STATISTICAL POSTPROCESSING OF WEATHER PARAMETERS FOR THE DWD HIGH-RESOLUTION MODEL LM

S. Theis* U. Damrath† A. Hense* V. Renner†

*Meteorologisches Institut, University Bonn, Germany
†Deutscher Wetterdienst, Offenbach, Germany

Abstract

Methods for the statistical postprocessing of the weather parameters predicted by a high resolution model are being developed. A neighbourhood method transforms the model output into a probabilistic forecast at every grid point. The probabilistic forecast at a certain gridpoint ist based on the model output within a spatial region surrounding the grid point and within a certain time interval. Pre-operational tests of the method are under way. Additionally, the development and investigation of an alternative method based on wavelet analysis has started.

1 Introduction

At the German Weather Service (DWD) a new numerical weather forecasting system became operational in December 1999. The system consists of the global model GME and the non-hydrostatic limited area model "Lokal-Modell" (LM). The LM has a horizontal resolution of Δx =7 km and produces forecasts for a time period of 48 hours. The DWD is planning to further increase the horizontal resolution of the LM to 2.8 km within the next few years.

Since the weather does exhibit considerable variability even on spatial scales of only a few kilometers, the increase of model resolution is certainly a project worth striving for. When doing so, an eye has to be kept on the predictability of the atmosphere, though. With increasing resolution of the forecast model, the ratio of the lead time of the forecast to the minimum spatial scale of model-resolved weather patterns increases as well.

Thus, in short-range forecasts small spatial features of only a few gridboxes' scale might not always be predictable in a deterministic way. In a high-resolution model output small-scale features rather contain a certain amount of stochastic information. For some variables like precipitation or cloud cover it has to be expected that these stochastic characteristics will noticeably influence the characteristics of the forecast.

Whenever a forecast is prone to contain a non-negligeable amount of stochastic information, it has to be interpreted in a statistical way. The DWD in co-operation with the University of Bonn has started to develop methods that interpret the LM direct model output statistically. Our method of statistical interpretation aims at the transformation of the direct model output into a probabilistic forecast.

2 Methodology

A well established method to gain statistical information for medium-range weather forecasts is the application of an ensemble prediction system (EPS). The EPS generates a sample of simulations with the aim to issue a probabilistic forecast. The probabilistic form of the forecast takes account of the fact that the atmosphere is not deterministically predictable.

The operational use of a high-resolution, limited-area, short-range ensemble prediction system would create an almost prohibitive demand on computer capacities. Therefore, at present, we restrict ourselves to the statistical postprocessing of the operational LM run only.

Two methods are being developed: the *neighbourhood method* (Section 2.1) and the *wavelet method* (Section 2.2). Whereas the wavelet method is still in its infancy, the neighbourhood method is already undergoing a pre-operational test phase.

2.1 Neighbourhood Method

The method intends to estimate properties of the probability distribution of the forecast at each grid point of the LM. In order to obtain information about the probability distribution of the forecast at a certain grid point, an appropriate sample of this forecast is needed. Since we have to restrict ourselves to only one forecast simulation, we need to make a general assumption in order to generate the sample.

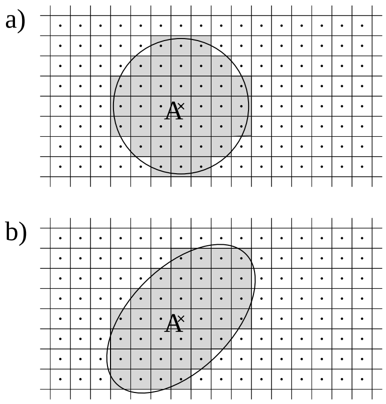

Figure 1: *The spatial neighbourhood of gridpoint A. a) isotropic neighbourhood, b) anisotropic neighbourhood. Grid point values of the shaded gridboxes are assumed to form a random sample of the forecast at gridpoint A.*

2.1.1 Key Assumption

Imagine a certain grid point called A. Around grid point A we define a spatial *neighbourhood* (Figure 1) of circular shape or elliptic shape (cf. Section 2.1.2). All grid point values are modified values of the LM direct model output (cf. Section 2.1.2).

The key assumption of the neighbourhood method is as follows: Grid point values of the shaded gridboxes are assumed to constitute the sample of the forecast at gridpoint A. In other words, all grid point values within the neighbourhood are assumed to be independent and identically distributed according to the probability density function of the forecast at gridpoint A.

Currently, the diameter of the circle enclosing the spatial neighbourhood is chosen to be $6 \times \Delta x = 42\,\mathrm{km}$ in case of an isotropic neighbourhood. In the case of an anisotropic neighbourhood the area of the ellipse is approximately the same as the area of the isotropic neighbourhood.

