11B.8

VERIFICATION OF HIGH-RESOLUTION PRECIPITATION FORECASTS OVER GERMANY

Marcus Paulat* [1], Christoph Frei [2], Martin Hagen [3] and Heini Wernli [1]

[1] University of Mainz, Germany
[2] MeteoSwiss Zürich, Switzerland
[3] DLR Oberpfaffenhofen, Germany

1 Introduction

The verification of numerical model forecasts is a key element of operational Numerical Weather Prediction (NWP) and essential for its continuous improvement. It serves to identify systematic model errors and forecast uncertainties (e.g. which parameters are difficult to predict in which regions?), and this is essential information both for the forecaster and the model developer. This study focuses on the verification of quantitative precipitation forecasts (QPF) from the "Local-Model" (LM) over Germany. The LM has been developed at the German Weather Service (DWD) and is in operational use since 1999 at both DWD and MeteoSwiss. The quality of QPF over central Europe has been identified as one of the central and problematic issues of NWP and led to the establishment of a German Priority Programme devoted to this subject (Hense et al. 2003). Research in our project aims on the one hand at identifying systematic errors of LM QPFs with standard error scores on high spatial and temporal scales and on the other hand at developing novel error measures that take into account the low predictability of the exact timing and location of small-scale convection.
Section 2 provides an overview of the datasets used in this study. In section 3 the disaggregation technique is introduced, which combines rain-gauge and Radar data in order to get a temporally high-resolution dataset of precipitation. A selection of results of a standard verification analysis for the LM is shown in section 4 and section 5 presents a summary and a short outlook.

2 The datasets

This section gives a short overview of the datasets used in this study: LM forecast data, rain gauge observations and the German radar composit. So far, all three datasets are available for the time period from January 2001 to December 2003. For these three years a preliminary verification of LM precipitation has been done (cf. section 4).

2.1 LM forecast data

The LM is a non-hydrostatic grid point model with a horizontal resolution of 7 km and 40 vertical layers based on the fully compressible dynamic equations (see Steppeler (2003) for a detailed model description). The LM domain currently covers the region of western and central Europe. Each day two operational 48-hour forecasts are started at 00 UTC and 12 UTC, respectively. From the operational runs, accumulated precipitation forecasts are available every hour. Figure 1 shows the LM orography for the region of Germany. The main orographic features are the Alps and the Alpine Foreland in the south, the Black Forest in the southwest and from there to the north the Rhine valley. Shown in blue are the North Sea (in the northwest) and the Baltic Sea (northeast).

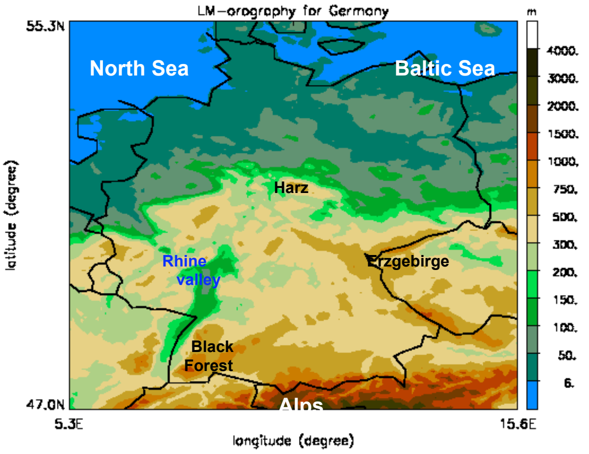

Figure 1: Orography of the LM for Germany [in m]. The highest mountains are shown in dark brown colors.

---

* *Corresponding author address:* Marcus Paulat, Institute for Atmospheric Physics, University of Mainz, Becherweg 21, 55099 Mainz, Germany. Telephone: +496131-3924152, e-mail: paulat@uni-mainz.de.

