Received: 27 November 2017 | Revised: 11 July 2018 | Accepted: 5 August 2018
DOI: 10.1002/we.2267

WILEY

RESEARCH ARTICLE

Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast IJmuiden at the North Sea

Peter Kalverla[1] | Gert-Jan Steeneveld[1] | Reinder Ronda[2] | Albert AM Holtslag[1]

[1]Wageningen University and Research, Meteorology and Air Quality Section, Wageningen, The Netherlands
[2]Koninklijk Nederlands Meteorologisch Instituut, R&D Observations and Datatechnology, De Bilt, Netherlands

Correspondence
Peter Kalverla, Wageningen UR, Meteorology and Air Quality Section, PO Box 47, 6700AA Wageningen, Netherlands.
Email: peter.kalverla@wur.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: SH-312-15; Stichting voor de Technische Wetenschappen, Grant/Award Number: 14158

Abstract

Numerical weather prediction models play an important role in the field of wind energy, for example, in power forecasting, resource assessment, wind farm (wake) simulations, and load assessment. Continuous evaluation of their performance is crucial for successful operations and further understanding of meteorology for wind energy purposes. However, extensive offshore observations are rarely available. In this paper, we use unique met mast and Lidar observations up to 315 m from met mast "IJmuiden," located in the North Sea 85 km off the Dutch coast, to evaluate the representation of wind and other relevant variables in three mainstream meteorological models: ECMWF-IFS, HARMONIE-AROME, and WRF-ARW, for a wide range of weather conditions. Overall performance for hub-height wind speed is found to be comparable between the models, with a systematic wind speed bias <0.5 m/s and random wind speed errors (centered RMSE) <2 m/s. However, the model performance differs considerably between cases, with better performance for strong wind regimes and well-mixed wind and potential temperature profiles. Conditions characterized by moderate wind speeds combined with stable stratification, which typically produce substantial wind shear and power fluctuations, lead to the largest misrepresentations in all models.

KEYWORDS

Met mast IJmuiden, Model evaluation, North Sea, Numerical weather prediction

1 | INTRODUCTION

Flow modeling is an integral activity in the development of wind energy systems. Applications range from resource assessment[1,2] to load calculations[3-5] and from wake simulations[6-8] to power forecasting.[9,10] A comprehensive review of flow modeling in the wind energy sector was recently published by Sanz Rogdrigo et al.[11] In that study, the authors distinguish between mesoscale and microscale models.

Mesoscale models typically use a grid spacing of a few km and support domains that are large enough to incorporate large-scale weather patterns such as low-pressure systems, fine enough to resolve regional circulation patterns such as the sea breeze, but too coarse to resolve turbulence and its interaction with wind turbines.[12] These models are typically used to refine global modeling products, which are run on coarser grids, for a limited area. Important applications of mesoscale models in the wind energy sector are in forecasting and resource assessment.[1,10] Currently (01-2018), several institutes are working on the production of a New European Wind Atlas[13] based on mesoscale model results. At the same time, these models are further developed, eg, in the context[14] of the Wind Forecast Improvement Project 2.

Microscale models use finer grids in order to resolve turbulence and interactions with wind turbines, but their domains are too small to incorporate regional or large-scale circulation patterns. These models typically use idealized, stationary forcing, which poses limitations when dealing with transient phenomena such as fronts.[15] Coupling between mesoscale and microscale models is an active research field[11,12,15-17] that is expected to substantially reduce uncertainties originating from the use of idealized inflow fields.

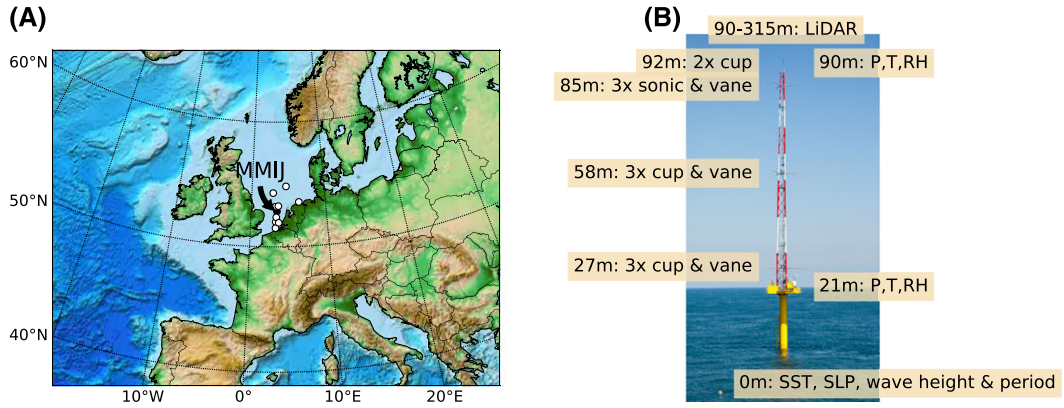

FIGURE 1 A, Map showing the location of the met mast and available automated weather stations (white dots) in the North Sea; B, a photograph of the mast with annotations of the instrumentation. MMIJ, met mast IJmuiden; SST, sea-surface temperature; SLP, sea-level pressure [Colour figure can be viewed at wileyonlinelibrary.com]

For a successful downscaling procedure and other applications, it is first and foremost important that all models are thoroughly validated. Since model performance varies with location (orography and land use), atmospheric conditions, and quantity of interest, and the models themselves are continuously improved, evaluation is an ongoing activity. For example, a downscaling product for resource assessment from 2012 was based on a 45-km grid,[18] whereas the New European Wind Atlas[13] will be based on a 3-km grid. The growing interest in offshore wind farm development also calls for model evaluation that focuses on offshore conditions, which have received considerably less attention than onshore conditions in the past due to a lack of observations.

In this study, we use a unique dataset that was obtained over a 4-year period between 2012 and 2015 to evaluate the performance of three mainstream numerical weather prediction models. "Met Mast IJmuiden" (MMIJ) was installed in the North Sea between the Netherlands and the United Kingdom (Figure 1). The exceptional distance to shore (85 km), length (4 years) and observation height (90 m mast, remote sensing up to 315 m) make the dataset well suited for an evaluation tailored to wind energy purposes at the Southern North Sea. The models included in this study are as follows:

1. The integrated forecasting system (IFS), the global operational weather model of the European Centre for Medium-range Weather Forecasts (ECMWF).
2. HARMONIE-AROME, a short-range regional weather forecasting model developed and used operationally by a consortium of European weather institutes.
3. The Weather Research and Forecasting (WRF) model, a community model maintained by the US University Corporation and National Center for Atmospheric Research (UCAR/NCAR).

The results presented in this paper address the following research question: "What is the typical model performance concerning (the vertical structure of) wind, temperature, and humidity and to what extent does this performance depend on the model itself and on atmospheric conditions?" A wide variety of cases is analyzed, each characterized by a unique combination of wind speed, direction and large-scale atmospheric stability. This characterization provides an interesting and novel perspective on the results.

More details on the models and observations are given in sections 2.1 and 2.2. Subsequently, we explain the case-selection procedure and the statistics used (sections 2.3 and 2.4). The results are presented and discussed in Section 3. The paper concludes with a summary and conclusions in Section 4.

2 | METHODOLOGY

2.1 | Description of the models

We use three mainstream numerical weather prediction models (IFS, HARMONIE, and WRF) that are all widely used for wind energy purposes but have a different focus.

The IFS is a global model that provides forecasts of up to 15 days ahead of time on a grid of approximately 16-km horizontal spacing (at the time of the MMIJ observations). It is a hydrostatic model, which means that vertical motions are not prognostically resolved. Instead, a subgrid parametrization is used to represent the effect of the unresolved convection. This approach is acceptable for relatively coarse grids but becomes invalid for finer grids where convection (strong updrafts and downdrafts such as in thunderstorms), local circulations, and turbulence start to be resolved. The latest operational analysis, which constitutes the best known state of the atmosphere at that time, is used as initial field. During the MMIJ observation period, the model received some updates, most notably a change from 91 to 137 vertical levels in June of 2013. Also, a

change in the formulation of turbulent mixing in stable conditions was made[19] in November of 2013. Appendix B shows which cases are before and after these changes. Further documentation of the IFS (versions 37-41) can be found on the ECMWF website.*

HARMONIE is a limited-area model that produces short-range forecasts of up to 48 hours ahead of time for the region of Northwest Europe at a horizontal grid spacing of 2.5 km. It uses 65 vertical levels with the model top at 10 hPa (for comparison, the IFS model top is at 0.01 hPa). Initial and lateral boundary conditions are taken from the IFS; however, additional regional observations from satellites, surface weather stations, buoys, ships, and airplanes are assimilated into the initial condition before the forecast is started. In contrast to the IFS, HARMONIE is a nonhydrostatic model, and it is assumed that deep convection reaching far above the boundary layer is completely resolved. However, the model still includes a subgrid parameterization for shallow convection at boundary-layer scales. An extensive description of the HARMONIE model can be found in Bengtsson et al.[20] For the studied period, the model version was 36.

