Deterministic Wind Speed Predictions with Analog-Based Methods over Complex Topography

IRIS ODAK PLENKOVIĆ

Meteorological and Hydrological Service, Zagreb, Croatia

LUCA DELLE MONACHE

National Center for Atmospheric Research, Boulder, Colorado

KRISTIAN HORVATH AND MARIO HRASTINSKI

Meteorological and Hydrological Service, Zagreb, Croatia

(Manuscript received 1 June 2017, in final form 11 May 2018)

ABSTRACT

The performance of analog-based and Kalman filter (KF) postprocessing methods is tested in climatologically and topographically different regions for point-based wind speed predictions at 10 m above the ground. The results are generated using several configurations of the mesoscale numerical weather prediction model ALADIN. This study shows that deterministic analog-based predictions (ABPs) improve the correlation between predictions and measurements while reducing the forecast error, with respect to both the starting model predictions and the KF-based correction. While the KF generally outperforms the ABPs in bias reduction, the combination of the KF and analog approach can be similarly successful. In the coastal complex area, characterized with a larger frequency of strong wind, the ABPs are more successful in reducing the dispersion error than the KF. The application of the KF algorithm to the analogs in the so-called analog space (KFAS) is the least prone to standard deviation underestimation among the ABPs. All ABPs improve the prediction of larger-than-diurnal motions, and KFAS is superior among all ABPs in predicting alternating wind regimes on time scales shorter than a day. The ABPs better distinguish different wind speed categories in the coastal complex terrain by using a higher-resolution model input. Differences among starting model and postprocessed forecasts in other types of terrain are less pronounced.

1. Introduction

The skill of numerical weather prediction models has improved at both global and regional scales. Their ability to simulate and forecast winds in complex terrain and coastal areas is, however, still largely affected by insufficient resolution, imperfect boundary and initial conditions, simplification of physical processes, and numerical approximations. It is often considered that the higher the model resolution, the more accurate the forecast, because of better resolved lower boundary conditions and flow adaptation when decreasing the grid spacing. These benefits are not always evident (e.g., Mass et al. 2002; Rife and Davis 2005). Even at the sub-kilometer grid spacing, state-of-the-art mesoscale models still exhibit considerable errors, especially in complex terrain (Horvath et al. 2012). This is particularly relevant for operational weather prediction systems that are constrained by the available computing resources. It is thereby useful to develop postprocessing methods that reduce starting model errors at locations where measurements are available, in addition to improving the model itself (e.g., using higher-resolution or an improved parameterization package).

The idea that analogies (i.e., to similar past forecasts, measurements, or analysis) can be used for forecasting future weather has been explored for decades. It is based on an assumption that if two atmospheric states are initially very close, they will remain somewhat close for

ⓐ Denotes content that is immediately available upon publication as open access.

Corresponding author: Iris Odak Plenković, odak@cirus.dhz.hr

some time in the future. It is hard to identify any state in the past that can be considered a good match to the present large-scale flow pattern, except for mediocre analogs (Lorenz 1969). Ruosteenoja (1988) and Lorenz (1969) quote an astronomical number of years that one has to wait until it is to be expected to find two atmospheric states that are within present-day observational error. The use of analogs for short-range weather forecasting thus is practically discarded. Van den Dool (1989), however, shows that it is possible to find useful analogies if the number of the degrees of freedom in the matching procedure is reduced. The author uses analyses over a localized area [i.e., not the entire Northern Hemisphere as in Lorenz (1969)] and then uses the 12-h subsequent analysis to each analog as a plausible 500-hPa height forecast. Various procedures are formulated afterward, including different predictors and analog selection criteria. The use of analogs for forecasting of meteorological fields is limited because of excessive degrees of freedom of the problem at stake. However, applications including long-range weather predictions using National Oceanic and Atmospheric Administration (NOAA) outgoing longwave radiation fields (Xavier and Goswami 2007) and very short-term orographic precipitation predictions using radar observations (Panziera et al. 2011) are skillful. The Southern Oscillation index (SOI) forecasts using SOI measurements (Drosdowsky 1994) and point wind speed forecasts using wind speed measurements (Klausner et al. 2009) also exhibit satisfactory results.

Besides predicting the weather using past measurements or analyses, analogies can be employed to reduce the errors in the numerical weather prediction model simulations. This approach utilizes achievements of numerical modeling in predicting the future state of the atmosphere. Additionally, it can reasonably absorb the information of the analogs in historical data (statistical model) in order to improve forecast skill as shown for idealized cases with low-order models (Ren and Chou 2006) and general circulation modeling (Gao et al. 2006; Ren and Chou 2007). The pioneering contribution of Van den Dool (1989) reveals the ability to predict the forecast skill of a numerical weather prediction (NWP) model. Analogies are also used to calibrate ensemble forecasts (Hamill and Whitaker 2006; Hopson and Webster 2010).

Delle Monache et al. (2011) propose two analog-based postprocessing methods to improve deterministic NWP forecasts of 10-m wind speed, based on a historical dataset including NWP data and observations at a single site. The weighted mean of the analog ensemble (AN) is tested and compared to a linear, adaptive, and recursive Kalman filter

(KF) postprocessing approach (Delle Monache et al. 2006, 2008, 2011). Another approach is to apply a Kalman filter to the historical set of (starting) model forecasts in the analog space, ordered from the worst to the best analog [Kalman filter in analog space (KFAS); Delle Monache et al. 2011]. With that approach, the correction of the current forecast is based on a higher weight to the analog forecasts closer to it. The authors demonstrate that both approaches increase correlation and reduce random and systematic errors. Similar approaches are used for predicting other variables as well. Djalalova et al. (2015) show similar results predicting $PM_{2.5}$ concentration, while Nagarajan et al. (2015) test the techniques across several models and meteorological variables. Additionally, Djalalova et al. (2015) apply the Kalman filter algorithm to the time series of the AN, resulting in a new deterministic forecast called the KFAN. Delle Monache et al. (2013) explore benefits from using the analog ensemble (AnEn) approach to produce probabilistic 10-m wind speed and 2-m temperature forecasts from deterministic numerical model predictions. The authors show that the AnEn exhibits high statistical consistency, reliability, and the ability to capture the flow-dependent behavior of errors. The use of the analog-based approach to produce probabilistic output is not limited to short- or medium-range forecasts. Vanvyve et al. (2015) provide high-quality long-term wind resource estimates, characterized by an accurate wind time series and frequency distribution. In addition to using probabilistic analog-based predictions to gain wind resource estimates (Vanvyve et al. 2015; Zhang et al. 2015), they are also used to downscale precipitation (Keller et al. 2017) and to predict solar irradiance (Alessandrini et al. 2015a) and wind power (Alessandrini et al. 2015b; Junk et al. 2015).

In this study we propose an in-depth analysis of analog-based methods, for the first time over a coastal region characterized by complex topography (Fig. 2). The target area of this study is located in Croatia, where mesoscale wind circulations include strong bora downslope windstorms [which may reach hurricane-scale strength, e.g., see review by Grisogono and Belusic (2009)], mountain valley and slope winds, and thermally induced land–sea breeze (e.g., Telišman Prtenjak and Grisogono 2007; Horvath et al. 2011). Because of the importance of model resolution necessary to represent wind processes in the target area, we study whether the postprocessing improves results when using a higher-resolution starting model. We thus test the role of the 8- and 2-km grid spacing full-physics Aire Limitée Adaptation Dynamique Développement International (ALADIN) model and a dynamical adaptation of the 8-km ALADIN output to the 2-km grid spacing, which is used for operational wind forecasting in the ALADIN

consortium (e.g., Žagar and Rakovec 1999; Ivatek-Šahdan and Tudor 2004).

We study the performance of different postprocessing methods using metrics that consider wind speed as both a continuous and categorical predictand. These include AN, KF, KFAS, and KFAN, as described above. We analyze the results across three regions with distinct wind regimes:

(i) Coastal complex terrain where the most significant portion of mesoscale energy is carried by strong downslope windstorms as well as thermally induced land–sea circulations

(ii) Mountain complex terrain where the most significant portion of mesoscale energy is carried by weak-to-moderate valley and slope mountain winds

(iii) Continental, nearly flat terrain where the motions are predominantly of synoptic-scale variability and origin (Zaninović et al. 2008; Horvath et al. 2011)

Finally, we study the importance of the starting model resolution and formulation by using three versions of ALADIN focusing on coastal complex terrain characterized by the mesoscale wind processes.

In section 2 the postprocessing methods are described, while section 3 introduces the NWP model data. In section 4 the experimental setup is explained in detail, followed by an observational analysis in section 5. The results are presented in section 6, and conclusions are highlighted in section 7.

2. Postprocessing methods

a. Kalman filter

The KF approach is a recursive postprocessing method used to estimate a signal from noisy measurements. The KF method uses recent past forecasts and observations at a given location. The KF method computes past prediction errors and estimates the future bias in the current raw forecast afterward.

The optimal recursive predictor of forecast bias x_t at time t is derived by minimizing the expected mean-square error. Kalman (1960) shows that x_t at time t can be written as a combination of the previous bias estimate and the previous forecast error y_t [the hat (^) indicates the estimate]:

$$\hat{x}_{t+\Delta t|t} = \hat{x}_{t|t-\Delta t} + K_t(y_t - \hat{x}_{t|t-\Delta t}). \tag{1}$$

The K_t is a weighting factor called Kalman gain and can be calculated from

$$K_t = \frac{p_{t-\Delta t} + \sigma_{\eta,t}^2}{(p_{t-\Delta t} + \sigma_{\eta,t}^2 + \sigma_{\varepsilon,t}^2)}. \tag{2}$$

The expected mean-square error p can be computed as

$$p_t = (p_{t-\Delta t} + \sigma_{\eta,t}^2)(1 - K_t). \tag{3}$$

The $\sigma_{\eta,t}^2$ and $\sigma_{\varepsilon,t}^2$ are variances of the noise term and the unsystematic error term, respectively. Their ratio is set to a 0.01 value, following other authors (i.e., Delle Monache et al. 2006, 2011). For any plausible estimate of p_0 and K_0, the KF algorithm converges promptly. Additional details of the procedure and algorithm applied in this study can be found in Delle Monache et al. (2006). Advantages of the KF approach are the short training period and the ability to adapt to changing synoptic conditions. The disadvantage is that it is less likely to predict a large forecast error when all errors for the past few days have been smaller, affecting rapid weather changes or extreme events forecasts (Delle Monache et al. 2011).

b. Analog-based deterministic predictions

The AnEn can be used to estimate the probability distribution $f(y|x^f)$ of the observed future value of the variable y at a given time and location. The x^f represents k variables (predictors) from the deterministic (starting) model prediction $x^f = (x_f^1, x_f^2, \ldots, x_f^k)$. To generate y samples, the AnEn method uses historical data within a specified analog training period for which both the deterministic NWP (starting model) and the verifying observation are available. The best-matching historical forecasts to the current prediction (analogs) may originate in any past date in the training period. The quality of the analog is evaluated by the following metric:

$$\|F_t, A_{t'}\| = \sum_{i=1}^{N_A} \frac{w_i}{\sigma_i} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{i,t+j} - A_{i,t'+j})^2}, \tag{4}$$

where F_t is the current NWP deterministic forecast at the given location, valid at the future time t, whereas $A_{t'}$ is an analog at the given location with the same forecast lead time, but valid at a past time t'. The N_A is the number of predictors used in the search for analogs, and w_i are the weights corresponding to a particular predictor, normalized with the standard deviation of the time series of past forecasts of a given variable at the same location is σ_i. The $\tilde{t}$ is equal to half the number of additional times over which the metric is computed (the half of the time window of any specified width); therefore, $F_{i,t+j}$ and $A_{i,t'+j}$ are the values of the forecast and the analog in the time window for a given variable, respectively. Analogs are found independently for every forecast time and location, narrowing the search around a particular time of day by a time window. Therefore, the number of degrees of freedom in the

analog-finding procedure is reduced. The verifying observations of the best-matching analogs are the members of AnEn. The assumption is that the errors of the good (quality) analog forecasts are likely to be similar to the error of the current forecast (Delle Monache et al. 2011). Once the AnEn is formed, it can be used to produce the deterministic analog-based prediction (ABP), as well as the probabilistic forecast (e.g., to estimate the probability of a predefined event).

The AN_t forecast for the future time t at a given location is an average (weighted, if $\gamma \neq 1$) of the observations O_i corresponding to N most similar analogs $A_{t'}$ [measured by metrics previously defined in Eq. (4)]:

$$AN_t = \frac{1}{N} \sum_{i=1}^{N} \gamma O_i(A_{t',i}). \qquad (5)$$

In another words, the AN_t is a (weighted) mean of N-sized AnEn for a (future) time t. Several authors, such as Delle Monache et al. (2013), state that the AnEn rank histograms are uniform. Every member of the AnEn is thus an equally probable outcome, even though some analogs are closer to the current forecast than the others (measured by previously defined metrics). Hence, the value assigned to the weights γ is 1. Forecasting the median of the AnEn (ANM) is additionally used as an alternative to the AN that is less sensitive to the assumptions about the overall nature of the data (e.g., robust; Wilks 2011) and to the small number of outliers (e.g., resistant; Wilks 2011) . The analogs are searched in forecast space only, for both the AN and ANM. Therefore, no observations are used to select the best analogs, and some sort of correction in real time is desired.

The KF uses all the available information to estimate the error of the current forecast, recursively giving higher weights to the most recent data. However, the KF alone is not able to predict large day-to-day changes in the prediction error, as discussed thoroughly in Delle Monache et al. (2011). The benefits and shortcomings of analog-based and KF approaches complement one another; hence, combining them seems like a reasonable choice. In this paper, two different ways to combine these methods are tested and schematically presented in Fig. 1.

The first combination of analog and the KF approach includes running algorithms independently. First, the AN forecasts are issued (or already saved), completing the time series of the AN forecasts. The last member of the AN time series is valid at the future time t. Then, the KF algorithm is applied (in time) to the time series of the AN forecasts. The KF of the AN forecasts is created (the KFAN forecast). In other words, the KF is applied to the time series of the mean AnEn values. Hereby, every ensemble consists of observations corresponding only to the N best analogs. The KF algorithm gives more weight to the recent AN forecast than the AN forecasts issued at some time in the past. The hypothesis is that the KFAN forecast is as adaptable as the AN forecast (e.g., when large day-to-day changes in the prediction error are present), but as unbiased as the KF forecast.

