



On the skill of various ensemble spread estimators for probabilistic short range wind forecasting

A. Kann

Central Institute for Meteorology and Geodynamics, Vienna, Austria

Correspondence to: A. Kann (alexander.kann@zamg.ac.at)

Received: 13 January 2012 – Revised: 14 March 2012 – Accepted: 3 May 2012 – Published: 15 May 2012

Abstract. A variety of applications ranging from civil protection associated with severe weather to economical interests are heavily dependent on meteorological information. For example, a precise planning of the energy supply with a high share of renewables requires detailed meteorological information on high temporal and spatial resolution. With respect to wind power, detailed analyses and forecasts of wind speed are of crucial interest for the energy management. Although the applicability and the current skill of state-of-the-art probabilistic short range forecasts has increased during the last years, ensemble systems still show systematic deficiencies which limit its practical use. This paper presents methods to improve the ensemble skill of 10-m wind speed forecasts by combining deterministic information from a nowcasting system on very high horizontal resolution with uncertainty estimates from a limited area ensemble system. It is shown for a one month validation period that a statistical post-processing procedure (a modified non-homogeneous Gaussian regression) adds further skill to the probabilistic forecasts, especially beyond the nowcasting range after +6 h.

1 Introduction

The analysis and nowcasting system INCA (Integrated Nowcasting Through Comprehensive Analysis), developed at the Austrian national weather service (Central Institute for Meteorology and Geodynamics, ZAMG), provides, among other quantities, short range deterministic analyses and forecasts of the horizontal 10-m wind components on very high resolution in time (60 min) and space (1×1 km) with special emphasis on the nowcasting range (usually 0–6 h ahead). In the analysis mode, it combines numerical weather prediction (NWP) forecast fields from the operational limited area model ALADIN-AUSTRIA (Aire Limitée Adaptation Dynamique Développement InterNational) with high-resolution topographic data and integrates about 250 stations within the operational domain. However, although the system shows high skill in the nowcasting range, it is affected by uncertainties mainly due to errors in the initial condition, model formulations and physical parameterizations. The paper compares three methods to quantify these uncertainties and to provide sharp and reliable probabilistic short range 10-m wind speed forecasts. The first method

combines the deterministic INCA forecast with the ensemble variance from ALADIN-LAEF (Aire Limitée Adaptation Dynamique Développement InterNational Limited Area Ensemble Forecasting), the operational limited area ensemble prediction system (LAM-EPS) at ZAMG. The second one derives uncertainty estimation from the skill of retrospective INCA forecasts. The third approach uses the uncertainty information of ALADIN-LAEF, but additionally applies a statistical calibration procedure.

2 The INCA system

The analysis and nowcasting system INCA, which is in operation at ZAMG, provides improved numerical forecasts especially in the nowcasting range on a very high horizontal resolution (1×1 km). A problem for NWP data at all scales is the comparatively low skill within the nowcasting range (0–6 h). To overcome this problem, INCA has been developed to provide temporarily and spatially highly resolved analyses and nowcasts taking into account regional and small scale topographic effects. The basic idea of INCA is to complement and improve NWP direct model output using real-time

observations, remote sensing data and high-resolution topographic data (Haiden et al., 2011). The INCA system provides near-real-time analyses and forecasts for the parameters temperature, humidity, wind, precipitation amount, precipitation type, cloudiness, and global radiation.

