Tethys

Journal of Mediterranean
Meteorology & Climatology

Dynamic downscaling of the NCEP EPS forecast using the PROMES limited area model

E. Hagel, V. E. Gil, C. Tejeda and M. de Castro

Instituto Meteorológico Regional de Castilla-La Mancha (iMetCam), Avda. Carlos III, s/n, 45071, Toledo

Correspondence to: edit.hagel@uclm.es

Abstract

When making numerical weather predictions, it is important to forecast not only the future state of the atmosphere, but also to predict the uncertainty related to this forecast. Keeping this in mind, research has started at iMetCam in order to develop a limited area ensemble prediction system. As a start, the simple dynamical downscaling approach was tried. Initial conditions and lateral boundary conditions were provided by the global ensemble system of NCEP and the PROMES limited area model was used for the downscaling. In this paper the first results and the future plans of this experiment are presented. Results show that both systems (global and limited area) are lacking spread, at least for the verification area in question, which indicates that additional perturbations are desirable, which will be the direction of our future work.

Key words: ensemble prediction system, probabilistic forecast, limited area model, dynamical downscaling, verification

1 Introduction

In order to predict the future state of the atmosphere, the use of mathematical models is required. Forecasts are made by solving a set of partial differential equations, the so-called primitive equations. These equations are nonlinear and, therefore, impossible to solve analytically. Because of the nonlinear nature of the equations, even small errors in the initial conditions can lead to large forecast errors. A possible solution to this problem is to run a set, or as usually called, an ensemble of forecasts, each starting from a slightly different initial condition (IC), thus, they are equally likely realizations of the "true" atmospheric state. The advantage of this method is clear: the spread of the ensemble members can provide useful information on the predictability of the atmospheric state, and a probability value can be assigned to different weather events. In other words, not only the future state of the atmosphere can be forecasted, but one can also predict the uncertainty related to this forecast. One should not forget however, that the IC error is not the only source of possible forecast errors, but there are also errors in the models themselves.

Since its first operational application in 1992 (Buizza et al., 1993; Toth and Kalnay, 1997), ensemble forecasting has become a widely used technique by many meteorological services around the world (e.g. Australia: Bourke et al., 2004; Canada: Charron et al., 2010; France: Nicolau, 2002; United Kingdom: Bowler et al., 2008; USA: Wei et al., 2008; European Centre for MediumRange Weather Forecasts (ECMWF): Buizza et al., 2007). Despite its obvious benefits, it was used only on global scales for a long time. In the last decade, however, intensive research has started to apply the ensemble method in limited area forecasting as well (see e.g. Molteni et al., 2001; Marsigli et al., 2008; Gebhardt et al., 2008; Bowler et al., 2008; Frogner et al., 2006; Iversen et al., 2009; Hagel, 2010; García-Moya et al., 2011; or Wang et al., 2010 and 2011).

Keeping all this in mind, since the beginning of its activity, iMetCam (Regional Meteorological Institute of Castilla-La Mancha, http://imetcam.uclm.es) is offering probability information on precipitation existence. At present, this probability is calculated using a time-lagged ensemble. In order to improve these forecasts and to have better representation of the forecast uncertainties, research started with the dynamical downscaling of the
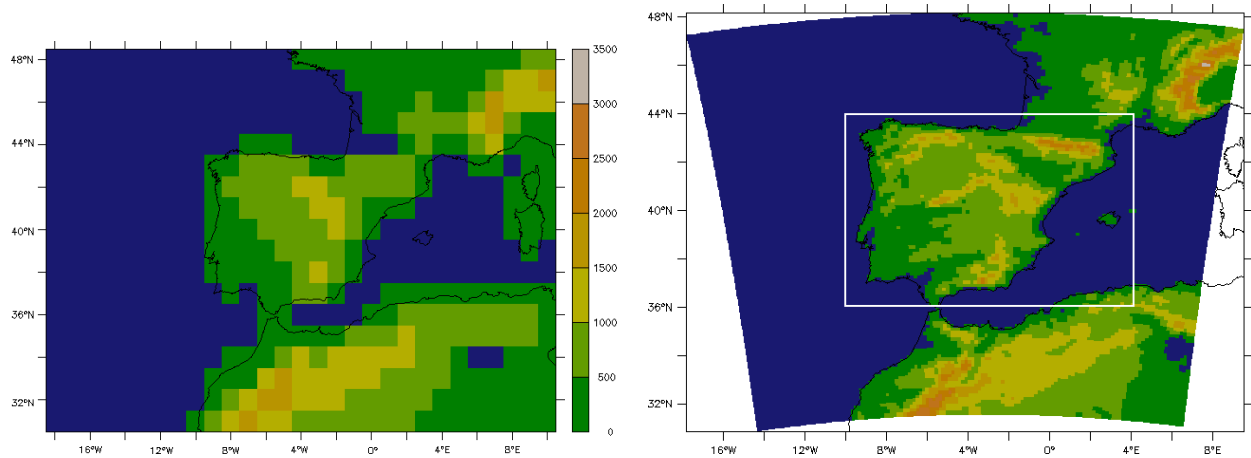
Figure 1. (a, left) Orography of the global model (1 degree resolution) over the area of interest. (b, right) Orography of the PROMES model (15 km resolution) and verification area (white rectangle).

National Centers for Environmental Prediction (NCEP) global Ensemble Prediction System (EPS) forecasts using the limited area model PROMES (is an acronym for *PROnóstico a MESoescala*). The final aim is to establish an operational Limited Area Model Ensemble Prediction System (LAMEPS) for the Iberian Peninsula, and to see how it can improve the predictions of the existing global ensemble forecasting systems.

In this article we present the very first results of this LAMEPS. The applied models and the characteristics of the system are described in Section 2. Verification results are presented in Section 3. Finally, conclusions and future plans are outlined in Section 4.

