**RMetS**

Royal Meteorological Society

# Spatial techniques applied to precipitation ensemble forecasts: from verification results to probabilistic products

Zied Ben Bouallègue* and Susanne E. Theis
*Deutscher Wetterdienst (DWD), Offenbach, Germany*

**ABSTRACT:** Spatial techniques have been developed to quantify the performance of a system beyond the classical point-to-point comparison with observations. Including spatial neighbourhood information in the verification process, the quality of a forecast can be better characterized. Guidance for the interpretation of deterministic forecasts can also be delivered. This paper investigates the application of spatial techniques to ensemble forecasts. The aim is to assess ensemble forecast skills better and to provide improved guidance to the forecasters in the form of refined probabilistic products. Two spatial techniques are applied to precipitation forecasts derived from an ensemble system at the convective scale (COSMO-DE-EPS). The first technique is a smoothing method which enlarges the ensemble sample size by neighbouring forecasts. The resulting forecasts are called fuzzy probabilistic forecasts. The second method is an upscaling procedure which modifies the reference area of the probabilities. Fuzzy and upscaled probabilistic forecasts are assessed over a 3 month period covering summer 2011. The impact of smoothing and upscaling is investigated for a range of neighbourhood sizes and spatial scales respectively. Based on the verification results, recommendations are drawn how to use these techniques in optimally presenting COSMO-DE-EPS probabilistic products to forecasters who issue weather warnings.

## 1. Introduction

In recent years, numerical weather prediction (NWP) models have gone to smaller and smaller grid sizes. Within the last 5 years, the grid size of many limited-area models has become fine enough to allow an explicit representation of convective processes (Baldauf *et al.*, 2011, and references within). This aims at an improved simulation of convection-related weather such as strong wind gusts and heavy precipitation. These improvements are highly relevant for the quality of severe weather warnings.

However, apart from the obvious benefits, the smaller grid size creates a challenge in terms of predictability. The ability of a system to resolve small scales results in forecast errors that grow rapidly (Lorenz, 1969). Convective processes are non-linear and strongly affected by uncertainties. Therefore, precipitation-related forecasts of convection-permitting models should be produced and interpreted within a probabilistic framework.

On the numerical weather prediction side, ensemble forecasting is today a standard strategy adopted to deal with forecast uncertainties (Lewis, 2005). For limited-area models, variations in boundary condition, initial condition, physics parameterization and/or dynamics formulation aim to reflect the uncertainties related to the forecasting process. Ensemble forecasting derives a probabilistic view from a sample of deterministic forecasts, thereby providing information about the degree of predictability.

Many weather prediction centres are therefore developing ensemble prediction systems (EPS) at the convective scale (Clark *et al.*, 2009; Vié *et al.*, 2011). At the German weather

service (DWD), an EPS based on the convection-permitting model COSMO-DE has been operational since May 2012 (Gebhardt *et al.*, 2011). This is one of the first operational convection-permitting EPSs worldwide. Because it is a new development and since the ensemble forecasts certainly do not have perfect quality, it is now necessary to learn how to use the forecasts optimally for weather warnings.

On the verification side, spatial verification methods have been developed, to assess precipitation forecasts from high-resolution models, accounting for limited predictability (Ebert, 2008). They put a grid point forecast into its spatial context. The idea is to relax the necessity of exact matching between forecast and observation. The uncertainty inherent to the forecast is integrated *a posteriori* following approaches inspired by 'fuzzy logic' (Zadeh, 1965). An event is seen as occurring somewhere within an area rather than at an exact location or with a certain probability of occurrence rather than in binary terms (yes or no). Neighbourhood approaches, which compare statistical properties of forecast and observation fields within a spatial neighbourhood (Gilleland *et al.*, 2009), are explored in this paper.

The same spatial technique can be used in the context of verifying a forecast and in the context of providing forecast guidance to the forecaster. For example, comparing the fractional occurrence of events within a spatial window, Roberts and Lean (2008) define a new metric to assess the performance of a deterministic forecast: the Fractions Skill Score. Their approach derives a 'scale of usefulness' in order to avoid a naive point-based interpretation of a deterministic precipitation forecast (Roberts, 2008; Roberts and Lean, 2008). This is a syncretic example of the duality of the spatial techniques applications: forecast verification and forecast guidance.

Moreover, the generically similar technique (statistics within a spatial window) can be applied in a different manner leading to complementary information. Considering the near neighbourhood forecasts as possible realizations of a local grid

* Correspondence: Z. Ben Bouallègue, DWD, Frankfurter Straße 135, 63067 Offenbach, Germany. E-mail: zied.ben-bouallegue@dwd.de

point forecast, Theis *et al.* (2005) derive a probabilistic forecast guidance from a single forecast. The spatial technique aims here at representing the spatial uncertainty at the grid-scale while the Roberts and Lean (2008) approach estimates the smallest scale at which the spatial variability can be considered as useful. Scale of usefulness and grid point probabilistic forecasts derived from spatial neighbourhoods are complementary guidances that contribute to the forecast interpretation. However, Theis *et al.* (2005) go one step further in terms of forecast guidance by generating refined forecast products presented to the forecaster. The present paper follows along this line and explores spatial techniques in terms of verification and refined forecast products, now with the focus on a convective-scale ensemble system.