Another possibility is to run the KF through an ordered set of (all) analog forecasts, rather than in time. The entire time series of analogs is ordered from the least similar (worst analog) to the most similar (best analog) model forecast to the current one, forming an analog space for every future time t. Then, the KF is applied to the ordered set of analogs in analog space (KFAS). The KFAS algorithm weights closeness in analog space and not proximity in time (as in the KFAN forecast). Therefore, the starting model forecast (issued in the past) that is the most similar to the current starting model forecast is given the most weight. This procedure should be able to cope even with drastic changes in both the starting model and the AN forecast error.

3. NWP model data

Three configurations of the operational limited-area mesoscale NWP model ALADIN (ALADIN International Team 1997) of the Croatian Meteorological and Hydrological Service are used to generate 10-m wind speed forecasts:

- The operational limited-area mesoscale ALADIN model is launched twice a day (0000 UTC and 1200 UTC) at 8-km horizontal grid spacing (A8). The A8 model uses the hydrostatic dynamics with spectral solver on 37 hybrid sigma-pressure vertical levels (Tudor et al. 2013). The initial conditions are based on a variational data assimilation scheme for the upper-air fields and optimal interpolation for surface variables (Stanešić 2011). The lateral boundary conditions are taken from the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) global model, which is run operationally at Météo France. Vertical transfer of momentum, heat, and moisture are based on a scheme that uses prognostic-turbulence kinetic energy (Geleyn et al. 2006) combined with modified Louis et al. (1982) stability dependency in the surface layer (Redelsperger et al. 2001). Contribution of shallow convection to evolution of prognostic fields is calculated within the turbulence parameterization according to Geleyn (1987). Deep convection is described by the modified diagnostic Kuo scheme (Geleyn et al. 1994). Microphysics parameterization (Catry et al. 2007) includes prognostic treatment of cloud water/ice, rain, and snow, as well as a statistical
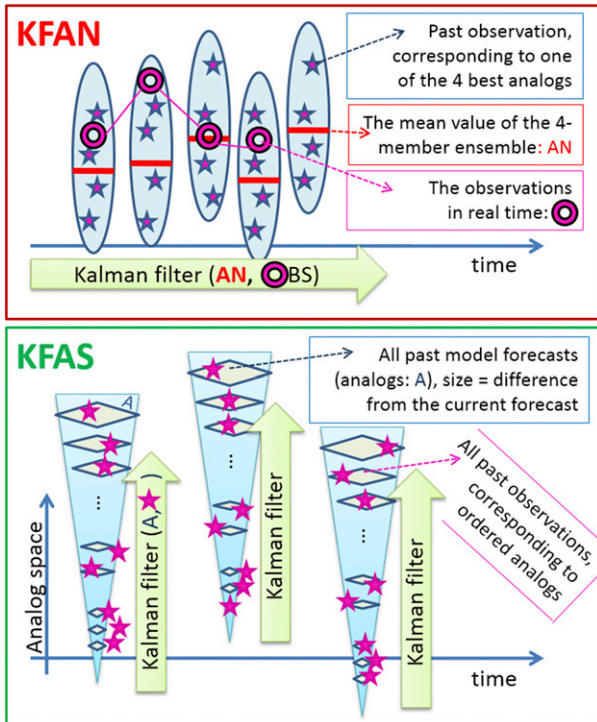
FIG. 1. The schemes for the KFAN and the KFAS forecasts in real time. For the KFAN forecasting, the last member of the AN time series is created, while previously issued AN forecasts are saved. The AN is hereby the mean of the N-member ensemble ($N = 4$ in this example). The KF is then applied to the time series of AN values and real-time measurements, recursively giving the highest weight to the most recent AN (i.e., closest in time). For the KFAS forecasting, the entire time series of previously issued model forecasts (analogs) are sorted by their similarity to the current model forecast, thus forming an analog space. Then, the KF is applied to the analogs and corresponding measurements in the analog space, giving the most weight to the most similar forecast.

approach for sedimentation of precipitation (Geleyn et al. 2008). Radiation effects are described according to Geleyn and Hollingsworth (1979) and Ritter and Geleyn (1992). The impact of soil processes on prognostic model fields is accounted by a two-layer Interactions between Soil, Biosphere, and Atmosphere (ISBA; Noilhan and Planton 1989) scheme, which is also used for the surface data assimilation (Giard and Bazile 2000). The physics contribution is coupled to the dynamics via interface based on a flux-conservative set of equations (Catry et al. 2007).

- An operational ALADIN high-resolution dynamical adaptation (DA) model. The DA procedure (Žagar and Rakovec 1999) takes the output fields from the A8. The DA dynamically adapts wind fields to the higher-resolution horizontal terrain (2-km grid spacing) by adopting the model field to reach a quasi-stationary state forced by time-invariant lateral

boundary conditions (Ivatek-Šahdan and Tudor 2004). Vertical levels in the planetary boundary layer are approximately at the same heights as in the A8 model (the lowest level is about 17 m above ground). The vertical levels in the upper troposphere and stratosphere are reduced; that is, the DA is run on 15 levels in the vertical. The wind field is interpolated to the height of measurements using the stability functions and Monin–Obukhov similarity theory (Geleyn 1988). Turbulence is the only parameterization scheme used in the DA, while contributions of moist and radiation processes are neglected. This cost-effective forecast refinement is run operationally twice a day (0000 and 1200 UTC run) for 72 h ahead with a 3-h model output frequency. In complex terrain the DA improves near-surface wind predictions, as described in a number of studies such as Tudor and Ivatek-Šahdan (2002), Ivatek-Šahdan and Tudor (2004), Ivatek-Šahdan and Ivančan-Picek (2006), Bajić et al. (2007, 2008), Horvath et al. (2011), and so on. The DA is used for operational wind forecasting in several countries that are members of the ALADIN consortia.

- ALADIN at 2-km horizontal grid spacing (A2) is configured similarly to the A8, but with nonhydrostatic dynamics. Physics parameterizations include a full parameterization set as in the A8, with an upgrade of a deep convection parameterization. Unlike the A8, the deep convection in the A2 is a prognostic mass-flux-type scheme (Gerard and Geleyn 2005; Gerard 2007). The convective processes in the A2 are accounted with the use of prognostic variables for updraft and downdraft vertical velocities and mesh fractions (Gerard et al. 2009). The A2 is initialized from the 6-h forecasts of the operational A8 0000 UTC run, and it is run with the scale-selective digital filter initialization (Termonia 2008). This high-resolution forecast is run once daily for 24 h in advance (until 0600 UTC of the following day), with 1-h model output frequency on 37 vertical levels (Tudor et al. 2013).

All three ALADIN configurations (A8 0000 UTC, DA 0000 UTC, and A2 0600 UTC) are used to prepare forecasts for the period 2010–12. The domains for all configurations are shown in Fig. 2. For every location of the analyzed measurement stations, the closest model grid point (on land) is chosen from the four grid points surrounding the observation location.

4. Experimental setup

Model and observation datasets over the 2010–12 period are divided into training and verification periods. The training period is from 2010 to 2011, and 2012 is used as the
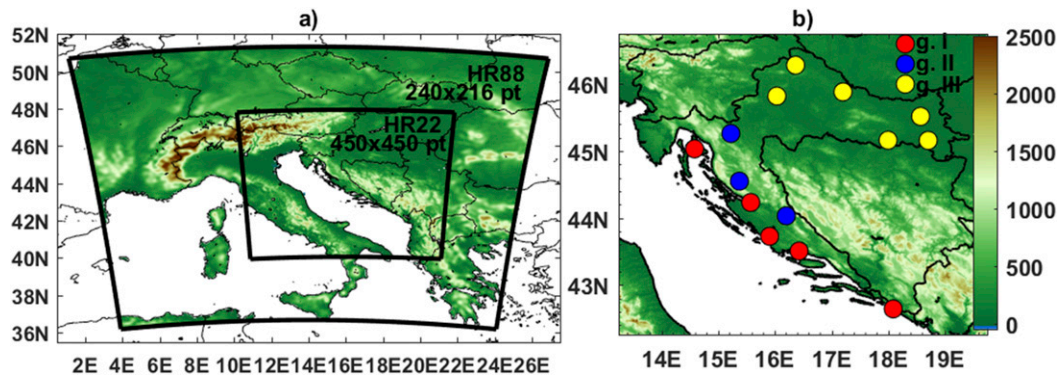
FIG. 2. (a) The ALADIN model domains and terrain—larger with 8-km horizontal grid spacing (A8) and smaller with 2-km horizontal grid spacing (A2; DA). (b) Topography and spatial distribution of the 14 stations providing the 10-m wind speed observations used in this study. The stations are divided in three groups: coastal complex terrain (group I; red markers), mountain complex terrain (group II; blue markers), and nearly flat continental terrain (group III; yellow markers).

verification period. The training period increases gradually after every forecast. As the newer observations might be available in real-time operational setting, they are added to the training database together with the corresponding NWP model forecast. Therefore, the training period is initially 24 months long (for the first verified forecast initialized 1 January 2012) and then prolonged on a daily basis up to 36 months (for the last forecast initialized 31 December 2012). Delle Monache et al. (2006) show that there is an improvement in skill for longer training datasets. The improvement is intense when increasing the training period, especially for training periods up to 6 months. The improvement in skill becomes less notable at around a yearlong dataset. Thus, a dataset ranging from 2 to 3 years should be long enough for this method in our opinion. Furthermore, the ABPs work best with consistent model setup. Since (operational) model setup changes every once in a while, in our opinion it would be better to develop a methodology that can easily adapt to those changes. It is, however, possible that by using a longer training dataset, the prediction of rare events such as extremely strong wind would be even better.

When using the A8 or the A2 as the starting model, five predictors are used: wind speed and direction logarithmically interpolated to 10-m height, air temperature at 2-m height, air pressure, and relative humidity. The DA does not include moist and radiation physics. Hence, only physical variables related to wind fields are included in the search for the best analogs: wind speed and direction logarithmically interpolated to 10-m height, and vorticity and divergence at the lowest vertical level (~17 m). The weight assigned to wind speed and direction is 1, and it is 0.8 for the all other predictor variables. The time window used to find the most similar analogs is defined by 1 time step before and after the lead

time of interest. For instance, in Eq. (4), $\tilde{t}$ is equal to 1, hence forming a 6-h time window for the A8 and the DA models, or a 2-h time window for the A2 model. The time window, the predictors, and the corresponding weights used to find the most similar analogs are the same for the KFAN and the KFAS as for the AN and the ANM. The same recursive algorithm is used for generating the KFAN and the KFAS as for the KF.

5. Observational analysis

The postprocessed forecasting methods described in section 2 are tested at 14 locations in Croatia, covering different climatological regions (cf. Fig. 2). Our goal is to compare and contrast the performance of the different methods, generated from different NWP models, and at different complex terrain and coastline sites. The locations are thereby divided into three groups:

1) Group I is a coastal complex terrain region that includes the locations near the coastline and near the western slopes of the Dinaric Alps. The prominent wind in this area is the bora, a strong and gusty downslope windstorm [e.g., see review by Grisogono and Belusic (2009)]. The bora wind is more frequent in the northern than in the southern Adriatic. Nevertheless, its maximal strength is similar in both regions (Horvath et al. 2009). Other mesoscale wind circulations are also notable and are governed by the surface inhomogeneity (e.g., land–sea breeze) and vicinity of the mountains (e.g., mountain–plain circulation, gap flows, weak downslope flows). Therefore, the diurnal cycle is shaped by the proximity of the sea and terrain elevation. The highest wind speeds analyzed in this study are recorded in this area (Fig. 3a), and the mean wind speed is 4.0 m s^{-1}.
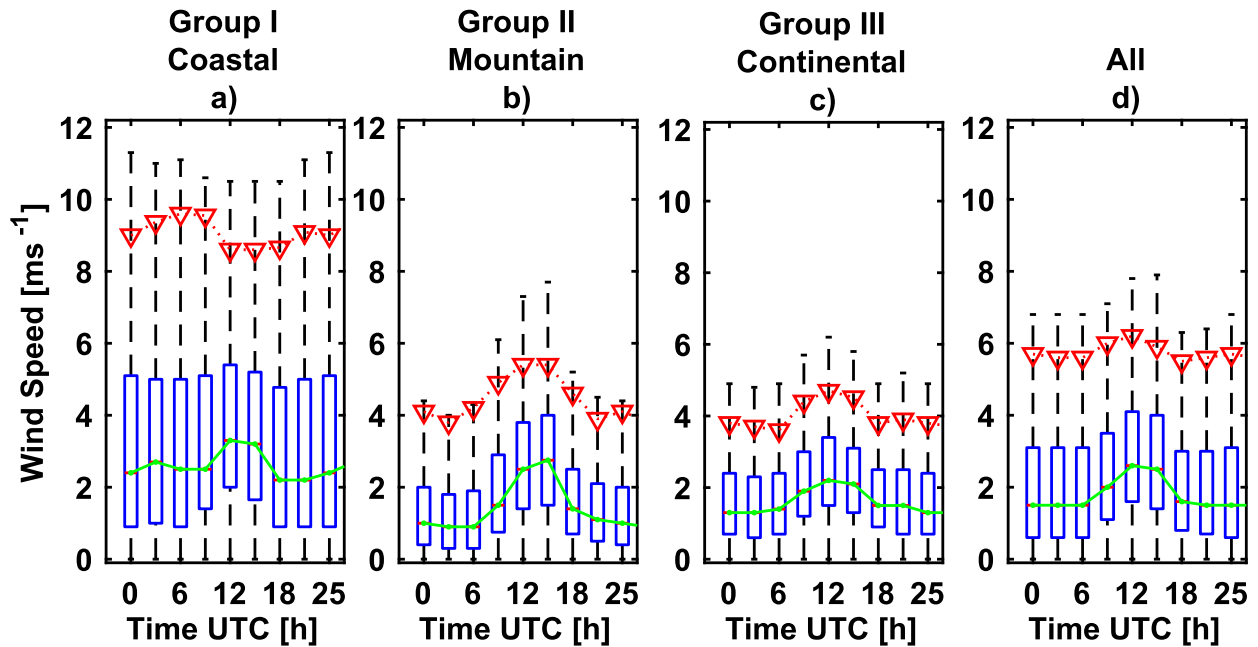
FIG. 3. The boxplots of the observed data (outliers are not shown), depending on time of the day. The data are measured during the 2010–12 period at 14 stations in Croatia. In addition to (d) the boxplot for all the data available, the data are sorted into (a)–(c) groups based on terrain type and basic climatological features. The green lines represent the 50th percentile and red triangle markers the 90th percentile. Those values are used as thresholds between categories in the verification procedure.

