Deriving Probabilistic Precipitation Forecasts from LM Simulations

Susanne Theis, A. Hense, U. Damrath[1)] and V. Renner[1)]

Meteorologisches Institut, University Bonn, Germany
[1)] *Deutscher Wetterdienst, Offenbach, Germany*

Abstract

The fact that numerical forecasts of precipitation are often not entirely deterministic is taken into account by two alternative approaches: (1) experimental ensemble predictions with the LM and (2) postprocessing the direct model output of a single LM simulation. While the ensemble approach has a scientifically more sound basis, the postprocessing method is designed to meet criteria of operational feasibility. The ensemble setup includes perturbations of initial conditions, model physics and roughness length. The outcome of the experimental ensemble predictions crucially depends on the input parameters of the stochastic physics scheme. The postprocessing procedure transforms the direct output of a single simulation into a probabilistic forecast, consisting of the mean, quantiles and probabilities of exceedance. The benefit of the postprocessed forecast is mainly contained in the probabilistic forecast, whereas the mean does not outperform the raw model output in every respect.

1 Introduction

It has been recognized that the atmosphere is of inherently chaotic nature (Lorenz, 1963). Precipitation is not only especially prone to chaotic behaviour, it is also the element of the weather forecast which is perhaps most relevant to users. Precipitation-related forecast products and information must be addressed in a probabilistic framework that accounts for their intrinsic uncertainty (Fritsch et al., 1998). Probabilistic information can increase the forecast value to those users who employ the forecast as an input into their decision-making process (Murphy, 1977).

In an attempt to derive probabilistic information from LM forecasts, two different approaches are pursued. The first approach is the development of an experimental ensemble prediction system which runs several LM simulations with perturbed model physics, roughness length and intial conditions. Its drawback is the prohibitive demand of computer capacities for the multiple ensemble member forecasts. The other, more pragmatic approach is the development of a statistical postprocessing technique which transforms the model output of a single LM simulation into a probabilistic quantitative precipitation forecast (QPF) at a given model grid point. If the postprocessing products prove worthwhile, the method is envisaged to become operational. The findings from the ensemble experiments are intended to be used as a guidance for the development of the postprocessing method.

Section 2 describes the design of the ensemble prediction system and presents first results from a pilot study. In section 3 the postprocessing method is introduced and resulting forecast products are assessed in terms of skill and value. A summary is given in section 4.

2 Experimental Ensemble Predictions

As an ideal ensemble design should involve perturbations of all elements of the model which are subject to uncertainty (Balzer, 1998), many different sources of uncertainty have been investigated in short-range mesoscale forecasting. They comprise initial conditions, model physics (e. g. Stensrud et al. (2000)), lateral boundary conditions (e. g. Marsigli et al. (2000)), surface parameters (e. g. Mölders (2001)) and combinations thereof.

In the present study, roughness length, model physics and initial conditions of the LM are perturbed. The conception of perturbing the roughness length has been presented in Theis et al. (2002). Methodology and impact of perturbing model physics is presented in subsection 2.1. The additional inclusion of perturbed initial conditions is described in subsection 2.2.

2.1 Stochastic Physics Scheme

When the large-scale forcing is weak, model physics appear to play a larger role than initial conditions (Stensrud et al., 2000). In recent research, the influence of model physics has mostly been represented by exchanging parametrization schemes or by exchanging the entire model (Du et al., 1997; Stensrud et al., 2000; Hou et al., 2001). However, a significant part of random model error may be caused by neglect of subgrid-scale variability which should be parametrized by stochastic parametrization schemes (Palmer, 2001). In the present study, stochastic physics have been implemented in the LM with the aim to quantify the impact of sub-grid scale variability on the uncertainty of the short-range precipitation forecast. The scheme closely follows the methodology of Buizza et al. (1999).

Methodology

The stochastic parametrization scheme simulates uncertainties caused by the parametrized processes. Using a similar notation as Buizza et al. (1999), the methodology of the stochastic scheme is as follows.

The model equations $\frac{\partial \mathbf{e}}{\partial t}$ of the state vector $\mathbf{e}$ can be seen as the sum

$$\frac{\partial \mathbf{e}}{\partial t} = \mathbf{A}(\mathbf{e}; t) + \mathbf{P}(\mathbf{e}; t),$$

where $\mathbf{A}$ and $\mathbf{P}$ identify the non-parametrized and the parametrized processes, respectively. In case of the Lokal-Modell, parametrized processes are covered by the cloud microphysics, the turbulence, the convection and the radiation scheme.

