On the Impact of the Choice of Global Ensemble in Forcing a Regional Ensemble System

Florian Weidle and Yong Wang

Department of Forecasting Models, Central Institute for Meteorology and Geodynamics, Vienna, Austria

Geert Smet

Royal Meteorological Institute, Brussels, Belgium

(Manuscript received 10 August 2015, in final form 14 January 2016)

ABSTRACT

It is quite common that in a regional ensemble system the large-scale initial condition (IC) perturbations and the lateral boundary condition (LBC) perturbations are taken from a global ensemble prediction system (EPS). The choice of global EPS as a driving model can have a significant impact on the performance of the regional EPS. This study investigates the impact of large-scale IC/LBC perturbations obtained from different global EPSs on the forecast quality of a regional EPS. For this purpose several experiments are conducted where the Aire Limitée Adaption dynamique Développement International–Limited Area Ensemble Forecasting (ALADIN-LAEF) regional ensemble is forced by two of the world's leading global ensembles, the European Centre for Medium-Range Weather Forecasts' Ensemble Prediction System (ECMWF-EPS) and the Global Ensemble Forecasting System (GEFS) from the National Centers for Environmental Prediction (NCEP), which provide the IC and LBC perturbations. The investigation is carried out for a 51-day period during summer 2010 over central Europe. The results indicate that forcing of the regional ensemble with GEFS performs better for surface parameters, whereas at upper levels forcing with ECMWF-EPS is superior. Using perturbations from GEFS lead to a considerably higher spread in ALADIN-LAEF, which is beneficial near the surface where regional EPSs are usually underdispersive. At upper levels, forcing with GEFS leads to an overdispersion of ALADIN-LAEF as a result of the large spread of some parameters, where forcing ALADIN-LAEF with ECMWF-EPS provides statistically more reliable forecasts. The results indicate that the best global EPS might not always provide the best ICs and LBCs for a regional ensemble.

1. Introduction

A proper way of perturbing the initial conditions (ICs) and lateral boundary conditions (LBCs) is one of the key factors for a skillful regional ensemble. There have been various approaches designed for generating IC and LBC perturbations for the regional ensemble prediction system (EPS; e.g., Marsigli et al. 2005; Bowler et al. 2008; Garcia-Moya et al. 2011; Hamill and Colucci 1997; Wang et al. 2011). In almost all of those methods, the large-scale IC perturbations and LBC perturbations provided by a global EPS are the essential part of the IC and LBC perturbations in the regional EPS.

The most common approach is the dynamical downscaling of a global EPS, in which the regional IC and LBC perturbations are provided by interpolating the ICs and forecasts from the different ensemble members of the driving global EPS. This method is widely used in operations, for example in the Consortium for Small-Scale Modeling's Limited Area Ensemble Prediction System (COSMO LEPS; Marsigli et al. 2005; Montani et al. 2011) or in the Met Office Global and Regional EPS (MOGREPS; Bowler et al. 2008). Because of its simplicity, its low computational costs, and its good performance, dynamical downscaling of a global EPS for IC and LBC perturbations is also used in a number of other regional EPS systems, for example, the pan-European Grand Limited Area Ensemble Prediction System (GLAMEPS; Iversen et al. 2011; Smet et al. 2012).

Another strategy that exists in regional EPS systems is to use LBCs from different deterministic global models,

Corresponding author address: Yong Wang, Dept. of Forecasting Models, Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, A-1190 Vienna, Austria.
E-mail: yong.wang@zamg.ac.at

DOI: 10.1175/WAF-D-15-0102.1

like in the Short-Range Ensemble Prediction System (SREPS) developed at the Spanish Meteorological Service (Garcia-Moya et al. 2011). This system uses five different global deterministic models to provide different LBCs for the different SREPS members. A similar approach is used in the convection-permitting EPS of the German Weather Service (COSMO-DE-EPS; Gebhardt et al. 2011) where four global deterministic models are interpolated to the target domain and provide the LBCs for the regional EPS ensemble members.

A few methods have been proposed for producing regional IC perturbations, such as breeding (Hamill and Colucci 1997; Stensrud et al. 2000; Du and Tracton 2001; Du et al. 2003; Wang et al. 2011) and ensemble transformation (ET; Bishop and Toth 1999; Wei et al. 2008; Bishop et al. 2009). They have been successfully implemented in the Short-Range Ensemble Forecast system (SREF) at NCEP and in the Aire Limitée Adaption dynamique Développement International–Limited Area Ensemble Forecasting (ALADIN-LAEF) at ZentralAnstalt für Meteorologie und Geodynamik (ZAMG). An alternative method is the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003), which has been tested by a number of studies. Bowler and Mylne (2009) compared the performance of a regional ETKF and the dynamical downscaling of a global EPS. They found that in MOGREPS the use of the ETKF for the regional ensemble does not outperform dynamical downscaling of the global ensemble. However, Hacker et al. (2011) showed that a regional ETKF is more skillful than the dynamical downscaling in a short-range ensemble prediction system based on the Weather Research and Forecasting (WRF) Model. Another method of generating IC perturbations for the regional EPS is the targeted singular vectors (SVs; Buizza and Palmer 1995; Leutbecher and Palmer 2008; Li et al. 2008; Frogner et al. 2006; Iversen et al. 2011), but up to now there is no documentation on the operational use of SVs in regional EPSs.

It is common practice in regional EPSs to use LBC perturbations from the forecasts of a global EPS. If the IC perturbations treatment in the regional EPS is not coherent with the global driving EPS, the regional IC perturbations might conflict with perturbations coming from the LBCs and would introduce numerical noise at the lateral boundaries (Warner et al. 1997; Bowler and Mylne 2009; Wang et al. 2011; Caron 2013). Recently, Wang et al. (2014) developed a blending method to generate the regional IC perturbations using the large-scale IC perturbations of the driving global EPS and the small-scale regional IC perturbations. This method has been implemented in the SREF at NCEP (J. Du 2015, personal communication) and in the regional EPS at the China Meteorological Administration (Zhang et al. 2015).

Another simple approach to generating perturbations is to randomly perturb a deterministic forecast. Torn et al. (2006) showed that different methods of generating LBCs by perturbing a deterministic forecast randomly lead to a comparable quality of the regional ensemble. However, Torn et al. (2006) used this method only to perturb LBCs and the usability of this method for producing appropriate IC perturbations for a regional EPS was not investigated.

The discussion above shows that the global EPS that provides the large-scale IC perturbations and LBC perturbations plays a crucial role in determining the forecast quality of the regional EPS. The focus of this paper is on assessing the impact of the global EPSs used to provide the IC and LBC perturbations on the forecast quality of a regional EPS.

There are very few documented studies on the impact of the driving global EPS in regional EPSs. In the work presented, the single-model-based regional ensemble ALADIN-LAEF is forced by two global ensemble systems: the ECMWF-EPS and GEFS. Their impact on the forecast performance of ALADIN-LAEF is evaluated.

For a 1.5-month test period during summer 2010 the full datasets of ECMWF-EPS and GEFS, which are two of the world's leading global EPSs (Bougeault et al. 2010), were available at ZAMG, where ECMWF-EPS is often acknowledged as the best global EPS (Park et al. 2008; Froude 2010). Buizza et al. (2005) compared ECMWF-EPS and GEFS and showed that ECMWF-EPS has the best overall performance. However, the performance of GEFS is comparable for the first few forecast days, which is the forecast range over which ALADIN-LAEF is evaluated and forced by the global ensembles. The comparison of Buizza et al. (2005) focused on the forecast performance in the free atmosphere and did not consider near-surface weather parameters.

It is of interest here to compare the results of using the IC and LBC perturbations from either ECMWF-EPS or GEFS in a single-model-based regional EPS, such as ALADIN-LAEF. The aim of this study is to investigate whose IC and LBC perturbations are more appropriate for the ALADIN-LAEF forecast of near-surface weather parameters, which are of most interest for a regional EPS.

Section 2 briefly describes the ALADIN-LAEF system, and the global EPSs from ECMWF and NCEP, as well as the experiment design. The verification results of the experiments for a 1.5-month summer period during 2010 are presented in section 3. Summary and conclusions follow in section 4.

2. Ensemble prediction systems and experiment design

This section gives a brief description of the ensemble systems used in this study, the experiments that have been conducted, and the verification strategy.

a. ALADIN-LAEF

ALADIN-LAEF is a regional EPS developed within the framework of the European ALADIN consortium for short-range numerical weather prediction and the Regional Cooperation for Limited-Area Modeling in Central Europe (RC LACE). It has 16 perturbed members, using the ALADIN Limited-Area Model forced by the first 16 members of the global ECMWF-EPS, and—for the experiments presented in this study—is running at a horizontal resolution of 18 km with 37 levels. The integration domain covers Europe and large parts of the North Atlantic (Fig. 1), and the maximum forecast range is 54 h. While ALADIN-LAEF is in operational mode, a breeding–blending technique is used to generate the initial perturbations (Wang et al. 2011); in this study the downscaling configuration (of ECMWF-EPS or GEFS) is used to clearly separate the effects of forcing the regional EPS with different global EPSs from other effects. For the same reason the multiphysics scheme of ALADIN-LAEF described in detail in Wang et al. (2011) is not employed, but instead the same physics and model setup is maintained for all members. Consequently, the noncycling surface breeding (NCSB) method introduced in ALADIN-LAEF to generate surface perturbations (Wang et al. 2010) is not implemented, but instead the surface analysis of the Action de Recherche Petite Echelle Grande Echelle model (ARPEGE), the global deterministic model of Météo-France, is used for all members. This is motivated by the fact that different models use different surface schemes that can cause large systematic errors due to inconsistencies when nesting a regional model into a global one. The ARPEGE surface has the same physical parameterizations for the surface as ALADIN; hence, inconsistencies between the surface of the global model and the regional model are avoided. Consequently, differences in the near-surface parameters cannot be caused by inconsistent surface schemes from NCEP/ECMWF models and ALADIN.

b. ECMWF-EPS

The operational ECMWF-EPS consists of 50 perturbed members. In summer 2010 the ECMWF-EPS ran operationally on a Gaussian grid with a horizontal resolution of T639L62 (approximately 30 km) with 62 vertical levels. The generation of initial perturbations in ECMWF-EPS is based on a combined use of SVs and ensemble data assimilation (EDA; Bonavita et al. 2011). SVs are computed to maximize the total energy norm for an optimization time of 48 h at a resolution of T42L40 (Buizza and Palmer 1995). This means that the initial perturbations are generated such that they are maximized within the first 2 days of the forecast and therefore can be small at the initial time. The introduction of EDA (Buizza et al. 2008, 2010; Isaksen et al. 2010) leads to an increase in spread in ECMWF-EPS, especially over the first few days of the forecast. To account for model uncertainties, the stochastically perturbed parameterization tendencies scheme (SPPT) was in used in ECMWF-EPS during summer 2010 (Buizza et al. 1999; Palmer et al. 2009).

