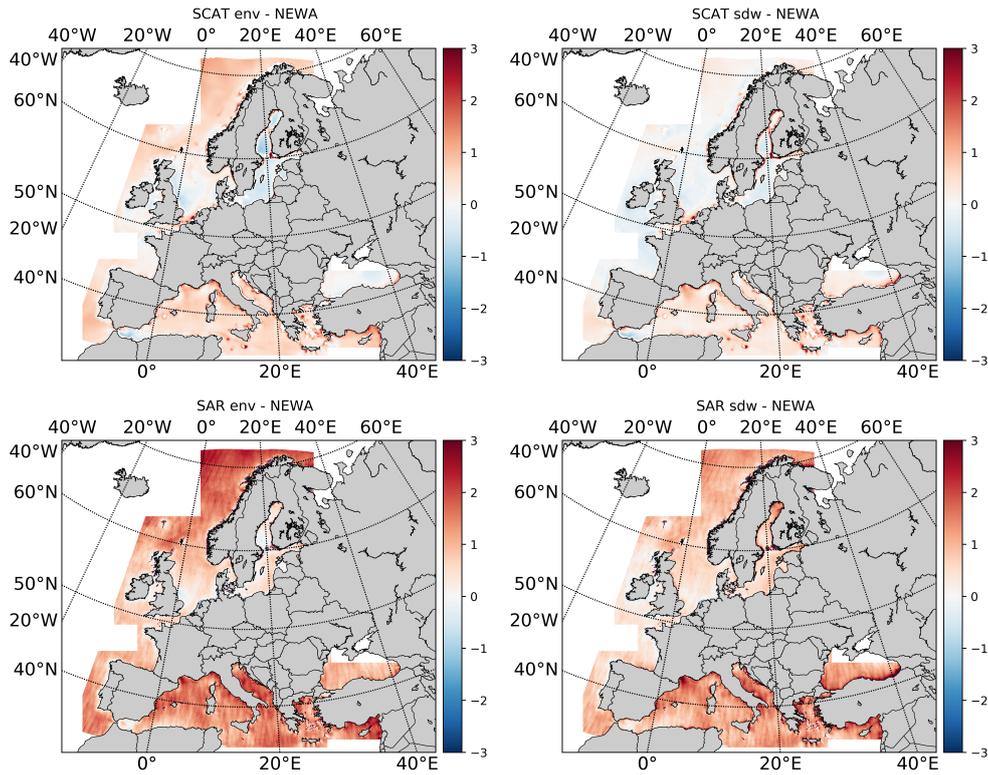


Figure 74: Comparison of the 30 year mean wind speed (1989-2018) at 100 m height from NEWA and ASCAT (top), and SAR (bottom).

The bottom left panel of Figure 74 shows the SAR ENW mean wind at 100 m (extrapolated assuming neutral conditions) minus the mean wind from the production run. Large positive biases are identified in most of the domain of interest, indicating higher SAR mean winds. These biases are significantly reduced when examining the difference between SAR SDW (extrapolated using a stability correction) and NEWA, see Figure 74 bottom right panel, although they do remain high in the Mediterranean basin and higher compared to the ASCAT case.

SAR wind retrievals are currently being re-calibrated following the method of Badger et al. (2019). Once inter-calibration amongst different SAR sensors is finalised, the resource maps will be re-evaluated and from results found already, it is expected to have a reduction in the biases with the NEWA production run. Nonetheless, SAR winds are valuable for characterising small scale flow features that are typically not resolved by mesoscale models, e.g. analysing the interaction between large-scale flow and orography around Crete (Hasager et al., 2019).

Furthermore, analysis of the conditional error statistics for 10 m winds from WRF ensemble members and ASCAT is being conducted within the framework of a MSc thesis (Ribauda, 2019) and results derived are informative regarding the behaviour of various options used to produce the ensemble members.

3.6 Model comparison with ERA5 reanalysis data and Global Wind Atlas data

To show the added value of the NEWA mesoscale wind atlas we compared the long-term wind climate with that of the ERA5 reanalysis data (which was used to force the WRF simulations that form the basis of the mesoscale wind atlas). Figure 75 presents a visual comparison of the 30 year (1989–2018) mean wind speed at 100 m height over Europe derived from ERA5 and NEWA. The mean wind speed over sea is very similar for both datasets. Both the vast open seas of the North Atlantic as well as the narrow Baltic Sea match very well, also the main flow features over the Mediterranean Sea can be seen in both pictures. Over land NEWA shows much higher wind speeds than ERA5. This is especially true for mountainous regions which are much better resolved in NEWA (see e.g. the Scandinavian Mountains or the Alps) but also for most of the other regions, even over very flat terrain. Note that the spatial resolution of NEWA is about 3 km while the resolution of the ERA5 reanalysis is about 30 km (0.3°).

Figure 76 shows the relative difference between both datasets ($(\text{ERA5} - \text{NEWA}) / \text{NEWA}$). To derive it, both datasets have been interpolated to a common $3 \text{ km} \times 3 \text{ km}$ grid. As discussed before, the relative differences over the seas are very low with a slight tendency towards lower wind speeds in the NEWA dataset. Nearly everywhere over land, the NEWA data shows higher wind speeds, mostly 10–20% higher than ERA5. Over mountainous regions the differences can be as large as 50%.

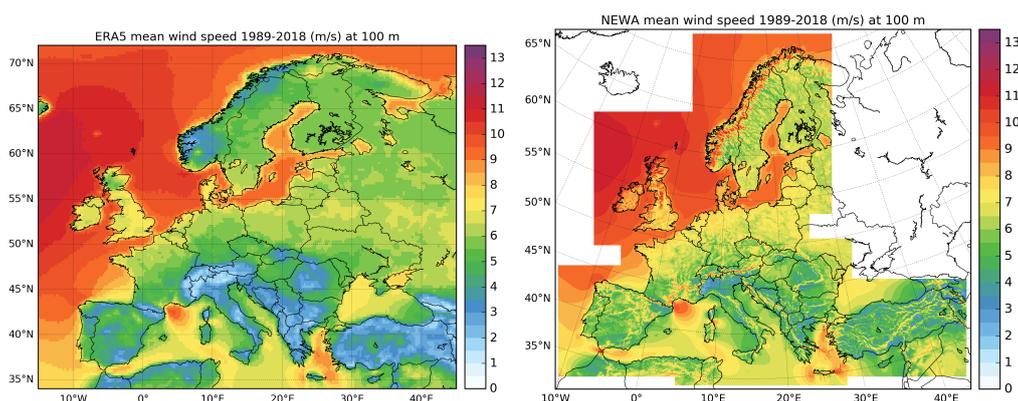


Figure 75: Comparison of the 30 year mean wind speed (1989–2018) at 100 m height: ERA5 (left) and NEWA (right).