Schwartz *et al.* (2010) have already taken the step to apply the Theis *et al.* (2005) approach to ensemble forecasts. The resulting probabilities correspond to a spatially smoothed version of the raw probabilities which are directly derived from the ensemble members at each grid point. This smoothing procedure can be seen as a computationally inexpensive method to enlarge the ensemble sample size by including the spatial neighbourhood forecasts of all members in the probability computation (Ben Bouallègue *et al.*, 2013). It has been shown (Schwartz *et al.*, 2010; Ben Bouallègue *et al.*, 2013) that the smoothing has a positive impact on the probabilistic forecast skill, in particular in terms of reliability but also to some extent in terms of resolution. However, the relationship between benefits and size of the spatial neighbourhood as well as the limit of the method still have to be explored.

Smoothing inevitably reduces the sharpness in the forecasts, i.e. probabilities close to 0 or 100% will occur in fewer cases. This may be the correct thing to do, because it corresponds to the inherently low predictability. However, in many situations the forecast needs to reach some level of certainty before it may be used for a weather warning in practice. Therefore, this paper investigates yet another spatial technique which we call 'upscaling'.

'Upscaling' aims to alleviate the problem of low predictability by changing the spatial scale of the forecast output. In weather forecasting, a spatial scale and a time window are often associated with the prediction, e.g. the probability that it will rain anywhere within a specific region and anytime within a specific time interval. This reference area and time must be known in order to interpret the forecast correctly (cf Gigerenzer *et al.*, 2005). For example, Epstein (1966) described the relationship between point and area probabilities for idealized cases and warned against confusion between these two kinds of forecast. The reference area of an ensemble probabilistic forecast can be modified through an upscaling procedure as described for verification purposes by Marsigli *et al.* (2008). Choosing the maximum value of each member within pre-defined spatial windows, new probabilistic products can be derived and interpreted as the probabilities that an event occurs anywhere within the selected windows. The forecast is still produced by the fine-scale model and still retains its benefits such as the occurrence of heavy precipitation which may only be captured by a convection-permitting model. However, the resulting forecast is formulated for a larger area and time window than the original grid size and the original time interval of the model output. For example, one could look at the probability of heavy precipitation anywhere within the region of Berlin and anytime within the afternoon.

This paper explores how two spatial techniques can better characterize the performance of an ensemble forecasting system and how they can be used to provide guidelines for the generation of more skillful probabilistic products. The first technique is the spatial neighbourhood and aims at improving the probabilistic forecast at grid scale. The second technique is the spatial upscaling procedure and aims at omitting the fine-scale information when issuing a probabilistic forecast product. The techniques are applied independently and are meant to provide two separate types of products. These products may then combine well to form a consolidated forecast guidance.

Smoothing and upscaling are applied here to precipitation forecasts derived from the COSMO-DE-EPS, an ensemble prediction system at the convective scale. Verification is performed for a range of spatial parameters, i.e. neighbourhood environment and window sizes, over a 3 month period covering summer 2011.

The rest of this manuscript is organized as follows: Section 2 describes the convection permitting ensemble COSMO-DE-EPS and the application of the two spatial techniques. Section 3 presents the dataset and verification methodology. Section 4 shows and discusses the results. Section 5 concludes and gives an outlook.

## 2. Ensemble and spatial techniques

### 2.1. COSMO-DE-EPS

An EPS at the convective scale has been developed at DWD and has been operational since May 2012 following a pre-operational phase of one and a half years. Based on the convection-permitting model COSMO-DE, a 2.8 km grid-spacing configuration of the COSMO model (Steppeler *et al.*, 2003; Baldauf *et al.*, 2011), COSMO-DE-EPS resembles the basic ideas of a multi-model approach including variations of lateral boundary conditions, model physics and initial conditions. Details about the variations setup can be found in Gebhardt *et al.* (2011) and Peralta *et al.* (2012). The ensemble forecasts cover Germany and have lead times up to 21 h. The forecast update follows a cycle of 3 h. The results within this paper focus on the 0000 UTC run.

The (pre-)operational version of COSMO-DE-EPS comprises 20 members. Probabilities are generated by applying a frequentist approach: all members are considered as equally probable and then contribute to the probability calculation with equal weights. An example of a probabilistic forecast derived from COSMO-DE-EPS is provided in Figure 1. It shows the probability of precipitation exceeding $10 \, \text{mm} \, (6 \, \text{h})^{-1}$, valid on 22 June 2011 at 1800 UTC. The probabilistic forecasts based on a sample size of 20 members and with a spatial scale corresponding to the model grid size (2.8 km) are called hereafter 'original' probabilistic forecasts.

### 2.2. Smoothing

The smoothing procedure followed here is an application of the neighbourhood method (Theis *et al.*, 2005). It consists of considering spatial neighbourhood forecasts as possible realizations of a forecast at a particular grid point. Schwartz *et al.* (2010) have shown that neighbourhood ensemble-based probabilities can be defined as the mean probabilities within a given environment around each grid point. From the ensemble forecast perspective, the neighbourhood method can be considered as a way to increase the sample size of the ensemble at low computational cost (Ben Bouallègue *et al.*, 2013).