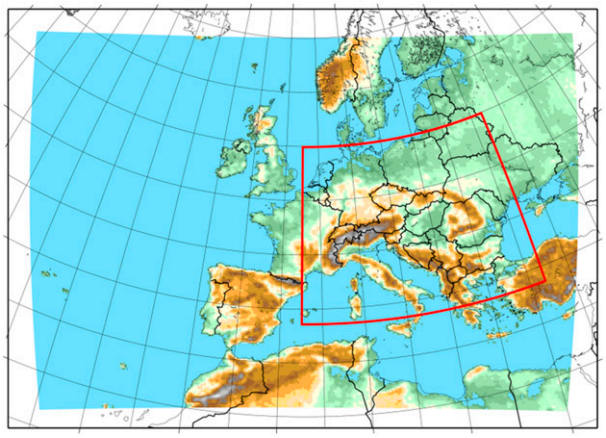

FIG. 1. ALADIN-LAEF domain and model topography. The inner limited-area domain in red depicts the verification domain, which covers central Europe.

c. GEFS

The GEFS is a global EPS with 20 perturbed members, running on a Gaussian grid at a resolution of T190L28 (horizontal resolution of approximately 73 km, with 28 vertical levels) in summer 2010. With the increase in horizontal resolution that took place in February 2010, a stochastic total tendency perturbation scheme (STTP) was included in GEFS to account for random model uncertainties (Hu et al. 2008). To perturb the initial conditions, an ensemble transform technique with rescaling (ETR) was used (Wei et al. 2008), which is an improved version of the original breeding vector (BV) technique (Toth and Kalnay 1993, 1997). The BVs, closely related to the Lyapunov vectors in dynamical systems, use pairs of members from the previous ensemble run to obtain the fastest-growing perturbations of the analysis error. The ETR is an extension of the breeding method, where the perturbed forecasts from

previous runs are not rescaled pairwise but by using orthogonal vectors (Wei et al. 2008).

d. Experiment design

To compare the performance of ALADIN-LAEF driven by ECMWF-EPS and GEFS, two experiments (hereafter referred to as ECMWF and NCEP, respectively) are performed using dynamical downscaling of ECMWF-EPS and GEFS for IC and LBC perturbations of ALADIN-LAEF. In the experiments, the first 16 ensemble members of both global EPSs are downscaled for IC and LBC perturbations of ALADIN-LAEF. The same method is applied in operational ALADIN-LAEF, where the first 16 members of ECMWF-EPS are used for regional IC and LBC perturbations. This arbitrary choice of selecting members from the ECMWF-EPS to provide IC and LBC perturbations was justified by Weidle et al. (2013), who compared the performance of the same basic version of ALADIN-LAEF, as it is used in this study, with ALADIN-LAEF runs where the forcing ECMWF-EPS members were selected by an objective cluster method. It was shown that using different strategies of clustering does not improve the ALADIN-LAEF forecast performance considerably compared to an arbitrary choice of the first 16 members.

e. Verification strategy

The experiments described above are evaluated by means of the following standard probabilistic verification scores: bias and root-mean-square error (RMSE) of the ensemble mean, ensemble spread, reliability diagrams, and the continuous ranked probability score (CRPS). For a complete description of these scores, the reader is referred to Wilks (2006). The upper-air variables wind speed (WS), temperature T, relative humidity (RH), and geopotential Z at 850 and 500 hPa are verified against the deterministic analyses of both NCEP and ECMWF. For this purpose, both forecasts and analyses were interpolated onto a common regular $0.15° \times 0.15°$ grid on a subdomain of the limited-area forecast domain. This subdomain, hereafter called the verification domain, which covers central Europe, is shown in Fig. 1. Since there were few qualitative differences between the scores when using either ECMWF analyses or NCEP analyses, only the results of the verification against ECMWF analyses will be shown. The few differences in the verification results that occur when using the NCEP analyses will be described in the text.

The surface weather variables 10-m wind speed (WS10m), 2-m temperature (T2m), 12-h accumulated precipitation (PREC), and mean sea level pressure (MSLP) are verified against 1219 surface synoptic observation stations in the verification domain. Forecasts are matched with observations by interpolating the smoothly varying forecast fields (WS10m, T2m, and MSLP) to the observation locations, and by using the grid point closest to the observation site for PREC.

The experiments were all initialized at 0000 UTC and integrated up to +54 h over the full verification period from 1 July 2010 through and including 20 August 2010. Verification scores were calculated for 6-hourly output and averaged over the whole verification domain and over the full verification period.

As mentioned before, all experiments use dynamical downscaling of the global EPS (i.e., without the upper-air breeding–blending cycle and the noncycling surface breeding that are implemented in the operational ALADIN-LAEF) and without using multiphysics. Hence, all members are dynamically downscaled with the same ALADIN model configuration.

3. Results

Below, we present the results of the experiments where in the first part the performance for near-surface variables is evaluated and in the second part upper-air variables are examined.

a. Surface weather variables

As the surface initial conditions are identical among ensemble members in all experiments, big differences in the scores for 2-m temperature are not to be expected, especially for short-range forecasts.

Figure 2 shows the RMSE of the ensemble mean and ensemble spread of the NCEP and ECMWF experiments for the variables WS10m, T2m, PREC, and MSLP. The RMSEs of the two experiments are very similar. Differences are visible in the spread for all parameters where forcing with GEFS produces larger values than with ECMWF-EPS. This difference is most pronounced at early lead times and gradually decreases with increasing lead time, but does not completely disappear (except for PREC). Only MSLP exhibits a consistently larger spread in NCEP over the whole forecast range, with the difference being quite constant. In both experiments the spread is significantly smaller than the RMSE, except for MSLP in NCEP at the longest forecast ranges, which points to statistical inconsistencies in the ensembles (Buizza et al. 2005). Because of the larger spread, the NCEP experiment has a better RMSE-to-spread relation than does the ECMWF.

The bias of the ensemble means is shown in Fig. 3 for the verified surface parameters. As for RMSE the differences between NCEP and ECMWF for WS10m, T2m, and MSLP are rather small. The biggest difference

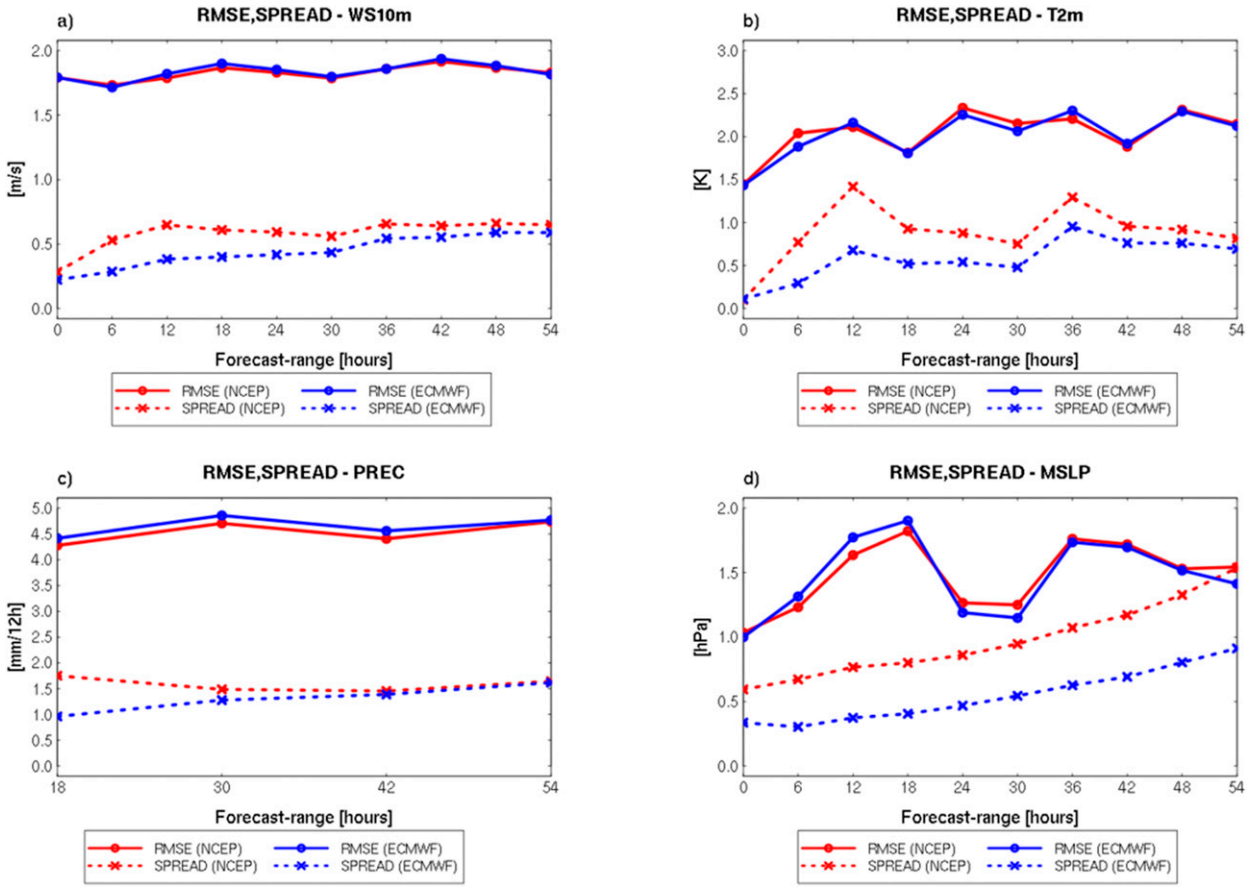

FIG. 2. RMSEs (solid lines) of the ensemble mean and spread (dashed) of the NCEP (red) and ECMWF (blue) experiments for (a) WS10m, (b) T2m, (c) PREC, and (d) MSLP, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

between NCEP and ECMWF occurs for precipitation where using the GEFS results in a negative bias for the 30- and 54-h forecasts, hence during night, whereas forcing with ECMWF-EPS produces a positive bias for precipitation during day- and nighttime (Fig. 3c). Notable differences can also be seen in the bias for 2-m temperature (Fig. 3b), where the forcing by ECMWF-EPS reduces the pronounced diurnal cycle in NCEP to some extent. While the bias for the ensemble mean in NCEP and ECMWF is comparable during the daytime, it is clearly better in ECMWF at night (forecast hours +6, +24, and +30 h). Both experiments show a negative bias for the whole forecast range, except for +36- and +42-h forecasts in NCEP.