Figure 77 shows the normalised 30 year standard deviation for both datasets. Note that the ERA5 values are based on hourly data while the NEWA values are based on 30 min data. All in all NEWA features higher standard deviations nearly everywhere in Europe. Over sea the difference to ERA5 is not that big, just a few percent. Over land, and especially in orographically complex terrain, the NEWA standard deviation reaches up to more than 100% while in ERA5 it does not exceed 70–80 %.

Figure 78 presents the maximum wind speed during the 30 year period 1989–2018 for each grid point and both datasets. Again and not surprisingly, NEWA yields higher wind maxima almost everywhere and in particular over land. The differences are specifically striking over Southern Europe where in some regions the maximum wind speed is increasing from 15 m s^{-1} in ERA5 to $30\text{--}40 \text{ m s}^{-1}$ in NEWA.

Finally, we present the comparison to the Global Wind Atlas (GWA2, DTU, 2019) which is based on 9 km WRF simulations using ERA-Interim forcing, but that includes the downscaling. Figure 79 shows that also compared to GWA2, the NEWA simulations feature higher wind speeds nearly everywhere. However, the broad picture looks very similar. It is striking that NEWA yields

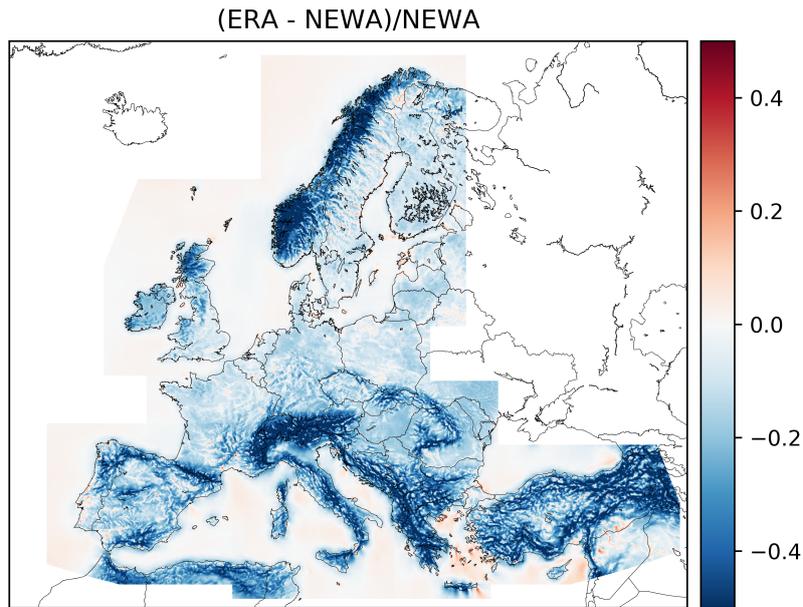


Figure 76: Comparison of the 30 year mean wind speed (1989–2018) at 100 m height between ERA5 and NEWA. Shown is the relative difference as $(ERA5 - NEWA)/NEWA$

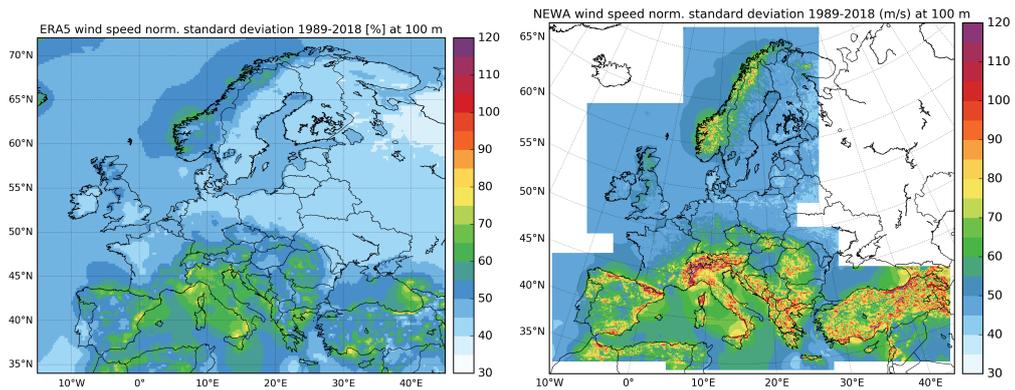


Figure 77: Comparison of the 30 year normalised standard deviation (1989–2018) at 100 m height: ERA5 (left) and NEWA (right).

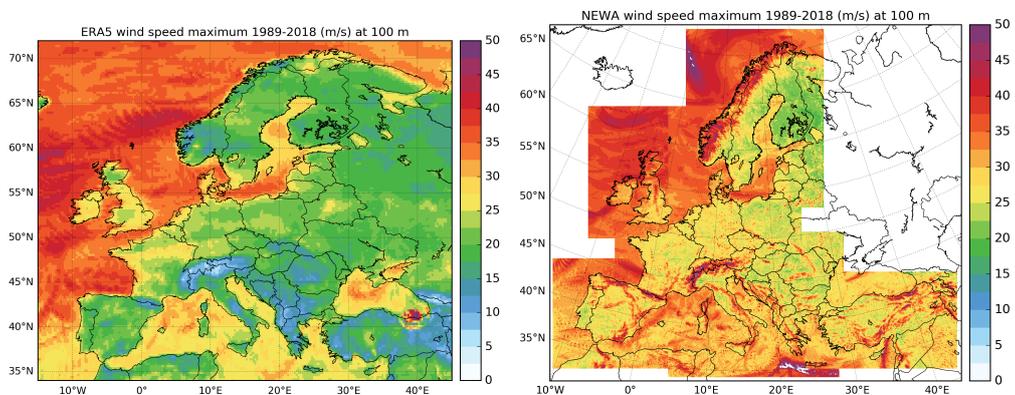


Figure 78: Comparison of the maximum wind speed during the 30 year period 1989–2018 at 100 m height: ERA5 (left) and NEWA (right).