For wind, z -coordinates with $\Delta z = 125$ m are used in the vertical, where z denotes the height above mean sea level. The three-dimensional wind analyses are based on NWP forecast fields as a first guess and observations to correct them. Operationally, the deterministic limited area model ALADIN-AUSTRIA is used, which runs on 9.6 km horizontal resolution with 60 vertical levels (Wang et al., 2006). ZAMG operates about 250 automated weather stations (Teilautomatisches Wetterstationsnetz, TAWES) in Austria. The wind analysis is constructed by a first guess of the NWP model (ALADIN-AUSTRIA), a 2-D (10-m wind vector) and 3-D (lowest model level wind vector) component. A factor, which transforms a model level wind into a 10-m wind, determines the differences between model-level wind and a 10-m wind observation. After multiplying the observed wind by this factor, differences of the u and v components between the model and the observations are interpolated using a modified inverse-distance weighting (IDW). An iterative relaxation algorithm is applied in order to provide mass consistent fields (Wang et al., 2005; Sherman, 1978). Additionally the generally higher wind speeds over lakes compared the surrounding land due to reduced roughness length is taken into account using boundary-layer similarity theory, as described in Haiden et al. (2010). In Fig. 1 an example of the resulting INCA 10-m wind analysis is shown in comparison to the first guess NWP (ALADIN) 10-m wind field.

3 The limited area ensemble system ALADIN-LAEF

The ALADIN-LAEF system, which runs in operational mode since 2007 at ZAMG, provides ensemble forecasts up to +60 h, the horizontal resolution is 18 km with 37 vertical levels. It uses the spectral limited area model ALADIN (Wang et al., 2006) in order to produce 17 ensemble members (16 perturbed members, 1 control). The perturbations for the initial conditions of the 16 ensemble members are composed by a large scale part using 16 ECMWF EPS members which are created with the singular vector method (Leutbecher and Palmer, 2008) and a small scale component derived with the breeding technique (Wang et al., 2011). Both perturbations are finally combined by applying a spectral blending procedure (Brozkova et al., 2001). Model errors are taken into account by a multi-physics approach. A detailed description of the ALADIN-LAEF system can be found in Wang et al. (2011).

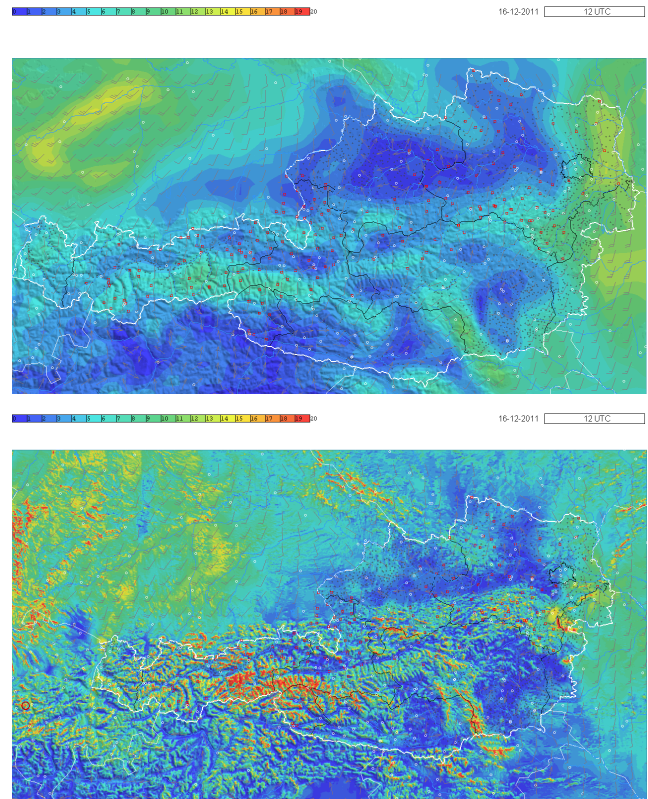


Figure 1. Example of a 10-m wind analysis provided by the INCA wind module issued for 16 December 2011, 12:00 UTC (bottom). The top figure shows the corresponding first guess field taken from the ALADIN-AUSTRIA model run initialized at 16 December 2011, 00:00 UTC, +12 h forecast.