For the verification, we use LM forecasts from DWD since October 2003. For the earlier part of the time period considered (Jan 2001 till Sept 2003) DWD forecasts were no longer available and therefore data from the aLMo (the LM operational at MeteoSwiss) are used. DWD and MeteoSwiss are developing the LM in close collaboration and in essence are using the same model code. The slight differences in the setup of the two operational models (different model domains, different methods of mesoscale data assimilation) should still permit to consider the 3-year dataset as relatively consistent.

2.2 The rain gauges

A very dense network of about 4000 rain gauge stations is available over Germany with daily accumulated precipitation measurements (from 0630 to 0630 UTC). The rain gauge data are made available by the DWD. In order to compare this dataset with the LM forecasts, the rain gauge data have been interpolated on the LM-grid using the method from Frei and Schär (1998).

2.3 The radar composit

Hourly composits from the 16 operational precipitation radars over Germany are computed from 15-minute composits, made available by DWD. The radar data are originally given in six reflectivity classes, in units of dBZ. Using a standard Z-R relationship, the classes are associated with precipitation values as follows: 7 dBZ=0.1 mm, 19 dBZ=0.3 mm, 28 dBZ=0.9 mm, 37 dBZ=2.5 mm, 46 dBZ=14.0 mm and 55 dBZ=40.0 mm. The original radar composit has a horizontal resolution of 1 km and in regions with overlapping radar observations the largest value has been used when constructing the composit. For our purpose, also the radar data have been averaged onto the LM grid with a resolution of 7 km.

3 The disaggregation technique

3.1 Motivation

Summertime precipitation in central Europe associated with deep convection has a distinct daily cycle and its correct representation is a difficult challenge for current NWP models. For instance, for the LM predecessor model of DWD and MeteoSwiss it has been shown that the model produces too much rain too early in the day (around noon), in particular in elevated Alpine regions (Kaufmann et al. 2003). It is therefore important to not only verify daily-accumulated QPFs, but also the daily cycle of precipitation or, more generally, QPFs with a high temporal resolution of 1 hour. However, for such a verification, there is no observational dataset routinely available with a temporal resolution of 1 hour and a spatial resolution comparable to the LM forecasts. (In Germany there are a few rain gauge stations with hourly measurements, but their spatial density is poor compared to the LM resolution.) Therefore novel techniques have to be developed that combine, for instance, rain gauge and radar data in a meaningful way, in order to produce a high-resolution dataset that can be used for verification of high-resolution QPFs.

3.2 The technique

Here we apply a temporal disaggregation technique of 24-hour accumulated rain-gauge analyses using radar to produce hourly precipitation fields. The technique has been developed within the Mesoscale Alpine Programme for the region of the European Alps (Hagen et al. 2003). The rationale of the technique is to combine the advantages of the rain-gauge observations (dense observation network, relatively high accuracy) with the advantages of the radar composit (high spatial and temporal resolution). Figure 2 illustrates the disaggregation procedure.

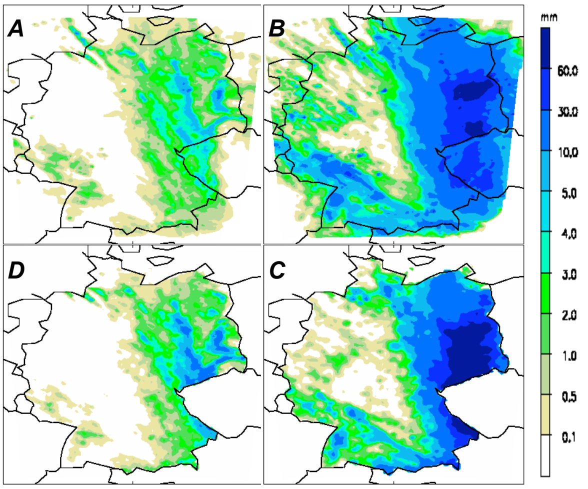

Figure 2: Precipitation over Germany for August 12th 2002 gridded on the LM grid. A) 1-h radar composit, B) 24-h accumulated radar composit, C) 24-h accumulated precipitation analysis from rain gauges, and D) is the 1-h precipitation field obtained from the disaggregation technique for the same 1-h time period as in A).