Additionally, forecasts of neighbouring points in time are included in the sample. Thus, the neighbourhood also extends into the time dimension and the circle in Figure 1 is actually a sphere or an ellipsoid (not shown in Figure 1). Only forecasts on the full hours are taken into account. Currently, the temporal diameter of the neighbourhood is chosen to be 3 hours.

2.1.2 Deterministic Factors

In some cases the assumption (Section 2.1.1) obviously would not hold, if a circular neighbourhood was used or if the original LM forecast values directly served as the the grid point values within the neighbourhood. For example, imagine grid point A is located over the sea and a coastline runs through its neighbourhood. Then the temperature forecasts on grid points which are located over land are generally *not* representative of the temperature forecast on grid point A. A similar example is the dependence of the forecast on orography. The problem occurs whenever there are variables that

- are dynamically relevant and
- vary on a small spatial or temporal scale and
- are independent of the short-range forecast.

We refer to such variables as *deterministic factors.*

The disturbing influence of the deterministic factors has to be alleviated. This is done in several ways:

1. The shape of the spatial neighbourhood (Figure 1) is modified according to the orography. We allow the spatial neighbourhood to have the form of an ellipse. The shape of the ellipse minimizes the variance of orography within the neighbourhood. The centre of the ellipse is restricted to remain at grid point A and the area of the ellipse is forced to be the same as the area of the original circle.

2. We modify the forecasts in the neigbourhood with respect to the deterministic factors. For each neighbourhood a linear regression between forecast and orography is performed. Then the forecasts of the neighbourhood are linearly adjusted to the orography at gridpoint A.

3. The neighbourhood is cut off at coastlines. If gridpoint A is situated on land, then all gridpoints situated on the sea are excluded and vice versa.

4. For the statistical postprocessing of temperature the neighbourhood is cut off at boundaries of snow cover. If gridpoint A is covered by snow, then all gridpoints without snow are excluded and vice versa.

2.1.3 Quantile Estimator

The linearly modified forecasts within the elliptic neighbourhood serve as a random sample of the forecast at grid point A. Once we have generated a sample, we can estimate properties of its probability distribution function F.

We have decided to estimate the quantile function x. The quantile function x is a function of probability p. It is the inverse of the probability distribution function F. Hence, there is a probability p that the quantile $x(p)$ is greater than the correct forecast.

The estimation of the quantile function follows a method by Müller (1991). Müller (1991) proposes a non-parametric kernel quantile estimator. The calculation of the estimator is directly based on the sample. Moon und Lall (1994) present an application of this method.

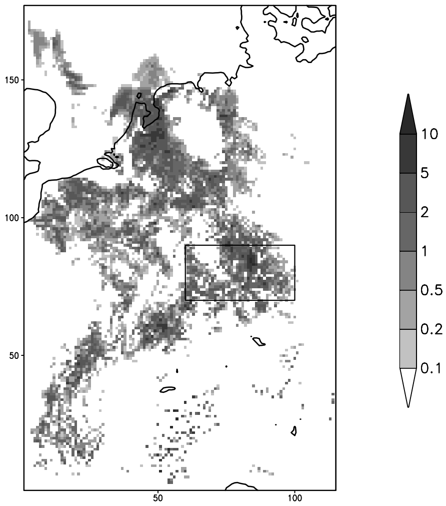

Figure 2: *LM direct model output of a convective situation over Central Europe at 12 UTC on July 4^{th}, 1994. Forecasted variable: Accumulated precipitation of the preceding hour. Unit: mm. Lead time of the forecast: 12 hours. Horizontal resolution: 7km.*

2.2 Wavelet Method

The wavelet method assumes that small spatial scales of the forecast field are not deterministically predictable. Therefore, they must be removed from the forecast in order to achieve a more reliable forecast.

The LM model output is transformed with respect to a wavelet basis. Then all wavelet coefficients which belong to wavenumbers greater than a certain threshold value are set to zero. The threshold value may vary with location in space. Finally, the inverse wavelet transformation is carried out, so we get the grid point values of the postprocessed forecast field. For the wavelet transformation we use the *fast wavelet transform* by Sweldens (1996).

3 Example

An example of the statistical postprocessing via the neighbourhood method is presented. At each grid point the LM direct model output is transformed into a probabilistic forecast. We choose a forecast of heavy convective precipitation in the summer (see Figure 2). The enlarged area in Figure 3 demonstrates the extreme spatial variability of the direct model output. The size of a grid box is $7\,\mathrm{km} \times 7\,\mathrm{km}$. It is unlikely indeed that the exact spatial location of these small-scale precipitation patterns can be predicted in a deterministic way.

