Model Wind Field Forecast Verification Using Multiple-Doppler Syntheses from a National Radar Network

Jeffrey Beck, Mathieu Nuret, Olivier Bousquet

Model Wind Field Forecast Verification Using Multiple-Doppler Syntheses from a National Radar Network

JEFFREY BECK AND MATHIEU NURET

CNRM-GAME, Météo-France and CNRS, Toulouse, France

OLIVIER BOUSQUET

CNRM-GAME, LACy - UMR 8105, Météo-France, CNRS, Université de La Réunion Saint-Denis, La Réunion, France

(Manuscript received 25 June 2013, in final form 8 January 2014)

ABSTRACT

Model verification has traditionally relied upon in situ observations, which typically exist on a sparse network, making nonsurface model forecast verification difficult. Given increasing model resolution, supplemental observational datasets are needed. Multiple-Doppler wind retrievals using a national network of radars provide an opportunity to assess the accuracy of wind forecasts at multiple levels, as well as verification within a three-dimensional domain. Wind speed and direction verification results are presented for a 9-day period of forecasts from the French Application of Research to Operations at Mesoscale-Western Mediterranean (AROME-WMED) model using multiple-Doppler retrievals from the French Application Radar à la Météorologie Infrasynoptique (ARAMIS) network. For the analyzed period, relationships were found that suggest that errors are not only linked to forecasted evolution of meteorological phenomena, but are sensitive to terrain height below the analyzed level as well as mesoscale processes. The spatial distribution of errors at initialization and forecast times shows that biases are generally independent of location and terrain height at initialization, but that the impact of terrain below the analysis level affects the forecasted wind magnitude and direction over time. These comparisons illustrate that multiple-Doppler wind retrieval measurements accurately identify model error and can serve as an invaluable dataset for model verification.

1. Introduction

Operational numerical weather prediction (NWP) continues to benefit from enhanced computing power, enabling higher-resolution simulations to be produced on increasingly shorter time scales. Nonhydrostatic operational NWP is now possible with horizontal resolutions approaching the boundary between the need for parameterization and the ability to explicitly resolve convective phenomena. For example, the National Severe Storms Laboratory (NSSL) runs operational Weather Research and Forecasting Model (WRF) simulations covering the continental United States with a 4-km horizontal resolution (Skamarock et al. 2008). In Europe, the Met Office runs a version of the Unified Model called the UKV at 1.5-km resolution over the United Kingdom and Ireland (Lean et al. 2011), while Météo-France currently operates the Applications of Research to Operations at Mesoscale (AROME) model at 2.5-km resolution over all of France, including parts of surrounding countries (Seity et al. 2011).

Increased resolution and the ability to explicitly resolve convective phenomena enable highly detailed model output and have the potential to verify and potentially improve forecast accuracy. However, comparison of these forecasts against available observations becomes a significant challenge. While the use of spaceborne instrument data will likely become more prevalent in the future, model verification has generally been limited to the use of surface-based observations and the sparse network of global upper-air observations. Higher-resolution simulations accentuate the need for dense, uniformly spaced datasets for verification, in order to assess the accuracy of smaller-scale features previously unresolved by lower-resolution model forecasts. In addition, model verification above the network of surface-based observing stations is extremely limited. Therefore, the availability of high-resolution 3D observational data above the ground represents a key necessity in the verification of current and future numerical weather prediction.

Corresponding author address: Jeffrey R. Beck, Météo-France/Centre National de Recherches Météorologiques, 42 Ave. G. Coriolis, 31057 Toulouse, CEDEX 1, France.
E-mail: jeff.beck@meteo.fr

DOI: 10.1175/WAF-D-13-00068.1

Past methods of model verification have required interpolation of model fields to nonuniformly spaced observation locations in order to facilitate comparison. When assessing near-surface temperature and wind field accuracy, smoothing of differences in model surface height and actual topography is often required (e.g., Accadia et al. 2007). In addition, model values are sometimes not observed at the standard 10-m ground-based observation height. In these cases, further interpolation is necessary, with some studies incorporating a power-law reduction to arrive at 10-m values (Mass et al. 2002). Interpolation in both the horizontal and vertical, in addition to any use of assumptions regarding boundary layer structure to retrieve observation-level model variables (e.g., Geleyn 1988), may introduce error and inhibits direct comparison and verification of model forecasts. Other verification techniques have been developed to improve and evaluate NWP forecasts more objectively than solely at grid scale. These methods include object-oriented techniques (e.g., Ebert and McBride 2000; Davis et al. 2009); neighborhood verification, including the ability of the model to identify precipitation displacement, orientation, and spatial extent (e.g., Ebert 2009; Gilleland et al. 2010); and scale decomposition techniques. The latter approach consists of explicit determination of scale resolution for a given NWP model, as well as evaluation of forecast error scale dependence. Such analysis is often performed with observed precipitation fields (e.g., Harris et al. 2001; Vasic et al. 2007; Bousquet et al. 2006; Roberts and Lean 2008), but can also be conducted with wind field data (Bousquet et al. 2008a). To conduct these types of verification experiments, it is therefore important to obtain a sufficiently broad network of observations, even more so than for traditional point-based verification.

Options for uniform observation fields over a large domain are extremely limited. Remotely sensed fields offer one possibility, such as near-surface wind data calculated by the Advanced Scatterometer (ASCAT), on board the *Meteorological Operational-A* satellite (Gelsthorpe et al. 2000). ASCAT provides data over sea surfaces with a relatively high wind speed accuracy ($\pm 2\,\mathrm{m\,s^{-1}}$), but with $\pm 20°$ wind direction accuracy. While ASCAT provides a uniform field of wind observations and offers an opportunity for expanded low-level model verification, the data have a coarse temporal resolution due to the polar orbit of the satellite and a nominal spatial resolution of 50 km, well below that of mesoscale models.

As an alternative, ground-based Doppler radars provide the desired broad spatial coverage with the advantage of much higher resolution than that of spaceborne instrumentation. Doppler radars can provide reflectivity and polarimetric variables for precipitation forecast verification, as well as radial velocity data, which can be incorporated into wind field forecast verification (e.g., Salonen et al. 2011). Given a sufficiently dense network of radars, multiple-Doppler syntheses can be conducted to retrieve a three-dimensional view of the airflow. Bousquet et al. (2008a) used multiple-Doppler synthesis data near Paris, France, to assess the spatial orientation and accuracy of the low-level horizontal wind field for two case studies (a frontal passage and squall line) modeled within a prototype version of AROME. Results from this study show that multiple-Doppler wind fields successfully allow for high-resolution (~1–2 km) model accuracy assessment and can serve as a viable complementary dataset for use in forecast verification techniques.

Radars used by Bousquet et al. (2008a) are part of the French Application Radar à la Météorologie Infrasynoptique (ARAMIS) radar network, which consists of 24 radars nationwide, and have recently been included in a real-time, nationwide 3D wind retrieval system (Bousquet and Tabary 2013). The availability of these data provides an excellent opportunity for use in verification of high-resolution model output over multiple forecast periods. While more advanced verification techniques such as object-oriented and scale-based analysis will be incorporated in the future, point verification is used in the present study to illustrate the utility and value of the high-resolution ARAMIS radar network.

The verification period chosen for analysis spans nine days, from 1 to 9 November 2011, with multiple-Doppler syntheses calculated every 15 min for all of southern France. These data were then compared to AROME-Western Mediterranean (AROME-WMED) model forecasts issued every day at 0000 UTC for 48 h. The analysis period included the unique evolution of an upper-level low into a warm-core, tropical-like cyclone that affected the southeastern coast of France. The study was conducted within the framework of the Hydrological cycle in the Mediterranean Experiment (HyMeX), which focused on the analysis of terrain-induced high-precipitation events in the Mediterranean basin (Ducrocq et al. 2014; HyMeX 2013). Significant amounts of precipitation were generated over regions of complex terrain, making the chosen case study ideal for field campaign preparation. Results are shown for model bias, mean absolute error (MAE), and normalized MAE for domain-averaged wind speed and direction for 2 km above mean sea level (MSL) as a function of forecast lead time as well as spatial distribution at initialization and all forecast times. Error statistics were also produced based on multiple-Doppler-defined wind speed categories and regional distribution (as a function of terrain height below or above 1 km MSL). Conclusions are drawn regarding error values, and potential future research and operational applications are then discussed.