4 Probabilistic (very) short range forecasting approaches

The basic strategy has been developed and implemented for the purpose of providing sharp and reliable probabilistic short range 2-m temperature forecasts and has been described more detailed in Kann et al. (2011). The main idea is to treat the deterministic INCA 10-m wind speed forecast as ensemble mean and to apply an ensemble variance by different techniques. A statistical method (EPS_{stat}), a dynamical one (EPS_{dyn}) and a coupled dynamic-statistical approach ($\text{EPS}_{\text{dynstat}}$) are developed to generate an 18-member ensemble providing probabilistic short range wind speed forecasts up to +36 h.

The statistical method EPS_{stat} uses the deterministic INCA run as ensemble mean and its relative root mean squared error (RMSE) of the past 50 training days as ensemble variance. The ensemble size of 18 members is obtained by extracting the quantile values from a truncated (i.e. cut off at zero) Gaussian cumulative distribution function (CDF) centralized about the mean in such a way that the ensemble spread is determined by a fraction of the relative (i.e. with

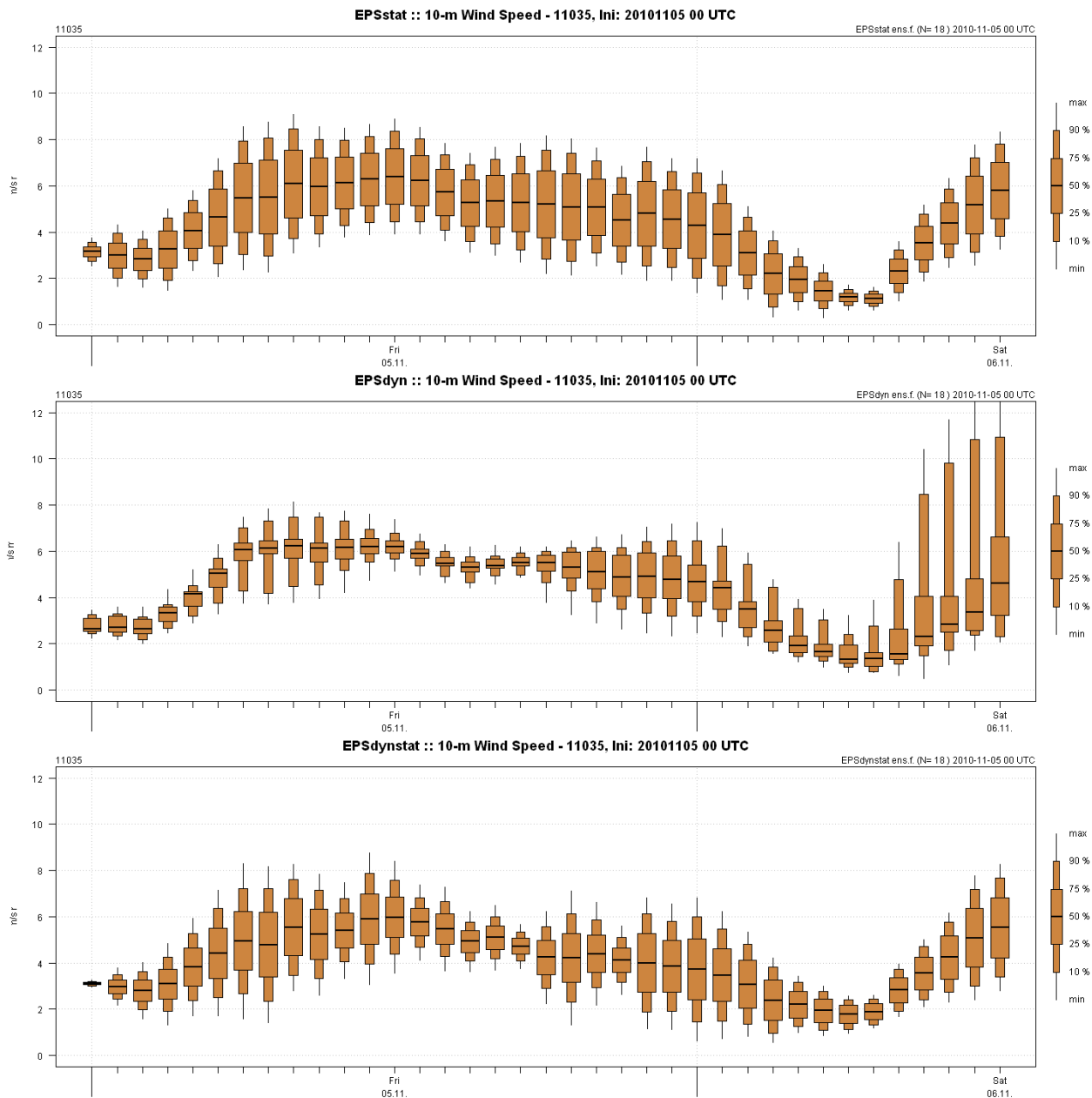


Figure 2. EPSgrams for 10-m wind speed, calculated with the three different methods, forecast initialized at 5 November 2010, 00:00 UTC, for station Vienna Hohe-Warte.

respect to the observed average) RMSE of the training data (Kann et al., 2009). Several experiments with varying rescaling factors revealed a value of 3/4 to perform best in terms of the Continuous Ranked Probability Score (CRPS) (not shown).

In the dynamical approach EPS_{dyn} , the ensemble variance from ALADIN-LAEF provides the uncertainty information and is superimposed on the deterministic INCA run (Kann et al., 2011). Practically, the differences between the individual ensemble members and the ensemble mean of ALADIN-LAEF are added to the INCA run. Thus, an 18-member en-