Each member $\mathbf{e}_j$ of the ensemble may be seen as the time integration

$$\mathbf{e}_j(t) = \int_{t=0}^{t} \{\mathbf{A}(\mathbf{e}_j; t) + \mathbf{P}'(\mathbf{e}_j; t)\} \mathrm{d}t$$

of the perturbed model equations

$$\frac{\partial \mathbf{e}_j}{\partial t} = \mathbf{A}(\mathbf{e}_j; t) + \mathbf{P}'(\mathbf{e}_j; t).$$

At every grid point r the perturbed parametrized tendency is defined as

$$\mathbf{P}'(\mathbf{e}_j; t) \equiv \langle x_j(r, t) \rangle_{D,T} \cdot \mathbf{P}_j(\mathbf{e}_j; t)$$

for any component of the state vector, where x_j denotes a random number which has been sampled uniformly from the interval $[1 - a, 1 + a]$ $(0 < a \leq 1)$. The notation $\langle \cdot \rangle_{D,T}$ implies

Table 1: The parameter settings of the first four experimental ensembles with stochastic forcing. Δx denotes the grid spacing of the model and Δt denotes the time step in the model integration.

	amplitude a	spatial autocorrelation D	temporal autocorrelation T
LLL	0.25	$5 \cdot \Delta x = 35\,\mathrm{km}$	$4 \cdot \Delta t = 160\,\mathrm{s}$
MLL	0.50	$5 \cdot \Delta x = 35\,\mathrm{km}$	$4 \cdot \Delta t = 160\,\mathrm{s}$
HLL	1.00	$5 \cdot \Delta x = 35\,\mathrm{km}$	$4 \cdot \Delta t = 160\,\mathrm{s}$
HHH	1.00	$10 \cdot \Delta x = 70\,\mathrm{km}$	$16 \cdot \Delta t = 640\,\mathrm{s}$

that the same random number x_j is used for all grid points inside a $D \times D$ box and over T timesteps.

Additionally, the solar radiation flux at the ground is altered, so as to be physically consistent with the perturbed temperature tendency from the radiation scheme.

Pilot Study

Optimal values for the amplitude a and the spatio-temporal autocorrelations D and T are not known a priori. Four different configurations have been used so far (Table 1). The configurations are specified in order of amplitude, spatial and temporal autocorrelation parameters. They are termed LLL, MLL, HLL and HHH, where L denotes "low", M denotes "medium" and H denotes "high".

Ensemble simulations have been generated for the case study of July 10, 2002. This particular day is characterized by intense convective precipitation and strong wind gusts. The forecasts have been started on July 10, 2002 at 00 UTC. The lateral boundary conditions are provided by LM analyses. The model domain has been reduced to 177×177 grid points. Version 2.14 of the LM was used. Each ensemble comprises ten members, consisting of nine perturbed simulations and the control simulation.

Results are presented for the LLL and the HHH configurations (Fig. 1). In the HHH ensemble the impact of the perturbations is much higher. While the standard deviation between the LLL ensemble members reaches approximately 30% of the precipitation amounts in the

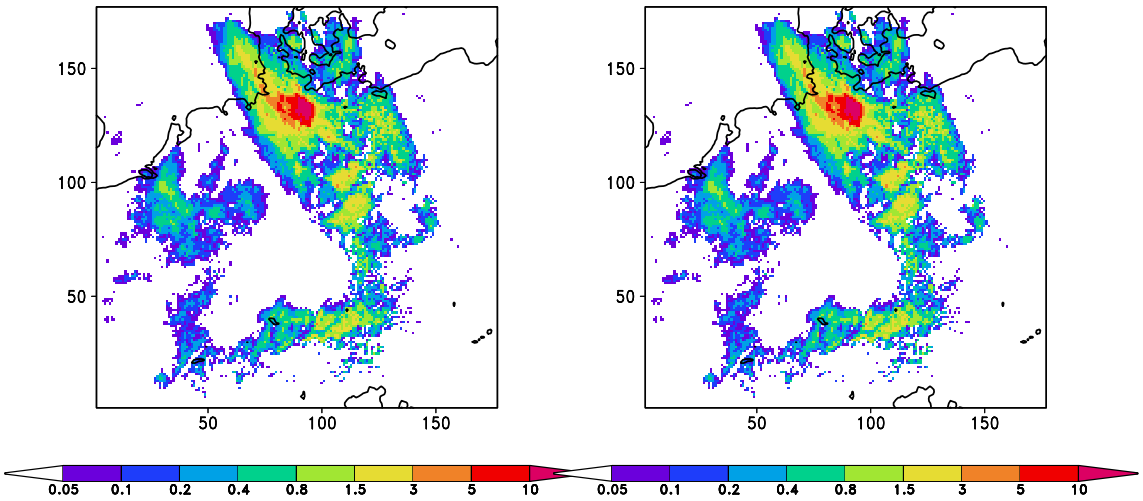

Figure 1: 1h-accumulations of precipitation (mm/h) at 17–18 UTC on July 10, 2002. Standard deviation between the precipitation forecasts of the ensemble members. The two ensembles differ by their configurations of the stochastic physics. Left: LLL configuration. Right: HHH configuration.

unperturbed forecast, the standard deviation between the HHH ensemble members is often as high as the precipitation amounts in the unperturbed forecast itself. The perturbations affect relatively small spatial scales of the forecast.