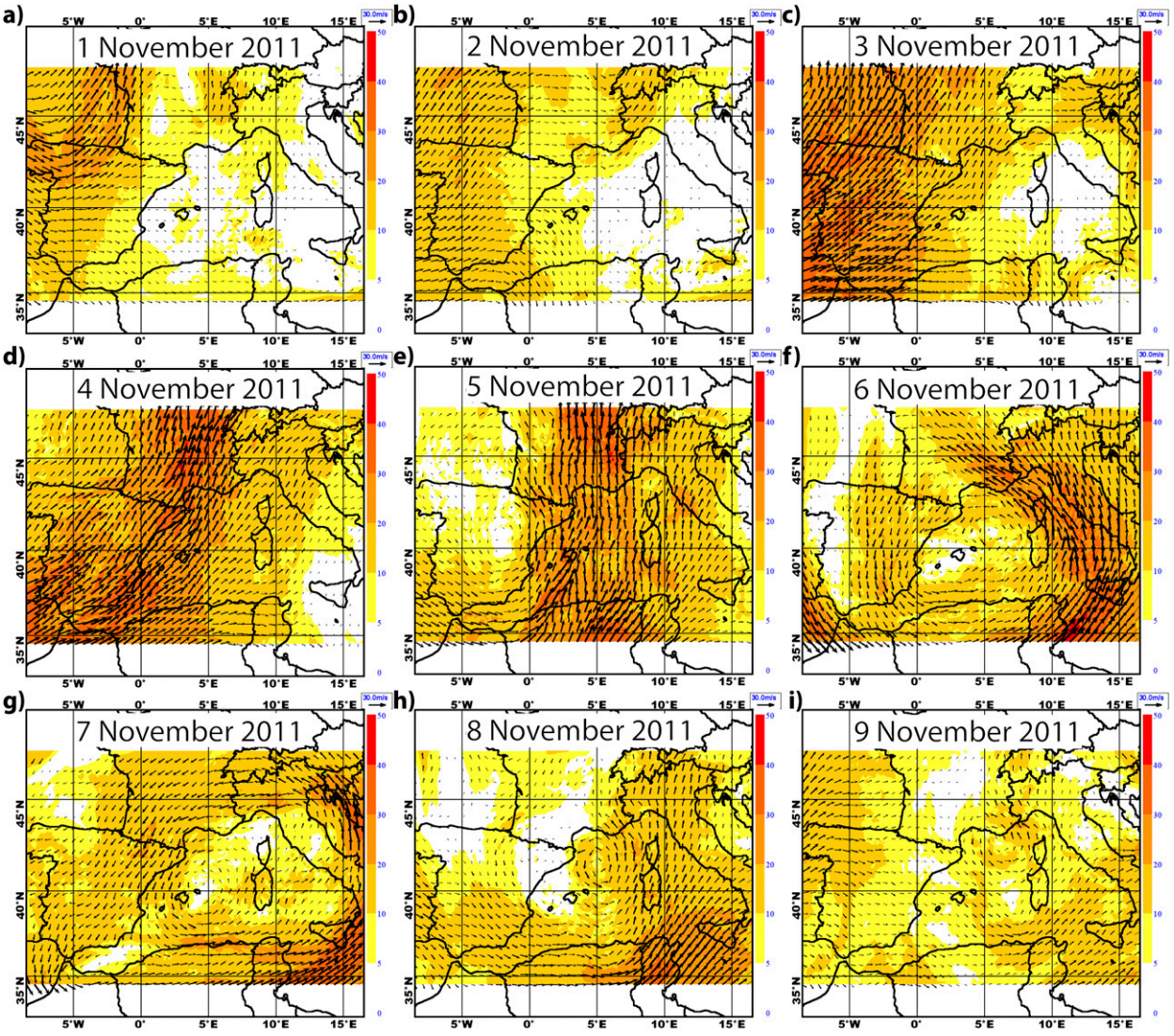

FIG. 1. AROME-WMED 0000 UTC analysis wind magnitude (color) and horizontal vectors (black, every sixth vector shown) at 500 hPa for (a)–(i) 1–9 Nov 2011.

2. Methodology

a. Case study description

During the period of 1–9 November 2011, two strong upper-level troughs traversed southern France (Fig. 1). Between 1 and 3 November 2011, nearly 500 mm of rain fell over southern portions of the Cevennes region of France (Fig. 2a), as southerly flow produced continued upslope, orographically induced convection along the southeastern escarpment of the Massif Central elevated terrain (near Alès, France). The second trough generated broad regions of precipitation with accumulations totaling more than 250 mm between 3 and 6 November 2011 near the French Pyrenees, the southeastern escarpment of the Massif Central, and the southern Alps (Fig. 2b). As the second upper-level trough developed into a closed upper-level low, a Mediterranean tropical-like cyclone, or medicane (e.g., Moscatello et al. 2008a,b), was produced that affected the southeast coast of France between 6 and 9 November 2011 with precipitation amounts for the period reaching 200 mm southeast of Marseille, France (Fig. 2c).

b. Multiple-Doppler syntheses

An extended period of precipitation over a broad portion of southern France provided an excellent opportunity to test the use of multiple-Doppler syntheses for model wind assessment over numerous forecast runs. These syntheses are possible on a real-time basis using the dense ARAMIS national French radar network and recently

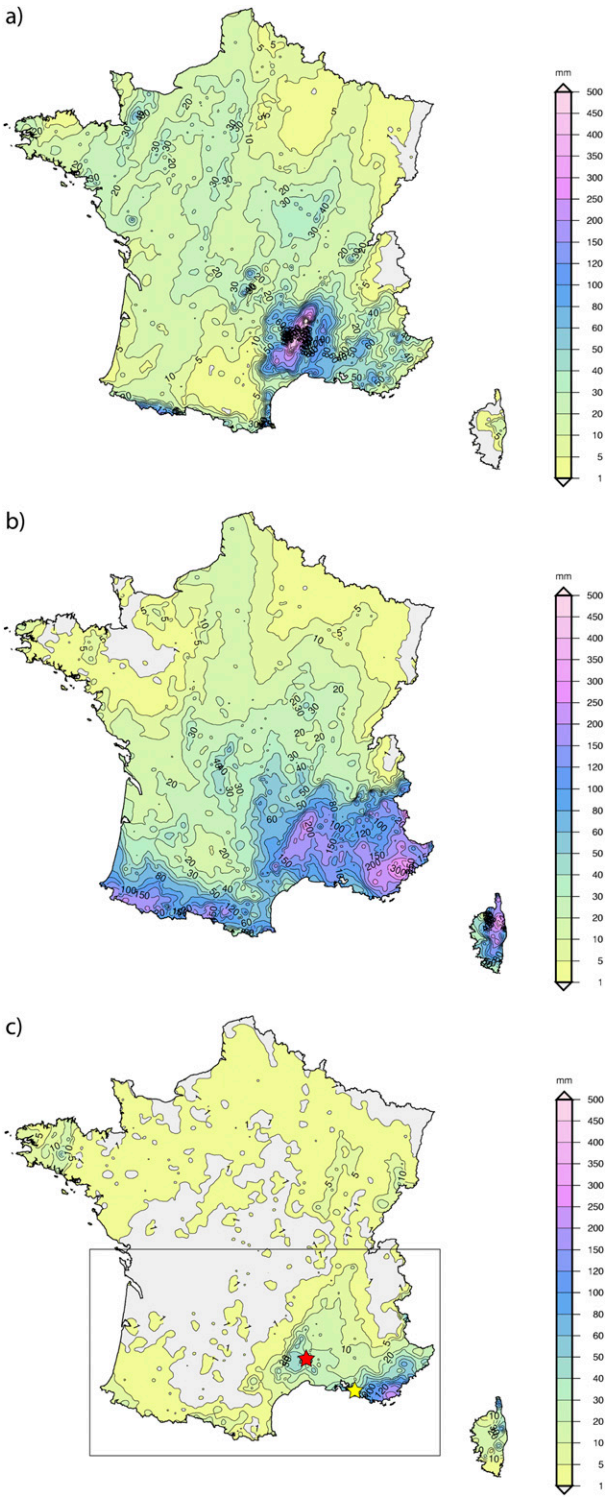

FIG. 2. Precipitation totals (mm) for (a) 1–3, (b) 3–6, and (c) 6–9 Nov 2011. The red and yellow stars in (c) represent the locations of Alès and Marseille, France, respectively, and the black box outlines the verification domain.

installed Risques Hydrométéorologiques en Territoires de Montagnes et Mediterranéens (RHYTMME) radars (Westrelin et al. 2012; Beck and Bousquet 2013). Together, these radars provide coverage for over 90% of the country (Beck and Bousquet 2013, Fig. 1). Volumetric radar data are available every 15 min, with a triple pulse repetition time (PRT) scheme (Tabary et al. 2006) allowing for a maximum unambiguous Nyquist velocity of 60 m s^{-1}. A fuzzy logic algorithm (Gourley et al. 2007) utilizes dual-polarization variables to classify and filter nonprecipitating echoes, providing clean radial velocity fields for automated multiple-Doppler syntheses.

Wind retrievals were conducted every 15 min over the 9-day period from 1 to 9 November 2011 using the 11 southernmost ARAMIS radars. Locations of the radars in addition to dual- and multiple-Doppler coverage for the ARAMIS network can be seen in Fig. 2 from Bousquet and Tabary (2013). Radial velocity data were first converted into a Cartesian grid using a Cressman interpolation scheme (Cressman 1959), with horizontal and vertical resolutions of 2.5 and 0.5 km, respectively. The created grid covers a region of 875 × 500 × 12 km^3, centered on southern France. Figure 3 shows the multiple Doppler domain at 0000 UTC 6 November 2011, as the upper-level trough undergoes a transition into a cutoff low. Multiple-Doppler syntheses were created using the Multi-Scale Chemistry Aerosol Transport (MUSCAT) software suite (Bousquet and Chong 1998). The reader is referred to Bousquet et al. (2008b) for more information about data processing and operational wind field retrieval within the ARAMIS network.

For the purposes of this study, two-dimensional wind vectors were chosen at a height of 2 km MSL to compare with output from model data. This height, in particular, was chosen in order to provide sufficient radar coverage above the ground, yet close enough to the ground to investigate any potential topographic impacts. In addition, 2 km MSL wind syntheses are constructed from relatively low radar elevation angles, minimizing the vertical velocity contribution that must be removed from the retrieved radial velocity and thus improving accuracy. Assessment of model forecast and multiple-Doppler synthesized winds was limited to regions where precipitation was occurring during the period of study, since ARAMIS radar winds generally cannot be retrieved in clear air. Multiple-Doppler wind vectors, while generally accurate, can contain some error due to their remotely sensed nature. Since wind retrievals represent the observational truth in this study, it is worthwhile to describe the limitations inherent in wind syntheses. Bousquet et al. (2008b) conducted an error assessment of ARAMIS-synthesized horizontal winds,

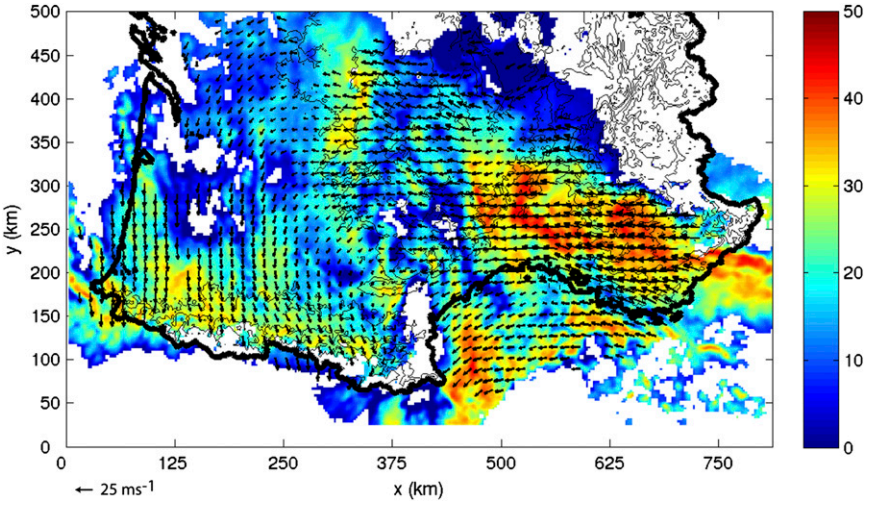

FIG. 3. Multiple-Doppler synthesis at 0000 UTC 6 Nov 2011 for 2 km MSL, showing the domain used for model comparison. Reflectivity (dBZ) is in color and horizontal wind vectors ($m\,s^{-1}$) are shown in black, with every fifth vector displayed. Black contours indicate topography.