2) Group II is a mountain complex terrain region with highly complex topographical features. Locations in this area are farther from the coastline and at higher elevation than the locations in any other group, with mountain tops reaching 1500 m above mean sea level (MSL). Because of terrain complexity and low population density, the measurements are coarse in space in this area. The measurements may also be prone to longer data gaps due to the remoteness of locations and generally more severe winter climate. After our analysis, we therefore chose three locations that satisfy the basic quality requirements within this area (e.g., that there are no gaps longer than a few weeks). This area is characterized by a significant portion of energy variance due to mountain slope and valley winds. Wind speeds in the mountain complex terrain are lower than in the coastal complex terrain (Fig. 3b) and the mean wind speed is 2.0 m s^{-1}.

3) Group III stations are located in the nearly flat inland continental climatological region of Croatia. The terrain elevation is up to 100-m MSL. The diurnal cycle is shaped mainly by the gentle microscale variations of the topography. The region is still influenced by nonlocal effects of the Dinaric Alps mountain system to the west and southwest, since these mountains affect predominant westerly flow through channeling, blocking, and other mesoscale

processes. A strong wind is very rare in the continental area, and it occurs during the cold-air outbreaks from polar or Siberian areas in winter or during rough weather in summer (Zaninović et al. 2008). The wind speeds are relatively low (Fig. 3c), with a mean wind speed of 2.0 m s^{-1}.

Mean wind speed for all 14 stations is 2.7 m s^{-1}. The maximum of the diurnal cycle occurs around 1200 UTC on average for all stations (Fig. 3d). However, different processes contribute to the average daily cycle at different locations.

6. Evaluation of the deterministic forecasts

a. Wind speed as a continuous predictand

To evaluate the performance of the different deterministic postprocessing methods, wind speed can be considered as a continuous or categorical predictand. Considered as a continuous variable, wind speed forecast error is quantified by root-mean-square error (RMSE), which penalizes a larger discrepancy more than a smaller one. It is possible to specify the source of the error by decomposing the RMSE to the bias of the mean (or simply bias), bias of the standard deviation (STD), and dispersion (phase) error (e.g., Murphy 1988; Horvath et al. 2012). Since the sum of these three terms is exactly

the RMSE value, it is enough to provide information about two out of these three terms to describe the dominant source of the error (the third term is the RMSE value reduced by the value of the other two terms). The term describing the dispersion error involves the Pearson correlation coefficient weighted with the STD of both the forecasts and measurements. Correlation coefficient and dispersion error are thus closely related: the smaller the correlation coefficient, the larger the dispersion error term in RMSE decomposition. In this work, the rank correlation coefficient (RCC) is used as a robust and resistant alternative to Pearson correlation, appropriate if dealing with non-Gaussian-distributed variables such as wind speed. Unlike the Pearson correlation coefficient, the RCC is a nonparametric statistic. The RCC therefore allows a nonlinear relationship between predictions and observations (Wilks 2011; Jolliffe and Stephenson 2011).

The first step in testing an ensemble-based method is to select a number of ensemble members N. For that purpose we analyze the RMSE averaged over all locations and all lead times (Fig. 4a). The optimal N is presented and determined for the A8 starting model. The results for the A2 and the DA are not shown here because they are similar to the A8 results. Specific differences between the results generated with different models, however, will be discussed afterward. The mean confidence intervals shown here are estimated with bootstrapping. The starting model forecasts (A8) yield RMSE of $2.35\,\mathrm{m\,s}^{-1}$, RCC of 0.58, and almost nonexisting bias of $-0.01\,\mathrm{m\,s}^{-1}$. These values are determined by the wind climate, complexity of topography, and the low resolution of the driving mesoscale model. All tested postprocessing methods (PPMs) averaged over the three studied regions improve results of the A8 model. The KF significantly reduces RMSE (Fig. 4a) and improves correlation (Fig. 4b), while bias remains small (Fig. 4c) when compared to the A8. Using analogs improves results even further than using just the KF, as can be seen in the KFAS. This ABP uses the entire analog space and therefore does not depend on the N. The other ABPs (AN, ANM, and KFAN) produce similar results as the KFAS for about 10 or more ensemble members. Furthermore, the AN, the ANM, and the KFAN show similar behavior—the RMSE is reduced at first by increasing the N, but then it increases again for $N > 15$. By increasing the N, correlation also improves, while bias slightly worsens. The mean of the observed wind speed during the verification period differs from the mean during the training period for approximately $0.2\,\mathrm{m\,s}^{-1}$. The bias is likely converging to that value when increasing the N. Even though the PPMs' biases are significantly different than the bias for the A8, one should

take into consideration that the bias under $0.5\,\mathrm{m\,s}^{-1}$ can be considered relatively small. It is an order of magnitude smaller than the other two terms in RMSE decomposition and comparable to observational error (up to $0.5\,\mathrm{m\,s}^{-1}$ or even higher; WMO 2008). Given the RMSE and bias growth for the larger N, the optimal number of ensemble members is set to 15, which is used hereinafter.

It can be noticed that the ANM has the highest RMSE and the highest bias if different ABPs are compared. Since the other ABPs produce better results than the ANM, and specific benefits are not achieved in tested cases presented in this work, results for the ANM are not shown hereinafter. Both AN and KFAN considerably reduce the RMSE (as evident from Fig. 4a) better than any other technique tested here. At the same time they improve correlation (Fig. 4b). Both AN and KFAN have a very small negative bias, mostly between -0.1 and $-0.2\,\mathrm{m\,s}^{-1}$. The AN has slightly better correlation and worse bias results than the KFAN, resulting in indistinguishable RMSE.

Since the KFAN forecast is created by applying the KF to the AN forecast, the differences between the KFAN and the AN in the correlation and bias results may be expected. The KF algorithm updates its estimate of the future bias by using the old bias plus uncertainty. The estimate is corrected by a linear function of the difference between the previous prediction and the verifying bias. It is therefore very successful in removing the systematic errors (such as bias of the mean) if the bias does not change rapidly (i.e., large day-to-day variations). By reducing the noise, the application of the KF algorithm can also lead to the decrease of the correlation coefficient [i.e., increase of the dispersion error, as explained in Delle Monache et al. (2006, 2008)].

A more detailed insight into the performance of the PPMs can be gained by analyzing the metrics in topographically different regions and at different lead times. The A8 model has the highest RMSE for the coastal complex terrain among all groups of stations (Fig. 5a). Besides the increasing trend for longer lead times, the A8 RMSE error is typically the largest during nighttime and peaks at 0600 UTC at coastal areas. While during nighttime the A8 exhibits maximum correlation (Fig. 5e), it underestimates the mean (Fig. 5i) and the STD (Fig. 6a) more than during daytime. While the observed wind speed shows the highest variability at 0600 UTC (Fig. 6a), the A8 forecasts almost do not show STD diurnal cycle. That result suggests a systematic source of the errors for the diurnal shape of the A8 RMSE (Fig. 5a). It is possible that the A8 model underestimates the land breeze, the combination of a land
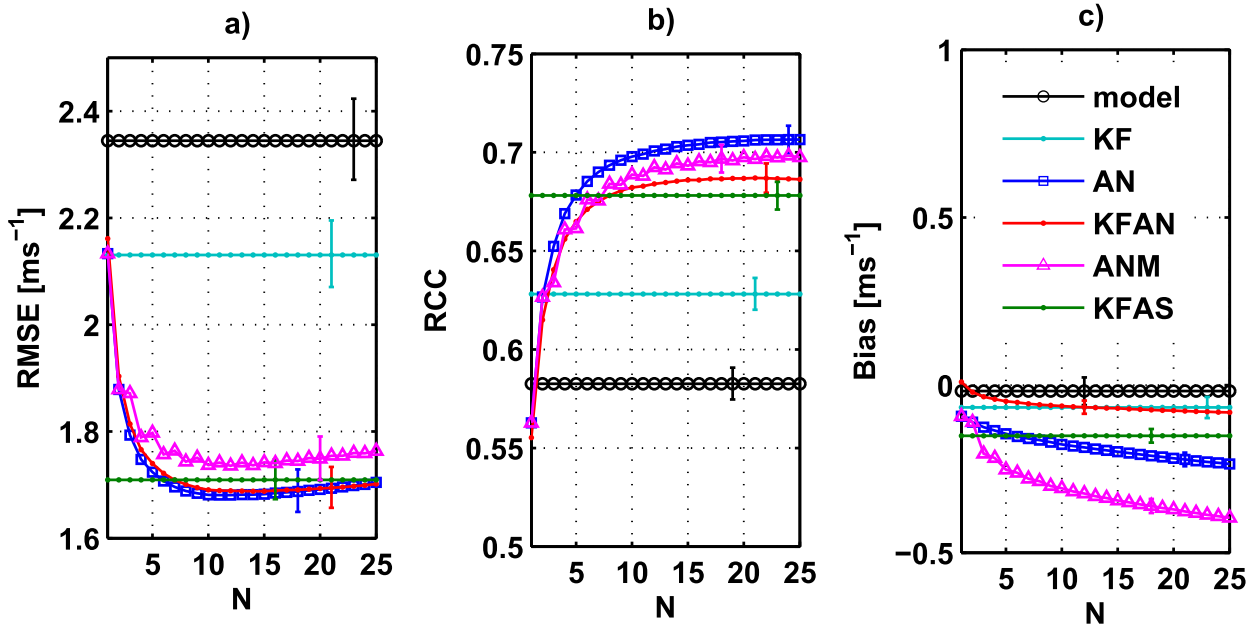
FIG. 4. (a) RMSE, (b) RCC, and (c) bias dependency on number of analog ensemble members N for the AN, the ANM, and the KFAN ABPs. The results are generated with the A8 and averaged over all lead times and all locations during 2012. The AN, ANM, and KFAN results are then compared to the A8 model, the KF, and the KFAS forecasts, which do not depend on N. The mean values of the 95% bootstrap confidence intervals are indicated by the error bars.

breeze and downslope wind called burin (Poje 1995), or underestimates both mean speed and variability of the strong bora wind, which can be determined with an analysis by season (e.g., bora occurs mostly during wintertime and is variable and intense, while the land breeze can be dominant during summertime stable conditions) or by examining case studies, but that is out of the scope of this work.

It is crucial to determine which PPM is the most successful in the error reduction, especially in the coastal group of the stations where the error is the largest. Additionally, it is important to demonstrate which term of the RMSE decomposition is reduced by which PPM. The results are presented in such manner that one can thus decide which PPM is the most applicable for a specific situation after a simple statistical analysis of the potential starting model.

The KF reduces RMSE and bias (Fig. 5i) while increasing STD (Fig. 6a), maintaining very similar dependency on lead time as the A8. The other ABPs (AN, KFAN, and KFAS) improve the A8 results even further—reducing RMSE and bias while STD is even closer to the STD of the measurements. Moreover, even though the STD is still a bit underestimated, the diurnal cycle of STD is more similar to the diurnal cycle of the measurements than for the A8. Previously mentioned systematic A8 error (possibly unresolved land breeze, underestimation of

burin wind, etc.) is thus reduced or removed completely. The STDs of the ABPs (AN, KFAN, and KFAS) are very close to the STD of the measurements available over the training period. The AN, the KFAN, and the KFAS underestimation of the STD is therefore partially explained by the fact that there is an STD difference between training and testing period.