Mathematically, the disaggregation technique works as follows: at every data grid point and for every hourly accumulated radar precipitation value (Fig. 2A) its contribution to the daily total radar precipitation (Fig. 2B) has been computed. This contribution is denoted here as A/B. For every hour of the day such a ratio can be calculated at every grid point. The daily total precipitation from the rain gauge analysis is denoted as C (Fig. 2C). The final step of the disaggregation is to multiply for every grid point and for every hour of the day C with the hour's contribution to daily precipitation (A/B). The resulting hourly precipitation field (C*A/B) is shown in Fig. 2D. The temporal variability of the disaggregated hourly dataset (Fig. 2D) is consistent with the radar data, whereas the daily total of the hourly fields corresponds exactly to the rain gauge measurements (Fig. 2C).

Due to inconsistencies between rain gauge and radar data, there are three problems that can occur when applying this disaggregation technique: (i) the radar indicates no precipitation but the rain gauges measured precipitation greater than zero, (ii) the rain gauge measured no precipitation but the radar precipitation was greater than zero and (iii) both measurements are non-zero, but one is much larger than the other. Problem (i) can occur due to failures of radar operation and/or to radar shadowing in complex orography. The second problem occurs in case of spurious radar echoes from the ground, or if precipitation observed by radar does not reach the ground and/or if the spatial resolution of the rain gauge network is too coarse to capture the event. Finally, problem (iii) can be due to a combination of the points mentioned above and due to the relatively coarse categories of radar precipitation inferred from reflectivity (cf. section 2.3). For the years 2001-2003, problem (i) occurs with a frequency of 4% whereas problem (ii) is more frequent (15%). In case of (i) disaggregation is not possible and the grid point values were considered as missing data. Detailed analysis has shown that problem (i) occurs almost exclusively if the observed precipitation is less than 2 mm. Also it occurs mainly near the German border in regions that are rather remote from the radar stations. In case of problem (ii) we decided to trust the rain gauge observations and all disaggregated values are set to zero. This problem occurs mainly in preferred regions (see Fig. 3) that are either close to the radar locations (mainly ground echoes) or in the range of older radar systems (located at Frankfurt, Essen and Munich). For problem (iii) a more detailed analysis has been made (see next subsection).

3.3 Statistical analysis of the ratio (C/B)

For problem (iii) it must be decided how large a difference between daily-accumulated radar and rain gauge observations can be tolerated in order to perform a useful disaggregation. Figure 4 shows the statistical distribution of the ratio of rain gauge and radar precipitation for the four seasons. Independent of the time of year, most values are smaller than 1. The maximum in the frequency distribution occurs near 0.3 during summer and about 0.5 during the other seasons. This indicates that the radar typically overestimates precipitation by at least a factor of 2. Note that at least partially this might be due to the coarse categories of reflectivity obtained from the radar. For instance, no detailed quantification is possible for intense events between 14 and 40 mm per day, leading in some cases to significant errors in the estimated precipitation values from radar. It has been subjectively decided that all ratios between 0.2 and 2.5 can be regarded as reliable and useful for the disaggregation. For values outside this interval, no disaggregation is performed and the grid point values are considered as missing data.

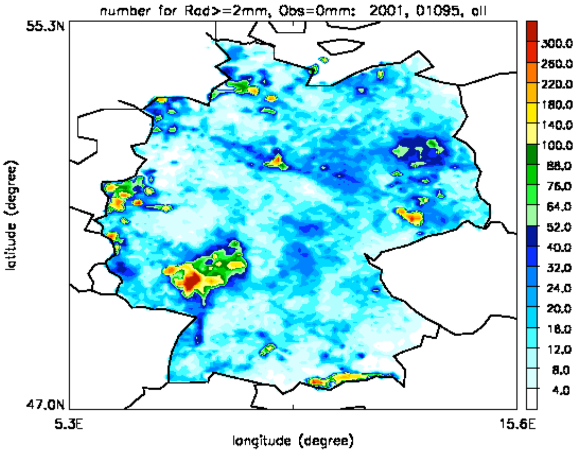

Figure 3: Frequency map for the years 2001-2003 of the situations where daily precipitation values from radar ≥ 2 mm and from rain gauge = 0 mm.