semble with a high resolution deterministic component and an uncertainty part derived from a limited area ensemble system is created.

The dynamic-statistical method $EPS_{dynstat}$ extends the aforementioned dynamical approach with an additional statistical post-processing application. The EPS_{dyn} is further calibrated by applying a cut-off variant of the non-homogeneous Gaussian regression (NGR) technique (Thorarinsdottir and Gneiting, 2010; Gneiting et al., 2005; Hagedorn et al., 2008; Kann et al., 2009). This technique statistically calibrates the mean and the ensemble variance

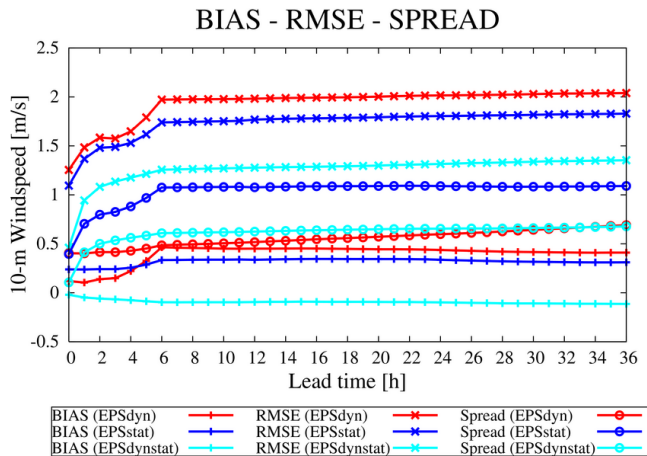


Figure 3. Bias, root mean square error (RMSE) of the ensemble mean, and ensemble spread of the 3 experimental INCA ensembles (dynamical: red, statistical: dark blue, dynamic-statistical: light blue).

by minimizing the Continuous Ranked Probability Score (CRPS) within a certain training period. The variance of the forecast distribution is a linear function of the ensemble variance. Again, the final 18-member ensemble is created by calculating the quantile values from the Gaussian CDF, centred about the mean with respect to the spread re-scaling factor.

The three sets of ensembles described above are tested for one month from 1 to 30 November 2010. The hourly 10-m wind speed forecasts of INCA and of ALADIN-LAEF are bilinearly interpolated separately for each station and lead time to the locations of the surface weather stations in Austria.

Note that the variance used for the experiments EPS_{dyn} and $EPS_{dynstat}$ is calculated from the latest available ALADIN-LAEF run, valid for the same target time as the INCA run (e.g.: The variance of the INCA ensemble issued at 1 November 2010, 11:00 UTC + 6h is provided by the ALADIN-LAEF issued at 1 November 2010, 00:00 UTC + 17h).