The AN forecast is the most prone to systematic underestimation of the STD. This reduction of the forecast variability is due to averaging of AnEn members while predicting the mean of the ensemble. This averaging naturally reduces the variability and might partially be improved by using the lower number of ensemble members N. This systematic error is partially removed by the application of the KF algorithm in the KFAN forecast. The KFAS forecast exhibits the highest STD among the ABPs. The simplified schematic example for improving the adaptability of the KFAS forecast is provided in Delle Monache et al. (2011). The hypothesis is that applying the KF algorithm in analog space (rather than in time) results in higher forecast variability during alternating wind regimes. The higher KFAS STD than the KFAN STD in the coastal area supports this hypothesis. The difference in the STD between the KFAN and the KFAS does not necessarily mean that the higher variability for the KFAS is occurring during alternating wind regimes (i.e., on the time scales shorter than a day).
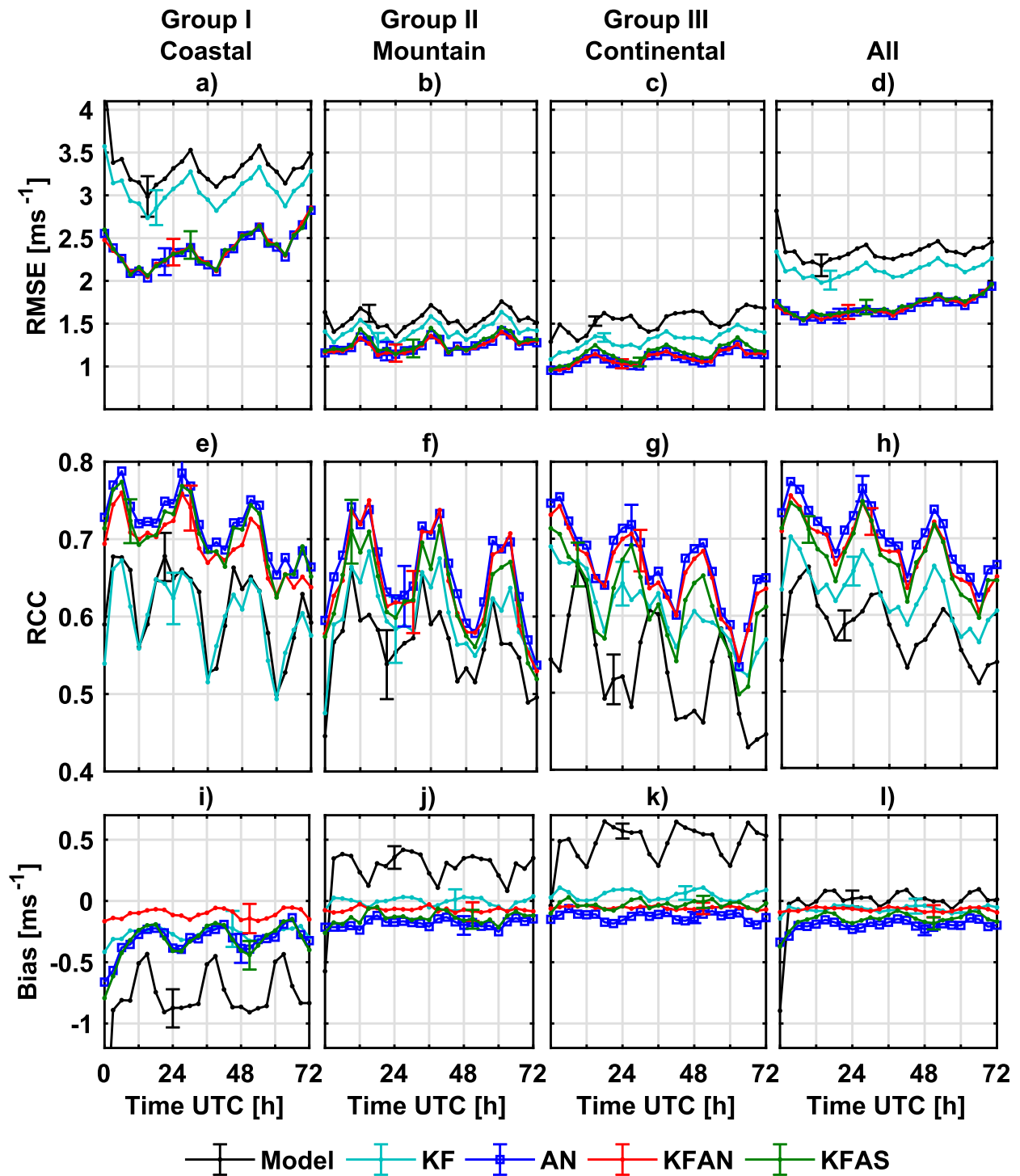
FIG. 5. (a)–(d) RMSE, (e)–(h) RCC, and (i)–(j) bias dependency on forecast lead time for the A8 starting model and the corresponding PPMs (KF, AN, KFAN, and KFAS). The results are averaged over corresponding group and for all locations during 2012. The mean values of the 95% bootstrap confidence intervals are indicated by the error bars.
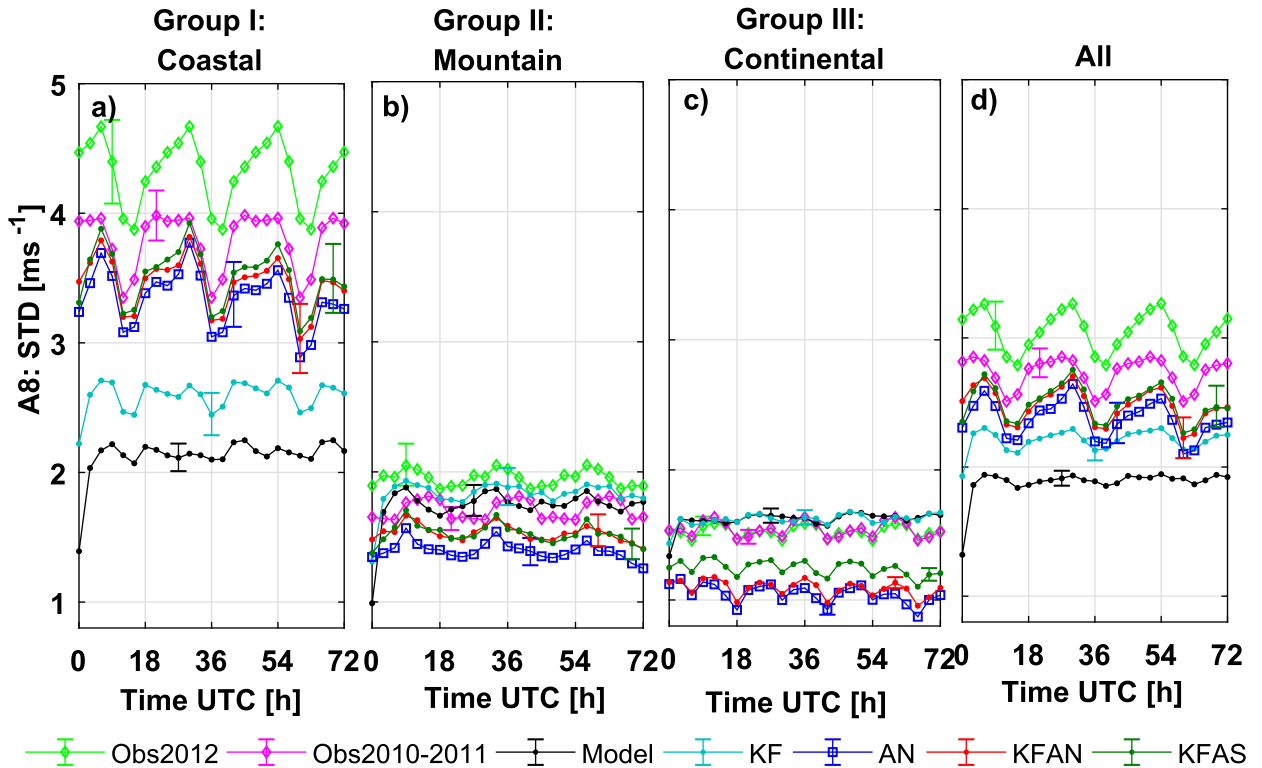
FIG. 6. The dependency of the standard deviation on forecast lead time for the observations during the training (2010–11) and the verification period (2012), the A8 starting model (2012), and the corresponding PPMs (KF, AN, KFAN, and KFAS). The results correspond to the (a)–(c) three groups and to (d) all locations. The mean values of the 95% bootstrap confidence intervals are indicated by the error bars.

The remaining underestimation of STD depends on other aspects such as the variability of starting model forecasts and fine tuning of the analog search setup [e.g., choice of predictors and corresponding weights, as shown by Junk et al. (2015)]. The variability in the training period might be enlarged by prolonging the period itself (i.e., including El Niño–Southern Oscillation variations). Finally, the variability of the PPMs' forecasts might be further improved by additional calibration, but that is beyond the scope of this paper.

The KF has smaller RCC than the A8, unlike all the ABPs that have higher RCC than the A8. Improving the RCC shows that by using analogs and measurements to build a prediction the random error is reduced, suggesting that additional information on physical processes is included in the ABPs. Among different ABPs, the AN seems to have the highest correlation, while the KFAN reduces the bias the most, as previously described in the more general case. The KFAS exhibits the highest STD among the ABPs, supporting the hypothesis that using the analog space improves variability during alternating wind regimes. After all, there are no significant differences in reduction of RMSE for the AN, the KFAN, and the KFAS.

The A8 exhibits considerably smaller RMSE for the mountain complex (Fig. 5b) and nearly flat terrain (Fig. 5c) than is the case for the coastal complex terrain area. The smaller A8 RMSE is predominantly due to lower, less underestimated STD of measured wind speed for these groups (Figs. 6b,c) than for the coastal complex terrain (Fig. 6a). Even though the A8 error is smaller than in the coastal complex terrain, it is still very important to determine which term in the RMSE decomposition is dominant and how it can be reduced by postprocessing. Instead of underestimation of (on average) higher wind speed in the coastal terrain, the A8 overestimates (on average) lower wind speed (Figs. 5j,k) in the mountain complex and the nearly flat terrain, exhibiting similar absolute value of the bias. The A8 STD is much closer to measured wind speed STD for the mountain complex (Fig. 6b) and the nearly flat (Fig. 6c) than the coastal complex terrain. The A8 RCC is lower for the mountain (Fig. 5f) and for the nearly flat terrain (Fig. 5g) than for the coastal complex terrain, decreasing with measured wind speed and corresponding STD. It seems that the lower the average wind speed for a certain group, the lower the correlation of measurements and predictions, implying that weak wind is less

predictable than strong wind. This especially makes sense for wind speeds that are comparable to the observational error (up to $0.5 \, \mathrm{m \, s^{-1}}$ or even higher; WMO 2008). Even though some statistical properties of the A8 predictions are similar for the mountain and nearly flat terrain, the physical processes influencing the flows are different. This is due to different dominant topographic characteristics, as explained in section 5. For this reason it is interesting to compare the effect of PPM in a certain group of stations.

The KF exhibits significantly lower RMSE than the A8 in the mountain and nearly flat terrain. The A8 bias is almost completely removed by the KF, regardless if the A8 is underestimating (Fig. 5i) or overestimating (Figs. 5j,k) wind speed. The KF STD in the mountain and the nearly flat terrain is almost the same as the A8, and very close to measured STD as well. In addition to reducing the A8 bias of the mean and maintaining the bias of the STD (which is almost nonexistent), the KF also improves the RCC for all lead times in the mountain and the nearly flat terrain. Unlike for the coastal complex terrain, dispersion error is therefore reduced by the KF, especially for the nearly flat terrain. Furthermore, the KF dependency on lead time is different than for the A8 in the nearly flat terrain. The KF exhibits a local RCC maximum around 0000 UTC, while the A8 exhibits a local minimum (Fig. 5g).

The ABPs (AN, KFAN, and KFAS) in the mountain complex terrain and the nearly flat terrain reduce the A8 RMSE even better than the KF, further improving RCC and reducing bias. The RMSE, RCC, and bias dependencies on lead time are similarly shaped as for the KF. This is especially interesting in the nearly flat terrain, where previously mentioned improvement of the A8 RCC is even more indicated when using analogs than for the KF. The analog approach selects similar numerical predictions (not necessarily recent) for assessment of the starting model error, unlike nonselectively using previously predicted (recent) values in the KF algorithm. The KF would be capable of improving persistent error in predicting stable boundary layer flow once it is started, as previously mentioned for the application of the KF algorithm. Analog approach would be more adaptable and capable of predicting the beginning of the flow, thus resulting in even higher RCC than for the KF. Similarly to the coastal complex terrain, in the mountain complex and the nearly flat terrain the AN seems to be the most highly correlated with measurements. The KFAN has slightly lower RCC, but is almost unbiased. Unlike the A8 and the KF, the ABPs exhibit slight underestimation of STD in the mountain complex (Fig. 6b) and nearly flat terrain (Fig. 6c). The underestimation of the STD is the smallest for the

KFAS and the largest for the AN, for the same reasons as previously mentioned. The results for the KFAN are mostly in between these two (AN and KFAS), which may be explained by the fact the KFAN shares important features with both methods.

Overall, the A8 RMSE is significantly reduced by every PPM tested for all lead times, more by the ABPs (AN, KFAN, and KFAS) than for the KF (Fig. 5d). All PPMs reduced the A8 bias, which is evident for specific group and lead time (Figs. 5i,j,k), even though it seems nonexistent on average for the A8 (Fig. 5l). The KFAN predictions seem to be the most successful in removing bias, while the AN appears to exhibit the highest correlation (Fig. 5h). Measured wind speed STD is underestimated on average by the A8 model and all PPMs (Fig. 6d), mostly because of the STD underestimation in the coastal area (group I). Overall, the KFAS STD is the closest to the observed value.

To investigate the influence of the starting model used to generate analogs, results are averaged over all lead times for every group of stations. A reasonable hypothesis could be that the more physical processes that are directly simulated in the starting model (e.g., with higher resolution), the better the forecast will be. The RMSE (Fig. 7a) and bias (Fig. 7i) are lower for the A2 and the DA models than for the A8 in the coastal complex terrain, empirically supporting this hypothesis. The RCC does not differ significantly among different models (Fig. 7e). It must be noted that it is difficult to quantify the improvement of more detailed forecasts over coarser ones using point-based verification metrics (Rossa et al. 2008; Jolliffe and Stephenson 2011). Point-based verification metrics tend to penalize spatial and phase errors, contaminating finer-resolution simulations more than coarser ones. Hence, it might be challenging to easily demonstrate the true benefits of using higher-resolution forecast. To determine if that is the case, it would be advisable to do case studies and some sort of special verification (for gridded forecasts). That is, however, beyond the scope of this paper.

All PPMs tested in this study improve model predictions. This type of improvement is clearly evident, for example, for the AN results in the coastal complex terrain. The results show that the differences for using different starting model configurations are much smaller after postprocessing than for the three starting models. It is to be expected that the ABPs (AN, KFAN, and KFAS) also achieve better results when using the A2 or the DA than when using the A8. The quality of an analog should be better the more physical processes are simulated in the starting model (i.e., with higher-resolution, nonhydrostatic dynamics in the A2, etc.). However, the RMSE (Figs. 7a–d), RCC (Figs. 7e–h),

and bias scores (Figs. 7i–l) are similar for the PPMs applied to all three starting models. For some scores, such as the RMSE, the ABPs have the best results when applied to the A8 model.