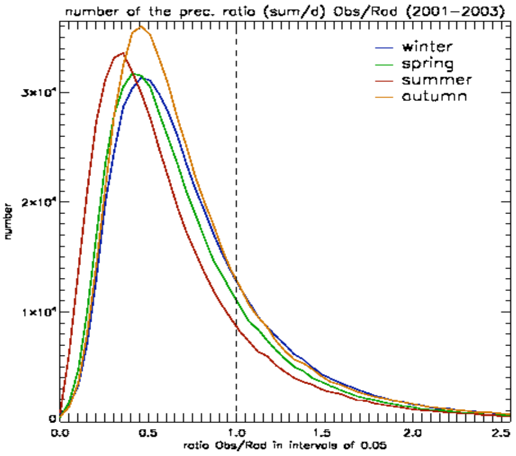

Figure 4: Statistical frequency distribution for the four seasons of the ratio between 24h-accumulated precipitation values from rain gauge observations and Radar for the years 2001-2003. Considered are only events with observed daily precipitation from rain gauges larger than 2 mm and from radar larger than 0 mm.

4 Standard verification of precipitation

The standard verification of the LM performed so far contains three parts: 1) analysis of total annual and seasonal precipitation, 2) calculation of error scores for 24-hour accumulated precipitation from the rain gauge analysis, and 3) calculation of error scores for precipitation accumulated on shorter time-scales (6 and 1h) using the disaggregated dataset introduced in section 3.

4.1 Total annual and seasonal precipitation

A straightforward way of validating model precipitation is to compare accumulated values from the model with observations. Figure 5 shows total annual precipitation averaged over the three years 2001-2003 from rain gauge observations and the LM forecasts (the considered forecast time period is from 6 to 30h from the 00 UTC runs). To compare more easily the two precipitation fields the difference field (LM - rain gauge observations) is also shown.

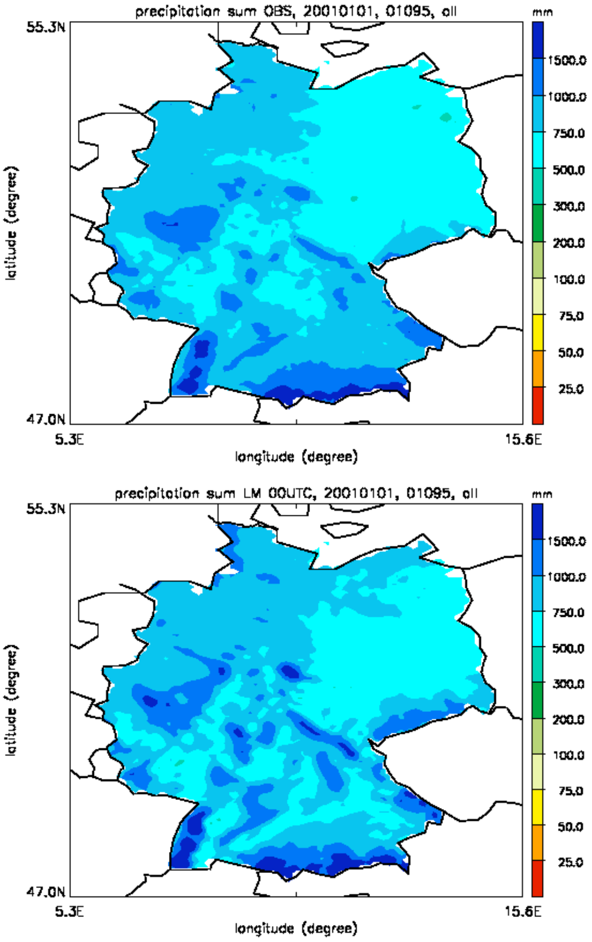

Figure 5: Total annual precipitation from rain gauge observations (top), LM forecasts (forecast time periods 6 to 30 hours from 00UTC runs, middle), averaged over the three years 2001-2003.