In addition to the deterministic verification of ensemble means, a number of probabilistic scores have also been evaluated, such as CRPS and reliability diagrams.

Figure 4 shows the difference in CRPS between NCEP and ECMWF, together with the 95% confidence intervals based on the bootstrap method (Wilks 2006). Since CRPS is negatively oriented, negative values of the difference indicate a better CRPS in NCEP, and positive values a superior ECMWF. Using GEFS as the driving model turns out to give consistently better CRPS results than does using ECMWF-EPS for WS10m and PREC at all lead times, except for +54 h (Figs. 4a,c), and for T2m and MSLP at most lead times (Figs. 4b,d). This is probably a consequence of the better RMSE-to-spread ratio of NCEP, as mentioned above (cf. Fig. 2). Consistent with the previous results (RMSE and spread), the difference in CRPS in the experiments seems to decrease with lead time. However, for T2m a clear diurnal cycle is present (Fig. 4b), with the difference larger at noon (+12 and +36 h) and a smaller result at midnight (+24 and +48 h). For T2m (Fig. 4b), the difference in CRPS is significant at the 95% level at noon and during the evening (+12, +18, +36, and +42 h), but not during the night and morning (+24 and +30 h). For MSLP, the picture is not as clear, with NCEP having a significantly lower CRPS at early lead times (up to +18 h), and ECMWF having a significantly

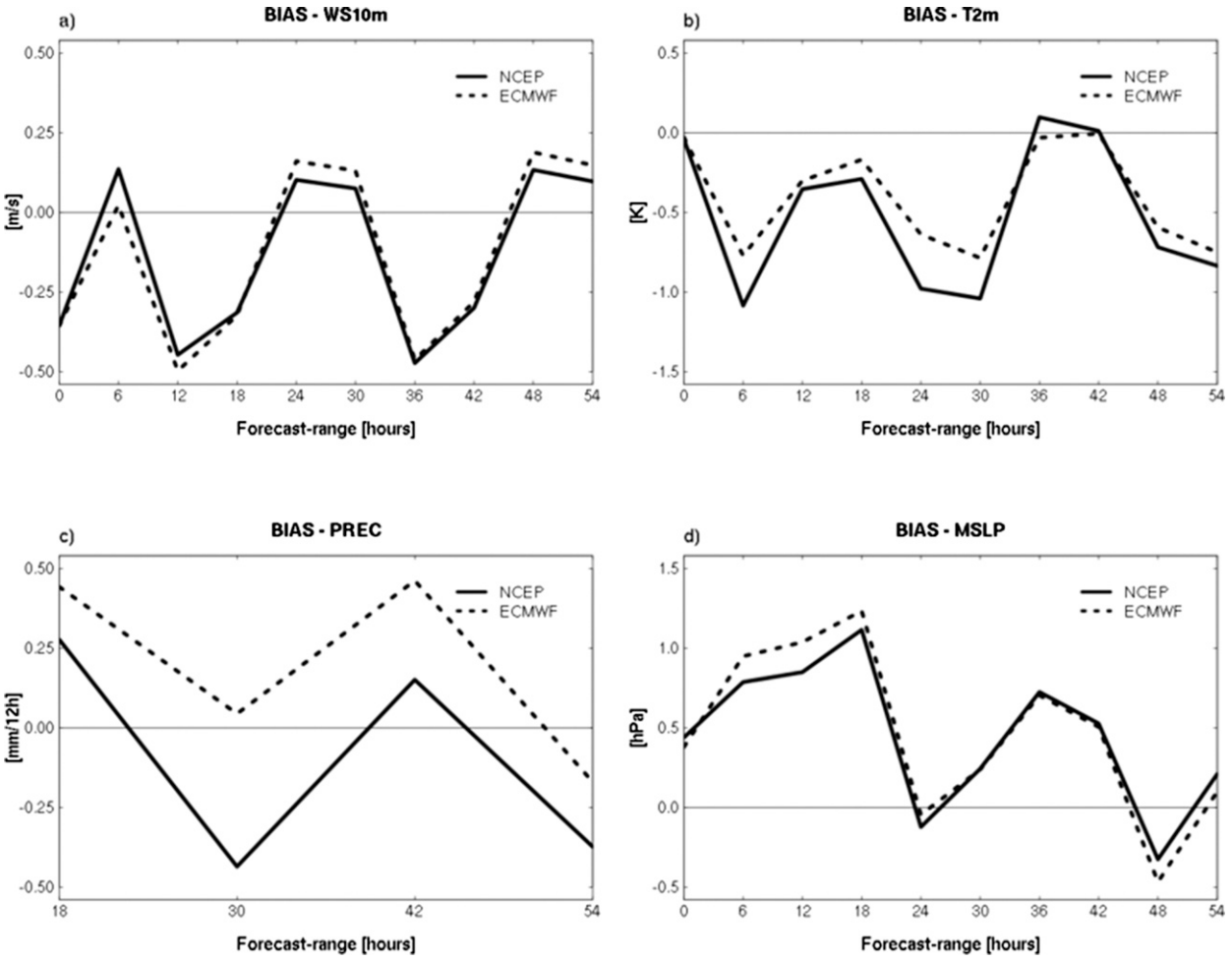

FIG. 3. Biases in the ensemble mean of the NCEP (solid lines) and ECMWF (dotted lines) experiments for (a) WS10m, (b) T2m, (c) PREC, and (d) MSLP, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

lower CRPS at +54 h, with no significant differences in the intermediate forecast range.

Reliability diagrams provide a measure of the accuracy of the ensemble system. For a perfect system the forecasted probability for a certain event equals its frequency of observation. Thus, when plotted against each other, a system will have perfect reliability when its points lie along a diagonal of slope 1. Figure 5 shows the reliability diagrams for NCEP and ECMWF at +42-h lead time and the verifying events WS10m $> 4\,\mathrm{m\,s^{-1}}$ (Fig. 5a), T2m anomaly $> 0°\mathrm{C}$ (Fig. 5b), PREC $> 5\,\mathrm{mm}\,(12\,\mathrm{h})^{-1}$ (Fig. 5c), and MSLP $> 1015\,\mathrm{hPa}$ (Fig. 5d). For WS10m, T2m, and PREC, forcing by GEFS leads to better reliability for the given thresholds. Only for MSLP are the results less clear, with experiment ECMWF being better for forecast probabilities between 0.2 and 0.5, and experiment NCEP better for forecast probabilities of 0.6 and above. Similar results are seen for WS10m and PREC for other lead times and thresholds with neutral or slightly more reliable forecasts in NCEP. For T2m a diurnal cycle is present with better reliability in ECMWF during the night; hence, forecast ranges 24, 30, 48, and 54 h and the diurnal cycle is most pronounced for forecasted probabilities between 0.3 and 0.7 (not shown).

The small differences in reliability are not in contradiction with the previous results; for example, the difference in CRPS for WS10m is significant at all lead times. One has to keep in mind that the reliability diagrams are for one specific threshold, while the CRPS score is an overall score, which can actually be interpreted as the integral of the Brier score over all possible thresholds (Hersbach 2000).

b. *Upper-air weather variables*

To investigate the upper-air performance, WS, T, RH, and Z are evaluated at 850 and 500 hPa. As mentioned above, the experiments were verified against the deterministic analyses of both NCEP and ECMWF, but