The Weather Research and Forecasting (WRF) model, specifically the Advanced Research WRF (ARW) version 3.9,[21] is a very flexible model which, in addition to forecasting and downscaling, is well suited for case studies, sensitivity analysis, and for developing and testing new physical parameterizations. WRF is also popular among researchers and industrial users because of its flexibility and the fact that it is completely open source. WRF is a nonhydrostatic model that we force with initial and lateral boundary conditions from the IFS. However, where HARMONIE used the IFS forecasts as lateral boundaries (and assimilates additional observations), we use the subsequent operational analyses instead (and did not use data assimilation). For our reference simulations, we use 3-km horizontal grid spacing and 91 vertical levels up to 50 hPa (corresponding to the lowest levels in the 137 level formulation of IFS). In WRF, many subgrid parameterizations are available. We use a scale-adaptive convection parametrization,[22] which gradually deactivates as grid spacing is refined. This scheme was found to perform well for resolutions up to 3 km.[23] Further details of the WRF configuration can be found in Appendix A.

It is clear that the model setup for each model is different. The WRF simulations were performed in-house, with a setup that is commonly used for case studies. The IFS and HARMONIE data were obtained from the operational archives of ECMWF and KNMI (the Dutch national weather service), respectively, and as such represent the operational model setup at the time of the MMIJ observations. Consequently, we evaluate the performance of each model in its typical setup, as if it were to be used straight away. However, to bridge the gap between the horizontal grid spacing, we performed an additional set of WRF simulations at a resolution (and vertical grid spacing) comparable with the IFS. These simulations will be referred to as "WRF-coarse." Additional sensitivity runs with WRF-coarse were performed concerning the strength of the turbulent mixing in stable conditions, the impact of prognostic sea-surface temperature (SST), and the impact of momentum entrainment at the boundary-layer top.

2.2 | Observations and model data

The IJmuiden dataset is described in detail in Werkhoven and Verhoef[24] and is previously analyzed in Kalverla et al[25] and Holtslag et al.[26] It is a 4-year dataset (2012-2015) with wind observations from vanes and cup anemometers at 27, 58, and 92 m, from sonic anemometers at 85 m, and from a Zephir Lidar between 90 and 315 m at 25-m intervals. All instruments have an uncertainty in the order of 1% to 2%, and they correlate very well[27] ($R^2 = 0.99$). Additionally, pressure, temperature, and humidity were observed at 21 m (platform) and 90 m (tower top). All data are freely available from http://www.meteomastijmuiden.nl/data/. For more details, we refer to Kalverla et al.[25] In a preliminary study, we found that the model performance in terms of wind speed and direction at this location is very similar to other locations at the North Sea (also indicated in Figure 1). However, the observation height in combination with the distance to shore and the high quality of the MMIJ data is quite unique, and therefore, we focus on MMIJ in the presentation of the results.

The observations are available as 10-minute averages, whereas the model data from IFS and HARMONIE were available as hourly instantaneous model fields. WRF data for the MMIJ location was available at each time step, but for our consistency in the comparison, we used hourly instantaneous fields as well. The model data were aligned with the corresponding observational data (ie, the 10-min average closest to the full hour). This approach was validated using the WRF dataset. By using the high-frequency output and trying various combinations of averaging and interpolation, it was found that this approach leads to the smallest "artificial" model error statistics. Note that the instantaneous model values can be interpreted as spatial analogs of the corresponding time-averaged observations, although the exact analogy is dependent on the mean wind speed (eg, at a mean wind speed of 10 m/s, a grid cell of 3 km would represent the average wind over an interval of 300 s).

Model data from the nearest neighboring grid cells were linearly interpolated between the original model levels to match the MMIJ observation levels. All models were initialized at 12 UTC the day before the selected case date and forecast time series from 12 to 36 hours ahead were used for validation. This is a relevant time scale for power forecasts (eg, for power trading) and moreover, for this relatively short forecast horizon the effect of nonlinearities in the model equations is still small. As such, varying model performance will mainly be due to the (in)ability of the models to accurately capture the various atmospheric conditions.

2.3 | Case selection: the $uvs \cdot t_2$ method

The WRF simulations are run in house on a rather large domain, which limits the number of cases to be analyzed. Since we expect that the model performance will be different for different weather situations, we use a clustering algorithm to objectively select a wide variety of weather types. From previous experience,[25] we know that wind speed, direction, and atmospheric stability are useful parameters to distinguish between

* https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation

various wind regimes. However, these parameters can vary substantially in time or space within one case. In an intermittent turbulence regime, for example, there is no single characteristic Obukhov length to characterize the stability of the flow. This is because the Obukhov length is a very sensitive, *internal* parameter, based on local fluxes or gradients that are the result of (rather than the precondition for) the manifestation of boundary-layer processes.

To avoid this issue, the case selection is based on larger scale, *external* variables instead. The idea is that the observed conditions (eg, intermittent turbulence) are the manifestation of the external forcing: an airmass driven over a water surface by an ambient pressure field. It follows that for similar characteristics of the airmass, sea state, and pressure field, we expect similar flow regimes. This rationale is common in meteorological (downscaling) studies.[28] However, existing classifications are not optimized for the current application. First of all, the correlation with local observations is not always clear. For example, the Großwetterlagen classification[29] is not suitable because there can be many wind directions over the North Sea within one weather type. Moreover, existing classifications do not usually account for stability, while we know that stratification can make a large difference for the manifestation of the wind field. Therefore, we introduce a new clustering algorithm that better suits our needs.

We find that a suitable balance between large-scale weather type and the local flow is achieved if wind speed and direction are characterized in terms of the zonal and meridional geostrophic wind components u and v. Following Jones et al,[30] we use the south-north pressure difference over 10° centered over the area of interest to calculate the west-east (u) wind component and vice versa for v. To account for large-scale density stratification s, we use the difference between the potential temperature at the 850-hPa level and the SST. Here, the 850-hPa potential temperature comprises information about the air mass, while the SST characterizes the sea state. The difference between the two was found to be a more robust clustering parameter than each of them separately (based on the criteria presented in section 3.1).

Daily averages of large-scale u, v, and s are derived from ERA-interim[31] mean sea-level pressure, SST, and 850-hPa potential temperature fields in the 2012 to 2015 MMIJ observation period. To improve the characterization of the airmass, the history of the flow is also relevant.[28] Therefore, we additionally include the previous day to obtain a six-parameter dataset (hence the subscript in $uvs \cdot t_2$). We use a hierarchical clustering algorithm that iteratively (1) splits the largest cluster (judged by average cluster radius) along a plane perpendicular to its principle component, (2) computes new cluster centers, and (3) reassigns all points to the closest cluster center.[32]

Once the algorithm has identified 30 clusters, we select the cases that are closest to the cluster centers. The number 30 is similar to other weather classification such as the Lamb Weather Types[30] and Großwetterlagen[29] and is aimed at selecting a wide range of characteristic weather conditions rather than exactly representing the climatology. Nevertheless, we perform several checks to verify the representativeness of our cases, which will be further discussed in sections 3.1. For HARMONIE, model data for six of these cases were unavailable. Therefore, the second closest (to the cluster center) dates from these clusters were used instead (see Appendix B).

2.4 | Statistical analysis: error diagrams

The difference between model predictions p_i and observations o_i is defined as $e_i = p_i - o_i$ for each record i. A visualization in the form of a histogram or boxplot (Figure 2A) immediately provides insight into the nature of the errors. If the errors follow a normal distribution (which is a reasonable first approximation for our data), the complete set can be characterized by a mean error or bias, μ, and a standard deviation of the error, σ. It can be shown that these two fundamental properties are directly related to a third commonly used error statistic, the root mean square error (RMSE) as[33]

$$RMSE^2 = \mu^2 + \sigma^2. \tag{1}$$

We can use this property to concisely visualize model performance for many cases in a single diagram by plotting systematic errors μ and random errors σ on the x- and y-axes, respectively (Figure 2B). Interestingly, the distance to the origin in such diagrams represents the root mean square error, by virtue of Pythogoras' theorem. For convenience, we will refer to this type of figure as *error diagrams* in the remainder of this paper.

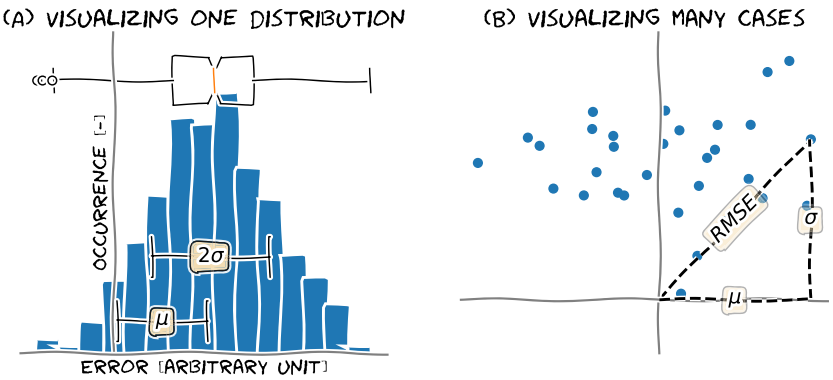

FIGURE 2 Various representations of a conceptual error distribution (A). In the boxplot visualization, the middle vertical bar shows the median, the width of the box represents the interquartile-range (IQR), the whiskers, located 1.5*IQR outside away from Q1 and Q3, delimit the total range, and circles denote outliers. The histogram visualization closely resembles the corresponding Gaussian probability distribution function characterized by mean μ and standard deviation σ. These can be used to visualize many cases in an *error diagram* (B) [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS AND DISCUSSION

In this section, the results are presented and discussed, including cross-comparison with existing studies. The results will be presented in the following order. Section 3.1 deals with the case selection. The overall results of the model validation are presented in section 3.2. Special attention is paid to the role of resolution in section 3.3 and to the representation of vertical gradients in section 3.4. Finally, the results are related to large-scale characteristics in section 3.5.