In this study the spatial environment is defined as circular and it is characterized by a size parameter called radius of
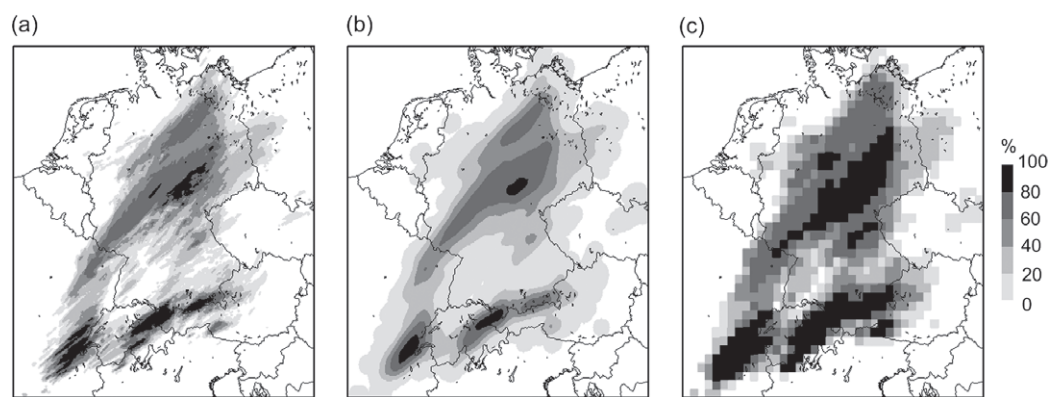
Figure 1. Example of 22 June 2011 valid at 1800 UTC for a threshold of 10 mm (6 h)$^{-1}$. (a) COSMO-DE-EPS probabilistic forecast, (b) fuzzy probabilistic forecast for a radius of influence of 10 grid points (28 km), (c) upscaled probabilistic forecast for windows of 10 by 10 grid points (28 by 28 km).

influence. An example of neighbourhood ensemble-based probabilistic forecast considering a radius of influence of 10 grid points (28 km) is shown in Figure 1(b). This forecast is a smoothed version of the original field (Figure 1(a)). Both of them have the same spatial reference (2.8 km) and it can be noted that all the information needed to construct the smoothed field is already contained in the original one. The neighbourhood ensemble-based probabilistic forecasts are hereafter called 'fuzzy' probabilistic forecasts.

### 2.3. Upscaling

The upscaling approach followed here consists of changing the reference area of a probabilistic forecast in terms of spatial scale. Practically, an event defined as 'precipitation exceeding (below) a certain threshold' is considered, the maximum (minimum) precipitation value of each ensemble member within predefined windows is taken as the precipitation sample field at the new spatial scale. New probability fields are then calculated from those values following a frequentist approach similar to that for the original probabilistic forecasts. The derived probabilities can be interpreted as referring to an event that occurs anywhere within the selected windows.

An example of upscaling is provided in Figure 1(c) for windows of 10 by 10 grid points (28 by 28 km). The COSMO-DE domain is divided into squared windows. The scale of validity of the probabilistic forecast after upscaling is 28 km, while it is 2.8 km for the original probabilistic forecast (Figure 1(a)). In contrast to the smoothing procedure which can be applied directly to a probability field (by taking the average probability over a given region, Section 2.2), the upscaling procedure cannot infer the resulting probabilities directly from the grid point probabilistic forecasts. The probability that an event will occur anywhere within a given region is not just a function of the probabilities that the event will occur in its sub-regions. The upscaling procedure really has to draw the information from the single ensemble members (not shown), because the spatial coherence within the precipitation field plays an additional role. The probabilistic forecasts at the new selected scale are hereafter called 'upscaled' probabilistic forecasts.

Note that the choice of the squares for the upscaled product allows the probabilities to be plotted on a coarse grid (squares) which directly marks the reference areas of the probabilities without any spatial overlap (Figure 1(c)). This kind of visualization has shown to be a good method in practice, because it

reminds the forecasters that these are 'area' probabilities instead of 'grid point' probabilities. For the smoothing approach, such a 'reminder' is not necessary, because the resulting probabilities still refer to the model grid points anyway. For the smoothing, circles are used instead of squares, because the resulting fields look better, without any artificial edges in the probability field.

## 3. Dataset and verification measures

### 3.1. Dataset

The COSMO-DE-EPS forecasts are assessed over a period of 3 months (June, July, August) which covers summer 2011. Precipitation accumulated over 6 h is investigated and verified against gauge adjusted radar precipitation estimates. This observational dataset combines hourly values point-measured at the precipitation stations with the areal precipitation data of 16 weather radars (Weigl and Winterrath, 2009). Verification of original and fuzzy forecasts is performed at the model grid scale. When upscaled probabilistic forecasts are assessed, the upscaling procedure is also applied to the observations: the maximum value of the observation field within the predefined windows is taken as the observation field at the new scale. This requires a sufficiently dense observation network, which motivates the choice to verify the forecasts against radar-type observations rather than rain gauges.

### 3.2. Verification measures

The effect of the spatial techniques is measured by standard tools of probabilistic verification (Wilks, 2006). In the framework of probabilistic verification, the two main attributes that contribute to the quality of a forecast are reliability and resolution. The reliability of a forecast assesses the agreement between the forecast probability and the mean observed frequency (conditional on the forecast probability), while the resolution describes the ability of the forecast to distinguish between event and non-event. More attention is given to resolution than to reliability, because resolution is a more fundamental property. Toth *et al.* (2003) state that 'the intrinsic value of forecast systems lies not in their reliability [ . . . ] but in the resolution [ . . . ]'. Reliability can still attain a substantial improvement through post-processing techniques using training data from past forecasts and observations (Gneiting *et al.*, 2007).

