

Assessing and Improving the Local Added Value of WRF for Wind Downscaling

M. GARCÍA-DÍEZ

Meteorology Group, Department of Applied Mathematics and Computer Science, University of Cantabria, Santander, and Climate Dynamics and Impacts Unit, Catalan Institute of Climate Sciences, Barcelona, Spain

J. FERNÁNDEZ

Meteorology Group, Department of Applied Mathematics and Computer Science, University of Cantabria, Santander, Spain

D. SAN-MARTÍN

Predictia Intelligent Data Solutions, Santander, Spain

S. HERRERA

Meteorology Group, Department of Applied Mathematics and Computer Science, University of Cantabria, and Predictia Intelligent Data Solutions, Santander, Spain

J. M. GUTIÉRREZ

Meteorology Group, Institute of Physics of Cantabria, Superior Council for Scientific Research–University of Cantabria, Santander, Spain

(Manuscript received 12 June 2014, in final form 26 November 2014)

ABSTRACT

Limited area models (LAMs) are widely used tools to downscale the wind speed forecasts issued by general circulation models. However, only a few studies have systematically analyzed the value added by the LAMs to the coarser-resolution-model wind. The goal of the present work is to investigate how added value depends on the resolution of the driving global model. With this aim, the Weather Research and Forecasting (WRF) Model was used to downscale three different global datasets (GFS, ERA-Interim, and NCEP–NCAR) to a 9-km-resolution grid for a 1-yr period. Model results were compared with a large set of surface observations, including land station and offshore buoy data. Substantial biases were found at this resolution over mountainous terrain, and a slight modification to the subgrid orographic drag parameterization was introduced to alleviate the problem. It was found that, at this resolution, WRF is able to produce significant added value with respect to the NCEP–NCAR reanalysis and ERA-Interim but only a small amount of added value with respect to GFS forecasts. Results suggest that, as model resolution increases, traditional skill scores tend to saturate. Thus, adding value to high-resolution global models becomes significantly more difficult.

1. Introduction

In recent years, the growing relevance of wind power has caused a surge of interest in improving the accuracy of existing surface wind forecasts (Costa et al. 2008). This variable is strongly affected by local topography and

other natural or human obstacles. In this context, dynamical and/or statistical downscaling are essential tools to improve the forecasts issued by coarse-resolution global models. The dynamical approach is usually based on the use of limited area models (LAM) operating on a region of interest and driven at the boundaries by the output of the global model. LAMs provide appropriate resolution and parameterization schemes to resolve mesoscale phenomena and orographic forcing.

There are relatively few studies in the literature analyzing the added value of dynamical downscaling as

Corresponding author address: M. García-Díez, Unitat de Dinàmica i Impacte Climàtic (UDIC) Institut Català de Ciències del Clima Doctor Trueta, 203 3a 08005 Barcelona, Catalonia, Spain.
E-mail: markel.garcia@ic3.cat

compared with the driving global model wind output. Over sea, Winterfeldt et al. (2011) and Feser et al. (2011) found added value in regions close to the coast, using satellite measures and the Regional Model (REMO) driven by a coarse-resolution reanalysis, the NCEP–NCAR, with grid cells of 1.875° . This was confirmed also with buoy data (Winterfeldt and Weisse 2009). Menéndez et al. (2014) carried out a longer-term study considering the WRF Model driven by both the NCEP–NCAR reanalysis and ERA-Interim. Using satellite observations as a reference, they found substantial added value with respect to the NCEP–NCAR reanalysis, as in previous studies. However, with respect to the higher-resolution (0.7°) ERA-Interim, added value was found to be confined to a few grid points close to the coastline. Over land, Jiménez et al. (2010) showed added value of WRF at very high resolution (2 km) over a small region with complex terrain. Again, the reference driving fields were relatively coarse reanalysis (1.125°) and analysis (1°) data from the European Centre for Medium-Range Weather Forecasts. It is clear that the added value of wind downscaling depends on the geographical complexity of the area under study. It is hard to add value over flat regions (e.g., over the sea) and easier over mountainous regions (Mass et al. 2002). Another recent study (Pryor et al. 2012) showed not much gain in skill by increasing the resolution of a regional model from 50 to 6.5 km over flat terrain in northern Europe. Therefore, we selected a relatively large region and evaluated the added value on complex and flat terrain and also over sea (Fig. 1).

Although there is a consensus about the capability of regional models to add value to the wind simulated by coarser global models, it is not clear how it depends on the resolution of the driving model, and whether it is possible to detect it with traditional skill scores such as root-mean-square error (RMSE). These scores are very sensitive to phase errors. As the higher-resolution models reproduce sharper changes in the fields, RMSE and other traditional scores can appear to be worse than in coarser models [see Fig. 16 of Mass et al. (2002) for a clear example]. Some attempts to solve this include object-based validation (Rife and Davis 2005), allowing phase shifts to find added value for 3.5-km forecasts in a complex terrain region. However, added value was found only in mountain stations, being very small for valley stations. Horvath et al. (2012) evaluated LAM simulations with different resolutions using spectral analysis and RMSE decomposition. They found that phase errors accounted for most of the RMSE, and that their effects increased with resolution. They also found that, in terms of variance, the resolution improved the results in the three bands that they defined (synoptic, diurnal, and subdiurnal). Herein we compare 6-h reanalysis data,

so such a detailed spectral analysis is out of the scope of this paper. Thus, the main goal is to check how the added value, detected using traditional skill scores, diminishes as one considers higher-resolution driving GCMs.

Moreover, although LAM outputs should be evaluated using areal-representative gridded observations (Osborn and Hulme 1997), to our knowledge, no gridded products based only on observations exist for daily or hourly wind. Because of the very local features of this variable, the spatial interpolation remains as a challenge. Therefore, in this paper we use local observations in Spain (from both stations and buoys) and focus in the assessment of the added value of LAM for wind speed prediction using the WRF Model and considering three widely used global datasets with different resolutions (~ 30 , 80, and 200 km). This will allow us to analyze the added value as a function of the resolution of the driving global model and estimate the potential improvement that could be obtained considering state-of-the-art global predictions (such as the GFS model used in this paper). The effect of the time aggregation (to 6-h and daily scale) in the assessment of the added value is also studied. Last, a modification to the subgrid orography drag parameterization, which significantly improves the model bias, is proposed.