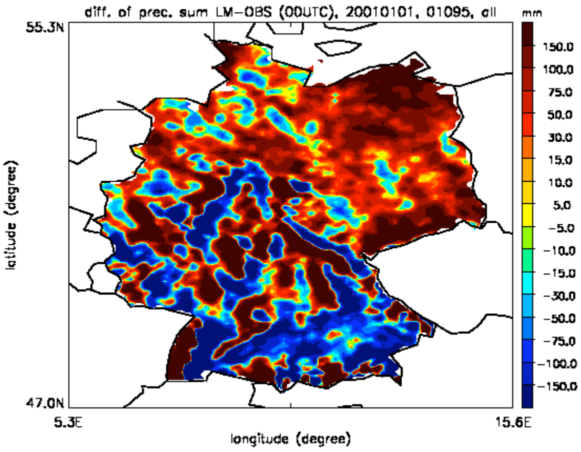

Figure 5 (continued): Difference between total annual precipitation from LM and rain gauge observations (red colors indicate that the LM overestimates annual precipitation).

Most of the precipitation maxima are located in mountainous regions, in particular in the Alps and the Black Forest region (annual mean > 1500 mm, see Fig. 1 for orography). The regions of eastern and central Germany are relatively dry (annual mean < 600 mm). Figure 5 also indicates that the annually accumulated precipitation distribution is generally well captured by the LM. It reproduces most of the local maxima, for instance in the northern Alpine foreland and the Black Forest. It captures also the drier areas in eastern and parts of central Germany. However, there are considerable quantitative deviations in some areas: in eastern Germany the LM precipitation is too high by typically 30-100 mm per year. In the central and southern part there are localized adjacent bands of large over- and underestimations, for instance in the luv and lee of the Black Forest. This luv/lee effect is a systematic error that indicates significant problems of the LM in correctly simulating precipitation near mountains of modest elevation. It occurs during all seasons (not shown) and is regarded as a major shortcoming of current QPF in Germany.

4.2 BIAS scores for daily total precipitation

Categorical error scores like for instance the BIAS score are dependent on a threshold. These scores are computed from the numbers of a so called contingency table which has four entries for hits (model and observations are above threshold), false alarms (only model is above threshold), misses (only observations are above threshold) and correct negatives (see Table 1). Several of these error scores exist (e.g. Ebert 2005) with different interpretations and generally it is better to look at more than one single score to infer the quality of a forecast model.

Contingency table	**Observation > threshold**		
		yes	*no*
Forecast > threshold	*yes*	hits	false alarms
	no	misses	correct negatives

Table 1: Contingency table which is the basis for the calculation of standard verification scores that depend on a chosen threshold.

Nevertheless, here we focus on the BIAS score, which is defined as

$$BIAS = \frac{hits + falsealarms}{hits + misses}.$$

A BIAS score > 1 indicates an overestimation and a BIAS score < 1 an underestimation of the frequency of precipitation events above the chosen threshold. Note that a perfect BIAS (=1) can result for the wrong reason, if false alarms and missing events happen to cancel each other. In dry regions/periods and for high thresholds, that is when only a few events of hits, false alarms and misses occur, the BIAS calculation might be noisy due to the small sample size. In Figure 6 the BIAS score is mapped for the three years over Germany for two different intensity thresholds of 24-hour accumulated precipitation (1.0 mm and 8.0 mm). The frequency of low threshold events is fairly well captured by the LM in most areas. The exceptions are the Black Forest in the southwest of Germany and parts of eastern Germany where over- and underestimations of 20-50% occur. For larger thresholds, the results are worse, especially in the regions with elevated orography. It clearly turns out, that prediction of intense precipitation events in the mountainous regions is a particularly difficult challenge for the LM. For example near the Black Forest there is a clear overestimation in the luv and also a clear underestimation in the lee, consistent with the general discussion in section 4.1. This effect is seen in most of the other mountainous regions in Germany (compare with Fig. 1).