indicating that errors decrease with increasing altitude (with bias of less than $\pm 1\,m\,s^{-1}$ and RMSE of less than $2\,m\,s^{-1}$ above 2 km MSL). Importantly, synthesized wind fields were found to be consistent with previous high-resolution radar observations, showing that ARAMIS radar data produce reliable wind retrievals. Therefore, errors of less than $\pm 1\,m\,s^{-1}$ at 2 km AGL are considered to be negligible, and are likely nullified in this study, given that the data are averaged over significant temporal–spatial scales and numerous forecast runs.

c. *AROME-WMED*

Given the southerly, onshore flow and subsequent orographic precipitation generated during the 1–9 November 2011 event in southern France, the AROME-WMED model was an ideal candidate for comparison with the multiple-Doppler syntheses and for HyMeX campaign preparation. Similar to AROME, but designed specifically to support HyMeX, AROME-WMED covers a geographical domain over the entire western Mediterranean basin, as well as Portugal, Spain, Italy, and northern Africa (Fig. 4a). In anticipation of HyMeX, the AROME-WMED initialization utilizes a background error covariance matrix that catered more toward convective phenomena than that of AROME [the reader is referred to Brousseau et al. (2011) for more information about the operational AROME model]. In addition, the AROME-WMED initialization relies on more satellite observations than does AROME given that much of the domain is located over water.

Using a centered 3-h assimilation cycle, AROME-WMED incorporates values closest in time to the analysis, including radial velocity when precipitation is present (Montmerle and Faccani 2009). However, the full wind vector is not retrieved from radar data for initialization, as only radial velocity from the nearest radar is assimilated for each grid point. For more information on initialization techniques, the reader is referred to Seity et al. (2011). Simulations are conducted every 6 h with a forecast range of 48 h for this study, compared to the 30-h AROME. Therefore, following this forecast cycle, data were saved every 3 h for simulations beginning at 0000 UTC 1 November until 0000 UTC 9 November 2011. In total, 33 model runs were considered, with 417 individual wind field forecasts used for comparison with the multiple-Doppler syntheses during the 9-day period.

Wind synthesis data from the MUSCAT software suite are created using a Lambert II extended projection and thus were interpolated onto the AROME-WMED model grid, with a resolution of 0.025° latitude × 0.025° longitude. The resulting dataset allowed for comparison between model wind fields and MUSCAT syntheses at ~2.5-km resolution (Fig. 4b). Since the AROME-WMED model does not provide height-level data, 800-hPa level winds were used as a proxy for 2 km MSL. Unfortunately, model levels above and below 800 hPa are too sparsely separated to apply vertical interpolation techniques, as this would in turn introduce more error. However, geopotential height deviation of the 800-hPa surface from the radar data level of 2 km is minor during the event studied (approximately ±200 m). In addition, small errors that may result from a deviation of the 800-hPa surface from 2 km MSL are minimized, since the analysis data are created from domain-wide or multiple forecast averages.

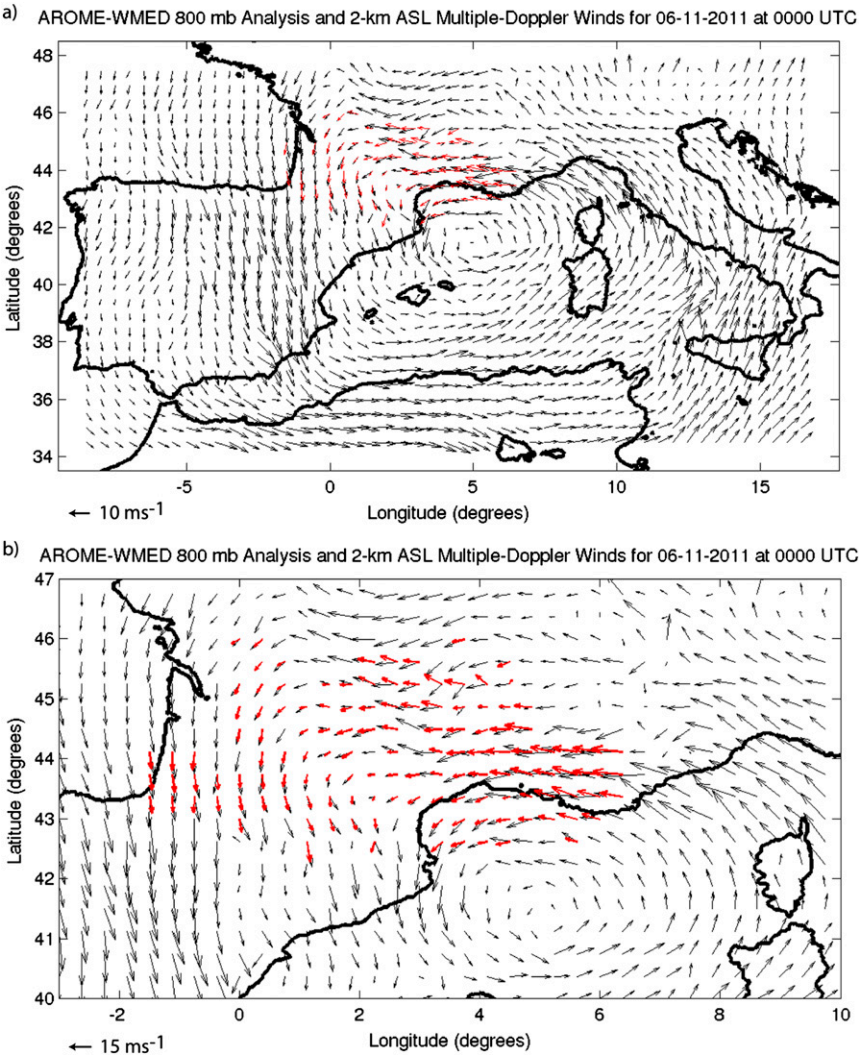

FIG. 4. (a) AROME-WMED model domain (sampled at 0.05° latitude × 0.05° longitude) for initialization at 0000 UTC 6 Nov 2011. Wind vectors are shown in black for 800 hPa. Multiple-Doppler synthesis wind vectors are overlaid in red for 2 km MSL. Every 10th vector is shown. (b) As in (a), but for the full AROME-WMED resolution, zoomed in on the analysis region (every 15th vector shown). Multiple-Doppler synthesis data were only available where red vectors are plotted.

3. Analysis

The distribution of precipitation during the comparison period varied as a function of the upper-level trough and medicane locations as they affected southern France. Figure 5 shows the number of multiple-Doppler points retrieved between 0000 UTC 3 November and 0000 UTC 9 November 2011. Beginning on 3 November 2011, the second upper-level trough begins to affect the multiple-Doppler domain, and the number of points increases to a maximum of over 30 000 after 0000 UTC 6 November 2011. Accounting for terrain blockage and areas inaccessible to the ARAMIS radars, 30 000 points represent precipitation coverage over more than 60% of the multiple-Doppler domain. As the upper-level trough becomes cut off and weakens, the number of points falls until the medicane makes landfall, with values increasing to near 10 000 points on 8 November 2011.

To assess the quantitative differences between model forecasts and multiple-Doppler syntheses, error statistics for both wind speed and direction were first calculated using an average over the full analysis domain for each forecast at each forecast hour. Bias and MAE were calculated (Fig. 6) to enable analysis of how AROME-WMED handled the progression of the main upper-level trough during the 9-day period. An error convention of

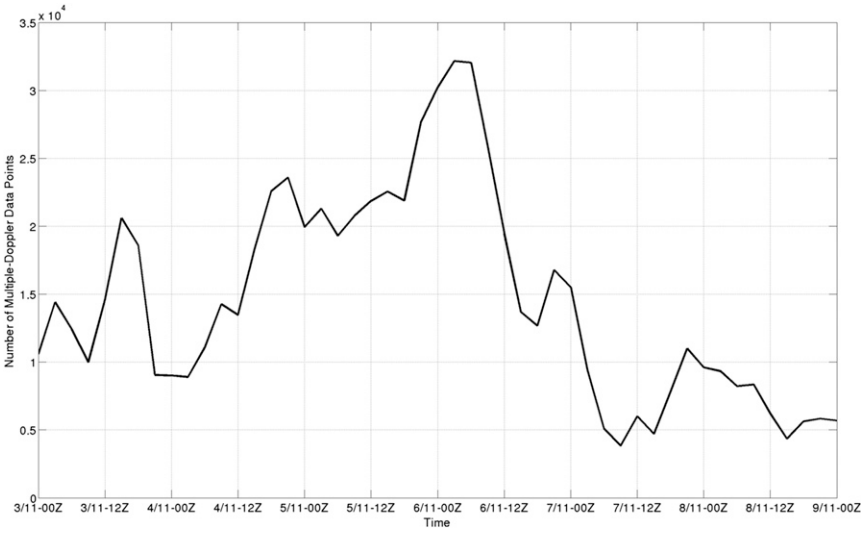

FIG. 5. Number of multiple-Doppler synthesis data points as a function of time throughout the analysis period between 3 and 9 Nov 2011.

subtracting the multiple-Doppler data from the model fields was adopted and is used for all statistics. For wind direction, any differences greater than ±180° were converted by subtracting ±360° and changing the sign of the difference. Normalized bias was calculated for the domain-wide, average wind speed at each forecast hour to assess the percentage error as a function of time. Further research focused on the distribution of normalized MAE as a function of location (regions with an altitude above or below 1 km MSL), as well as forecast wind speed categories with increments of 5 m s^{-1}.

