Received: 1 August 2019 | Revised: 22 January 2020 | Accepted: 21 February 2020

DOI: 10.1002/qj.3769

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Analog-based post-processing of the ALADIN-LAEF ensemble predictions in complex terrain

Iris Odak Plenković[1] | Irene Schicker[2] | Markus Dabernig[2] | Kristian Horvath[1] | Endi Keresturi[1]

[1]SMIR, Croatian Meteorological and Hydrological Service, Meteorological Research and Development Sector, Zagreb, Croatia

[2]VHMOD, Zentralanstalt für Meteorologie und Geodynamik, Vorhersagemodelle, Vienna, Austria

Correspondence
I. Odak Plenković, SMIR, Croatian Meteorological and Hydrological Service, Državni hidrometeorološki zavod, Grič 3, 1000 Zagreb, Croatia.
Email: odak@cirus.dhz.hr

Abstract

The main goal of this study is to assess the performance of the analog-based post-processing method applied to the Austrian ALADIN-LAEF wind speed ensemble predictions through a set of sensitivity experiments. Evaluation of several analog-based configurations using various meteorological variables as predictors is therefore conducted. The results of those experiments are compared to the ensemble model output statistics (EMOS) baseline model. The hypothesis further investigated is that using summarized measures, such as mean and standard deviation of an ensemble for several meteorological variables, is comparable to the analog post-processing using all of the ensemble members. Results show that both analog-based and EMOS experiments considerably improve the raw model forecast. Even though the improvement over raw model forecast is evident, large differences among nearby stations are noticed in the highly complex terrain. The processes at lower stations seem to be better represented by the raw model, which leads to a better input forecast to the post-processing and a better overall result than for the mountain stations. The analog-based method is overall comparable to or even outperforms the EMOS. Assessing the post-processing performance for high wind speeds shows that the analog experiments can improve the raw forecast, exhibiting significantly higher skill than the EMOS. The difference among all analog experiments is less pronounced, especially the experiment using all of the raw model ensemble members and the one using summarized measures. Furthermore, it is demonstrated that the usage of summarized ensemble measures is an optimal way to improve the forecast skill, compared to the other analog-based experiments and the EMOS model. Therefore, it is suggested that it is not necessary to increase the computational costs by using the full input spectrum of a raw probabilistic model, that is, all ALADIN-LAEF members as predictors, as the summarized metric suffices.

KEYWORDS

analog approach, complex terrain, ensemble calibration, ensemble model output statistics, ensemble prediction, mesoscale model, statistical post-processing, wind ensemble forecast

1 | INTRODUCTION: THE ANALOGIES AS A PART OF A WEATHER PREDICTION SYSTEM

Analogies between, for example, similar past forecasts, measurements or analyses as a potentially useful tool for forecasting the future state of the weather have been explored for decades, in both positive and pessimistic ways. Lorenz (1969) and Rousteenoja (1988), for instance, claimed that one needs to wait an astronomical number of years until the likelihood of finding two atmospheric states that differ less than the present-day observational error is sufficiently high enough to be considered as usable. Back then, the applicability of analogues for short-range weather forecasting was discarded. Even so, Van den Dool (1989) showed that if the number of degrees of freedom in the matching procedure is reduced, finding suitable analogies is possible. In contrast to Lorenz (1969) who searched the entire Northern Hemisphere, Van den Dool (1989) used data centred over a localized area for the analogue search.

In the past, a set of different analogue search procedures was defined. This was done mainly because the use of analogues for forecasting of meteorological fields is limited due to excessive degrees of freedom of the problem at stake. Xavier and Goswami (2007) used the National Oceanic and Atmospheric Administration (NOAA)'s outgoing long-wave radiation fields for long-range weather predictions, whereas Panziera *et al.* (2011) performed very short-term orographic precipitation predictions using radar observations. Besides single fields, the use of spatially correlated observational variables (Wu *et al.*, 2012) also proved to be suitable. Satisfactory results were also achieved for the Southern Oscillation Index (SOI) forecasts using SOI measurements (Drosdowsky, 1994), point wind-speed forecasts using wind speed measurements (Klausner *et al.*, 2009), for idealized cases with low-order models (Ren and Chou, 2006), and general-circulation modelling (Gao *et al.*, 2006; Ren and Chou, 2007).

As a very successful continuation of the aforementioned studies, Delle Monache *et al.* (2011) proposed two deterministic analog-based post-processing methods using historical data that includes both observations and numerical weather prediction (NWP) data. They applied the analog method to a single site to improve 10 m wind-speed deterministic NWP forecasts. In contrast to a recursive and linear Kalman filter post-processing approach (KF), the deterministic approaches using analogs both had a higher correlation and lower random and systematic error than the KF method (Delle Monache *et al.*, 2006; 2008; 2011). Similar approaches were used for predicting other variables, such as PM2.5 (fine particulate matter) concentrations (Djalalova *et al.*, 2015), or even across several models and meteorological variables (Nagarajan *et al.*, 2015).

Van den Dool (1989) revealed that analogues can be used to predict the forecast skill of an NWP model. Hamill *et al.* (2006) and Hopson (2005) extended the idea and applied the analogues to ensemble forecasts. Hamill and Whitaker (2006) stated that, when comparing the pattern match of the historical local ensemble-mean forecast to the current ensemble-mean forecast in the same region, it is possible to find many similar and useful analogues within a few decades of reforecasts. Their study focused on probabilistic forecasts of 24 hr precipitation. All the aforementioned analogue-techniques were able to improve the Brier skill score, resulting in a skill comparable to a logistic regression technique. The authors, while comparing different analogue-techniques, also concluded that selecting analogues for each member rather than for the ensemble mean generally decreased the forecast skill. Another successful example of a calibrating ensemble forecast can be found in Hopson and Webster (2010). The authors sought analogues to generate the final set of discharge ensembles accounting for all aspects of discharge forecast uncertainty (meteorological and hydrological). This part of the fully automated operational 1–10-day multi-model ensemble forecasting scheme for the major river basins of Bangladesh helped to evacuate many thousands of people and livestock during flood events in 2007.

Delle Monache *et al.* (2013) applied the analog ensemble (AnEn) approach to produce probabilistic 10 m wind speed and 2 m temperature forecasts using only one deterministic NWP model as input. They showed that the AnEn exhibits high statistical consistency and reliability. Similarly, Vanvyve *et al.* (2015) provided high-quality long-term wind resource estimates. The probabilistic analog-based predictions were also successfully used to gain wind resource estimates (Vanvyve *et al.*, 2015; Zhang *et al.*, 2015), to predict solar irradiance (Alessandrini *et al.*, 2015a), wind power (Alessandrini *et al.*, 2015b; Junk *et al.*, 2015), downscale precipitation (Keller *et al.*, 2017) and 10 m wind speed (Sperati *et al.*, 2017).

Additional to using a deterministic NWP to create AnEn (as described in Delle Monache *et al.*, 2011; 2013), the same approach could also be applied using an NWP ensemble. The AnEn ability to capture the flow-dependent error growth would be complemented with the aspects of error growth that could be represented dynamically by the multiple model runs of an NWP ensemble. Following that idea, Eckel and Delle Monache (2016) produced m analogs for each member of the n-member NWP ensemble, resulting in an $m \times n$ "hybrid" analog ensemble. The approach yielded mixed results for 10 m wind speed forecasts, while the application for the 2 m temperature forecast was more successful. Mugume *et al.* (2017), who used the analog approach to post-process ensemble members who use different convection parametrization schemes, also explored the same idea. The authors demonstrated a root-mean-square error (RMSE) and bias reduction in rainfall prediction when using corresponding predictions of the (starting) ensemble mean analog as a forecast. Slightly better results (e.g. significant reduction of negative bias error) were achieved when seeking the analog for every (starting) ensemble member and then averaging the analogs.

In this study we propose an in-depth analysis of different analog-based configurations applied to the Austrian ALADIN-LAEF ensemble forecasts. Following the work of Eckel and Delle Monache (2016) and Mugume *et al.* (2017), the main goal of this study is to significantly improve the ALADIN-LAEF ensemble 10 m wind speed forecast while maintaining low computational cost for the analog search. To test the performance of the analog-based post-processing and determine the optimal configuration, several experiments using different sources of information available to the ALADIN-LAEF ensemble forecasts are performed. The experiments include using one or more ALADIN-LAEF meteorological variables as predictors. Through performed analysis, the experiments including only information about the ALADIN-LAEF ensemble mean (as suggested by Hamill and Whitaker, 2006) or every ensemble member (similar to Mugume *et al.*, 2017) are also tested. A novelty in this study is the usage of the starting model ensemble uncertainty through its standard deviation (σ) in addition to ensemble mean (μ). The hypothesis additionally explored in this study is that using a summarized measure, as σ, is the optimal way to dynamically represent the aspects of error growth of the input ensemble model to the flow-dependent error growth, which is already captured by the analog approach. This hypothesis is investigated and evaluated against other experiments using 29 meteorological observation sites (TAWES) in Austria for a winter (January) and summer (July) month of 2018. The ensemble model output statistics post-processing approach (EMOS: Gneiting *et al.*, 2005) is used as a reference model in order to better understand the analog-search impact on the raw forecasts. All experiments provide a 17-member wind-speed analog ensemble forecast, as well as the ALADIN-LAEF forecast.

In Section 2 the data are described, while Section 3 introduces the post-processing methods used and explains the experimental set-up in detail. The results are presented in Section 4 and conclusions are highlighted in Section 5.

2 | DATA

2.1 | Observations and climatology

The Austrian meteorological observation network, TAWES, consists of more than 300 sites across Austria. In this work, 29 TAWES sites are used representing the different Austrian climate zones. All sites monitor temperature, wind speed and direction, relative humidity, pressure, precipitation, and, depending on the site, different radiation measurements are carried out. Here, only 10 m wind speed observations are used. The 2015 and 2016 wind speed observations are used for the analog-based method training period in this study. For the performance testing, two target months are chosen, January and July 2018. These months are selected to investigate the forecast performance during winter and summer periods. The January and July 2017 wind speed observations are used for independent sensitivity testing (weight optimization).

The observed average monthly wind speed is slightly higher in January (2.88 $\mathrm{m \cdot s^{-1}}$) than in July (2.22 $\mathrm{m \cdot s^{-1}}$), across all available stations and lead-times. Additionally, the standard deviation of the wind speed measurements is also higher on average in January (3.27 $\mathrm{m \cdot s^{-1}}$) than in July (1.92 $\mathrm{m \cdot s^{-1}}$).