Figure 2 shows an example of a probabilistic 10-m wind speed forecast initialized at 5 November 2010, 00:00 UTC at the location of Vienna Hohe-Warte (11035) obtained by the three different methods.

5 Validation

To assess the performance of the three probabilistic methods, a comparative evaluation was carried out for the one month period (1–30 November 2010). The 10-m wind speed forecasts are verified against measurements of approx. 250 automatic surface weather stations in Austria. Hourly initializations with lead times up to +36h are considered. A set of standard ensemble and probabilistic forecast verification methods is consulted to evaluate the performances: ensemble spread, ensemble root-mean square error (RMSE), the continuous ranked probability score (CRPS), percent-

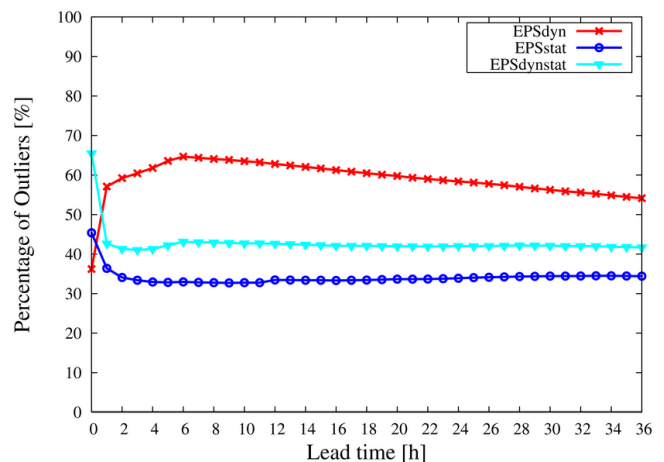


Figure 4. Percentage of outliers as a function of lead time, for the dynamical (red), statistical (dark blue) and dynamic-statistical (light blue) approach.

age of outliers, area under relative operating characteristic (ROC) curve, and the reliability diagram. A detailed description of these verification measures is found in Wilks (2006). In Fig. 3, the error of the ensemble mean in terms of bias and root mean squared error (RMSE) and the ensemble standard deviation is illustrated for the three experiments. Both the statistical and dynamical methods convey a slight positive bias, whereas the combined statistic-dynamical is almost biasfree. The RMSE is substantially reduced to about $1\text{--}1.4\text{ m s}^{-1}$ by for the combined approach compared to the pure statistical and dynamical methods, respectively, where the RMSE is about $1.5\text{--}2\text{ m s}^{-1}$. Another measure of statistical reliability is the percentage of outliers. It expresses the number of cases where the verifying analysis at any location falls outside the ensemble. A reliable system should have a score close to an expected value of $2/(n+1)$ which corresponds to about 10% for an 18-member EPS. The pure dynamical ensemble is affected by the highest percentage of outliers (approx. 60%), but the number of outliers decreases for the dynamic-statistical approach (about 42%) and the statistical method (35%) (Fig. 4). The ROC curve and the area below it (Zhu et al., 2002), respectively, offers another methodology for a quantitative evaluation of an ensemble system. With respect to the ROC area, the dynamic-statistical method outperforms the statistical and especially the dynamical approach (Fig. 5). A further assessment of forecast quality with regard to reliability and resolution is given by the reliability diagram which indicates quantitative evidence about the ability of the probability forecasts to reflect the observed relative frequency (Stanski et al., 1989). Figure 6 reveals the high reliability of the dynamic-statistical combination compared to both the dynamical and statistical approach for different lead times (6h (Fig. 6a) and 24h (Fig. 6b), respectively). The latter two methods are characterized by the fact that low as well as high observed frequencies are overforecast (apart from the