The ensemble mean and the control simulation have been verified against rain gauge data within a region of 47.0 N–55.5 N and 5.5 E–15.5 E and a time period of (July 10, 06 UTC) – (July 11, 06 UTC). The data set contains 4174 observations of 1 hour accumulations. While the ensemble mean of the LLL configuration renders almost identical verification results as the control simulation, the ensemble mean of the HHH configuration appears to be of higher quality than the control simulation. One has to bear in mind though that the verification period is very short and that the results might be influenced by chance.

2.2 Perturbation of Initial Conditions

In order to assess the relative impact of stochastic physics compared to the impact of initial conditions, another ensemble setup has been constructed. The nudged LM assimilation run provides LM analyses for arbitrary points of time. In an experimental setting the LM simulation is initialized by an LM-analysis of a slightly shifted time. The LM analyses of 01 UTC, 00 UTC and the analysis of 23 UTC of the previous day are used, though the LM simulation is actually started at 00 UTC. Together with the analysis of 00 UTC, the LM analyses of 23 UTC and 01 UTC constitute three different initial conditions. This can be thought of a "poor man's method" to perturb the initial conditions.

The new ensemble setup combines the variation of initial conditions with the previous ensemble setup (Fig. 2). From each of the three initial conditions, three simulations have been started. They differ by the stochastic physics and the roughness length. In the stochastic physics scheme the LLL configuration was used (cf. Table 1). Altogether the ensemble comprises nine perturbed simulations plus the control simulation.

The standard deviation between the ensemble members (Fig. 3) is about as large as in the HHH ensemble. Whereas the stochastic physics scheme mainly affects small scales of the forecast, the initial conditions affect scales up to 200 km. The initial conditions of 23 UTC (cf. Fig. 2) results in a southward shift of the precipitation maxiumum by 100 km, while the initial conditions of 01 UTC results in a northward shift of the maximum by 100 km.

Verification results for the ensemble mean are not considerably better than for the HHH stochastic physics ensemble. The ensemble mean does not outperform the control simulation in every respect. For example, compared to the control simulation the frequency bias of the ensemble mean is mostly deteriorated. This might be due to the fact that ensemble averaging is a smoothing operation (Du et al., 1997).

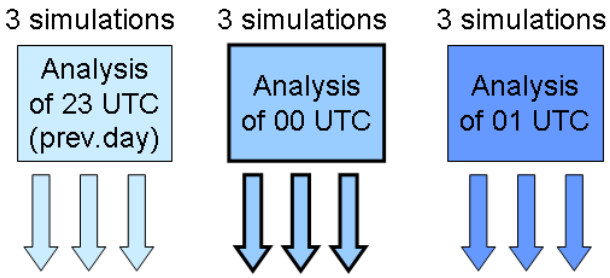

Figure 2: Setup of the ensemble including perturbations of initial conditions. Three simulations per initial condition have been started, differing by the stochastic physics and roughness length. Together with the control simulation, the ensemble consists of ten members.

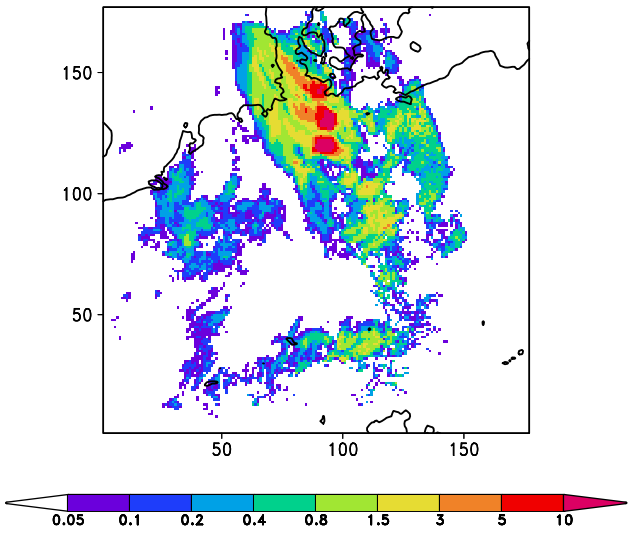

Figure 3: 1h-accumulations of precipitation (mm/h) at 17–18 UTC on July 10, 2002. Standard deviation between the precipitation forecasts of the ensemble members. In addition to perturbations of roughness length and stochastic physics, the initial conditions have been varied according to Figure 2.