In this study, several quality measures are applied. The decomposition of the Brier Score provides the reliability and resolution. The reliability diagram directly shows the agreement between forecast probability and relative frequency of the observed event. The relative operating characteristic (ROC) is used to evaluate the resolution of forecasts at different scales. The ROC curve plots hit rate *versus* false alarm rate using a set of increasing probability thresholds. The ROC area, the area under the ROC curve, is calculated using a trapezoidal approximation. A probabilistic forecast with a ROC area greater than 0.7 is generally considered 'useful' while a ROC area greater than 0.8 indicates a 'good' prediction (Mullen and Buizza, 2002).

A third attribute, the sharpness, is also computed (following Mason, 2004). Sharpness is a property of the forecast only and measures the ability of a system to provide forecasts far from the climatological frequencies.

Reliability and resolution gains as well as sharpness loss (Ben Bouallègue *et al.*, 2013) are computed to compare the new products to a reference forecast. It is important to note that the probabilistic forecasts are always compared at the same scale since the BS decomposition is only meaningful if all samples share the same climatology (Hamill and Juras, 2006).

Reliability diagram, Brier Score decomposition and ROC area estimation require a discretization of the issued probability forecasts. A fixed number of probability categories is considered here (11 bins) and they are defined as $P < 0.05$, $0.05 \leq P < 0.15$, ..., $0.85 \leq P < 0.95$, $P \geq 0.95$. This binned approach alleviates the sensitivity of the scores to the ensemble size (Buizza *et al.*, 1999).

## 4. Results and discussion

### 4.1. Fuzzy probabilistic forecasts

The smoothing has substantial impact on reliability and sharpness. This is demonstrated in Figure 2 which shows reliability diagrams for two precipitation thresholds (1 mm $(6 h)^{-1}$ and 10 mm $(6 h)^{-1}$). The original probabilistic forecast is compared to fuzzy probabilistic forecasts considering two smoothing

intensities (radii of 30 and 60 grid points). For a threshold of 1 mm $(6 h)^{-1}$, nearly perfect reliability (reliability curve near the diagonal) is reached when applying smoothing with a size parameter of 30 grid points. A loss of sharpness is simultaneously perceived, in that fewer probabilistic forecasts fall in high probability categories and more in categories near the climatological frequency. An increase of the radius of influence to 60 grid points decreases further the sharpness and rotates the reliability curve anti-clockwise around the climatological frequency. This indicates a tendency to move from underdispersion towards overdispersion. For a threshold of 10 mm $(6 h)^{-1}$, the smoothing has a positive impact on the reliability for both radii (30 or 60 grid points) but the loss in sharpness is much more pronounced for a radius of 60 grid points.

So far, these results reflect the typical trade-off between improved reliability and decreased sharpness. Figure 3 further elucidates this trade-off, showing reliability gain and sharpness loss for two thresholds (1 mm $(6 h)^{-1}$ and 10 mm $(6 h)^{-1}$), as a function of the radius of influence. The gain in reliability increases rapidly with the radius of influence and reaches its maximum at around 40 grid points for both thresholds while the sharpness loss increases linearly with the radius of influence.

Figure 3 also shows the gain in resolution which deserves our special attention (Section 3.2). Resolution improves by the neighbourhood method but the maximum gain is much smaller (less than 20%). Apparently, the COSMO-DE-EPS probabilistic forecast slightly benefits from the uncertainty information which is mimicked by the spatial variability within the individual ensemble members. In other words, COSMO-DE-EPS is improved by adding some 'uncertainty in location'.