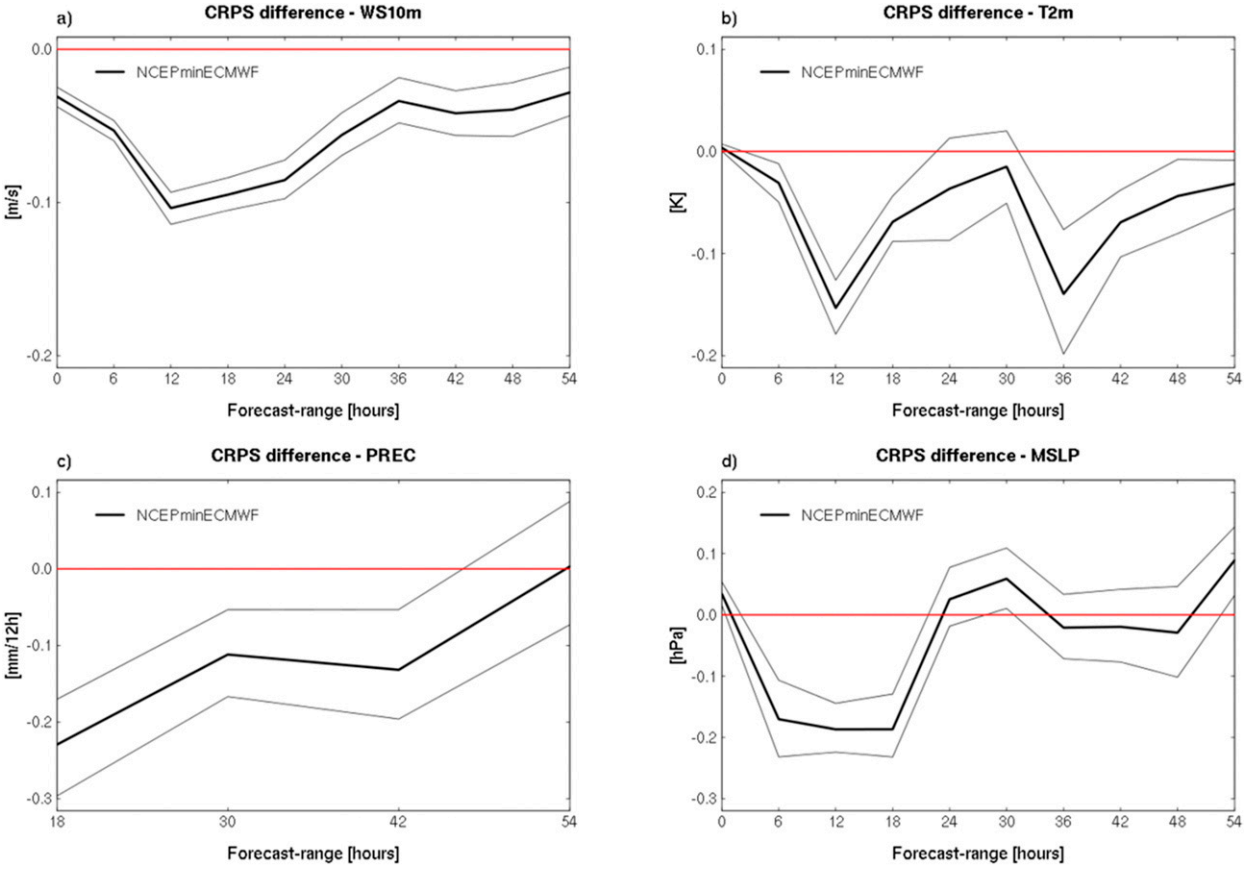

FIG. 4. Differences in CRPSs between NCEP and ECMWF (thick solid line) for (a) WS10m, (b) T2m, (c) PREC, and (d) MSLP, averaged over the whole verification domain and over the whole verification period (0000 UTC runs). The corresponding 95% confidence interval is shown with thin lines.

since the differences in results are usually very small, only the verification scores with ECMWF analyses as a reference will be presented and differences in the scores when verified against NCEP analyses will be addressed in the text.

Figure 6 shows the RMSE and spread at 850 hPa for the four verified parameters, while Fig. 7 shows the bias of the ensemble means of the experiments. For T850 and WS850 the RMSE and spread resemble those for the surface parameters with a comparable RMSE in both experiments but a consistently higher spread when ALADIN-LAEF is driven by GEFS. A better RMSE can be seen for RH in ECMWF for a forecast range up to +36 h and for Z at forecast ranges longer than 1 day ahead (Figs. 6c,d). The superior RMSE for RH in ECMWF is caused by a much smaller bias than in the NCEP experiment in the first forecast hours (see Fig. 7c). For Z the better RMSE in ECMWF at longer forecast ranges is a combination of a slightly smaller negative bias in ECMWF (Fig. 7d) and a smaller variance in the bias of the ensemble mean compared to NCEP. In the NCEP experiment the bias of the ensemble mean ranges in the verification period (e.g., for forecast range +48 h) from −180 to 70 $m^2 s^{-2}$ while for ECMWF the bias ranges from −150 to 60 $m^2 s^{-2}$. Since the RMSE is affected more by large errors, the events with the large negative bias in NCEP lead to considerably worse RMSEs than in ECMWF.

As for the surface parameters, the spread is consistently higher in NCEP and closer to the RMSE of the ensemble mean than in ECMWF (Fig. 6) for all variables. This leads to a better RMSE-to-spread ratio and, hence, in total to a better statistical reliability in NCEP rather than in the ECMWF experiment as a result of a more adequate representation of the forecast uncertainty. However, for the first forecast hours of RH and for the whole forecast range of Z at 850 hPa the forcing by GEFS leads to an overdispersion of the regional ensemble, which means that the forecasts must be treated as unreliable (Wilks 2006). The biases for T and WS at 850 hPa in both experiments (Figs. 7a,b) are close to each other with only a small positive bias for

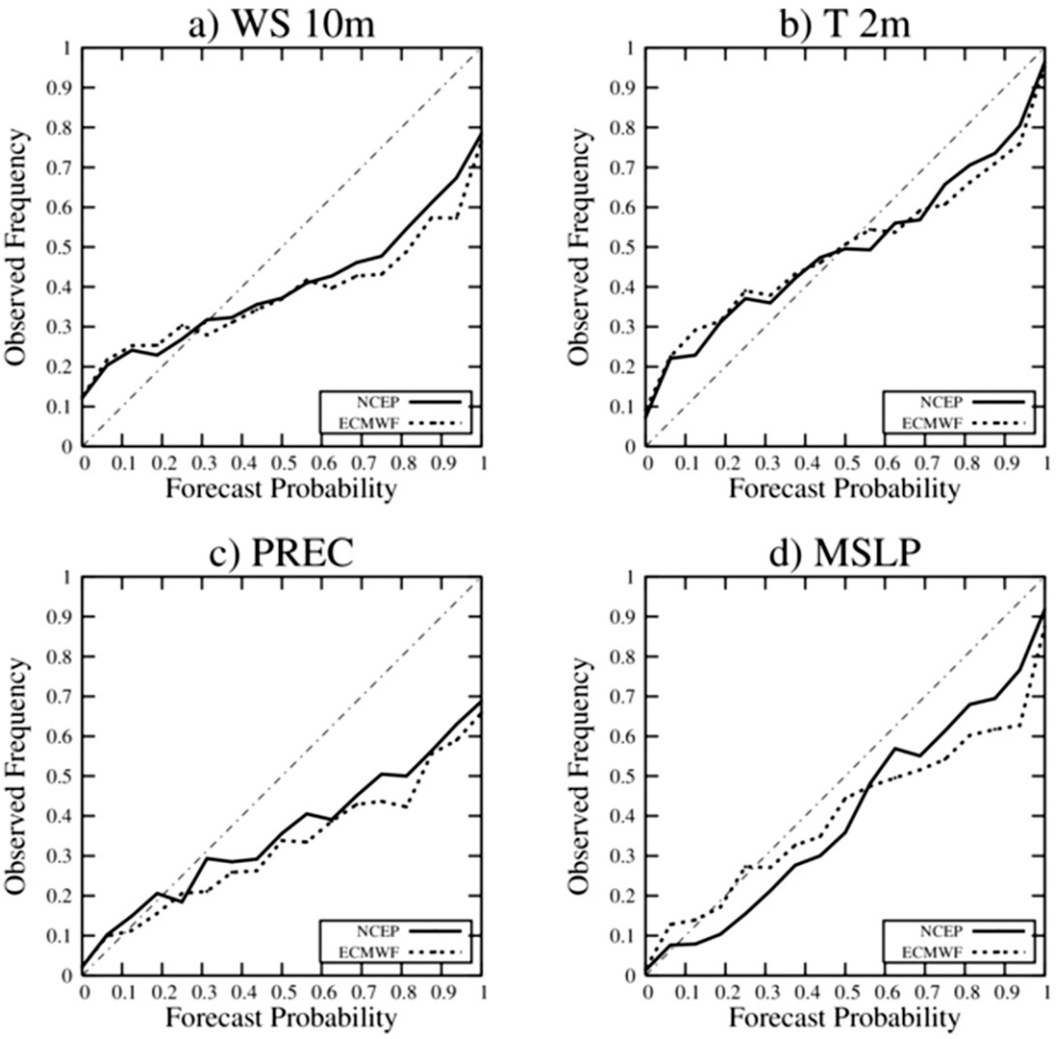

FIG. 5. Reliability diagrams of the NCEP (solid line) and ECMWF (dashed line) experiments for (a) WS10m > 4 m s^{-1}, (b) T2m anomaly > 0°C, (c) PREC > 5 mm (12 h)$^{-1}$, and (d) MSLP > 1015 hPa averaged over the whole verification domain and over the whole verification period (0000 UTC runs, +42-h lead time).

temperature that is slightly larger in ECMWF than in NCEP, and almost no bias for wind speed. As already stated above, NCEP has a strong negative bias for RH, where ECMWF exhibits a much smaller negative bias for short forecast times (Fig. 7c), with the difference almost disappearing for the longest forecast ranges. For *Z*, NCEP is clearly better on the first day, while ECMWF becomes superior for longer forecast ranges (Fig. 7d).

The difference in CRPS with the 95% confidence interval at 850 hPa is presented in Fig. 8. As for the surface verification, positive (negative) values denote a better CRPS for ECMWF (NCEP). For T850 the difference in CRPS is significant in the evenings (+18 and +42 h), in favor of NCEP, as is also the case for T2m (cf. Fig. 8b with Fig. 4b). However, the superiority in CRPS of NCEP for +12-h forecasts is not significant at 850 hPa and actually disappears when using the NCEP analyses as a reference (not shown).

The CRPS differences for wind speed at 850 hPa are only significant at +6 and +54 h with a better CRPS in ECMWF (Fig. 8a). This is in contrast to the results for WS10m at the surface, where NCEP was significantly better for the whole forecast range. The significance disappears when using NCEP analyses as a reference (not shown), with the sole exception being for wind speed at +54 h.

