Strategy for the long-term evaluation of the model output from the Lokal Modell

By performing an evaluation of the Lokal Modell (LM) for a period of several months (long-term evaluation; LTE), systematic biases in this model can be identified. As a first step, case studies will be performed in order to see whether there are problems that can be solved easily in the model and to help us with developing a strategy for the LTE. After this an evaluation for a period of 1 week in autumn 2004 will be performed. We need to finish a detailed analysis of these case studies and the 1-week period before the DFG-Colloquium of expertise on 10-11 March 2005. At this meeting we have to show that we are very close to starting the three-month evaluation and that it is feasible to finish that within the time scale of the project.

For case studies with precipitation we have selected:

19 September 2001:	case with frontal precipitation in the Netherlands
8 July 2004:	case with strong precipitation over southern Germany
12 august 2004:	case with strong thunderstorm over southern Germany

In addition, of relevance are the two BBC (Baltex Bridge Campaign) cases selected for the WMO workshop.

23 September 2001:	low-level clouds in the morning transferring into shallow cumulus in the early afternoon in the Netherlands
21 May 2003:	two cloud layers during daytime with shallow cumulus and stratocumulus in the Netherlands

When the AQUARADAR proposal gets funded (this will be known at the beginning of 2005), it would be most suitable to do the long 3-month integration for the period when AQUARAD is operational (May-July 2005). This gives us only a short period to do the evaluation (8 months), but the AQUARAD will give such a wealth of information on the vertical distribution of the hydrometeors, which enables us to evaluate the model in more detail. When AQUARAD does not get funded, it is better to do the LTE for the summer of 2004, since this gives us more time to analyse the results.

What are we looking for by doing the long-term evaluation of forecasts of the Lokal Modell?

- What are the systematic errors in the precipitation forecasts of the Lokal Modell?
- What are the typical conditions in which these systematic errors can be most clearly detected?
- How well do the models need to represent the clouds in order to get a good forecast of precipitation; having found situations where the precipitation is wrong, we want to look for correlated errors in cloud and water vapour to see if we can trace the precipitation errors back through the water cycle.

How do we set-up the model for the case studies, evaluation of 1 week and LTE?

Region

A relative large area will be chosen including Lindenberg and Cabauw (see figure for the preferred model grid). A grid spacing of 2.8km will be used, which is identical to the resolution of LMK. The reason for using this resolution is that the model resolves the process of deep convection and that therefore a parameterisation of deep convection is not needed in the model. In addition, radar data form an important source for the evaluation of the model. On a resolution coarser than 2.8km, many of the features detected by the radar cannot be resolved. If it turns out to be unfeasible to do integrations including the Alps due to computational costs, the southern boundary can be shifted northwards. The model domain will be about $(1000km)^2$ which is $(330 \text{ grid points})^{2,}$ which is computationally very expensive and after trials it has to be decided whether this is feasible. If such a large domain is not feasible, a possible solution would be to do integrations for sub domains around Cabauw, Lindenberg and Southern Germany, but using a larger domain is preferred since the lateral boundaries are located further away from the region of interest, and the model output is more dependent on the model physics, and less dependent on what is prescribed at the lateral boundaries than for the smaller domain. In addition, the region used for the operational forecast is also relatively large. A third arguments it that, for the evaluation with satellite data, the large domain is certainly more interesting and statistical test on the 2-dimensional fields are more robust.

Spin-up times
Especially for convective precipitation, the model is dependent on the initialisation time. The convective precipitation has a clear daily cycle. From literature, it is known that there are cases where the model can only represent the convective precipitation properly when it was initialised at 18UTC on the day before or at 6UTC on the same day, but not when it was initialised at 00UTC. The reason is that, for this specific case, initialisation at 00UTC does not give the model enough time to get the conditions right, which are relevant for the convective precipitation, and neither are these conditions correctly represented in the initial conditions, as is the case for the 06UTC initialisation. For the LTE, we will start the integrations at 18UTC and perform a 36hr (30hr) forecast from which the first 12 hours (6 hours) will be disregarded. We will start a forecast once a day, so that the entire 3-month period is covered by forecasts from T_0+12 until T_0+36. If it turns out from literature study or from the case studies that this is not a good strategy, the following strategy can be adopted, to make our results less dependent on the start time of the integration:
For a 90-day period, 120 forecasts of 30 hours will be performed, where T_0 varies according to the following equation as a function of the forecast number (i):
$T_{0,i} = (i-1)\cdot 18$,
where $T_{0,i}$ is given in hours. The full 90-day period then consists of the 120 forecasts from T_0+12 until T_0+30.