2. Data and methods

The period of analysis in this work consists of one annual cycle (from March 2011 to February 2012), in which both subdaily wind observations and predictions were available.

a. Observations

Hourly observational data of instantaneous wind speed in a number of locations (land stations and buoys) were obtained for the analysis period from the Spanish Meteorological Agency (AEMET) and the National Port Authority (Puertos del Estado), respectively. Most of these stations are automatic, and, although they are regularly maintained, there are many sources of possible errors: calibration problems, missing data, inadequate station location, and so on. Only those locations with less than 25% of missing data in the target period were considered in this work. Apart from missing values, other quality checks were also performed to discard suspicious locations. These checks were based on the presence of too many outliers and/or a large frequency of zero wind speed values (above 15% of the total records). In total, after this quality check, data from 152 land stations and 9 buoys were used in this work (see Fig. 1, bottom). All the land stations measure the wind at 10-m height, but the buoys do it at 3 m. Thus, the buoy records were extrapolated to

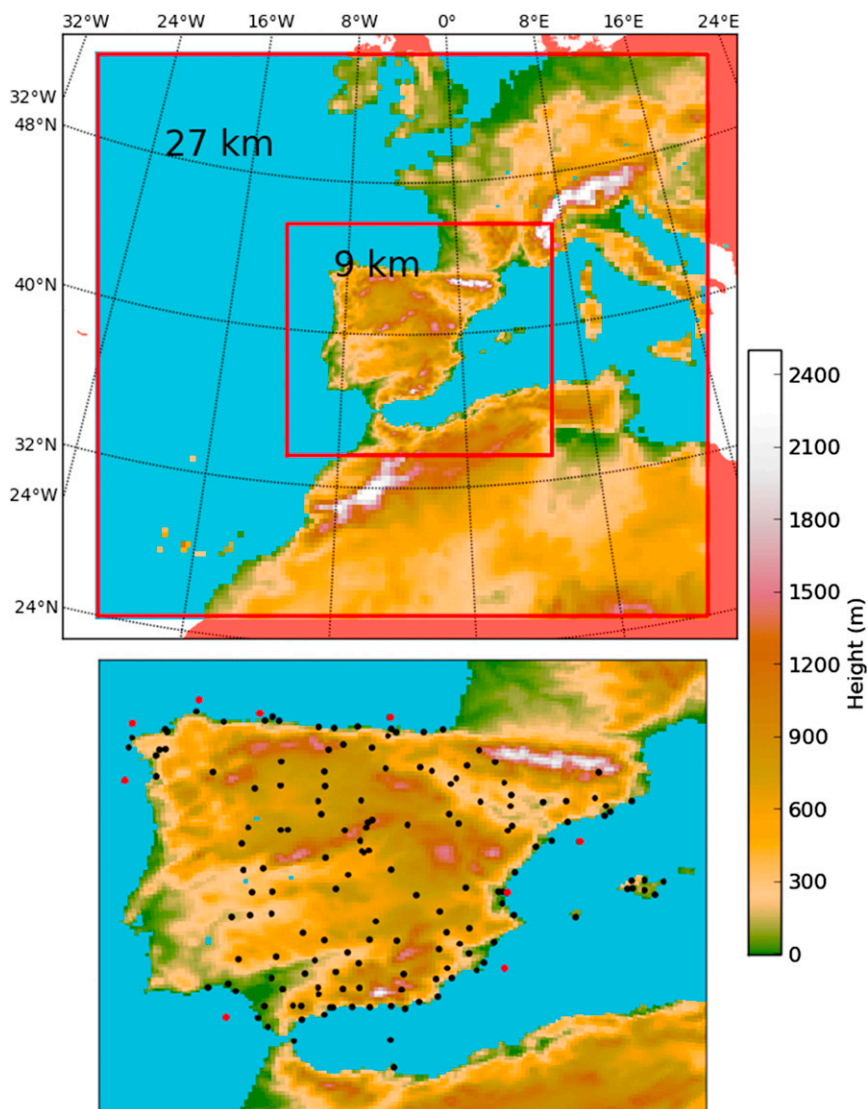


FIG. 1. (top) Two nested model domains at 27- and 9-km resolution and the corresponding model orography. (bottom) Location of the land stations (black dots) and buoys (red dots) in the inner domain.

10 m using the wind profile power law with an exponential coefficient of 0.11, following [Hsu et al. \(1994\)](#).

The overall skill of the models has been assessed using robust statistics (median, quantiles) over the whole set of individual statistics of each station, represented as box plots. This way we overcome the effect of dubious or possible miscalibrated individual stations. Stations with very low correlation or high bias were inspected for possible mistakes. However, unless they showed clear symptoms of being wrong, they have been retained.

b. WRF Model configuration

In the present work, we used the WRF Model ([Skamarock et al. 2008](#)). Namely, the Advanced Research

WRF (ARW), version 3.4 (V3.4), was used. The WRF Model is a community-developed LAM led by NCAR. In contrast with most other models, WRF is an open-source model that can be configured in many different ways. This includes choosing among a large set of parameterization schemes. In the present work, the standard parameterization schemes have been configured for the analysis region (the Iberian Peninsula) using the experience of previous works ([García-Díez et al. 2012](#); [Menéndez et al. 2014](#)): the Yonsei University scheme for the planetary boundary layer ([Hong et al. 2006](#)), WRF single-moment 5-class scheme for microphysics ([Hong et al. 2004](#)), Kain–Fritsch scheme for cumulus ([Kain 2004](#)), Rapid Radiative Transfer

Model for longwave radiation (Mlawer et al. 1997), Dudhia scheme for shortwave radiation (Dudhia 1989), and Noah land surface model for land soil processes (Chen and Dudhia 2001).

In addition to that, a new subgrid orography parameterization (SOP; Jiménez and Dudhia 2012, hereinafter JD2012) was introduced in WRF V3.4 to remove the large wind biases found in previous versions—surface wind speed was overestimated in flat and valley regions. This was attributed to a lack of representation of the effect of the unresolved orography in the surface drag. JD2012 also found an underestimation of the wind over hills and mountain ridges. To solve this, they introduced a parameterization of the surface drag that was originally represented by a sink term in the momentum equation, dependent on the friction velocity. JD2012 introduced a weighting factor C_t in this term that depends on the subgrid topography variance (σ_{sso}) and the Laplacian of the terrain height ($\Delta^2 h$). This factor is defined using thresholds, so C_t is larger over valleys and in areas with large σ_{sso} , and tends to zero when $\Delta^2 h < -20$ m, assuming no drag over hills and mountaintops. JD2012 found that this correction was successful in reducing the large biases found. Furthermore, they found significant added value when comparing their 2-km simulation with the driving model, which was ERA-40 (Uppala et al. 2005) at 1.125° resolution.

To assess the added value of this new parameterization, some experiments have been duplicated in this paper by switching it on and off. Note that JD2012 is not activated by default in WRF V3.4, so the latter option would correspond to the default WRF V3.4 configuration.

c. Global atmospheric models

Three global atmospheric models have been considered in this work as boundary conditions to drive WRF to assess the value added to them:

- The GFS (<http://www.emc.ncep.noaa.gov/GFS>) is a global model developed by NCEP (Yang et al. 2006). Currently, this model runs 4 times per day with T574 resolution (≈ 27 km), and it is freely available on the Internet in a 0.5° (≈ 50 km) regular grid. The frequency of the data is 3 h.
- ERA-Interim (Dee et al. 2011) is a third-generation reanalysis with T255 resolution ($\sim 0.7^\circ$) and a temporal frequency of 6 h. It is widely used for downscaling [e.g., in the Coordinated Regional Downscaling Experiment (CORDEX) framework; Giorgi et al. 2009].
- The NCEP–NCAR reanalysis (Kalnay et al. 1996) is a classical and somewhat outdated dataset with a coarse T62 resolution ($\approx 1.875^\circ$) and a temporal frequency of 6 h. This dataset is still widely used because it covers a

long period (more than 60 yr) and it is free and lightweight to download.