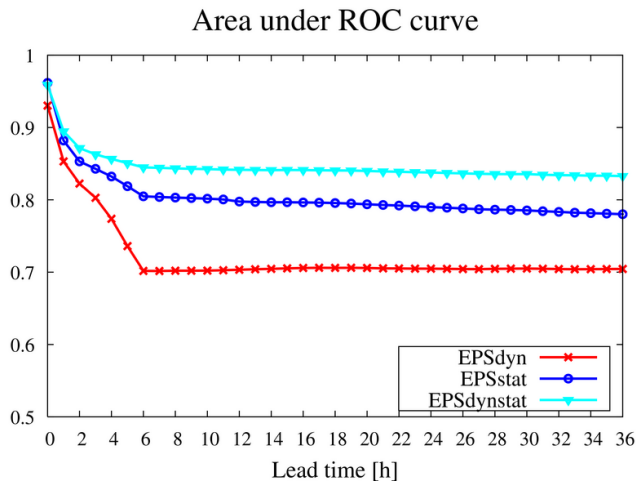


Figure 5. Area under ROC curves for 10-m wind speed $>2 \text{ m s}^{-1}$ as a function of lead time, from the dynamical (red), statistical (dark blue) and dynamic-statistical (light blue) method.

very low frequencies which are underforecast). The Continuous Ranked Probability Score (CRPS), which can be decomposed into a reliability, resolution and uncertainty term, is the integral form of the discrete ranked probability score over all (possible) thresholds (Hersbach, 2000). It compares a full distribution with the observation, where both are represented as cumulative distribution functions (CDFs). In case of a specific observation value, the corresponding CDF is a single step-function with the step from 0 to 1 at the observed value of the considered parameter (Heaviside function). Figure 7 confirms previous findings that the combination of dynamical and statistical methods outperforms both the pure statistical and dynamical approaches. This is especially true for lead times beyond the nowcasting range, where the CRPS for $\text{EPS}_{\text{dynstat}}$ is substantially reduced to about 0.7 m s^{-1} compared to the CRPS for EPS_{dyn} (1.2 m s^{-1}) and EPS_{stat} (0.9 m s^{-1}). However, also within the nowcasting range up to +6h, the dynamic-statistical method shows highest skill of all three methods. A further elaboration of the statistical approach has been performed by applying the non-homogeneous Gaussian regression to EPS_{stat} . Figure 7 confirms that the calibration is able to add further skill to the statistical method (experiment $\text{EPS}_{\text{statcal}}$, black line), although the error reduction is not pronounced. However, the calibration of the dynamical approach ($\text{EPS}_{\text{dynstat}}$) allows for a higher skill in terms of the CRPS. In other words, the spread obtained from ALADIN-LAEF provides additional information compared to the RMSE spread estimator.

6 Conclusions

Three different methods for constructing an ensemble of 10-m wind speed forecasts are developed and validated for a one month period: a statistical, a dynamical method and a