The wind speed is weak and moderate (i.e. <8 $\mathrm{m \cdot s^{-1}}$) for both January (Figure 1a) and July (Figure 1b) at the majority of the stations. The average monthly wind speed

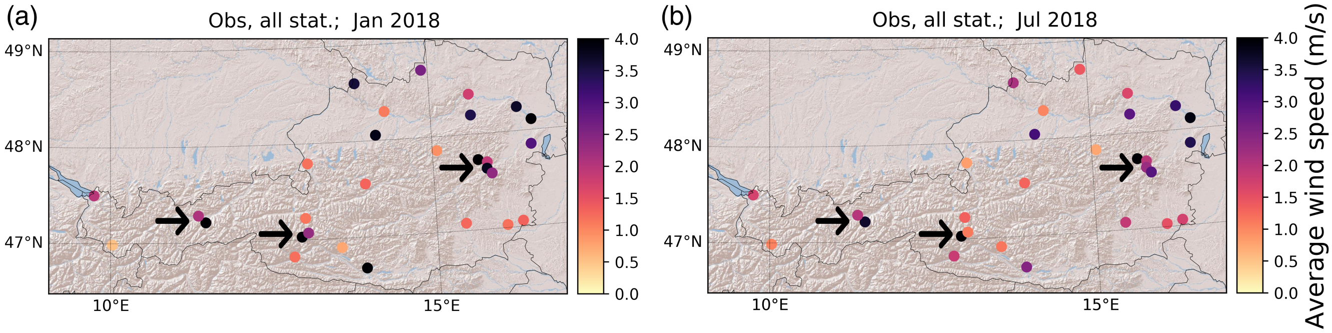

FIGURE 1 The spatial distribution of the observed monthly mean wind speed in (a) January and (b) July, 2018. The arrows mark mountain stations for later comparison

TABLE 1 The summary information for the experiments tested in this work

Name	Meteorological variables used	ALADIN-LAEF input (predictors)	No. of analog searches per lead-time
LAEFws	*ws*	X	X
EMOSws	*ws*	Ensemble μ and σ for one parameter, wind speed (2 predictors)	X
EMOSstd	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	Ensemble μ and σ for six parameters (12 predictors)	X
AnEnCtrl	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	Control ensemble member for six parameters (6 predictors)	1
AnEnWs	*ws*	17 ensemble wind speed members (17 predictors)	1
AnEnMu	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	Ensemble μ for six parameters (6 predictors)	1
AnEnStd	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	Ensemble μ and σ for 6 parameters (12 predictors)	1
AnEnAll	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	17 ensemble members for 6 parameters (6 × 17 predictors)	1
AnEnMem	*ws*, *dir*, *t2m*, *rh*, *p*, *prec*	1 ensemble member for every parameter (6 predictors)	17

increases towards the northeastern part of Austria (Pannonian basin) for both January and July. Exceptions are the three mountain stations (arrows in Figure 1), where the average wind speeds are much higher if compared to the neighbouring valley stations.

Most of the stations are located in or near the Alps, which significantly modulates the related local wind regimes. The complex terrain of the Alpine area is characterized by a variety of different wind processes such as föhn and downslope windstorms, gap winds, valley, and slope winds, flow blocking and others. To investigate those phenomena, among others, the Alpine region has been the target area for several major field experiments, such as ALPEX, MAP and TEAMx (e.g. Kuettner, 1986; Bougeault *et al.*, 2001; Lehner and Rotach, 2018; Serafin *et al.*, 2018). Nevertheless, many challenges related to NWP in complex terrain still exist (e.g. Arnold *et al.*, 2012), including modelling wind climatology of the Alpine areas prone to such downslope windstorms (Horvath *et al.*, 2011) and objective föhn wind classification (e.g. Mayr *et al.*, 2018).

2.2 | Model data

The numerical model used in this study is the ALADIN-LAEF (Wang *et al.*, 2011) ensemble forecasting system. It is adjusted to fit Austrian purposes and has been running in operational mode since 2013. The ALADIN-LAEF consists of 17 ensemble members: 16 perturbed members and one control run. The 16 perturbed members are driven by 16 European Centre for Medium-range Weather Forecasts Ensemble Prediction System (ECMWF-EPS) members. Given the structure and composition of the LAEF ensemble, it can be considered as a non-exchangeable ensemble. However, as could be shown by Baran and Lerch (2015), the differences between the treatment of a non-exchangeable ensemble as fully exchangeable did not worsen the results to a statistically relevant size. Therefore, we decided to treat the LAEF ensemble as exchangeable. The resolution of 10.9 km on a Lambert conformal grid is used in the horizontal. In the vertical, 45 terrain-following pressure-based hybrid coordinate levels with on average nine levels within the lowest 1,000 m above ground level are used. It is initialized daily at 0000 and 1200 UTC with one hourly lead-time, up to 72 hr. Only the dataset corresponding to the model run initialized at 0000 UTC is used in this work. A pre-selected subset of six ALADIN-LAEF parameters (temperature (*t2m*), wind speed (*ws*) and direction (*dir*), relative humidity (*rh*), pressure (*p*) and precipitation (*prec*)) is used in different combinations (summarized in Table 1). The datasets correspond to the previously mentioned observation datasets. The 2-year long dataset (2015–2016) is used for training. The 2-month period (January and July 2017) is used for weight optimization. Finally, the results are given for the independent dataset consisting of January and July 2018.

3 | METHOD

3.1 | Reference method

To assess the performance of the proposed analog ensemble methods, a reference is needed. The reference in this article is the ensemble model output statistics (*EMOS*). The *EMOS* is introduced by Gneiting *et al.* (2005) and adapted for wind by Messner *et al.* (2014). Therefore, a non-homogeneous regression with a 30-day rolling training window is fitted on every lead-time and station. To capture the natural boundary of wind at 0 m·s^{-1}, a left-censored logistic regression is used. In the *EMOS* the observed wind speed (y) is explained by a logistic distribution censored at 0 ($\mathcal{L}_0$) with μ as a mean and σ as a spread. A logistic distribution has a similar bell shape as a Gaussian distribution but with slightly heavier tails. Additionally, censoring at zero states that no negative wind values can occur. Further details can be found in Messner *et al.* (2014). Censoring and the linear regressions for μ and σ are defined as follows:

$$\text{wind speed} = \begin{cases} 0 & \text{if } y \leq 0 \\ y & \text{else'} \end{cases} \tag{1}$$

$$y \sim \mathcal{L}_0(\mu, \sigma), \tag{2}$$

$$\mu = \beta_0 + \beta_1 \; ws_\mu, \tag{3}$$

$$\log(\sigma) = \gamma_0 + \gamma_1 \; \log(ws_\sigma), \tag{4}$$

with β_* and γ_* as regression coefficients, ws_μ as ensemble mean, and ws_σ as ensemble spread of the wind speed members. The logarithmic link function is used to ensure positive values. Further applications of the *EMOS* to wind speed can be found in Thorarinsdottir and Gneiting (2010), Baran and Lerch (2015) or Scheuerer and Möller (2015).

The 30-day rolling training window is used for the *EMOSws* experiment, making it a good reference for the analog experiment that uses only the raw model wind-speed data. However, since the other analog experiments use all available variables, a second reference is added. The second experiment (*EMOSstd*) uses all available variables. The boosting method of Messner *et al.* (2017), which is implemented in the R-package "*crch*", is applied to all variables and the whole dataset, instead of the rolling training window. Additionally, annual and biannual harmonic functions are added to capture a seasonal bias. A variable selection method, such as boosting, is needed to prevent overfitting. The boosting is able to choose the most important variables and exclude the other variables using zero value. As a result, a single fit per station and lead time can be used to forecast both test months.

Concluding, whereas the *EMOSws* only uses the last 30 days as training and only the wind speed as an input, the *EMOSstd* uses all available training data and all variables including seasonal functions.

3.2 | The analog ensemble method

The probability distribution $f(y|x^{\mathrm{f}})$ of the observed future value of a variable y at a given time and location can be estimated by the analog ensemble (AnEn) using x^{f} representing k predictor variables from the starting (deterministic or ensemble) model prediction $x^{\mathrm{f}} = (x_{\mathrm{f}}^1, x_{\mathrm{f}}^2, \ldots, x_{\mathrm{f}}^k)$. The AnEn method uses historical point-based data within a specified analog training period for which both the starting model and the verifying observation are available. The so-called analogs (best-matching historical forecasts to the current prediction) may originate in any past date in the training period. The assumption is that the error of the good (quality) analog is likely to be similar to the error of the current forecast (Delle Monache *et al.*, 2011). As proposed in Delle Monache *et al.* (2013), the quality of the analog is evaluated by the following metric (Equation 1):

$$\|F_t A_{t'}\| = \sum_{i=1}^{N_A} \frac{w_i}{\sigma_{\mathrm{f}i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (\mathrm{F}_{i,t+j} - A_{i,t'+j})^2}, \tag{5}$$

where F_t is the current NWP forecast at a given location, valid at the future time t. The $A_{t'}$ is an analog at a given location with the same forecast lead-time, but valid at a past time t'. The N_A represents the number of predictor variables used in the analog search and w_i are the weights corresponding to a particular predictor. The absolute value of the metric is not important as such since it is only used for the intercomparison of analogs when used for sorting of the distances. Therefore, the weights are not constrained (i.e. their sum does not need to be fixed). For the fair comparison between different meteorological parameters, however, the weights w_i are normalized using the standard deviation, $\sigma_{\mathrm{f}i}$, of the time series of past forecasts of a given variable at the same location. The $\tilde{t}$ is equal to half of the time window width over which the metric is computed, 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. The time window is used to account for shifts and/or trends in the starting model forecast. The analog search is independent for every forecast time and location while limited around a particular time of day by a time window width. The width used in this work is 2 hr, including the ±1 lead time step, as proposed by Delle Monache *et al.* (2013). The number of degrees of freedom in the analog-finding procedure is therefore reduced (as proposed in Van den Dool, 1989). The members of the AnEn are verifying observations of the best-matching analogs. Once the AnEn is formed, it can be used to produce the deterministic analog-based prediction (e.g. forecasting AnEn mean value), as well as the probabilistic forecast (e.g. to estimate the probability of a predefined event). Several authors state that the AnEn rank histograms are uniform (e.g. Delle Monache *et al.*, 2013). Therefore, every member of the AnEn is an equally probable outcome, even though, measured by previously defined metrics, some analogs are closer to the current forecast than the others are.

Instead of assigning the same importance to each predictor variable, the brute-force weight optimization can

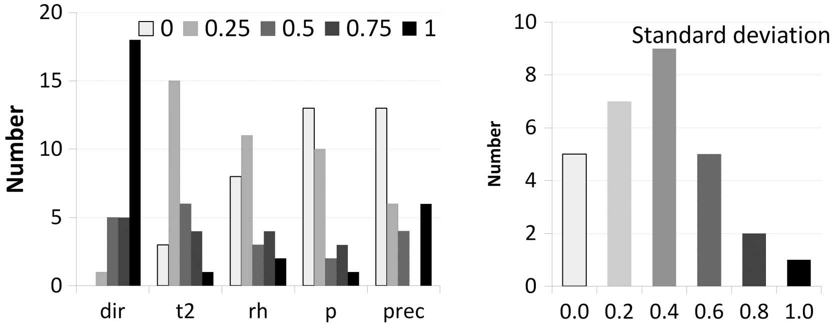

FIGURE 2 The histogram of the optimized weights for each predictor tested, at 29 stations in Austria (January and July 2017). (a) The predictors include wind direction (dir), temperature (t2), relative humidity (rh), pressure (p), and precipitation (prec). (b) The standard deviation predictor is calculated as one multiplying factor for several meteorological parameters

increase the AnEn performance. This is demonstrated in several applications, such as Junk *et al.* (2015) and Alessandrini *et al.* (2015a). The weights' optimization is based on choosing the combination that minimizes the error (measured by the continuous rank probability score). Due to the limited computational resources, not all the possible combinations are tested in this work. The forward selection algorithm is used instead, starting with weight value fixed at 1 for wind speed parameter. Then, one by one (ensemble mean) predictor is added from a pre-selected subset of six ALADIN-LAEF parameters, in the following order: *dir*, *t2*, *rh*, *p* and *prec* (as listed in Section 2.2). The weights are optimized independently at each location by error minimization. The forward selection algorithm is computationally less demanding than testing all the possible combinations independently at each location. However, it needs to be noted that the algorithm makes a key assumption that is often not true – assuming that all predictors are independent of each other, which is generally not the case. Five possible weight values (0.00, 0.25, 0.50, 0.75 and 1.00) are investigated for January and July 2017. Therefore, the optimization procedure uses a completely independent dataset from the period when training, as well as when forecasting is performed. The results show that the wind direction is the most important predictor in addition to wind speed (Figure 2). It is followed by temperature and relative humidity parameters, which carry more weight, especially in the more complex terrain, such as the Alpine area (not shown). The pressure and precipitation parameters are often optimized with the 0.00 weight, meaning that they are not carrying additional benefits at certain stations. But that is not always the case. For instance, the precipitation parameter often is optimized by weight value 1, mostly at the southeastern side of the Alpine area, where convective precipitation often occurs (not shown).