Note that the GCM data used as boundaries are of different nature. ERA-Interim and NCEP–NCAR are reanalyses, which assimilate observations, while the GFS data are $D + 1$ (where D is day) forecasts, which used observations only in the initial condition. Therefore, an additional degradation of skill is present in GFS. As we will show later, even with this degradation, GFS outperforms both low-resolution reanalyses when compared to observations.

d. Prediction experiments

The model was initialized each day at 1200 UTC and ran for 36 h, with the first 12 h being used as spinup. Previous studies found that this running scheme performs well against observations (Lenderink et al. 2009; Jiménez et al. 2010; Jiménez and Dudhia 2012; García-Díez et al. 2012) and provides improved day-to-day correspondence compared to continuous runs, even when nudged toward reanalysis data (Menéndez et al. 2014). This running scheme is computationally cheap, even though it sacrifices the small-scale realism of slow-varying variables, such as soil moisture. For atmospheric variables, Skamarock (2004) analyzed the kinetic energy spectra of WRF and found that 6–12 h were enough for the model to generate finescale variability. The model domain (Fig. 1) encompasses the whole Iberian Peninsula, an area with complex orography and different climate regimes. The horizontal resolution chosen is 9 km, with an intermediate domain of 27 km. This setup allows us to cover a large domain with a reasonable computational cost.

Five simulations have been produced following this scheme (see Table 1 for more details). WRF-G, WRF-E, and WRF-N denote the runs performed with the above WRF V3.4 configuration (activating JD2012) driven by GFS, ERA-Interim, and NCEP–NCAR fields, respectively. Moreover, to test the effect of the JD2012 SOP, a replica of WRF-G was produced but in this one we switched off the JD2012 parameterization (hereinafter WRF-G0). WRF-GM denotes an alternative replica obtained with a modified version of the JD2012 SOP proposed in the present paper (section 3b). The added value of WRF for the different resolutions of the driving fields can be analyzed by comparing WRF-G, WRF-E, and WRF-N, whereas the effect of the SOP parameterization can be analyzed by comparing WRF-G, WRF-G0, and WRF-GM.

e. Intercomparison issues

When dealing with different datasets, care must be taken so the comparison is fair. In this work we use as

TABLE 1. Characteristics of the model data used in this paper: topo_wind refers to the subgrid orography drag parameterization; Cy31r2 is the code version of the Integrated Forecast System, the GCM used to produce ERA-Interim. The 1994 means that NCEP–NCAR was carried out with the 1994 version of the NCEP–NCAR spectral model (Kalnay et al. 1996).

Label	Model version	Global model	Run resolution (km)	Data resolution	topo_wind
WRF-N	3.4	NCEP–NCAR	9	9 km	JD2012
WRF-E	3.4	ERA-Interim	9	9 km	JD2012
WRF-G	3.4	GFS	9	9 km	JD2012
WRF-G0	3.4	GFS	9	9 km	None
WRF-GM	3.4	GFS	9	9 km	Modified JD2012
GFS	9.0.1	—	T574 \approx 27	0.5°	—
ERA-Interim	IFS Cy31r2	—	T255 \approx 79	0.7°	—
NCEP–NCAR	1994	—	T62 \approx 210	1.875°	—

boundary conditions the $(D + 1)$ forecasts produced at 1200 UTC from the GFS model and the $(D + 0)$ analysis fields from the ERA-Interim and NCEP–NCAR reanalysis. Note that we are interested in the added value of WRF with respect to the corresponding global wind predictions/analysis. Therefore, we will mainly focus on measures of relative improvement to minimize the impact of the different nature of the driving global fields. However, when comparing with observations, the runs driven by GFS (as well as the corresponding direct model outputs) will have the disadvantage of being affected by the $D + 1$ global forecast error.

In the present study, the different frequencies of the data considered may also influence the final results. Station data from AEMET are available with a 10-min frequency, data from buoys and WRF are saved hourly, data from GFS are available every 3 h, and finally data from NCEP–NCAR and ERA-Interim are only available every 6 h. Thus, we present the results with three different time aggregations: instantaneous 6-h data (subsampling WRF, GFS, and observations), averaged 6-h data (not available in the case of NCEP–NCAR and ERA-Interim), and averaged daily data.

Part of the AEMET station data used were included in the surface synoptic observations (SYNOP) reports. Thus, the evaluation has been carried out using data that are partly assimilated in the GCMs. This issue would be relevant depending on the results. If WRF could not add value to ERA-Interim or NCEP–NCAR reanalyses, it could be argued that it is due to the assimilation of observations. However, as we will see, WRF is able to add value to these products and the only challenge is to add value to GFS. This product only used observations to produce the analysis used as initial condition. Therefore, our results hold even taking into account that part of the observations was assimilated in the reanalyses.

Finally, the comparison with point observations is also a relevant issue, since the datasets have different

horizontal resolutions. As mentioned in the introduction, models should be evaluated by comparing with gridded observations representing gridcell averages. However, that kind of interpolation would require a very dense wind station network. Another possibility is to interpolate the model data to the station point, weighting the nearest grid points and/or using a correction accounting for the representation error. However, our aim here is to study the model output as is it, including the representation error, which should be a source of added value for the higher-resolution simulations. Therefore, the models were not interpolated to the station data, but data from the nearest neighbor grid point were used instead. The land–sea mask of each model was used to filter out sea (land) grid points when comparing with inland (buoy) stations. This procedure significantly improved the results for coastal stations.

3. Results

a. Dependence on driving model resolution

We first analyze the local added value of WRF for wind forecasting in terms of the resolution of the driving global fields. To this end, we validate the experiments WRF-G, WRF-E, and WRF-N and compare the results with those corresponding to the driving global model outputs (GFS, ERA-Interim, and NCEP–NCAR, respectively). In particular, we consider the bias (model minus observed means), variability (model to observed variance ratio), and temporal (Spearman) correlation. The box plots in Fig. 2 summarize the results of these scores for the different locations shown in Fig. 1. The box edges represent the first and third quartiles (Q25 and Q75), the midline represents the median, and the whiskers reach the maximum and minimum values, provided they depart less than 1.5 times the interquartile range from the edge of the box. Beyond that limit, individual values are represented as crosses. The three box plots shown for each model refer, from left to right,