Lateral boundary conditions
Lateral boundary conditions from the Global-Modell (GME) of the Deutsche Wetterdienst will be used for the LTE since we want our integrations to be as close as possible to the operational forecasts from LM.

Initial conditions
For the same reason as described above, initial conditions will be used from GME.

Radiation
Radiation is calculated every hour (as in the operational setting), but the sensitivety for this will also be tested for the selected cases.

Organisation of the model output
Model output will be organised in the following way:
- Time series of model prognostic variables and important diagnostic variables will be stored at every time step at the location of selected stations: Cabauw, Lindenberg, München, Oberpfaffenhofen, AQUARAD network.
- Standard fields (similar to fields selected by DWD) will be stored in the GRIB format at every hour at all grid boxes. All fields needed for SynPolRad will be stored in this format. If storage of these fields is a problem, sub regions around Cabauw, Lindenberg, and Munich /Oberpfaffenhofen, AQUARAD region, will be selected and data will only be stored for those sub regions.
- Rain rate at 1km height will be stored every 5 minutes to compare with the rain radar.
- Cloud cover, integrated water vapour and liquid water path (and preferably radiance at infrared channels and cloud optical thickness) will be stored every 15 minutes to facilitate comparison with Meteosat-8´s (MSG), which has this high temporal resolution.
- Area averages of the important hydrometeors will be stored every 15 minutes, this is important for the sensitivity runs for the cases only and not so much for the LTE

How do we treat representativeness of spatial and temporal scales?

Satellite remote sensing products
Since the resolution of the satellite products is higher than the resolution of the model output, we can use the satellite data to get an indication of the sub-grid scale variability. We will investigate this variability by creating resolution dependent histograms. Subsequently, the satellite data will be interpolated on the model grid. We will calculate the average value over several grid boxes to see which spatial scales the model correctly resolves. We will start with averaging over an area of 2x2 grid boxes, then we continue with averaging over an area of nxn grid boxes where n increases from 3 to 10. This intercomparison is done at the overpass time of the satellite. From Meteosat-8 we have measurements every 15 min and will be able to study the diurnal cycle.

Polarimetric Radar
Radar measurements are done and stored in polar coordinates, which makes it almost impossible to perform statistics on these data. SynPolRad starts on the model (hence Cartesian coordinates) to calculate the electromagnetic interactions of the radar beam with the hydrometeors and transfers afterwards the modelled reflectivities to polar coordinates in order to get the beam propagation and attenuation included. We decided to use an already existing algorithm to transfer, respectively retransfer, the measured and the synthetic radar images from the polar coordinates to a Cartesian grid for further evaluation. The dimension of this grid can then be chosen freely but a sensible choice would be something in the order of the LMK grid. As for the evaluation, we will start with histograms of reflectivity and the area of precipitation. The use of the polarimetric variables depends on the availability of the LMK version including graupel. Further statistics will be done on the classification scheme comparing for example ratios of rain and graupel or the appearance of the bright band. Averaging over areas respectively volumes will be performed and radar data will be compared with the satellite remote sensing data.

Upward looking ground based remote sensing data and the surface-based in-situ data
The upward looking ground based remote sensing data and the surface-based in-situ data have to be treated in the same way. An advection time scale (t_a) needs to be calculated for comparison of model output with these data. This advection time scale is the time that is needed for an air parcel to move through a grid box. A complication of this method is that a wind shear exists; the wind speed and direction are a function of height above the surface. As a first guess we take the wind speed at 1 km height to calculate t_a, since this is the level with high moisture and liquid water content (and therefore most relevant for the variables that we would like to evaluate). Once we have calculated t_a, we use t_a to average the high-resolution time series in time, by using a running mean window. Note that the t_a and thus the size of the running mean window are varying in time. In this way, we have interpolated the observed time series at the site to the model grid. Sub grid scale variability can be studied by comparing the time-averaged observed time series with the raw observed time series. A similar approach as described for the satellite remote sensing data will be used. The running means over an nxt_a time window will be calculated where n ranges from 2 to 10, to see whether there is any spatial scale at which the model represents clouds and precipitation better than at the spatial scale of the grid boxes.