2   Applied models and the characteristics of the system

For the present study, ICs and lateral boundary conditions (LBCs) were provided by the EPS based on the Global Forecast System (GFS) of NCEP. The global EPS of NCEP runs four times a day. For the present study, the 00 UTC predictions were used. At the time of the experiments, the model was integrated at T126 spectral resolution with 28 levels in the vertical. Through the Internet, data was available on $1° \times 1°$ grid (Figure 1a) and 26 vertical levels. The forecast length was 16 days with a 6 hours output frequency. The system had 20 perturbed members and an unperturbed control member. Initial condition perturbations were generated using the Ensemble Transform with Rescaling (ETR, Wei et al., 2008) method. The optimization time was 48 hours and the whole globe was perturbed. The model errors were also taken into account by applying stochastic perturbation during the forecast.

The dynamical downscaling of the global EPS forecasts was performed by the PROMES limited area model (LAM). PROMES (Castro et al., 1993) is a state-of-the-art primitive

equation model, which is hydrostatic and fully compressible. ICs and LBCs were provided every 6 hours. The integration domain was covering the Iberian Peninsula, the adjacent sea areas, part of North Africa and the south of France (Figure 1b). The applied horizontal resolution was 15 km (number of gridpoints: 133×121) and the number of vertical levels was 37. Runs were made once a day starting from 00 UTC up to 7 days for the whole month of February, 2010. Simple dynamical downscaling was applied, no local perturbations were added. When starting the experiments, the possibility of clustering the global ensemble and downscaling only some representative members (RMs) was considered (like, for example in Molteni et al., 2001), but finally rejected. The main reasons of the rejection were the following. Starting from a 20-member global ensemble, in most of the cases it would be difficult to form enough clusters (i.e. select enough RMs) in order to have a good representation of the forecast uncertainties. One should also keep in mind that the downscaling may affect the relevance of the global ensemble forecasts: a RM of the global ensemble system might not be a representative member (after the downscaling) of the limited area system (Branković et al., 2008). Considering all these, it was decided to downscale all global NCEP EPS members, therefore the PROMES ensemble also had 20+1 members.

3   Verification results

One of the essential issues when talking about verification is the definition of "truth". Most often SYNOP (surface synoptic observations) and TEMP (upper level temperature, humidity, wind and geopotential height) data are used to describe the true state of the atmosphere. However, the number of observations is variable in time and space, which means that over specific areas, such as oceans and/or in specific synoptic times (e.g. at 06 and 18 UTC) our knowledge about the atmosphere might be insufficient. Therefore
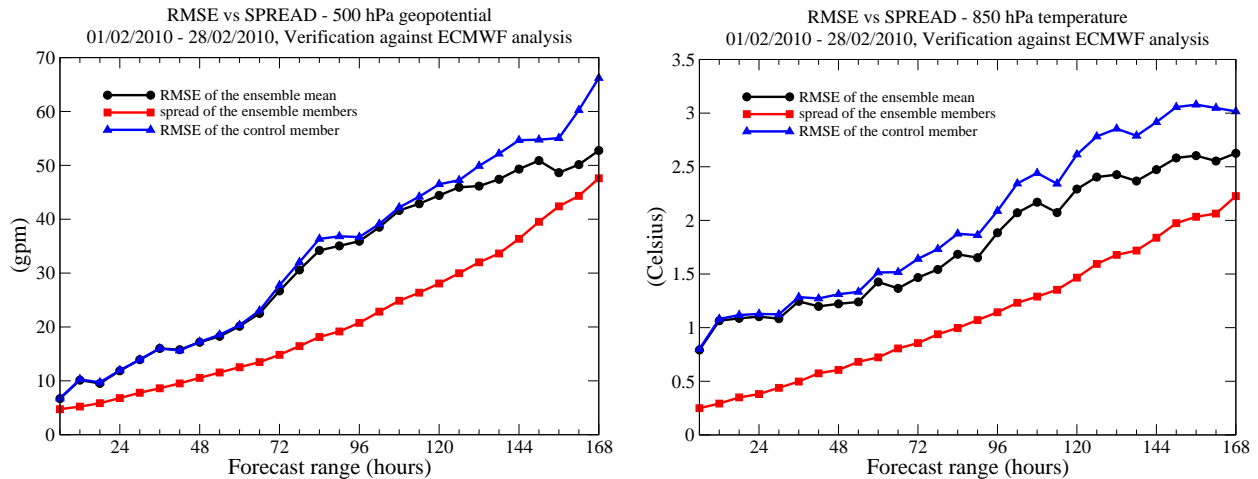
Figure 2. RMSE of the ensemble mean and control member and spread of the ensemble members for PROMES. (a, left) 500 hPa geopotential and (b, right) 850 hPa temperature. Verification was performed on a $0.15° \times 0.15°$ grid.

analyses coming from numerical models or even short-range forecasts are often used for verification purposes. As observation data over the Iberian Peninsula was not publicly available at the time of the experiments, verification of the global and limited area EPS was performed against the operational, high resolution (∼16 km) ECMWF analysis. The applied verification methods included (i) bias and Root Mean Squared Error (RMSE) computations, (ii) comparison of the ensemble spread and the error of the ensemble mean, (iii) ranked histograms and percentage of outliers diagrams and (iv) Relative Operating Characteristics (ROC) and reliability diagrams (for a detailed description of these verification methods, the reader is referred to Jolliffe and Stephenson, 2003). In case of the PROMES EPS, two different horizontal grids were used for verification: $0.15° \times 0.15°$ grid, corresponding to the original resolution of the LAM, and for the comparison with the NCEP EPS system a $1° \times 1°$ grid (the ECMWF analysis was also interpolated to these grids). In case of the global ensemble system only the $1° \times 1°$ grid was used. The verified parameters were temperature, geopotential, specific humidity and wind speed on several pressure levels (500, 700, 850, 925 and 1000 hPa). The verification area was covering the Iberian Peninsula (Figure 1b).