With respect to bias in the wind speed forecasts, the strength of the main upper-level trough and medicane is underestimated as they approach southern France, resulting in a negative bias for all forecast hours (Fig. 6a). Up to 10% error can be seen in the domain-wide, averaged normalized wind speed bias between forecast hours 9 and 30 (Fig. 7). This trend is reversed at later forecast times (>36 h), when model wind speed forecasts become more accurate. The domain-wide, averaged normalized wind speed biases at these forecast hours are less than 5%.

The mean absolute error (Fig. 6b) and normalized MAE (Fig. 7) for wind speed and direction as a function of forecast hour show a steady increase until about the 30-h forecast. After this point, the domain-wide average MAE in both wind speed and direction remain fairly constant at about 4.2 m s^{-1} and 28°, respectively, while the domain-wide averaged normalized MAE for wind speed stays just above 30%. This plateau in MAE coincides with the same forecast times at which bias errors improve (Fig. 6a). It is important to note that even at initialization, the average MAE rates are 3.4 m s^{-1} and 21° and the domain-wide average normalized wind speed MAE is 25%. These initial errors are likely associated with a lack of observational data and/or a compromise (minimization) between the previous model run used as a first guess and assimilated observations within the AROME-WMED three-dimensional variational data assimilation (3DVAR) initialization system. In addition, while radial velocities are used for assimilation, they are screened for differences of more than 20 m s^{-1} between the model and observations, then thinned onto a 15 km × 15 km grid in order to avoid cross-correlation errors during initialization. Screening and thinning are detrimental to high-resolution radial velocity measurements and likely reduce the potential corrections imparted to the wind field during initialization.

In addition to bias and MAE, a spatially dense and extensive observational dataset (such as the ARAMIS multiple-Doppler wind syntheses) provides the opportunity to analyze error characteristics in closer detail. For example, data were analyzed as a function of topography to assess the impact of elevated terrain on the domain-wide average MAE (Fig. 8). Results initially indicate that for both wind speed (Fig. 8a) and wind direction (Fig. 8b), the height of the terrain has an impact on the amount of error seen between the AROME-WMED forecasts and the multiple-Doppler syntheses at 2 km MSL. For wind speed MAE in regions where the topography is less than 1 km MSL, early forecast times show nearly 1 m s^{-1} lower MAE than for regions where terrain is above 1 km MSL. This discrepancy in wind speed MAE as a function of terrain height is reduced with increasing forecast times and, eventually, becomes nearly identical for all topographical heights at 48 h. For wind speed MAE in regions with terrain above 1 km

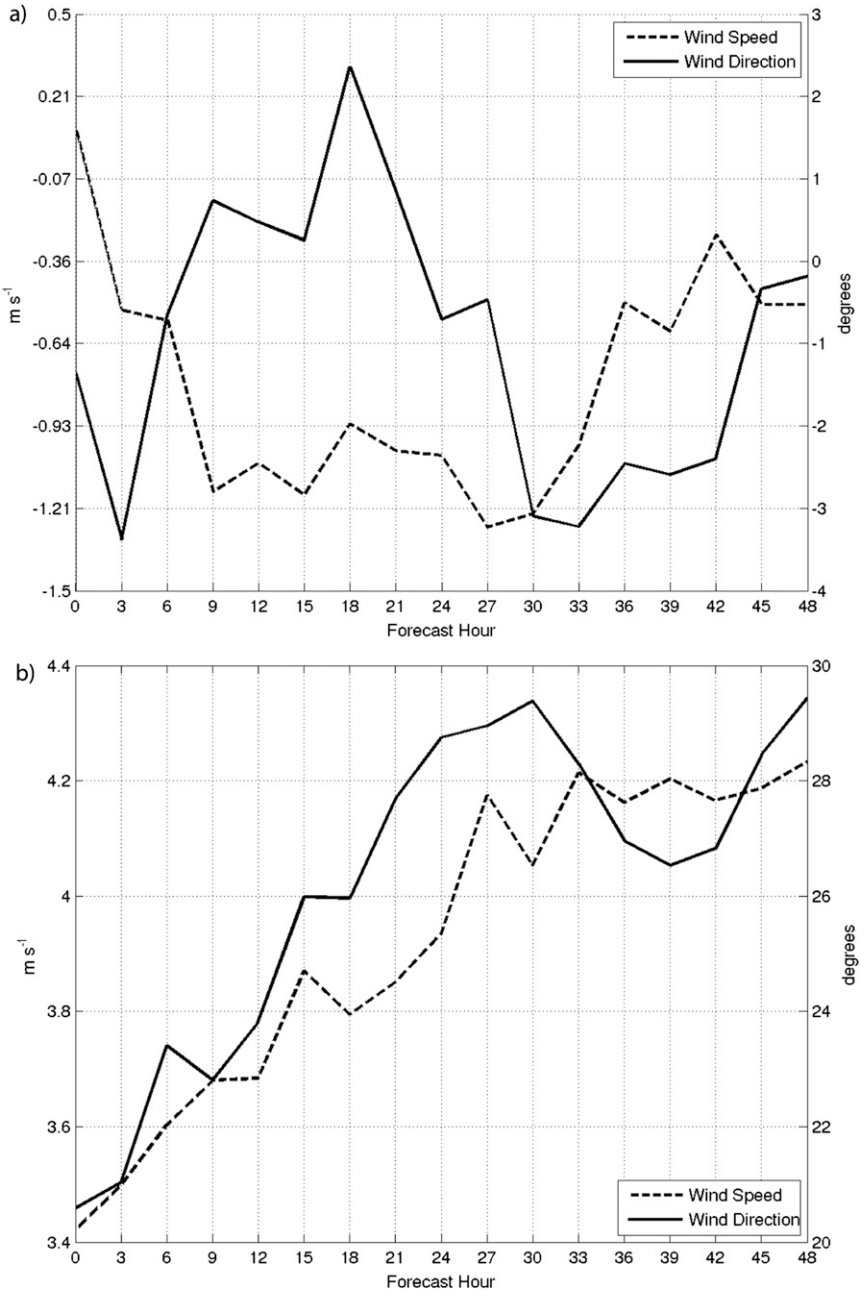

FIG. 6. (a) Domain-wide averaged bias and (b) mean absolute error for wind speed (dashed line) and direction (solid line) between AROME-WMED forecasts and multiple-Doppler syntheses at 2 km MSL. Error convention is (model − analysis).

AGL, the error remains relatively constant at about 4.4 m s^{-1} throughout all forecast times. However, in order to ensure that these errors are truly a result of location and not just elevated wind speeds over higher terrain causing higher errors, wind speed MAE were normalized using the domain-wide average wind speed for regions above and below 1 km MSL (Fig. 9). Normalized MAE are also lower for areas where the elevation is less than 1 km MSL, up to 36 h, with an average difference of 2%–3% error.

A Student's *t* test was used to assess whether the differences in the two different error distributions below and above 1 km MSL are statistically significant. With 95% confidence, the *t* test indicated that the differences are significant when data from all forecast hours are used. A separate *t* test was calculated for the period

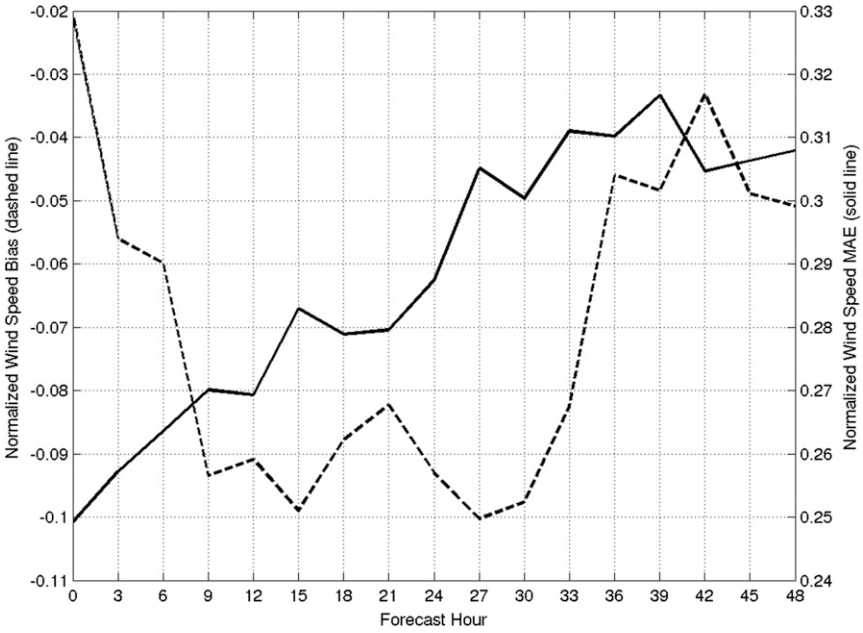

FIG. 7. Normalized wind speed bias (dashed line) and normalized wind speed MAE (solid line) between AROME-WMED forecasts and multiple-Doppler syntheses at 2 km MSL.

between 30 and 48 h, where the errors are nearly similar at about 30%. Between these times, there is no statistical significance. Therefore, wind speed error is indeed enhanced in areas with terrain above 1 km MSL, specifically between initialization and the 30-h forecast time. It is possible that these differences in wind speed MAE may be the result of wind flow perturbations caused by rapid changes in topography or localized terrain-induced mesoscale phenomena that may not be well captured by AROME-WMED. Another possibility is the topography resolution, which may differ enough from that of the actual terrain to impede the ability to capture small-scale orographically induced changes to the wind field that are propagated upward to 2 km MSL. These errors may then be translated downstream. Finally, for some regions above 1 km MSL, the 800-hPa surface may approach or enter the boundary layer, where the role of turbulence and shallow convection parameterizations in error generation at 2 km MSL may be more important than for terrain less than 1 km MSL.

For domain-wide average wind direction MAE as a function of terrain height (Fig. 8b), a similar trend is seen when compared with that of wind speed MAE. Regions with terrain below 1 km MSL have lower wind direction MAE than areas with topography above 1 km MSL. However, wind direction MAE at initialization are nearly identical, with error increasing rapidly for regions where the topography is greater than 1 km MSL, while MAE increase more slowly for lower terrain. Both regions show that wind direction MAE do not change much after the 12-h forecast time, with the average error for terrain below 1 km AGL stabilizing near 16°, and about 22° for higher terrain. Therefore, forecast times approaching 48 h do not see convergence of wind direction MAE based on terrain, as was the case for wind speed. Again, a Student's t test revealed that the two distributions of wind direction MAE are statistically significant from one another, confirming elevated wind direction errors for regions above 1 km AGL. Location error of the main upper-level trough, cutoff low, and medicane at extended forecast times may explain this difference. As was the case for wind speed MAE based on topography, the presence of complex terrain below the analysis level of 2 km MSL is implicated in wind direction MAE discrepancies.