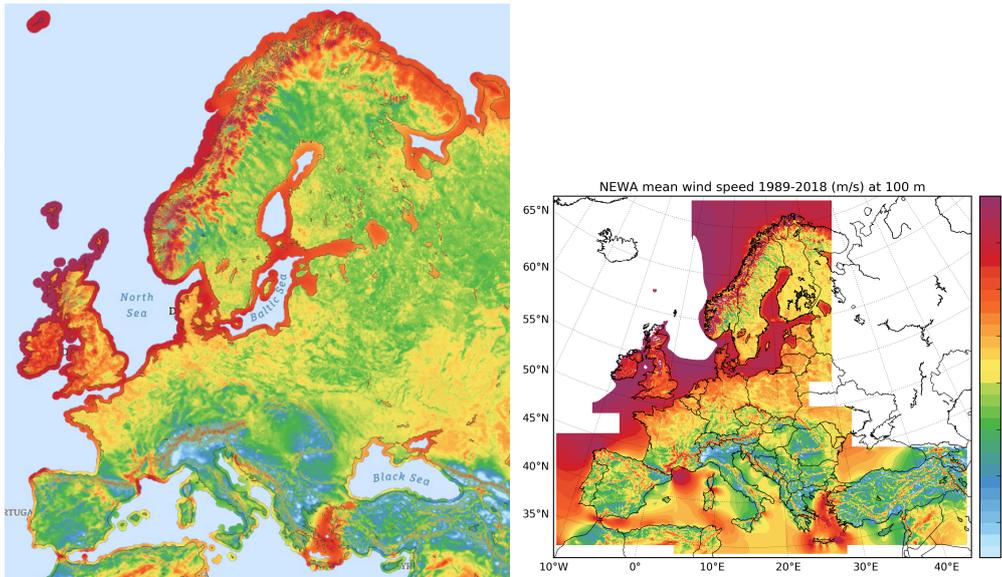


Figure 79: Comparison of the 30 year mean wind speed (1989–2018) at 100 m height: Global Wind Atlas 2 (left) and NEWA (right). The colour scale is the same for both plots.

significantly higher wind speeds over the seas (at least for the coastal waters included in GWA2).

4 Conclusions

This report explores the quantification of uncertainty in the winds simulated in the NEWA project. Uncertainty is understood here as the result of the contributions of model sensitivity to different model setups, and of model errors in a model-data comparison framework.

The first part of this report (Section 2 of this report explores the uncertainty derived from model sensitivity subjected to the decisions taken regarding the use of different models setups and how these produce variability in model output. The range of this variability has been regarded as spread in model output and has been quantified in various manners.

The NEWA project produced a large mesoscale multi-physics ensemble. Its purpose is twofold: on one hand allowing for a selection of an optimal setup for the production run (Witha et al., 2019); and on the other hand allowing for an estimation of model uncertainties. To achieve the second goal (Section 2.1), the multi-physics ensemble generated in NEWA has been analysed and the role of the principal sources of uncertainty in the model has been assessed. The differences between ensemble members have been quantified considering wind speed, wind direction and atmospheric stability. The analysis has allowed for identifying some of the ensemble members that do not differ significantly from the base run and do not contribute meaningfully to model spread. Such members were excluded from further analysis. In the context of this work the main sources of uncertainty are the different representations of subgrid-scale physical processes in thermally driven or orographically forced mesoscale processes.

Results indicate that applying different reanalysis products as boundary conditions, does not result in significant changes in mesoscale wind climate. This means that synoptic scale uncertainty is either too small or cannot currently be represented in the mesoscale models. On the other hand a change in the Sea Surface Temperature boundary dataset in at least one case (OISST) significantly changed the wind speed climate. This underscores the impact of the correct description of the surface processes. The analysis of wind direction distributions shows that differences between members are moderate, while changes in stability distributions depend on the land use class.

On the basis of the previous results over a specific sub-domain, the goal of selecting a small number of ensemble members that were to be calculated over the whole domain of the wind atlas was established based on both computing resources and identifying the most important model configurations contributing to spread (Section 2.2). During preliminary analysis it was concluded that it is very likely that ensemble properties would be different in different geographical regions of Europe, with contributing factors such as large scale circulation and general climate characteristics (northern Europe vs. southern Europe), the orography (flat or mountainous terrain) and typical land use (e.g. forests) affecting the ensemble. Therefore, the ensemble members that were expected to contribute most to spread were additionally calculated for a single year for another very different sub-domain in southern Europe, the GR (Greece) domain. For the calculations over the whole domain, a methodology to select the ensemble members according to their contribution to the total spread expressed in the space of the cumulative distribution function was developed. Using this methodology, based on the EMD metric as a descriptor of the differences between two probability distributions, it was established that the two most different ensemble members (in averaged sense over the two test domains), were those where the PBL and surface layer parameterisation scheme was changed to YSU-MM5 and MYJ-MO, respectively.

The advantage of defining the spread in *cdf* space is that the spread can be analysed separately for low-wind speeds and high-wind speeds and the analysis suggests that the spatial distribution of the spread is very different in both of these cases. Finally, the total spread was translated into five uncertainty categories. The final wind atlas will include information about the uncertainty category for each mesoscale grid point.

An alternative methodology to estimate the model spread based on the sensitivity experi-

ments with different WRF model configurations and use of varying physical parameterisations is demonstrated in Section 2.3 over the Iberian Peninsula sub-domain. Therefore, a pool of WRF simulations over the Iberian Peninsula subdomain making use of different parameterisation schemes and alternative setups was implemented. A methodology based on the rotation of the Empirical Orthogonal Functions of the original wind field over the region under study was implemented, obtaining with such a procedure several subregions that shared homogeneous wind variability.

A high resolution domain in the northeastern Iberian Peninsula showed a number of principal rotated modes of wind circulation that showed common variations of the simulated wind field. The differences among the set of simulations over this subregion were explored based on the projections of the wind time series onto the several subregional modes. In such a way the regional model spread is determined by differences among principal components or among differences between regional averages in different simulations. It has been shown that the spread varies with time within each region of homogeneous wind behaviour. This variability in the spread is likely related to synoptic configurations and will be addressed elsewhere.

This method focuses on anomalies of normalised data. Therefore, the biases studied in Sections 2.1 and 2.2 are filtered out and the focus is put on the time variability that purportedly relates to changes in the large scale circulation. The regional model spread over the Alaiz IB (northeastern IB) domain is in general low (Section 2.3.2), except for some synoptic transients. The impacts during these events are not negligible when for instance, different LSMs are used. In general, spread tends to increase during periods with larger wind variability. The YSU PBL scheme, the CLM or the Noah-Multiphysics LSMs options are the largest contributors to spread.

It has also been discussed that there are factors that have not been systematically considered in their contribution to spread. The model running strategy, can play a larger role than assessed herein and should be further considered in the future in correctly addressing model spread. The running strategy resets the initial conditions every week, thus hampering the memory of the subsurface processes (temperature and humidity) to develop at longer timescales and therefore hinders the real potential to develop the influence of the physics of the LSMs. The use of such initialisation strategies with distributed computing intervals of one week is totally justified from the perspective of computing resources and demand for the NEWA production run. However, it deserves further attention in the future as computing resources increase.

The second part of this report (Section 3) addressed how model performance can be characterised with the data at hand and whether decisions regarding selection of a given model setup for a production run can be taken on the basis of model performance in a variety of situations, using different variables and datasets as observational targets: tall wind masts from the Vestas database; wind profiles from tall masts and lidar over the sea and simple terrain; surface wind data; satellite data and reanalysis outputs.

The NEWA model-chain was firstly evaluated (Section 3.1) using measurements from 291 tall masts distributed across Europe, including Turkey. The average wind speed bias for the 291 masts is $0.28 \pm 0.76 \text{ m s}^{-1}$ for WRF-WAsP compared to $-1.50 \pm 1.30 \text{ m s}^{-1}$ and $0.02 \pm 0.78 \text{ m s}^{-1}$, for ERA5 and WRF, respectively.

