Upscaled and fuzzy probabilistic forecasts: verification results

Zied Ben Bouallègue

Deutscher Wetterdienst (DWD), Frankfurter Str. 135, 63067 Offenbach, Germany

1 Introduction

The combined use of high resolution models and ensemble forecasting techniques is expected to be an optimal framework to provide probabilistic quantitative precipitation forecasts [3]. High resolution models simulate convective processes explicitly and ensemble forecasting techniques take into account the sources of uncertainty.

However, during a verification procedure, high resolution model forecasts suffer of the well known double penalty problem [5]. Small displacements in space or time between forecasted and observed precipitation events penalize twice the forecast in a point to point comparison. This limitation in terms of predictability should be reflected by an ensemble forecast which aims to provide information about the uncertainty of the prediction. Nevertheless, ensemble forecasts of surface variables present very often an important drawback: underdispersiveness.

The double penalty problem affects probabilistic forecasts as long as the uncertainty in position is not well represented. Then, it is still meaningful to use spatial verification techniques in order to better characterize the potential of a probabilistic forecast. Taking into account the spatial environment of each grid point forecast is a simple way to address this problem [2] and to perform scale analysis [7]. Neighborhood approaches can also be used to derive or improve existing forecasts [15, 13].

Two spatial techniques and their derived products are investigated here. First, the neighborhood method, which smoothes the probabilistic forecast, produces a fuzzy probabilistic forecast. This method is a cheap solution to enlarge the sample of an ensemble forecast. Secondly, the upscaling technique, which modifies the scale of the forecast, produces an upscaled probabilistic forecast. The upscaling procedure allows to provide information at different scales of interest.

The purpose of this report is double: firstly to evaluate the performance of an ensemble forecasting system, secondly to provide a guideline for the generation of new probabilistic products. We focus here on daily precipitation.

2 Dataset and Methodology

Dataset

The COSMO-DE-EPS is an ensemble prediction system using the convection-permitting model COSMO-DE, a 2.8 km grid-spacing configuration of the COSMO model covering the area of Germany. An experimental version of the system is used where boundary conditions and physics perturbations are applied without initial condition variation [4]. Note that the introduction of the initial condition perturbations in the new version of the system is described in [11].

The ensemble forecasts start at 00UTC and comprise 15 members. 55 days are available with this configuration during the summer of 2009 from 14 June 2009 to 30 September 2009. The precipitation observations are radar data. The observations possibly affected by bright band effects are rejected from the verification process.

The neighborhood method: fuzzy probabilistic forecasts

The first spatial technique applied here is the so-called neighborhood method, developed by [15]. Originally, this method was designed to derive probability forecasts from deterministic ones. [13] describe how to apply this method to ensemble forecasts. The processed probability at a given grid point corresponds to the mean probability within a given environment. This environment is a circular neighborhood defined by a radius of influence (called hereafter size parameter of the process).

The upscaling process: upscaled probabilistic forecasts

The second spatial procedure performed here consists of dividing the domain into squared windows. In this procedure, the length of the window is the size parameter. The maximum value of each member within each window represents the precipitation field on the new spatial scale. The upscaled probability fields are calculated from those values and refer then to an event that occurs anywhere within the defined window.

For verification purposes, other types of upscaling can be investigated. Rather than the maximum, the 90 percent quantiles within the windows (or the 95%, 99%) can be chosen to represent the realization of a member at the new spatial scale. The use of the quantiles instead of the maximum allows to alleviate the sensitivity to possible outliers in observation data and small scale variability [14].