Further analysis of error characteristics also included the assessment of domain-wide, averaged MAE based on observed wind speed classification (Fig. 10). Differences in MAE were calculated for both wind speed (Fig. 10a) and wind direction (Fig. 10b) based on categories of the observed wind speed between 10 and 25 m s^{-1}, in increments of 5 m s^{-1}. Looking at the general trend in MAE within classified wind speeds (Fig. 10a), the error magnitude increases with increasing observed wind speed. However, in order to get a better idea of the MAE within classified wind speed categories, the MAE was normalized by the domain-wide, mean wind speed within each category (Fig. 11). The results show that the

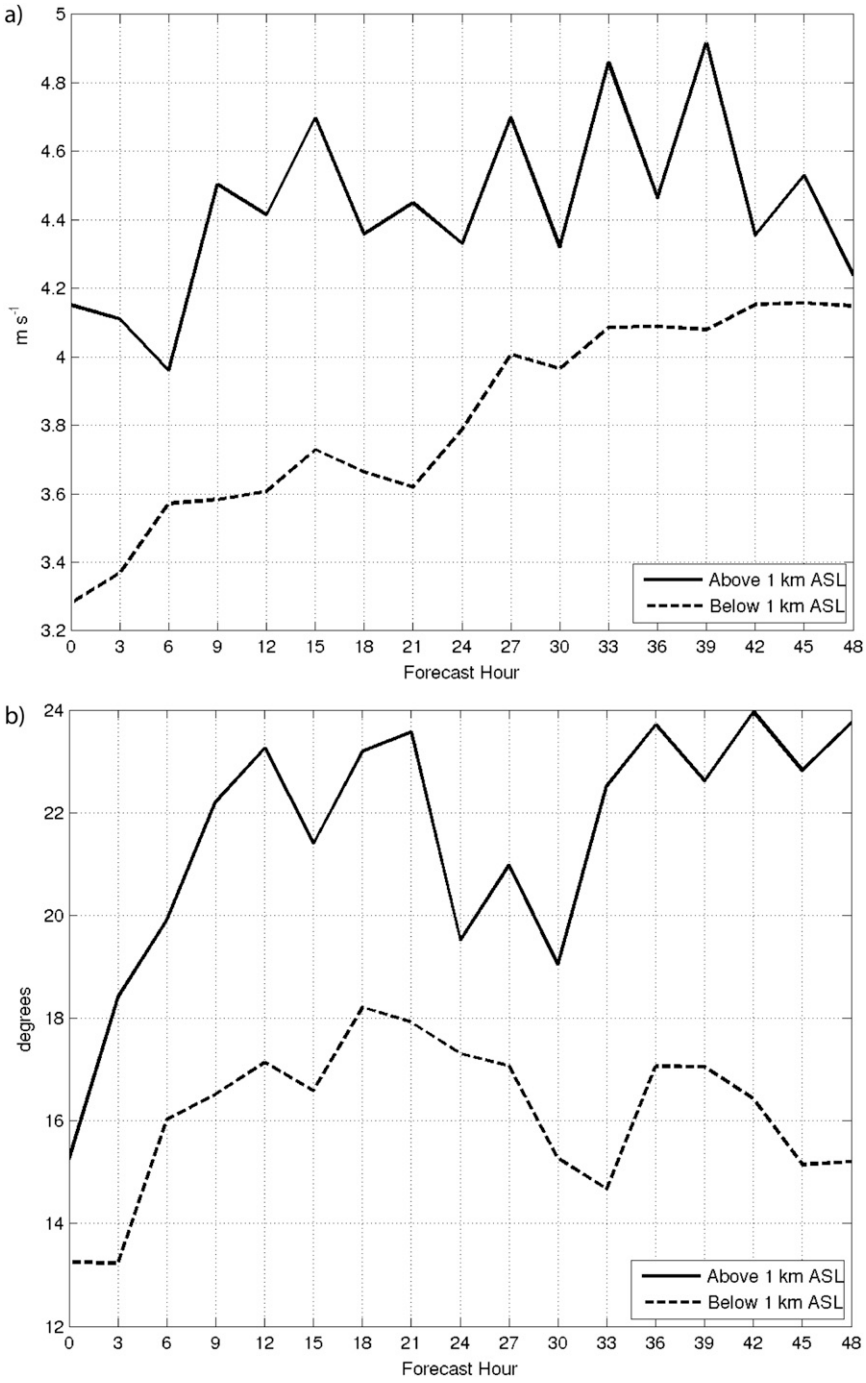

FIG. 8. Domain-wide averaged MAE for regions with terrain below (dashed lines) and above (solid lines) 1 km MSL for (a) wind speed and (b) direction between AROME-WMED forecasts and multiple-Doppler syntheses at 2 km MSL. Error convention is (model − analysis).

percentage MAE increases with decreasing wind speed. Therefore, the largest normalized MAE in wind speed occurs where the weakest wind speeds were observed. Up to a 50% error in normalized MAE is observed in the lowest wind speed category, while areas where winds are greater than 25 m s^{-1} contain normalized MAE of less than 20%.

To determine whether AROME-WMED errors identified as a function of normalized wind speed were truly significant, *t* tests were again applied. With 95% confidence,

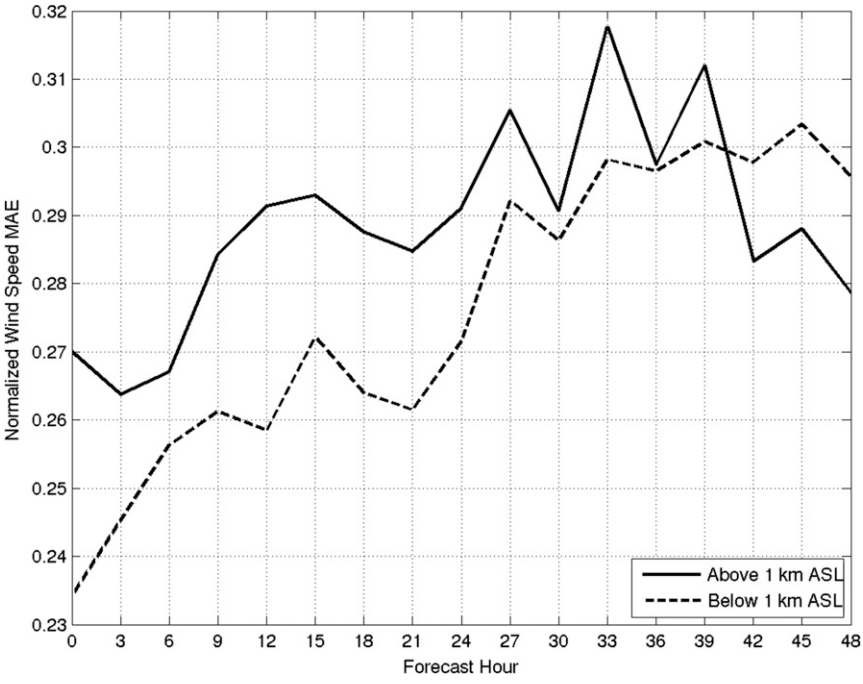

FIG. 9. Domain-wide averaged normalized wind speed MAE for regions with terrain below (dashed line) and above (solid line) 1 km MSL for wind speed between AROME-WMED forecasts and multiple-Doppler syntheses at 2 km MSL.

each category of normalized wind speed MAE was shown to be statistically significant from the others. Therefore, AROME-WMED indeed showed a relation between the normalized wind speed strength and the associated error. These errors are likely linked with the ability to forecast different meteorological patterns. Strong synoptic-scale temperature gradients and associated low pressure systems will tend to have smaller normalized wind speed MAE than regions with weaker temperature gradients, or synoptic high pressure. However, mesoscale phenomena will influence the error in all wind speed categories as they can be underresolved or poorly forecasted. Areas of synoptic high pressure may also be more susceptible to mesoscale error. Therefore, the ability to detect differences between sources of synoptic and mesoscale error within the model is important. The ARAMIS radar network provides the high-resolution, homogenous observational network necessary to identify these scale-dependent errors.

When assessing the impact of categorizing observed wind speed on the associated domain-averaged wind direction MAE (Fig. 10b), an inverse relationship is found compared with the wind speed MAE. The strongest wind speed categories are associated with the smallest amounts of wind direction MAE. For wind speeds of less than $10\,m\,s^{-1}$, wind direction MAE show no particular trend, with errors reaching nearly 28° at 18 h. Yet, higher wind speed categories show more moderate increases in wind direction MAE as a function of forecast hour. Observed wind speeds greater than $25\,m\,s^{-1}$ are related to wind direction MAE of less than 10° at initialization and do not exceed 14° at any forecast hour.

When all forecast hours are considered together, the Student's t test shows that each distribution of error in wind direction MAE is statistically significant from the others, fortifying the findings above. However, when analyzed separately, three periods exist where the wind direction MAE for wind speeds less than $10\,m\,s^{-1}$ show no difference between the other distributions (at 3 h, between 27 and 33 h, and between 42 and 48 h). Therefore, while stronger wind categories show consistent and significant differences in wind direction MAE, the weakest winds are only sometimes linked to higher wind direction MAE. However, overall, these results may suggest that meteorological phenomena with strong winds are likely better forecasted and influence these statistics. In addition, regions of high pressure may cause particular difficulty in wind direction forecasting, and may explain generally higher wind direction MAE values associated with the weakest wind speeds categories.