The CRPS for RH in ECMWF is significantly better in the early forecast period and is a result of the larger RMSE of NCEP at the beginning. When using NCEP analyses as a reference, the difference in CRPS is only significant up to +12 h. For forecast ranges at and

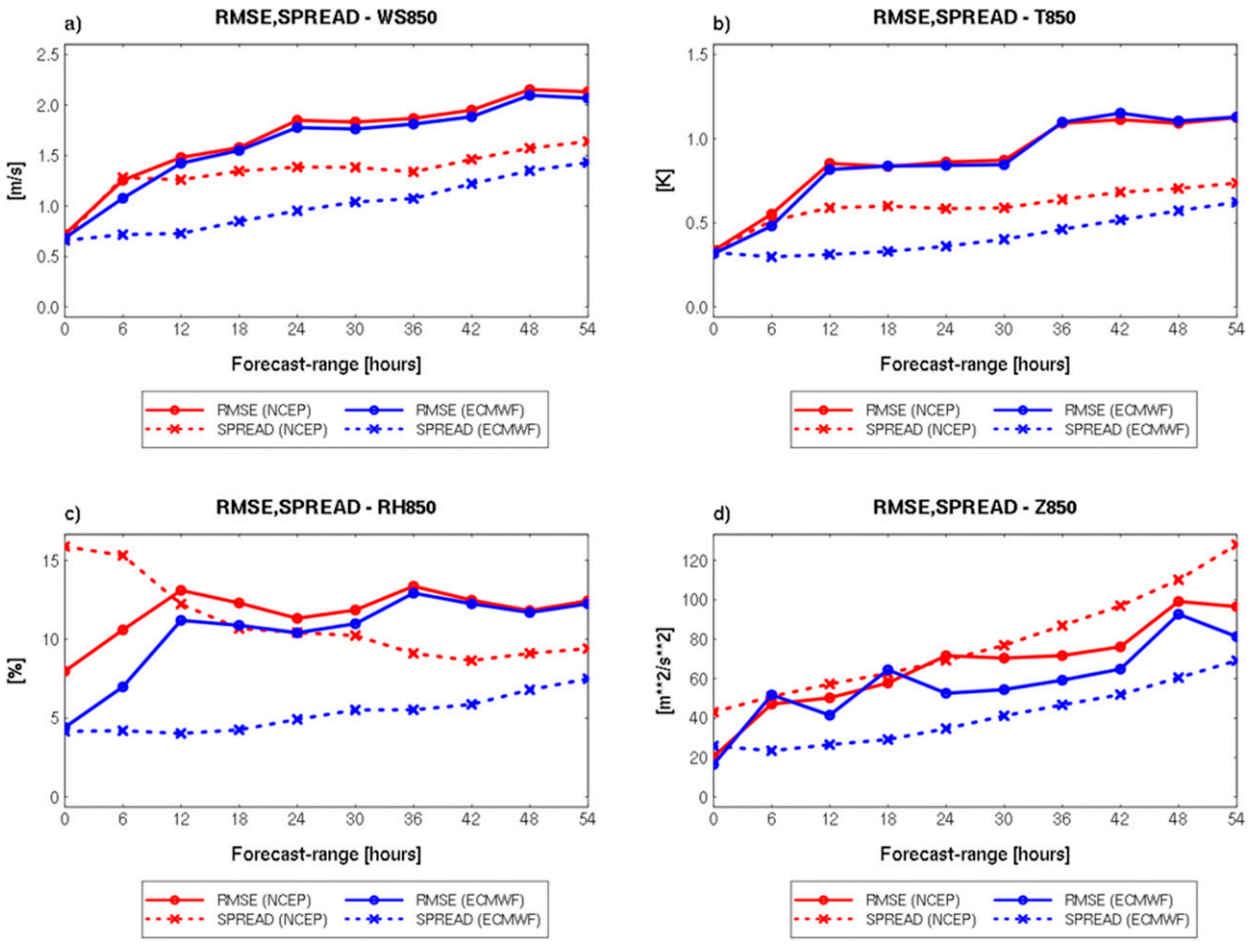

FIG. 6. RMSEs (solid lines) of the ensemble mean and spread (dashed) of the NCEP (red) and ECMWF (blue) experiments for (a) WS850, (b) T850, (c) RH850, and (d) Z850, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

beyond +36 h the difference is significant at the 95% level in favor of NCEP, although the difference is rather small.

The difference in CRPS for geopotential at 850 hPa for lead times from +24 h up to +42 h, and for lead time +54 h is significant in favor of ECMWF (Fig. 8d). At shorter lead times none of the experiments clearly outperforms the other, and there is no significant difference when NCEP reference analyses are taken into account (not shown).

Figure 9 shows the RMSE and the spread at 500 hPa for the two experiments and Fig. 10 shows the bias of the ensemble means of the experiments. At 500 hPa the results differ for temperature and relative humidity from those at 850 hPa. The RMSE for the temperature of the ensemble mean in experiment ECMWF is considerably smaller than for NCEP, whereas at 850 hPa the RMSEs were almost equal (cf. Fig. 6b with Fig. 9b). This can be explained by a distinctively smaller bias in ECMWF compared to NCEP (Fig. 10b). In addition, the difference in spread at 500 hPa is smaller than at 850 hPa, which leads, together with the smaller RMSE, to a better RMSE-to-spread relation in ECMWF (Fig. 9b).

For RH at 500 hPa the difference between spread and RMSE for ECMWF is also comparable to values at 850 hPa (cf. Fig. 6c with Fig. 9c) whereas for NCEP the spread is larger than the RMSE; hence, the forecasts are overdispersive up to +24 h. In addition to the inappropriate spread, the NCEP results also show a considerably high positive bias (Fig. 10c), in contrast to the large negative bias at 850 hPa, especially in the short forecast range. In contrast, forcing by ECMWF leads to a more or less bias-free ensemble mean (Fig. 10c).

The patterns of behavior of RMSE and spread for geopotential and wind speed at 500 hPa are comparable to those discussed for 850 hPa (Figs. 9a,d). This also holds for the bias of wind speed with a small negative bias for ECMWF, whereas both experiments show a

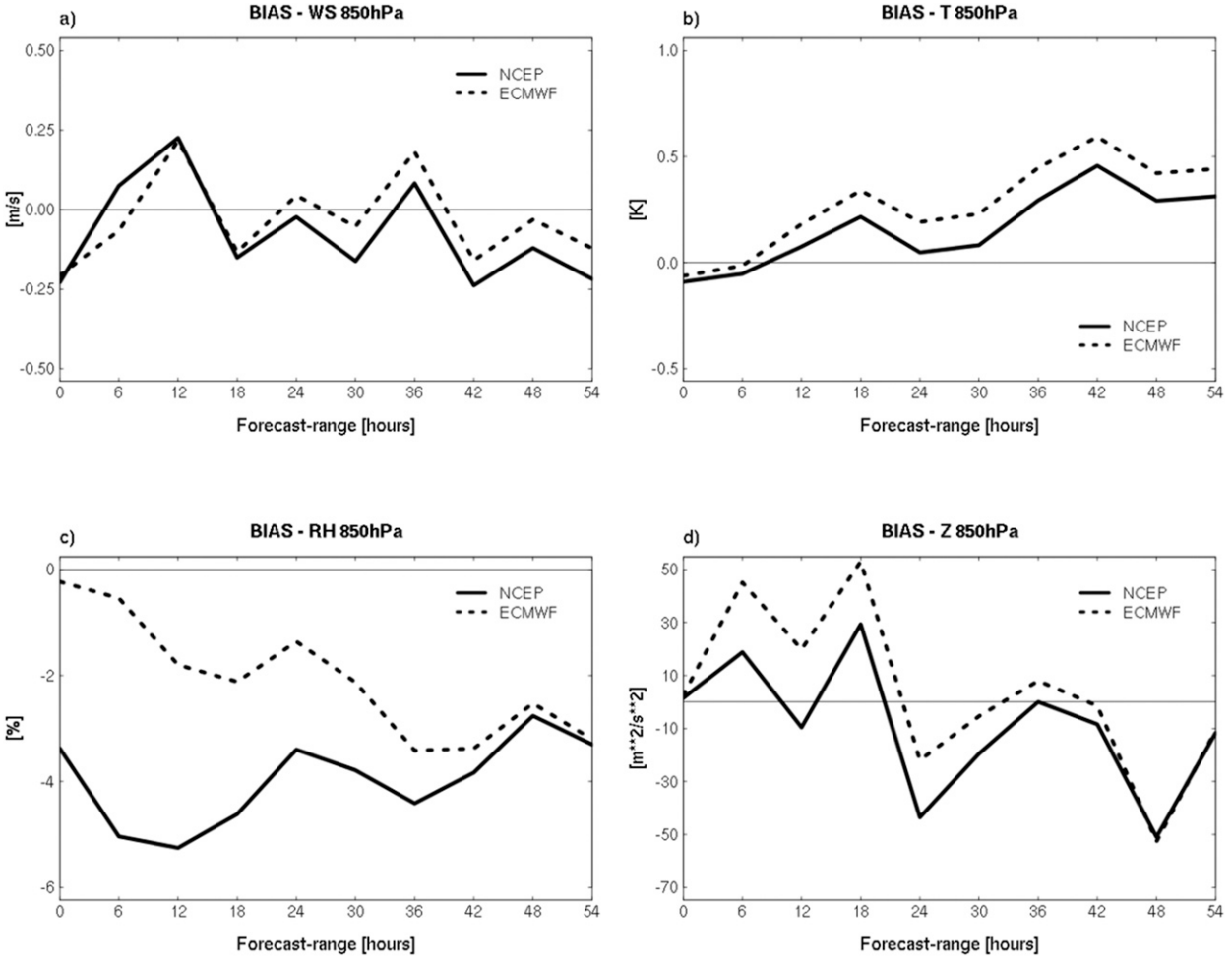

FIG. 7. Biases of the ensemble mean of the NCEP (solid line) and ECMWF (dotted line) experiments for (a) WS850, (b) T850, (c) RH850, and (d) Z850, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

positive bias in Z at 500 hPa in contrast to a tendency toward a negative bias at 850 hPa (cf. Fig. 7d with Fig. 10d).