3 Statistical Postprocessing

The drawback of the ensemble approach is its prohibitive demand of computer capacities for the multiple ensemble member forecasts. As a computationally cheap alternative, a statistical postprocessing procedure has been developed which transforms a single LM simulation into a probabilistic forecast.

3.1 Methodology

The postprocessing procedure intends to estimate properties of the probability distribution of the QPF 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. This sample is generated by making a simplistic assumption.

Assumption

When postprocessing the model QPF at a given location (x_0, y_0) of the model grid and for a given forecast lead time T_0, a *neighbourhood* around this grid point is defined. It extends both into space (x, y) and time T. Figure 4 shows a schematic view of the neighbourhood in the (x, y)-plane and in the (x, T)-plane, where Δx and Δy denote the size of a gridbox and ΔT denotes the time step between successive model output times. For an operational LM simulation $\Delta x = \Delta y = 7\,\mathrm{km}$ and $\Delta T = 1\,\mathrm{h}$. Shaded grid boxes belong to the neighbourhood.

The key assumption of the postprocessing procedure is as follows. Model QPF at grid points within the neighbourhood are assumed to constitute a sample of the QPF at location (x_0, y_0) and forecast lead time (T_0). In other words, the model QPFs within the neighbourhood are assumed to be independent and identically distributed according to the probability density function of the QPF at location (x_0, y_0) and forecast lead time T_0. The shape and the size of the neighbourhood is held fixed on the entire model domain and at all lead times of the simulation. Harold E. Brooks (NSSL, Oklahoma) presented a similar approach at the 1998 Workshop on Mesoscale Model Verification (Davis and Carr, 2000), confined to a spatial neighbourhood.

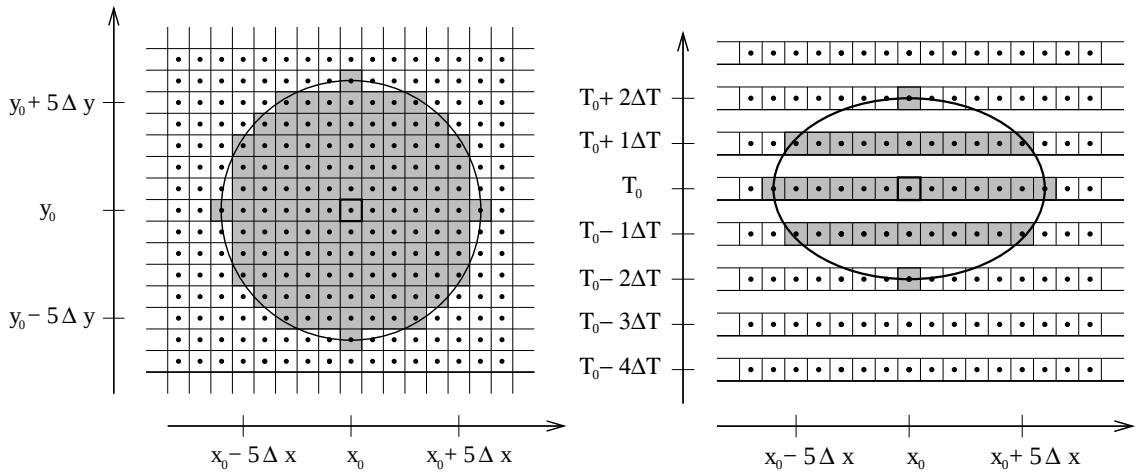

Figure 4: Example of a spatio-temporal neighbourhood of a given grid point at location (x_0, y_0) and forecast lead time T_0. Left: The spatial neighbourhood in the (x, y)-plane. Right: The spatio-temporal neighbourhood in the (x, T)-plane. Δx and Δy denote the size of a gridbox and ΔT denotes the time step between successive model output times. Shaded grid boxes belong to the neighbourhood.

A crucial issue is determining an optimal size and shape of the neighbourhood. Some indication of the order of magnitude may be inferred from verification and predictability studies, especially those which use scale decomposition. Several sizes have been tested empirically. The trial led to an ambiguous outcome and did not allow to decide on an optimum size yet.

Attempts have been made to vary the shape of the neighbourhood in order to account for terrain variability (Theis et al., 2002). This appeared successful when applied to surface temperature, but rendered no improvements in the postprocessed model QPF. Therefore, the standard shape of the neighbourhood (Fig. 4) is still used for QPF.

In addition, the forecasts have been linearly adjusted to orography (Theis et al., 2002) so as to account for orographically induced extended predictability. Again, this has only been proven successful when applied to temperature forecasts and is omitted when postprocessing QPF.