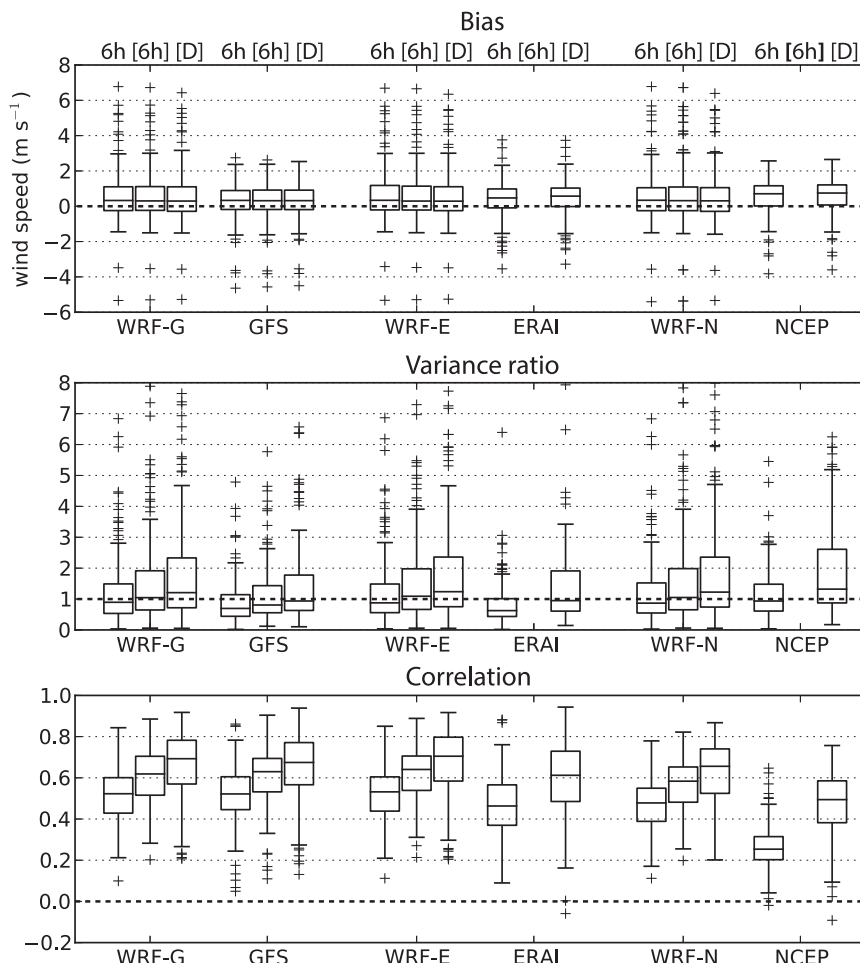


FIG. 2. Box plots representing the (top) bias (model – observed), (middle) variance ratio (model/observed), and (bottom) correlation of the local model predicted vs observed wind speeds during the period of study. The three box plots for each model on the x axis correspond to 6-hourly instantaneous data (6h), 6-hourly averaged data ([6h]), and daily averaged data ([D]), respectively. In the case of ERA-Interim and NCEP–NCAR, as only 6-hourly instantaneous data are available, the second box plot ([6h]) cannot be computed and is left blank. Box plots represent the statistics of the results for the 161 locations—152 land stations and 9 buoys. Plus signs denote outliers that deviate more than 1.5 times the interquartile distance from the closest quartile. Whiskers (vertical lines) extend to the minimum/maximum value that is not an outlier.

to 6-h instantaneous data (6h), 6-h averaged data ([6h]) and daily averaged data ([D]).

Unlike the other scores, bias is not influenced by the temporal aggregation of the data in any of the experiments. Global models exhibit a positive bias pattern indicated by the shift of the interquartile box over the zero axis (i.e., the bias is positive in approximately 75% of the locations). This shift is larger for NCEP–NCAR and ERA-Interim than for GFS. The WRF experiments also exhibit a positive bias pattern, which is independent of the driving global model. WRF-E and WRF-N slightly improve the median bias of their driving models (ERA-Interim and NCEP–NCAR, respectively), whereas

WRF-G is very similar to GFS. Two stations appear as lower outliers with pronounced negative biases in the WRF experiments: Cabo Villano and Estaca de Bares. Both are windy stations located on capes and surrounded by cliffs, obstacles that are not resolved by the model. At the upper side of the distribution, WRF largely overestimates wind in about 8–10 stations, which appear as upper outliers in the three WRF experiments. This problem is related to the JD2012 SOP parameterization and will be analyzed in detail later.

Regarding the variance ratio, in general the median is close to one for all simulations, but exhibits an increasing trend—together with the variability—as data

are temporally aggregated. This yields a systematic overestimation of the variance for the daily averaged data, which is more evident in the WRF experiments (with the exception of the NCEP–NCAR case). There are also a number of outliers, which correspond to the same outlier locations reported for the bias.

Finally, correlation increases when aggregating the data; for instance, correlation is about 0.1–0.2 larger for the daily data than for the instantaneous ones. This is consistent with the smaller predictability of the intradaily variability. The stations with low correlation values (below 0.2) have been individually checked, and no suspicious behavior has been found. These results seem related to the poor representativeness of the model topography for those stations. As expected, NCEP–NCAR exhibits the lowest correlations among all datasets because of its coarse resolution, and a great improvement is achieved using WRF in this case. This is in agreement with [Menéndez et al. \(2014\)](#). On the other hand, the correlations for ERA-Interim and GFS are very similar and the added value of WRF is still appreciable in the former case, but minor in the latter.

To better analyze the added value of the WRF experiments—with respect to the corresponding global outputs—[Fig. 3](#) shows the box plots of Brier skill score (BSS), computed following the definition given by [Winterfeldt et al. \(2011\)](#):

$$\text{BSS} = \begin{cases} 1 - \sigma_F^2 \sigma_R^{-2} & \text{for } \sigma_F^2 \leq \sigma_R^2 \\ \sigma_R^2 \sigma_F^{-2} - 1 & \text{for } \sigma_F^2 > \sigma_R^2, \end{cases} \quad (1)$$

where σ_F^2 and σ_R^2 are the error variances of the model forecast (e.g., WRF-G) and the reference forecast (e.g., GFS), respectively. Error variance is defined as $\sigma^2 = (1/N) \sum_{i=1}^N (x_i - f_i)^2$, where x_i are forecast data and f_i are observed data. BSS takes values in $[-1, 1]$, where positive values indicate added value and negative values imply that the forecast model performs worse than the reference. Note that this score mixes the errors in representing the mean (bias), variability, and phase (correlation), used in the previous analysis ([Murphy and Epstein 1989](#)). Therefore, to assess the added value in the temporal correlation, we also considered the correlation differences (CD), obtained as the correlation of WRF minus the correlation of the driving global model.

In agreement with [Winterfeldt et al. \(2011\)](#), WRF improves the performance of the NCEP–NCAR reanalysis for both BSS and CD in the vast majority of the stations. In the case of WRF-E, added value is substantially smaller but still appreciable for CD, since the difference is positive in over 75% of the stations. In case

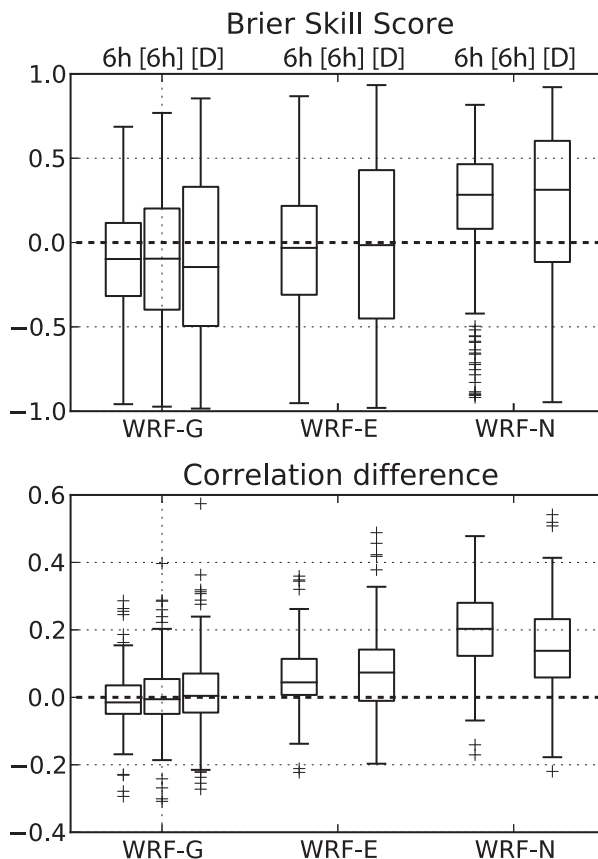


FIG. 3. Box plots of added value for the WRF-G, WRF-E, and WRF-N models with respect to GFS, ERA-Interim, and NCEP–NCAR models, respectively, using (top) BSS and (bottom) correlation difference. The plotting convention for 6 h, 6-hourly averaged data [6 h], and daily averaged data [D] is as in Fig. 2.

of WRF-G no appreciable added value is found, and both models exhibit a similar performance in terms of correlation, with WRF even worse in terms of BSS.

b. Modification of the subgrid orography drag parameterization

To understand the reasons for the large biases found in some stations for the WRF simulations (represented as outliers in [Fig. 2](#)), we analyze the performance of the [JD2012](#) parameterization in our particular case study. This parameterization was shown to successfully correct the problems related to wind bias reported in the literature [see [Jiménez and Dudhia \(2012\)](#) for more details]. However, an individual analysis of the outliers found in our study reveals that the large biases correspond to stations close to mountain ranges, suggesting that orographic representativeness problems remain.