The bias statistics can be segregated according to the complexity of the terrain at the mast site locations. In simple terrain (*low* RIX), and terrain with mostly gently sloping hills (*medium* RIX), the wind speed bias $\overline{U}_{\text{bias}}$ estimated by WRF-WAsP has the smallest spread of any of the models ($\sigma=0.43 \text{ m s}^{-1}$ in *low* RIX and $\sigma=0.76 \text{ m s}^{-1}$ in *medium* RIX), and the average wind speed bias is low in simple terrain (*low* RIX) $\mu=0.06 \text{ m s}^{-1}$, compared to $\mu=0.21 \text{ m s}^{-1}$ for WRF. In *medium* RIX terrain WRF-WAsP overestimates the mean wind speed by 0.23 m s^{-1} on average, while the mean bias of WRF is 0.03 m s^{-1} . In terrain with the greatest occurrence of steep hills (*high* RIX) WRF-WAsP overestimates the mean wind speed by 0.62 m s^{-1} . The WAsP flow model assumes attached flow, which is an inaccurate assumption in steep terrain, which may explain the overestimation of mean wind speed by WRF-WAsP. A similar trend, greater complexity means larger spread, was seen for roughness complexity. The smallest spread for WRF-WAsP was seen for the lowest roughness complexity.

The EMD was used to measure the similarity between the observed and modelled wind direction PDFs. On average, the WRF and WRF-WAsP wind directions are very similar, with a similar level of accuracy, and ERA5 has slightly less accurate wind direction PDFs. However, when only masts in the *low* RIX locations are considered, the wind direction PDFs for ERA5 is slightly more accurate than WRF and WRF-WAsP. For *high* RIX locations the opposite is true (WRF and WRF-WAsP are slightly more accurate).

The wind speed and wind direction data obtained from the NEWA production run was compared to 14 tall masts in Central Europe, most of them near the coast and offshore but some also far onshore and in more complex terrain. In summary, the NEWA results are close to the observations, especially for the near-coastal and offshore locations. Over land in more complex terrain, NEWA is typically over-predicting the wind speed in 100 m height. The highest correlations are obtained offshore (around 0.9), the lowest onshore in complex terrain (around 0.7). For the onshore locations clear annual and diurnal cycles of the correlation have been found.

Validation of the wind speeds simulated in the NEWA production run was carried on with mostly Lidar and mast wind data from the North and Baltic Seas and some land-based sites in Denmark, Sweden and Holland (Section 3.3). In summary, the NEWA simulations compare quite well against wind speeds from tall masts and Lidar offshore, except when the time series are short. At turbine-rotor-level ($\sim 100\text{m}$) the overall absolute biases are $<3\%$ and correlations >0.86 . For taller measurement heights ($>150\text{m}$), a systematic under-prediction of the mean wind speed by the NEWA simulations is seen in the evaluation, but it could be partially due to Lidar tendency to sample more often at higher wind speeds, which are common at these heights.

The ideal situation to validate the NEWA wind atlas is to make use of wind speed and direction observations at the hub height. However this type of records are typically scarce. Thus, the evaluation of the mesoscale simulation ability to reproduce the observed wind fields needs further comparisons with other available wind observations. To overcome this difficulty, a surface wind dataset over the whole European domain was compiled in the frame of the NEWA wind atlas development, the so-called WiSEd database, with a dense network of stations with wind speed and direction values over the vast European domain.

As a first step, the simulations were evaluated over a high resolution domain that comprised the northeastern part of the Iberian Peninsula and southwestern France, using the WiSEd data and complemented with wind mast data at some locations. It was shown that the mean wind bias tends to be positive at the surface and negative when compared with observations in the tall masts, generally contained in the range $[-2, 2]\text{ m s}^{-1}$.

Importantly it has been shown that the differences between simulations and observations are in general larger than differences between the simulations, i.e., errors tend to be larger than model spread. The intermodel differences are more relevant at the surface than at height in the case of simulations using different LSMs.

The underestimation of the mean wind is corroborated by the fact that regional model greatly underestimates the variance for three of the masts in the region. The correlation values between the simulations and the observations is higher in the case of the tall mast series compared to the surface stations as expected, since the effect of the non-resolved orography, among other issues, is largest closer to the ground. Therefore, although some issues can complicate the representation of the correct levels of variance by the model at the surface, comparatively with the simulated wind above, still it can be said that in general the WRF model adequately represents the variability of the observed wind over the region.

Additionally, the surface WiSEd observations and the tower records have been compared to the NEWA production simulation at the broader European scale, to analyse the skill of the official WRF model configuration in reproducing the observed wind field at the surface. The areas showing highest wind speed are northern Europe and there, better scores for the co-variability between observation and the simulation were found. Nevertheless, these areas showed also some difficulties in reproducing the largest variance observed comparatively with regions where the mean wind speed is generally lower. The simulated wind at the towers apparently outperforms that at the surface sites but does not seem to improve as the height increases, rather statistics are homogeneous across the different heights within the tower.

Additionally to land surface data, satellite winds at 10 m above the ocean surface have been used to assess WRF simulations in NEWA (Section 3.5). Although the temporal availability of satellite wind retrievals is reduced compared to in-situ measurements and model simulations, mean wind speed over the entire period of satellite data availability is descriptive of the spatial variability over the NEWA domain.

When compared to the WRF model simulations for 2015, ASCAT mean wind speeds showed higher spatial variability, although wind speed magnitude differences were on average up to 0.3 m s^{-1} , with exceptions being associated to artefacts in the ASCAT mean wind speed. WRF mean wind speeds simulated using different PBL schemes showed differences of 0.4 m s^{-1} at 10 m which decreased to 0.25 m s^{-1} at 100 m. Furthermore, an increased sensitivity to the atmospheric stability was observed and very frequent unstable conditions occurred in the North Sea domain. Sea surface temperature fields were used to interpret such frequently unstable conditions, with a good spatial agreement between relatively large day-time increase of the SST and the frequency of unstable WRF simulations.

Satellite winds were extrapolated to 100 m using the logarithmic wind profile (ENW) and a long-term stability correction (SDW) derived from WRF simulations. Comparisons with the NEWA production run revealed increased mean wind speeds for most offshore areas, particularly apparent in coastal areas while highlighting spatial features connected with topography and persistent wind patterns, e.g. the Mistral wind in the Gulf of Lyon and the Etesians in the Aegean Sea. Smaller biases were identified between the NEWA production run and the ASCAT stability dependent winds at 100 m for the largest part of the domain, although in the Mediterranean and especially in coastal regions ASCAT winds were higher.