Products of the Postprocessing Procedure

The model QPFs within the neighbourhood serve as a sample of the QPF at the central grid point. Once a sample has been generated, properties of the probability distribution function may be estimated and used as probabilistic forecast products. As the main product of the postprocessing procedure, quantiles are estimated for various probabilities p. The quantiles are calculated using a non-parametric kernel quantile estimator (Moon and Lall, 1994). The fact that precipitation has a mixed discrete-continuous distribution is also taken into account.

Figure 5 shows an example how the quantiles enhance the information content of a model QPF at a given grid point. The bars represent the direct model QPF of the operational LM simulation for July 10–11, 2002 at a given grid point near (6.8° E, 51.6° N). As the LM is a deterministic model, the direct model output does not contain any forecast of uncertainty or risk and therefore lacks consistency (Murphy, 1993). Adding quantiles for different probabilities might improve the forecast in this respect. If the forecast is correct, then there is a probability of 50% that the observed precipitation will be between the 25%- and the 75%-quantile. The difference between the 25%- and the 75%-quantile gives an idea of forecast uncertainty. The 90%-quantile forecasts the amount which will be exceeded with

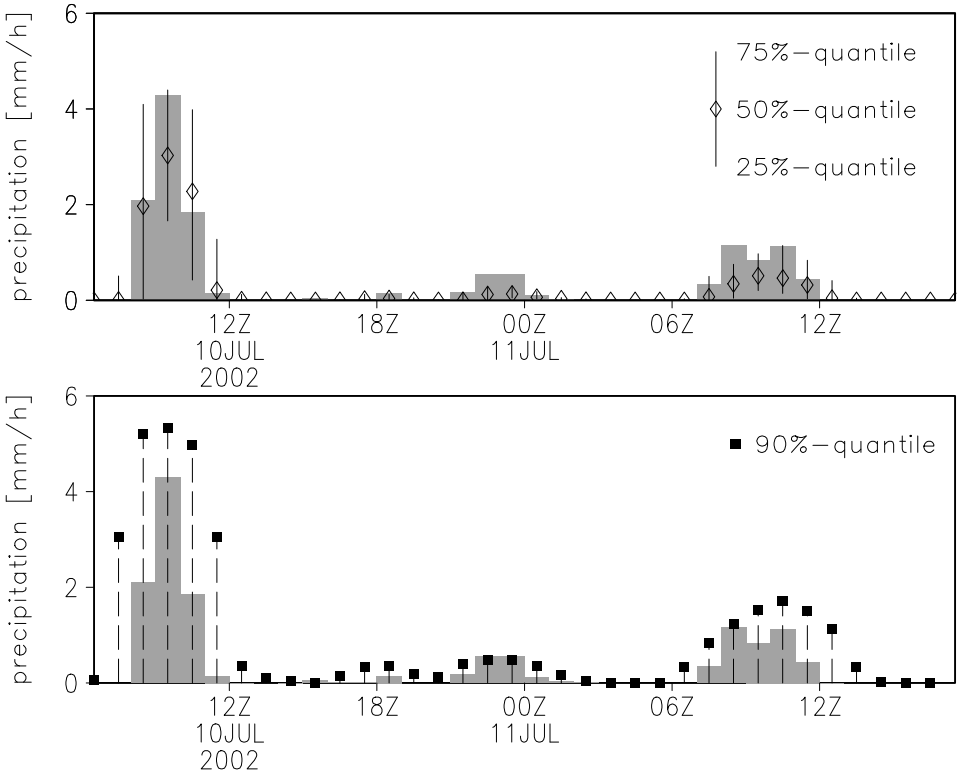

Figure 5: QPF for July 10–11, 2002 at a given grid point near (6.8° E, 51.6° N). The bars represent the direct model output of the LM. Top: The diamond and the endpoints of the corresponding vertical line represent the 25%-, 50%-, and 75%-quantiles respectively. Bottom: The filled squares represent the 90%-quantiles.

a probability of 10%. This could be useful for forecasting risk and issuing warnings.

In addition, the arithmetic mean of the neighbourhood and the probability of exceedance have been calculated at every grid point. The probability of exceedance is just the number of model QPFs within the neighbourhood which are greater than a given threshold, divided by the total number of grid points within the neighbourhood.

3.2 Verification

In the previous section some advantages of the postprocessed forecast have already been mentioned. It still has to be shown whether the products of the postprocessing procedure are of good quality and value. In the following, the mean of the neighbourhood, the probability of exceedance and the 90%-quantile are evaluated by comparing them with observational data.