3.1 | Case selection

We start with an inspection of the selected cases. A summary of each case, comprising case date; large-scale wind speed, direction, and stability; and number of cases in the corresponding cluster, is available in Appendix B. Figure 3 shows all 30 cases on a *v* versus *u* diagram. We find that the strongest winds are southwesterly, which corresponds to the typical climatology of the North Sea. Observed static stability at MMIJ varies within each case, but the relative occurrence of stable stratification is consistently larger for cases with a stronger large-scale density stratification (the latter is indicated by the color bar). The Weibull shape and scale parameters for both the full dataset (2.11 and 11.80 m/s) and the subset of 30 cases (2.23 and 11.87 m/s) turn out to be very similar. We have further checked that the cases adequately sample the annual cycle of sea-surface temperature: the cases are equally distributed over the year. Surface pressure charts (see Appendix B) indicate that we have sampled a broad range of synoptic circulation types, and moreover, the surface pressure chart for each case does not deviate too much from the average surface pressure chart for its cluster. The correlation coefficients between the large-scale *u* and *v* and the IJmuiden observations are about 0.9, demonstrating that these clustering variables are indeed a very suitable "golden mean" between large-scale circulation and local conditions.

To assess the uncertainty in the computed error metrics due to the case selection, we evaluated additional IFS forecasts for the full MMIJ period and calculated error statistics for a varying number of cases (Figure 4). The full lines indicate the algorithm used in the $uvs \cdot t_2$ selection (principal component–based partitioning, see Section 2.3), while the shading is based on a k-means algorithm (using the same input variables). The latter involves a random initialization step allowing us to estimate to some degree the uncertainty in the estimate. The figure demonstrates that this uncertainty is relatively small for most standard deviations but can be relatively large with respect to the bias since the bias is often very small. For wind speed and potential temperature, the uncertainty has more or less converged at 30 cases, while for mixing ratio, our partitioning algorithm deviates from the k-means uncertainty bands. With respect to the representation of vertical gradients, the uncertainty is larger, and especially, the standard deviation converges slowly.

Summarizing, we have sampled a variety of cases that allows us to investigate model performance under a wide range of forcing conditions. Care should be taken to generalize these results quantitatively, because the uncertainty due to the limited amount of cases cannot always be neglected. However, the quick convergence of the error statistics is promising and suggests that this method could be extended to downscaling climatologies as well. Perhaps our choice of large-scale variables could be combined with the subset selection strategy of Rife et al,[34] who demonstrate that it is possible to obtain a fully representative subset of a climatology using 180 days.

3.2 | Overall performance

Figure 5 shows the error distribution of wind speed, wind direction, potential temperature, and mixing ratio at different altitudes for each model. Means and standard deviations of these distributions are listed in Table 1. These results are weighed for the cluster size (see Appendix B). Note that the boxplots are quite symmetric, which supports the assumption that the errors approximately follow a normal distribution.

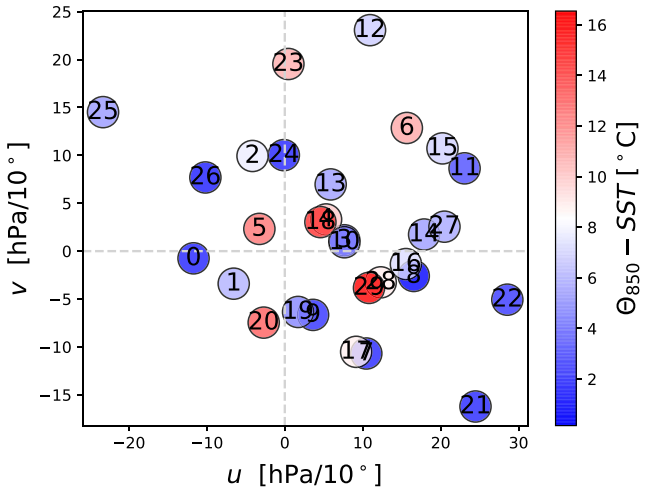

FIGURE 3 Illustration of the clustering results. For each of the selected cases, the *x*- and *y*-axes show the *u* and *v* pressure gradients, respectively. The color bar indicates the large-scale stability [Colour figure can be viewed at wileyonlinelibrary.com]

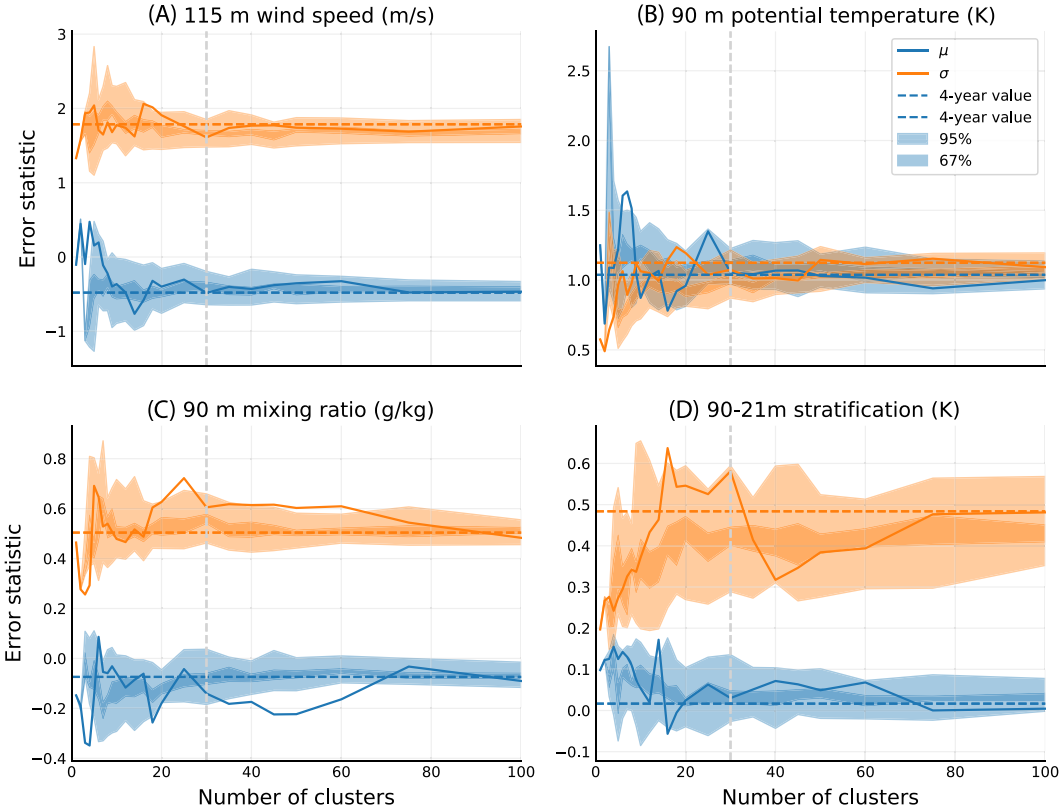

FIGURE 4 Uncertainty in selected error metrics as a function of the number of cases included. A, 115-m wind speed; B, 90-m potential temperature; C, 90-m mixing ratio; and D, vertical difference in virtual potential temperature between 315 and 90 m [Colour figure can be viewed at wileyonlinelibrary.com]