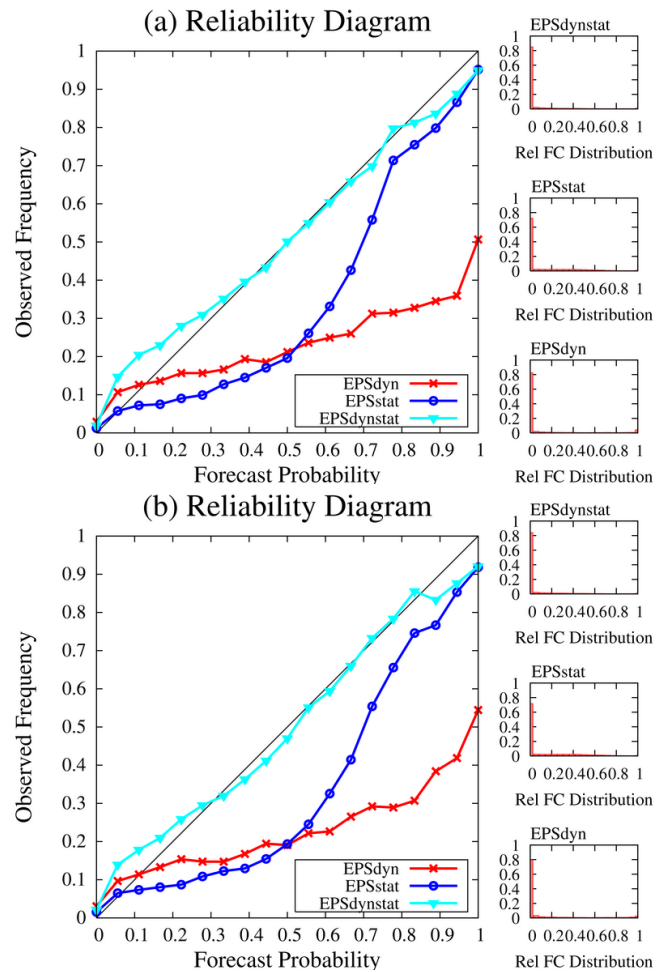


Figure 6. Reliability diagram for +6h (a) and +24h (b) from the dynamical (red), statistical (dark blue) and dynamic-statistical (light blue) method. Threshold: 5 m s^{-1} .

combination of both. The common idea, which has originally been designed for 2-m temperature, is to unite the strengths of a very high resolution nowcasting system with a limited area ensemble system and to derive reliable and sharp probabilistic, site-specific short range 10-m wind speed forecasts. The evaluation conveys that a combined dynamic-statistical approach, where the high resolution deterministic ensemble mean from INCA and the ensemble standard deviation from ALADIN-LAEF are further statistically downscaled, is of superior skill than the pure dynamical and statistical methods. Especially beyond the nowcasting range, the underdispersive behavior is reduced significantly by the statistical adaptation. It has also been demonstrated by a further experiment that the ALADIN-LAEF spread provides additional information compared to the RMSE estimator of spread.

For an operational implementation, some issues have to be still considered: are the results and the conclusion generalizable to other seasons? At least periods in different wind regimes should be studied in order to confirm the results of

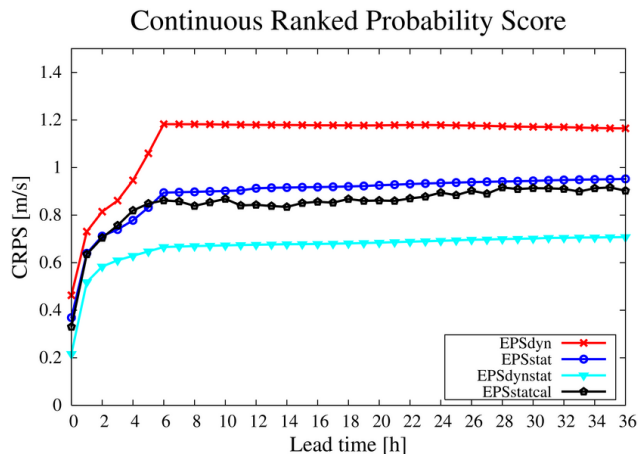


Figure 7. CRPS as a function of lead time, from the dynamical (red), statistical (dark blue) and dynamic-statistical (light blue) method. Furthermore, a calibrated-statistical approach indicating the skill of a further calibration of the statistical method is added (black).

the present paper. Is the re-scaling factor used in the experiment $EPS_{dynstat}$ already applicable to other seasons or still improvable?

Both the INCA and the ALADIN-LAEF systems are under permanent development at ZAMG. In INCA, work on the implementation of dynamical and high resolution topographic effects (e.g. by dynamical downscaling techniques) is ongoing and will potentially lead to more realistic wind analyses and short range forecasts. Research activities in ALADIN-LAEF are focusing on the revision of the multi-physics approach and on a stochastic surface perturbation scheme. These activities will have a direct positive impact on such applications which combine these systems algorithmically.