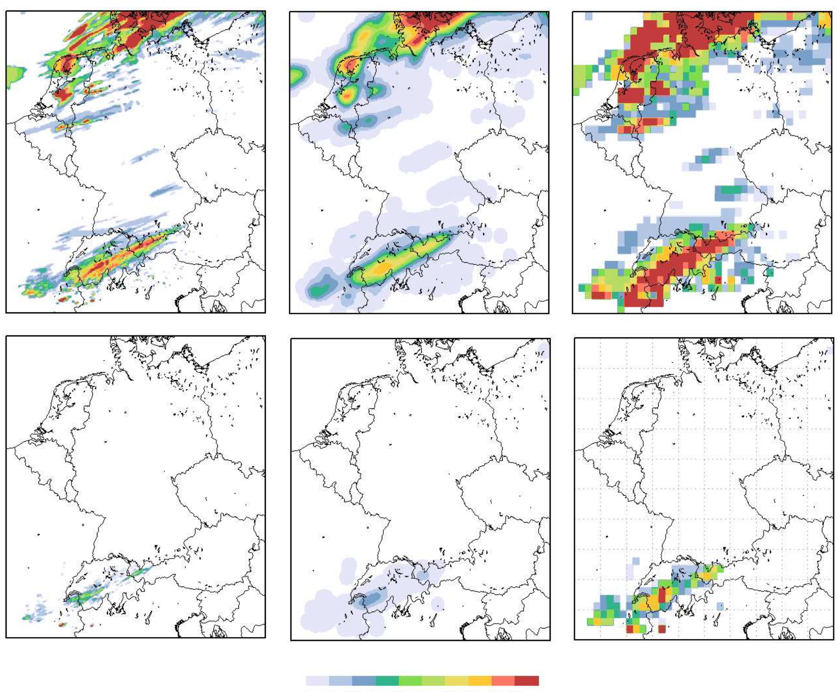

Figure 1: Example of 18 June 2009: Probability of precipitation exceeding 1mm/24h (top) and 10mm/24h (bottom). From left to right : original products, fuzzy products (the neighborhood method is applied with a size parameter of 10 grid points), upscaled products (within 10×10 grid points boxes).

Example of processed probability fields

An example of fuzzy and upscaled probabilistic forecasts is shown in Figure 1. The fuzzy probabilistic forecasts are a common application in image processing which is called low pass filtering or convolution kernel [12]. The smoothing of the original field can be done by 'eye' since all the information needed to construct the smoothed field is already contained in the original one. This is not true for the upscaled probability forecast where the spatial variability of each precipitation forecast (of each member) must be known beforehand.

3 Verification measures

Score measure

Among the numerous existing verification scores, the most common one for the evaluation of probabilistic forecasts is the Brier score (BS). It is defined as the average square difference between the forecast probability and the observation. The BS can be decomposed in three terms [10]: reliability (REL), resolution (RES) and uncertainty (UNC).

The reliability term measures the statistical consistency between the probabilistic forecast and the frequency of occurrence of the observed event given the forecast probability. The resolution term measures the capacity of the system to distinguish between event and non event. The uncertainty is a function of the observation only. Another attribute, the sharpness (SHP), which depends only on the forecast is also examined here. The sharpness is defined as the mean squared departure of the forecast probabilities from the climatological probability. It corresponds to the reliability term of a random forecast [16].

Skill score

We use Brier skill scores (BSS) for the evaluation of the impact of the processes described above. The BSS is defined as (Wilks 1995):

$$BSS = \frac{BS_{ref} - BS}{BS_{ref}}, \tag{1}$$

where BS_{ref} is the Brier score of a reference forecast, the object of the comparison. Similarly, reliability, resolution and sharpness gain are defined as:

$$G_{REL} = \frac{REL_{ref} - REL}{REL_{ref}}, \qquad G_{RES} = \frac{RES - RES_{ref}}{RES_{ref}}, \qquad G_{SH} = \frac{SH - SH_{ref}}{SH_{ref}},$$

since the reliability is counted negatively (the lower the better) and the resolution and sharpness positively (the higher the better).

The BS is not well adapted to compare forecast performance at different scales, as we do with upscaled probabilistic forecasts. In fact, its uncertainty component, which is a function of the observation only, differs from scale to scale. It is then worth to use a BSS as a measure of the performance at each scale. Mason [9] has shown that using a random forecast as reference for the calculation of the BSS leads to a measure of the 'usefulness of the information' adapted to forecasts with high sharpness. We define then the gain in skill score as:

$$G_{SS} = \frac{BSS^{ran} - BSS^{ran}_{ref}}{BSS^{ran}_{ref}},$$

where BSS^{ran} is the BSS of the upscaled forecast compared to a random forecast at the corresponding scale and BSS^{ran}_{ref} is the BSS of the original ensemble forecast compared to a random forecast at the original scale.