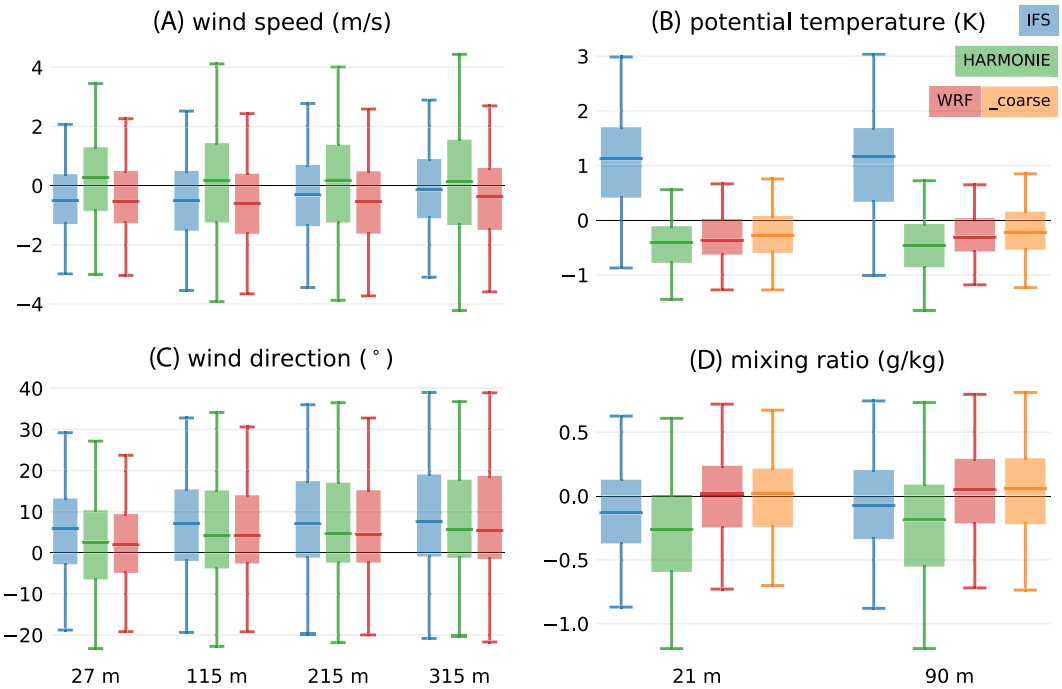

FIGURE 5 Boxplots of the error distribution for wind speed, direction, temperature, and humidity for all models at different heights. Lead time 12 to 36 hours. See Figure 2 for the definition of the boxplot. Outliers not shown [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Mean error (μ) and centered root mean square error (σ) for each model for different variables: wind speed v (m/s) and direction d (°), (virtual) potential temperature $\theta_{(v)}$ (K), and mixing ratio r (g/kg); subscripts are altitudes[a]

Model	IFS		HARMONIE		WRF		WRF coarse	
Variable	μ	σ	μ	σ	μ	σ	μ	σ
v_{27}	−0.48	**1.26**	**0.19**	1.87	−0.25	1.77	−0.24	1.95
v_{115}	−0.48	**1.59**	**0.02**	2.17	−0.49	1.93	−0.51	2.13
v_{215}	−0.33	**1.61**	**−0.02**	2.23	−0.47	2.02	−0.60	2.17
v_{315}	−0.12	**1.59**	**0.05**	2.27	−0.33	2.05	−0.50	2.19
d_{27}	5.97	**20.22**	**1.49**	25.77	2.07	23.71	3.12	26.66
d_{115}	8.60	**22.24**	**5.74**	25.48	6.52	24.42	6.83	26.41
d_{215}	9.74	**23.57**	**8.41**	27.68	8.67	26.71	9.01	27.14
d_{315}	1.02	**24.40**	**9.00**	26.77	10.27	28.19	9.02	29.93
θ_{21}	1.00	0.94	−0.43	**0.58**	−0.19	0.75	**−0.10**	0.87
θ_{90}	1.03	1.06	−0.54	0.95	−0.17	**0.83**	**−0.03**	1.03
r_{21}	−0.12	**0.40**	−0.27	0.51	**−0.01**	0.48	**−0.01**	0.45
r_{90}	−0.14	0.61	−0.24	**0.60**	−0.03	0.64	**−0.02**	0.62
$\theta_{v,21}$	0.99	0.94	−0.48	**0.57**	−0.20	0.75	**−0.10**	0.89
$\theta_{v,90}$	1.01	1.04	−0.58	0.97	−0.18	**0.80**	**−0.04**	0.99
Δv	0.37	**1.35**	**−0.06**	1.72	0.07	1.76	−0.13	1.82
Δd	**5.69**	14.12	7.87	20.87	7.30	17.40	6.30	**13.56**
$\Delta\theta_v$	0.03	0.58	−0.10	0.70	**0.02**	0.62	0.07	**0.57**

[a]Δ indicates difference between 315 and 27 m for wind and between 90 and 21 m for θ_v. Bold numbers emphasize best scores.

Typical wind speed biases are within 0.5 m/s, and σ is approximately 2 m/s. On average, all models slightly underestimate the wind speed. The IFS is virtually unbiased at 315 m but underestimates winds closer to the surface. HARMONIE has the smallest wind speed bias but, shows a larger spread than the other models. These findings are in line with previous results. The ECMWF publishes standard validation metrics, demonstrating that the mean bias and standard deviation of the errors in 10 m wind speed over Europe at forecast hour 60 are within 0.5 and 2.0 m/s (see Haiden et al,[35] figure 27). For the IFS ensemble system, the RMSE for short-range wind speed forecasts over land is reported to be around 2 m/s.[36] For (earlier versions of) WRF, Hahmann et al[37] find small systematic errors (0.09 m/s) in 100-m wind speed predictions at the FINO-1 mast, and Krogsæter and Reuder[38] find mean absolute errors of about 2 m/s. Similar errors are reported in Draxl et al,[39] who find a wind speed bias ranging from −0.33 to +0.68 m/s and RMSE ranging from 1.95 to 2.69 m/s. Carvalho et al[40] report wind speed biases larger than 0.5 m/s and σ roughly between 1.5 and 2.5 m/s. While Carvalho et al[40] attribute the positive bias they find to an excessively smooth sea surface, Hahmann et al[37] and Draxl et al[39] find their results to be sensitive to the boundary-layer parameterization used. Given this sensitivity, recent changes in the WRF model and the MYNN boundary-layer scheme might well explain the minor differences we see with our results. The smaller wind speed bias in the HARMONIE simulations might be the result of the additional data assimilation, although we were not able to run additional experiments to verify this hypothesis. With respect to the modifications to the IFS[19] of November 2013, it is worth mentioning that in our results, there is no marked improvement in the representation of surface wind speed. Observed differences are generally small, not consistent with height, and may be related to the frequency of stable conditions or other factors rather than the modifications to the IFS.

In all models, the wind direction is veered (turned clockwise) with respect to the observations by up to 10° on average (Table 1). Performance is better near the surface than at 315 m. At 27 m, HARMONIE and WRF represent the wind direction better than the IFS, but at the other altitudes, performance is similar. The misrepresentation of surface winds has previously been attributed to excessive mixing under stable conditions, and the recent modifications to the IFS are expected to improve this deficiency.[19] The ECMWF website reports a decrease in the systematic wind direction error near the surface over land since then (<5° after 60 hr). Also, in our results, there seems to be an improvement, but we cannot conclude that this is due to the modifications, because the frequency and intensity of stable stratification are generally lower for the cases after November 2013.

The IFS shows a warm bias of roughly 1 K and also a large spread ($\sigma \approx 1$ K). Oppositely, WRF and HARMONIE underestimate the temperature (at 21 m by 0.19 and 0.43 K, respectively) and show a smaller spread. Draxl et al[39] also evaluated temperature at 2 and 100 m from WRF simulations and found a positive bias at 2 m and either a positive or negative bias at 100 m depending on the boundary-layer parameterization. They report typical RMSE scores of about 1.3 K. Thus, our findings show improved performance. The most important difference is that Draxl et al[39] only used data from October 2009. The ECMWF high-resolution verification chart of 2-m temperature shows that the bias is typically within −0.5 and +0.5 K, which is substantially smaller than our results (see Haiden et al,[35] figure 24). This is an average over a large area and based on synop observations over land, so it might not be directly comparable. Nevertheless, we had expected that offshore performance would be relatively well represented, because the surface is much more homogeneous. Mixing ratio is best represented by WRF. Both IFS and HARMONIE results show a dry bias, approximately 0.1 and 0.2 g/kg, respectively. Standard deviation amounts to about 0.5 g/kg for all models.

3.3 | Role of model resolution

Since we include model results from both the relatively coarse grid of the IFS and the refined grids of HARMONIE and WRF, the questions arise whether differences in results can be attributed to differences in model resolution and whether a higher resolution contributes to an improved forecast. Theoretically, we would expect that the higher resolution models are able to incorporate local features that are not resolved by the IFS. Also, a coarser resolution would effectively smooth the time series at a given location, as it can only represent the grid-cell average. Indeed, Table 1 shows that the IFS has the lowest standard deviation of the error for most variables, which is consistent with a smoother signal.

However, there are many other differences between the three models, so in order to isolate the effect of increased resolution, we performed additional WRF simulations on a resolution of 16 km, comparable with the IFS. Some of these additional results are included in the right subplots of Figure 5 and in Table 1. It turns out that the change in resolution has only a small impact on these results. For most variables, the spread of the errors is larger for the coarser resolution runs, instead of smaller.

Thus, the difference in model resolution does not seem to be responsible for differences between model results in this study and increased resolution does not necessarily lead to better results.