Site-specific, optimized probabilistic wind speed forecasts are of increasing interest in a variety of applications. For example, in the renewable energy sector, precise deterministic and probabilistic wind forecasts for wind farms are of crucial relevance for a detailed planning of the energy production and power management. Thus, improved wind forecasts have a direct impact on a cost-efficient power industry.

Acknowledgements. The author is thankful to Christoph Wittmann, Yong Wang and Florian Weidle for fruitful discussions about ALADIN-LAEF and to Thomas Haiden, Georg Pistotnik, Christine Gruber and Benedikt Bica for comments on scientific and operational INCA aspects. The author is also grateful to two anonymous reviewers for their suggestions and constructive comments that led to improvements of the manuscript.

Edited by: R. Hagedorn

Reviewed by: two anonymous referees

References

- Brozkova, R., Klaric, D., Ivatek-Sahdan, S., Geleyn, J. F., Casse, V., Siroka, M., Radnoti, G., Janousek, M., Stadlbacher, K., and Seidl, H.: DFI Blending, an alternative tool for preparation of the initial conditions for LAM, PWRP Report Series No. 31 (CAS/JSCWGNE Report), WMO-TD, No. 1064, 1.7., 2001.
- Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, *Mon. Weather Rev.*, 133, 1098–1118, 2005.
- Haiden, T., Kann, A., Pistotnik, G., Stadlbacher, K., and Wittmann, C.: Integrated Nowcasting through Comprehensive Analysis (INCA) – System description, ZAMG report, 60 pp., available at: http://www.zamg.ac.at/fix/INCA_system.pdf, 2010.
- Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., and Gruber, C.: The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the Eastern Alpine region, *Weather Forecast.*, 26, 166–183, doi:10.1175/2010WAF2222451.1, 2011
- Hagedorn, R., Hamill, T. M., and Whitaker, J. S.: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts, Part I: temperature, *Mon. Weather Rev.*, 136, 2608–2619, 2008.
- Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, 15, 559–570, 2000.
- Kann, A., Wittmann, C., Wang, Y., and Ma, X.: Calibrating 2 m-temperature of limited area ensemble forecasts using high-resolution analysis, *Mon. Weather Rev.*, 137, 3373–3387, 2009.
- Kann, A., Haiden, T., and Wittmann, C.: Combining 2-m temperature nowcasting and short range ensemble forecasting, *Nonlin. Processes Geophys.*, 18, 903–910, doi:10.5194/npg-18-903-2011, 2011.
- Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, *J. Comput. Phys.*, 227, 3515–3539, 2008.
- Sherman, C. A.: A mass-consistent model for wind field over complex terrain, *J. Appl. Meteorol.*, 17, 312–319, 1978.
- Stanski, H. R., Wilson, L. J., and Burrows, W. R.: Survey of common verification methods in meteorology, WMO/WWW Tech. Rep. 8, 114 pp., 1989.
- Thorarindottir, T. L. and Gneiting, T.: Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroskedastic censored regression, *J. R. Stat. Soc. A. Sta.*, 173, 371–388, 2010.
- Wang, Y., Williamson, C., Garvey, D., Chang, S., and Cogan, J.: Application of a multigrid method to a mass-consistent diagnostic wind model, *J. Appl. Meteorol.*, 44, 1078–1089, 2005.
- Wang, Y., Haiden, T., and Kann, A.: The operational Limited Area Modelling system at ZAMG: ALADIN-AUSTRIA, *Österr. Beiträge Meteorol. Geophys.*, 37, 1016–6254, 2006.
- Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Sahdan, S., Tian, W., Ma, X., Tascu, S., and Bazile, E.: The Central European limited-area ensemble forecasting system: ALADIN-LAEF, *Q. J. Roy. Meteor. Soc.*, 137, 483–502, doi:10.1002/qj.751, 2011.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, 2nd Edn., Academic Press, 627 pp., 2006.
- Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Mylne, K.: The economic value of ensemble based weather forecasts, *B. Am. Meteorol. Soc.*, 83, 73–83, 2002.