Model and Observational Data

Verification has been carried out for several time periods: 15–30 September 2000, 1–15 May 2001, 1–15 August 2001, 1–15 September 2001 and 10–24 July 2002. The findings from the different verification periods are similar. In the following, the specific verification procedure and the results are presented for the period 10–24 July 2002.

Observed 1h-accumulations of precipitation data from rain gauges have been collected within an area of 47.0 N–55.5 N and 5.5 E–15.5 E. The data set contains 60 239 values. These precipitation data have been compared to both the raw and the postprocessed LM output.

Table 2: The three different configurations of the postprocessing procedure which have been applied and evaluated for the verification period 10–24 July, 2002. They differ in the size of their spatio-temporal neighbourhood.

Configuration	Spatial Diameter of Neighbourhood	Temporal Diameter of Neighbourhood
"Small"	$7 \times \Delta x \approx 50\,\mathrm{km}$	$4 \times \Delta T \approx 4\,\mathrm{h}$
"Medium"	$13 \times \Delta x \approx 90\,\mathrm{km}$	$5 \times \Delta T \approx 5\,\mathrm{h}$
"Large"	$21 \times \Delta x \approx 140\,\mathrm{km}$	$7 \times \Delta T \approx 7\,\mathrm{h}$

The model data consist of the operational LM simulations which have been run at the German Weather Service during that time period. The lateral boundary conditions have been provided by forecasts of the global model GME. The verification procedure has been applied to LM simulations which have been started at 00 UTC on every day. Only forecasts with a lead time of 6–30 hours have been used. For each observation, the nearest grid point of the model output has been selected for comparison.

In the postprocessing procedure three different sizes of the neighbourhood have been tested. They have been termed "small", "medium" and "large" and are listed in Table 2. Figure 4 depicts the "medium" size.

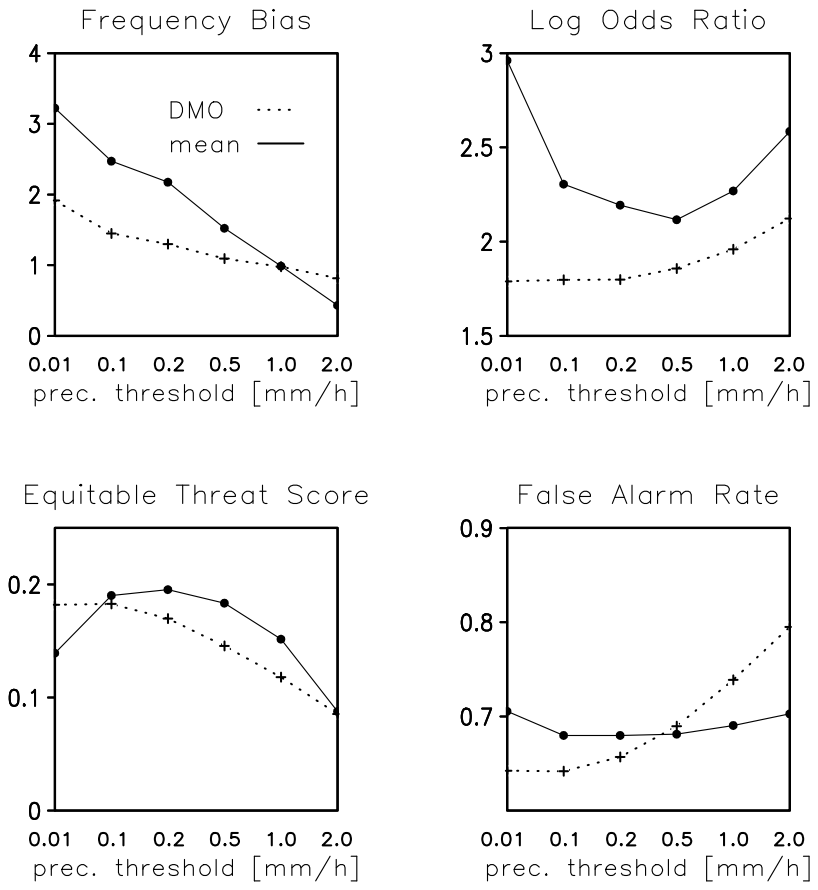

Figure 6: Measures of forecast quality derived from the contingency table for seven precipitation thresholds. The results for the direct model output (DMO) are compared to those of the arithmetic mean of the neighbourhood (mean). The mean has been calculated by applying a "large" neighbourhood at every grid point (Tab. 2).