3.1 RMSE and the spread of the ensemble

After an initial (linear) phase of the forecast, it is expected that the ensemble mean has lower RMSE values than the control forecast, since the averaging has the effect of filtering out the less predictable features and leaving only the more predictable ones. To test whether the PROMES EPS fulfils this basic requirement, the RMSE of the control forecast and the ensemble mean was compared. For all parameters the ensemble mean had equal or lower RMSE values than the control member. For geopotential, they were almost identical in the first 2.5-3 days (Figure 2a shows the results

for 500 hPa). For temperature, specific humidity and wind speed (only 850 hPa temperature is shown, Figure 2b) the two values started to differ much earlier. This behaviour may suggest that nonlinearity has a stronger effect on these parameters and the initially small perturbations grow faster for temperature, humidity and wind speed than for geopotential.

Another important feature of an ensemble system is the relationship of the spread (standard deviation of the ensemble members, computed around the ensemble mean) and the error (e.g. RMSE) of the ensemble mean. In case of large error, large spread is expected as a sign of less predictability. On the other hand, if the spread is small, it is expected that the situation has good predictability, therefore the error should be small as well. For all parameters the spread was smaller than the RMSE of the ensemble mean (Figure 2; only geopotential and temperature are shown). Although the system lacks spread, it can be positively mentioned that it is continuously growing with time and for most of the parameters it is approaching the RMSE value of the ensemble mean.

The above mentioned behaviours are not unique features of the PROMES EPS, they were detected in other limited area ensemble prediction systems as well. As an example, the ALADIN LAMEPS system of the Hungarian Meteorological Service can be mentioned (Hagel, 2010). Not only the behaviour of the RMSE and the spread shows similarity, but also the values themselves. One should not forget either, that if the global EPS lacks spread (see results later in Section 3.4), it is very difficult for the LAM - without the introduction of local perturbations - to correct this behaviour and produce a good spread-skill relationship.

3.2 Ranked histograms and percentage of outliers diagrams

Ranked histograms and percentage of outliers are useful tools to analyse different characteristics of an ensemble
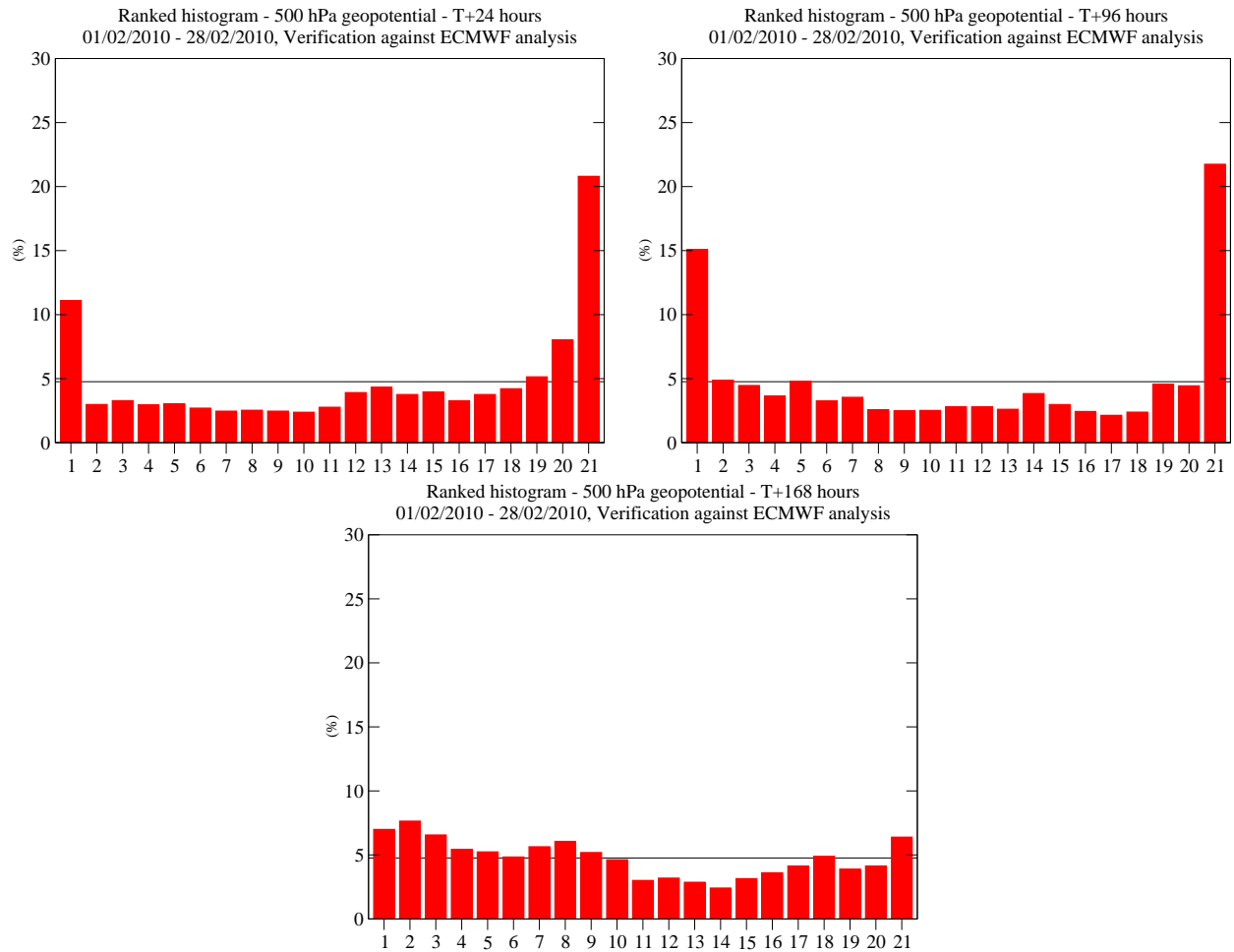
Figure 3. Ranked histograms of the PROMES ensemble system for 500 hPa geopotential at different forecast ranges: (a, upper left) T+24h, (b, upper right) T+96h and (c, lower) T+168h. Verification was performed on a $0.15° \times 0.15°$ grid. The thin horizontal line is the expected value.

system. They show how often the verifying analysis lies out of the interval defined by the (sorted) ensemble members. In an ideal case the distribution of the ranked histogram should be flat. Different shapes of the distribution indicate different behaviours like BIAS, too small or too large spread. Figure 3 shows the ranked histograms of 500 hPa geopotential for different time ranges. One can observe that the U shape is present in every time range, but much less pronounced around day 7. Similar behaviour was found for all examined parameters, except for specific humidity, where the histograms had an L shape, which indicates overestimation (not shown).