Overall, it seems that even though the higher-resolution A2 and DA models achieve better results, the ABPs based on the A2 and the DA do not statistically outperform the ABPs based on the A8 (Figs. 7d,h,l). This does not necessarily mean that improvements are not made at all. The benefits might be partially hidden because of the imperfections of the verification metrics used. To investigate the benefits of using higher resolution further, one can analyze the forecasts categorically (i.e., to examine the forecasts of the rare events such as strong wind), perform a spectral analysis, or look at the case study. It is decided that the case studies are beyond the scope of this paper, while the categorical verification results and spectral analysis are presented in the next sections.

b. Wind speed as a categorical predictand

Wind speeds are divided into three categories: weak (or no wind at all), moderate, and strong wind, depending on the climatology of the corresponding group of stations. For each lead time, thresholds are determined as the 50th and 90th percentile of the entire group, so they vary according to diurnal cycle (Fig. 3). The categorical verification procedure includes frequency bias (Fbias), critical success index (CSI), and polychoric correlation coefficient (PCC). The choice of these measures is consistent with the continuous case. The CSI measures the fraction of observed forecast events that are correctly predicted. It can be thought of as the relative accuracy when correct negatives are removed from consideration. The CSI therefore measures the error (similar to the RMSE in continuous case). Sensitive to hits, the CSI penalizes both misses and false alarms. It does not distinguish the source of forecast errors and hence additional verification measures need to be examined (Wilks 2011; Jolliffe and Stephenson 2011). The Fbias, similarly to bias, measures the tendency to forecast too often (Fbias greater than 1) or too rarely (Fbias less than 1) a particular category (Wilks 2011; Jolliffe and Stephenson 2011). The PCC measures association of forecasts and observations in the contingency table. The idea behind the PCC is to assign a density function to the contingency table, and then cut the domain into rectangles corresponding to the cells of the contingency table. The PCC is the parameter value of the bivariate normal density function for which the volumes of the discretized distribution are equal to the corresponding joint probabilities of the contingency table (Juras and Pasarić 2006). Ekström (2011) shows the (asymptotical) equivalence of the RCC and the PCC

under several conditions including that the number of categories is as large as the number of measurement–forecast pairs, the underlying joint distribution is binormal, and so on. Even though a "simplified" RCC can be recalculated if the ordinal variables arise from discretization, such as groupings of values into categories (as in this section), it has some undesirable properties. For instance, it can achieve a value of 1 even if nondiscretized empirical variables are not perfectly dependent. The PCC is therefore considered to be more conservative and better suited for statistical inference concerning the association of the underlying, nondiscretized variables than the RCC.

The PCC results for different forecasts (Figs. 8a–d) do resemble the RCC results (Figs. 7e–h) when results are averaged for all lead times in a certain group. The DA and the A2 exhibit a similar association as the A8, regardless of the terrain, as previously discussed. Association is significantly improved by almost all PPMs in all groups of stations and overall, as already presented. The ABPs achieve better RCC and PCC results than the KF in general, particularly the AN. There are some differences between the RCC and PCC results that need to be highlighted in order to determine the origin—whether it is due to statistical properties of the verification measure used or it is a direct consequence of discretization (i.e., grouping of wind speed into three categories). If both RCC and PCC are calculated for the same (ordered) data and grouped into identical categories, the RCC would have a slightly higher value (Ekström 2011). The PCC shows higher values than the RCC calculated for the continuous variable, hence confirming the assumption that it is easier to predict the category than the exact (continuous) value of wind speed.

There are a variety of Fbias results depending on the exact model, group of stations, and wind category. For instance, the DA predicts category 2 too often (Fig. 9e), while predicting the other two categories (Figs. 9a,i) too rarely in the coastal area. The Fbias results for the A8 model are somewhat similar, while the A2 is almost unbiased in this case. All starting models underforecast weak wind (Figs. 9b,c) while overforecasting moderate and strong wind in the mountain complex and the nearly flat terrain (Figs. 9f,g,j,k). The exact values differ for different models and categories, yielding mixed results in terms of determining the best-performing starting model. The KF only slightly impacts the A8 Fbias by decreasing the value for the weak wind (Fig. 9a), while only indicating the increased value for the moderate and the strong wind (Figs. 9e,i) in the coastal area. More generally, besides the Fbias reduction for the weak winds (Figs. 9a–d), the KF does not have a noticeable impact on the starting model results. Unlike the coastal area, in the
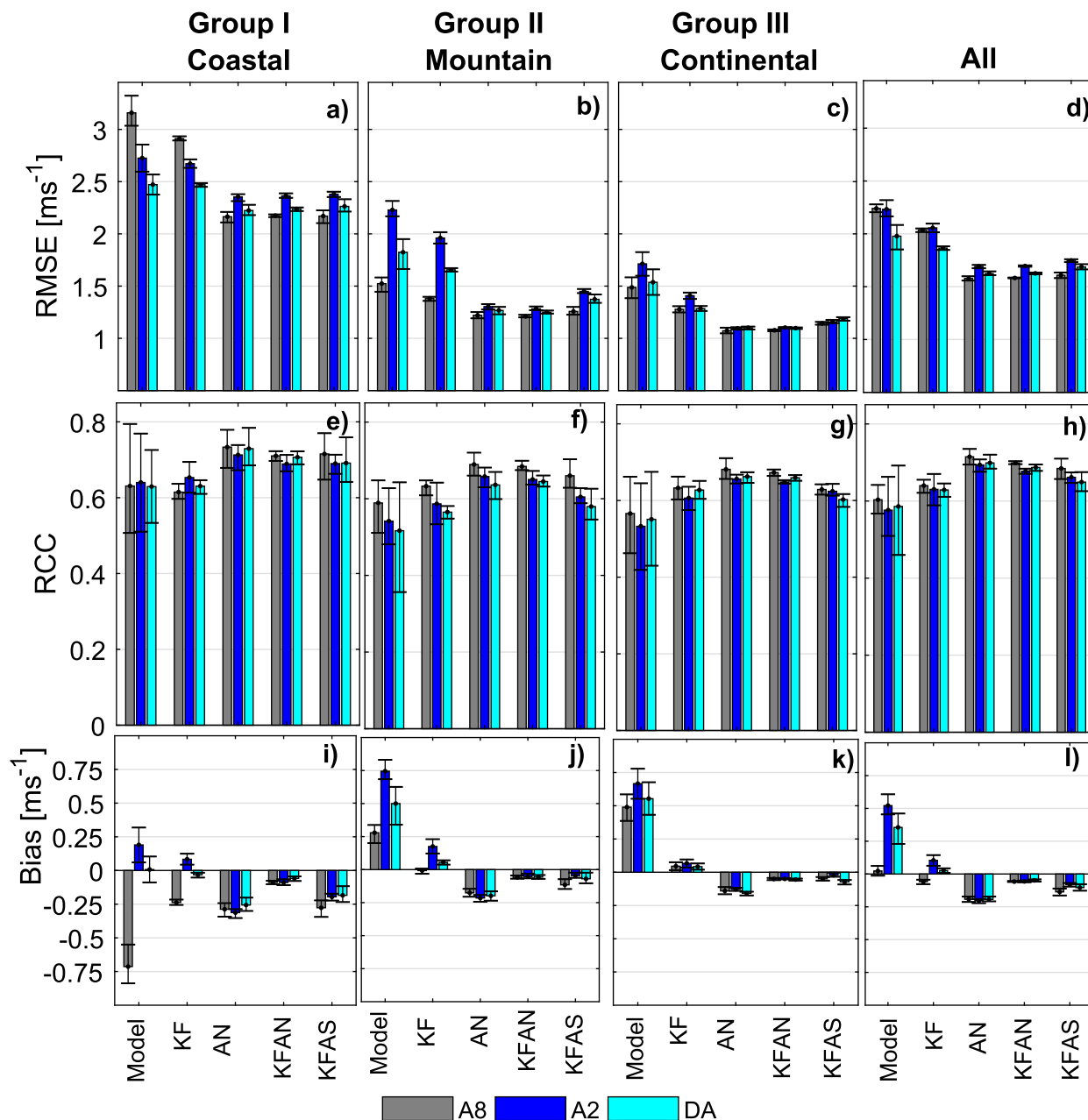
FIG. 7. (a)–(d) The average RMSE, (e)–(h) RCC, and (i)–(l) bias for three different starting models and the corresponding PPMs (KF, AN, KFAN, and KFAS). The results are averaged over three groups and over all available locations during 2012. The colors represent the starting model used (A8, A2, and DA), while the *x* tick labels stand for the forecasting method (model and corresponding PPMs). The values of the 95% bootstrap confidence intervals are indicated by the error bars.

mountain complex and the nearly flat terrain the KF seems to be less biased than the corresponding (starting) model for all cases tested. This is indicated by the significantly smaller bias for the weak winds, and smaller confidence intervals near the zero value for the moderate and strong winds. The smaller confidence interval referring to the same sample size means smaller variability within the results. The Fbias results for the ABPs (AN,

KFAN, and KFAS) seem to exhibit much less variety depending on different group of stations. The results are indistinguishable among different starting models, especially for the moderate and strong winds (Figs. 9e–l). For any given group, the ABPs consistently overpredict moderate wind speeds, while underpredicting rarer and stronger wind. These ABPs sometimes even underpredict the occurrence of weak wind. The KFAS seems to be the
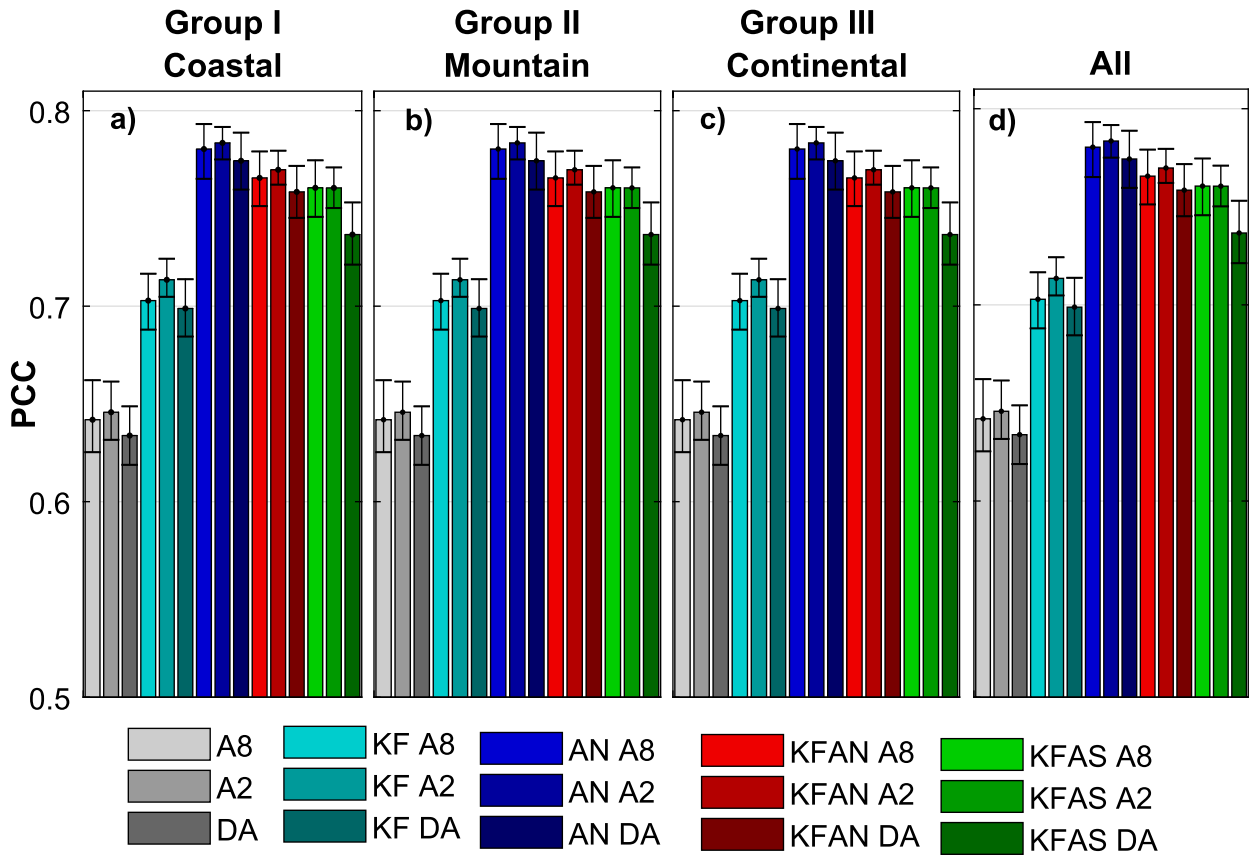
FIG. 8. The PCC for three different starting models (A8, A2, and DA) and the corresponding PPMs (KF, AN, KFAN, and KFAS). The results are averaged for (a)–(c) three groups of stations and for (d) all locations during year 2012. The PCC is calculated using three different categories, divided by the 50th and 90th percentiles. The values of the 95% bootstrap confidence intervals are indicated by the error bars.

least biased ABP, showing the highest values for strong wind, while being as unbiased as the AN in the other two categories. However, it needs to be mentioned that these differences are not statistically significant, partially because of the small sampling size. Overall, the PPMs' forecasts reduce bias for climatologically common wind speed category (weak wind). The ABPs' Fbias results are not as variable as for the starting model and the KF, inheriting only a slight difference from the corresponding model for an exact method (AN, KFAN, or KFAS). The main deficiency of the PPMs seems to be underforecasting the occurrence of strong wind, with the KFAS being the most successful (Fig. 9l).