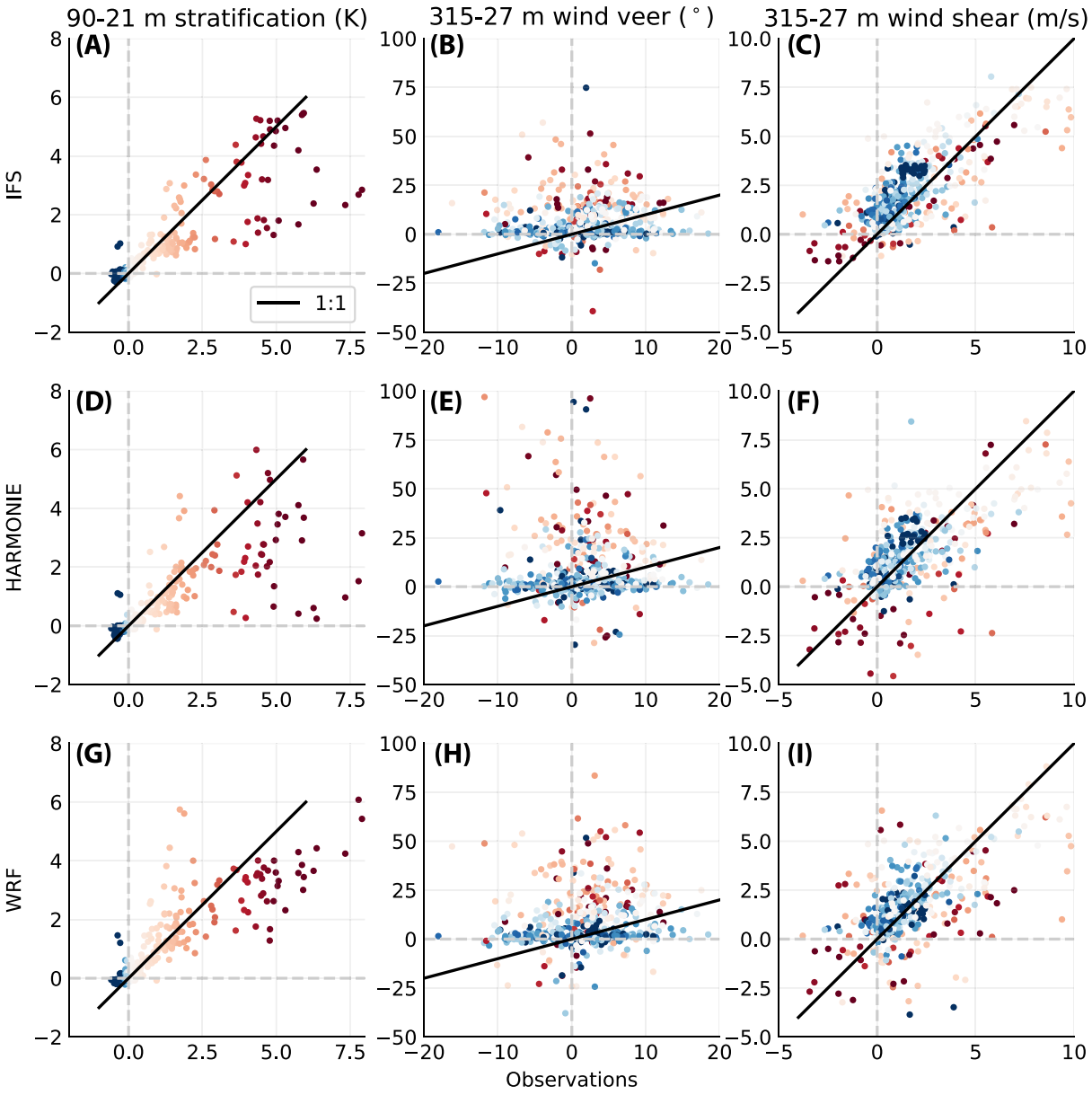

FIGURE 6 Modeled versus observed: virtual potential temperature difference between 90 and 21 m (left), wind direction difference between 315 and 27 m (middle), and wind speed difference between 315 and 27 m (right) for the IFS (top), WRF (middle), and HARMONIE (bottom). Red and blue are used, respectively, for stable and unstable observed stratification (the color scale is not linear) [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Boundary-layer structure

In the previous sections, we evaluated the model performance at individual altitudes. For wind energy purposes it is especially important that the models adequately reproduce the vertical boundary-layer structure as well. Misrepresentation of wind shear or density stratification can have adverse effect on turbine loads[3] and power yield.[8] Besides, the gradients of wind and temperature are intimately linked through the process of turbulent mixing. Table 1 also lists statistics of vertical gradients, but in this case, scatter plots are helpful to further understand the model performance (Figure 6).

We first discuss the temperature stratification, which was calculated as the difference in virtual potential temperature between the top of the mast (90 m) and the platform (21 m). While neutral conditions represent most of the data, most scatter occurs for stable conditions. In the IFS results, we observe two different types of behavior: in one branch, the strong stable stratification is well represented, whereas in the other branch, the stratification is underestimated (Figure 6A). The HARMONIE results also show a small branch in which the stratification is not captured by the model, while WRF shows a more continuous behavior where intermediate to strong stratification is generally underestimated.

The difference in wind direction with height (calculated between 315 and 27 m) is not well represented by any of the models. Mostly, the models do not capture any wind veer at all, irrespective of the observed wind veer. If they do represent wind veer, it is often far outside the range of observed veer. As indicated by the color scale, this is usually associated with stable conditions.

Modeled wind (speed) shear over the same layer correlates reasonably with the observations, with most scatter during stable conditions. The bottom-left quadrant in these figures indicates negative shear that is captured by the model. This is a characteristic feature of low-level jets, which are always associated with stable conditions.

The misrepresentation of stable boundary layers has received considerable attention in the past. For example, Brown et al[41] evaluated the representation of wind veer over sea in the IFS model in 2005 and found a significant underestimation of wind turning across the boundary layer. The largest misrepresentation occurred for stable conditions or cases with strong baroclinity (horizontal temperature contrasts). Svensson and Holtslag[42] related this issue to excessive mixing under stable conditions. This excessive mixing, which was also reported by Holtslag et al,[43] inspired Sandu et al[44] to make the aforementioned modifications to the IFS in November 2013, which were expected to improve the representation of wind under stable conditions.[19] In our results, we do not see a clear difference in the representation of vertical gradients for the cases before and after November 2013. However, the frequency and strength of stable stratification were considerably lower in the cases after this update, so we cannot conclude whether the representation of (very) stable conditions improved or not based on our results. With respect to the other modification, the addition of vertical levels in June 2013, there is one case with very stable stratification after this date (but before November 2013), which is reasonably well represented. Obviously, this is not enough to be conclusive, but it supports the hypothesis that the diverging model behavior, which we noticed in the scatter plot of temperature stratification, is related to modifications to the model.

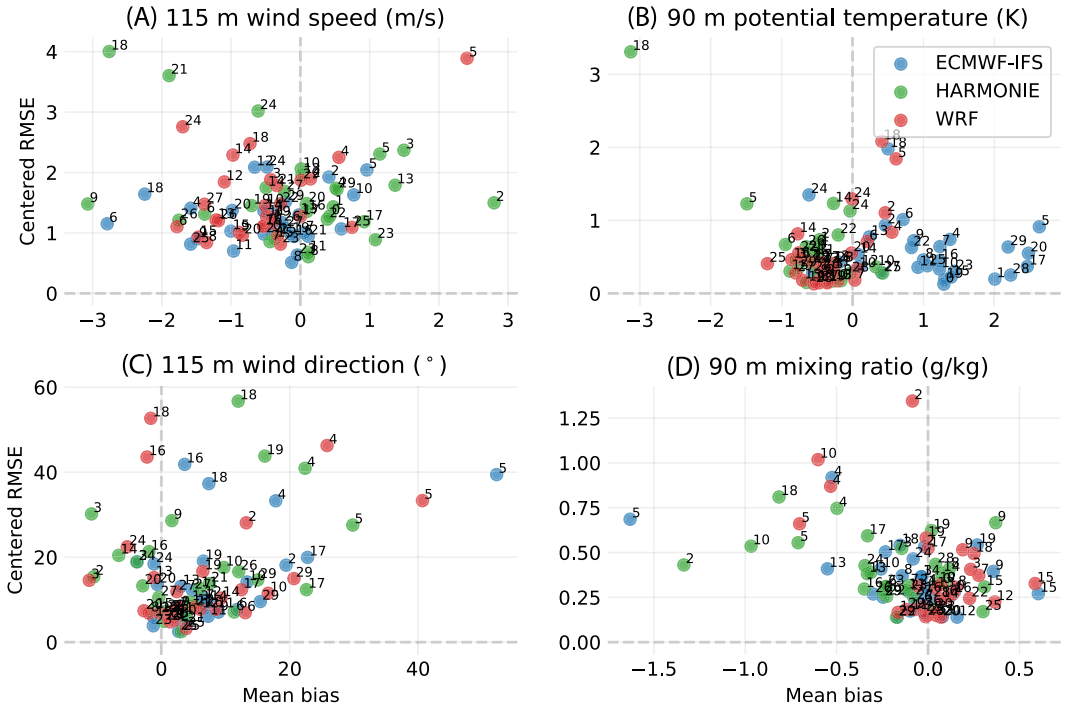

FIGURE 7 (A, C) Error diagrams of 115-m wind speed and direction and (B, D) 90-m potential temperature and mixing ratio at met mast IJmuiden for the IFS, HARMONIE, and The WRF models for lead times up to 36 hours. See section 2.4 for an explanation of the error diagrams [Colour figure can be viewed at wileyonlinelibrary.com]

3.5 | A case against the average

Up to this point, we have focused on overall error distributions and obtained an impression of average model performance. However, if we focus on individual cases, we find substantial variability in model performance. To gain insight in the variability between individual cases, we will inspect the error diagrams of 115 m (≈ hub height) wind and 90-m potential temperature and humidity (Figure 7). Recall that the absolute distance to the origin represents the overall RMSE.