Simply stated, [JD2012](#) introduced a multiplicative factor in the wind drag as a function of the Laplacian of the topography $\Delta^2 h$ as follows:

$$C_t = \begin{cases} 1 & \text{if } \Delta^2 h > -20 \text{ and } \sigma_{\text{SSO}} < e \\ \ln \sigma_{\text{SSO}} & \text{if } \Delta^2 h > -10 \text{ and } \sigma_{\text{SSO}} > e \\ \alpha \ln \sigma_{\text{SSO}} + (1 - \alpha) & \text{if } \Delta^2 h \in [-20, -10) \sigma_{\text{SSO}} > e \\ \delta_{\text{mod}} + \frac{\Delta^2 h + 30}{10} (1 - \delta_{\text{mod}}) & \text{if } \Delta^2 h \in [-30, -20) \\ \delta_{\text{mod}} & \text{if } \Delta^2 h < -30 \end{cases}, \quad (2)$$

where σ_{SSO} is the variance of the subgrid orography, computed from a $\simeq 100$ -m-resolution dataset, and $\alpha = (\Delta^2 h + 20)/10$. The original definition of JD2012 corresponds to $\delta_{\text{mod}} = 0$, which effectively leads to a linear decrease of the drag toward zero when the Laplacian is smaller than -20 . This limit was considered to identify grid cells on mountaintops by JD2012, who set up the parameterization with 2-km resolution. However, when using this parameterization at a lower resolution (e.g., 9-km resolution in this paper), the smoothing effect of the topography interpolation causes serious representativeness errors. For such low resolution, the Laplacian criterion selects an area that extends well beyond the real mountaintops (not shown). This leads to extended suppression of wind drag and yields an unrealistic wind speed overestimation. Therefore, in the present paper we propose a modification of JD2012 parameterization by not removing the drag in those grid boxes, but keeping it unaltered with respect to the default WRF configuration. Note that this is achieved by using $\delta_{\text{mod}} = 1$ in Eq. (2).

To analyze the effect of the above (original and modified) parameterizations, Fig. 4 shows the results of three experiments running different versions of WRF driven by the GFS model output: WRF-G, WRF-G0, and WRF-GM. This allows comparing the same experiment, but conducted with the JD2012 parameterization [Eq. (2) with $\delta_{\text{mod}} = 0$], with no SOP parameterization (WRF-G0), and with the modified SOP parameterization proposed in this paper [Eq. (2) with $\delta_{\text{mod}} = 1$, WRF-GM], respectively.

As found by JD2012 and others (e.g., Carvalho et al. 2014), the default WRF configuration (WRF-G0, no subgrid orography drag) suffers from a pronounced overestimation of the surface wind speed, leading to systematic positive biases and large variances. However, the bias outliers found in the WRF-G experiment do not appear in WRF-G0 (Fig. 4, top), indicating that this problem is caused by the JD2012 parameterization. This was the motivation for the modified parameterization run in experiment WRF-GM, where the factor weighting the surface drag is kept constant ($C_t = 1$) for the grid

points with Laplacian below -20 m (note that this is equivalent to using WRF with no SOP in these points). As can be seen in Fig. 4, this modification not only removes the outliers, but it also clearly improves the spatial bias distribution, which is now centered around zero.

Even if biases improved with the proposed modification, variability is underestimated for intradaily data and correlation remains mostly unaffected. Wind speed variability in WRF-GM improves as we move from instantaneous to daily mean data. This is consistent with the WRF underestimation of the daily wind cycle, reported in other studies (Dudhia 2013). WRF-G, however, shows better variance for the time scales resolving the daily cycle (6 h, [6 h]) and overestimates the variance when it is averaged out.

Regarding the added value of the modified parameterization (Fig. 5), it shows a small improvement for both BSS and CD only for the daily mean values. Given the resolution difference between WRF (9 km) and the GFS data used ($0.5^\circ \simeq 50$ km), more (local) added value was expected. These results highlight the need of complex metrics/scores involving spatial patterns in the analysis of added value at this range of resolutions.

c. Spatial distribution of the added value

Figure 6 shows, for each station, the added value (or lack of, in red) of the different WRF experiments except WRF-G0. As shown before, the added value of the WRF downscaling (white dots) is only evident and widespread when nested into low-resolution boundary data (WRF-E and WRF-N), especially regarding the correlation difference. The spatial pattern of the added value does not exhibit a clear connection with the orography or the degree of continentality. For the WRF simulations nested into GFS, the improvement of the modification proposed (WRF-GM) with respect to the original (WRF-G) is also evident. However, the value added to the driving data does not show any specific pattern and in roughly half of the stations WRF worsens the results according to these added value scores.

The stations affected by the large bias associated with the wind acceleration on mountaintops can be spotted in

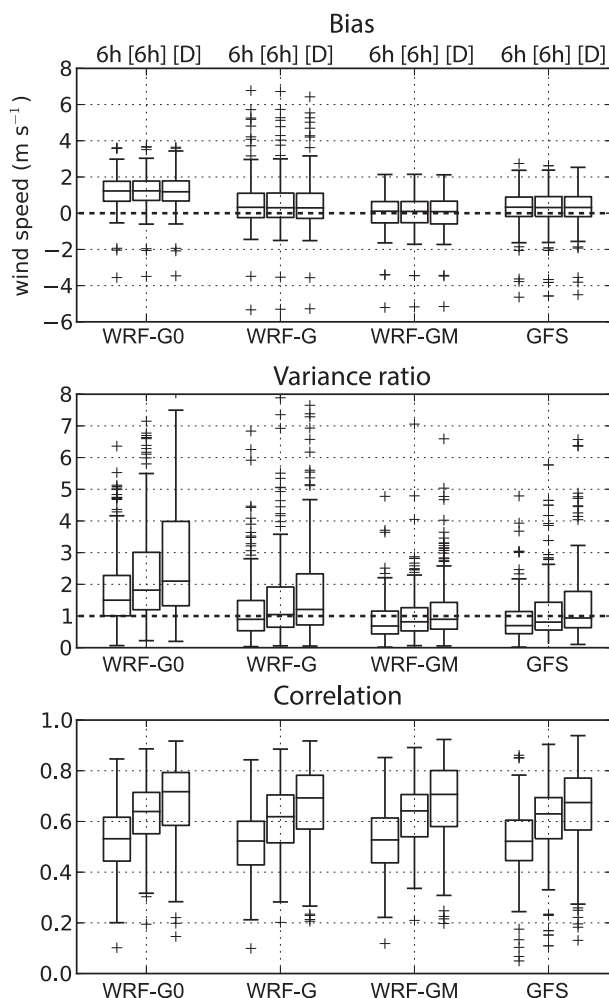


FIG. 4. As in Fig. 2, but for WRF-G0, WRF-G, WRF-GM. The box plots for the GFS data used as boundaries and WRF-G are the same as in Fig. 2. They are reproduced here to ease the comparison with the modified version of the SOP.

the WRF-N BSS maps, as they are the only ones where the BSS is negative (red dots). These stations are limiting with mountain ranges (Fig. 1) and show the larger BSS improvement when applying our modification (WRF-GM) to the JD2012 SOP (WRF-G). Therefore, even though the modification proposed was not tested on NCEP–NCAR- and ERA-Interim-driven simulations, the few negative BSS spots on WRF-N and WRF-E are likely to improve with the proposed modification.