Furthermore, Figure 3 shows that the maximum resolution gain is reached for smaller size parameters than for the reliability gain. This raises the question of an optimal size parameter. An optimal choice of the smoothing intensity certainly depends on the user's individual ability to tolerate a lack in reliability, sharpness or resolution. Further guidance from verification results is provided by Figure 4. It compares the performance of the ensemble system (COSMO-DE-EPS) to the performance of the deterministic one (COSMO-DE) after applying the same smoothing to both issued forecasts. Probabilities derived from
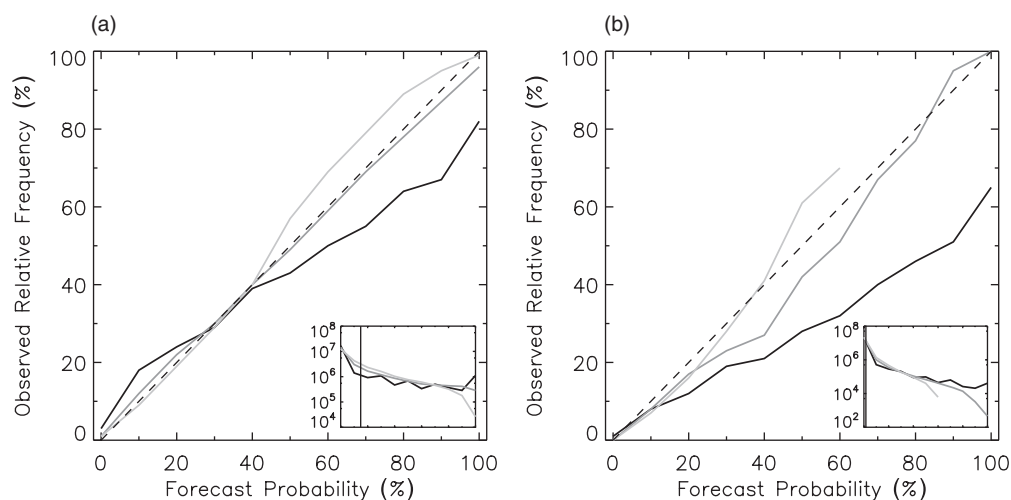
Figure 2. Reliability diagrams for thresholds of (a) 1 mm $(6 h)^{-1}$ and (b) 10 mm $(6 h)^{-1}$. The black lines represent the original COSMO-DE-EPS reliability curves. The dark grey and light grey lines correspond to fuzzy probabilistic forecasts derived by smoothing considering radii of influence of 30 grid points (84 km) and 60 grid points (168 km) respectively. The inset plots show the frequency of usage of each probability category where the vertical lines represent the climatological frequency.
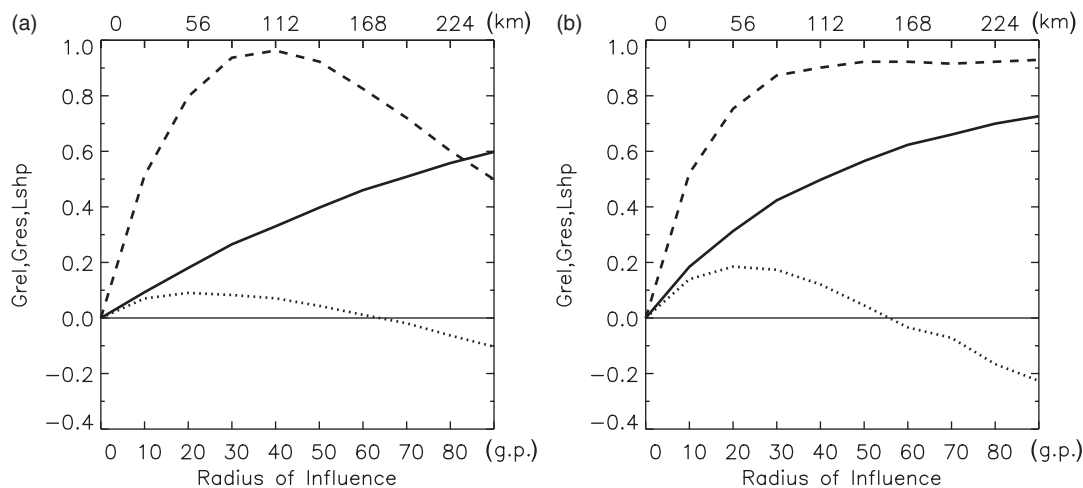
Figure 3. Reliability gain (dashed lines), resolution gain (dotted lines) and sharpness loss (full lines) as a function of the radius of influence, size parameter of the smoothing procedure (expressed in grid points and km) for thresholds of (a) 1 mm [6 h]$^{-1}$ and (b) 10 mm [6 h]$^{-1}$. The reference forecast is the original COSMO-DE-EPS forecast.
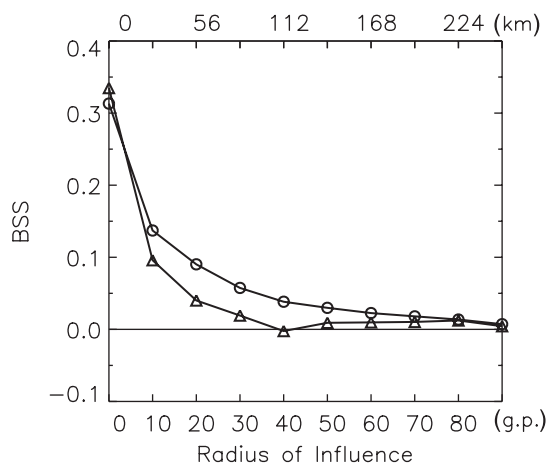


Figure 4. Brier Skill Score as a function of the radius of influence, size parameter of the smoothing procedure. The ensemble COSMO-DE-EPS forecasts are compared to the deterministic COSMO-DE forecasts when the neighbourhood method is applied simultaneously to both of them. The circles refer to a threshold of 1 mm (6 h)$^{-1}$ and the triangles to a threshold of 10 mm (6 h)$^{-1}$.

a sample based on ensemble members and neighbourhood forecasts are compared to probabilities derived from a sample based on neighbourhood forecasts only. Figure 4 shows Brier Skill Scores of the fuzzy probabilistic forecasts considering the 'fuzzy deterministic' forecasts as reference.

The results show that the Brier Skill Score deteriorates with increasing smoothing intensity. The smoothing leads to a decreased benefit of the ensemble relative to the single forecast. The results imply that the application of the neighbourhood to the single forecast already explains a great fraction of the uncertainty which is represented in the smoothed probabilities of the ensemble.

For the higher precipitation threshold, the Brier Skill Score becomes zero at a radius of around 40 grid points (Figure 4). At this point the fuzzy probabilistic forecasts have the same quality as the 'fuzzy deterministic' forecasts. The smoothing intensity starts to dominate and conceal the quality of the uncertainty information coming from the ensemble. This is

not desirable, because the smoothing is only meant as an add-on to the ensemble technique and not as a replacement. It is noted that 'the benefit of the ensemble system is lost' when a forecast of similar quality can be achieved by simply applying the neighbourhood method to a single forecast. Similarly to the result in resolution gain (Figure 3), the results in Brier Skill Score (Figure 4) indicate that a size of 40 grid points would be too large and recommend a choice which is more moderate.