Not shown here but important to note is that there is a significant seasonal variability. Summer and spring are similar with an underestimation in southeastern and an overestimation in central and northern Germany (especially for larger thresholds). Different results are found for winter (adjacent bands of over- and underestimations) and autumn (general tendency for underestimation).

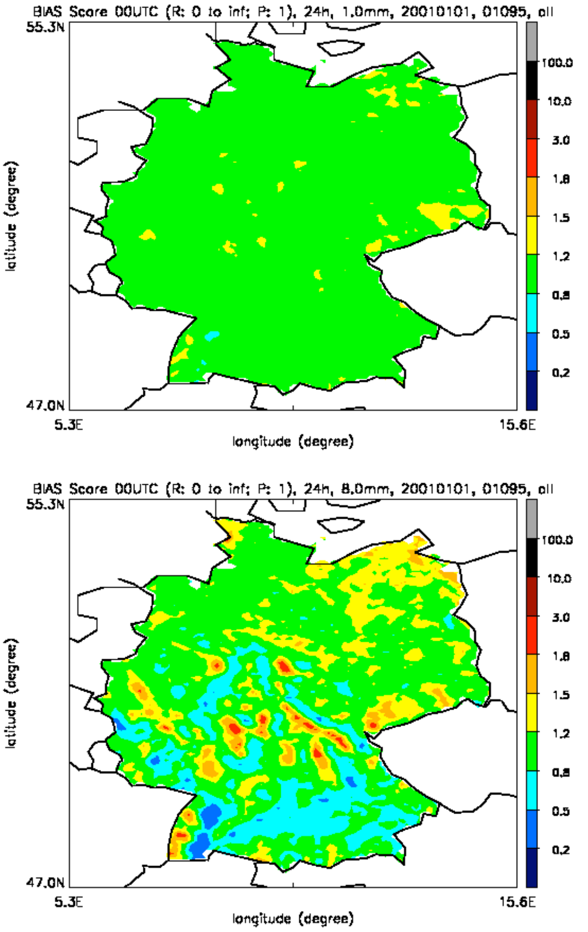

Figure 6: BIAS score for the three years 2001-2003 over Germany for different intensity thresholds: 1.0 mm (top) and 8.0 mm (bottom).

4.3 BIAS scores for shorter accumulation times

Now we investigate the BIAS score of LM precipitation forecasts accumulated over shorter time scales (1 and 6 hours), using the disaggregation dataset (cf. section 3) for summer 2002. For a threshold of 0.5 mm, Figure 7 shows that the BIAS score generally increases with decreasing accumulation time. This indicates that whilst the daily precipitation amount from the LM might be generally correct, its temporal distribution during the day has larger errors. However, due to the fixed threshold, the comparison in Figure 7 looks at moderate daily precipitation events and rather intense (and rare) hourly events and should therefore be interpreted with caution. Figure 8 shows a similar comparison but now for different thresholds such as to capture similarly frequent events. The pictures now look fairly similar, i.e. during the considered summer season there is a significant positive BIAS score in central Germany, in the northeast and the Black Forest region for daily precipitation larger than 5 mm, 6-hourly