Amplitude distribution

We use in this study two more verification measures. The first one focuses on the amplitude distributions. The discrepancy from uniformity (D) measures the deviation from a uniform rank histogram. It is defined as [1]:

$$D = \sum_{i=1}^{K+1} \left| p_i - \frac{1}{K+1} \right| ,$$

where K is the number of members and p_i is the relative frequency of rank i. This measure summarizes the information contained in a rank histogram. In other words, it is an estimation of the fit between forecasted and observed amplitude distributions. A value of 0 indicates a flat histogram. This measure is applied hereafter to the probability integral transform (PIT) histogram, equivalent to the rank histogram in probability space [6].

Spatial distribution

The second tool allows spatial structure analysis: the empirical (semi-)variogram. It is complementary to the first one since an amplitude distribution conveys no information about the spatial structure of a field. The empirical variogram is a well known function of geostatistics which is also commonly used for meteorological applications (see Marzban and Sandgathe [8] and references therein). It can be seen as a tool to gauge the texture of a field. It is defined as:

$$\gamma(y) = E\left(|z(i) - z(j)|^2\right) ,$$

where $z(i)$ is the value of the field at a location i, y is the distance between the points i and j and E is the expected value operator. The variogram quantifies the spatial extent of correlation.

4 Results and Discussion

Fuzzy probabilistic forecasts

We investigate the impact of the neighborhood method in function of its size parameter. First, we compare the processed fields to the original probabilistic forecast in terms of accuracy, sharpness, reliability and resolution. The original probabilistic forecast is the reference for skill scores and gain calculation.

The general impact of the neighborhood method is shown in Figures 2(a) and 2(b). The method allows to improve the score and we note that an optimal size parameter exists in terms of BS. The optimum size parameter is similar for all the thresholds, around 40 grid points. On the other hand, the sharpness decreases linearly with the radius of influence. Figures 2(c) and 2(d) focus on the two main attributes of the forecast: reliability and resolution. Fuzziness slightly increases the resolution but has a large positive impact on the reliability. We can also note that the maximum gain in resolution is reached for smaller size parameters compared to the maximum gain in reliability.

To go further in the description of the neighborhood method impact, we compare the amplitude distribution of the processed fields and the observation distribution. The discrepancy from uniformity is shown in Figure 3(a). The fit between the amplitude distributions is improved and has an optimum for a size parameter around 40 grid points. The neighborhood method enlarges the spatial spread and has a positive impact on the pointwise amplitude distribution.

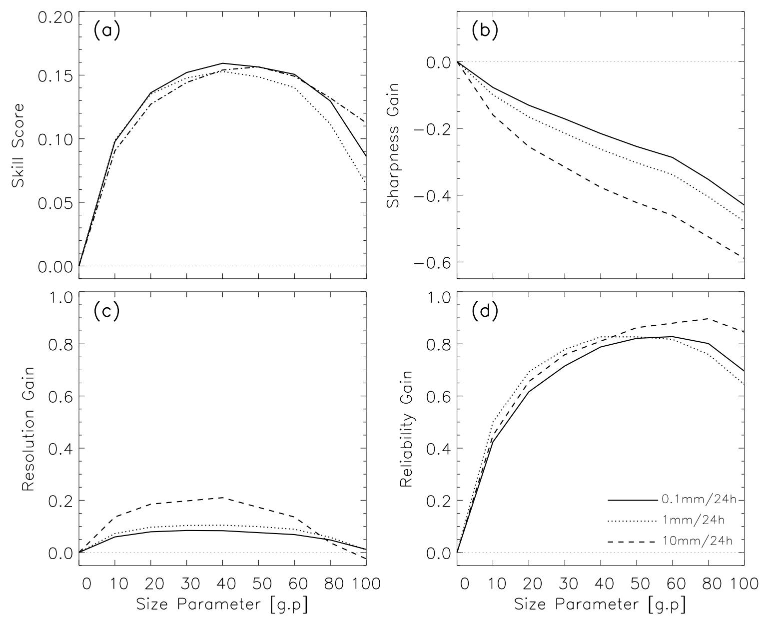

Figure 2: (a) Brier skill score, (b) sharpness gain, (c) resolution gain and (d) reliability gain in function of the neighborhood method size parameter (in grid points). The reference for the skill and gain calculation is the original probabilistic forecast.