In addition to changes in model error as a function of forecast hour, it is informative to assess the spatial distribution of error given the complex terrain of southern

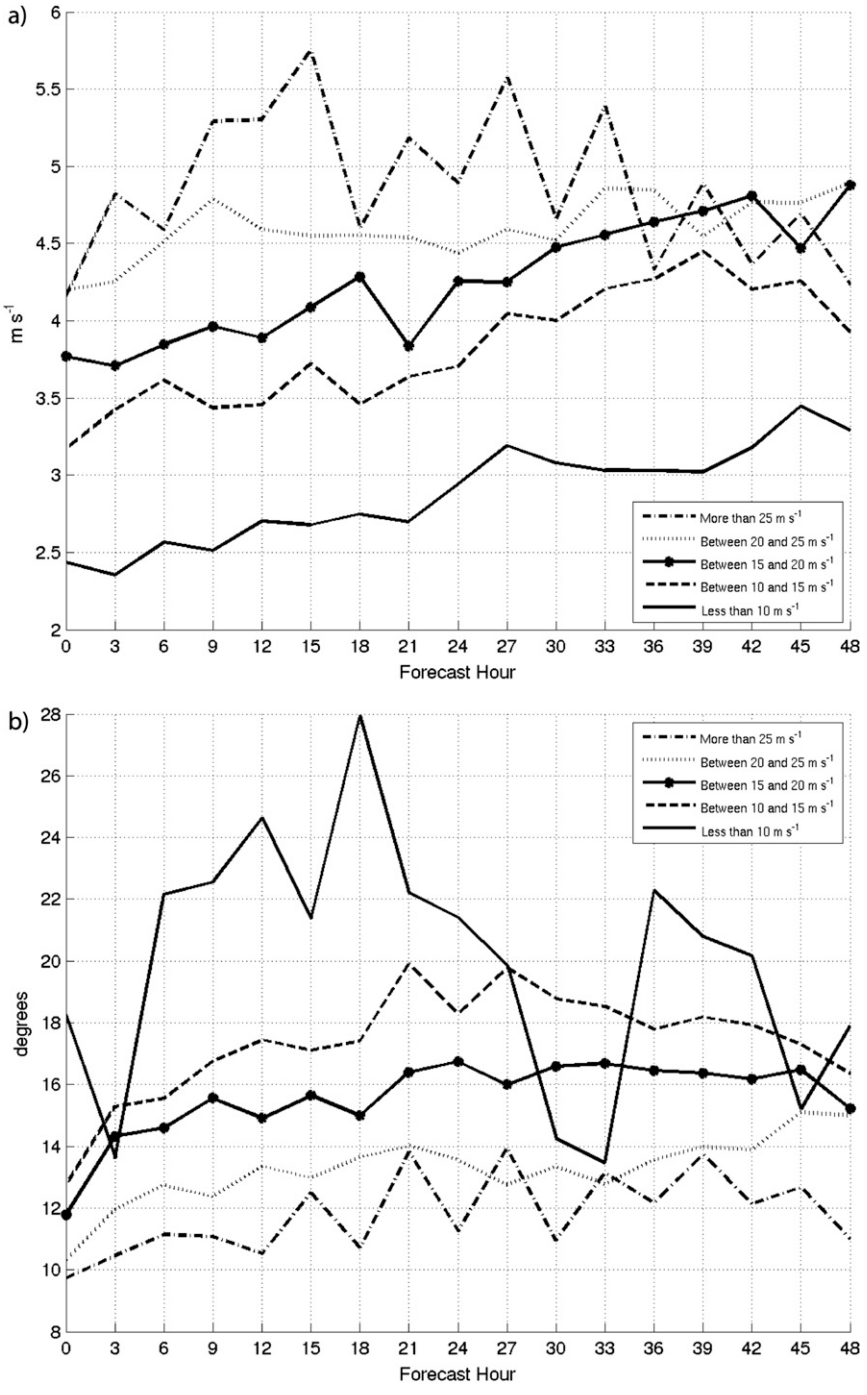

FIG. 10. Domain-wide averaged MAE for AROME-WMED (a) wind speed and (b) wind direction compared to multiple-Doppler syntheses as a function of observed wind speed categories. Error convention is (model − analysis).

France (Fig. 12) and the coverage of the multiple-Doppler dataset. Data from the analysis period at 0000 UTC and all simulation forecasts were separated to assess the spatial distribution at initialization compared to forecast lead times. The availability and distribution of multiple-Doppler data are shown later (see Fig. 15). A total of 25 initializations were incorporated (Fig. 13a) with 392 different forecast times (Fig. 13b). It should be noted that in addition to precipitation patterns and evolution, data availability is also a function of radar

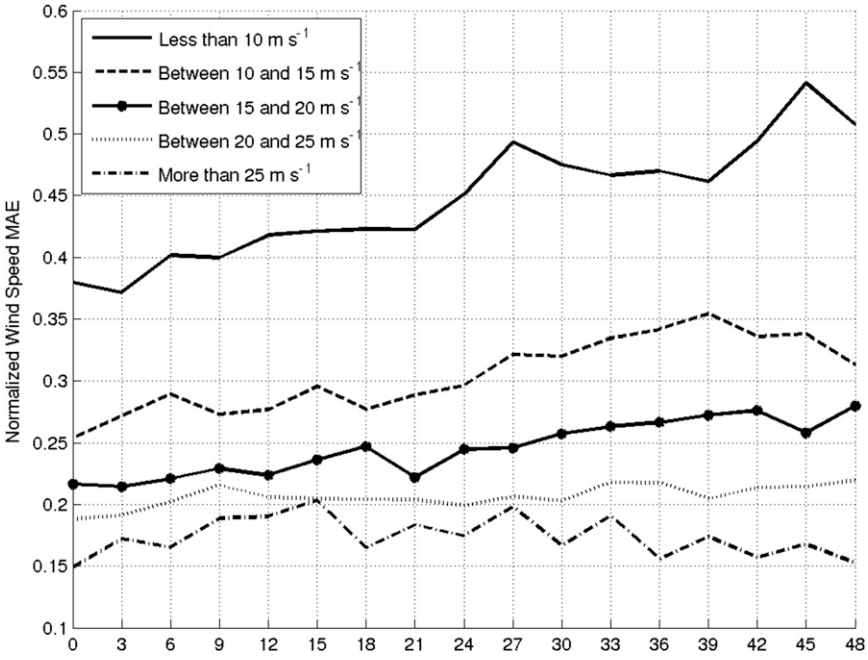

FIG. 11. Domain-wide averaged normalized wind speed MAE for AROME-WMED compared to multiple-Doppler syntheses as a function of observed wind speed categories.

location, given the need for (at least) two radars to retrieve the wind field. Consequently, boundaries may appear where this limit is reached. Nevertheless, it is clear that certain regions received greater amounts of precipitation than others throughout the multiple-Doppler analysis (extended periods of orographically induced, convective precipitation over the southeastern Massif Central are associated with the highest concentration of data points).

In an effort to assess different sources of model error and their spatial distribution, the normalized bias for wind magnitude and wind direction bias were calculated as point averages over all initialization times and then for all forecasts (Figs. 14 and 15). By assessing only initialization times, it is possible to isolate the error produced due to the availability of unevenly distributed and/or sparse observations as well as 3DVAR and, therefore, deviation from the true atmospheric state (Figs. 14a and 15a). The evolution of error resulting from the progression of the model state can then be assessed by looking at all forecast times together (Figs. 14b and 15b). To produce a more statistically robust analysis, locations with fewer than 5 data points for all initializations and 85 points for all forecasts were excluded (thresholds above these values greatly decreased the spatial extent of the dataset).

The normalized wind magnitude bias at initialization (Fig. 14a) shows a somewhat random error distribution, with respect to both land and water. However, several locations over high terrain indicate a positive normalized bias of up to 30% (e.g., Massif Central and southern Alps) when compared to other regions. Overall, the normalized bias magnitudes are generally ±20% (errors of 1–3 m s^{-1}; not shown) with double these values present in localized regions. The wind speed biases for all forecast times (Fig. 14b) are much smoother (attributable, at least in part, to a larger sample size than at initialization), with generally smaller errors than those seen in the initialization, especially in regions below 1 km MSL. However, locations that do show elevated normalized bias are situated over complex terrain, specifically with negative errors over the Massif Central, and positive errors over the southern Alps and eastern Pyrenees. In addition, the Massif Central undergoes the biggest shift in normalized bias within the domain, from about 20% to −20% error from initialization to forecast period. The best improvement occurs over low-lying areas, in particular the western portion of the domain, where error rates at initialization approach 30% and during the forecast period drop to nearly 0%. Overall, during the analysis period, the wind magnitude is generally underforecasted at 2 km MSL, with elevated terrain showing a tendency for more error.

A similar error trend is found when analyzing wind direction bias (Fig. 15). Initialization bias shows no tendency as a function of location or terrain height (Fig. 15a), with values generally smaller than ±30°. Larger biases are, however, concentrated in localized regions.

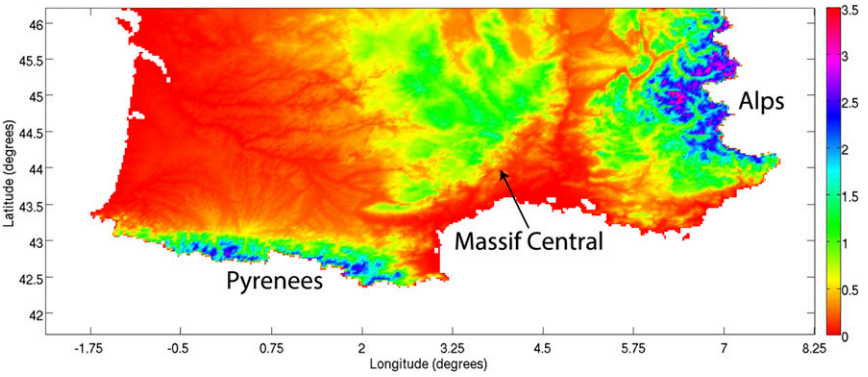

FIG. 12. Topography of southern France. Height (km) MSL is shown in color. Areas of elevated terrain are labeled.