Radiosondes
Radiosondes cannot be used to address the problems related to the spatial scale. The only way in which a comparison between model output and radiosonde measurements can be made is by extracting the information from the model column at the time of radiosonde launch.

List of instruments to compare the model output with:

Radar
Compare the radar images with the products of the forward operator tool

- DLR polarimetric radar
- DWD weather radars

Satellite remote sensing
- IR radiances (use the forward operator tool to calculate the IR radiance from the model output; Christian Keil has experience with using this tool)
- Cloud mask
- Integrated water vapour for cloud free images in both model and reality (the IWV below the cloud is not seen by the satellite and when clouds occur either in reality or in the model, these grid boxes need to be excluded)
- Liquid water.
- Cloud optical thickness

In-situ
- Rain gauge data: these can be obtained from Deutsche Wetterdienst or the Local Water Authorities (Birgitte Fuchs)
- Radiosonde soundings

Ground based remote sensing:
- Microwave radiometer
- Cloud radar
- Micro rain radar (MRR)
- Wind profiles
- Lidar

Important deadlines

Mid November 2004:	output from LM for selected cases will be ready for radar forward operator and comparison with satellite measurements
Mid February 2004:	meeting with QUEST participants to discuss the progress, make a plan for the meeting in March and make a plan for writing the next proposal
10/11 March 2005:	DFG-Colloquium of expertise at the Center of Physics in Bad Honnef
End March 2005:	make a detailed plan for using AQUARAD in the LTE of QUEST
May-Sept 2005:	AQUARAD operational (?): this should also be the time period for the long-term evaluation
September 2005:	proposal for year 3-4 needs to be finished

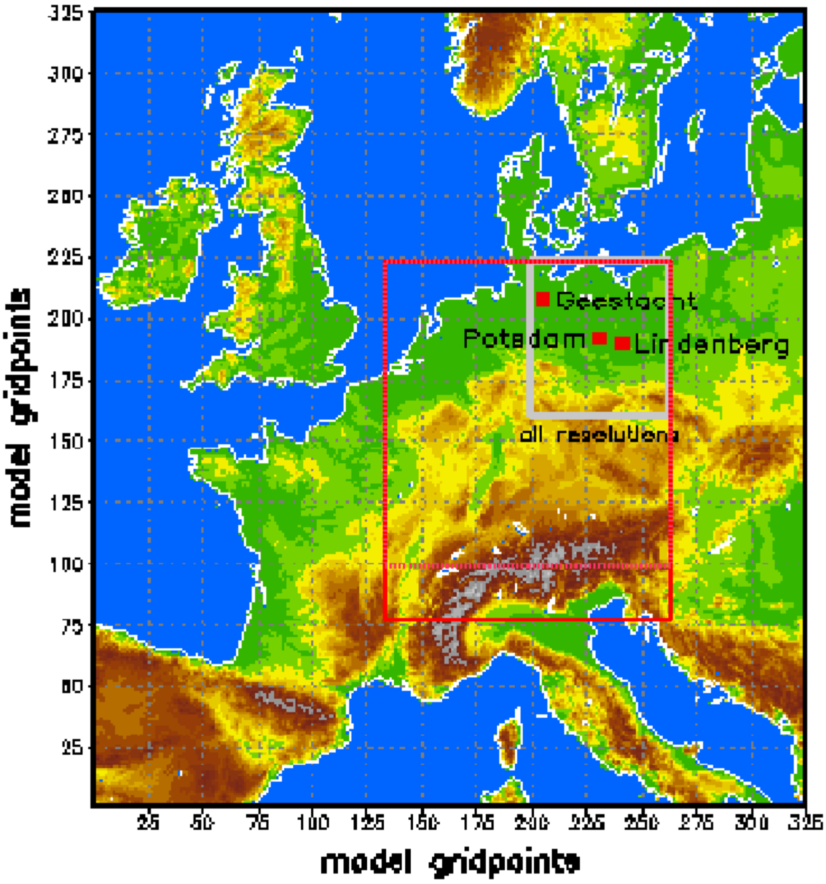

Figure 1: *The proposed model domain. Solid red line, preferred model grid (910 km x 1050 km = 303 x 350 grid points). Dashed red line: possible model grid (910 km x 910 km = 303 x 303 grid points), which is somewhat less computationally expensive.*