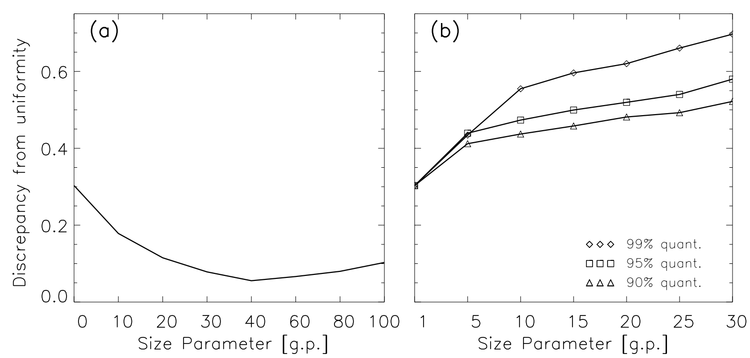

Figure 3: Discrepancy from uniformity (a) in function of the size parameter (in grid points) of the neighborhood method and (b) for the upscaling processes.

In order to highlight the reason of the improvement due to the application of the neighborhood method, we analyze the spatial distribution of the original error fields. The error is defined as the absolute difference between the observation and the original probabilistic forecast. The empirical variogram in Figure 4(a) represents the spatial correlation of this error. We note that above 40 grid points the spatial correlation of the error is no more significant. The neighborhood method introduces spatial correlation in the probability field (and then in the uncertainty representation) that corresponds to the spatial correlation of the original probabilistic forecast error.

Considered as products, we have finally to quantify the usefulness of the fuzzy probabilistic forecast compared to cheaper solutions. The reference for the computation of the BSS is no more the raw original probabilistic forecast but the deterministic forecast. To make a fair comparison, the smoothing is contemporarily applied to the reference and to the probabilistic forecast. Figure 5(a) shows that the BSS tends to zero in this case. The information within a

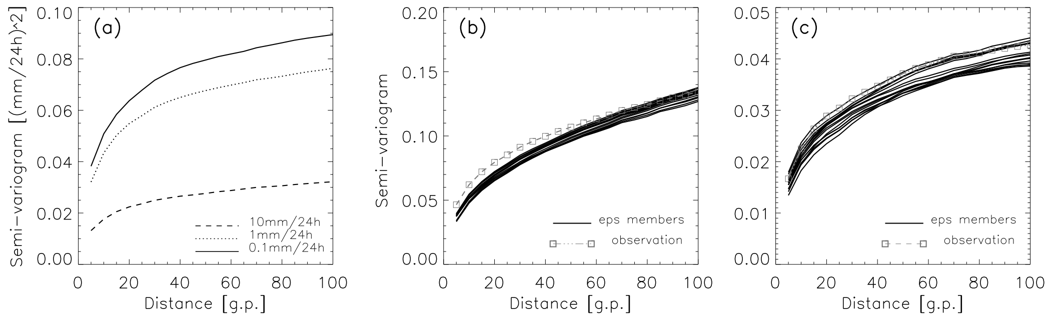

Figure 4: (a) Empirical semi variogram of the absolute difference between probabilistic forecast and binary observation in function of the distance (in grid points). Empirical semi variogram of binary precipitation fields defined with a threshold of (b) 0.1mm/24h and (c) 10mm/24h.