Percentage of outliers diagrams were also plotted (Figure 4). These diagrams show the sum of the two outermost intervals of the ranked histograms throughout the whole forecast interval. One can observe that in case of geopotential (Figure 4a), after an initial increase, the percentage of outliers is decreasing and at T+168h is close to the expected value. For the rest of the parameters (only wind speed is shown, Figure 4b) the outlier values are quite high in the ini-

tial forecast ranges, but are decreasing throughout the whole forecast interval. At T+168h, with the exception of 500 and 700 hPa specific humidity, they are all close to the expected value.

All these results - in accordance with the spread-skill diagrams of Figure 2 - suggest that the spread of the ensemble is not sufficient, especially in the early forecast ranges. This indicates that additional (local) perturbations, introduced in the LAM, might be desirable to improve the quality of the LAMEPS system.

3.3 ROC and reliability diagrams

ROC diagrams give us information about the skill of the ensemble forecasts. Hit rates and false alarm rates are calculated for different probability thresholds and entered into a diagram with hit rate on the y-axis and false alarm rate on the x-axis. A perfect system would have all its points in the upper left corner with hit rate equals 1 and false alarm rate equals 0. The integral area under the ROC curve can be cal-
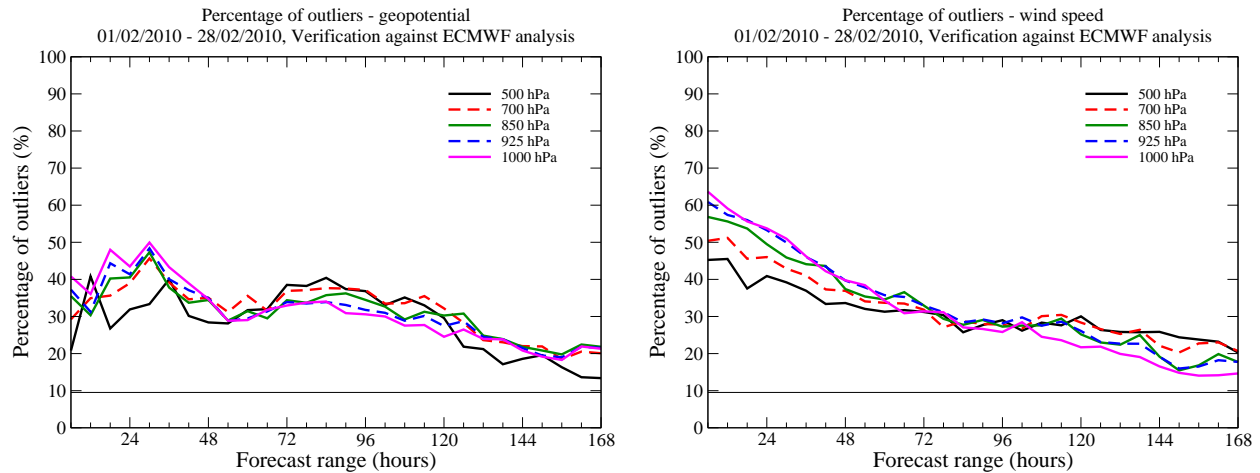
Figure 4. Percentage of outliers diagrams of the PROMES ensemble system for (a, left) geopotential (b, right) wind speed at 500 hPa, 700 hPa, 850 hPa, 925 hPa and 1000 hPa. Verification was performed on a $0.15° \times 0.15°$ grid. The thin horizontal line is the expected value.

culated to represent the skill of the forecast. An integral area of 1 represents perfect forecast, while an integral area less than 0.5 means that the forecast has no skill compared to the use of climatological statistics.

Another useful verification characteristic is the reliability diagram. On average, when an event is forecasted with a given probability, it should occur with the same frequency in the reality. On the reliability diagram, the forecast probabilities are displayed along the x-axis and the observed frequencies for each forecast probability are on the y-axis. If the forecasted probabilities and the observed frequencies agree, the curve lies along the diagonal. Also under- and overestimation can be easily read from the diagram.

ROC and reliability diagrams were plotted for wind speed (Figure 5) and temperature anomaly (Figure 6) with thresholds 5, 10, 15, 20 m s^{-1} and 2°C, 4°C and 8°C respectively. Comparison was made between (i) different thresholds, (ii) different forecast ranges and (iii) different vertical levels.

In case of wind speed the following conclusions can be drawn. The comparison between different thresholds reveals, that the ROC area is larger for low (5 m s^{-1}) and middle (10 m s^{-1}) thresholds, and smaller for high (15 m s^{-1}, 20 m s^{-1}) thresholds. The difference between low and middle thresholds is mainly in the shape of the curve. For low thresholds the Hit Rate (HR) is higher, but also the False Alarm Rate (FAR). For middle thresholds both HR and FAR are somewhat lower, resulting in a very similar ROC area. For high thresholds the FAR is very low, but also the HR; this is in connection with the sample size, which - for such rare events - was too small (Figure 5a). Concerning the reliability diagrams, best results were obtained for lower thresholds (5 m s^{-1}, 10 m s^{-1}) and worst results for the higher ones (Figure 5d). In case of 20 m s^{-1} we can conclude that the sample size was too small. The comparison between different forecast ranges shows that the ROC area is clearly

decreasing with time (Figure 5b), but in terms of reliability, the curves have a similar shape: slight underestimation for lower probabilities and more pronounced overestimations for higher ones (Figure 5e). As for the comparison between different vertical levels, ROC curves are very similar for all levels (Figure 5c), while the reliability diagrams show more variance. However, the curves had similarity in their general behaviour: underestimation for lower probabilities and overestimation for higher ones (Figure 5f).