Mean of the Neighbourhood

The arithmetic mean of the neighbourhood has been calculated at every grid point. This is in a way comparable to the ensemble mean in ensemble forecasting. The quality of the mean has been assessed by inspecting the 2×2-contingency table. The frequency bias, the equitable threat score, the false alarm rate (Wilks, 1995) and the logarithm of the odds ratio (Stephenson, 2000) have been calculated for both the direct model output and the arithmetic mean of the neighbourhood (Fig. 6). Figure 6 shows the results for the "large" neighbourhood (Tab. 2). When a "small" or "medium" size is used, the differences between the results for the direct model output and for the mean are similar, but less pronounced.

The mean is not necessarily a better forecast compared to the raw model QPF. It suffers from overforecasting low precipitation amounts and from underforecasting high precipitation amounts. This is a result of effectively smoothing the forecast. Broadly speaking, the equitable threat score and the false alarm rate show that the mean achieves both more hits and more false alarms at low precipitation amounts and it achieves less hits and less false alarms at high precipiation amounts. Such a forecast can be beneficial to some users and disadvantageous to others, depending on their ability of tolerating high false alarm rates or low hit rates.

The odds ratio indicates though, that the mean achieves a better combination of hits and false alarms than the direct model output. The differences between the log odds ratio of the mean and of the direct model output are statistically significant to the 95% confidence level for all thresholds shown in Figure 6. It is important to note that the odds ratio benefits from a large bias. When the bias is artificially removed in both the contingency tables of the mean and the direct model output, the odds ratios change considerably. After bias removal, only for thresholds 0.5–2.0 mm/h the odds ratio of the mean is still larger than the odds ratio of the direct model output.

Probability of Exceedance

For verification purposes, the probability of exceedance has been calculated, because numerous verification procedures exist in the literature and they are widely used in many applications. The probability of exceedance is the inverse of the quantile function. As the probability of exceedance is derived from the same sample as the quantiles, its evaluation also gives some indication on the quality and value of the quantiles.

Several standard verification procedures have been applied, such as calculating the brier skill score, decomposing the brier score, generating the attributes diagram (Wilks, 1995) and displaying the relative operating characteristics (Stanski et al., 1989). As to infer the usefulness of the forecast, also the relative value has been computed as a function of the user's cost-loss ratio (Richardson, 2000). This has led to the following key findings:

- Among the three postprocessing configurations tested, the one with the largest spatio-temporal neighbourhood achieved best verification results.
- The main advantage of the probability forecast consists in its variety of forecasts for different users and their different needs.
- In comparison with a climatological forecast the probability of exceedance has proven to be both skilful and valuable when the large neighbourhood is used.
- For high precipitation thresholds and cost-loss ratios near 10% the exceedance probability possesses considerably higher value than the raw model output.

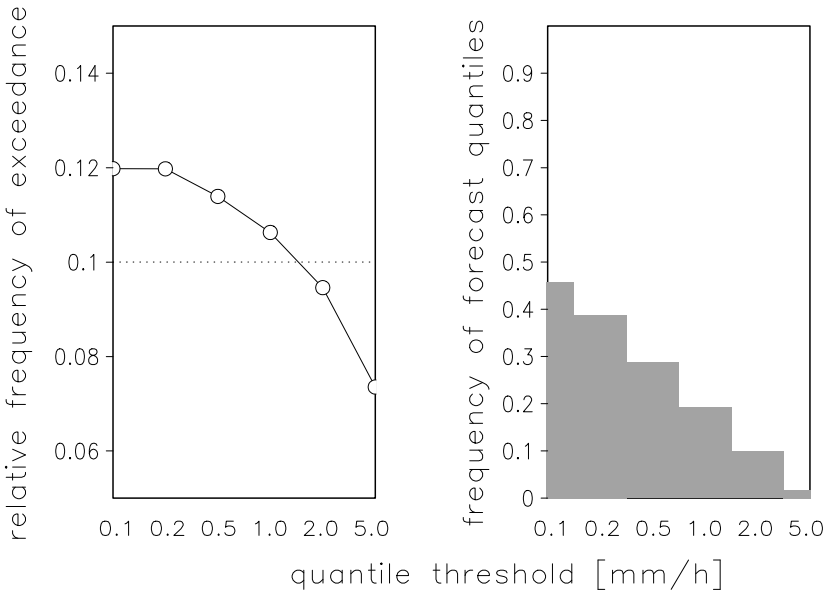

Figure 7: Left: Relative frequency of observed precipitation amounts exceeding the 90%-quantile, conditional on the forecast 90%-quantile. The 90%-quantile has been classified according to six threshold values. The dotted line represents perfect reliability. Right: The frequency of the 90%-quantile exceeding the respective threshold values. The 90%-quantile has been calculated using a "large" neighbourhood at every grid point.

- Calibration of the probability forecasts appears promising, since forecast skill and value still suffer from lack of reliability.