If results among different starting models are compared, it can be seen that for the weak wind the A2 produces higher CSI than the A8 and the DA in the coastal complex terrain (Fig. 10a). Furthermore, finer horizontal resolution slightly improves relative accuracy for strong wind (Fig. 10i). The results for moderate wind are similar across the different starting models (Fig. 10e). Increasing the horizontal resolution does not

necessarily improve CSI in the other groups of stations. Because of small sample size, the results rarely differ significantly (Figs. 10b,c,f,g,j,k). The CSI results are considerably higher for the KF than for the starting models (A8, A2, and DA) for the weak wind in the mountain complex and the nearly flat terrain, but not as much in the coastal area. The indication is that the KF is the most successful in predicting the strong winds in nearly flat continental terrain, and even though it is not statistically significant, it might still suggest a dominant systematic error in the models' predictions of the strong wind. The Fbias is lower for the KF than for any starting model, which, combined with higher CSI, indicates that the number of false alarms is reduced. Analysis suggests that ABPs outperform starting models and corresponding KF forecasts for all categories and all groups of stations except the strong winds in the nearly flat continental terrain (Fig. 10k). The improvement of CSI value is the most evident, and statistically significant, for the most common weak wind (Figs. 10a–d). However, the larger sample is needed to prove such a
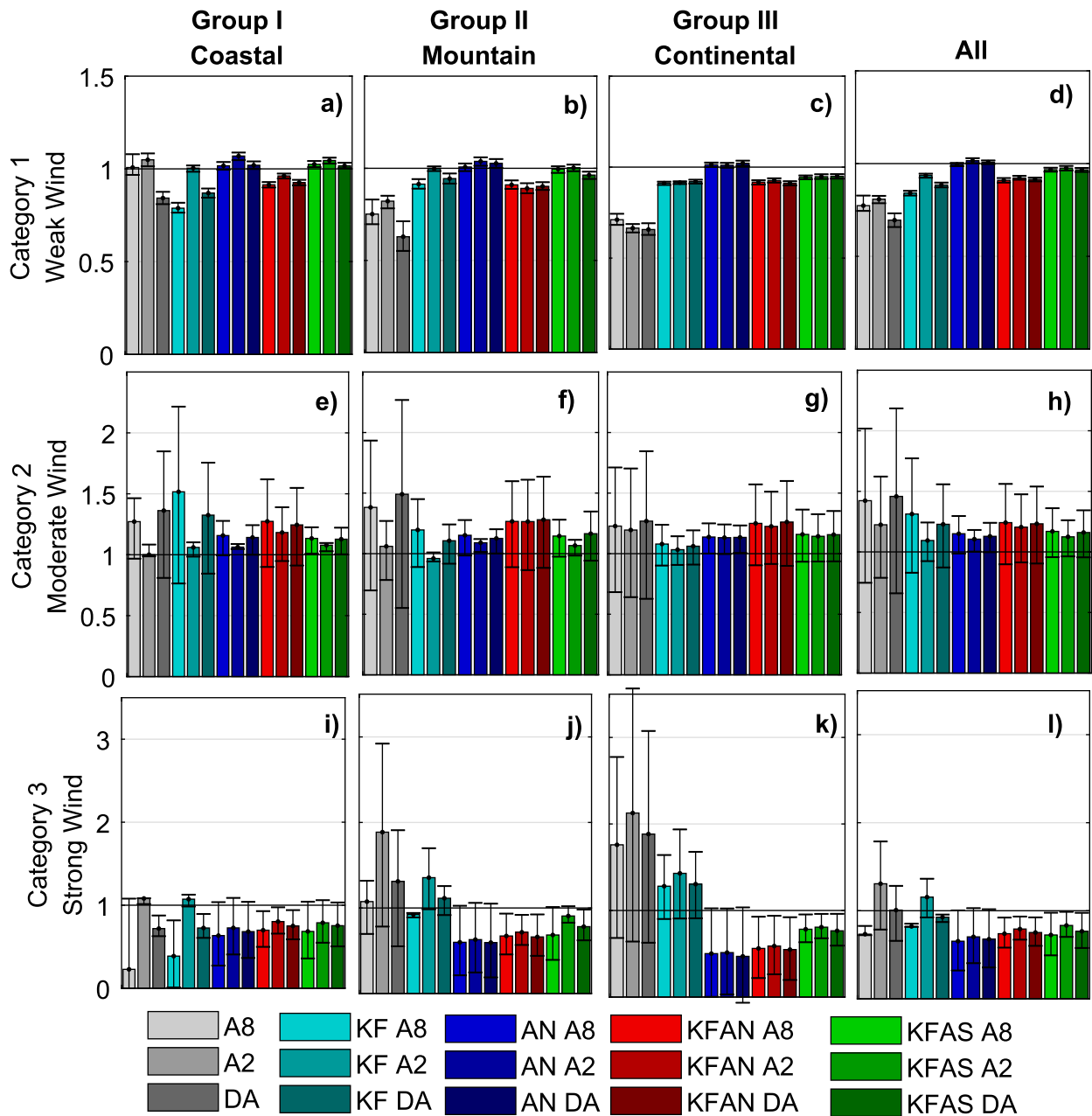
FIG. 9. As in Fig. 8, but for Fbias.

statement for the moderate and strong wind. Overall, all PPMs improve the CSI value. The AN forecasts achieve the best result for predicting weak wind (Fig. 10d), while the KFAN and the KFAS produce slightly better results than the KF and the AN for the other two categories (Figs. 10h,l). It needs to be noted that the results for the moderate and strong wind are rarely statistically significant, partially because of small sample size. However, analysis suggests that the best results are achieved when using the A2 as the starting model, mostly because of the

higher CSI in the coastal complex terrain than when using a coarser-resolution starting model. It is possible that additional improvements may be generated by increasing the resolution (1 km or less) in the complex terrain. The necessity to use even 2-km grid spacing is, however, questionable and might be reexamined for nearly flat continental terrain (i.e., by spectral analysis). In addition to improving the relative accuracy in coastal complex terrain, the categorization suggests the higher association for the full-physics A2 model and
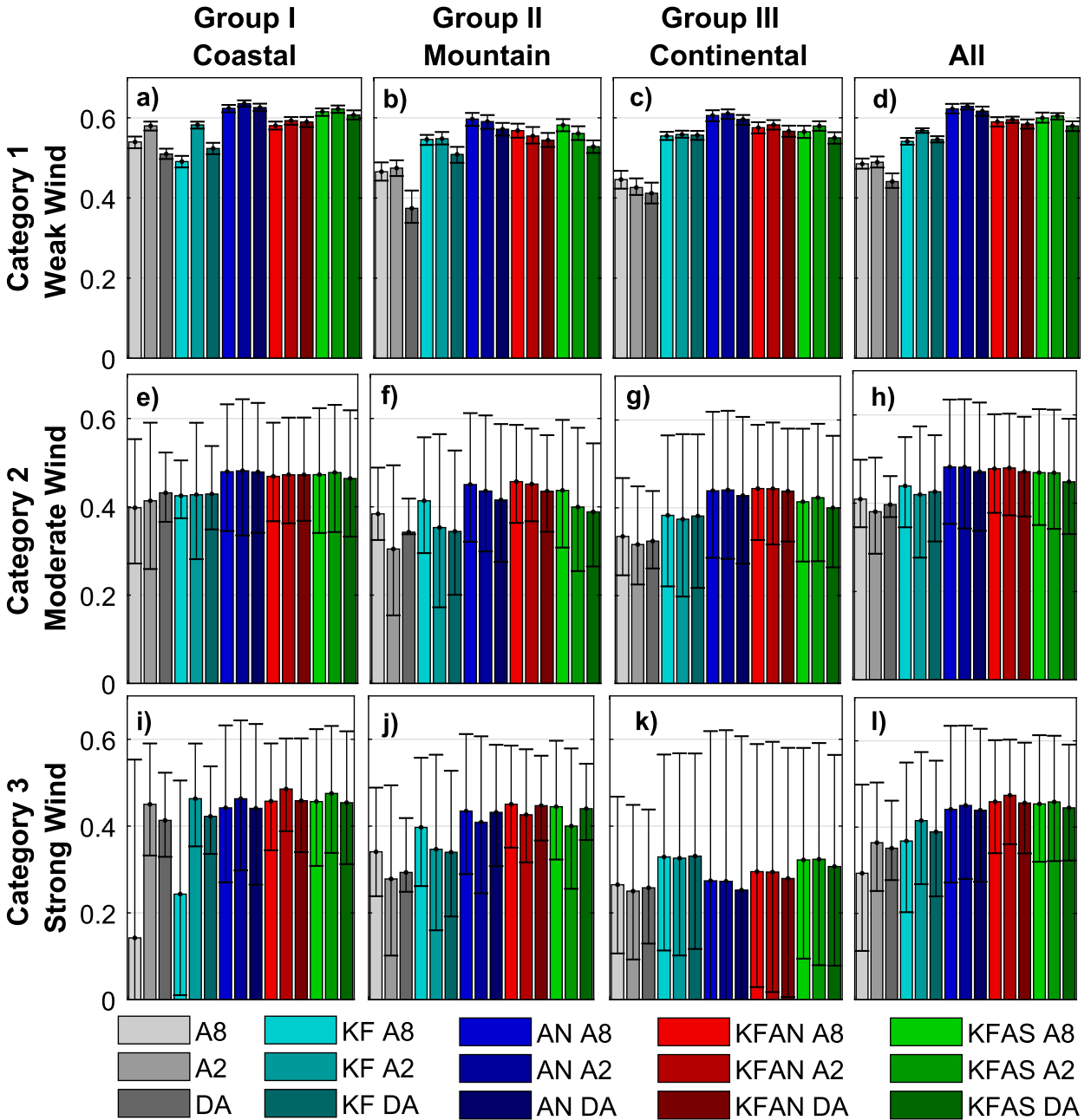
FIG. 10. As in Fig. 8, but for CSI.

corresponding PPMs in the coastal complex and the nearly flat continental terrain, as shown before. These results combined might suggest that the higher-resolution full-physics A2 model is better capable in distinguishing low from moderate or unusually strong wind, especially in the coastal complex terrain. This capability is then mostly inherited by the different PPMs, including the ABPs.

There is a decrease of the CSI values for moderate and, in particular, strong wind, regardless of the exact group of the stations or the forecast. It should be mentioned that that decrease is partially the direct consequence of sensitivity of the CSI metrics to climatological probability of the predefined category that is being evaluated, and therefore it should be analyzed with caution. The sensitivity to climatology (or base rate) is due to counting the portion of correct forecasts that can be accurately predicted by random chance. Also, the different values across different groups for the same category (e.g., strong winds in Figs. 10i,j,k) might mean

that unusually strong and rare wind is predicted more easily in coastal than in continental areas, regardless of the exact forecast.

c. Spectral analysis of wind speed

The small spatial and temporal errors of (generally) well-simulated phenomena can profoundly change the verification results (Mass et al. 2002; Rife et al. 2004). For that reason, spectral analysis in the frequency domain is utilized to provide a scale-dependent measure of different PPMs' performance. Spectral analysis allows quantification of power distribution among different temporal scales. It is relevant to determine the exposure of particular stations to longer-than-diurnal (LTD), diurnal (DIU), and shorter-than-diurnal (SD) motions and the forecast ability to simulate these motions.

Spectral decomposition of detrended time series is performed using the Welch periodogram-based method (Welch 1967) with 50% overlapping segments. The length of the Hamming spectral window ($L = 256$) is adjusted to optimally emphasize the difference among tested PPMs. Missing data are provided by using linear interpolation. It should be noted that power spectral density (PSD) analysis performed contains the effect of aliasing, necessarily contaminating all scales by oscillations with periods shorter than 6 h (here corresponding to the Nyquist frequency). Testing this effect on measured data suggests that it is rather small on LTD scales. Significant effects may be found on SD scales, especially near the periods corresponding to the Nyquist frequency (Žagar et al. 2006; Hrastinski et al. 2015). Since the A8 and the DA forecasts are archived every 3 h (the A2 and the measurements are adjusted to the same frequency), it is not possible to circumvent this effect. However, it may be noted that all the forecasts tested (and measurements) are aliased in a similar manner; therefore, the effect is not crucial for the intercomparison.

The PSD of 10-m wind speed is calculated for all forecasts (all starting models and all corresponding PPMs) at all locations for the entire year of 2012. The forecast frequency is $(3\,\mathrm{h})^{-1}$ for all forecasts, and only a 24-h forecast period is considered. It should be noted that the typical diurnal rotation of winds in the Adriatic partially hides the diurnal spectral peak if the analysis is performed using wind speed values (Telišman Prtenjak and Grisogono 2007). However, the preferred spectral analysis of wind components is not possible, as the ABPs predict only the wind speed (not the direction). The spectral analysis is performed for all forecasts and locations included in this paper. However, it is decided that it is more comprehensive to show the results for several representative locations, instead of any sort of averaging or summarizing the results. The particularities

that could not be easily seen in figures are pointed out and explained in the text. Two locations (Dubrovnik and Jasenice stations) correspond to the coastal group of stations, covering the northern and the southern parts of the coastline. The reason for including these two stations is that the governing processes somewhat differ [e.g., processes that lead to bora windstorms as explained in Horvath et al. (2009)]. The representative station for nearly flat continental terrain is Osijek station, while Ogulin represents the mountain complex terrain.

The PSD functions for the A8 and corresponding PPMs' wind speed are shown in Fig. 11. The KF approach influences the motions on the time scales longer than 10 days if the model's PSD function is biased. The KF therefore enlarges the energy of these large-scale motions in the coastal area and reduces the energy at the nearly flat continental terrain. Besides the large-scale motions, the KF does not significantly influence the shorter time scale. Similarly, the KFAN is almost exactly the same as the AN spectra, except with rarely significant differences for large-scale motions. Very small differences among spectra before and after application of the KF algorithm might mean that the ratio of the variances used in the algorithm is not optimal. If the ratio of the variances is set too high, the filter puts excessive confidence on the past forecasts, therefore failing to remove any error. On the other hand, if the ratio is too low, the filter will be unable to respond to changes in bias (Delle Monache et al. 2006). The sensitivity of these results to changing the ratio of the variances used in the algorithm therefore might be tested in the future work. The KF spectra is the same as model spectra and the KFAN spectra is the same as the AN spectra for the scales shorter than 10 days. The same conclusions regarding spectral analysis are therefore valid hereinafter. Hence, the analysis for the KF or the KFAN will not be explicitly mentioned.

It can easily be seen that the largest portion of measured power at all stations is associated with the LTD motions. These LTD motions are more energetic for the coastal area (Jasenice and Dubrovnik stations, Figs. 11a,d) than for the mountain complex (Fig. 11g) and nearly flat continental terrain (Fig. 11j). As shown by several other authors, this is related to the strong and gusty bora wind (e.g., Horvath et al. 2009, 2011; Hrastinski et al. 2015). The LTD motions are severely underestimated with the A8 model in the coastal area (Figs. 11a,d). The LTD motions in the AN and the KFAS data contain more energy compared to the model PSD, therefore improving the model. This shows a great potential for the ABPs to improve the model forecast when there is a model underestimation of LTD motion, even in the complex terrain. In the nearly flat terrain, the A8 model

simulates well, or overestimates, the energy of LTD motions (Fig. 11j). The ABPs (AN, KFAN, and KFAS) lower the energy of LTD motions if they are well simulated or overestimated by the model. This sometimes leads to underestimation of LTD motions, especially for the AN. The KFAS exhibits LTD PSD spectra very similar to measurements. Thus, the KFAS shows the greatest potential for the forecast improvement if the model overestimates the energy of LTD motions in the nearly flat terrain. The A8 results for LTD motions in the mountain complex consist of all previously mentioned scenarios, depending on the location and the exact time scale. For instance, it is well simulated for periods longer than three days and underestimated for shorter time scales at Ogulin station (Fig. 11g). The ABPs act similarly to previous types of terrain, exhibiting more energy if they are underestimated by the A8 model, or less if they are not.