The results for temperature show many similarities. The comparison between different thresholds reveals that results are better for 4°C than for 2°C, which is mainly due to the lower FAR values. As for the 8°C threshold, the FAR values are low, but also the HR values, which is because of the small sample size (Figure 6a). On the reliability diagrams one can observe underestimation for lower probabilities and overestimation for high probabilities in case of 2°C and 4°C thresholds. For the highest threshold, just like the ROC diagram shows, the sample size was too small (Figure 6d). Comparing different forecast ranges it can be said that the ROC area (Figure 6b) is decreasing with time (more than in case of wind speed), but in terms of reliability (Figure 6e) the curves show a more complex picture. What is common in all of them is the overestimation for higher probabilities, which is most likely in connection with the small sample size of rare events. Finally, the comparison between different vertical levels results in very similar curves in terms of ROC diagram (Figure 6c) and more variance in terms of reliability (Figure 6f). Again, a slight underestimation can be observed for all levels for low probabilities and a more pronounced overestimation for high probabilities.

3.4 Comparing the performance of the limited area and the global EPS systems

Verification results of the global and the limited area ensemble systems were compared to each other (and to the
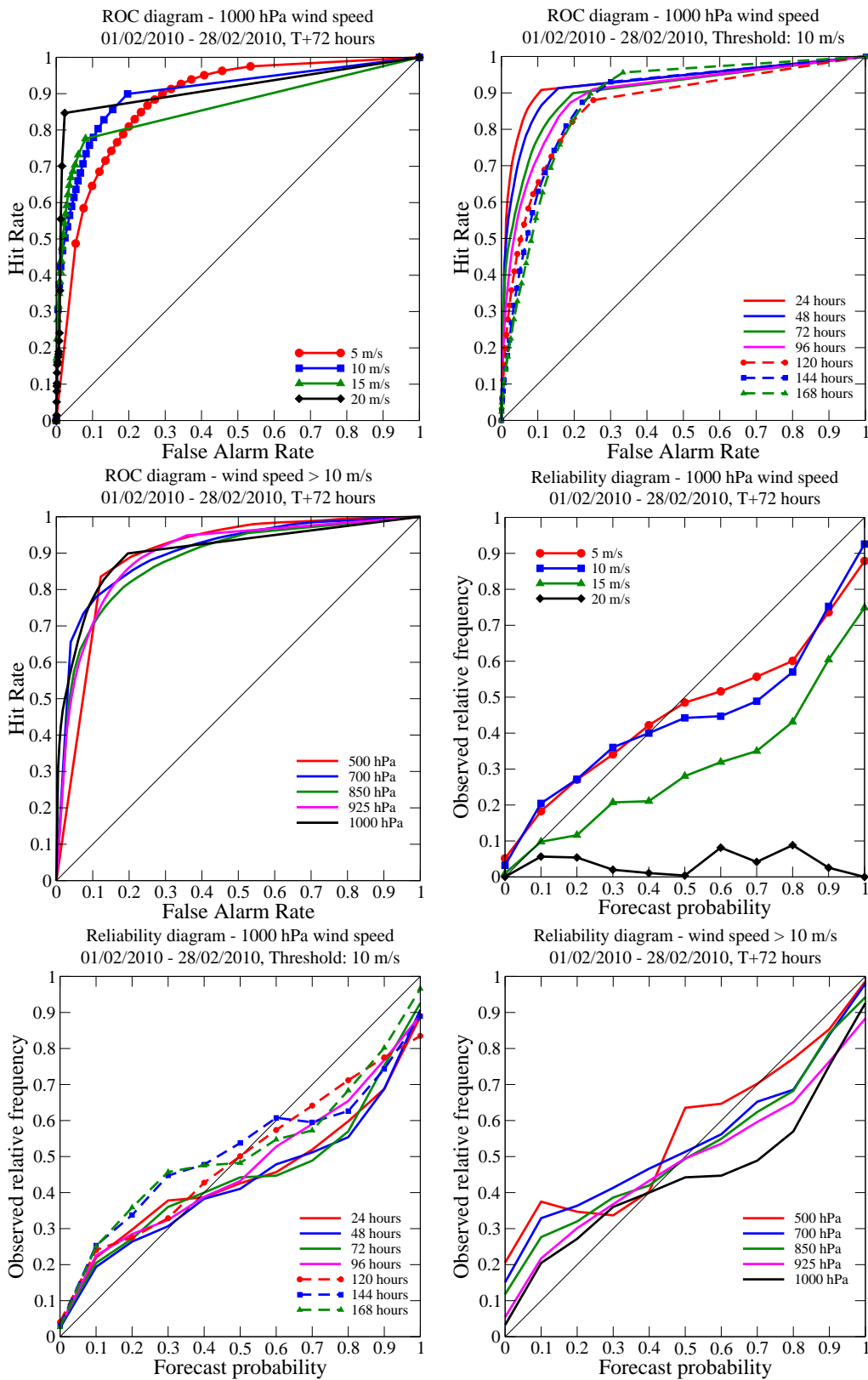
Figure 5. ROC (a, upper left, b, upper right, c, middle left) and reliability (d, middle right, e, lower left, f, lower right) diagrams for wind speed. (a and d) Diagrams for thresholds 5 m s^{-1}, 10 m s^{-1}, 15 m s^{-1}, and 20 m s^{-1} at 1000 hPa and T+72 hours. (b and e) Diagrams for time range T+24 hours, T+48 hours, T+72 hours, T+96 hours, T+120 hours, T+144 hours and T+168 hours at 1000 hPa and 10 m s^{-1} threshold. (c and f) Diagrams for 500 hPa, 700 hPa, 850 hPa, 925 hPa and 1000 hPa at T+72 hours and 10 m s^{-1} threshold. Verification was on a 0.15° × 0.15° grid.
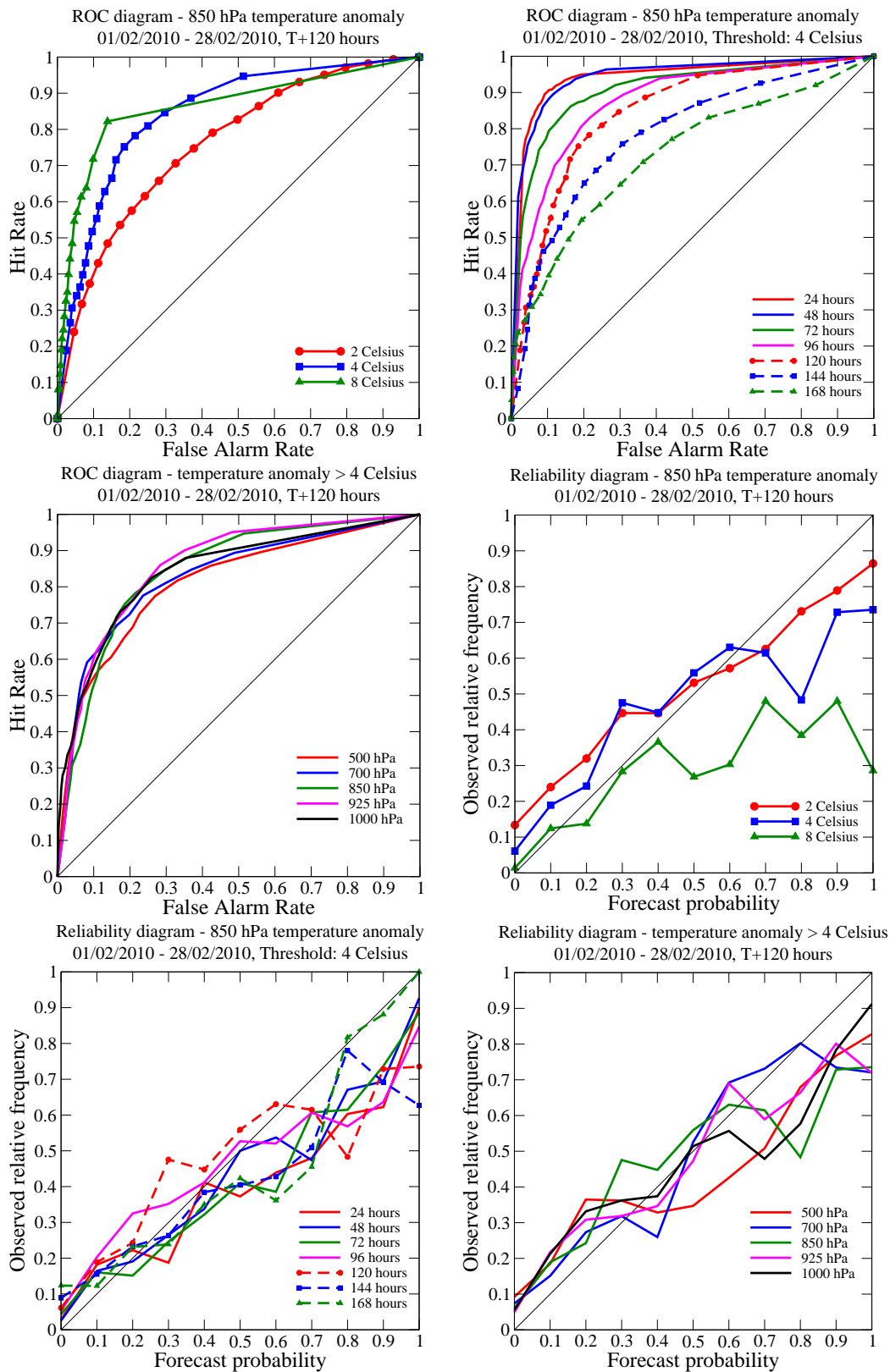
Figure 6. ROC (a, upper left, b, upper right, c, middle left) and reliability (d, middle right, e, lower left, f, lower right) diagrams for temperature anomaly. (a and d) Diagrams for thresholds 2°C, 4°C, and 8°C at 850 hPa and T+120 hours. (b and e) Diagrams for time range T+24 hours, T+48 hours, T+72 hours, T+96 hours, T+120 hours, T+144 hours and T+168 hours at 850 hPa and 4°C threshold. (c and f) Diagrams for 500 hPa, 700 hPa, 850 hPa, 925 hPa and 1000 hPa at T+120 hours and 4°C threshold. Verification was on a 0.15° × 0.15° grid.