As mentioned in the introduction, Spain has a complex topography leading to rich mesoscale features such as locally channeled winds, sea and mountain breezes, and coastal areas. Thus, the natural candidates to find added value in WRF experiments would be for specific regions related to these features. However, no clear spatial patterns were found (Fig. 6). Moreover, there are many cases of neighboring stations with opposite added

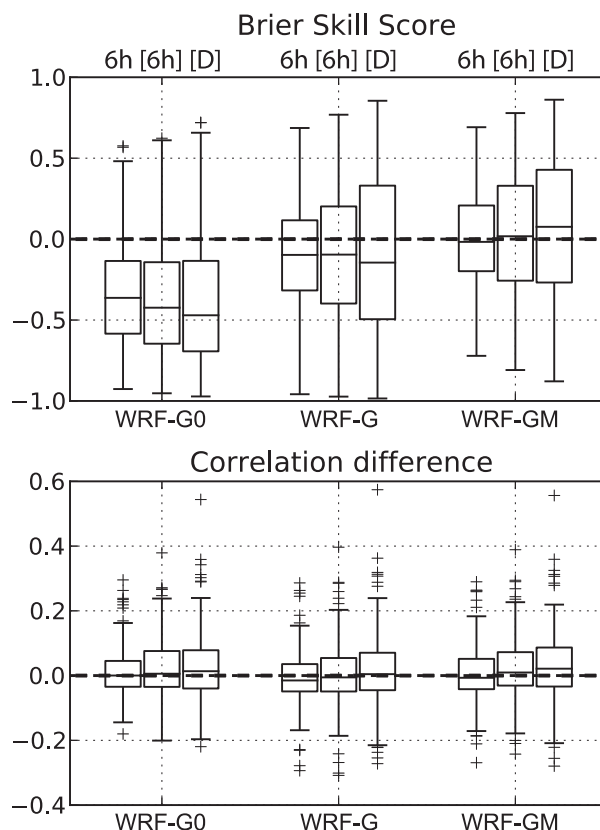


FIG. 5. As in Fig. 3, but for the added value of WRF-G0, WRF-G, and WRF-GM with respect to GFS.

value scores (close red and white spots in Fig. 6). This suggests that the added value depends on the representation of very local obstacles. Furthermore, results show that it is not possible to generally assess the added value with a small amount of stations. For example, choosing a subset of 4–5 stations (e.g., central northern coastal stations), we could conclude that WRF-G is adding substantial value. When considering the whole dataset, this is not the case. In the case of offshore buoy data, substantial added value is found in WRF-N. However, no added value is found for WRF-E in the buoys on the Atlantic Ocean, where ERA-Interim already has very large values of correlation, above 0.9 (see below). That is, the added value also depends on the reference skill: where the correlation between the global model and observations is small, WRF is more prone to add value. Contrarily, it is hard to add value to a station where the correlation of the driving data is already high. To check this, CD of each station was compared with the correlation between the global model output and the observations in that station (Fig. 7). The dashed diagonal line marks the maximum CD achievable, taking into account that correlation cannot exceed 1.

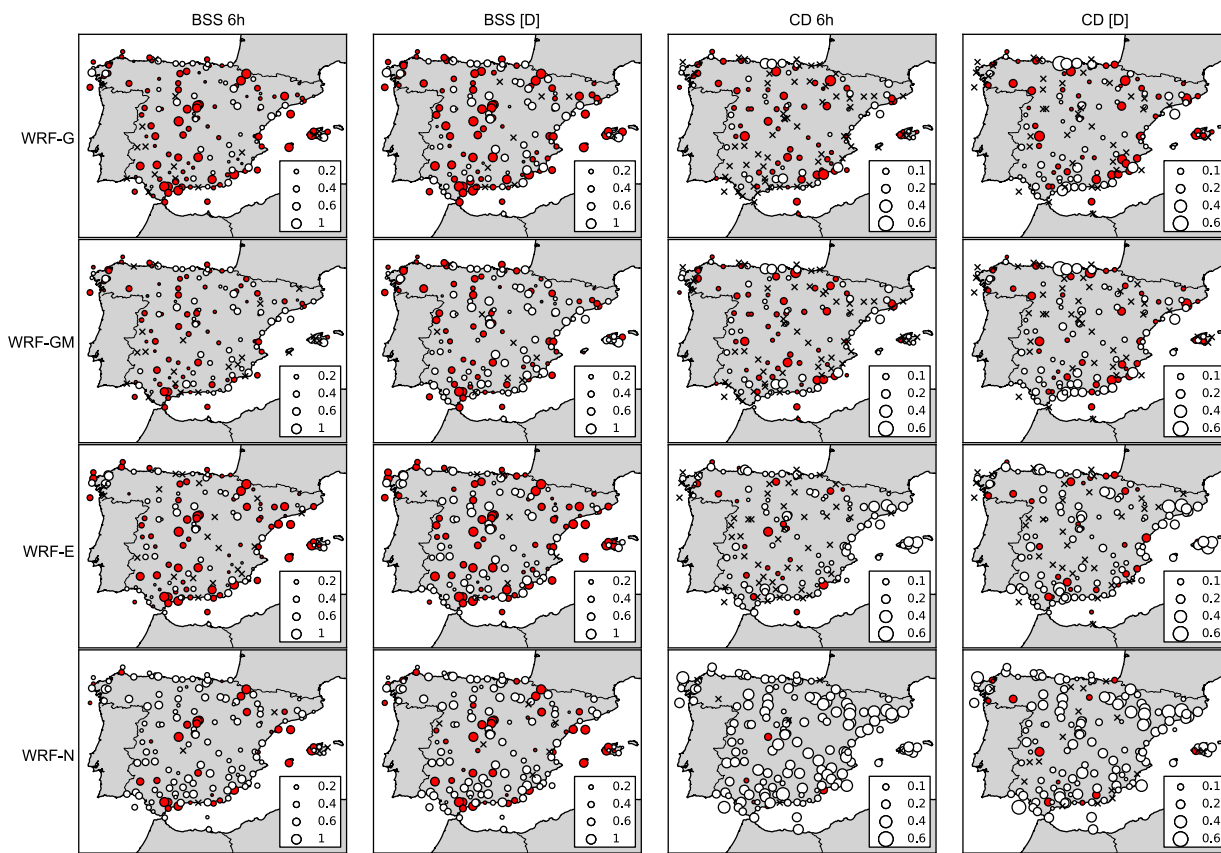


FIG. 6. Added value maps for the WRF experiments (except WRF-G0) using the (first and second columns) BSS and (third and fourth columns) correlation difference. The added value is shown at 6-h instantaneous and daily mean time scales. Each point represents one station or buoy. Red (white points) represent negative (positive) values, and times signs identify locations with small absolute values (<0.05). The size of the point represents its value as given in the legend.