Supplementary to using the mean value of 17 ALADIN-LAEF ensemble members for each meteorological parameter, the standard deviation of those 17 members can also be used as an additional predictor. Thus, the information on the starting model ensemble uncertainty is included in the analog search. The standard deviation predictors are optimized as one multiplying factor to all the pre-calculated weights for meteorological parameters, independently for each location. Five possible values of this multiplying factor are tested: 0.2, 0.4, 0.6, 0.8 and 1.0. If using none of the values results in a forecast improvement, the value 0.0 is used as the best fit. In the following illustrative example, it is assumed that the optimal weight for the ALADIN-LAEF temperature ensemble mean predictor is 0.75 at a particular location. Similarly, the weight for the relative humidity is optimized as 0.50, for precipitation as 0.00, etc. Then, the weight for the six ALADIN-LAEF ensemble standard deviation predictors is optimized as 0.2. The w_i in Equation 4 would be 0.2×0.75 for the temperature standard deviation predictor, 0.2×0.50 for the relative humidity standard deviation predictor, 0.2×0.00 for the precipitation, etc. The distribution for the optimized standard deviation multiplying factors is given in Figure 2. The result shows that the optimal contribution of the standard deviation predictors is about 40% of the ensemble mean predictors' contribution to the majority locations tested.

The AnEn can be affected by a conditional negative bias, especially when predicting events in the right tail of the forecast distribution. For that reason, a novel bias correction method is applied, as proposed by Alessandrini *et al.* (2019). The method is based on a correction factor proportional to the linear regression coefficient between the wind speed observations and raw model forecast (i.e. ALADIN-LAEF wind speed ensemble mean) during training, as well as to the distance between the current raw model forecast and the average value of the previous raw model forecasts that correspond to the currently selected analogs in the AnEn. The lead-time-independent correction factor is added to all the members of the AnEn if the current raw model forecast is above a certain threshold value. If the threshold is set too low, the bias correction adjustment can become small and noisy, leading to forecast performance degradation. After the simple AnEn testing (minimizing the RMSE; not shown), the 95th percentile of the climatological raw model forecast distribution (during the training period) is chosen as a threshold in this work.

3.3 | Description of experiments

In total, six different input configurations using the observations and the ALADIN-LAEF ensemble data are investigated in this study. All six investigated configurations provide an analog ensemble forecast consisting of the past observation corresponding to the 17 most similar past ALADIN-LAEF ensemble predictions. Thus, the new analog-based ensemble forecast provides the 17 ensemble members, equivalent to the original ALADIN-LAEF model. The chosen ensemble size does not only reflect the input NWP ensemble but is close to the optimal size of 15 members for the deterministic application of the analog ensemble found by Odak Plenković *et al.* (2018).

Dabernig *et al.* (2015) show the value of an ensemble forecast compared to its deterministic control run. Therefore, the first experiment, the *AnEnCtrl*, uses the ALADIN-LAEF control member for the six meteorological parameters available as six predictors. The *AnEnWs* uses all 17 ALADIN-LAEF ensemble member wind speed predictions (*LAEFws*) as 17 predictors. More meteorological variables are exploited in the *AnEnMu* experiment. In contrast to the *AnEnWs*, in the *AnEnMu* experiment, only the ensemble mean μ for every parameter is used as a predictor. For the *AnEnStd* ensemble forecasts, the ALADIN-LAEF ensemble uncertainty (σ) and the ensemble mean (μ) of the defined six meteorological parameters are used. The *AnEnStd* includes the aspects of error growth, represented dynamically by the ensemble model used, as explained in Eckel and Delle Monache (2016). This adds additional information to the flow-dependent error growth already captured by the analog approach (e.g. in *AnEnMu*).

In addition to the aforementioned experiments, two diverging ways of including all the ALADIN-LAEF information available are investigated. The first additional experiment, the *AnEnAll*, uses every member of the ALADIN-LAEF ensemble for every defined meteorological predictor. Thus, in this study, 6 variables and 17 ensemble members are used, which equals 6 × 17 predictors. An important goal of this study is to evaluate if all probabilistic information is needed or summary measures, such as mean or spread, are already sufficient. The second additional experiment is the “member by member” approach *AnEnMem*. Here, the analog search procedure is carried out for every ALADIN-LAEF member separately. Therefore, each raw model member is now distinguishable from the others. The analog-search procedure is independently done for each set of six pre-defined meteorological parameters, corresponding to the same raw model member. Thus, in our study, the search procedure is performed 17 times in total. Only one analog is chosen in every analog search procedure per ensemble member, with verifying observation chosen as the member in the *AnEnMem* ensemble. This is the most demanding configuration presented in this study. An analog experiment similar to the *AnEnMem* experiment, but using more than one analog (e.g. five analogs) for each of the ALADIN-LAEF ensemble members, is also investigated. However, besides being even more computationally demanding, it did not provide any benefits justifying the additional computational costs. Therefore, these results are not shown here.

All experiments use an analog search time window fixed at every lead-time individually, including one time step before/after to account for a trend and produce a 17-member ensemble output.

To determine if the difference in scores between the experiments is statistically significant, the moving-block bootstrap technique, following the procedure of Wilks (1997) and using 1,000 re-samples at a confidence level of 95%, was applied, except for correlation where pair bootstrap technique was used (Wilcox, 2009; see his section 4.2).

The forecast performance of the different experiments is evaluated in the following section.

4 | RESULTS

4.1 | Overall results

In total, six different analog-based ensemble experiments (see Table 1 for a summary) are carried out in this study. Results are evaluated against observations, the raw ensemble model, the ALADIN-LAEF (*LAEFws*) and the variations of the *EMOS* forecasts. The novelty of this approach is the usage of different types and set-ups of the probabilistic input model to give new insights into the analog-based methodology. Summarizing, all analog forecasts show an improvement compared to the raw forecasts (Table 2). Moreover, most analog forecasts perform similar or even better than the *EMOS* methods. Furthermore, distinct differences between the analog configurations are found.

The source of error of a model can be specified when decomposing the RMSE to the bias of the mean (or simply bias), the bias of the standard deviation (σ bias) and the dispersion (phase) error (e.g. Murphy, 1988; Horvath *et al.*, 2012):

$$RMSE^2 = (\overline{F} - \overline{O})^2 + (\sigma_f - \sigma_o)^2 + 2\sigma_f\sigma_o(1 - r_{fo}), \qquad (6)$$

where F represents forecast, O observations, σ_f is the standard deviation of the forecast F, σ_o is the standard deviation of observations O, and r is the correlation coefficient between the forecast and observed data; all referring to the same period (i.e. month). In this work, the σ bias is defined

TABLE 2 The average values and confidence interval (0.95 sig. level) of several verification measures for the different forecasts at all available stations and all lead-times during January (top half) and July (bottom half), 2018. The best result among compared forecasts is underlined (the spread is better when closer to the RMSE value). The values significantly different from the *AnEnStd* forecast (0.05 sig. level) are marked with an asterisk

January	*LAEFws*	*EMOSws*	*EMOSstd*	*AnEnCtrl*	*AnEnWs*	*AnEnMu*	*AnEnStd*	*AnEnAll*	*AnEnMem*
Bias [m·s^{-1}]	−0.210* [−0.232, −0.185]	−0.053* [−0.069, −0.039]	−0.160* [−0.174, −0.146]	−0.060* [−0.072, −0.046]	−0.036 [−0.048, −0.022]	−0.029 [−0.042, −0.016]	−0.023 [−0.035, −0.011]	−0.061* [−0.075, −0.048]	−0.048* [−0.061, −0.034]
CC	0.378* [0.371, 0.385]	0.831* [0.826, 0.835]	0.841* [0.837, 0.845]	0.841* [0.837, 0.845]	0.845* [0.841, 0.849]	0.861* [0.858, 0.865]	0.863 [0.858, 0.865]	0.863 [0.860, 0.867]	0.856* [0.852, 0.860]
Disp. Err [m·s^{-1}]	2.670* [2.645, 2.696]	1.801* [1.784, 1.826]	1.705* [1.681, 1.733]	1.694* [1.672, 1.715]	1.705* [1.682, 1.727]	1.613 [1.593, 1.633]	1.608 [1.589, 1.626]	1.596* [1.573, 1.618]	1.634* [1.612, 1.654]
σ bias [m·s^{-1}]	−1.501* [−1.545, −1.458]	−0.322* [−0.378, −0.278]	−0.454* [−0.505, −0.404]	−0.495* [−0.546, −0.444]	−0.391* [−0.444, −0.340]	−0.386 [−0.438, −0.328]	−0.372 [−0.433, −0.314]	−0.405* [−0.455, −0.352]	−0.420* [−0.483, −0.367]
RMSE [m·s^{-1}]	3.070* [3.029, 3.111]	1.831* [1.812, 1.851]	1.772* [1.748, 1.795]	1.766* [1.743, 1.792]	1.749* [1.729, 1.771]	1.659 [1.639, 1.677]	1.650 [1.632, 1.672]	1.647 [1.624, 1.667]	1.688* [1.670, 1.707]
Spread [m·s^{-1}]	0.850* [0.846, 0.854]	1.611* [1.599, 1.622]	1.605* [1.592, 1.617]	1.776* [1.750, 1.779]	1.663 [1.650, 1.675]	1.672 [1.660, 1.686]	1.667 [1.655, 1.679]	1.641* [1.629, 1.654]	1.728* [1.714, 1.742]
BSS (>5 m·s^{-1})	−0.075* [−0.093, −0.059]	0.490* [0.479, 0.500]	0.515* [0.505, 0.524]	0.520* [0.510, 0.529]	0.513* [0.504, 0.523]	0.546 [0.537, 0.555]	0.549 [0.541, 0.558]	0.555 [0.546, 0.563]	0.526* [0.517, 0.535]
CRPS [m·s^{-1}]	1.631* [1.613, 1.648]	0.883* [0.875, 0.892]	0.823* [0.815, 0.831]	0.814* [0.806, 0.820]	0.823* [0.816, 0.831]	0.777 [0.770, 0.784]	0.772 [0.765, 0.779]	0.769 [0.762, 0.776]	0.816* [0.809, 0.823]