Table 1. Summary of the results of the different systems (global vs limited area system).

Verification measure/Parameter	Geopotential	Temperature	Wind speed	Spec. humidity
RMSE - ensemble mean vs control member	Larger error of PROMES in the initial part of the forecast range; later very similar RMSE values (Figure 7a)		PROMES has equal or smaller error than GFS	Larger error of PROMES in the upper levels, larger error of GFS near the surface
	Control member has equal (initial part) or larger error than the ensemble mean for all parameters & levels and both models			
Spread skill relationship	Somewhat larger spread of GFS (Figure 7a)		Larger spread of GFS in the higher levels	Somewhat larger spread of PROMES
	Spread is smaller than the RMSE of the ensemble mean for all parameters & levels and both models			
Ranked histogram, percentage of outliers	Larger outlier values for PROMES, especially in the first 2-2.5 days	Larger outlier values for PROMES in the higher levels, very similar values in the middle and lower levels	Larger outlier values for PROMES in the higher levels, very similar values in the middle levels, better values of PROMES at low levels (Figure 7b)	Larger outlier values for PROMES in the higher levels, similar values in the middle levels, better values of PROMES near the surface
	The percentage of outliers is higher than the expected value for both models and all time steps			
ROC diagram	Better results of GFS for higher levels (500, 700 and 850 hPa, see Figure 7c). For 925 and 1000 hPa results of PROMES and GFS are more similar, with PROMES being slightly better in some forecast ranges			
Reliability diagram	No clear conclusion can be drawn, but PROMES is slightly better for some time ranges and thresholds (Figure 7d)			

ECMWF analysis as "truth") on the resolution of the global system ($1° \times 1°$) for a verification area covering the Iberian Peninsula (Figure 1b).