For hub-height wind speed (Figure 7A), most cases cluster in an area with a mean bias between −1.5 and +1 m/s and σ of about 1.5 m/s. Again, we find a tendency to underestimate wind speeds on average. Exceptionally poor performance is found for cases 5 (WRF); 6 (IFS); 2, 9, and 21 (HARMONIE); 24 (WRF and HARMONIE); and 18 (all models). These are almost exclusively spring and summer cases, in which the sea water is relatively cold and stable conditions occur frequently. For wind direction (Figure 7C), two third of the cases are clustered within 15° on both axes, without obvious differences between models. Cases 5 and 18, which are among the worst represented cases, are both characterized by weak-wind regimes, for which wind direction is usually less well defined. Nevertheless, the positive direction bias (modeled wind direction is veered w.r.t. the observations) is clearly present for nearly all individual cases.

For temperature (Figure 7B), we find a remarkable difference in performance between the three models. The cold bias of about 0.5 K in WRF and HARMONIE is clearly evident for all cases, and very well defined. In contrast, the IFS bias is very "case-sensitive": It varies between −1 and +3 K. Cases 5 and 18, which we already highlighted for poorly representing wind speed, are also outliers in the temperature plot. These cases

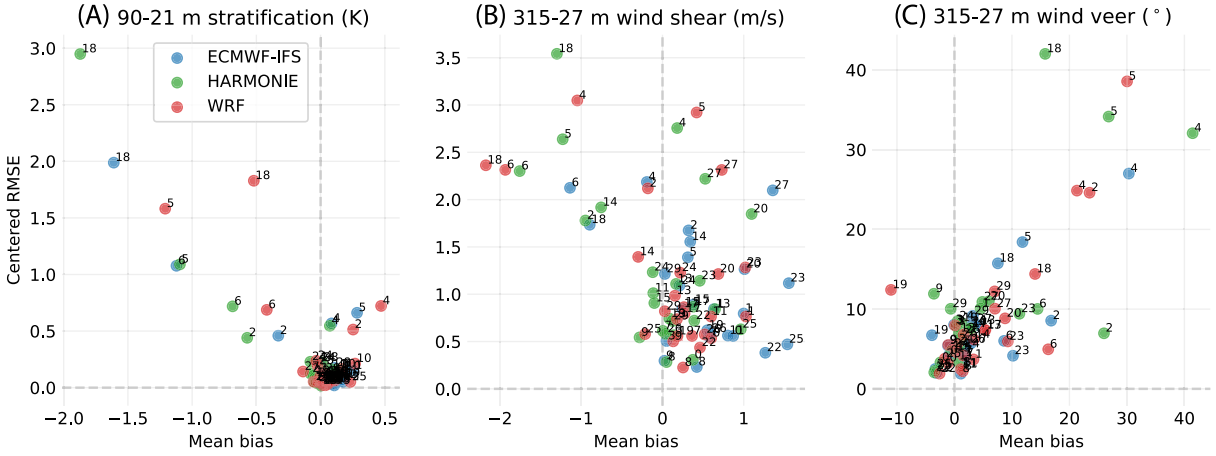

FIGURE 8 Error diagrams of A, 90- to 21-m virtual potential temperature and B, 315- to 27-m wind speed and C, wind direction differences [Colour figure can be viewed at wileyonlinelibrary.com]

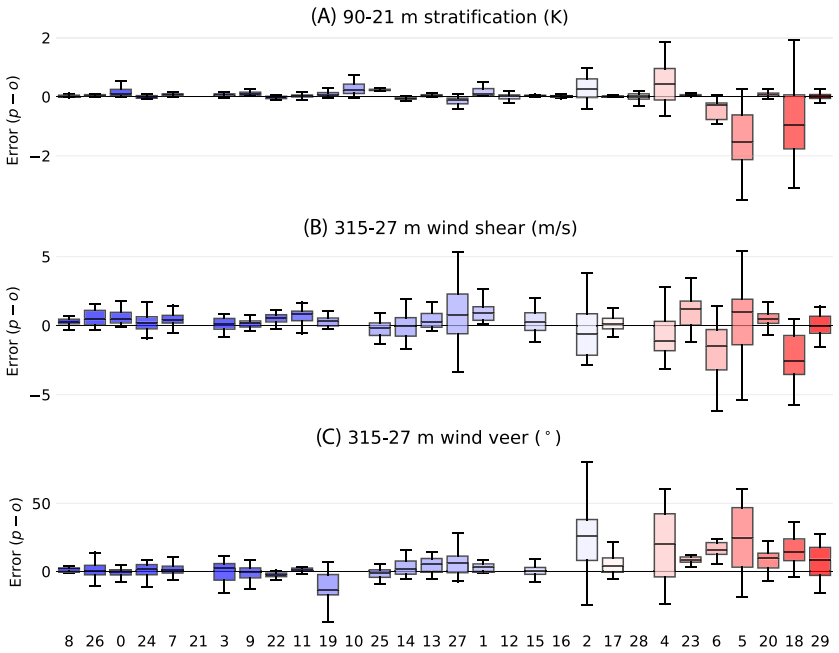

FIGURE 9 Error distributions of A, 90- to 21-m virtual potential temperature and B, 315- to 27-m wind speed and C, direction differences for WRF, sorted by increasing large-scale stability ($\theta_{850} - SST$) [Colour figure can be viewed at wileyonlinelibrary.com]

share a characteristic large-scale stable stratification ($\theta_{850} - SST > 10$ K). Errors in humidity are quite concentrated (Figure 7D), but remarkably, all cases that have a relatively large variability also have a negative bias. This points to an underlying process that creates variability in moisture and the misrepresentation of which is associated with too dry boundary layers. Examples of such processes are frontal passages, precipitation, or dry-air entrainment at the boundary-layer top.

A similar figure for the vertical difference in virtual potential temperature (Figure 8A) indicates that all models typically perform very well (see also Table 1) but a few cases are actually responsible for almost all of the overall error. Cases 2, 4, 5, 6, and 18 perform particularly poorly. Exactly these cases also underestimate the wind speed gradient (Figure 8B) and have a large positive bias and spread in the wind direction gradient (ie, too little backing or too much veering with height; Figure 8D). These cases are characterized by (strong) stable stratification, or weak winds, or both.

The role of stability becomes especially clear from Figure 9, where we sorted the boxplots by the characteristic large-scale stability for that case. The representation of vertical boundary-layer structure is especially poor if the large-scale stability is strong, ie, when relatively warm air mass is advected over a cold sea surface. This is indisputable for the vertical differences shown but also apparent for 90-m potential temperature and 115-m wind direction (not shown). For 115-m wind speed and 90-m mixing ratio, the difference in model performance between stable and unstable conditions is less pronounced. Similar results are found for the other two models. It is worth noting that within each case, we find a both stable and unstable conditions locally (based on the temperature gradient near the surface). However, stable conditions are much more frequently observed for cases characterized by a strong large-scale stratification.

4 | SUMMARY AND CONCLUSIONS

We have presented an evaluation of three mainstream meteorological models against unique observations up to 315 m at the North Sea. The IFS is the operational global weather model of the ECMWF, HARMONIE is the operational limited-area model used by a number of European weather agencies, and WRF-ARW is a community research model. These models are used in wind energy systems for forecasting, resource assessment, and increasingly to provide input for microscale simulations of wind-turbine interactions. The observations from MMIJ provide a unique opportunity to validate these models at a far (85 km) offshore location up to 315 m.

Our results show that there is substantial variability in model performance. Both WRF and the IFS consistently underestimate the wind speed by about 0.5 m/s. The spread of the errors, expressed as standard deviation of the error distribution, is below 2 m/s. HARMONIE has a smaller wind speed bias than the other models, which may be due to the additional data assimilation used in this model. The spread in the errors is smallest for the IFS, and we have shown that this is not due to the coarser grid.

Wind direction is found to be biased by up to 10°, ie, the model winds are slightly veered (turned clockwise) with respect to the observations. One might argue that a wind direction mismatch of $<10°$ is not particularly harmful. However, as Gaumond et al[45] point out, the wake losses are very sensitive to wind direction due to the typical offshore wind farm layouts with close turbine distances for certain directions. Furthermore, both WRF and HARMONIE show a consistent temperature bias of about −0.5 K, while the IFS shows a positive bias of about 1 K.

Additionally, we find that the model results are very "case-sensitive." All models struggle with the representation of stable conditions, leading to unrealistic boundary-layer structures. This is especially important because both power production (through enhanced wake losses) and turbine loads (through enhanced wind shear) are affected. Recently, Baas et al,[46] in line with Kalverla et al,[25] reported that over sea, strong wind shear can be maintained even under moderate to strong wind conditions if the density stratification is strong enough.