The SD motions are severely underestimated by the A8 model for the majority of locations, regardless of the terrain (e.g., Horvath et al. 2011). Only at a few stations (e.g., Osijek in Fig. 11j) is the amount of energy at these scales comparable to measured values. The AN forecast is, once again, the most prone to energy underestimation. The SD KFAS spectra, on the other hand, seem very similar to model spectra. Moreover, it seems that the KFAS exhibits PSD values similar to the AN and observations for longer time scales, but it is similar to model values at shorter time scales. Finally, it is interesting to note that even though energy of the SD motions is underestimated, the harmonics of the diurnal cycle (24-, 12-, and 8-h periods) are very well simulated by the A8 model and all of the PPMs.

Introducing the higher-resolution orography affects the dynamical processes and increases the amount of energy at all temporal scales (e.g., Žagar et al. 2006). Therefore, the difference between the A8 and the DA is that there is almost no underestimation of the LTD motions, even in the coastal complex area (e.g., Fig. 11c). The exception is Dubrovnik station (Fig. 11f), which is very similar as for the A8 model (Fig. 11d). The energy simulated by the DA is higher at the mountain complex station (Fig. 11i) than when simulated with the A8 (Fig. 11g), overestimating the LTD motions. Introducing the higher-resolution orography into nearly flat continental terrain results in very similar PSD curves for the DA, as is the case for the A8 (e.g., cf. Figs 11l and 11j). This is to be expected because the flatter the terrain, the smaller the number of details added by increasing horizontal resolution. In the mountain complex terrain (e.g., Ogulin station), results may be improved by using even finer model resolution to represent local flows. However, the need for using 2-km as opposed to

8-km grid spacing for weak wind in the nearly flat continental terrain (e.g., Osijek station) may be questioned. Naturally, the PPMs are also exhibiting a similar effect, as is the case for the A8 model.

Similarly, introducing higher-resolution fields into the A2 forecasts increases the power at all time scales. All the conclusions regarding PSD spectra that are valid for the DA LTD motions are valid for the A2 model as well. Additionally, because of a more complete package of physics parameterizations and nonhydrostatic effects, the A2 model SD part of PSD spectra contains more energy than for the A8 and the DA models. Both the A8 and the DA models severely underestimate the power at scales below diurnal, as reported by Žagar et al. (2006). Unlike the A8 and the DA models, the A2 sometimes overestimates the SD motions. That is especially the case in the coastal complex area (e.g., Fig. 11b). When the model overestimates the SD motions, the ABPs reduce the SD power, often leading to underprediction of SD motions. The AN forecast often severely lacks power for these SD motions, whereas the KFAS performs better. There is no energy overestimation by the A2 model for SD motions in the mountain complex terrain nor in the nearly flat continental terrain. The energy is well simulated or underpredicted by all forecasts tested, and the best results are produced when using either the A2 model or the KFAS (Figs. 11e,h,k). Finally, as it can be seen at Fig. 11b, if a model overpredicts the amplitudes of diurnal cycle harmonics, the analog approach is able to reduce them.

7. Conclusions

Different postprocessing methods (PPMs) are compared in this study, based on a historical dataset including three different mesoscale model runs with 8-km grid spacing [ALADIN (A8)] and 2-km grid spacing [ALADIN (A2); dynamical adaptation (DA)], as well as verifying observations of 10-m wind speed. By analyzing root-mean-square error (RMSE), rank correlation coefficient (RCC), and bias of the mean, it is shown that all tested PPMs improve results of the A8 starting model. The best results are obtained for the analog-based predictions (ABPs) when using 15 analog ensemble members. The RMSE and bias growth are noticed for larger ensembles, probably because of climatological differences between training and verification periods.

The PPMs reduce RMSE and bias when compared to the starting model. The KF and the forecasting KF of the analog ensemble mean (KFAN) are the most successful PPMs for the bias reduction. That is the expected result since the KF algorithm is constructed to remove the systematic error if it does not change rapidly (i.e., large
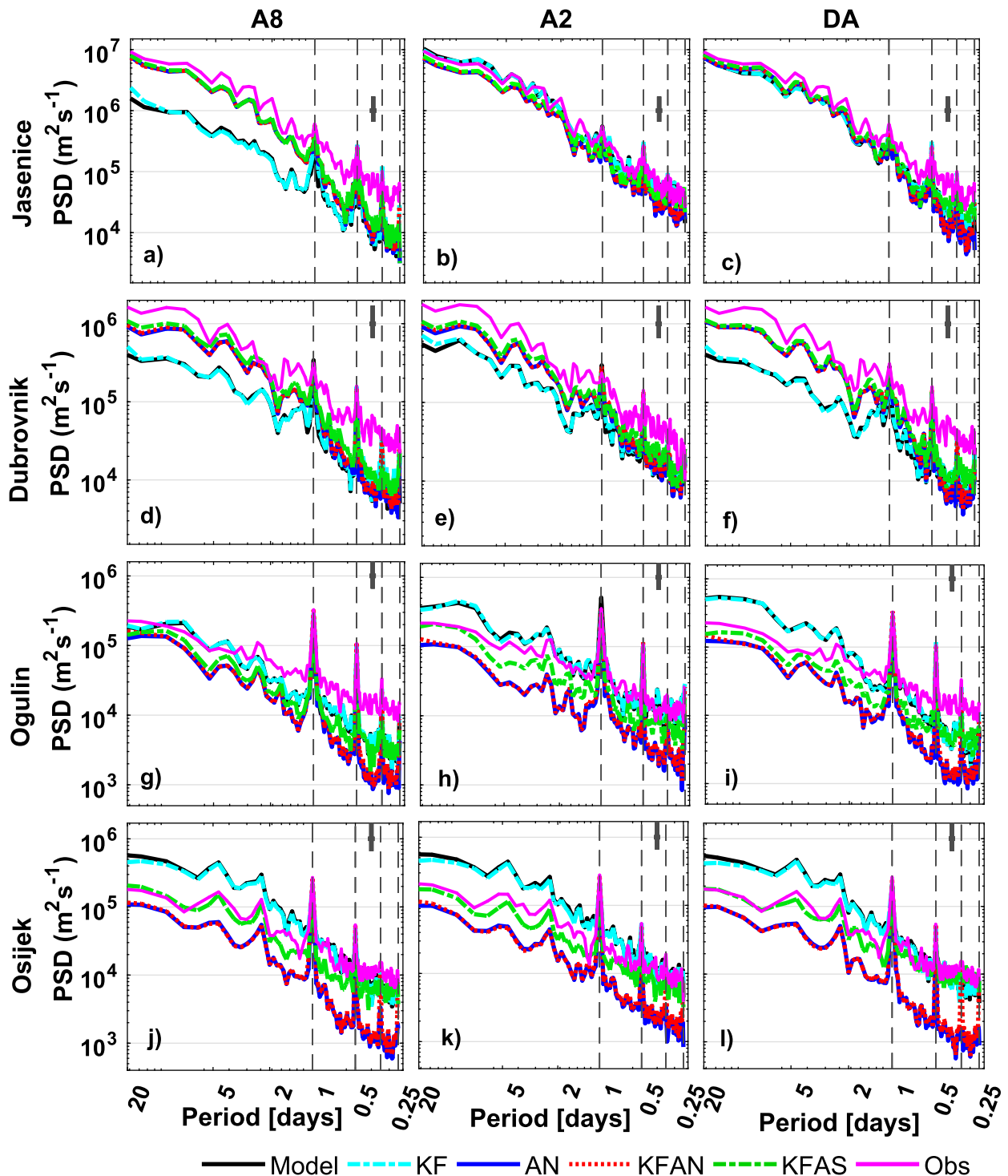
FIG. 11. The PSD of the observed 10-m wind speed, starting model forecasts (A8, A2, and DA) and the corresponding PPMs (KF, AN, KFAN, and KFAS) for stations Jasenice, Dubrovnik, Ogulin, and Osijek during year 2012. The confidence intervals (in the logarithmic scale) are noted by the cross-like symbol in the upper-right corner of each plot.

day-to-day variations). However, the application of the KF algorithm can also lead to the decrease of the correlation coefficient (i.e., increase of the dispersion error). The dispersion error is noticeably reduced by the KF approach in the flat terrain, where there are some indications of a systematic error influencing large-scale (i.e., period longer than 10 days) strong wind in the model. The KF is not as successful in reducing the dispersion error in the coastal complex area. The ABPs reduce dispersion error (i.e., improve RCC) regardless of the terrain complexity, showing greater adaptability than the KF forecast. Forecasting the analog ensemble mean (AN) seems to be the most suitable PPM for the RCC improvement.

The standard deviation (STD) of the KF forecasts is closer to the observed STD than the ABPs. The underestimation of the measured STD for the ABPs is partially explained by climatological differences between training and testing periods. The AN forecast is the most prone to systematic underestimation of the STD. This is due to additional averaging when forecasting the ensemble mean, thus naturally reducing the variability of the forecast. This systematic error is partially removed by the application of the KF algorithm in the KFAN forecast. The KFAS forecast exhibits the highest STD among the ABPs because of better adaptability. Even though the ABPs affect different aspects of the starting model, the RMSE reduction is very similar among them and superior to the KF approach. This is especially the case in the coastal complex terrain.

Comparison of PPMs' performance with starting models at 2-km grid spacing (A2; DA) compared to the PPMs' performance with the A8 as a starting model shows that PPMs considerably improve numerical predictions for all tested model resolutions. We furthermore test the hypothesis that the greater the representation of physical processes directly simulated by the starting model, the better the quality of the analogs. Even though the higher-resolution starting models yield better statistical results themselves, it is not necessarily the case for the ABPs generated by the higher-resolution model. This may be due to the imperfections of the point-based verification metrics used that typically increase with resolution at near-kilometer-scale grid spacing of numerical models (i.e., high sensitivity to spatial and phase errors). Therefore, the categorical and spectral analyses are performed to explore the potentially undetected benefits of using the higher-resolution models further.

To assess the performance of forecasts for different wind speeds, we perform a categorical verification using three wind speed categories: weak, moderate, and strong wind. Overall, starting models forecast weak wind occurrence too rarely and moderate wind occurrence too often. For coastal complex terrain, different starting models yield different frequency bias. For other terrain types (mountain complex and nearly flat continental), all starting models tested in this study underforecast the frequency of weak wind and overforecast the frequency of moderate and strong wind. All PPMs significantly reduce the frequency bias for climatologically common wind speed category on average. The main challenge for ABPs seems to be the underforecasting of strong wind occurrence, while the results for the KF are slightly less biased. The KFAS seems to be the least biased ABP, showing the best result for strong wind, while being as unbiased as the AN in the other two categories. It has to be noted that because of the small sample size, the results in the moderate and strong wind speed categories exhibit very large confidence intervals, providing only indications of the PPMs' ability to improve the starting model forecast.

The KF has considerably higher critical success index (CSI) than the starting models for the weak wind category in the nearly flat continental and mountain complex terrain, but not as much in the coastal complex terrain. Results suggest that the CSI result is improved for the moderate and strong winds. The ABPs seem to outperform both starting models and corresponding KF forecasts for all the cases tested, except the prediction of the strong wind in the nearly flat continental terrain. The KF seems to be the best PPM in this case, once again suggesting consistent model error when predicting strong wind. The AN achieves the highest CSI for weak wind, while the KFAN and the KFAS seem to be better in predicting the other categories. However, the results corresponding to moderate and strong winds should be further reinforced using a larger sample size. Using finer horizontal resolution might lead to the improvements for the CSI for starting model predictions of the strong wind in the coastal complex terrain. This suggests that finer resolution might lead to a better ability of the forecast in distinguishing low from moderate or unusually strong wind. This horizontal resolution increase yields mixed results for other categories and terrain types, potentially due to statistical imperfections of the metrics. This property is then inherited by all of the PPMs.

The spectral analysis confirmed that the KF approach affects only large-scale motions (i.e., period longer 10 days) if the power spectral density (PSD) function is biased. The KF thus enlarges the energy of the large-scale motions in the coastal area and reduces the energy of the large-scale motions it the nearly flat continental terrain. The possibly too conservative KF approach might be slightly adjusted by optimizing the parameters of the KF algorithm. However, the KF does not significantly influence the shorter time scale.

Unlike the KF approach, introducing past similar situations in the ABPs leads to better forecasting processes on longer-than-diurnal (LTD) scales. The LTD scales are much more relevant than the larger scales (with a period longer than 10 days) for forecasts up to 72 h ahead. The ABPs improve model underestimation of the LTD motions in the coastal area and in the nearly flat terrain when the model overestimates the LTD motions. The KFAS method is superior to the other PPMs because it maintains the modeled energy for the shorter-than-diurnal (SD) part of the power spectra (unlike the AN), while it improves both under- and overestimation of the LTD motion energy (just as good as or better than the AN). Furthermore, higher-resolution models generally contain more energy. Consequently, there are fewer situations with the underprediction of large-scale motions. But when they do occur, the PPMs behave as presented for the lower-resolution A8 model. Even though the ABPs often underpredict the SD motions, they simulate the correct amplitude of diurnal cycle harmonics (24-, 12-, and 8-h) similarly to the model. Additionally, even if the model overpredicts the amplitudes of diurnal cycle harmonics, the analog approach reduces them.

The performed verification shows that all analyzed PPMs improve upon the starting model forecasts. The level of improvement depends on the type of terrain, starting model, and verification metric. Each tested PPM has its strengths and weaknesses, and the choice for operational use of those methods depends on the envisaged purpose. The results are presented in such manner that after a simple statistical analysis of the potential starting model, one can thus decide which PPM is the most applicable for a specific situation. Summarily, the AN exhibits the highest correlation with measurements. It is thus applicable if the model is unbiased, but the dispersion error needs to be removed. The KF and the KFAN are the most successful in removing bias, whereas the KFAN is better suited if the terrain is more complex. The ABPs exhibit better results than KF in the complex terrain in general, especially the coastal area. If the focus is on the prediction of the weak wind, then the AN is the most suitable, whereas for the strong wind the analog approach is better suited when combined with the KF (i.e., KFAS). The KF algorithm affects only the large-scale flows (i.e., enlarges the energy of these large-scale motions in the coastal area and reduces the energy at the nearly flat continental terrain for the periods longer than 10 days), while the ABPs affect smaller scales. If the starting model PSD spectra are biased, the KFAS method is superior to the other ABPs. The superior adaptability of the KFAS results in an STD that is the closest to the measured STD.