Wind direction biases during all forecast hours (Fig. 15b) show a smoother error field than at initialization (as was the case with wind magnitude bias) with smaller errors, generally less than ±20° over the whole domain. In particular, errors are minimized over elevations below 500 m MSL. Wind direction biases appear to be affected by certain regions of terrain, with positive wind direction biases associated with regions near the southwestern

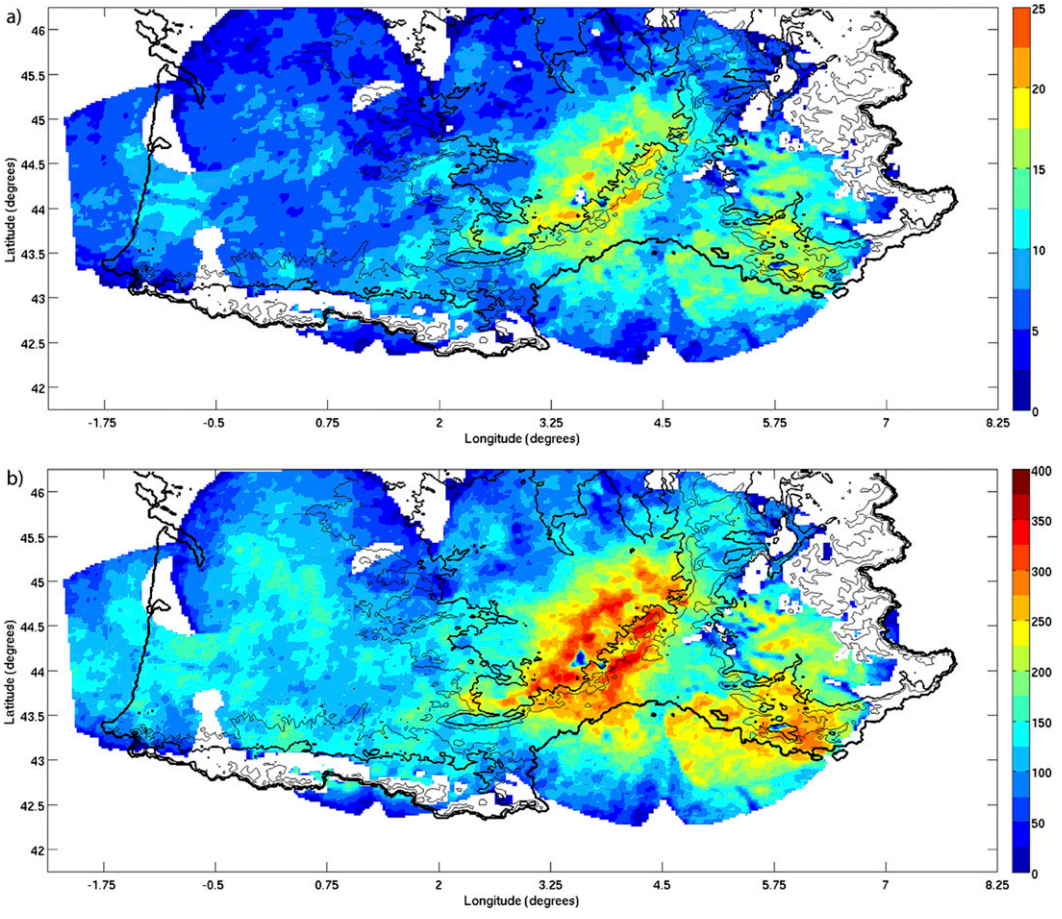

FIG. 13. Spatial distribution of multiple-Doppler data points available for (a) initialization and (b) all forecast hours combined. Black contour lines indicate 0.5, 1.0, and 2.0 km MSL. The 1.0 km MSL contour line is thickened to delineate regions above and below this altitude.

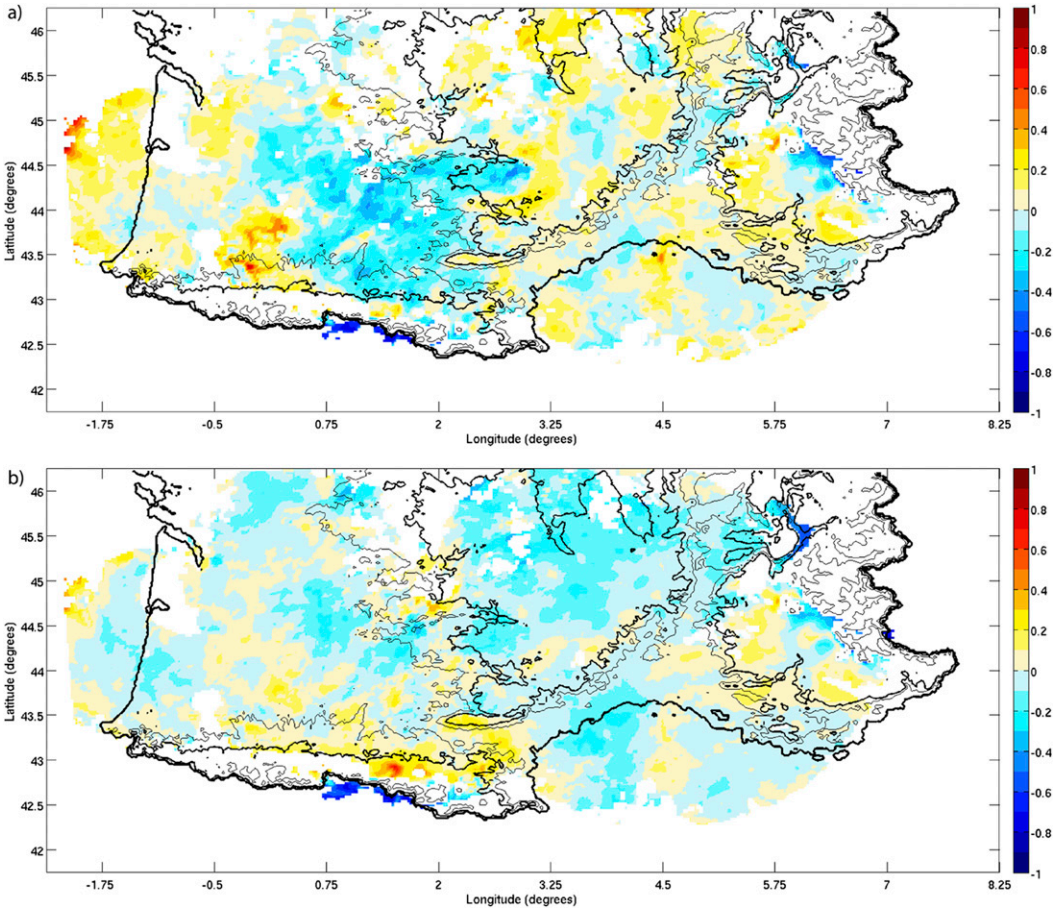

FIG. 14. Spatial distribution of normalized bias for wind magnitude ($m\,s^{-1}$) at (a) initialization and (b) all forecast hours combined. Contour lines are as in Fig. 13.

and southeastern foothills of the Massif Central, and negative wind direction biases over the Pyrenees and southern Alps.

These model error statistics represent a unique and potentially valuable resource for comparing the performance of a model with regard to changes in terrain, impacts of land–sea transition and interaction, and how the model handles physical processes related to complex wind flow, for example. The data and analyses presented serve as a proof of concept, given that the analysis period only spans 9 days, and therefore, the ability to draw conclusions regarding model behavior applies to this specific case. However, given a sufficiently long period, conclusive results could be obtained from such analyses and may offer new types of model comparison and error assessment that were previously unfeasible.

4. Conclusions

Results of our AROME-WMED wind magnitude and direction forecast verification at 2 km MSL are presented using multiple-Doppler syntheses from the ARAMIS national network of Doppler radars over a 9-day analysis period during 1–9 November 2011. Relationships are found suggesting that normalized wind magnitude/direction bias and MAE are not only linked to the forecasted evolution of meteorological phenomena, but errors at 2 km MSL are sensitive to terrain height below the analyzed level. Categorized forecasted wind speeds also illustrate error susceptibility to the strength of the forecasted events and wind direction uncertainties associated with regions where weak wind magnitudes are present. The temporal evolution of normalized MAE as a function of forecast lead time indicates that errors plateau near 30 h, with generally stable values thereafter. This finding provides an incentive for an extension of the operational, 30-h AROME simulations toward longer forecast times.

In addition to temporal error as a function of forecast lead time, spatial distributions of bias allow for the specific identification of regions prone to error, including tendencies in wind magnitude and direction over

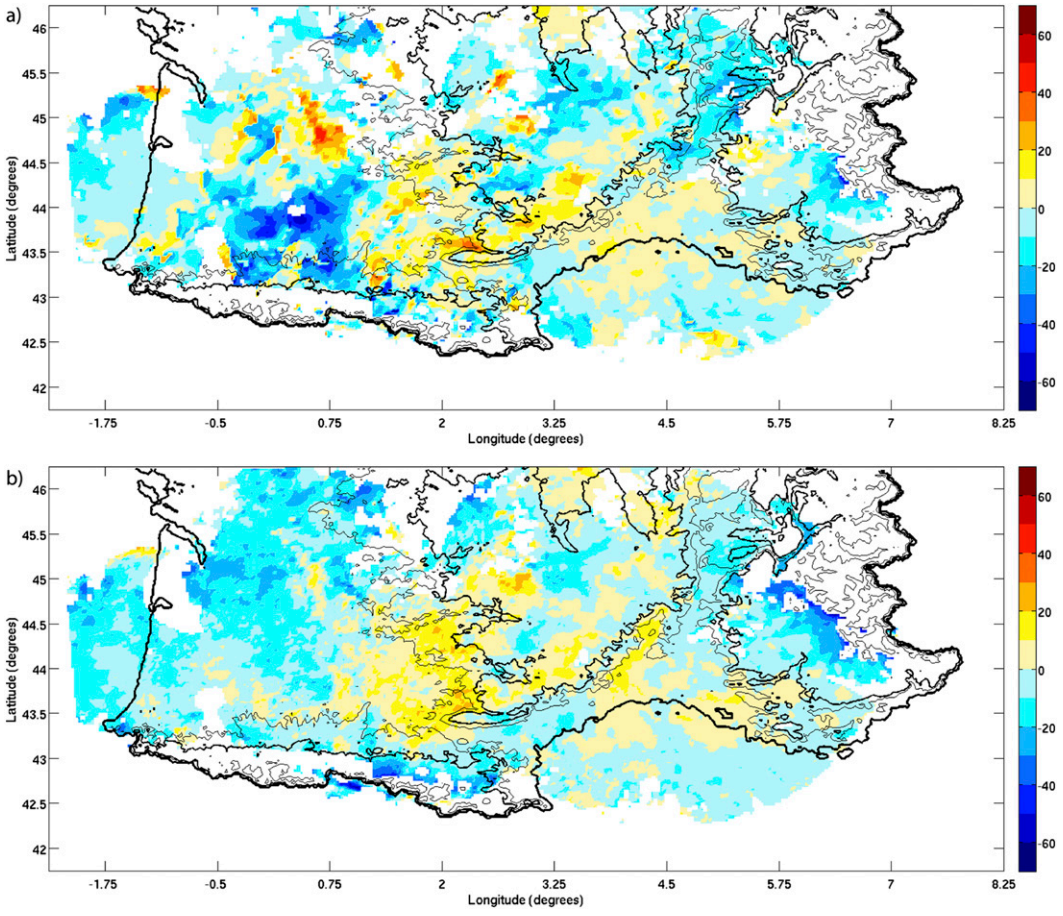

FIG. 15. Spatial distribution of bias for wind direction (°) at (a) initialization and (b) all forecast hours combined. Contour lines are as in Fig. 13. Error convention is (model − analysis).