Our case-selection strategy, based on the clustering of relevant large-scale variables, was not only able to select a wide range of circulation patterns but also provided a reasonable representation of the underlying climatology. However, with the limited number of 30 cases included in this study, one should be careful to generalize the overall model scores presented in this study. This is especially true because the results in section 3.5 illustrate that the overall model performance as listed in Table 1 cannot always be projected on individual cases.

Finally, our results indicate that the large-scale stability, expressed as the difference between the potential temperature at 850 hPa and the SST, can be used to estimate the accuracy of model results. Since the large-scale stratification is easily determined, and also more robust and generic than the local stability, it could serve as guidance during the interpretation of model results.

ACKNOWLEDGEMENTS

This research is supported by the Dutch Technology Foundation NWO-TTW, which is partly funded by the Ministry of Economic Affairs (TTW project number: 14185). IJmuiden observations are publicly available via ECN, IFS model results were available from the ECMWF operational archive, and thate HARMONIE results were kindly provided by the Dutch weather service KNMI. Personal communication with Sukanta Basu has greatly improved our insight into model evaluation statistics. The numerical simulations were performed with the supercomputer facilities at SURFsara and financially sponsored by the Netherlands Organisation for Scientific Research (NWO) Physical Science Division (project number SH-312-15)

ORCID

Peter Kalverla http://orcid.org/0000-0002-5025-7862
Gert-Jan Steeneveld http://orcid.org/0000-0002-5922-8179

REFERENCES

1. Rodrigues S, Restrepo C, Kontos E, Pinto RT, Bauer P. Trends of offshore wind projects. *Renewable Sustainable Energy Rev*. 2015;49:1114-1135.
2. Archer CL, Jacobson MZ. Evaluation of global wind power. *J Geophys Res Atmos*. 2005;110(D12):D12110.
3. Park J, Basu S, Manuel L. Large-eddy simulation of stable boundary layer turbulence and estimation of associated wind turbine loads. *Wind Energy*. 2014;17(3):359-384.
4. Schulz C, Letzgus P, Lutz T, Krämer E. Cfd study on the impact of yawed inflow on loads, power and near wake of a generic wind turbine. *Wind Energy*. 2017;20(2):253-268. we.2004.
5. O'Brien J, Young T, O'Mahoney D, Griffin P. Horizontal axis wind turbine research: a review of commercial cfd, fe codes and experimental practices. *Prog Aerosp Sci*. 2017;92:1-24.
6. Mehta D, van Zuijlen A, Koren B, Holierhoek J, Bijl H. Large eddy simulation of wind farm aerodynamics: A review. *J Wind Eng Ind Aerodyn*. 2014;133:1-17.
7. Sanderse B, van der Pijl S, Koren B. Review of computational fluid dynamics for wind turbine wake aerodynamics. *Wind Energy*. 2011;14(7):799-819.
8. Dörenkämper M, Witha B, Steinfeld G, Heinemann D, Kühn M. The impact of stable atmospheric boundary layers on wind-turbine wakes within offshore wind farms. *J Wind Eng Ind Aerodyn*. 2015;144:146-153. Selected papers from the 6th International Symposium on Computational Wind Engineering CWE 2014.
9. Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Rev*. 2014;31:762-777.
10. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renewable Energy*. 2012;37(1):1-8.
11. Sanz Rodrigo J, Chávez Arroyo RA, Moriarty P, et al. Mesoscale to microscale wind farm flow modeling and evaluation. *Wiley Interdiscip Rev Energy Environ*. 2017;6(2):e214.
12. Talbot C, Bou-Zeid E, Smith J. Nested mesoscale large-eddy simulations with wrf: Performance in real test cases. *J Hydrometeorology*. 2012;13(5):1421-1441.
13. Petersen EL, Troen I, Jørgensen HE, Mann J. Are local wind power resources well estimated? *Environ Res Lett*. 2013;8(1):011005.
14. McCaa J, Cline JW, Shaw WJ, et al. Overview of the Wind Forecasting Improvement Project in Complex Terrain (WFIP2). Presented at the 97th American Meteorological Society Annual Meeting, Seattle, WA; 2017.
15. Muñoz-Esparza D, Kosović B, Mirocha J, van Beeck J. Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models. *Boundary-Layer Meteorol*. 2014;153(3):409-440.
16. Mirocha J, Kosović B, Kirkil G. Resolved turbulence characteristics in large-eddy simulations nested within mesoscale simulations using the weather research and forecasting model. *Mon Weather Rev*. 2014;142(2):806-831.
17. Sanz Rodrigo J, Churchfield M, Kosovic B. A methodology for the design and testing of atmospheric boundary layer models for wind energy applications. *Wind Energy Sci*. 2017;2(1):35-54.
18. Petersen EL, Troen I. Wind conditions and resource assessment. *Wiley Interdiscip Rev Energy Environ*. 2012;1(2):206-217.
19. Riddaway B. ECMWF Newsletter No.138 - Winter 2013/14. https://www.ecmwf.int/en/elibrary/14581-newsletter-no138-winter-2013-14
20. Bengtsson L, Andrae U, Aspelien T, et al. The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System. *Mon Weather Rev*. 2017;145(5):1919-1935.
21. Skamarock WC, Klemp JB. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J Comput Phys*. 2008;227(7):3465-3485. Predicting weather, climate and extreme events.
22. Grell GA, Freitas SR. A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos Chem Phys*. 2014;14(10):5233-5250.
23. Fowler LD, Skamarock WC, Grell GA, Freitas SR, Duda MG. Analyzing the Grell-Freitas convection scheme from hydrostatic to nonhydrostatic scales within a global model. *Mon Weather Rev*. 2016;144(6):2285-2306.
24. Werkhoven EJ, Verhoef JP. Offshore Meteorological Mast IJmuiden – Abstract of Instrumentation Report. Technical Report ECN-Wind Memo-12-010, ECN; 2012.
25. Kalverla PC, Steeneveld G-J, Ronda RJ, Holtslag AA. An observational climatology of anomalous wind events at offshore meteomast IJmuiden (North Sea). *J Wind Eng Ind Aerodyn*. 2017;165:86-99.
26. Holtslag M, Bierbooms W, van Bussel G. Validation of surface layer similarity theory to describe far offshore marine conditions in the Dutch North Sea in scope of wind energy research. *J Wind Eng Ind Aerodyn*. 2015;136:180-191.
27. Maureira Poveda J, Wouters D. Wind measurements at meteorological mast IJmuiden. Technical Report ECN-E–14-058, Energieonderzoek Centrum Nederland; 2015.
28. Philipp A, Bartholy J, Beck C, et al. Cost733cat—a database of weather and circulation type classifications. *Phys Chem Earth Parts A/B/C*. 2010;35(9):360-373.
29. Hess P, Brezowsky H. Katalog der Großwetterlagen Europas; 1969.
30. Jones PD, Harpham C, Briffa KR. Lamb weather types derived from reanalysis products. *Int J Climatol*. 2013;33(5):1129-1139.

31. Dee DP, Uppala SM, Simmons AJ, et al. The era-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorolog Soc*. 2011;137(656):553-597.
32. Eggels AW, Crommelin DT, Witteveen JAS. Clustering-based collocation for uncertainty propagation with multivariate correlated inputs, arXiv preprint arXiv:1703.06112. Particularly section 3.2; 2017.
33. Murphy AH. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon Weather Rev*. 1988;116(12):2417-2424.
34. Rife DL, Vanvyve E, Pinto JO, Monaghan AJ, Davis CA, Poulos GS. Selecting representative days for more efficient dynamical climate downscaling: application to wind energy. *J Appl Meteorol Climatol*. 2013;52(1):47-63.
35. Haiden T, Bauer P, Bidlot J-R, et al. Evaluation of ecmwf forecasts, including 2013-2014 upgrades. *Technical Memorandum*. Tecnical Memorandum 742; 2014.
36. Pinson P, Hagedorn R. Verification of the ecmwf ensemble forecasts of wind speed against analyses and observations. *Meteorol Appl*. 2012;19(4):484-500.
37. Hahmann AN, Vincent CL, Peña A, Lange J, Hasager CB. Wind climate estimation using WRF model output: method and model sensitivities over the sea. *Int J Climatol*. 2015;35(12):3422-3439.
38. Krogsæter O, Reuder J. Validation of boundary layer parameterization schemes in the weather research and forecasting (WRF) model under the aspect of offshore wind energy applications–part II: boundary layer height and atmospheric stability. *Wind Energy*. 2015;18(7):1291-1302.
39. Draxl C, Hahmann AN, Peña A, Giebel G. Evaluating winds and vertical wind shear from weather research and forecasting model forecasts using seven planetary boundary layer schemes. *Wind Energy*. 2014;17(1):39-55.
40. Carvalho D, Rocha A, Gómez-Gesteira M, Santos CS. Offshore wind energy resource simulation forced by different reanalyses: comparison with observed data in the Iberian Peninsula. *Appl Energy*. 2014;134:57-64.
41. Brown AR, Beljaars ACM, Hersbach H, Hollingsworth A, Miller M, Vasiljevic D. Wind turning across the marine atmospheric boundary layer. *Q J R Meteorolog Soc*. 2005;131(607):1233-1250.
42. Svensson G, Holtslag AAM. Analysis of model results for the turning of the wind and related momentum fluxes in the stable boundary layer. *Boundary-Layer Meteorol*. 2009;132(2):261-277.
43. Holtslag AAM, Svensson G, Baas P, et al. Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. *Bull Am Meteorol Soc*. 2013;94(11):1691-1706.
44. Sandu I, Beljaars A, Bechtold P, Mauritsen T, Balsamo G. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (nwp) models? *J Adv Model Earth Syst*. 2013;5(2):117-133.
45. Gaumond M, Réthoré P-E, Ott S, Peña A, Bechmann A, Hansen KS. Evaluation of the wind direction uncertainty and its impact on wake modeling at the horns rev offshore wind farm. *Wind Energy*. 2014;17(8):1169-1178.
46. Baas P, Bosveld FC, Burgers G. The impact of atmospheric stability on the near-surface wind over sea in storm conditions. *Wind Energy*. 2016;19(2):187-198. WE-14-0090.R2.
47. Benjamin SG, Weygandt SS, Brown JM, et al. A north american hourly assimilation and model forecast cycle: the rapid refresh. *Mon Weather Rev*. 2016;144(4):1669-1694.