The difference in CRPS with the 95% confidence interval at 500 hPa is presented in Fig. 11. The difference in CRPS between NCEP and ECMWF at 500 hPa for wind speed is similar to the results shown at 850 hPa (cf. Fig. 8a with Fig. 11a), but with statistically significant differences only up to +18 h. For temperature and relative humidity at 500 hPa the difference is clearly positive, implying a better CRPS in ECMWF, and the results are statistically significant over the whole forecast range. This can be clearly attributed to the smaller RMSE in ECMWF compared to NCEP. The significance for relative humidity and temperature also holds when using NCEP analyses as a reference. For geopotential at 500 hPa, ECMWF has a better CRPS for the whole forecast range as well, although it is not always statistically significant. The significance for 12-h forecasts is not valid when verifying against NCEP analyses.

Overall, the experiment ECMWF performs clearly better at 500 hPa, while at 850 hPa, experiment NCEP is somewhat better. The biggest difference in the performance of the experiments between 500 and 850 hPa is in the temperature. Taking into account the results for the surface variables, it may be concluded that forcing of the ICs and LBCs by ECMWF-EPS performs better than using GEFS in the high atmosphere, while close to and at the surface forcing by GEFS can be advantageous.

4. Summary and conclusions

In this study the forecast performance of ALADIN-LAEF using ICs and LBCs from two of the world's leading global EPSs, ECMWF-EPS and GEFS, was evaluated to investigate the impact of the driving global

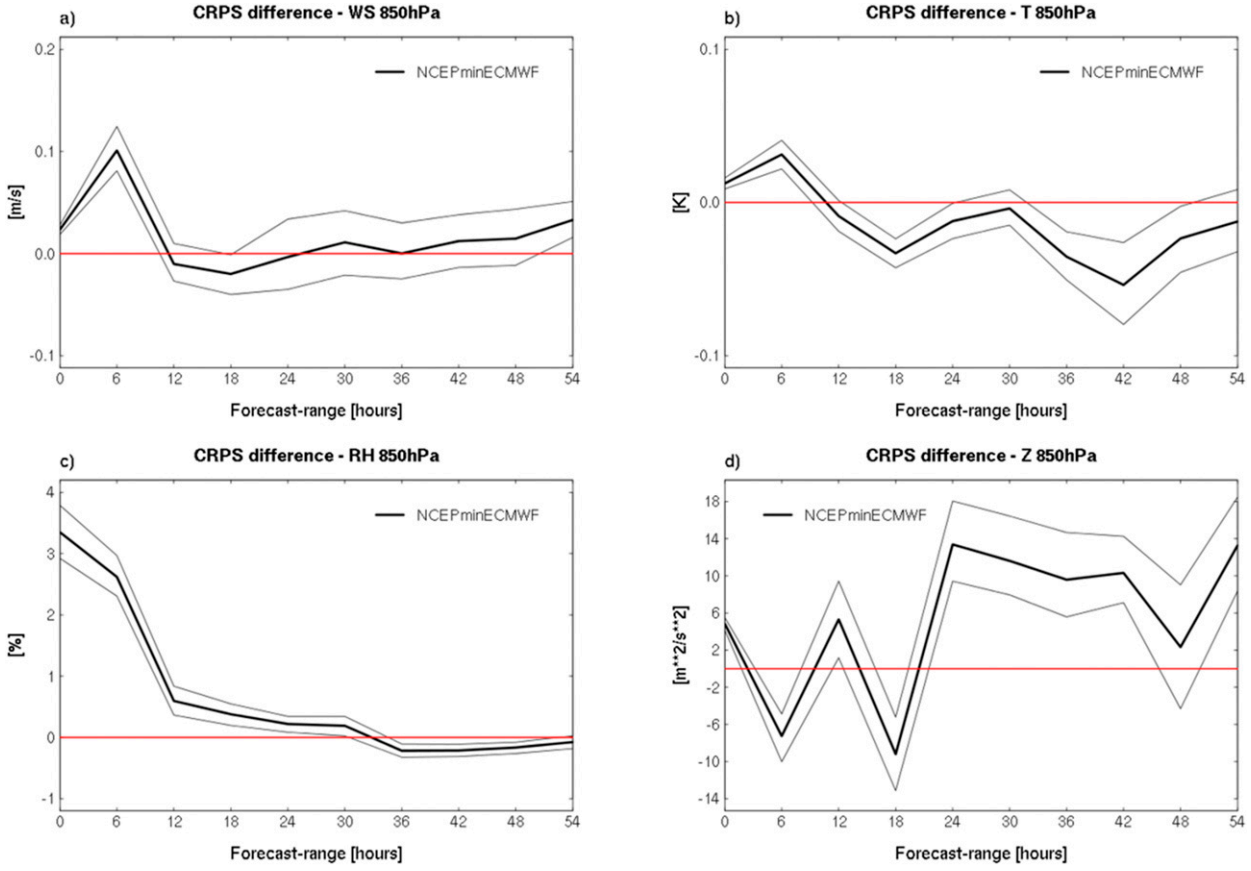

FIG. 8. Differences in the CRPSs of NCEP minus ECMWF (thick solid line) for (a) WS850, (b) T850, (c) RH850, and (d) Z850, averaged over the whole verification domain and over the whole verification period (0000 UTC runs). The corresponding 95% confidence interval is shown with thin lines.

ensemble for a regional EPS. The ALADIN-LAEF dynamical downscaling configuration was applied over a 51-day period during summer 2010 in this study.

A comparison of two experiments, the first using ECMWF-EPS as initial and lateral boundary conditions (ECMWF) and the second using GEFS as driving global EPS (NCEP), is presented. So that scores especially close to the surface are not influenced by a possible inconsistency between the surface scheme of the driving model and the ALADIN model, the ARPEGE surface scheme was used in all experiments. Since ARPEGE is the global counterpart to ALADIN, it uses the same surface scheme as ALADIN and therefore is more consistent than the simulations used in NCEP and ECMWF.

The comparison of using ECMWF-EPS and GEFS (experiments ECMWF and NCEP, respectively) both as initial and boundary conditions showed that the forecast quality of ALADIN-LAEF differs in the upper atmosphere from the lower atmosphere and near the surface. For the near-surface level where 10-m wind speed, 2-m temperature, 12-hourly precipitation, and mean sea level pressure were verified, forcing by GEFS leads to better performance of ALADIN-LAEF than does forcing by ECMWF-EPS. While the errors (RMSE and bias) are comparable in both experiments, the spread is considerably larger in the NCEP experiment. This leads to better statistical consistency for the ALADIN-LAEF forecasts, which can be also seen in CRPS and reliability diagrams. For CRPS the superiority of NCEP is shown to be statistically significant at the 95% level for 10-m wind speed and precipitation, as well as for some forecast ranges of 2-m temperature and mean sea level pressure.

At 850 hPa, wind speed, temperature, relative humidity, and geopotential are verified. The results are rather neutral. Except for temperature, the RMSEs are smaller in ECMWF. However, the spread is still considerably larger and more ideal in the NCEP experiment, except for geopotential where there is a clear overdispersion. The ECMWF experiment has a better CRPS for wind speed, but the NCEP experiment has a better CRPS for 850-hPa temperature.

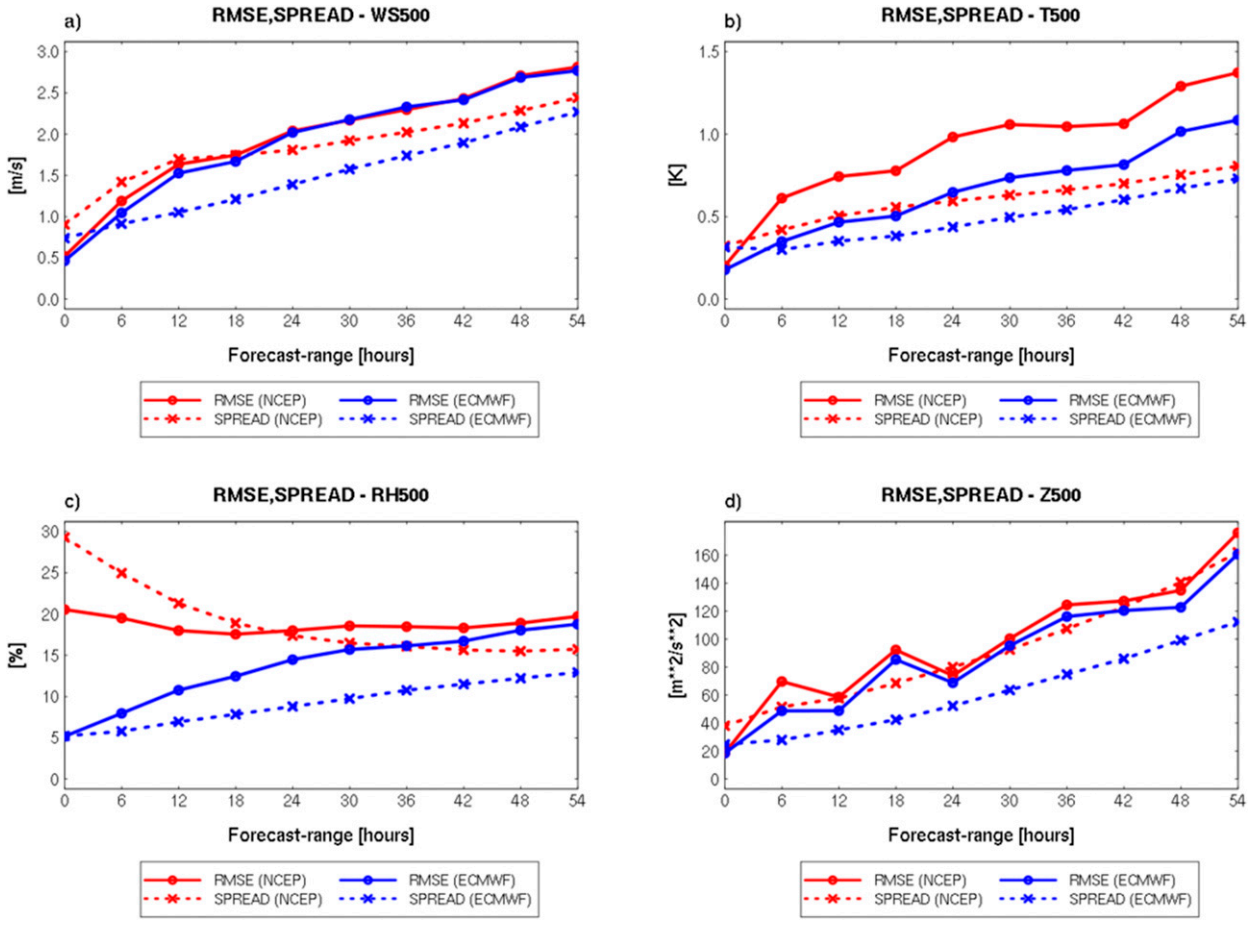

FIG. 9. RMSEs (solid lines) of the ensemble mean and spread (dashed) of the NCEP (red) and ECMWF (blue) experiments for (a) WS500, (b) T500, (c) RH500, and (d) Z500, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