TABLE 2 Continued

July	*LAEFws*	*EMOSws*	*EMOSstd*	*AnEnCtrl*	*AnEnWs*	*AnEnMu*	*AnEnStd*	*AnEnAll*	*AnEnMem*
Bias [m·s^{-1}]	−0.229* [−0.242, −0.215]	−0.001* [−0.008, −0.010]	−0.119* [−0.129, −0.111]	−0.012 [−0.021, −0.001]	−0.090* [−0.099, −0.080]	−0.055 [−0.063, −0.046]	−0.063 [−0.072, −0.054]	−0.088* [−0098, −0.080]	−0.043* [−0.053, −0.033]
CC	0.415* [0.406, 0.422]	0.750* [0.745, 0.754]	0.764* [0.759, 0.768]	0.752* [0.748, 0.757]	0.739* [0.735, 0.744]	0.770* [0.766, 0.774]	0.774 [0.769, 0.778]	0.774 [0.770, 0.778]	0.759* [0.754, 0.763]
Disp. Err [m·s^{-1}]	1.602* [1.589, 1.616]	1.229* [1.215, 1.240]	1.144* [1.132, 1.154]	1.229* [1.216, 1.241]	1.262* [1.250, 1.273]	1.156* [1.144, 1.167]	1.145 [1.136, 1.157]	1.148* [1.138, 1.159]	1.183* [1.172, 1.194]
σ bias [m·s^{-1}]	−0.773* [−0.794, −0.754]	−0.344* [−0.368, −0.325]	−0.474* [−0.494, −0.452]	−0.344* [−0.364, −0.323]	−0.331* [−0.353, −0.308]	−0.400* [−0.418, −0.377]	−0.409 [−0.429, −0.387]	−0.396* [−0.416, −0.375]	−0.403* [−0.423, −0.383]
RMSE [m·s^{-1}]	1.794* [1.775, 1.813]	1.276* [1.262, 1.288]	1.244* [1.234, 1.256]	1.272* [1.261, 1.284]	1.307* [1.294, 1.321]	1.225 [1.213, 1.237]	1.219 [1.208, 1.229]	1.218 [1.206, 1.228]	1.251* [1.238, 1.262]
Spread [m·s^{-1}]	0.651* [0.648, 0.654]	1.170* [1.164, 1.176]	1.138* [1.133, 1.144]	1.318* [1.311, 1.326]	1.256* [1.248, 1.263]	1.253 [1.246, 1.261]	1.244 [1.236, 1.250]	1.190* [1.184, 1.197]	1.301* [1.294, 1.308]
BSS (>5 m·s^{-1})	0.032* [0.009, 0.055]	0.329* [0.314, 0.345]	0.337 [0.322, 0.353]	0.329* [0.313, 0.344]	0.319* [0.303, 0.335]	0.349 [0.334, 0.365]	0.355 [0.341, 0.369]	0.353 [0.338, 0.369]	0.325* [0.310, 0.340]
CRPS [m·s^{-1}]	1.032* [1.022, 1.042]	0.648* [0.643, 0.653]	0.624* [0.619, 0.629]	0.636* [0.631, 0.641	0.650* [0.645, 0.656]	0.613 [0.608, 0.618]	0.610 [0.605, 0.615]	0.612 [0.606, 0.617]	0.635* [0.630, 0.640]

as the bias of the standard deviation of the ensemble mean (regardless of the ensemble spread).

Results show that the average bias of the *LAEFws* ensemble is small, underestimating the wind speed by 0.21 m·s^{-1} in January and 0.23 m·s^{-1} in July. The same results are found for the σ bias in July with 0.77 m·s^{-1}, while it is a slightly more dominant source of error in January with −1.50 m·s^{-1}. Also, the other evaluated scores such as the correlation coefficient (CC), which is on average higher in July than in January with 0.37, or the RMSE with 3.07 m·s^{-1} in January and 1.79 m·s^{-1} in July, indicate that the *LAEFws*, in general, has realistic results, especially for the summer month. However, there are still some unresolved processes, as can be seen by the results of the dispersion error.

The main aim of any kind of the NWP model post-processing is improving the results of the original model. This is the case here, too. The *EMOS* post-processing experiments are applied successfully, reducing all three error sources: the bias, the σ bias and the dispersion error in comparison to *LAEFws*. The *EMOSws* is more successful in removing a systematic source of the error, while the *EMOSstd* is better at removing the dispersion error. All six analog-based experiments are able to outperform the *LAEFws* as well. Specifically, they can reduce all three error sources for the ensemble mean. Already the first and most "simple" experiments in terms of input data, the *AnEnCtrl* and the *AnEnWs*, successfully remove the systematic errors in the bias and σ bias similar to the *EMOS* approach. Even more successful in removing the predominant dispersion source are the experiments with the additional predictors: *AnEnMu*, *AnEnStd* and *AnEnAll*.

In addition to improving the results for the ensemble mean, the average ensemble spread matches the average RMSE better after any post-processing. The *AnEnStd* exhibits the best spread among analog-based experiments in July, while *AnEnAll* shows better results in January. This might be related to the fact that wind speed shows greater variability (higher standard deviation of observations) and is probably harder to predict correctly in January. For that reason, using more information from the raw model adds more variety to the ensemble members. This result also indicates that in the convective season most likely a horizontally and vertically higher-resolved convection-permitting NWP model might add some additional information not present in the coarser *LAEFws*.

The Brier Skill Score (BSS) is a commonly used metric for the probabilistic forecast of a binary event that uses climatology as a reference (Jolliffe and Stephenson, 2011; Wilks, 2011). It is calculated using the following expression:

$$BSS = 1 - BS/BS_{\mathrm{clim}}, \tag{7}$$

where the Brier score ($BS = \sum_i (p_i - o_i)^2/n$) averages the squared differences between pairs of forecast probabilities p and the subsequent binary observations o over all n forecast–observation pairs.

A binary event is defined using an exceedance threshold, that is, of wind speed forecasted higher than 5 m·s^{-1}. The closer the BSS is to the perfect number 1, the better the skill of the forecast is. Here, a threshold of 5 m·s^{-1} is chosen for the BSS as it is reasonably high while, on the other hand, not being too rare. In the selected two months, the observed frequency of the wind speed exceeding 5 m·s^{-1} is higher for January with 18% cases than for July with 9%. Based on these observed numbers, the BSS value of the original ensemble (*LAEFws*) is −0.08 for January and 0.03 for July, indicating that the small differences are already present in the input data. It is shown that the BSS is improved by all post-processing experiments. This is especially the case in January, where the underlying climatology shows that the higher wind speed is more frequently observed than in July and the wind speed variance (higher standard deviation of observations) is higher. The *AnEnMu*, *AnEnStd* and *AnEnAll* experiments show a nearly similar improvement. The other post-processing approaches improve BSS less.

The continuous rank probability score (CRPS) is a summary metric that can be interpreted as the integral of the Brier score over all possible threshold values for the parameter under consideration:

$$CRPS = \int_{-\infty}^{\infty} [P_{\mathrm{f}}(x) - P_{\mathrm{o}}(x)]^2 \, dx, \tag{8}$$

where P_{f} stands for forecasted probability (cumulative distribution), while P_{o} is a cumulative-probability step function that jumps from 0 to 1 at the point where the forecast variable equals the observation. The CRPS is a negatively oriented (the lower, the better) accuracy measure that is equivalent to the mean absolute error for deterministic forecast and also has a value of 0 for the perfect forecast. The *LAEFws* shows a higher CRPS (1.63 m·s^{-1}) for January than for July (1.03 m·s^{-1}). Again, the CRPS value is improved by all post-processing experiments, exhibiting better overall results for July than in January, when wind speed and its variance is higher on average. Similar to the BSS, the *AnEnAll* shows the highest skill during the winter month, while the *AnEnStd* is slightly better during the summer month. This indicates that adding more input from the raw model does not just increase the ensemble spread, but it also improves its accuracy. The *AnEnMu* follows both *AnEnAll* and *AnEnStd* results closely. The other post-processing experiments are not as successful (see Table 2), exhibiting significantly worse overall results for both months investigated.

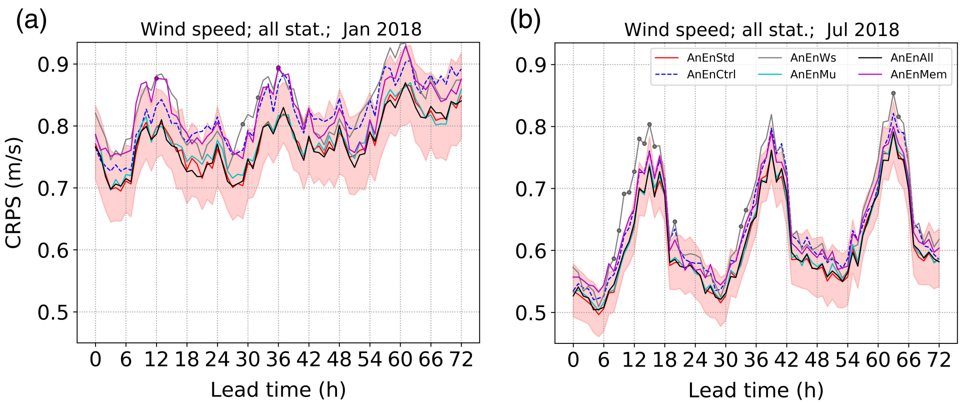

FIGURE 3 Continuous rank probability score (CRPS) depending on lead-time for five different analog-based ensemble experiments during (a) January and (b) July 2018 at all stations tested in this study. The markers are set for the results significantly different from the *AnEnStd* forecast (95% confidence level), while the red shaded area represents the *AnEnStd* 95% confidence interval calculated by bootstrap percentile method (Jolliffe, 2007)

4.2 | Lead-time performance

To investigate six analog-based ensemble experiments comparison further, a summary score CRPS is considered for the individual lead-times (Figure 3). The result shows that there is no significant difference between the *AnEnMu*, *AnEnStd* and *AnEnAll* performance during either winter or summer month. The *AnEnCtrl*, *AnEnWs* and *AnEnMem* are slightly outperformed by other analog-based experiments, especially for January. Even though the *AnEnCtrl*, *AnEnWs*, and *AnEnMem* are able to improve the raw NWP forecasts, comparable to the *EMOS* approach, they are less promising than other analog-based experiments. The *AnEnWs* results show that it is essential to use more than one meteorological variable as a predictor in the analog approach. This can be explained by the better ability of the analog method to distinguish different seasonal and synoptic situations. The analog-search pool in the *AnEnMem* experiment is smaller than in other analog experiments since the search is performed dependently for the same ensemble member. Possibly, that is why the *AnEnMem* would not increase the skill of the raw probabilistic input, as one would inherit undesirable properties of the input model, such as under-dispersion and lower-resolution issues. Additionally, *AnEnMem* is the most computationally expensive set-up. For these reasons, it is not shown or discussed further in this article. Finally, even though the *AnEnCtrl* and the *AnEnMu* use the same number of the meteorological parameters as predictor variables, the *AnEnMu* performs better for both months and at all lead times tested. Similar results are shown in Dabernig *et al.* (2015), where the *EMOS* results based on ensemble forecasts outperformed the forecasts using only the control run.