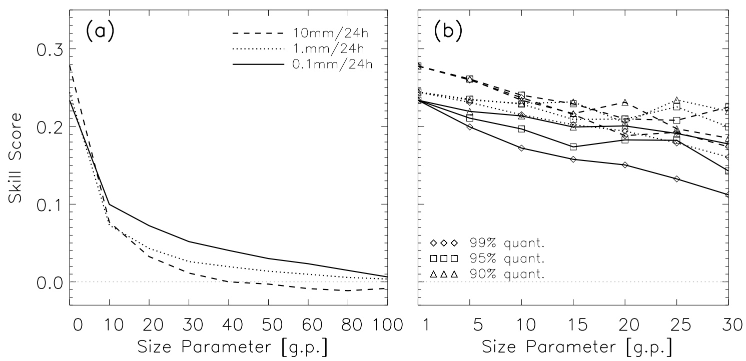

Figure 5: Brier skill scores in function of the size parameters (in grid points) of (a) the neighborhood method and (b) the upscaling process. The reference is the deterministic forecast. The neighborhood method or the upscaling process is applied contemporarily to the probabilistic forecast and to the reference.

fuzzy ensemble forecast is not useful if the filtering process is too strong: the same information can be provided by a cheaper fuzzy deterministic forecast.

Upscaled probabilistic forecasts

We analyze now the results of the probabilistic forecast at different scales. Three types of upscaling are compared using the 90%, 95% and 99% quantiles within a window as representative of the variable at the new scales. We first comment the results independently of the choice of quantile.

The impact of the upscaling processes is shown in Figure . The reference for the gain calculation is the probabilistic forecast at the model grid resolution. In Figure (a), we see an improvement of the forecast (excepted for the lowest threshold) and in Figure (b) an increase of the sharpness with the size parameter. The impact of the upscaling is especially remarkable for high thresholds. On Figure (c), the upscaling has a clear positive impact on the resolution term for medium and high thresholds. The benefit of the upscaling in terms of resolution increases with the threshold. In terms of reliability (Figure (d)), a loss of quality is registered for all the thresholds.

The spread reduction is intrinsic to the upscaling technique. The uncertainty concerning the exact location of an event is reduced as the size parameter increases. The width of the forecast amplitude distribution is then reduced. The discrepancy from uniformity measured in Figure 3(b) shows that the observations tend to fall more often outside of the forecasts distribution

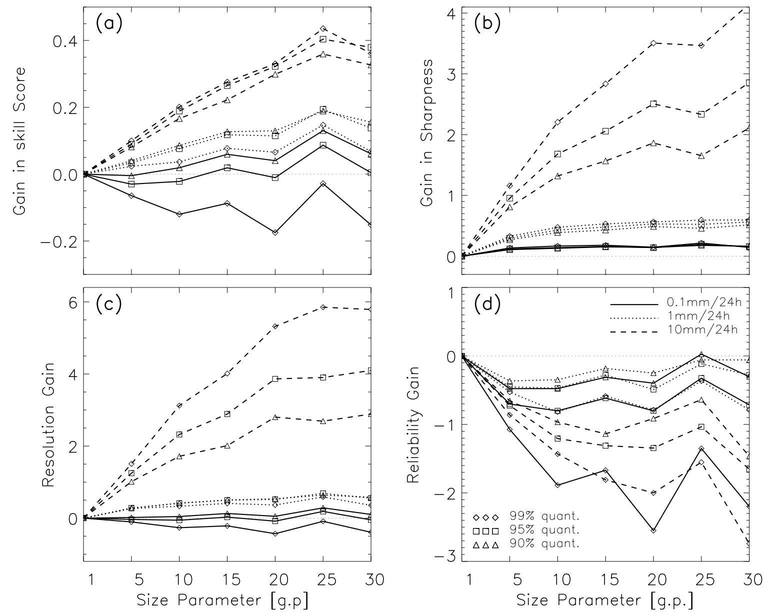

Figure 6: (a) Gain in BSS, (b) sharpness gain, (c) resolution gain and (d) reliability gain in function of the upscaling processes size parameter (in grid points). The gains compare the upscaled probabilities to the original forecast.

as the scale increases. The spread reduction is too severe compared with observation and induces then a more pronounced underdispersive situation.

The general usefulness of the upscaled forecast is investigated comparing the upscaled probabilistic forecasts to upscaled deterministic forecasts. We see in Figure 5 that the BSS is positive for all the upscaling size parameters and for all the thresholds. The upscaled probabilistic forecasts at larger scales can then be considered as better than the deterministic upscaled forecast.