At 500 hPa, forcing by ECMWF-EPS is clearly superior to forcing by GEFS. This is mainly due to better RMSEs in ECMWF across almost all variables and forecast ranges. The outcome of this is a better CRPS for ALADIN-LAEF forecasts driven by ECMWF-EPS. Its superiority is statistically significant for all variables in the early forecast ranges (up to 24 h). When ALADIN-LAEF is forced by GEFS, an overdispersion can be found for all verified variables at 500 hPa at the initial time and for wind speed and relative humidity up to a forecast range of 1 day ahead.

The major difference identified when using ECMWF-EPS and GEFS as forcing for the global EPS is that the ALADIN-LAEF forecasts have larger spread with GEFS. This may be due to the methods employed by the GEFS and ECMWF-EPS in the generation of the IC perturbations. GEFS uses the ensemble transform technique, which can be considered to be an improved breeding scheme, while ECMWF-EPS applies the SV technique combined with EDA. Magnusson et al. (2008) found that the breeding vector method produces much larger spread at the earlier forecast range than do the singular vectors, in order to achieve sufficient spread in the medium range. When serving as the driving model for a regional ensemble, this turns out to be beneficial for the performance at the surface and in the lower atmosphere, up to approximately 850 hPa, where most regional EPSs are quite underdispersive, as in ALADIN-LAEF. The larger spread leads to better statistical reliability in the system, by demonstrating an improved RMSE-to-spread relation. Aloft (e.g., 500 hPa), the underdispersion in ALADIN-LAEF is not so strong and the forcing by GEFS results in an ''overshooting'' of spread for some variables and forecast ranges, most pronounced for wind speed and relative humidity in the first forecast hours. Using ECMWF-EPS as the driving ensemble shows a smaller error at 500 hPa,

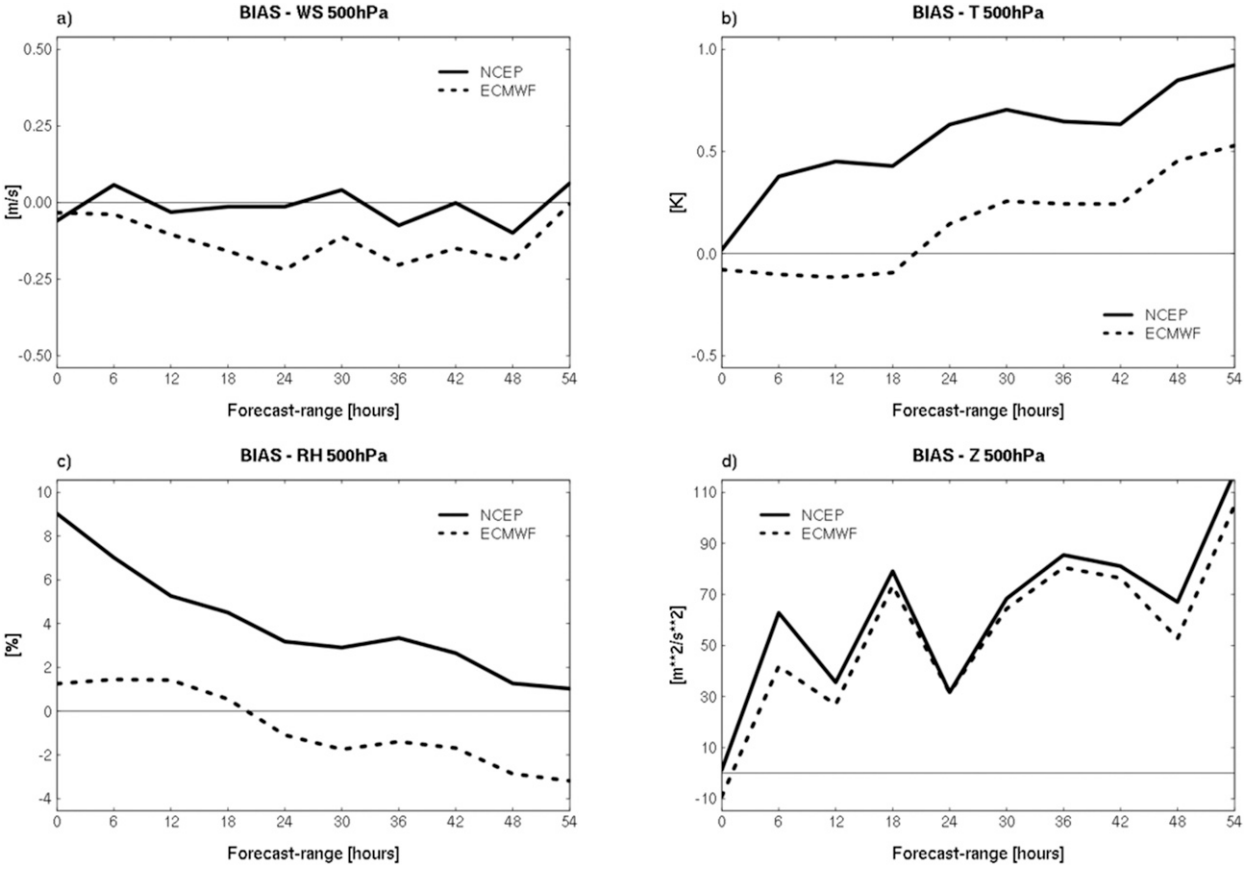

FIG. 10. Biases of the ensemble mean of the NCEP (solid line) and ECMWF (dotted line) experiments for (a) WS500, (b) T500, (c) RH500, and (d) Z500, averaged over the whole verification domain and over the whole verification period (0000 UTC runs).

which is clearly seen in RMSE, resulting in better forecast performance for ALADIN-LAEF.

Two additional experiments were designed but not shown to distinguish between the impact of downscaled regional IC and LBC perturbations from different global EPSs. In one experiment the IC perturbations from ECMWF were used while the LBC perturbations were retrieved from GEFS, and the opposite is true for the second experiment.

It was found that the skill of ALADIN-LAEF is largely determined by the ICs while an impact of LBCs in the spread especially for sea level pressure and geopotential was clearly visible with increasing forecast length. But, in general, the combination of NCEP ICs and ECMWF LBCs or vice versa does not have a clear impact on the forecast performance of ALADIN-LAEF. Certainly, the impact of LBCs on the quality of a regional ensemble is clearly dependent on the domain size and the weather situations. For example, in stable conditions with a weak jet the information from the boundaries will take a long time to propagate deep into the integration domain of the model. In such cases the ICs have a bigger impact on the forecast quality than the LBCs.

The results of this study suggest that the best global EPS might not always provide the most appropriate regional IC and LBC perturbations for regional ensembles. Taking into consideration the strong underdispersive character of the regional ensembles near the surface, a global ensemble, having IC perturbations that effectively sample the analysis error and grow fast from initial time, may considerably improve the forecast quality of the regional EPS.

The results presented in this study are only based on a 1.5-month summer period, and therefore investigations of other time periods might lead to quantitatively different results, but we assume that they qualitatively hold also for other periods. This is based on the fact that the results presented can be attributed to known features of the global EPS used (e.g., the bigger spread in the first forecast days of GEFS compared to ECMWF-EPS), which are independent from seasons.

For institutes running a regional EPS operationally and that have different global EPS available to provide ICs and