90%-Quantile

The 90%-Quantile has been selected as an especially interesting product of the postprocessing procedure, since it might be relevant for forecasting risk and issuing warnings. Therefore, it receives special attention in the verification.

Reliability has been tested by constructing the diagram depicted in Figure 7. Observed precipitation amounts should exceed the forecast 90%-quantile with a frequency of 10%, if the probabilistic forecasts were completely reliable. This is met fairly well, even though the frequency decays towards higher thresholds. The declining of the curve can be qualitatively explained by arguments of Bayesian statistics (Theis and Hense, 2002). It might be caused by the finite size of the underlying sample, similarly to the effect of finite sampling in ensemble prediction (Richardson, 2001).

In the future, more detailed verification will be carried out, for example estimating forecast value for decision makers. Encouraged by the verification results for the exceedance probability, positive results are expected.

4 Summary and Conclusions

The fact that numerical forecasts of precipitation are often not entirely deterministic is taken into account by (1) experimental ensemble predictions with the LM and (2) postprocessing the direct model output of a single LM simulation. While the ensemble approach has a scientifically more sound basis, the postprocessing method is designed to meet criteria of operational feasibility.

The ensemble setup includes perturbations of initial conditions, model physics and roughness length. The outcome of the experimental ensemble predictions crucially depends on the input parameters of the stochastic physics scheme. Before the ensemble prediction can serve as a guidance during the development of the statistical postprocessing method, an optimal ensemble configuration must be identified.

In the postprocessing procedure all grid points within a spatio-temporal vicinity around a central grid point are considered equally likely forecasts of the event for the central grid point. The direct model output of a single simulation is transformed into a probabilistic forecast, consisting of the mean, quantiles and probabilities of exceedance. Even without ensemble guidance in the development stage, the postprocessing procedure already produces skilful and valuable forecasts. The benefit of the postprocessed forecast is mainly contained in the probabilistic forecast, whereas the mean does not outperform the raw model output in every respect.

References

Balzer, K., 1998: Aktuelle Herausforderungen und erste Antworten. *In* K. Balzer, W. Enke and W. Wehry (eds), *Wettervorhersage*. Springer-Verlag, Berlin Heidelberg, pp. 71–92.

Buizza, R., M. Miller and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Met. Soc.*, 81, 809–819.

Davis, C. and F. Carr, 2000: Summary of the 1998 Workshop on Mesoscale Model Verification. *Bull. Amer. Meteor. Soc.*, 81, 809–819.

Du, J., S. L. Mullen and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, 125, 2427–2459.

Fritsch, J. M., R. A. Houze Jr., R. Adler, H. Bluestein, L. Bosart, J. Brown, F. Carr, C. Davis, R. H. Johnson, N. Junker, Y.-H. Kuo, S. Rutledge, J. Smith, Z. Toth, J. W. Wilson, E. Zipser and D. Zrnic, 1998: Quantitative Precipitation Forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.*, 79(2), 285–299.

Hou, D., E. Kalnay and K. K. Droegemeier, 2001: Objective Verification of the SAMEX '98 Ensemble Forecasts. Mon. Wea. Rev., 129, 73–91.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20, 409–418.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Met. Soc.*, 127, 2095–2115.

Mölders, N., 2001: On the uncertainty in mesoscale modeling caused by surface parameters. *Meteorol. Atmos. Phys.*, 76, 119–141.

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

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss situation. *Mon. Wea. Rev.*, 105, 803–816.

Murphy, A. H., 1993: What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8, 281–293.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Met. Soc.*, 127, 279–304.

Richardson, D. S., 2000: Skill and economic value of the ECMWF Ensemble Prediction System. *Quart. J. Roy. Met. Soc.*, 126, 649–668.

Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Met. Soc.*, 127, 2473–2489.

Stanski. H. R., L. J. Wilson and W. R. Burrows, 1989: Survey of common verification methods in meteorology. *World Weather Watch Technical Report No. 8*, WMO/TD No. 358, Geneva, 114pp.

Stensrud, D. J. J.-W. Bao and T. T. Warner, 2000: Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems. it Mon. Wea. Rev., 128, 2077–2107.

Stephenson, D. B., 2000: Use of the "Odds Ratio" for Diagnosing Forecast Skill. *Weather and Forecasting*, 15, 221–232.

Theis, S. and A. Hense, 2002: Weiterentwicklung der Umgebungsmethode. Bericht zum Werkvertrag *Untersuchung zur statistischen Nachbearbeitung der Bodenwetterparameter des LM. Teil IV.*

Theis. S., A. Hense, U. Damrath and V. Renner, 2002: Ensemble Prediction and Statistical Postprocessing of Weather Parameters for the LM. *COSMO Newsletter No. 2*, pp 152–161.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, San Diego, London.