Overall, the *AnEnAll* performs the best in post-processing for January whereas the *AnEnStd* set-up performs the best for July. Similar can be concluded using additional metrics, such as BSS, reliability, spread-skill and relative operating characteristic (ROC) diagram (not shown). Among these experiments with a similar result, the *AnEnStd* is chosen as the best representative. The reason for this decision is that it is not as computationally demanding as the *AnEnAll*, while it includes the information about raw model spread (unlike the *AnEnMu*). The information about the raw model error growth is considered a very important aspect of the raw NWP ensemble forecast. Therefore, it is expected to be further developed in the near future, leading to greater differences between the *AnEnMu* and *AnEnStd* experiments. To determine if using summarized predictors, such as in the *AnEnStd* experiment, leads to information loss and decreases the forecast quality, the results are compared to the *AnEnAll* experiment.

In addition to overall comparison, the *AnEnStd* and *AnEnAll* experiments are also compared against the two different *EMOS* experiments and the *LAEFws*, separated into lead-times using several verification metrics.

The CRPS shows that the *LAEFws* exhibits a higher skill during daytime (i.e. 0600–1800 UTC) than during night-time, and higher during July (Figure 4b) than during January (Figure 4a). The *EMOS* and the analog-based experiments are more skilful during night-time than during daytime. The improvement over the *LAEFws* after post-processing is greater in January for both the *EMOS* and the analog approach since the *LAEFws* is worse than in July. However, the *EMOS* and the analog experiments are overall better in July, when the *LAEFws*, which also served as input, is better. These results imply that the best result is achieved when the input model is also working better. The *AnEnStd* and *AnEnAll* show almost no difference. They are both more skilful than the two *EMOS* experiments. Even though the differences are often subtle, they are significant for the *EMOSws* at almost all lead-times during January and at several lead-times during July, especially within the first 24 hr.

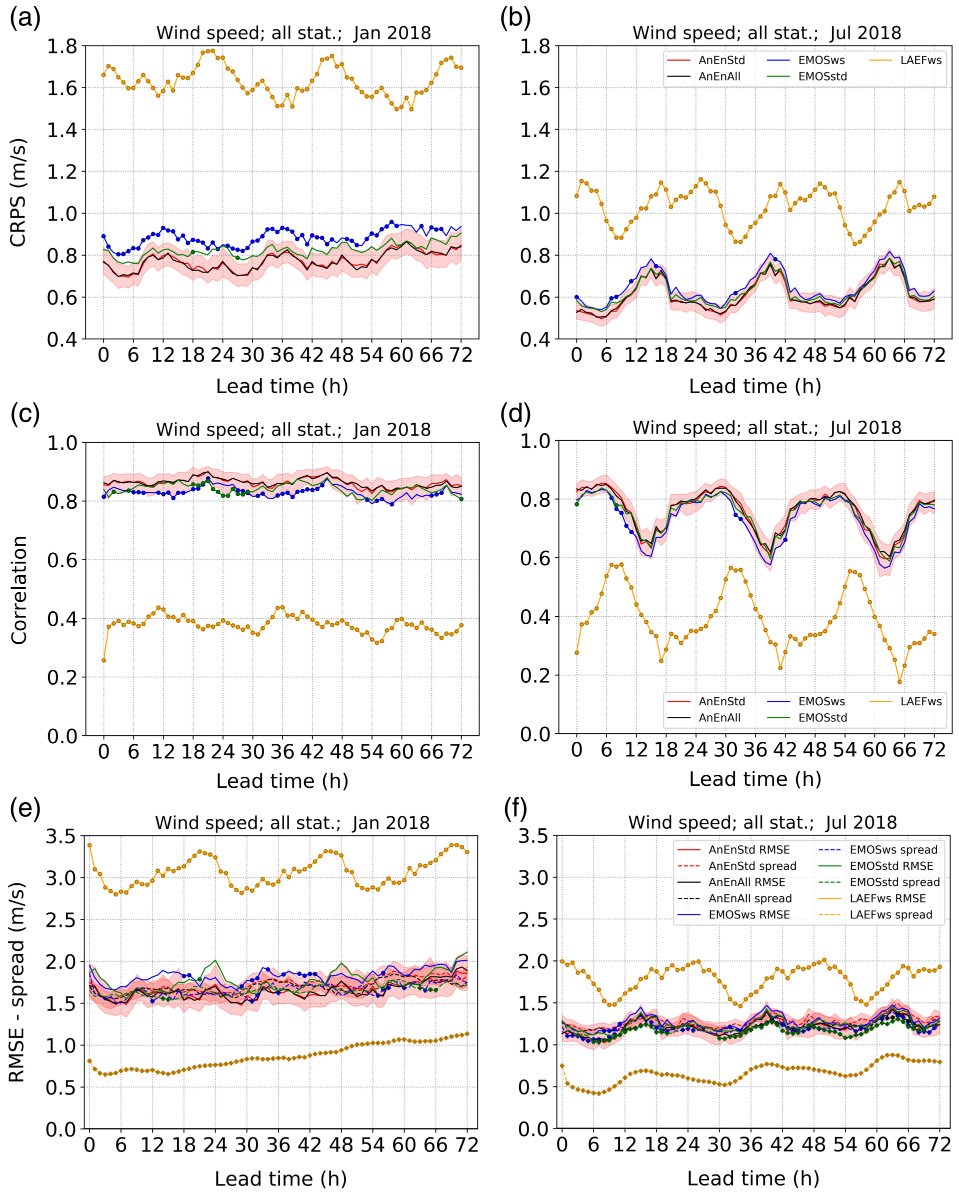

FIGURE 4 (a and b) Continuous rank probability score, (c and d) the correlation coefficient for the ensemble mean and (e and f) the spread-skill diagram depending on lead-time for the raw *LAEFws* ensemble, the *EMOS* and two different analog ensemble configurations at all the stations tested for January (left) and July (right) 2018. The markers are set for the results significantly different from the *AnEnStd* forecast (95% confidence level), while the red shaded area represents the *AnEnStd* 95% confidence interval calculated by bootstrap percentile method (Jolliffe, 2007)

Evaluating the dependency on the lead-time, the analog post-processing methods show considerable improvement over the *LAEFws* for both months tested with the CC (Figure 4c,d). The analog approach outperforms the *EMOS* methods in terms of correlation, often significantly. This is especially the case for January when the CC enlargement over *EMOSws* is significant for almost all lead-times and sometimes even over *EMOSstd* (i.e. during night-time).

The analog-based forecasts significantly reduce the *LAEFws* RMSE at all lead-times (Figure 4e,f), similarly to the *EMOS* approach, with very few significant differences. The improvement is the most evident for the *LAEFws* RMSE maxima at 0000 UTC.

Similar results can be found in the spread-skill diagrams. These diagrams test if the average ensemble spread matches the average RMSE, representing the forecast uncertainty appropriately. All post-processing methods satisfactorily increase the spread. Here, both analog-based forecasts are showing an almost perfect agreement between the RMSE and the spread, while the *EMOS* experiments are slightly under-dispersive, especially the *EMOSws* in January (Figure 4e). This can be related to the fact that it uses only the wind speed as a predictor and most likely, not enough dispersion information is given. Additionally, the *EMOSws* only uses a 30-day training window, which also results in a small under-dispersion.

4.3 | Spatial performance

The climatology in Figure 1 shows that the wind speed increases towards the northeastern part of Austria

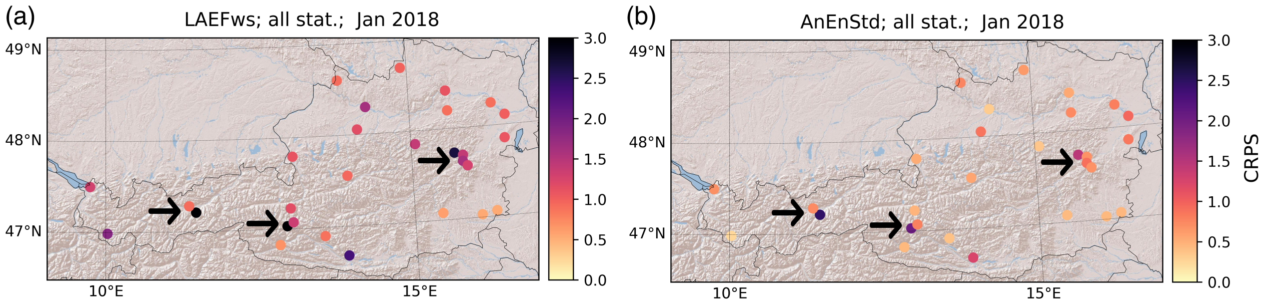

FIGURE 5 The spatial distribution of the monthly mean continuous rank probability score for (a) the raw *LAEFws* and (b) the *AnEnStd*, for January 2018. The arrows point to closely situated stations in the highly complex terrain, whereas the valley stations exhibit much better results than the mountain stations

(Pannonian plain) for both January and July, which also suggests a spatial pattern in forecast performance. Within this subsection, it is decided to show only results for January since the previous results suggested the better distinction in the performance after post-processing. Even though not shown here, the spatial distribution of results for July is very similar to the ones for January.

Additionally, due to very subtle and hardly noticeable differences among analog experiments, only the *AnEnStd* configuration is shown as a representative. The results for the *AnEnMu* and the *AnEnAll* experiments are almost indistinguishable from the *AnEnStd*, while the *AnEnCtrl*, *AnEnWs* and the *AnEnMem* are the same or slightly worse. Since the results for these experiments carry no new information within this subsection, they are not shown from this moment on.

The value for the *LAEFws* monthly mean CRPS follows the climatological wind speed pattern, having higher values at the stations prone to higher winds. The error is reduced for the analog experiments (Figure 5b) compared to the *LAEFws* (Figure 5a), following a similar pattern. Additionally, there are large differences for the nearby stations situated in highly complex terrain. The mapped CRPS values for any forecast tested show that the valley stations are better predicted than the mountain stations (arrows). The plains are better represented by the ALADIN-LAEF topography and, therefore, the performance of ALADIN-LAEF is, in general, better at lower altitudes and less complex terrain. This results in the *LAEFws* not being as successful at the mountain stations. A close look at the two stations in Innsbruck (arrow in the west of Austria) shows, for example, that the *AnEnStd* CRPS at the valley station is improved by around 20% compared to the *LAEFws*. As the *LAEFws* performance at mountain stations is not as efficient, this leaves room for improvement. Here, the CRPS can be improved by around 70% at for example, Patscherkofel, the mountain station close to Innsbruck. A similar pattern is shown at the station Sonnblick (arrow in the middle) where the mountain station has much higher CRPS values (raw and post-processed) compared to the valley station. As an example, for the three sites located in the Semmering region (most eastern arrow), a mountain pass in the east of Austria, the different settings of the sites can be one of the factors. The site located at the pass is prone to gap flows (e.g. Mayr *et al.*, 2007), whereas the site at the mountaintop is located within the skiing resort, somewhat shielded by the nearby hut and not represented in the model lower boundary conditions. The site located in the valley shows again the lower CRPS values. These differences in predictability are mainly related to the high wind speeds and the coarse resolution of the raw model. This suggests a large sensitivity of the models in the Alpine complex terrain to the exact details of the mountain height and shape, as well as the incoming background layer, where subtle differences can result in a large range of responses in the downslope wind regime. In contrast, the stations in the northeast of Austria (around Vienna) are also climatologically prone to high wind speeds but show much better CRPS values.