Figure 4 shows three quantile estimates that result from the statistical postprocessing of the precipitation forecast in Figure 3. The quantile estimators were calculated for the probabilities $p = 0.25$, $p = 0.5$ and $p = 0.75$, respectively. For example, at each grid-

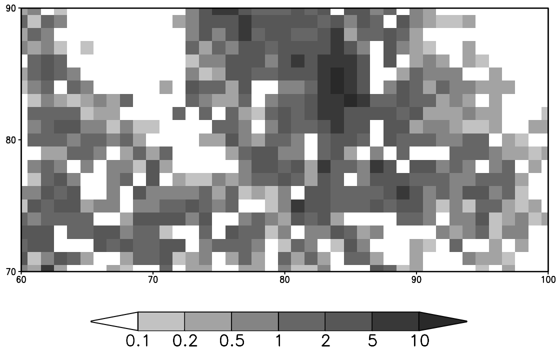

Figure 3: *Zoom of the area which is marked in Figure 2.*

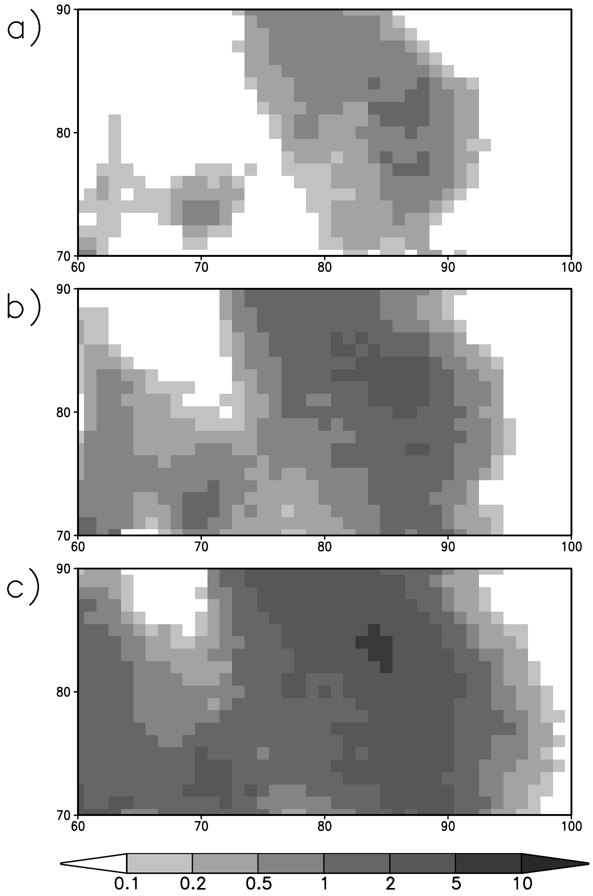

Figure 4: *Result of the statistical postprocessing: Quantile estimates of the precipitation forecast shown in Figure 3. a) Quantiles for the probability $p = 0.25$. b) Quantiles for the probability $p = 0.5$. c) Quantiles for the probability $p = 0.75$. Unit: mm.*

point there is a probability of 25% that the correct forecast is less than the corresponding 0.25-quantile in Figure 4a. Similarly, there is a probability of 50% that the correct forecast is less than the quantile in Figure 4b and there is a probability of 75% that the correct forecast is less than the quantile in Figure 4c.

The quantiles must be interpreted for each grid-box individually. Note that the chances are far less than 25% that the correct precipitation forecast will fall below the corresponding 0.25-quantile (Figure 4a) at *all* grid points simultaneously. We expect the correct precipitation forecast to exceed the 0.25-quantile only at about 75% of the grid points. Similarly, at about half of the grid points the correct precipitation forecast will be less than the 0.75-quantile and greater than the 0.25-quantile.

The differences between the three quantiles in Figure 4 reflect the range of uncertainty associated with the prediction of small-scale patterns. We would like to emphasize, though, that the differences between the three quantiles do not necessarily predict the *overall* skill of the LM forecast. The quantiles are restricted by the assumption that the large-scale precipitation forecast is perfect. For example, if the global model GME provides a wrong forecast at the lateral boundaries of the LM, then the resulting forecast errors of the LM will not become visible through the statistical postprocessing of the LM output.

4 Outlook

At present, pre-operational tests using the neighbourhood method are being conducted. This work mainly concentrates on the development of a procedure that will evaluate representative ("statistically smoothed") precipitation values. First results will be presented at the conference.

Subjects of the ongoing development of the methods are - amongst others - the simultaneous, consistent treatment of several parameters, the derivation of probabilistic information by the wavelet method, and the calibration of both methods by comparing their results to those of a larger set of experimentally conducted ensemble forecasts.

References

Moon, Y.-I. und Lall, U. (1994): Kernel quantile function estimator for flood frequency analysis. *Water Resour. Res.* **30**(11), 3095–3103.

Müller, H.-G. (1991): Smooth optimum kernel estimators near endpoints. *Biometrika* **87**(3), 521–530.

Sweldens, W. (1996): Wavelets and the lifting scheme: A five minute tour. *Z. Angew. Math. Mech.* **76 (Suppl. 2)**, 41–44.