The long-term wind climate as predicted by the NEWA production run has been compared to ERA5 reanalysis data (Section 3.6). NEWA indicates significantly higher wind speeds over land, especially in mountainous regions that are not resolved in such detail in ERA5. Over sea both datasets show very similar results except some deviations near islands or mountainous coasts. Generally, NEWA features much higher standard deviations of wind speed and also higher wind speed maxima.

A preliminary, purely visual comparison to the downscaled Global Wind Atlas results indicates that NEWA is predicting higher wind speeds not only over land but also over sea.

The research described in this report contains overall preliminary insights regarding the development of strategies for characterising uncertainty in terms of model spread and model-data comparison. More in depth analysis will be developed and subjected to peer review publication elsewhere.

Acknowledgements

The NEWA project counts with the support from an ERA-Net Plus consortium composed by the European Commission and 9 funding agencies from 8 member states:

- Public Service of Wallonia, Department of Energy and Sustainable Building (Belgium)
- Department of Economy, Science and Innovation Flemish Government (Belgium)
- Danish Energy Authority (Denmark)
- Federal Ministry for the Economic Affairs and Energy, on the basis of the decision by the German Bundestag (Germany)
- Latvijas Zinatnu Akademija (Latvia)
- Funda ção para a Ciênciã e a Tecnologia (Portugal)
- Ministerio de Economía y Competitividad (Spain)
- The Swedish Energy Agency (Sweden)
- The Scientific and Technological Research Council of Turkey (Turkey)

The tall mast data used for the verification has been kindly provided by the following people and organisations:

- Cabauw: Data provided by Cabauw Experimental Site for Atmospheric Research (Cesar), maintained by KNMI (KNMI, 2019)
- FINO 1,2,3: German Federal Maritime And Hydrographic Agency (BSH) (Forschungs- und Entwicklungszentrum Fachhochschule Kiel GmbH, 2019)
- Hamburg: Data from the Hamburg Weather Mast provided by Meteorological Institute at Hamburg University (Lange, 2018)
- Hohenpeißenberg: Data provided by the German Weather Service (DWD, Matthias Lindauer)
- Høvsøre: Data provided by Technical University of Denmark (DTU)
- Ijmuiden: Data from the Meteorological Mast Ijmuiden provided by Energy Research Center of the Netherlands (ECN) (Energy Research Centre of the Netherlands, 2018), processed data shared by Peter Kalverla from Wageningen University
- Jülich: Data from the weather station of Forschungszentrum Jülich (FZJ) provided by Axel Knaps
- Karlsruhe: Data provided by Martin Kohler, Institute of Meteorology and Climate Research at Karlsruhe Institute of Technology (KIT) (Kohler et al., 2018)
- Kassel: Data provided by Fraunhofer IEE
- Lindenberg: Data provided by the German Weather Service (DWD, Udo Rummel)
- OWEZ: Publicly available data (NordzeeWind, 2019)
- Tystofte: Data provided by Technical University of Denmark (DTU)
- Data from Lidar and masts in the North and Baltic Sea provided from the NORSEWInD project (Hasager et al., 2008).

We used data from the Global Wind Atlas 2.0 (DTU, 2019), a free, web-based application developed, owned and operated by the Technical University of Denmark (DTU) in partnership with the World Bank Group, utilising data provided by Vortex, with funding provided by the Energy Sector Management Assistance Program (ESMAP).

Satellite wind data from ASCAT are courtesy of EUMETSAT, the Copernicus Marine Environment Monitoring Service (CMEMS) and the Ocean & Sea Ice Satellite Application Facility (O&SI SAF). SAR data are courtesy of ESA and have been processed at DTU Wind Energy for the wind retrievals.

We acknowledge PRACE for awarding us access to MareNostrum at Barcelona Supercomputing Center (BSC), Spain, without which the NEWA simulations would not have been possible. Part of the simulations were performed on the HPC Cluster EDDY at the University of Oldenburg, funded by the German Federal Ministry for Economic Affairs and Energy under grant number 0324005. This work was partially supported by the computing facilities of the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund (ERDF), CIEMAT and the Government of Spain. In addition, simulations carried out as part of this work also made use of the computing facilities provided by CIEMAT Computer Center.

Most of the WRF simulations were initialised using ERA5 data, downloaded from ECWMF and Copernicus Climate Change Service Climate Data Store (CDS, Copernicus Climate Change Service (C3S), 2019).

References

- Argüeso, D., J. M. Hidalgo-Muñoz, S. R. Gámiz-Fortis, M. J. Esteban-Parra, J. Dudhia, and Y. Castro-Díez, 2011: Evaluation of WRF parameterizations for climate studies over southern Spain using a multistep regionalization. *Journal of Climate*, **24**, 5633–5651.
- Badger, J., H. Frank, A. N. Hahmann, and G. Giebel, 2014: 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.
- Badger, M., T. Ahsbahs, P. Maule, and I. Karagali, 2019: Inter-calibration of SAR data series for offshore wind resource assessment. *Remote Sensing of Environment*, **under review**.
- Badger, M., A. Peña Diaz, A. N. Hahmann, A. Mouche, and C. B. Hasager, 2016: Extrapolating satellite winds to turbine operating heights. *Journal of Applied Meteorology and Climatology*, **55**, 975–991.
- Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Monthly Weather Review*, **132**, 473–494.
- Bonan, G. B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R. E. Dickinson, and Z.-L. Yang, 2002: The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *Journal of Climate*, **15**, 3123–3149.
- Brinckmann, S., S. Krähenmann, and P. Bissolli, 2015: High-resolution daily gridded datasets of air temperature and wind speed for Europe. *Earth System Science Data Discussions*, **8**, 649–702.
URL <http://www.earth-syst-sci-data-discuss.net/8/649/2015/>
- Burlando, M., M. Antonelli, and C. Ratto, 2008: Mesoscale wind climate analysis: identification of anemological regions and wind regimes. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **28**, 629–641.
- Chang, R., R. Zhu, M. Badger, C. B. Hasager, R. Zhou, D. Ye, and X. Zhang, 2014: Applicability of synthetic aperture radar wind retrievals on offshore wind resources assessment in Hangzhou Bay, China. *Energies*, **7**, 3339–3354.
- Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Y. Duan, M. Ek, et al., 1996: Modeling of land surface evaporation by four schemes and comparison with field observations. *Journal of Geophysical Research: Atmospheres*, **101**, 7251–7268.
- Cheng, C. S., E. Lopes, C. Fu, and Z. Huang, 2014: Possible impacts of climate change on wind gusts under downscaled future climate conditions: Updated for Canada. *Journal of Climate*, **27**, 1255–1270.
- Chin, T., R. Milliff, and W. Large, 1998: Basin-scale, high-wavenumber sea surface wind fields from a multiresolution analysis of scatterometer data. *Journal of Atmospheric and Oceanic Technology*, **15**, 741–763.
- Cleugh, H., J. Miller, and M. Böhm, 1998: Direct mechanical effects of wind on crops. *Agroforestry Systems*, **41**, 85–112.
- Copernicus Climate Change Service (C3S), 2019: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Copernicus Land Monitoring Service, 2019: CORINE Land Cover. <https://land.copernicus.eu/pan-european/corine-land-cover>, accessed: 2019-04-15.