### 4.2. Upscaled probabilistic forecasts

Figures 5 and 6 show the impact of the upscaling on the ensemble probabilistic forecasts in terms of resolution (cf Section 3.2). Figure 5 shows ROC curves for two thresholds (1 mm (6 h)$^{-1}$ and 10 mm (6 h)$^{-1}$) at three different window sizes (2.8, 28, 56 km). The corresponding ROC areas are also noted. For a threshold of 1 mm (6 h)$^{-1}$, the upscaling has low impact: false alarm rate and hit rate increase with the window size in such a way that the ROC areas are similar for all cases. For a threshold of 10 mm (6 h)$^{-1}$, the ROC area increases from 0.81 at 2.8 km to 0.88 at 56 km. This is a substantial improvement, so the increase of the probabilistic forecast scale of interpretation leads to a much better differentiation between event and non-event. These results indicate that especially forecasts of large precipitation thresholds benefit from upscaling.

Figure 6 further elucidates the impact of the window size and concentrates on higher precipitation thresholds. ROC areas as a function of the upscaling window are shown for two precipitation thresholds (10 mm (6 h)$^{-1}$ and 20 mm (6 h)$^{-1}$). These thresholds are part of the warning criteria used by forecasters at DWD. For both thresholds, Figure 6 shows that the ROC areas increase linearly with the window size up to 25 by 25 grid points (70 km). For larger window sizes, a limit is reached. Forecasts at larger scales have no better performance in terms of resolution. For the higher threshold (20 mm (6 h)$^{-1}$), the ROC area exceeds 0.8 for window sizes greater than 15 by 15 grid points (42 km). In other words, the prediction system is able to show 'good' performance (cf Section 3.2) when the information is related to events occurring within squared boxes of length equal or greater than 42 km.
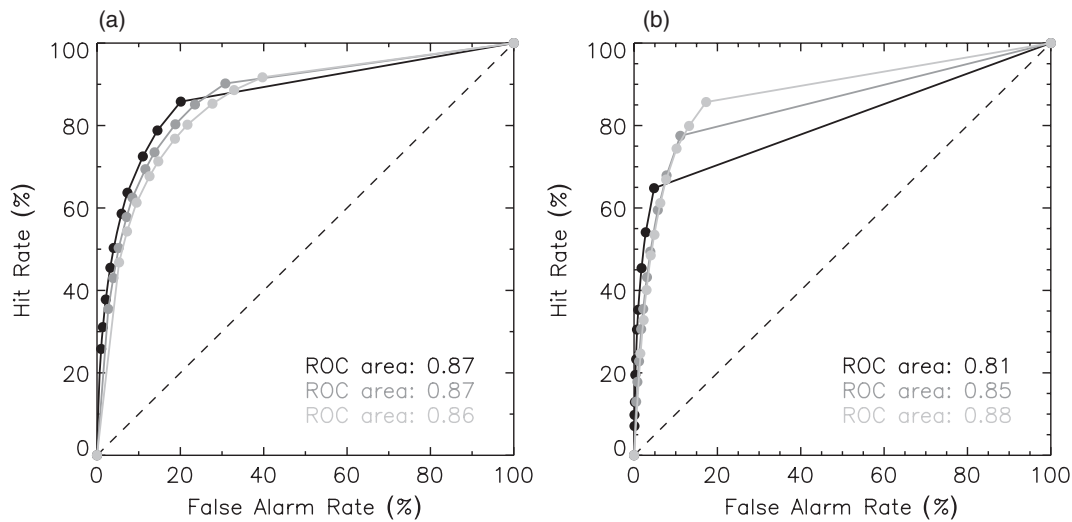
Figure 5. ROC curves and area for thresholds of (a) 1 mm $(6\,h)^{-1}$ and (b) 10 mm $(6\,h)^{-1}$. In black, the results for the raw COSMO-DE-EPS. In dark grey and light grey, the results of the upscaled probabilistic forecasts considering squared windows of 10 by 10 grid points (28 by 28 km) and 20 by 20 grid points (56 by 56 km), respectively.
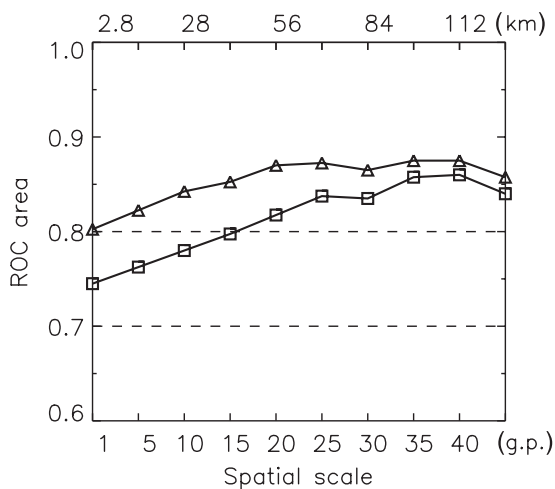