Additionally, results of the PPMs are further improved if higher-resolution (convection permitting) starting model data are used in the coastal complex terrain. Introducing the higher-resolution modeling in nearly flat continental terrain results in very similar PSD curves. The PPMs exhibit at least as good of a result when using the coarser horizontal resolution, if not better. Therefore the need for using a 2-km as opposed to 8-km grid spacing model may be questioned. On the other hand, the higher-resolution modeling increased the energy available for all of the time scales in the mountain complex terrain. The latter, however, yielded mixed results when using the other verification metrics for both the starting models and corresponding PPMs. In this case, the results may be improved by using even finer model resolution than 2 km to represent local flows.

Finally, future contributions could also focus on probabilistic predictions, taking full advantage of the distribution sampled by the ensemble, which may be more suitable, particularly for the prediction of rare events.

REFERENCES

ALADIN International Team, 1997: The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. *WMO Bull.*, **46**, 317–324.

Alessandrini, S., L. Delle Monache, S. Sperati, and G. Cervone, 2015a: An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy*, **157**, 95–110, https://doi.org/10.1016/j.apenergy.2015.08.011.

——, ——, ——, and J. Nissen, 2015b: A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, **76**, 768–781, https://doi.org/10.1016/j.renene.2014.11.061.

Bajić, A., S. Ivatek-Šahdan, and K. Horvath, 2007: Spatial distribution of wind speed in Croatia obtained using the ALADIN model. *Croat. Meteor. J.*, **42**, 67–77.

——, ——, and Z. Žibrat, 2008: ANEMO-ALARM iskustva operativne primjene prognoze smjera i brzine vjetra (ANEMO-ALARM operational use of wind speed and direction forecast). GIU Hrvatski cestar, 109–114.

Catry, B., J.-F. Geleyn, M. Tudor, P. Benard, and A. Trojakova, 2007: Flux conservative thermodynamic equations in a mass-weighted framework. *Tellus*, **59A**, 71–79, https://doi.org/10.1111/j.1600-0870.2006.00212.x.

Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, and R. B. Stull, 2006: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *J. Geophys. Res.*, **111**, D05308, https://doi.org/10.1029/2005JD006311.

——, and Coauthors, 2008: A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus*, **60B**, 238–249, https://doi.org/10.1111/j.1600-0889.2007.00332.x.

——, T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather

predictions. *Mon. Wea. Rev.*, **139**, 3554–3570, https://doi.org/10.1175/2011MWR3653.1.

——, T. Eckel, D. Rife, and B. Nagarajan, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, https://doi.org/10.1175/MWR-D-12-00281.1.

Djalalova, I., L. Delle Monache, and J. Wilczak, 2015: PM$_{2.5}$ analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmos. Environ.*, **108**, 76–87, https://doi.org/10.1016/j.atmosenv.2015.02.021.

Drosdowsky, W., 1994: Analog (nonlinear) forecasts of the Southern Oscillation index time series. *Wea. Forecasting*, **9**, 78–84, https://doi.org/10.1175/1520-0434(1994)009<0078:AFOTSO>2.0.CO;2.

Ekström, J., 2011: On the relation between the polychoric correlation coefficient and Spearman's rank correlation coefficient. UCLA Dept. of Statistics Paper, 15 pp.

Gao, L., H. Ren, J. Li, and J. Chou, 2006: Analogue correction method of errors and its application to numerical weather prediction. *Chin. Phys.*, **15**, 882–889, https://doi.org/10.1088/1009-1963/15/4/038.

Geleyn, J.-F., 1987: Use of a modified Richardson number for parameterizing the effect of shallow convection. *J. Meteor. Soc. Japan* (Suppl.: NWP Symp.), **64A**, 141–149, https://doi.org/10.2151/jmsj1965.64A.0_141.

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

——, and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitr. Phys. Atmos.*, **52**, 1–16.

——, and Coauthors, 1994: Atmospheric parametrization schemes in Météo-France's ARPEGE N.W.P model. *Proc. 1994 ECMWF Seminar on Parametrization of Sub-Grid Scale Physical Processes*, Reading, United Kingdom, ECMWF, 385–402, https://www.ecmwf.int/sites/default/files/elibrary/1994/9536-atmospheric-parametrization-schemes-meteo-frances-arpege-nwp-model.pdf.

——, F. Váňa, J. Cedilnik, M. Tudor, and B. Catry, 2006: An intermediate solution between diagnostic exchange coefficients and prognostic TKE methods for vertical turbulent transport. *Research Activities in Atmospheric and Oceanic Modelling*, J. Côté, Ed., World Meteorological Organization, 4.11–4.12.

——, B. Catry, Y. Bouteloup, and R. Brožkova, 2008: A statistical approach for sedimentation inside a microphysical precipitation scheme. *Tellus*, **60A**, 649–662, https://doi.org/10.1111/j.1600-0870.2008.00323.x.

Gerard, L., 2007: An integrated package for subgrid convection, clouds and precipitation compatible with the meso-gamma scales. *Quart. J. Roy. Meteor. Soc.*, **133**, 711–730, https://doi.org/10.1002/qj.58.

——, and J.-F. Geleyn, 2005: Evolution of a subgrid deep convection parametrization in a limited area model with increasing resolution. *Quart. J. Roy. Meteor. Soc.*, **131**, 2293–2312, https://doi.org/10.1256/qj.04.72.

——, J.-M. Piriou, R. Brozkova, J.-F. Geleyn, and D. Banciu, 2009: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model. *Mon. Wea. Rev.*, **137**, 3960–3977, https://doi.org/10.1175/2009MWR2750.1.

Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997–1015, https://doi.org/10.1175/1520-0493(2000)128<0997:IOANAS>2.0.CO;2.

Grisogono, B., and D. Belusic, 2009: A review of recent advances in understanding the meso- and micro-scale properties of the severe bora wind. *Tellus*, **61A**, 1–16, https://doi.org/10.1111/j.1600-0870.2008.00369.x.

Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, https://doi.org/10.1175/MWR3237.1.

Hopson, T. M., and P. J. Webster, 2010: A 1–10-day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003–07. *J. Hydrometeor.*, **11**, 618–641, https://doi.org/10.1175/2009JHM1006.1.

Horvath, K., S. Ivatek-Šahdan, B. Ivančan-Picek, and V. Grubišić, 2009: Evolution and structure of two severe cyclonic bora events: Contrast between the northern and southern Adriatic. *Wea. Forecasting*, **24**, 946–964, https://doi.org/10.1175/2009WAF2222174.1.

——, A. Bajić, and S. Ivatek-Šahdan, 2011: Dynamical downscaling of wind speed in complex terrain prone to bora-type flows. *J. Appl. Meteor. Climatol.*, **50**, 1676–1691, https://doi.org/10.1175/2011JAMC2638.1.

——, 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, https://doi.org/10.1029/2012JD017432.

Hrastinski, M., K. Horvath, I. Odak Plenković, S. Ivatek-Šahdan, and A. Bajić, 2015: Verification of the operational 10 m wind forecast obtained with the ALADIN mesoscale numerical weather prediction model. *Croat. Meteor. J.*, **50**, 105–120.

Ivatek-Šahdan, S., and M. Tudor, 2004: Use of high-resolution dynamical adaptation in operational suite and research impact studies. *Meteor. Z.*, **13**, 99–108, https://doi.org/10.1127/0941-2948/2004/0013-0099.

——, and B. Ivančan-Picek, 2006: Effects of different initial and boundary conditions in ALADIN/HR simulations during MAP IOPs. *Meteor. Z.*, **15**, 187–197, https://doi.org/10.1127/0941-2948/2006/0117.

Jolliffe, I. T., and D. B. Stephenson, Eds., 2011: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Wiley, 274 pp.

Junk, C., L. Delle Monache, S. Alessandrini, L. von Bremen, and G. Cervone, 2015: Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteor. Z.*, **24**, 361–379, https://doi.org/10.1127/metz/2015/0659.

Juras, J., and Z. Pasarić, 2006: Application of tetrachoric and polychoric correlation coefficients to forecast verification. *Geofizika*, **23**, 59–81, http://geofizika-journal.gfz.hr/vol_23/No1/juras.pdf.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, https://doi.org/10.1115/1.3662552.

Keller, D. E., A. M. Fischer, M. A. Liniger, C. Appenzeller, and R. Knutti, 2017: Testing a weather generator for downscaling climate change projections over Switzerland. *Int. J. Climatol.*, **37**, 928–942, https://doi.org/10.1002/joc.4750.

Klausner, Z., H. Kaplan, and E. Fattal, 2009: The similar days method for predicting near surface wind vectors. *Meteor. Appl.*, **16**, 569–579, https://doi.org/10.1002/met.158.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

Louis, J.-F., M. Tiedtke, and J.-F. Geleyn, 1982: A short history of PBL parameterization at ECMWF. *Proc. ECMWF Workshop on Planetary Boundary Layer Parameterization.* Reading,

United Kingdom, 59–79, https://www.ecmwf.int/sites/default/files/elibrary/1982/10845-short-history-pbl-parameterization-ecmwf.pdf.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecast? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Nagarajan, B., L. Delle Monache, J. P. Hacker, D. L. Rife, K. Searight, J. C. Knievel, and T. N. Nipen, 2015: An evaluation of analog-based postprocessing methods across several variables and forecast models. *Wea. Forecasting*, **30**, 1623–1643, https://doi.org/10.1175/WAF-D-14-00081.1.

Noilhan, J., and S. Planton, 1989: A simple parametrization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549, https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2.

Panziera, L., U. Germann, M. Gabella, and P. V. Mandapaka, 2011: NORA—Nowcasting of orographic rainfall by means of analogues. *Quart. J. Roy. Meteor. Soc.*, **137**, 2106–2123, https://doi.org/10.1002/qj.878.

Poje, D., 1995: Bura (bora) and burin in Split. *Croat. Meteor. J.*, **30**, 1–19.

Redelsperger, J. L., F. Mahé, and P. Carlotti, 2001: A simple and general subgrid model suitable both for surface layer and free-stream turbulence. *Bound.-Layer Meteor.*, **101**, 375–408, https://doi.org/10.1023/A:1019206001292.

Ren, H., and J. Chou, 2006: Analogue correction method of errors by combining statistical and dynamical methods. *Acta Meteor. Sin.*, **20**, 367–373.

——, and ——, 2007: Strategy and methodology of dynamical analogue prediction. *Sci. China*, **50D**, 1589–1599, https://doi.org/10.1007/s11430-007-0109-6.

Rife, D. L., and C. A. Davis, 2005: Verification of temporal variations in mesoscale numerical wind forecasts. *Mon. Wea. Rev.*, **133**, 3368–3381, https://doi.org/10.1175/MWR3052.1.

——, ——, and Y. Liu, 2004: Predictability of low-level winds by mesoscale meteorological models. *Mon. Wea. Rev.*, **132**, 2553–2569, https://doi.org/10.1175/MWR2801.1.

Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.*, **120**, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2.

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. Michaelides, Ed., Springer, 419–452.

Ruosteenoja, K., 1988: Factors affecting the occurrence and lifetime of 500 mb height analogues: A study based on a large amount of data. *Mon. Wea. Rev.*, **116**, 368–376, https://doi.org/10.1175/1520-0493(1988)116<0368:FATOAL>2.0.CO;2.

Stanešić, A., 2011: Assimilation system at DHMZ: Development and first verification results. *Croat. Meteor. J.*, **44/45**, 3–17.

Telišman Prtenjak, M., and B. Grisogono, 2007: Sea/land breeze climatological characteristics along the northern Croatian Adriatic coast. *Theor. Appl. Climatol.*, **90**, 201–215, https://doi.org/10.1007/s00704-006-0286-9.

Termonia, P., 2008: Scale-selective digital filter initialization. *Mon. Wea. Rev.*, **136**, 5246–5255, https://doi.org/10.1175/2008MWR2606.1.

Tudor, M., and S. Ivatek-Šahdan, 2002: The MAP-IOP 15 case study. *Croat. Meteor. J.*, **37**, 1–14.

——, ——, A. Stanešić, K. Horvath, and A. Bajić, 2013: Forecasting weather in Croatia using ALADIN numerical weather prediction model. *Climate Change and Regional/Local Responses*, P. Ray and Y. Zhang, Eds., InTech, 59–88, http://doi.org/10.5772/55698.

Van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247, https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

Vanvyve, E., L. Delle Monache, D. Rife, A. Monaghan, and J. Pinto, 2015: Wind resource estimates with an analog ensemble approach. *Renew. Energy*, **74**, 761–773, https://doi.org/10.1016/j.renene.2014.08.060.

Welch, P., 1967: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.*, **15**, 70–73, https://doi.org/10.1109/TAU.1967.1161901.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

WMO, 2008: Guide to meteorological instruments and methods of observation. 7th ed., WMO 8, 716 pp., http://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf.

Xavier, P. K., and B. N. Goswami, 2007: An analog method for real-time forecasting of summer monsoon subseasonal variability. *Mon. Wea. Rev.*, **135**, 4149–4160, https://doi.org/10.1175/2007MWR1854.1.

Žagar, M., and J. Rakovec, 1999: Small-scale surface wind prediction using dynamic adaptation. *Tellus*, **51A**, 489–504, https://doi.org/10.3402/tellusa.v51i4.14051.

Žagar, N., M. Žagar, J. Cedilnik, G. Gregorič, and J. Rakovec, 2006: Validation of mesoscale low-level winds obtained by dynamical downscaling of ERA40 over complex terrain. *Tellus*, **58A**, 445–455, https://doi.org/10.1111/j.1600-0870.2006.00186.x.

Zaninović, K., and Coauthors, 2008: *Klimatski Atlas Hrvatske 1961–1990, 1971–2000* (*Climate Atlas of Croatia 1961–1990, 1971–2000*). Meteorological and Hydrological Service, 200 pp.

Zhang, J., C. Draxl, T. Hopson, L. Delle Monache, and B.-M. Hodge, 2015: Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Appl. Energy*, **156**, 528–541, https://doi.org/10.1016/j.apenergy.2015.07.059.