- De Kloe, J., A. Stoffelen, and A. Verhoef, 2017: Improved use of scatterometer measurements by using stress-equivalent reference winds. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**, 7913–7927.
- Dee, D. P., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, et al., 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553–597.
- Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The operational sea surface temperature and sea ice analysis (ostia) system. *Remote Sensing of Environment*, **116**, 140–158.
- Doubrawa, P., R. J. Barthelmie, S. C. Pryor, C. B. Hasager, M. Badger, and I. Karagali, 2015: Satellite winds as a tool for offshore wind resource assessment: The Great Lakes wind atlas. *Remote Sensing of Environment*, **168**, 349–359.
- DTU, 2019: Global Wind Atlas 2.0. <https://globalwindatlas.info/>, accessed: 2019-06-06.
- Energy Research Centre of the Netherlands, 2018: Wind op Zee. <http://www.windopzee.net/>, accessed: 2019-06-06.
- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, doi:10.5194/gmd-9-1937-2016.
- Farquhar, G. and M. Roderick, 2005: Worldwide changes in evaporative demand. *Water and the Environment*, **12**, 81–99.
- Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, et al., 2013: *Evaluation of climate models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Floors, R., A. N. Hahmann, and A. Peña, 2018: Evaluating mesoscale simulations of the coastal flow using lidar measurements. *Journal of Geophysical Research – Atmospheres*, **123**, 2718–2736.
- Forschungs- und Entwicklungszentrum Fachhochschule Kiel GmbH, 2019: FINO1,2,3 - Forschungsplattformen in Nord- und Ostsee Nr. 1,2,3. <https://www.fino-offshore.de/en/index.html>, accessed: 2019-06-06.
- Fujiwara, M., J. S. Wright, G. L. Manney, L. J. Gray, J. Anstey, T. Birner, S. Davis, E. P. Gerber, et al., 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, **17**, 1417–1452.
- García-Bustamante, E., J. F. González-Rouco, J. Navarro, E. Xoplaki, P. a. Jiménez, and J. P. Montávez, 2011: North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula: uncertainty and long term downscaled variability. *Climate Dynamics*, **38**, 141–160.
- García-Bustamante, E., J. F. González-Rouco, J. Navarro, E. Xoplaki, J. Luterbacher, P. A. Jiménez, J. P. Montávez, A. Hidalgo, et al., 2012: Relationship between wind power production and North Atlantic atmospheric circulation over the northeastern Iberian Peninsula. *Climate Dynamics*, **40**, 935–949.
- Gryning, S. E., E. Batchvarova, B. Brümmner, H. Jørgensen, and S. Larsen, 2007: On the extension of the wind profile over homogeneous terrain beyond the surface layer. *Boundary Layer Meteorology*, **124**, 251–268.

- Hahmann, A., C. Lennard, J. Badger, C. Vincent, M. Kelly, P. Volker, B. Argent, and J. Refslund, 2014: Mesoscale modeling for the wind atlas for South Africa (WASA) project. Tech. rep., DTU Wind Energy.
- Hahmann, A. N., B. T. Olsen, P. J. H. Volker, N. G. Mortensen, and J. Badger, 2019: Generalization of WRF-derived wind climatologies for validation and coupling of mesoscale and microscale models. *Wind Energy Science (In preparation)*.
- Hahmann, A. N., C. L. Vincent, A. Peña, J. Lange, and C. B. Hasager, 2015: Wind climate estimation using WRF model output: Method and model sensitivities over the sea. *International Journal of Climatology*, **35**, 3422–3439.
- Halpern, D., M. H. Freilich, and R. A. Weller, 1999: Ecmwf and ers-1 surface winds over the arabian sea during july 1995. *Journal of Physical Oceanography*, **29**, 1619–1623.
- Hannachi, A., I. Jolliffe, and D. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *International journal of climatology*, **27**, 1119–1152.
- Hannachi, A., I. Jolliffe, D. Stephenson, and N. Trendafilov, 2006: In search of simple structures in climate: simplifying eofs. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **26**, 7–28.
- Hartmann, D. L., A. M. G. K. Tank, M. Rusticucci, L. V. Alexander, S. Bronnimann, Y. Charabi, F. Dentener, E. J. Dlugokencky, et al., 2013: *Observations: Atmosphere and Surface*. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Hasager, C. B., M. Badger, N. Nawri, B. Rugaard Furevik, G. N. Petersen, H. Björnsson, and N.-E. Clausen, 2015a: Mapping offshore winds around Iceland using satellite synthetic aperture radar and mesoscale model simulations. *I E E E Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **8**, 5541–5552.
- Hasager, C. B., M. Badger, A. Peña Diaz, X. G. Larsén, and F. Bingöl, 2011: SAR-based wind resource statistics in the baltic sea. *Remote Sensing*, **3**, 117–144.
- Hasager, C. B., A. N. Hahmann, T. Ahsbaks, I. Karagali, T. Sile, M. Badger, and J. Mann, 2019: Europe’s offshore winds assessed from SAR, ASCAT and WRF. *Wind Energy Science Discussions - under review and discussion*, **In review**.
- Hasager, C. B., A. Mouche, M. Badger, F. Bingöl, I. Karagali, T. Driesenaar, A. Stoffelen, A. Peña, et al., 2015b: Offshore wind climatology based on synergetic use of envisat ASAR, ASCAT and quikSCAT. *Remote Sensing of Environment*, **156**, 247–263.
- Hasager, C. B., A. Peña Diaz, M. B. Christiansen, P. Astrup, N. M. Nielsen, F. Monaldo, D. Thompson, and P. Nielsen, 2008: Remote sensing observation used in offshore wind energy. *I E E E Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **1**, 67–79.
- Hasager, C. B., D. Stein, M. Courtney, A. Penlfa, T. Mikkelsen, M. Stickland, and A. Oldroyd, 2013: Hub height ocean winds over the North Sea observed by the NORSEWInD lidar array: Measuring techniques, quality control and data management. *Remote Sensing*, **5**, 4280–4303.
- Hasager, C. B., P. Vincent, J. Badger, M. Badger, A. Di Bella, A. Peña Diaz, R. Husson, and P. Volker, 2015c: Using satellite SAR to characterize the wind flow around offshore wind farms. *Energies*, **8**, 5413–5439.
- Hersbach, H., 2010: Comparison of c-band scatterometer cmod5.n equivalent neutral winds with ecmwf. *Journal of Atmospheric and Oceanic Technology*, **27**, 721–736.

- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Weather Rev.*, **132**, 103–120.
- Horstmann, J., W. Koch, and S. Lehner, 2004: Ocean wind fields retrieved from the advanced synthetic aperture radar aboard envisat. *Ocean Dynamics*, **54**, 570–576.
- Hu, X.-M., P. M. Klein, and M. Xue, 2013: Evaluation of the updated YSU Planetary Boundary Layer Scheme within WRF for Wind Resource and Air Quality Assessments. *Journal of Geophysical Research: Atmospheres*, **118**, 10 490–10 505.
- Huber, P. J., 1973: The 1972 Wald memorial lectures: Robust regression: Asymptotics, conjectures, and Monte Carlo. *The Annals of Statistics*, **1**, 799–821.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, **113**, D13 103.
- IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. rep., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jerez, S., J. P. Montávez, J. J. Gómez-Navarro, P. Jiménez-Guerrero, J. M. Jiménez, and J. F. González-Rouco, 2010: Temperature sensitivity to the land surface model in regional climate simulations. *Meteorologische Zeitschrift*, **19**, 363–374.
- Jiménez, P., J. González-Rouco, E. García-Bustamante, J. Navarro, J. Montávez, J. de Arellano, J. Dudhia, and A. Muñoz-Roldan, 2010a: Surface wind regionalization over complex terrain: evaluation and analysis of a high-resolution WRF simulation. *Journal of Applied Meteorology and Climatology*, **49**, 268–287.
- Jiménez, P., J. González-Rouco, J. Navarro, J. Montávez, and E. García-Bustamante, 2010b: Quality assurance of surface wind observations from automated weather stations. *Journal of Atmospheric and Oceanic Technology*, **27**, 1101–1122.
- Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, **140**, 898–918.
- Jiménez, P. A., J. F. González-Rouco, J. P. Montávez, E. García-Bustamante, J. Navarro, and J. Dudhia, 2013: Analysis of the long-term surface wind variability over complex terrain using a high spatial resolution WRF simulation. *Climate Dynamics*, **40**, 1643–1656.
- Jiménez, P. A., J. Vilà-Guerau de Arellano, J. F. González-Rouco, J. Navarro, J. P. Montávez, E. García-Bustamante, and J. Dudhia, 2011: The effect of heat waves and drought on surface wind circulations in the northeast of the Iberian peninsula during the summer of 2003. *Journal of Climate*, **24**, 5416–5422.
URL <https://doi.org/10.1175/2011JCLI4061.1>
- Jolliffe, I., 2011: *Principal component analysis*. Springer.
- Jonkman, J. M., S. Butterfield, W. Musial, and G. Scott, 2009: Definition of a 5-MW reference wind turbine for offshore system development. Technical Report NREL/TP-500-38060, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, Colorado 80401-3393.
- Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *Journal of Applied Meteorology and Climatology*, **43**, 170–181.
- Kalverlaa, P. C., G.-J. Steeneveld, R. J. Rondab, and A. A. Holtslag, 2017: An observational climatology of anomalous wind events at offshore meteorological mast IJmuiden (North Sea). *Journal of Wind Engineering and Industrial Aerodynamics*, **165**, 86–99.

- Kara, A. B., A. J. Wallcraft, and M. A. Bourassa, 2008: Air-sea stability effects on the 10 m winds over the global ocean: Evaluations of air-sea flux algorithms. *Journal of Geophysical Research: Oceans*, **113**, C04009.
- Karagali, I., M. Badger, A. N. Hahmann, A. Peña Diaz, C. B. Hasager, and A. M. Sempreviva, 2013a: Spatial and temporal variability of winds in the Northern European Seas. *Renewable Energy*, **57**, 200–210.
- Karagali, I. and J. Høyer, 2014: Characterisation and quantification of regional diurnal SST cycles from SEVIRI. *Ocean Science*, **10**, 745–758.
- Karagali, I., J. Høyer, and C. B. Hasager, 2012: SST diurnal variability in the North Sea and the Baltic Sea. *Remote Sensing of Environment*, **121**, 159–170.
- Karagali, I., X. G. Larsén, M. Badger, A. Peña, and C. B. Hasager, 2013b: Spectral properties of ENVISAT ASAR and quikSCAT surface winds in the North Sea. *Remote Sensing*, **5**, 6096–6115.
- Karagali, I., A. Peña Diaz, M. Badger, and C. B. Hasager, 2014: Wind characteristics in the North and Baltic Seas from the QuikSCAT satellite. *Wind Energy*, **17**, 123–140.
- Kelly, M. and H. E. Jørgensen, 2017: Statistical characterization of roughness uncertainty and impact on wind resource estimation. *Wind Energy Science*, **2**, 189–209.
- Khanduri, A. and G. Morrow, 2003: Vulnerability of buildings to windstorms and insurance loss estimation. *Journal of wind engineering and industrial aerodynamics*, **91**, 455–467.
- KNMI, 2019: Cesar - Cabauw Experimental Site for Atmospheric Research. <http://www.cesar-database.nl/>, accessed: 2019-06-06.
- Kohler, M., J. Metzger, and N. Kalthoff, 2018: Trends in temperature and wind speed from 40 years of observations at a 200-m high meteorological tower in southwest germany. *International Journal of Climatology*, **38**, 23–34.
URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5157>
- Kotlarski, S., K. Keuler, O. B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, et al., 2014: Regional climate modeling on european scales: a joint standard evaluation of the euro-cordex rcm ensemble. *Geoscientific Model Development*, **7**, 1297–1333.
URL <https://www.geosci-model-dev.net/7/1297/2014/>
- Lange, I., 2018: Merkblatt: Daten vom Wettermast Hamburg. https://wettermast.uni-hamburg.de/Downloads/Wettermast_Datenbeschreibung.pdf, accessed: 2019-06-06.
- Lee, J. A., W. C. Kolczynski, T. C. McCandless, and S. E. Haupt, 2012: An objective methodology for configuring and down-selecting an NWP ensemble for low-level wind prediction. *Mon. Weather Rev.*, **140**, 2270–2286.
URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-11-00065.1>
- Leutbecher, M. and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539.
URL <http://10.0.3.248/j.jcp.2007.02.014>
<https://dx.doi.org/10.1016/j.jcp.2007.02.014>
- Liu, W. T. and W. Tang, 1996: Equivalent neutral wind. JPL Publication 96-17, NASA Jet Propulsion Laboratory, ISBN: 978-3-9818107-0-7.
- Lorente-Plazas, R., J. Montávez, P. Jimenez, S. Jerez, J. Gómez-Navarro, J. García-Valero, and P. Jimenez-Guerrero, 2015: Characterization of surface winds over the iberian peninsula. *International journal of climatology*, **35**, 1007–1026.