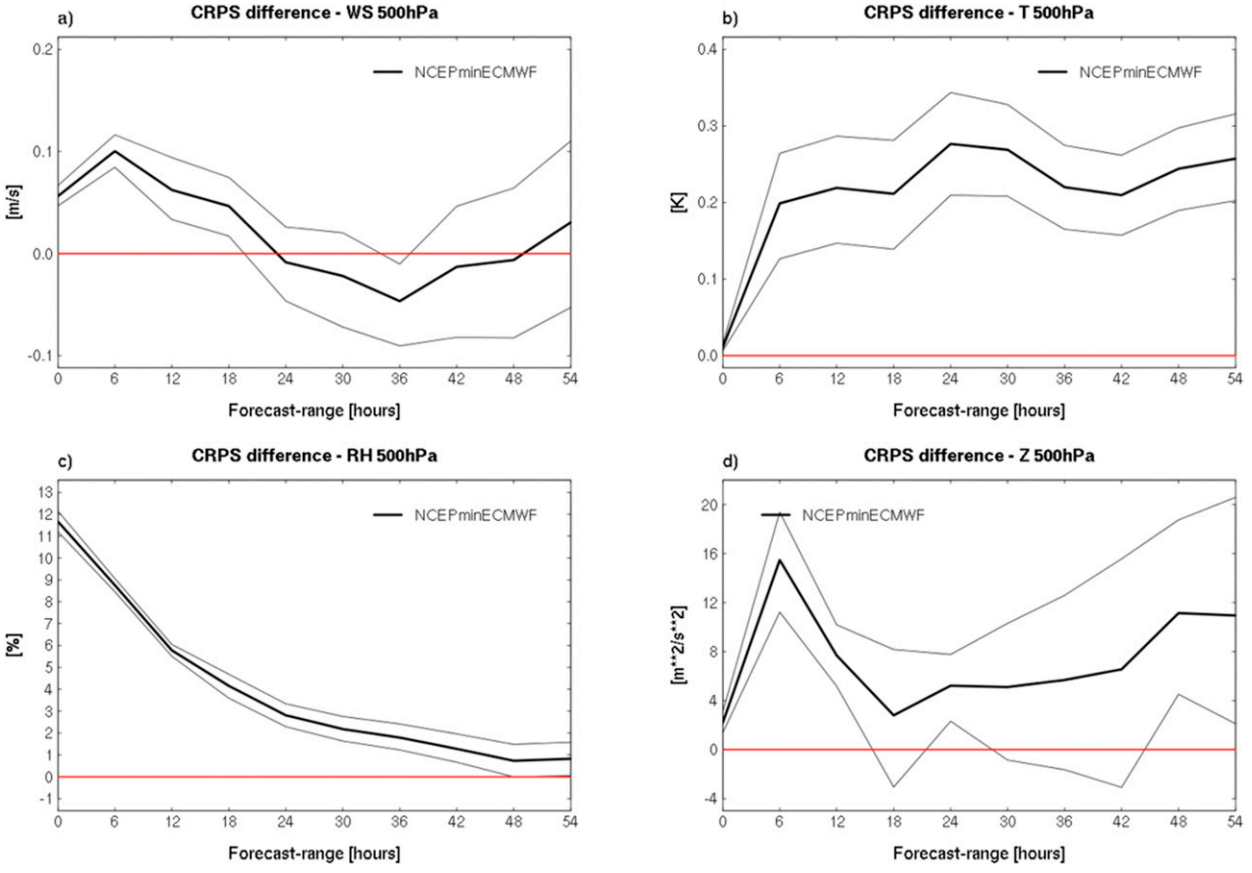

FIG. 11. Differences in the CRPSs of NCEP minus ECMWF (thick solid line) for (a) WS500, (b) T500, (c) RH500, and (d) Z500, averaged over the whole verification domain and over the whole verification period (0000 UTC runs). The corresponding 95% confidence interval is shown with thin lines.

LBCs, it is worth exploring in more detail which one leads to the best performance among the regional EPS. Although this article presented data from a few years ago and the global EPSs have improved in the meantime, their main features (e.g., the IC and model perturbation methods of both ECMWF-EPS and GEFS) have remained the same over the last five years. Therefore, similar results could still be expected. Furthermore, the main conclusion that the best global EPS is not necessarily the most appropriate EPS to force a regional EPS will continue to hold in any case.

In the operational ALADIN-LAEF, lateral boundary perturbations from ECMWF-EPS are, and will be, used because of operational restrictions. The initial perturbations are generated by a breeding–blending cycle in ALADIN-LAEF and a surface ensemble data assimilation method is implemented to mitigate the drawbacks pointed out in this study by using ECMWF-EPS to drive an EPS.

Acknowledgments. We gratefully acknowledge all of the RC LACE/ALADIN colleagues who have contributed to the development of ALADIN-LAEF. Special thanks to Yuejian Zhu for providing NCEP GFS information and to Martin Steinheimer from AustroControl for technical assistance. ECMWF has provided the computer facilities and technical help in implementing ALADIN-LAEF on the ECMWF High Performance Computing Facility (HPCF). The work has been partly funded by RC LACE and ALADIN.

REFERENCES

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765, doi:10.1175/1520-0469(1999)056<1748:ETAAO>2.0.CO;2.

——, B. Etherton, and S. Majumdar, 2001: Adaptive sampling with ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

——, T. R. Holt, J. Nachamkin, S. Chen, J. G. McLay, J. D. Doyle, and W. T. Thompson, 2009: Regional ensemble forecasts using the ensemble transform technique. *Mon. Wea. Rev.*, **137**, 288–298, doi:10.1175/2008MWR2559.1.

Bonavita, M., L. Raynaud, and L. Isaksen, 2011: Estimating background-error variances with the ECMWF ensemble of

data assimilations system: Some effects of ensemble size and day-to-day variability. *Quart. J. Roy. Meteor. Soc.*, **137**, 423–434, doi:10.1002/qj.756.

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, doi:10.1175/2010BAMS2853.1.

Bowler, N. E., and K. R. Mylne, 2009: Ensemble transform Kalman filter perturbations for a regional ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **135**, 757–766, doi:10.1002/qj.404.

——, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, doi:10.1002/qj.234.

Buizza, R., and T. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456, doi:10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

——, M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, doi:10.1002/qj.49712556006.

——, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, doi:10.1175/MWR2905.1.

——, M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066, doi:10.1002/qj.346.

——, ——, ——, and J. Haseler, 2010: Combined use of EDA- and SV-based perturbations in the EPS. *ECMWF Newsletter*, No. 123, ECMWF, Reading, United Kingdom, 22–28.

Caron, J. F., 2013: Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: A case study. *Mon. Wea. Rev.*, **141**, 356–374, doi:10.1175/MWR-D-12-00051.1.

Du, J., and M. S. Tracton, 2001: Implementation of a real-time short-range ensemble forecasting system at NCEP: An update. Preprints, *Ninth Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc., 355–356. [Available online at http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html.]

——, G. DiMego, M. S. Tracton, and B. Zhou, 2003: NCEP short-range ensemble forecasting (SREF) system: Multi-IC, multi-model and multi-physics approach. *Research Activities in Atmospheric and Oceanic Modeling*, J. Cote, Ed., CAS/JSC Working Group Numerical Experimentation (WGNE) Rep. 33, WMO/TD-No. 1161, 5.09–5.11.

Frogner, I. L., H. Haakenstad, and T. Iversen, 2006: Limited-area ensemble predictions at the Norwegian Meteorological Institute. *Quart. J. Roy. Meteor. Soc.*, **132**, 2785–2808, doi:10.1256/qj.04.178.

Froude, L. S. R., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **25**, 819–836, doi:10.1175/2010WAF2222326.1.

Garcia-Moya, J. A., A. Callado, P. Escriba, C. Santos, D. Santos-Munoz, and J. Simarro, 2011: Predictability of short-range forecasting: A multimodel approach. *Tellus*, **63A**, 550–563, doi:10.1111/j.1600-0870.2010.00506.x.

Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.*, **100**, 168–177, doi:10.1016/j.atmosres.2010.12.008.

Hacker, J. P., and Coauthors, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus*, **63A**, doi:10.3402/tellusa.v63i3.15814.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, doi:10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hu, D., Z. Toth, Y. Zhu, and W. Yang, 2008: A stochastic parameterization scheme in NCEP global ensemble forecast system. Preprints, *19th Conf. on Probability and Statistics*, New Orleans, LA, Amer. Meteor. Soc., 1.1. [Available online at http://ams.confex.com/ams/pdfpapers/134165.pdf.]

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Tech. Memo. 636, 44 pp.

Iversen, T., A. Deckmyn, C. Santos, K. Sattler, J. B. Bremnes, H. Feddersen, and I. L. Frogner, 2011: Evaluation of 'GLAMEPS'—A proposed multimodel EPS for short range forecasting. *Tellus*, **63A**, 513–530, doi:10.1111/j.1600-0870.2010.00507.x.

Leutbecher, M., and T. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, doi:10.1016/j.jcp.2007.02.014.

Li, X., M. Charron, L. Spacek, and G. Candille, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbation. *Mon. Wea. Rev.*, **136**, 443–462, doi:10.1175/2007MWR2109.1.

Magnusson, L., M. Leutbecher, and E. Källen, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **136**, 4092–4104, doi:10.1175/2008MWR2498.1.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: Validation of the methodology and verification. *Nonlinear Processes Geophys.*, **12**, 527–536, doi:10.5194/npg-12-527-2005.

Montani, A., D. Cesari, C. Marsigli, and T. Paccagnella, 2011: Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: Main achievements and open challenges. *Tellus*, **63A**, 605–624, doi:10.1111/j.1600-0870.2010.00499.x.

Palmer, T., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050, doi:10.1002/qj.334.

Smet, G., P. Termonia, and A. Deckmyn, 2012: Added economic value of limited area multi-EPS weather forecasting applications. *Tellus*, **64A**, 18 901, doi:10.3402/tellusa.v64i0.18901.

Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, doi:10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, doi:10.1175/MWR3187.1.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NCEP: The generation of perturbation. *Bull. Amer. Meteor. Soc.*,

74, 2317–2330, doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Wang, X., and C. Bishop, 2003: A comparison of breeding and ensemble Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

Wang, Y., A. Kann, M. Bellus, J. Pailleux, and C. Wittmann, 2010: A strategy for perturbing surface initial conditions in LAMEPS. *Atmos. Sci. Lett.*, **11**, 108–113, doi:10.1002/asl.260.

——, and Coauthors, 2011: The Central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quart. J. Roy. Meteor. Soc.*, **137**, 483–502, doi:10.1002/qj.751.

——, M. Bellus, J. F. Geleyn, X. Ma, W. Tian, and F. Weidle, 2014: A new method for generating initial perturbations in a regional ensemble prediction system: Blending. *Mon. Wea. Rev.*, **142**, 2043–2059, doi:10.1175/MWR-D-12-00354.1.

Warner, T., R. Peterson, and R. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617, doi:10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79, doi:10.1111/j.1600-0870.2007.00273.x.

Weidle, F., Y. Wang, W. Tian, and T. Wang, 2013: Validation of strategies using clustering analysis of ECMWF-EPS for initial perturbations in a limited area model ensemble prediction system. *Atmos.–Ocean*, **51**, 284–295, doi:10.1080/07055900.2013.802217.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.

Zhang, H., J. Chen, X. Zhi, and Y. Wang, 2015: A comparison of ETKF and downscaling in a regional ensemble prediction system. *Atmosphere*, **6**, 341–360, doi:10.3390/atmos6030341.