regions such as the Massif Central, the southern Alps, and the Pyrenees. Results indicate that for the analysis period, wind speeds are overestimated near the southern Alps and underestimated over the Massif Central. Positive wind direction biases near the southwestern and southeastern foothills of the Massif Central indicate that the gradient in topography below the forecast level may play a role in forecast wind direction errors. Regions below 1 km MSL showed the smallest wind magnitude and direction bias. Separation of spatial error between initialization and forecast times show that biases are generally independent of location and terrain height at initialization, but that the impact of terrain below the analysis level affects the forecasted wind magnitude and direction over time.

These error analyses show that wind syntheses derived from Doppler radar networks can serve as a robust dataset for the verification of model performance over large areas and at multiple levels above the ground. The advantage of multiple-Doppler syntheses over a large region is related to the high temporal and spatial resolution and homogeneous distribution of the data, which, at levels above ground, traditional wind observations are unable to match. For example, within the domain studied, only two sounding stations exist, separated by a distance of ~600 km and which have a temporal resolution of 12 h, compared to 2.5 km and 15 min for the radars. Therefore, these soundings only represent local conditions, and would be difficult to use in the assessment of the overall wind field characteristics for the domain analyzed. However, limitations exist in regard to the vertical extent of the verification domain for multiple-Doppler wind retrieval datasets. Surface wind model verification is not feasible due to radar beam blockage and ground clutter, and the lowest level possible for wind retrieval is a function of the baseline distance between the radars, with longer distances increasing this height. Therefore, use of radar wind retrieval data in conjunction with surface-based, radiosonde, and other in situ observations is optimal, and could improve the high-resolution model verification for other models run in regions containing dense radar networks.

With sufficient data storage availability (similar to that of model output when considering 3D radar wind retrieval), the automated production of these multiple-Doppler syntheses paves the way for verification to be conducted on a real-time basis. Additional fields such as vertical velocity, reflectivity, and polarimetric radar data may provide greater opportunity for model verification in the future. In addition, two-dimensional multiple-Doppler winds and/or radial velocities from multiple radars could be incorporated into the current 3DVAR methods (e.g., Gao et al. 2013) to improve the model initialization and subsequent forecasts.

Acknowledgments. The authors wish to thank the European Union, the Provence-Alpes-Côte d'Azur Region and the French Ministry of Ecology, Energy, Sustainable Development and Sea. Thanks are also due to Joël Stein and Nadia Fourrié for their help during this research.

REFERENCES

Accadia, C., S. Zecchetto, A. Lavagnini, and A. Speranza, 2007: Comparison of 10-m wind forecasts from a regional area model and QuikSCAT scatterometer wind observations over the Mediterranean Sea. *Mon. Wea. Rev.,* **135,** 1945–1960, doi:10.1175/MWR3370.1.

Beck, J., and O. Bousquet, 2013: Using gap-filling radars in mountainous regions to complement a national radar network: Improvements in multiple-Doppler wind syntheses. *J. Appl. Meteor. Climatol.,* **52,** 1836–1850, doi:10.1175/JAMC-D-12-0187.1.

Bousquet, O., and M. Chong, 1998: A multiple-Doppler synthesis and continuity adjustment technique (MUSCAT) to recover wind components from Doppler radar measurements. *J. Atmos. Oceanic Technol.,* **15,** 343–359, doi:10.1175/1520-0426(1998)015<0343:AMDSAC>2.0.CO;2.

——, and P. Tabary, 2014: Development of a nationwide real-time 3-D wind and reflectivity radar composite in France. *Quart. J. Roy. Meteor. Soc.,* doi:10.1002/qj.2163, in press.

——, C. Lin, and I. Zawadzki, 2006: Analysis of scale dependence of quantitative precipitation forecast verification: A case study over the Mackenzie River basin. *Quart. J. Roy. Meteor. Soc.,* **132,** 2107–2125, doi:10.1256/qj.05.154.

——, T. Montmerle, and P. Tabary, 2008a: Using operationally synthesized multiple-Doppler winds for high resolution horizontal wind forecast verification. *Geophys. Res. Lett.,* **35,** L10803, doi:10.1029/2008GL033975.

——, P. Tabary, and J. Parent-du-Châtelet, 2008b: Operational multiple-Doppler wind retrieval inferred from long-range radial velocity measurements. *J. Appl. Meteor. Climatol,* **47,** 2929–2945, doi:10.1175/2008JAMC1878.1.

Brousseau, P., L. Berre, F. Bouttier, and G. Desroziers, 2011: Background-error covariances for a convective-scale data-assimilation system: AROME-France 3D-Var. *Quart. J. Roy. Meteor. Soc.,* **137,** 409–422, doi:10.1002/qj.750.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374, doi:10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Davis, C., B. Brown, R. Bullock, and J. Halley-Gotway, 2009: The Method for Object-Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Wea. Forecasting,* **24,** 1252–1267, doi:10.1175/2009WAF2222241.1.

Ducrocq, V., and Coauthors, 2014: HyMeX-SOP1, the field campaign dedicated to heavy precipitation and flash-flooding in Northwestern Mediterranean. *Bull. Amer. Meteor. Soc.,* doi:10.1175/BAMS-D-12-00244.1, in press.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting,* **24,** 1498–1510, doi:10.1175/2009WAF2222251.1.

——, and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.,* **239,** 179–202, doi:10.1016/S0022-1694(00)00343-7.

Gao, J., and Coauthors, 2013: A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting,* **28,** 727–745, doi:10.1175/WAF-D-12-00093.1.

Geleyn, J. F., 1988: Interpolation of wind, temperature and humidity values from model levels to the height of measurements. *Tellus,* **40A,** 347–351, doi:10.1111/j.1600-0870.1988.tb00352.x.

Gelsthorpe, R. V., E. Schied, and J. J. W. Wilson, 2000: ASCAT-MetOp's advanced scatterometer. ESA Bull. 102, 9 pp. [Available online at http://www.esa.int/esapub/bulletin/bullet102/Gelsthorpe102.pdf.]

Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.,* **91,** 1365–1373, doi:10.1175/2010BAMS2819.1.

Gourley, J. J., P. Tabary, and J. Parent-du-Chatelet, 2007: A fuzzy logic algorithm for the separation of precipitating from non-precipitating echoes using polarmetric radar observations. *J. Atmos. Oceanic Technol.,* **24,** 1439–1451, doi:10.1175/JTECH2035.1.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydrometeor.,* **2,** 406–418, doi:10.1175/1525-7541(2001)002<0406:MSPOAH>2.0.CO;2.

HyMeX, cited 2013: Hydrological cycle in the Mediterranean Experiment. [Available online at http://www.HyMeX.org/.]

Lean, H. W., J. Bornemann, Y. Tang, M. Mittermaier, N. Roberts, and M. Trueman, 2011: Experiences with a 1.5 km version of the Met Office Unified Model for short range forecasting. *Proc. 24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., 4B.3. [Available online at https://ams.confex.com/ams/91Annual/webprogram/Paper177409.html.]

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.

Montmerle, T., and C. Faccani, 2009: Mesoscale assimilation of radial velocities from Doppler radars in a preoperational framework. *Mon. Wea. Rev.,* **137,** 1939–1953, doi:10.1175/2008MWR2725.1.

Moscatello, A., M. M. Miglietta, and R. Rotunno, 2008a: Numerical analysis of a Mediterranean ''hurricane'' over southeastern Italy. *Mon. Wea. Rev.,* **136,** 4373–4397, doi:10.1175/2008MWR2512.1.

——, ——, and ——, 2008b: Observational analysis of a Mediterranean ''hurricane'' over southeastern Italy. *Weather,* **63,** 306–311, doi:10.1002/wea.231.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of

convective events. *Mon. Wea. Rev.,* **136,** 78–97, doi:10.1175/2007MWR2123.1.

Salonen, K., S. Niemelä, and C. Fortelius, 2011: Application of radar wind observations for low-level NWP wind forecast validation. *J. Appl. Meteor. Climatol.,* **50,** 1362–1371, doi:10.1175/2010JAMC2652.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.,* **139,** 976–991, doi:10.1175/2010MWR3425.1.

Skamarock, W. C., 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.pdf.]

Tabary, P., F. Guibert, L. Perier, and J. Parent-du-Chatelet, 2006: An operational triple-PRT Doppler scheme for the French radar network. *J. Atmos. Oceanic Technol.,* **23,** 1645–1656, doi:10.1175/JTECH1923.1.

Vasic, S., C. A. Lin, I. Zawadzki, O. Bousquet, and D. Chaumont, 2007: Evaluation of precipitation from numerical weather prediction models using values retrieved from radars and satellites. *Mon. Wea. Rev.,* **135,** 3750–3766, doi:10.1175/2007MWR1955.1.

Westrelin, S., P. Mériaux, P. Tabary, and Y. Aubert, 2012: RHYTMME Project: Risk management based on a radar network. *Proc. Seventh European Conf. on Radar in Meteorology and Hydrology,* Toulouse, France, ERAD, 11.5. [Available online at http://www.meteo.fr/cic/meetings/2012/ERAD/short_abs/HS_404_sh_abs.pdf.]