In order to evaluate the performance for valley and mountain stations, the stations marked with arrows (Figures 1 and 5) are investigated separately. For the valley stations, the RMSE (1.50 m$\cdot$s^{-1}) shows that the *LAEFws* wind speed prediction performs adequately. However, for mountainous sites, the RMSE is 6.24 m$\cdot$s^{-1}, due to the aforementioned reasons. The RMSE is notably reduced by the analog approach, by 0.45 m$\cdot$s^{-1} in the valley and by 3.33 m$\cdot$s^{-1} at mountain stations. The RMSE decomposition (Figure 6) shows that the dispersion error is notably reduced by the analog approach, slightly more for the mountain than the valley sites. The *LAEFws* exhibits much larger systematic errors for the mountain than the valley stations. The *LAEFws* bias and the σ bias at the valley stations are very small, to begin with. The analog approach is therefore not able to make a large difference after post-processing. On the other hand, the *LAEFws* systematic sources of error at the mountain stations are much

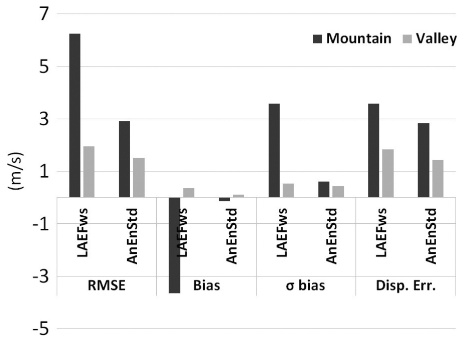

FIGURE 6 The *AnEnStd* and the *LAEFws* forecasts performance comparison at mountain and valley stations by root-mean-square error (RMSE) decomposition into bias, σ bias and dispersion error during January 2018

more pronounced than at the valley stations. These sources of error are yet again successfully removed by the analog approach. The RMSE reduction is therefore much more noticeable for the mountain stations than for the valley stations, due to the reduction of systematic sources of error, which are not as present in the raw model for the valley stations. However, the spatial distribution of forecast performance could be further investigated in future work.

4.4 | Special diagrams: Reliability, ROC and rank histograms

The reliability of a probabilistic forecast is the property of that forecast to predict probabilities that match the relative frequencies within the data. Here, it is evaluated for the probability of wind speed exceedance of >5 m·s^{-1}. Again, the *LAEFws* ensemble has lower reliability in January (Figure 7a) than in July (Figure 7b). Furthermore, it is below the no-skill line for the high probabilities in January. Both *EMOS* experiments improve *LAEFws* reliability, *EMOSstd* improving a bit more than *EMOSws*. However, the analog experiments show an even higher resolution and reliability across all experiments, especially for the winter month. The differences can be noticed for the probabilities up to a 50% chance of wind speed to exceed 5 m·s^{-1}, where the *EMOSstd* is slightly underconfident, or for the probabilities with a more than 40% chance, where the *EMOSws* is slightly overconfident. Between the analog experiments, only small and insignificant differences are found. Both analog-based experiments exhibit almost perfect reliability for the winter month, while being slightly overconfident during summer.

Besides a higher resolution of the analog experiments, one can notice that the sharpness diagram (upper-left corner of the reliability diagram) is reasonable for all approaches. However, the *LAEFws* is a bit sharper than the post-processing experiments, indicating a higher tendency to forecast extreme probabilities. This is preferable because of the better forecast usability if the forecasts are reliable. Still, the post-processing experiments are overall more accurate in terms of reliability.

The ROC curve shows a ratio of hit rate versus false-alarm rate using a pre-defined threshold. Again, the threshold of 5 m·s^{-1} is used. The ROC curve (Figure 7c,d) indicates that the analog methods, in general, improve the raw *LAEFws* forecasts comparable to or better than the *EMOS*. Unlike other measures, the reliability and discrimination property exhibit higher values for January than for July. However, this might be due to the higher climatological frequency of such wind speeds in January (18%) than in July (9%). For that reason, the differences among winter and summer months should not be investigated by using the fixed threshold. The results should be used for comparison among different experiments. The *AnEnStd* exhibits a slightly higher hit rate than the *AnEnAll* and *EMOS* experiments, especially for July.

Evaluating the rank histogram (Figure 8), a clear under-dispersion of *LAEFws* is found, especially for January. This is not the case for the post-processed forecasts. It shows that the analog method is able to improve the dispersion of the original NWP ensemble.

Finally, it is shown that the analog approach outperforms the raw *LAEFws* model in terms of better accuracy, reliability, resolution, discrimination and spread for both winter and summer months. The results are very similar to or better than the *EMOS* experiments shown, with the larger differences during the winter month. The difference among analog experiments (*AnEnAll* and *AnEnStd*) is barely notable. Therefore, it is indicated that using the summarized metrics of the raw model meteorological variable ensemble as a predictor in the analog approach barely sacrifices the forecast quality, while saving computational power.

4.5 | High wind speed predictions

The majority of measured wind speed values during the selected months are within 2–3 m·s^{-1} range (30–40%), while wind speeds higher than 5 or 10 m·s^{-1} are rare (Figure 9c,d). However, it is still important to properly forecast higher wind speeds because of their higher impact on people, damage to property, road and air traffic disruptions, wind energy production, etc. For this reason, it is important that a probabilistic forecast is consistently good for several different thresholds. Besides the exceedance of 5 m·s^{-1} the thresholds ranging from 0.5 to 20 m·s^{-1} are investigated (Figure 9a,b).

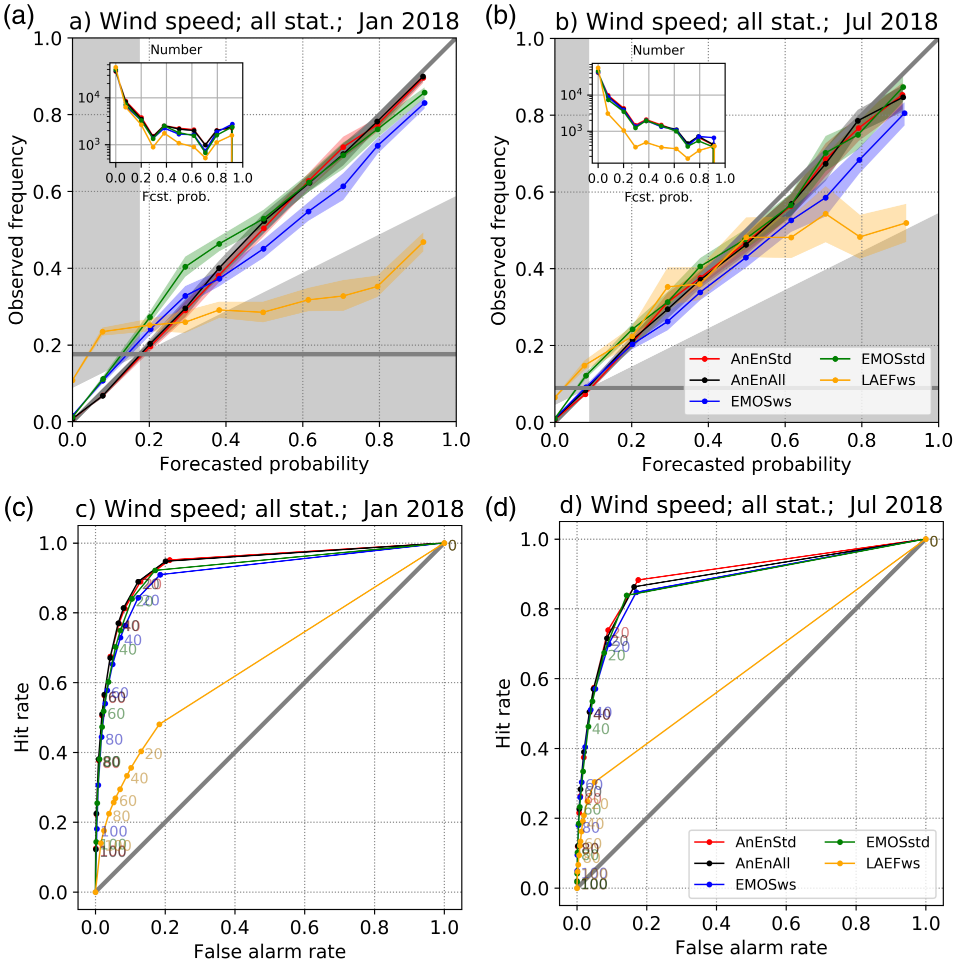

FIGURE 7 Reliability diagrams and relative operating characteristic (ROC) diagrams for two different analog forecasts and a threshold of >5 m·s^{-1}, compared to the raw *LAEFws* and the *EMOS* during (a) January and (b) July 2018 at all stations tested in this study. The dashed lines in the reliability diagrams show 95% confidence interval, while the sharpness diagrams are shown in the upper-left corners

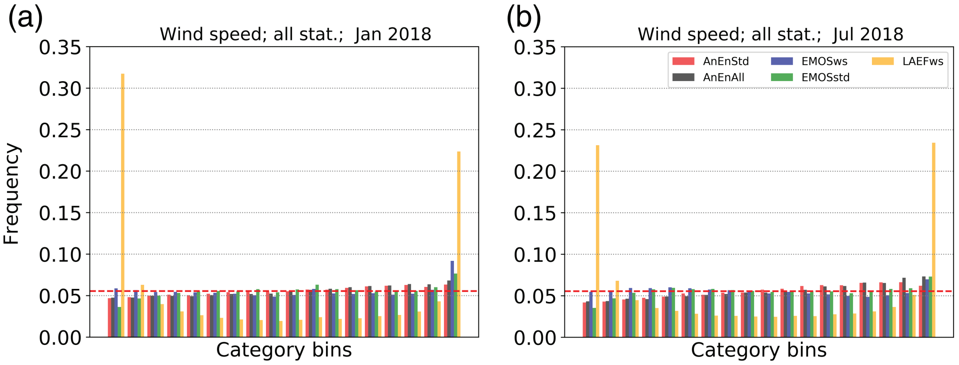

FIGURE 8 Rank histograms for the *AnEnStd* and *AnEnAll* forecasts compared to the raw ALADIN-LAEF *LAEFws and EMOS* forecasts during (a) January and (b) July 2018 at all stations tested in this study