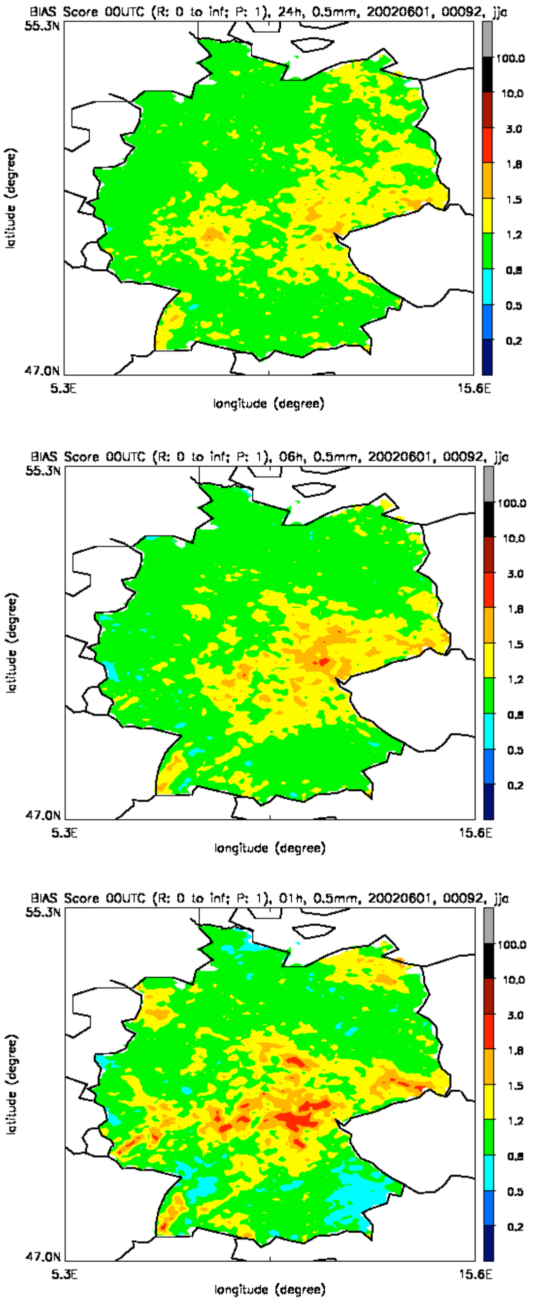

Figure 7: BIAS score over Germany for different accumulation times (top: 24 h, middle: 6 h, bottom: 1 h) and the identical intensity threshold of 0.5 mm.

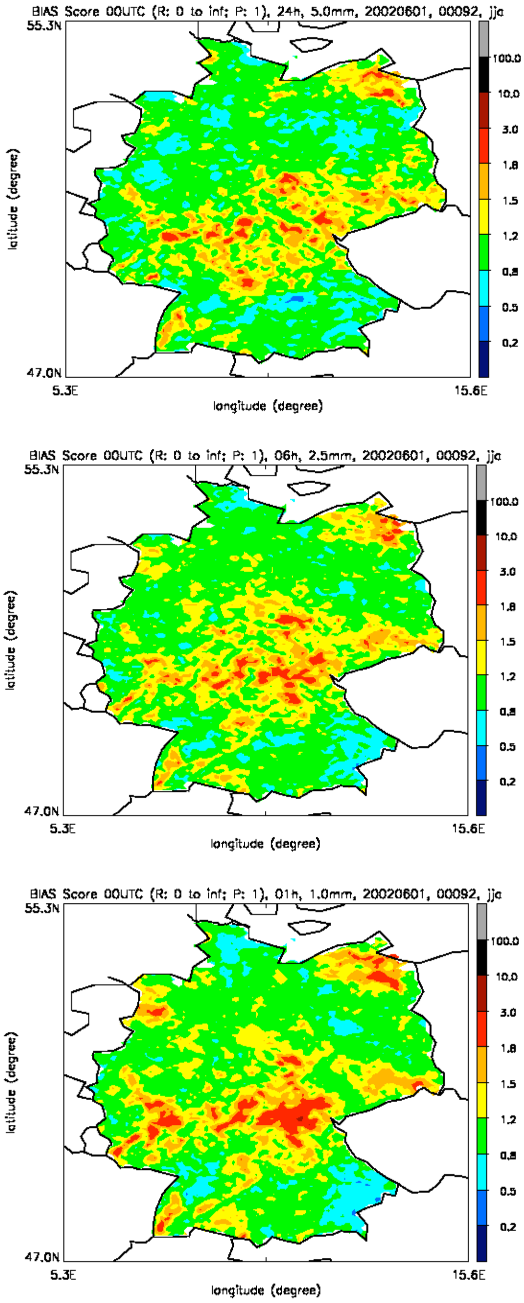

Figure 8: As Figure 5 but for different intensity thresholds: 5 mm within 24 h (top), 2.5 mm within 6 h (middle), 1 mm within 1 h (bottom).

precipitation larger than 2.5 mm and hourly precipitation larger than 1 mm. In summary, the BIAS score gets worse for increasing thresholds and decreasing accumulation time periods.