Finally, we can make some remarks concerning the impact of the choice of quantile. The results for the 90%, 95% and 99% quantiles show more significant differences for low thresholds. The spatial variability of the individual members and the observation are then analyzed. Figures 4(b) and 4(c) show the empirical variograms of the ensemble members and observation binary fields defined by two thresholds (respectively 0.1 and 10 mm/24h). For the lowest threshold, we note important differences between the observation spatial variability and the spatial variability of the different members. Small structures of low intensity are described within the radar observation fields which are not represented in the forecasts. This situation can explain the sensitivity of the results to the choice of quantile.

6 Summary and Recommendation

We investigated two spatial methods applied to ensemble probabilistic forecasts. Verification results have been shown for daily precipitation during a summer period. A guideline for the use of those methods to generate new probabilistic products can be drawn.

Concerning the application of the neighborhood method, an optimal size parameter in terms of BS was found. However, we noted that this optimal solution (40 grid points) leads to an important loss of sharpness and the resulting forecast has a similar quality as a smoothed deterministic forecast. Since the maximum gain, specially in terms of resolution, is realized at the beginning of the environment extension, we advocate the use of a smaller radius of influence. For example, a size parameter of 10 grid points maximizes the gain in resolution and improves the reliability up to a factor of 2.

Concerning the upscaling procedure, the choice of the size parameter relies on the expectation of the user. For all investigated scales, the upscaled probabilistic forecast performs better than the upscaled deterministic forecast. Moreover, for high thresholds, the forecast resolution is improved after upscaling compared to the original fine scale probabilistic forecast. The better discrimination at large scales between event and non event is then of high relevance for decision making. The negative impact of the upscaling on the reliability can be solved later by the application of a calibration technique.

Acknowledgements

We thank S. Theis and C. Peralta for very useful comments and remarks.

References

[1] V. J. Berrocal, A. E. Raftery, and T. Gneiting. Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts. *Monthly Weather Review*, 135:1386, 2007.

[2] E. E. Ebert. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15:51–64, March 2008.

[3] J. M. Fritsch and R. E. Carbone. Improving Quantitative Precipitation Forecasts in the Warm Season: A USWRP Research and Development Strategy. *Bull. Amer. Met. Soc.*, 85:955–965, July 2004.

[4] C. Gebhardt, S. E. Theis, M. Paulat, and Z. Ben Bouallègue. Experimental ensemble forecasts of precipitation based on a convection-resolving model. *Accepted for publication in Atmos. Res.*, 2010.

[5] E. Gilleland, D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert. Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5):1416–1430, 2009.

[6] T. Gneiting, A. E. Raftery, A. H. Westveld, and T. Goldman. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, 133:1098, 2005.

[7] D. Harris, E. Foufoula-Georgiou, K. K Droegemeier, and J. J. Levit. Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, 2(4):406–418, 2001.

[8] C. Marzban and S. Sandgathe. Verification with Variograms. *Weather and Forecasting*, 24:1102, 2009.

[9] S. J. Mason. On Using "Climatology" as a Reference Strategy in the Brier and Ranked Probability Skill Scores. *Monthly Weather Review*, 132:1891, 2004.

[10] A. H. Murphy. A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, 12:595–600, June 1973.

[11] C. Peralta and M. Buchhold. Initial condition perturbations for the COSMO-DE-EPS. *Submitted to COMSO newsletter*, 2011.

[12] N. M. Roberts and H. W. Lean. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, 136: 78, 2008.

[13] C. S. Schwartz, J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, and M. S. Wandishin. Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership. *Weather and Forecasting*, 25:263, 2010.

[14] M. S. Tesini, C. Cacciamani, and T. Paccagnella. Statistical properties and validation of Quantitative Precipitation Forecast. *COSMO newsletter*, 10:45–54, 2010.

[15] S. E. Theis, A. Hense, and U. Damrath. Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications*, 12:257–268, September 2005.

[16] D. S. Wilks. *Statistical methods in the atmospheric sciences*. Academic Press, New York, 467pp., 1995.