How to cite this article: Kalverla P, Steeneveld G-J, Ronda R, Holtslag AAM. Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast IJmuiden at the North Sea. *Wind Energy*. 2018;1–15. https://doi.org/10.1002/we.2267

APPENDIX A: WRF SET-UP

We used WRFV3.9.[21] The WRF settings were chosen such that they represent the latest trends and developments in mesoscale modeling. Most of the parametrization schemes have undergone active development in recent years. Settings are roughly similar to the NOAA high-resolution rapid-refresh model and the NCAR 3-km ensemble.[47] The following physics options were used:

- MYNN2.5 with mass-flux (stem) and updated length scale; tke-advection on
- MYNN surface layer
- RRTMG for both short- and longwave radiation
- Thompson microphysics
- Grell-Freitas scale-adaptive convection
- Noah land surface

The single domain consisted of 600-by-600 grid cells of each 3 × 3 km centered around the IJmuiden mast using a Lambert projection with true latitudes at 45 and 55°. In the vertical, we used the ECMWF L137 formulation up to 50 hPa (91 levels in total). Additional sensitivity tests on a coarser resolution used the ECMWF L91 formulation for cases before 2013/06/25 to maximize the similarity to the IFS simulations. Simulations were run with a time step of 10 s.

APPENDIX B: SELECTED CASES

Table B1 lists details of the selected cases. Synoptic charts for these dates are shown in Figure B1.

TABLE B1 All cases, sorted by month, with their average large-scale wind speed, direction and stability and number of days in the corresponding cluster[a]

#	Date	Wind Speed, hPa/10°	Wind Direction, °	Stability, K	N
29[b]	2013-01-05	11.5	289.5	16.5	12
14	2014-01-09	18.0	264.4	5.2	37
20[b]	2013-02-28	7.9	20.0	13.5	53
12	2014-02-25	25.5	205.4	6.7	33
25[b]	2013-03-23	27.5	121.9	5.1	3
28	2015-03-10	12.7	284.8	9.0	24
0[b]	2013-04-01	11.7	86.3	1.2	58
6[b]	2013-04-16	20.3	230.6	11.1	54
18[b]	2013-04-25	5.5	236.4	15.0	35
21	2015-04-01	29.4	303.5	1.4	16
19	2014-06-22	6.6	345.0	4.4	60
9	2014-06-30	7.6	331.4	1.7	75
5[c]	2013-07-23	4.0	125.4	12.7	68
10	2014-07-31	7.7	263.1	4.6	94
4	2015-07-11	6.2	238.0	10.0	106
2[b]	2012-08-14	10.8	157.4	7.8	65
7[c]	2013-08-13	15.0	315.6	1.3	55
24[b]	2012-09-26	10.0	179.3	1.3	25
8[b]	2012-09-29	16.7	278.9	0.2	54
1	2014-09-12	7.4	62.6	5.9	94
16	2014-09-26	15.6	275.0	7.2	33
3	2015-09-22	7.7	261.6	1.4	106
13	2014-10-17	9.1	220.2	5.2	95
15	2014-10-26	22.9	242.1	6.9	51
23	2015-10-30	19.5	181.4	10.9	28
17	2013-11-28	13.9	318.9	8.8	14
26	2014-11-17	12.8	127.0	1.0	29
22	2015-11-19	29.0	280.1	1.9	17
11	2013-12-20	24.6	249.5	2.7	41
27	2014-12-17	20.6	262.9	5.6	25

[a]Bold numbers are second choice cases, because HARMONIE data were missing for the first choice (ie, closest to cluster centers).

[b]Before update IFS of June 2013.

[c]Before update IFS of November 2013.

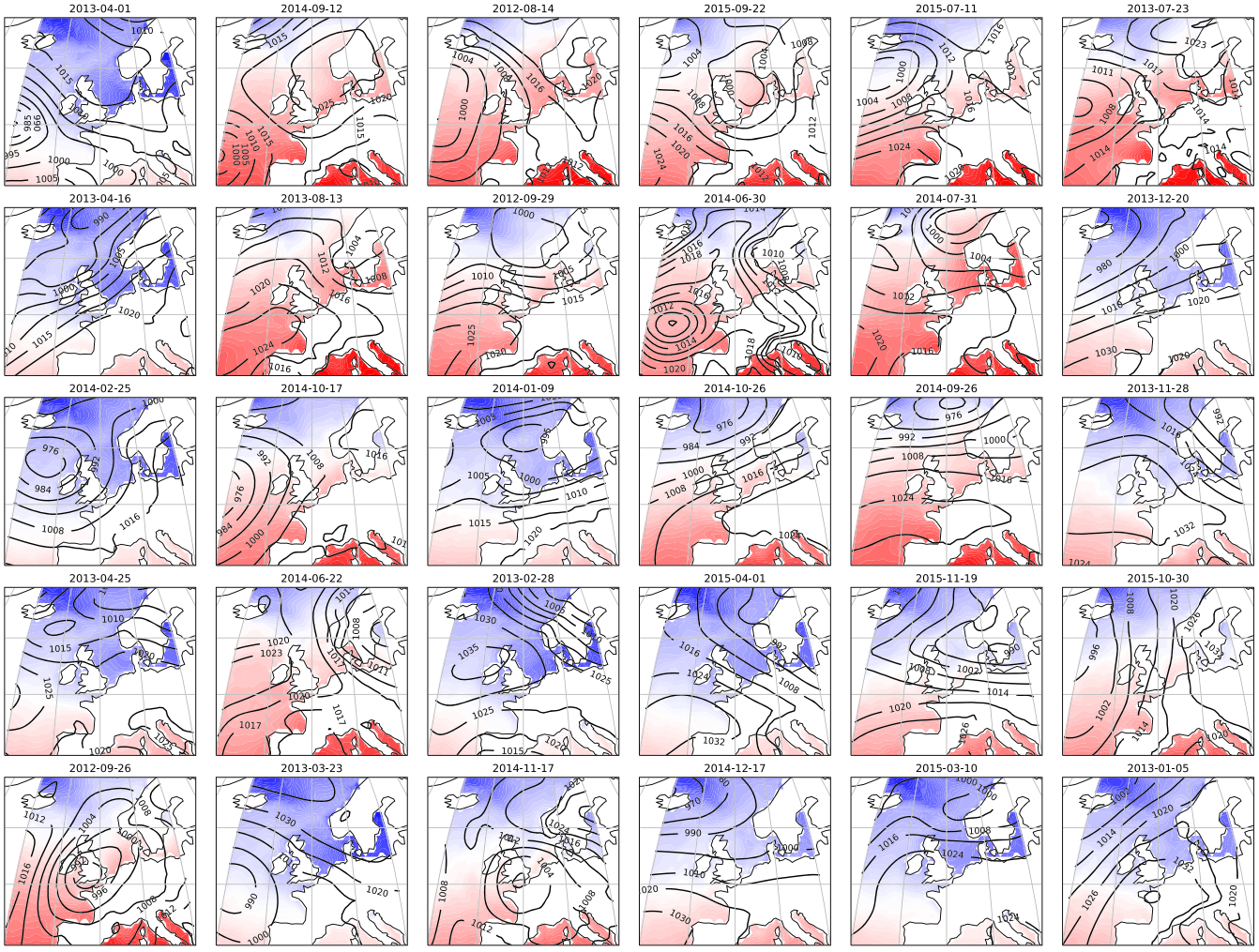

FIGURE B1 Synoptic charts of sea level pressure and sea-surface temperature for each case. Numbering 0 to 29 is from left to right and top to bottom [Colour figure can be viewed at wileyonlinelibrary.com]