Figure 6. ROC area as a function of the window size used for the upscaling process. The triangles refer to a threshold of 10 mm $(6\,h)^{-1}$ and the squares to a threshold of 20 mm $(6\,h)^{-1}$.
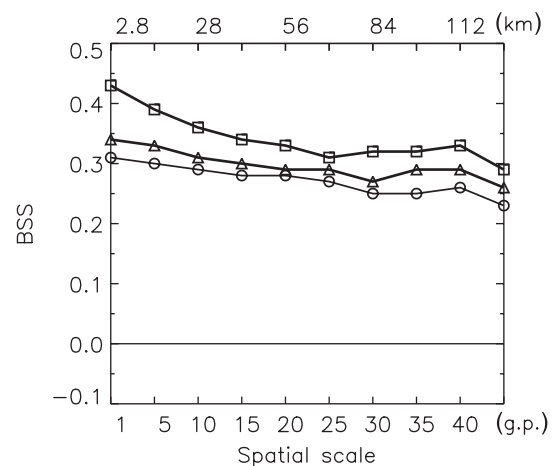


Figure 7. Brier Skill Score (BSS) as a function of the upscaling window size. BSS compares the ensemble COSMO-DE-EPS forecast to the deterministic COSMO-DE forecast. The upscaling is applied simultaneously to both of them as well as to the observation field. The circles refer to a threshold of 1 mm $(6\,h)^{-1}$, the triangles to a threshold of 10 mm $(6\,h)^{-1}$ and the squares to a threshold of 20 mm $(6\,h)^{-1}$.

Apparently, forecast quality benefits from shifting the focus away from the specific location towards a somewhat broader region. This, of course, comes at the cost of omitting fine-scale information. So the optimal choice of a window size would not only depend on verification results, but also on the user's need for high-resolution information which is only retained using a fairly small window size.

Similarly to the smoothing procedure (Section 4.1), also the upscaling is meant as an add-on to the ensemble technique which should not conceal the quality of the ensemble. So in analogy to Figure 4, it is checked whether the ensemble still outperforms the deterministic forecast after applying upscaling to both forecasts. Figure 7 shows Brier Skill Score of the upscaled ensemble forecast using upscaled deterministic forecasts as reference. Results for three thresholds (1 mm $(6\,h)^{-1}$, 10 mm $(6\,h)^{-1}$, 20 mm $(6\,h)^{-1}$) are plotted as a function of the upscaling window size. For all investigated scales, the upscaled probabilistic forecast performs better

than the upscaled deterministic forecast. The benefit of the ensemble technique is not affected by changing the reference area; it is still present even after the exclusion of small-scale variability in the forecast. So COSMO-DE-EPS clearly has value beyond representing 'uncertainty in location'.

Note that Figure 7 compares the ensemble to a single forecast while Figure 4 compares two probabilistic forecasts. The single deterministic forecast only produces probabilities of 0 and 100% without the chance to include any kind of uncertainty range, not even a very simple one. So it is very likely that the ensemble shows a benefit compared to the deterministic in Figure 7. However, such a comparison is still relevant, because models do become operational even before implementing any post-treatments for their probabilistic interpretation. Therefore, the comparison does look at forecast guidance which is provided to forecasters in practice.

### 4.3. Implications for probabilistic products

The verification results show how the spatial techniques affect the quality of probabilities derived from COSMO-DE-EPS. Based on these findings, we make recommendations how to use these techniques in optimally presenting COSMO-DE-EPS probabilistic products to forecasters who issue weather warnings.

The smoothing technique and the upscaling technique provide two separate types of products ('fuzzy' and 'upscaled' probabilities) which may be used simultaneously. Note again that the fuzzy probabilities refer to the event at the grid point while the upscaled probabilities refer to the event anywhere within a larger region (cf Section 2).

When producing grid point probabilities, spatial smoothing of the 'original' probabilities (cf Section 2.1) is recommended, because verification shows a gain in resolution. According to the maximum gain in resolution (Figure 3), a moderate radius of influence is recommended. In practice, the smoothing could be realized either as an automatic procedure or subjectively by the forecaster's eye, since the fuzzy probabilities can be directly inferred from the field of original probabilities (cf Section 2). The subjective approach would leave it to the forecaster to insert the 'uncertainty in location', so they have the chance to recognize fine-scale features associated with orography and to exclude them from smoothing.

As another probabilistic product, we strongly recommend upscaled probabilities, because verification shows a substantial quality gain (Figure 7) especially for higher precipitation thresholds which are relevant for weather warnings. This is further supported by our experience that many DWD forecasters explicitly favour the upscaled probabilities when issuing warnings.

For upscaling, the optimal choice of a window size is a trade-off between good verification results and the forecaster's need for high-resolution information (cf Section 4.2). The latter may be tied to the warning strategy of the weather service, such as the required size of alert areas. Depending on the forecaster's needs, providing several types of probabilistic products simultaneously is recommended, for example the fuzzy grid point probabilities as guidance for local information and the upscaled probabilities as guidance for alert areas.

### 5. Conclusion and outlook

Two spatial techniques are applied to precipitation forecasts derived from the short-range convection-permitting ensemble system COSMO-DE-EPS. Smoothing and upscaling are used to generate two new probabilistic products, fuzzy and upscaled probabilistic forecasts, respectively. Verification results over the summer period 2011, for a range of spatial parameters, help to better characterize the performance of the system and to draw a guideline for the generation of more skillful probabilistic products.