An overview of this comparison can be found in Table 1. In general, it can be said that by simply downscaling the global NCEP EPS forecasts using the higher resolution PROMES model it is very difficult to achieve significant - overall - improvements. For some parameters, vertical levels and verification measures the limited area ensemble forecasts performed better, in other cases the global forecasts were more skillful. Also in a couple of cases the two models had nearly the same scores (see Table 1 and Figure 7).

One should not forget however that it is a common phenomenon that high resolution (limited area) models might not perform better (on average, not for all individual cases) than the low resolution (global) models when usual verification measures are applied. Although the increased horizontal resolution of the LAM generally produces more realistic results, inevitable errors in timing and position can lead to larger RMSE values than for the smoother forecasts of the low resolution model. This is known as the double penalty problem.

However, as the representation of the orography is more realistic in LAMs - which is very important in the case of e.g. wind or precipitation forecasts - they can produce more reliable mesoscale structures. Therefore, it is important to run limited area models, because (at least) for near-surface variables they have added value compared with the control member and with the global model providing the LBCs.

4   Conclusions

In this paper the very first results of the PROMES EPS were presented. As a first step the simple dynamical downscaling was used in order to develop a LAMEPS with all the computational background needed (programs for running the ensemble, verification, visualization, etc). ICs and LBCs were provided by the NCEP global EPS and for downscaling the PROMES model was used. The LAMEPS was tested for a one month period. The applied verification methods included several measures (bias, RMSE, comparison of the spread and error, ranked histograms, percentage of outliers diagrams, ROC and reliability diagrams) and the global and the limited area EPS were compared as well.

From the results of the verification it was found that in both systems the spread of the ensemble members was smaller than the RMSE of the ensemble mean and the percentage of outliers was higher than the expected value. Thus, it can be concluded that both systems are lacking spread, at least for the verification area in question (Iberian Peninsula).
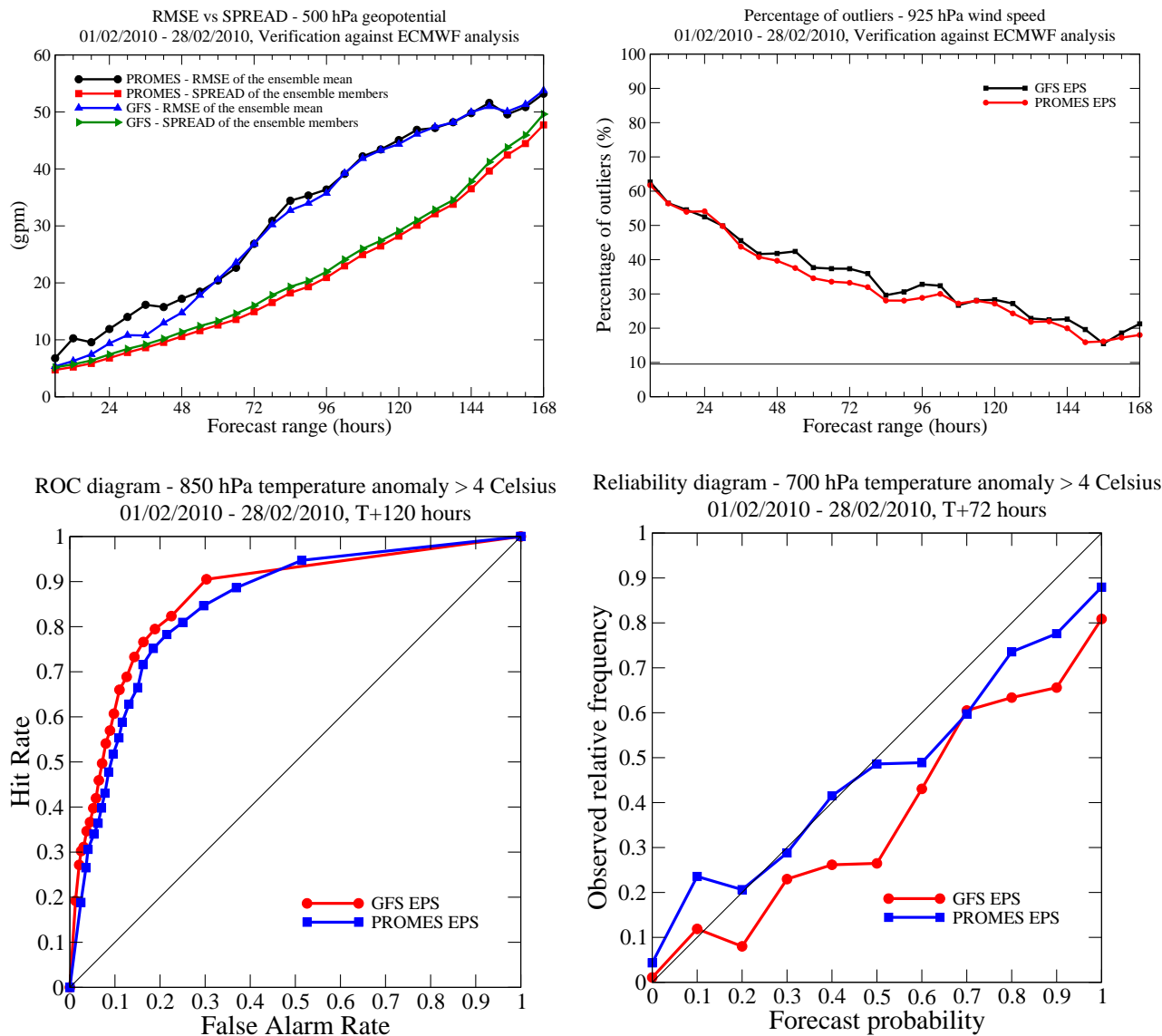
Figure 7. Comparison of the verification results of the global and the limited area system. (a, upper left) RMSE of the ensemble mean and spread of the ensemble members for 500 hPa geopotential. (b, upper right) Percentage of outliers diagram for 925 hPa wind speed. (c, lower left) ROC diagram for 850 hPa temperature anomaly > 4°C at T+120 hours. (d, lower right) Reliability diagram for 700 hPa temperature anomaly > 4°C at T+72 hours. Verification was performed on a $1.0° \times 1.0°$ grid.