The BSS indicates that the *LAEFws* forecast is somewhat skilful in reproducing wind speeds of the order of 3 m·s^{-1}, but shows much less skill, if any, for the higher and lower thresholds. The *EMOS* approach is more skilful than the *LAEFws* for any threshold value in January and up to 10 m·s^{-1} (*EMOSws*) or even 15 m·s^{-1} (*EMOSstd*) in July. The analog experiments are able to improve the forecast skills up to 10 m·s^{-1} significantly better than the *EMOS* experiments. Approaches as in Baran and Lerch (2015) could be used to adjust EMOS to higher wind speeds but have not been tried. Furthermore, the *AnEnStd* and *AnEnAll* improve the *LAEFws* forecasts for all thresholds investigated for January. Again, the *AnEnCtrl*, *AnEnWs* and *AnEnMem* do improve the *LAEFws* forecasts but are less skilful than the other analog experiments (not shown). However, *AnEnWs* still provides a good result. It is, thus, the recommended approach if only a reduced set of ensemble data is available or the computational resources are limited. These results reveal the potential for post-processing using the analog approach, even though one needs to be careful with the interpretation since the number of

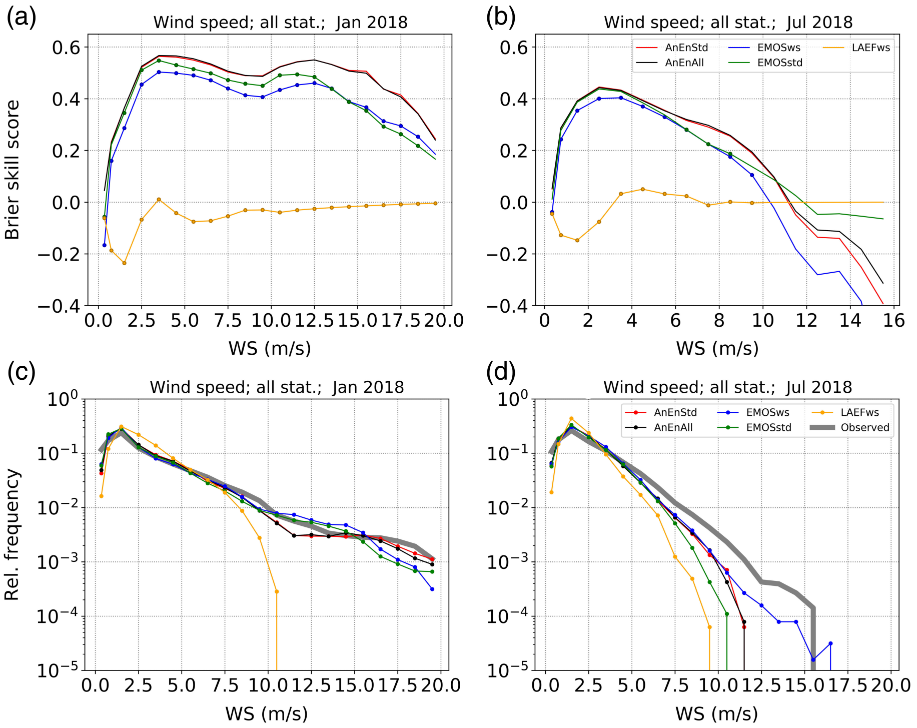

FIGURE 9 (a and b) Brier skill score and (c and d) relative frequency depending on a wind speed threshold. The analog probabilistic forecasts are compared to the raw *LAEFws* and the *EMOS* forecasts during January (left) and July (right) 2018 at all stations tested in this study. The markers are set for the BSS results significantly different from the *AnEnStd* forecast (95% confidence level)

occurrences of high wind speed (i.e. around 20 m·s^{-1}) is very small.

5 | SUMMARY AND CONCLUSIONS

In this study, the analog method is applied for the sites in Austria using a numerical weather prediction (NWP) ensemble as input. The aim of this work is to significantly improve the NWP ALADIN-LAEF ensemble forecasts for the 10 m wind speed (*LAEFws*) while maintaining low computational costs for the analog search. For that reason, several experiments using different forecast information of the Austrian ALADIN-LAEF ensemble as input to the analog method are thoroughly analysed. The experiments use 29 TAWES sites in Austria for winter (January 2018) and summer (July 2018). Using an NWP ensemble enables the use of more meteorological variables (predictors) in more than one realization as input to the analog search. In addition, using summarized information such as the ensemble mean and/or the standard deviation or every single ensemble member can provide useful insights. In total, six experiments are conducted in this study using a different set of input information from the ALADIN-LAEF model as predictors to the analog-based method. The choice of predictors from raw NWP model includes:

- The ensemble control member of all available parameters (*AnEnCtrl*)
- All wind-speed raw forecast ensemble member (*AnEnWs*)
- The ensemble mean of all available parameters (*AnEnMu*)
- The ensemble mean and spread of all parameters (*AnEnStd*)
- All ensemble members of all parameters (*AnEnAll*)
- All available parameters corresponding to only one (distinguishable) ensemble member (*AnEnMem*),

where the abbreviations for analog experiments are listed in the brackets.

In addition to the thorough *LAEFws* forecast improvement by the analog-based method, the experiments provide a thorough insight into subtler differences among the analog-based configurations. The additional hypothesis that using the summarized measures for several meteorological variables, such as the mean and the standard deviation of ensemble input, is tested and proved to be sufficient to improve the raw forecast. This computationally less demanding experiment shows no compromise on the accuracy of the post-processing in comparison to using all available ensemble members and variables. All experiments provide the 17 members wind speed analog ensemble forecast. To better understand the impact on the raw forecasts, the two experiments using the ensemble model output statistics post-processing approach (*EMOS*) are used as a baseline. The *EMOSws* only uses the last 30 days as training and only the wind speed as an input, whereas the *EMOSstd* uses all available training data and all variables including seasonal functions. The *EMOSws* is slightly more successful in removing the systematic source, the *EMOSstd* the dispersion source of the error.

Results show that all analog-based experiments improve the raw model forecast. However, the most computationally demanding "member by member" *AnEnMem* experiment proved to be the least successful. The undesirable properties of the raw model, such as under-dispersion and lower resolution, are inherited more easily for this than for the other analog experiments. That is probably due to the fact that the analog-search pool is smaller than when seeking among all members independently, as is the case in the other analog experiments. Using only one predictor variable as input (the 17 members of *LAEFws*) already improves the forecast skills and lowers the systematic error of the ensemble mean, better or comparable to the *AnEnMem* experiment. If the number of available parameters from the raw model is limited, the experiment using only wind speed ensemble members proved to be successful. Even better results are achieved when using more than one predictor variable. Therefore, similar or better results are achieved when using only the ensemble control member as input (*AnEnCtrl*). In addition, using more than one ensemble member within the analog search procedure improves results even more. However, it is shown that often there is no need to use the full input spectrum of a raw probabilistic model, that is, all ALADIN-LAEF members as predictors. Using basic information of an input ensemble, such as ensemble mean and standard deviation, improves the forecast skills almost as successful as using the full input spectrum of a raw probabilistic model as predictors, with very little significant differences, if any. Furthermore, it is computationally less demanding. The results confirm the hypothesis that the summary metric (e.g. mean and standard deviation) is the optimal way to add the aspects of error growth that can be represented dynamically by the input ensemble model to the flow-dependent error growth already captured by the analog approach. Therefore, it can be suggested as the most promising configuration among experiments tested in this work.

All post-processing experiments in this work provide better results than the raw input model, as expected, reducing the under-dispersion while increasing the reliability and discrimination. The best results for both analog approach and the *EMOS* are achieved in July when the raw model performs better. The raw model under-spread is almost completely removed by all experiments. The *EMOSws* approach is slightly under-dispersive, especially in January, probably due to using only the wind speed parameter and much shorter training than other post-processing experiments.

The accuracy of the ensemble forecast is measured by the root-mean-square error (RMSE) for the ensemble mean and the continuous rank probability score (CRPS). The analog-based experiments outperform the raw *LAEFws* forecast in terms of significantly better accuracy for all forecast lead-times during both winter and summer months. They are more skilful during night-time than during daytime. The analog-based method is comparable to or outperforms both *EMOS* experiments. The outperformance is noticed at short lead-times and during the winter month, especially in terms of correlation. The *EMOSws* is overconfident to a certain extent for the high-probability forecasts, while *EMOSstd* is underconfident for low-probability forecasts. The analog-based experiments are almost perfectly reliable. Additionally, discrimination is slightly better than the *EMOS* due to a higher hit rate. The difference among the analog experiments is less pronounced than when compared to the *LAEFws* and the *EMOS* experiments, confirming that using basic information of an input ensemble, such as an ensemble mean and a standard deviation, is often sufficient.

If considered spatially, the *LAEFws* error follows the climatological wind speed pattern, having higher values at the stations prone to higher winds. The accuracy is improved when compared to the raw model forecast, following a similar pattern. Additionally, even though an improvement over raw model forecast is evident, large differences among nearby stations are noticed in highly complex terrain. The valley stations seem to have more predictable weather, and the overall post-processing result is, therefore, better than at the mountain stations with the climatologically higher wind speeds. On the other hand, the relative improvement to the raw model is more pronounced at mountain stations due to the reduction of systematic sources of error by post-processing, which is not as present in the raw model for the valley stations.

Finally, it is very important to assess the post-processing performance for high wind speed because of the impact on people and property, even though the strong wind does not occur as often as the mild wind. For that reason, several thresholds ranging from 0.5 to 20 $\mathrm{m{\cdot}s^{-1}}$, are used to test the skill of the post-processed forecasts. The result shows that the *LAEFws* forecast is skilful in reproducing wind speeds of the order of 3 $\mathrm{m{\cdot}s^{-1}}$ threshold, but the same cannot be concluded at higher or lower thresholds. The analog experiments are able to improve the raw forecast, exhibiting significantly higher skill than the *EMOS*, up to 10 $\mathrm{m{\cdot}s^{-1}}$ wind speed threshold. Furthermore, both *AnEnStd* and *AnEnAll* experiments significantly improve the raw model results for all thresholds tested in January.

The results presented in this study prove that using the raw model ensemble mean and/or the ensemble spread for more than one meteorological variable as an input to the analog method delivers suitable ensemble post-processed forecasts, especially when computationally only limited resources are available. Moreover, this approach sometimes even outperforms other, more computationally demanding, analog-based configurations, such as the "member-by-member" approach, or other methods such as the *EMOS*.

ACKNOWLEDGEMENT

We gratefully acknowledge RC LACE for supporting this research. We thank the two anonymous reviewers whose suggestions helped improve this manuscript.

ORCID

Iris Odak Plenković https://orcid.org/0000-0002-6622-4918
Irene Schicker https://orcid.org/0000-0001-6401-2412
Markus Dabernig https://orcid.org/0000-0002-2238-4282
Endi Keresturi https://orcid.org/0000-0002-9626-1947

REFERENCES

Alessandrini, S., Delle Monache, L., Sperati, S. and Cervone, G. (2015a) An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy*, 157, 95–110.

Alessandrini, S., Delle Monache, L., Sperati, S. and Nissen, J. (2015b) Ensemble novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, 76, 768–781.

Alessandrini, S., Sperati, S. and Delle Monache, L. (2019) Improving the analog ensemble wind speed forecasts for rare events. *Monthly Weather Review*, 147, 2677–2692. https://doi.org/10.1175/MWR-D-19-0006.1.