- Lucio-Eceiza, E. E., J. F. González-Rouco, E. García-Bustamante, J. Navarro, and H. Beltrami, 2019: Multidecadal to Centennial Surface Wintertime Wind Variability Over Northeastern North America via Statistical Downscaling. *Climate Dynamics*.
- Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro, and H. Beltrami, 2018a: Quality Control of surface wind observations in North Eastern North America. Part I: Data Management Issues. *Journal of Atmospheric and Oceanic Technology*, **35**, 163–182.
- Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro, H. Beltrami, and J. Conte, 2018b: Quality Control of surface wind observations in North Eastern North America. Part II: Measurement Errors. *Journal of Atmospheric and Oceanic Technology*, **35**, 183–205.
- McVicar, T. R., M. L. Roderick, R. J. Donohue, L. T. Li, T. G. Van Niel, A. Thomas, J. Grieser, D. Jhajharia, et al., 2012: Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, **416-417**, 182–205.
- Mearns, L., F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, 2003: Guidelines for use of climate scenarios developed from regional climate model experiments. Tech. rep. URL http://www.ipcc-data.org/guidelines/dgm_no1_v1_10-2003.pdf
- Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2004: Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, **109**, D13 104.
- Monaldo, F. M., X. Li, W. G. Pichel, and C. R. Jackson, 2014: Ocean wind speed climatology from spaceborne SAR imagery. *Bulletin of the American Meteorological Society*, **95**, 565–569.
- Mortensen, N. G., A. Tindal, and L. Landberg, 2008: Field validation of the rix performance indicator for flow in complex terrain. *Paper presented at 2008 European Wind Energy Conference and Exhibition, Brussels, Belgium.*
- Murthy, K. and O. Rahi, 2017: A comprehensive review of wind resource assessment. *Renewable and Sustainable Energy Reviews*, **72**, 1320–1342.
- Nakanishi, M. and H. Niino, 2006: An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407.
- Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, et al., 2011: The community noah land surface model with multiparameterization options (Noah-MP): 1. model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, **116**.
- NordzeeWind, 2019: NordzeeWind. <http://www.noordzeewind.nl/en/knowledge/reportsdata/>, accessed: 2019-06-06.
- Nuño Martínez, E., P. Maule, A. N. Hahmann, N. A. Cutululis, P. E. Sørensen, and I. Karagali, 2018: Simulation of transcontinental wind and solar pv generation time series. *Renewable Energy*, **118**, 425–436.
- Oreskes, N., 1998: Evaluation (not validation) of quantitative models. *Environmental Health Perspectives*, **106**, 1453–1460.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz, 1994: Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646.
- Palutikov, J., P. Kelly, T. Davies, and J. Halliday, 1987: Impacts of spatial and temporal windspeed variability on wind energy output. *J. Clim. Appl. Meteor.*, **26**, 1124–1133.
- Peña Diaz, A., A. N. Hahmann, C. B. Hasager, F. Bingöl, I. Karagali, J. Badger, M. Badger, and N.-E. Clausen, 2011: South Baltic Wind Atlas. URL <https://orbit.dtu.dk/files/5578113/ris-r-1775.pdf>

- Peixoto, J. P. and A. Oort, 1984: Physics of climate. *Rev. Mod. Phys.*, **56**, 365–429.
- Pele, O. and M. Werman, 2008: A linear time histogram metric for improved sift matching. *European conference on computer vision*, Springer.
- Pryor, S. C., J. T. Schoof, and R. J. Barthelmie, 2005: Climate change impacts on wind speeds and wind energy density in Northern Europe: Empirical downscaling of multiple AOGCMs. *Climate Research*, **29**, 183–198.
- REN21, 2017: Renewables 2017 Global Status Report. Tech. rep., Paris: REN21 Secretariat, ISBN: 978-3-9818107-0-7.
- Ribaudo, S., 2019: Conditional error statistics for offshore winds, from WRF modelling and ASCAT satellite data. Master's thesis, DTU Wind Energy - submitted.
- Richman, M. B., 1986: Rotation of principal components. *Journal of climatology*, **6**, 293–335.
- Sanz Rodrigo, J., R. Chávez Arroyo, B. Witha, M. Dörenkämper, J. Gottschall, M. Avila, J. Arnqvist, A. Hahmann, et al., 2019: The New European Wind Atlas model chain. (*In preparation*).
- Sempreviva, A. M., S. E. Larsen, N. G. Mortensen, and I. Troen, 1990: Response to neutral boundary layers to change. *Boundary-Layer Meteorol.*, 205–225.
- Silva, J., C. Ribeiro, and i. Guedes, 2007: Roughness length classification of corine land cover classes. *roceedings of EWEC 2007*, **710**, 110.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 3. *Technical Report, NCAR, NCAR/TN-475+STR*.
URL http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.8/ARWUsersGuideV3.8.pdf
- Smits, A., A. M. Klein Tank, and G. P. Können, 2005: Trends in storminess over the Netherlands, 1962-2002. *International Journal of Climatology*, **25**, 1331–1344.
- Sotillo, M. G., A. W. Ratsimandresy, J. C. Carretero, A. Bentamy, F. Valero, and J. F. González-Rouco, 2005: A high-resolution 44-year atmospheric hindcast for the Mediterranean Basin: contribution to the regional improvement of global reanalysis. *Clim. Dyn.*, **25**, 219–236.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of cmip5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, DOI:10.1175/BAMS-D-11-00094.1.
- Tobin, I., S. Jerez, R. Vautard, F. Thais, E. Van Meijgaard, A. Prein, M. Déqué, S. Kotlarski, et al., 2016: Climate change impacts on the power generation potential of a European mid-century wind farms scenario. *Environmental Research Letters*, **11**.
- Troen, I. and E. L. Petersen, 1989: *European Wind Atlas*. Published for the Commission of the European Communities, Directorate-General for Science, Research, and Development, Brussels, Belgium by Risø National Laboratory.
URL https://orbit.dtu.dk/files/112135732/European_Wind_Atlas.pdf
- von Storch, H., 1995: Inconsistencies at the interface of climate impact studies and global climate research. *Meteorol. Zeitschrift*, **4**, 71–80.
- , 2010: Climate models and modeling: an editorial essay. *WIREs Clim. Change*, **1**, 305–310. doi: 10.1002/wcc.12.
- von Storch, H. and F. Zwiers, 1999: *Statistical analysis in climate research*. Cambridge University Press.
- White, D., M. Richman, and B. Yarnal, 1991: Climate regionalization and rotation of principal components. *International Journal of Climatology*, **11**, 1–25.

- Witha, B., A. N. Hahmann, T. Sile, M. Dörenkämper, Y. Ezber, E. G. Bustamante, J. F. González-Rouco, G. Leroy, et al., 2019: Report on WRF model sensitivity studies and specifications for the mesoscale wind atlas production runs: NEWA Deliverable D4.3. Tech. rep.
URL <https://doi.org/10.5281/zenodo.2682604>
- Zecchetto, S. and F. De Biasio, 2003: Wavenumber spectra of the mediterranean sea winds derived from the NASA QuikSCAT data. *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, vol. 3.