WRF-G (Fig. 7a) adds value to GFS only in about half of the stations. As expected, the larger added values (0.4–0.5) correspond to stations showing low correlation with GFS (below 0.4). There are, however, some stations

with low correlation with GFS and no gain from WRF downscaling. One of these stations is located in a small deep valley, not resolved by either WRF or GFS. Buoy data are very well correlated with GFS (~ 0.9) and,

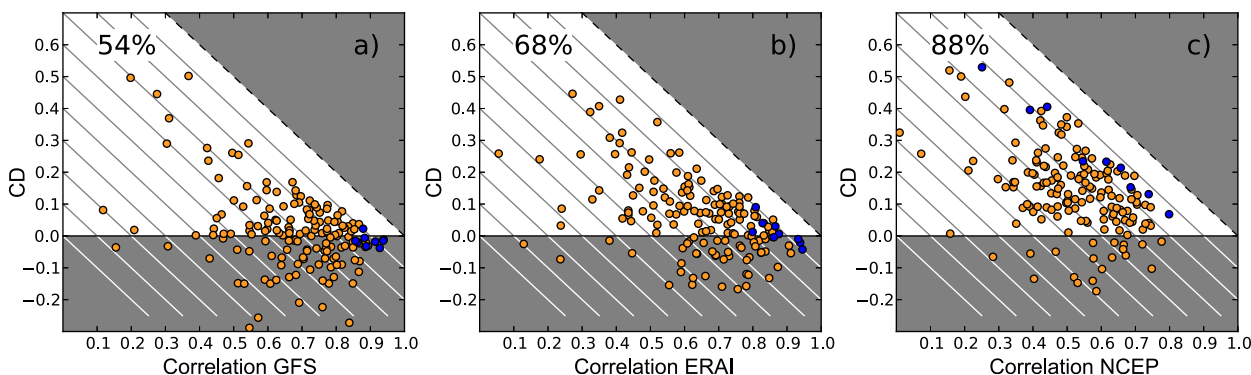


FIG. 7. Scatterplots of the correlation difference between WRF and the GCM against the correlation of the GCM for (a) GFS and WRF-G, (b) ERA-Interim and WRF-E, and (c) NCEP–NCAR and WRF-N. Diagonal lines are WRF correlation isolines, readable on the left scale. WRF adds value to the GCM in the stations lying in the white area. The percentage of stations where WRF adds value is shown on the top-left corner of each panel. Blue (orange) dots represent buoy (land) stations.

therefore, WRF adds no value there. WRF-N (Fig. 7c) adds value to NCEP–NCAR data at most points. Buoy data show mixed CD added value and correlation values with NCEP–NCAR. However, they are systematically well correlated with WRF-N (values around and above 0.8). For ERA-Interim and WRF-E (Fig. 7b), there is an intermediate situation between GFS and NCEP–NCAR.

In general, these results indicate that, with this resolution, WRF has the potential to improve the correlation to a maximum value of about 0.9. Additional analyses were carried out to compare CD added value to topography, topography standard deviation, variance, and variance in the high-frequency bands (not shown), but no significant patterns were found. Thus, CD seems linked to local differences between WRF and driving data orography that are challenging to identify systematically.

4. Conclusions

A set of 9-km-resolution simulations spanning a 1-yr period have been carried out with WRF to assess the dependence of added value on driving model resolution. Three popular global datasets were used as boundary conditions: GFS forecasts and ERA-Interim and NCEP–NCAR reanalyses. The study focused on wind speed at different time scales (instantaneous, 6-h, and daily averaged), which were compared to station data using two added value scores: BSS and CD.

Clear and large added value was found in the WRF downscaling of the coarse NCEP–NCAR reanalysis, as found in previous works (Feser et al. 2011). Smaller added value was found in the downscaling of ERA-Interim. However, very small or negligible added value was found in the wind speed downscaled from GFS.

The subgrid orography drag parameterization implemented by JD2012 to correct the WRF surface wind bias successfully removes the problem from most of the stations. However, the suppression of wind drag over grid points with strong negative Laplacian causes a large overestimation of the wind in a subset of the stations. While appropriate for high-resolution simulations (as shown by JD2012), we suggest using the modified JD2012 with $\delta_{\text{mod}} = 1$ for resolutions similar to this study (~ 10 km). As expected, the bias reduction greatly improved the BSS, but only slightly improved the correlation.

Given that the Iberian Peninsula is a region with a complex orography, the results found for GFS were unexpected. WRF, just with the improvement in the representation of the orography, should add value in most of the stations. However, according to scores, no added value was found with respect to the GFS wind speed.

This might be caused by a suboptimal configuration of the experiment or by more fundamental predictability issues. In this work, daily simulations were carried out to preserve day-to-day correspondence, but no data assimilation was used for model initialization. Data assimilation could improve the initial condition, making it balanced and removing or reducing the need of a spinup period (Pielke 2002). Also, the daily simulations were run in parallel, for computational efficiency, which prevented starting from warm soil. This affects the downscaling, since soil moisture is restarted from coarse data every day, and loses part of the potential regional detail that WRF can develop (Case et al. 2008). These improvements could probably increase WRF skill, but a dramatic improvement is unlikely, and the configuration used is widely used in the literature (Winterfeldt et al. 2011; Hu et al. 2010; Lenderink et al. 2009; Jiménez and Dudhia 2012). Also, GFS is very well tuned to its resolution. However, the flexibility of WRF, designed to run in a very wide range of resolutions, could be a shortcoming when compared to carefully tuned models.

The lack of added value with respect to GFS could also be the result of more fundamental predictability limitations. Short-lived features of the flow simulated by WRF might be realistic but not correlated with the observations. Some studies (Rife et al. 2004; Rife and Davis 2005), even at finer resolution, show that object-based evaluation is required to unveil added value. We showed that, in general, the model skill increases with the temporal aggregation. The value added to NCEP–NCAR decreased when considering averaged daily data, meaning that subdaily scales are an important contributor to the added value in this case. In the ERA-Interim case, the added value slightly increased for daily averaged data. Thus, intradaily variability is much better resolved by ERA-Interim than by NCEP–NCAR.

There are statistical tools, such as Kalman filters, that can correct model systematic errors (Cassola and Burlando 2012). As these tools are not able to improve correlation, it should be the main focus when looking for added value. We argue that, if correlation is lost at finer scales, for applications sensitive to phase shifts (e.g., short-term wind energy forecasts), increasing resolution to a few kilometers or even subkilometer levels might not be worth the extra computational power. The value of dynamical downscaling at those scales would then be confined to the study of the physical structures developing, and not to the increase of forecast accuracy itself. More research is needed to identify the sources of added value in wind speed correlation.