5 Summary and Outlook

In this study, preliminary verification of the precipitation forecast of the "Local-Model" (LM) from the German Weather Service (DWD) has been performed. To verify the model forecasts on the hourly time-scale, a novel observational dataset for precipitation over Germany has been computed with a time resolution of one hour. To this end a disaggregation technique introduced by Hagen et al. (2003) has been used, that combines 24-hour accumulated rain gauge data with radar information. It is shown that the method works well over Germany due to a dense network of rain gauge stations and an almost complete coverage with radars. Technical problems due to reduced radar visibility only occur in some border areas of Germany and in a few regions with elevated orography. Problems with ground echoes are confined to the close environment of the radar stations and to larger regions surrounding a few older radar systems.

This dataset offers exciting new possibilities for the analysis of precipitation characteristics in Germany and for the verification of high-resolution QPFs. It will be possible to investigate in detail the daily cycle of precipitation both in observations and the model and to identify specific problems associated with the onset and propagation of precipitation events in the LM forecasts.

So far, for a time period of three years (2001-2003) conventional verification of the LM has been performed and the BIAS score has been used to illustrate some of the basic problems of the model QPFs. It is shown that intense (presumably convective) events lead to significant model BIAS, with a non-trivial pattern of over- and underestimation. A particular “luv-lee problem” occurs during all seasons in southern Germany. To the west of the Alpine foreland mountain chains (e.g. Black Forest) precipitation is strongly overestimated whereas to the east there is a clear tendency for underestimation. This indicates a problem with the initiation of orographic precipitation and further research is required to identify the dominant flow patterns and meteorological conditions during the erroneous forecast events.

Also as an outlook it is mentioned that we start to develop and apply novel error scores that will be grid-point independent and particularly useful for short accumulation times. On the time scale of one hour there is very limited predictability for the exact location of single convective events. Using standard scores this inevitably leads to a poor model performance although the general QPF pattern might indicate the “right kind of weather”. Therefore, the novel error scores will focus on the location-independent intensity and shape distribution within river catchments.

Acknowledgements

The authors are grateful for the funding of the present study as part of the Priority Programme SPP 1167 of the German Science Foundation (DFG). We are also grateful to the DWD for providing access to the rain gauge and radar data and to the DWD and MeteoSwiss for the LM forecasts.

References

Ebert, B. et al., 2005: Forecast Verification – Issues, Methods and FAQ. http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

Frei, C. and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. Int. J. Climatol., **18**, 873-900.

Hagen, M., C. Frei and C. Schär, 2003: Temporal disaggregation of rain-gauge analyses using radar. Int. Conf. Alpine Meteorol. and MAP-meeting, Brig, Switzerland.

Hense, A., G. Adrian, C. Kottmeier, C. Simmer and V. Wulfmeyer, 2003. Initial proposal of the Priority Program 1167 from the Deutsche Forschungsgesellschaft (DFG).

Kaufmann, P., F. Schubiger and P. Binder, 2003: Precipitation forecasting by a mesoscale numerical weather prediciton (NWP) model: eight years of experience. HESS, **7(6)**, 812-832.

Steppeler, J., G. Doms, U. Schättler, H. W. Bitzer, A. Gassmann, U. Damrath and G. Gregoric, 2003 : Meso-gamma scale forecasts using the nonhydrostatic model LM. Meteorol. Atmos. Phys., **82**, 75-96.