This indicates that additional perturbations, introduced in the LAM, are desirable to improve the quality of the PROMES EPS. Possible methods for this are e.g. ET/ETKF (Ensemble Transform/Ensemble Transform Kalman Filter), LAM singular vectors (Stappers and Barkmeijer, 2011), breeding (Wang et al., 2011). However, most of these methods are computationally too expensive and/or require the operational run of a data assimilation system, thus they are not feasible for iMetCam in the short term.

At the same time, it should be considered that not only the IC error, but other uncertainties (e.g. model physics) should be addressed as well in order to create an efficient LAMEPS system. Possible directions for this future research

would be: multi-analysis and multi-boundary method (as an example see García-Moya et al., 2011); multi-parameter and/or multi-parametrization method (Wang et al., 2011), or the application of stochastic physics (Buizza et al., 1999).

It is planned for the near future to test the behaviour of the system for different seasons and extreme events. At present, the system (based on dynamical downscaling only) is running daily in order to build a larger dataset and gather more information about the behaviour of the LAMEPS in different weather conditions. Given that observation data became publicly accessible for the area of Spain, verification will be extended to surface parameters as well. This is especially important since the positive impact of dynamical

downscaling is usually more evident for surface parameters (e.g. 2 m temperature, 10 m wind) than for upper level ones.

References

Bourke, W., Buizza, R., and Naughton, M., 2004: *Performance of the ECMWF and the BoM Ensemble Prediction Systems in the Southern Hemisphere*, Mon Weather Rev, **132**, 2338–2357, doi: 10.1175/1520–0493(2004)132<2338:POTEAT>2.0.CO;2.

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

Branković, Č., Matjačić, B., Ivatek-Šahdan, S., and Buizza, R., 2008: *Downscaling of ECMWF Ensemble Forecasts for Cases of Severe Weather: Ensemble Statistics and Cluster Analysis*, Mon Weather Rev, **136**, 3323–3342, doi: 10.1175/2008MWR2322.1.

Buizza, R., Tribbia, J., Molteni, F., and Palmer, T., 1993: *Computation of optimal unstable structures for a numerical weather prediction model*, Tellus Ser A-Dyn Meteorol Oceanol, **45**, 388–407, doi: 10.1034/j.1600–0870.1993.t01–4–00 005.x.

Buizza, R., Miller, M. J., and Palmer, T. N., 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, Technical Memoranda 279, ECMWF.

Buizza, R., Bidlot, J. R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., and Vitart, F., 2007: *The new ECMWF VAREPS*, Q J R Meteorol Soc, **133**, 681–695.

Castro, M., Fernandez, C., and Gaertner, M. A., 1993: Description of a meso-scale atmospheric numerical model, Diaz JI and Lions JL (eds) Mathematics, climate and environment, Masson, 230-253.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gagnon, N., Mitchell, H. L., and Michelin, L., 2010: *Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System*, Mon Weather Rev, **138**, 1877–1901, doi: 10.1175/2009MWR3187.1.

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

García-Moya, J. A., Callado, A., Escribà, P., Santos, C., Santos-Muñoz, D., and Simarro, J., 2011: *Predictability of short-range forecasting: a multimodel approach*, Tellus Ser A-Dyn Meteorol Oceanol, **6**, 550–563, doi: 10.1111/j.1600–0870.2010.00 506.x.

Gebhardt, C., Theis, S., Krahe, P., and Renner, V., 2008: *Experimental ensemble forecast of precipitation based on a convection-resolving model*, Atmos Sci Lett, **9**, 67–72, doi: 10.1002/asl.177.

Hagel, E., 2010: *The quasi-operational LAMEPS system of the Hungarian Meteorological service*, Idojaras, **114**, 121–133.

Iversen, T., Bremnes, J. B., Santos Burguete, C., Deckmyn, A., Feddersen, H., Frogner, I. L., and Sattler, K., 2009: *Configuring GLAMEPS for operational production*, HIRLAM Newsletter, **55**, 37–50.

Jolliffe, I. T. and Stephenson, D. B., 2003: Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley.

Marsigli, C., Montani, A., and Paccagnella, T., 2008: *A spatial verification method applied to the evaluation of high-resolution ensemble forecasts*, Meteorol Appl, **15**, 125–143, doi: 10.1002/met.65.

Molteni, F., Buizza, R., Marsigi, C., Montani, A., Nerozzi, F., and Paccagnella, T., 2001: *A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments*, Q J R Meteorol Soc, **127**, 2069–2094, doi:10.1002/qj.49712757 612.

Nicolau, J., 2002: Short-range ensemble forecasting, In Proceedings WMO/CBS Technical Conferences On Data Processing and Forecasting Systems, Cairns, Australia, 2-3 December, 4p., available at: http://www.wmo.ch/pages/prog/www/DPS/TC-DPFS-2002/Papers-Posters/Topic1-Nicolau.pdf.

Stappers, R. and Barkmeijer, J., 2011: *Properties of singular vectors using convective available potential energy as final time norm*, Tellus Ser A-Dyn Meteorol Oceanol, **63**, 373–384, doi: 10.1111/j.1600–0870.2010.00 501.x.

Toth, Z. and Kalnay, E., 1997: *Ensemble Forecasting at NCEP and the Breeding Method*, Mon Weather Rev, **125**, 3297–3319, doi: 10.1175/1520–0493(1997)125<3297:EFANAT>2.0.CO;2.

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

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S., and Bazile, E., 2011: *The Central European limited-area ensemble forecasting system: ALADIN-LAEF*, Q J R Meteorol Soc, **137**, 483–502, doi: 10.1002/qj.751.

Wei, M., Toth, Z., Wobus, R., and Zhu, Y., 2008: *Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system*, Tellus Ser A-Dyn Meteorol Oceanol, **60**, 62–79, doi: 10.1111/j.1600–0870.2007.00 273.x.