These results illustrate how resolution yields diminishing improvements to the deterministic forecast skill. In this scenario, the added value of high-resolution

dynamical downscaling cannot be taken for granted, at least for deterministic measures. The great value added by LAMs to very coarse resolution models (Feser et al. 2011) does not hold for the higher resolutions reached by modern global models. The potential of LAMs to improve global models by getting to even higher spatial and temporal resolutions is limited by fundamental predictability problems of the smaller scales. This problem is not only limited to short-term forecasting but has also emerged in high-resolution regional climate models, where recent increases in resolution showed no clear added value as measured by standard evaluation scores (Vautard et al. 2013; Kotlarski et al. 2014).

Acknowledgments. The authors thank Puertos del Estado (Spanish National Ports and Harbour Authority) and AEMET (Spanish Meteorological Agency) for providing buoy and land observational records. This work was partly supported by the projects EXTREMBLES (CGL2010-21869) and CORWES (GL2010-22158-C02-01), funded by the Spanish R&D program. The WRF simulations performed in this study were managed by WRF4G, which is an open-source tool funded by the Spanish government and cofunded by the European Regional Development Fund under Grant CGL2011-28864.

REFERENCES

- Carvalho, D., A. Rocha, M. Gómez-Gesteira, and C. Silva Santos, 2014: WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for Portugal. *Appl. Energy*, **117**, 116–126, doi:10.1016/j.apenergy.2013.12.001.
- Case, J., W. Crosson, S. Kumar, W. Lapenta, and C. Peters-Lidard, 2008: Impacts of high-resolution land surface initialization on regional sensible weather forecasts from the WRF model. *J. Hydrometeorol.*, **9**, 1249–1266, doi:10.1175/2008JHM990.1.
- Cassola, F., and M. Burlando, 2012: Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. *Appl. Energy*, **99**, 154–166, doi:10.1016/j.apenergy.2012.03.054.
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.
- Costa, A., A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Freitas, 2008: A review on the young history of the wind power short-term prediction. *Renewable Sustainable Energy Rev.*, **12**, 1725–1744, doi:10.1016/j.rser.2007.01.015.
- Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.
- , 2013: Evaluation of WRF model surface and low-level winds in complex terrain and coastal regions. *NCAR–NCAS WRF Users Workshop*, Davis, CA, NCAR, 25 pp. [Available online at https://www.ncas.ac.uk/index.php/en/documents/doc_download/760-wrf-jimy-dudhia.]
- Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, 2011: Regional climate models add value to global model data. *Bull. Amer. Meteor. Soc.*, **92**, 1181–1192, doi:10.1175/2011BAMS3061.1.
- García-Díez, M., J. Fernández, L. Fita, and C. Yagüe, 2012: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quart. J. Roy. Meteor. Soc.*, **139**, 501–514, doi:10.1002/qj.1976.
- Giorgi, F., C. Jones, and G. Asrar, 2009: Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bull.*, **58**, 175–183. [Available online at http://www.wmo.int/pages/publications/bulletinarchive/archive/58_3_en/documents/58_3_giorgi_en.pdf.]
- Hong, S., J. Dudhia, and S. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.
- , Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.
- Horvath, K., D. Koracin, R. Vellore, J. Jiang, and R. Belu, 2012: Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models. *J. Geophys. Res.*, **117**, D11111, doi:10.1029/2012JD017432.
- Hsu, S. A., E. A. Meindl, and D. B. Gilhouse, 1994: Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. *J. Appl. Meteor.*, **33**, 757–765, doi:10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2.
- Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844, doi:10.1175/2010JAMC2432.1.
- Jiménez, P., and J. Dudhia, 2012: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *J. Appl. Meteor. Climatol.*, **51**, 300–316, doi:10.1175/JAMC-D-11-084.1.
- , J. González-Rouco, E. García-Bustamante, J. Navarro, J. Montávez, J. de Arellano, J. Dudhia, and A. Muñoz Roldan, 2010: Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation. *J. Appl. Meteor. Climatol.*, **49**, 268–287, doi:10.1175/2009JAMC2175.1.
- Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev. Discuss.*, **7**, 217–293, doi:10.5194/gmdd-7-217-2014.
- Lenderink, G., E. van Meijgaard, and F. Selten, 2009: Intense coastal rainfall in the Netherlands in response to high sea surface temperatures: Analysis of the event of August 2006 from the perspective of a changing climate. *Climate Dyn.*, **32**, 19–33, doi:10.1007/s00382-008-0366-x.
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful

- forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, doi:[10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2).
- Menéndez, M., M. García-Díez, L. Fita, J. Fernández, F. J. Méndez, and J. M. Gutiérrez, 2014: High-resolution sea wind hindcasts over the Mediterranean area. *Climate Dyn.*, **42** (7–8), 1857–1872, doi:[10.1007/s00382-013-1912-8](https://doi.org/10.1007/s00382-013-1912-8).
- Mlawer, E., S. Taubman, P. Brown, M. Iacono, and S. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated- k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:[10.1029/97JD00237](https://doi.org/10.1029/97JD00237).
- Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–582, doi:[10.1175/1520-0493\(1989\)117<0572:SSACCI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2).
- Osborn, T. J., and M. Hulme, 1997: Development of a relationship between station and grid-box rainfall frequencies for climate model evaluation. *J. Climate*, **10**, 1885–1908, doi:[10.1175/1520-0442\(1997\)010<1885:DOARBS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2).
- Pielke, R. A., 2002: *Mesoscale Meteorological Modeling*. Academic Press, 676 pp.
- Pryor, S. C., G. Nikulin, and C. Jones, 2012: Influence of spatial resolution on regional climate model derived wind climates. *J. Geophys. Res.*, **117**, D03117, doi:[10.1029/2011JD016822](https://doi.org/10.1029/2011JD016822).
- Rife, D. L., and C. A. Davis, 2005: Verification of temporal variations in mesoscale numerical wind forecasts. *Mon. Wea. Rev.*, **133**, 3368–3381, doi:[10.1175/MWR3052.1](https://doi.org/10.1175/MWR3052.1).
- , —, Y. Liu, and T. T. Warner, 2004: Predictability of low-level winds by mesoscale meteorological models. *Mon. Wea. Rev.*, **132**, 2553–2569, doi:[10.1175/MWR2801.1](https://doi.org/10.1175/MWR2801.1).
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, doi:[10.1175/MWR2830.1](https://doi.org/10.1175/MWR2830.1).
- , and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3_bw.pdf.]
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:[10.1256/qj.04.176](https://doi.org/10.1256/qj.04.176).
- Vautard, R., and Coauthors, 2013: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dyn.*, **41**, 2555–2575, doi:[10.1007/s00382-013-1714-z](https://doi.org/10.1007/s00382-013-1714-z).
- Winterfeldt, J., and R. Weisse, 2009: Assessment of value added for surface marine wind speed obtained from two regional climate models. *Mon. Wea. Rev.*, **137**, 2955–2965, doi:[10.1175/2009MWR2704.1](https://doi.org/10.1175/2009MWR2704.1).
- , B. Geyer, and R. Weisse, 2011: Using QuikSCAT in the added value assessment of dynamically downscaled wind speed. *Int. J. Climatol.*, **31**, 1028–1039, doi:[10.1002/joc.2105](https://doi.org/10.1002/joc.2105).
- Yang, F., H.-L. Pan, S. K. Krueger, S. Moorthi, and S. J. Lord, 2006: Evaluation of the NCEP Global Forecast System at the ARM SGP site. *Mon. Wea. Rev.*, **134**, 3668–3690, doi:[10.1175/MWR3264.1](https://doi.org/10.1175/MWR3264.1).