The smoothing consists of increasing the ensemble sample size by spatial neighbourhood forecasts and aims at improving the probabilistic forecast at grid-scale. The derived fuzzy probabilistic forecasts benefit from a better uncertainty representation. Smoothing improves the reliability of the forecast but, as a counterpart, implies a loss of sharpness. An optimal radius of influence in terms of reliability gain exists. However, this optimal solution leads to an important loss of sharpness and the resulting forecast is similar in quality to a smoothed deterministic forecast. Since the maximum gain in terms of resolution is reached for a smaller radius of influence, the use of a moderate size parameter is recommended.

The upscaling of ensemble forecasts delivers probabilistic forecasts at selected scales. An upscaled probability is interpreted as the probability that an event occurs anywhere within a given region which is larger than the grid size of the forecast model. The ensemble forecast resolution, its ability to discriminate between event and non-event, is substantially improved by increasing the interpretation scale up to a certain limit. The upscaling procedure does not harm the benefits of the ensemble compared to the deterministic forecast. Therefore, upscaling is recommended, because it improves the quality of the probabilistic precipitation forecasts in terms of resolution, especially for higher precipitation thresholds. Of course, the benefit comes with the cost of omitting the information about the specific location of an event.

As an outlook, the findings in this paper could be extended. It would be interesting to combine the spatial techniques with additional efforts to improve the probability forecasts. For example, a statistical postprocessing scheme (e.g. Ben Bouallègue, 2013) could be implemented for operational calibration of the probabilities. The spatial techniques could be combined with such a post-processing scheme and then the benefit of the spatial techniques could be explored again.

### References

Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Mon. Weather Rev.* **139**: 3887–3905.

Ben Bouallègue Z. 2013. Calibrated short-range ensemble precipitation forecasts using extended logistic regression with interaction terms. *Weather Forecast.* **28**: 515–524.

Ben Bouallègue Z, Theis SE, Gebhardt C. 2013. Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. *Meteorol. Z.* **22**: 49–59.

Buizza R, Hollingsworth A, Lalaurette F, Ghelli A. 1999. Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. *Weather Forecast.* **14**: 168–189.

Clark AJ, Gallus WA, Xue M, Kong F. 2009. A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Weather Forecast.* **24**: 1121–1140.

Ebert EE. 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorol. Appl.* **15**: 51–64.

Epstein ES. 1966. Point and area precipitation probabilities. *Mon. Weather Rev.* **94**: 595–598.

Gebhardt C, Theis SE, Paulat M, Ben Bouallègue Z. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.* **100**: 168–177.

Gigerenzer G, Hertwig R, Van Den Broek E, Fasolo B, Katsikopoulos KV. 2005. A 30% chance of rain tomorrow: how does the public understand probabilistic weather forecasts? *Risk Anal.* **25**: 623–629.

Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. 2009. Intercomparison of spatial forecast verification methods. *Weather Forecast.* **24**: 1416–1430.

Gneiting T, Balabdaoui F, Raftery AE. 2007. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc.* **69B**: 243–268.

Hamill TM, Juras J. 2006. Measuring forecast skill: is it real skill or is it the varying climatology? *Q. J. R. Meteorol. Soc.* **132**: 2905–2923.

Lewis JM. 2005. Roots of ensemble forecasting. *Mon. Weather Rev.* **133**: 1865–1885.

Lorenz EN. 1969. The predictability of a flow which possesses many scales of motion. *Tellus* **21**: 289–307.

Marsigli C, Montani A, Paccangnella T. 2008. A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorol. Appl.* **15**: 125–143.

Mason SJ. 2004. On using "Climatology" as a reference strategy in the brier and ranked probability skill scores. *Mon. Weather Rev.* **132**: 1891–1895.

Mullen SL, Buizza R. 2002. The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system. *Weather Forecast.* **17**: 173–191.

Peralta C, Ben Bouallègue Z, Theis SE, Gebhardt C. 2012. Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.* **117**: D07108.

Roberts NM. 2008. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorol. Appl.* **15**: 163–169.

Roberts NM, Lean HW. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather Rev.* **136**: 78–97.

Schwartz CS, Kain JS, Weiss SJ, Xue M, Bright DR, Kong F, Thomas KW, Levit JJ, Coniglio MC, Wandishin MS. 2010. Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather Forecast.* **25**: 263–280.

Steppeler J, Doms G, Schättler U, Bitzer HW, Gassmann A, Damrath U, Gregoric G. 2003. Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorol. Atmos. Phys.* **82**: 75–90.

Theis SE, Hense A, Damrath U. 2005. Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorol. Appl.* **12**: 257–268.

Toth Z, Talagrand O, Candille G, Zhu Y. 2003. Probability and ensemble forecasts. In *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Jolliffe IT, Stephenson DB (eds). John Wiley & Sons, Ltd.: Chichester; 137–164.

Vié B, Nuissier O, Ducrocq V. 2011. Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions. *Mon. Weather Rev.* **139**: 403–423.

Weigl E, Winterrath T. 2009. Radargestützte niederschlagsanalyse und -vorhersage (RADOLAN, RADVOROP). *Promet* **35**: 78–86.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Academic Press: Burlington, MA; 627.

Zadeh L. 1965. Fuzzy sets. *Inform. Control* **8**: 338–353.