Arnold, D., Morton, D., Schicker, I., Seibert, P., Rotach, M.W., Horvath, K., Dudhia, T., Satomura, T., Muller, M., Zangl, G., Takemi, T., Serafin, S., Schmidli, J. and Schneider, S. (2012) Issues in high-resolution modeling in complex topography – the HiRCoT workshop. *Croatian Meteorological Journal*, 47, 1–12.

Baran, S. and Lerch, S. (2015) Log-normal distribution based ensemble model output statistics models for probabilistic wind speed forecasting. *Quarterly Journal of the Royal Meteorological Society*, 141, 2289–2299. https://doi.org/10.1002/qj.2521.

Bougeault, P., Binder, P., Buzzi, A., Dirks, R., Houze, R., Jr., Kuettner, J., Smith, R.B., Steinacker, R. and Volkert, H. (2001) The MAP special observing period. *Bulletin of the American Meteorological Society*, 82(3), 433–462.

Dabernig, M., Mayr, G.J. and Messner, J.W. (2015) Predicting wind power with reforecasts. *Weather and Forecasting*, 30, 1655–1662. https://doi.org/10.1175/WAF-D-15-0095.1.

Delle Monache, L., Nipen, T., Deng, X., Zhou, Y. and Stull, R.B. (2006) Ozone ensemble forecasts: 2. A Kalman filter predictor bias-correction. *Journal of Geophysical Research*, 111, D05308.

Delle Monache, L., Wilczak, J., McKeen, S., Grell, G., Pagowski, M., Peckham, S., Stull, R., McHenry, J. and McQueen, J. (2008) A Kalman-filter bias correction of ozone deterministic, ensemble-averaged, and probabilistic forecasts. *Tellus B*, 60, 238–249.

Delle Monache, L., Nipen, T., Liu, Y., Roux, G. and Stull, R. (2011) Kalman filter and analog schemes to post-process numerical weather predictions. *Monthly Weather Review*, 139, 3554–3570.

Delle Monache, L., Eckel, T., Rife, D. and Nagarajan, B. (2013) Probabilistic weather prediction with an analog ensemble. *Monthly Weather Review*, 141, 3498–3516.

Djalalova, I., Delle Monache, L. and Wilczak, J. (2015) PM2.5 analog forecast and Kalman filter post-processing for the community multiscale air quality (CMAQ) model. *Atmospheric Environment*, 108, 76–87.

Drosdowsky, W. (1994) Analog (nonlinear) forecasts of the Southern Oscillation index time series. *Weather and Forecasting*, 9, 78–84.

Eckel, F.A. and Delle Monache, L. (2016) A hybrid NWP–Analog ensemble. *Monthly Weather Review*, 144, 897–911.

Gao, L., Ren, H., Li, J. and Chou, J. (2006) Analogue correction method of errors and its application to numerical weather prediction. *Chinese Physics*, 15, 882–889.

Gneiting, T., Raftery, A.E., Westveld, A.H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133, 1098–1118. https://doi.org/10.1175/MWR2904.1.

Hamill, T.M. and Whitaker, J.S. (2006) Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Monthly Weather Review*, 134, 3209–3229.

Hamill, T.M., Whitaker, J.S. and Mullen, S.L. (2006) Reforecasts: an important dataset for improving weather predictions. *Bulletin of the American Meteorological Society*, 87, 33–46.

Hopson, T.M. (2005) *Operational flood-forecasting for Bangladesh* (225 pp). Boulder: Ph.D. thesis, University of Colorado.

Hopson, T.M. and Webster, P.J. (2010) A 1–10-day ensemble forecasting scheme for the major river basins of Bangladesh: forecasting severe floods of 2003–07. *Journal of Hydrometeorology*, 11, 618–641.

Horvath, K., Bajić, A. and Ivatek-Šahdan, S. (2011) Dynamical downscaling of wind speed in complex terrain prone to bora-type flows. *Journal of Applied Meteorology and Climatology*, 50, 1676–1691.

Horvath, K., Koračin, D., Vellore, R., Jiang, J. and Belu, R. (2012) Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models. *Journal of Geophysical Research*, 117, D11111.

Jolliffe, I.T. (2007) Uncertainty and inference for verification measures. *Weather and Forecasting*, 22, 637–650.

Jolliffe, I.T. and Stephenson, D.B. (2011) *Forecast Verification: A Practitioner's Guide in Atmospheric Science,* 2nd edition. Chichester: Wiley.

Junk, C., Delle Monache, L., Alessandrini, S., von Bremen, L. and Cervone, G. (2015) Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteorologische Zeitschrift*, 24, 361–379.

Keller, D.E., Fischer, A.M., Liniger, M.A., Appenzeller, C. and Knutti, R. (2017) Testing a weather generator for downscaling climate change projections over Switzerland. *International Journal of Climatology*, 37, 928–942.

Klausner, Z., Kaplan, H. and Fattal, E. (2009) The similar days method for predicting near surface wind vectors. *Meteorological Applications*, 16, 569–579.

Kuettner, J.P. (1986) The aim and conduct of ALPEX (Alpine Experiment (ALPEX). *WMO Proceedings of the Conference on the Scientific Results of the Alpine Experiment (ALPEX)*, ICSU-WMO, GARP Publication Series, 27, 3–13.

Lehner, M. and Rotach, M.W. (2018) Current challenges in understanding and predicting transport and exchange in the atmosphere over mountainous terrain. *Atmosphere*, 2018(9), 276.

Lorenz, E.N. (1969) Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26, 636–646.

Mayr, G.J., Armi, L., Gohm, A., Zängl, G., Durran, D.R., Flamant, C., Gaberšek, S., Mobbs, S., Ross, A. and Weissmann, M. (2007) Gap flows: results from the Mesoscale Alpine Programme. *Quarterly Journal of the Royal Meteorological Society*, 133, 881–896. https://doi.org/10.1002/qj.66.

Mayr, G.J., Plavcan, D., Armi, L., Elvidge, A., Grisogono, B., Horvath, K., Jackson, P., Neururer, A., Seibert, P. and Steenburgh, J.W. (2018) The community foehn classification experiment. *Bulletin of the American Meteorological Society*, 99(11), 2229–2235. https://doi.org/10.1175/BAMS-D-17-0200.1.

Messner, J.W., Mayr, G.J., Wilks, D.S. and Zeileis, A. (2014) Extending extended logistic regression: extended versus separate versus ordered versus censored. *Monthly Weather Review*, 142, 3003–3014. https://doi.org/10.1175/MWR-D-13-00355.1.

Messner, J.W., Mayr, G.J. and Zeileis, A. (2017) Nonhomogeneous boosting for predictor selection in ensemble postprocessing. *Monthly Weather Review*, 145, 137–147. https://doi.org/10.1175/MWR-D-16-0088.1.

Mugume, I., Mesquita, M.D.S., Bamutaze, Y., Ntwali, D., Basalirwa, C., Waiswa, D., Reuder, J., Twinomuhangi, R., Tumwine, F., Jakob Ngailo, T. and Ogwang, B.A. (2017) Improving quantitative rainfall prediction using ensemble analogues in the Tropics: case study of Uganda. *Preprints*, 2017, 1–16. doi:10.20944/preprints201710.0199.v1.

Murphy, A.H. (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116, 2417–2424.

Nagarajan, B., Delle Monache, L., Hacker, J., Rife, D., Searight, K., Knievel, J. and Nipen, T. (2015) An evaluation of analog-based post-processing methods across several variables and forecast models. *Weather and Forecasting*, 30, 1623–1643.

Odak Plenković, I., Delle Monache, L., Horvath, K. and Hrastinski, M. (2018) Deterministic wind speed predictions with analog-based methods over complex topography. *Journal of Applied Meteorology and Climatology*, 57, 2047–2070. https://doi.org/10.1175/JAMC-D-17-0151.1.

Panziera, L., Germann, U., Gabella, M. and Mandapaka, P.V. (2011) NORA – nowcasting of orographic rainfall by means of analogues. *Quarterly Journal of the Royal Meteorological Society*, 137, 2106–2123.

Ren, H. and Chou, J. (2006) Analogue correction method of errors by combining statistical and dynamical methods. *Acta Meteorologica Sinica*, 20, 367–373.

Ren, H. and Chou, J. (2007) Strategy and methodology of dynamical analogue prediction. *Science in China Series D: Earth Sciences*, 50, 1589–1599.

Rousteenoja, K. (1988) Factors affecting the occurrence and lifetime of 500 mb height analogues: a study based on a large amount of data. *Monthly Weather Review*, 116, 368–376.

Scheuerer, M. and Möller, D. (2015) Probabilistic wind speed forecasting on a grid based on ensemble model output statistics. *Annals of Applied Statistics*, 9(3), 1328–1349. https://doi.org/10.1214/15-AOAS843.

Serafin, S., Adler, B., Cuxart, J., De Wekker, S.F.J., Gohm, A., Grisogono, B., Kalthoff, N., Kirshbaum, D.J., Rotach, M.W., Schmidli, J., Stiperski, I., Večenaj, Ž. and Zardi, D. (2018) Exchange processes in the atmospheric boundary layer over mountainous terrain. *Atmosphere*, 2018(9), 102.

Sperati, S., Alessandrini, S. and Delle Monache, L. (2017) Gridded probabilistic weather forecasts with an analog ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143, 2874–2885. https://doi.org/10.1002/qj.3137.

Thorarinsdottir, T.L. and Gneiting, T. (2010) Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173, 371–388. https://doi.org/10.1111/j.1467-985X.2009.00616.x.

Van den Dool, H.M. (1989) A new look at weather forecast through analogs. *Monthly Weather Review*, 117, 2230–2247.

Vanvyve, E., Delle Monache, L., Rife, D., Monaghan, A. and Pinto, J. (2015) Wind resource estimates with an analog ensemble approach. *Renewable Energy*, 74, 761–773.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Sahdan, S., Tian, W., Ma, X., Tascu, S. and Bazile, E. (2011) The central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 137(655), 483–502.

Wilcox, R.R. (2009) Comparing Pearson correlations: dealing with heteroscedasticity and nonnormality. *Communication in Statistics-Simulation and Computation*, 38, 2220–2234.

Wilks, D.S. (1997) Resampling hypothesis tests for autocorrelated fields. *Journal of Climate*, 10, 65–82.

Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*, 3rd edition. Oxford: Academic Press.

Wu, W., Liu, Y., Ge, M., Rostkier-Edelstein, D., Descombes, G., Kunin, P., Warner, T., Swerdlin, S., Givati, A., Hopson, T.M., Swerdlin, S., Givati, A., Hopson, T.M. and Yates, D.N. (2012) Statistical downscaling of climate forecast system seasonal predictions for the southeastern Mediterranean. *Atmospheric Research*, 118, 346–356.

Xavier, P.K. and Goswami, B.N. (2007) An analog method for real-time forecasting of summer monsoon subseasonal variability. *Monthly Weather Review*, 135, 4149–4160.

Zhang, J., Draxl, C., Hopson, T., Delle Monache, L. and Hodge, B.-M. (2015) Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Applied Energy*, 156, 528–541.

How to cite this article: Odak Plenković I, Schicker I, Dabernig M, Horvath K, Keresturi E. Analog-based post-processing of the ALADIN-LAEF ensemble predictions in complex terrain. *Q J R Meteorol Soc*. 2020;1–19. https://doi.org/10.1002/qj.3769