An Objective Quality Measure based on a Pattern Recognition Technique to validate Regional Ensemble Forecasts

Christian Keil, George Craig

Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Wessling, Germany

1 Introduction

The overall objective is an investigation into the predictability of convective storms, with the aim of determining the relative importance of several different sources of uncertainty. Such knowledge is needed for the design of an ensemble forecasting system, and eventually of an appropriate observing network. The work consists of ensemble simulations with differing sources of variability and development of objective methods of assessing accuracy of individual forecasts. Errors in regional forecasts often take the form of phase errors, where a forecasted weather system is displaced in space or time. For such errors, a direct measure of the displacement is likely to be more valuable than traditional measures, such as RMS error. A displacement measure is developed with a view to using it to explore the relative importance of various sources of uncertainty in regional ensemble forecasts.

2 The Regional Ensemble System

In general, a limited area model is influenced by the following sources of uncertainty:

(1) boundary conditions (uncertainty in the synoptic and meso-scale environment provided by a global model),

(2) initial conditions (uncertainty due to structures not seen by the observing system or due to limited resolution) and

(3) physical parameterizations (uncertainty resulting from the model formulation of convection, cloud microphysical, planetary boundary layer, or other processes).

The most obvious way to account for boundary condition uncertainty is to use a set of boundary conditions generated by a global ensemble forecasting system. Following the limited-area ensemble prediction system COSMO-LEPS methodology (Molteni et al., 2001; Marsigli et al., 2001), the high-resolution non-hydrostatic Lokal-Modell (LM) (Steppeler et al., 2003; Doms and Schättler, 2002) is nested on selected members of the global ECMWF EPS. It was found in this system that most of the variability in the 51 member ECMWF EPS for a region centered on the European Alps can be retained by as few as ten members (Marsigli et al., 2005). Here, 51 ECMWF EPS T255L40 ensemble members (Init. 2002070712 +72h fc) were down-scaled by a cluster analysis into 10 classes, for which LM experiments (Version 3.12) with 7 km horizontal resolution were conducted.

Secondly, using the new LM module LMSynSat (available from version 3.12 onwards) allows the production of synthetic satellite imagery. The synthetic satellite imagery is generated using the fast radiative transfer model for TIROS Operational Vertical Sounder (RTTOV-7), that allows fast simulation of brightness temperatures for various satellite radiometers

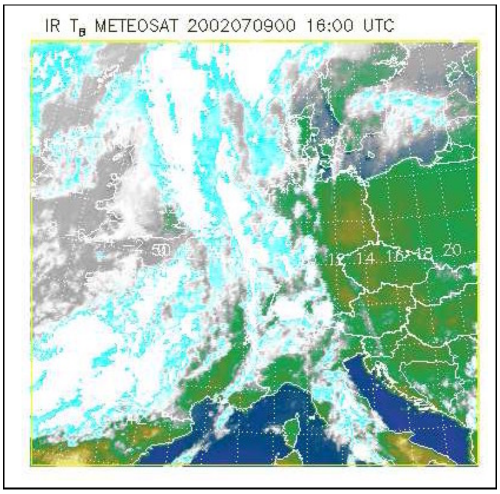

Figure 1: Observed Meteosat 7 IR imagery at 16:00 UTC 9 July 2002.

(e.g. Meteosat 7 MVIRI and Meteosat 8 SEVIRI). The input variables provided by LM are atmospheric profiles of temperature, specific humidity, cloud properties (cloud cover, cloud liquid water, cloud ice), specific content of snow and rain, and surface properties (skin temperature, temperature and specific humidity at 2m, land-sea mask). The output variables are clear and cloudy-sky radiance and brightness temperatures in IR and WV of Meteosat 7 and 8 channels of Meteosat 8. Sensitivity studies showed that a more realistic representation of clouds in LM can be achieved using the prognostic precipitation scheme (incl. precipitating snow) and a modified critical ice-mixing ratio (Keil et al., 2005).

Thirdly, using the model-forecast and the observed satellite image a field of displacement vectors is computed which 'morphs' the simulated image into a best match of the observed image. The magnitude of the mean displacement vector and the quality of the final match measured by the correlation give objective measures of the quality of the forecast.

After implementation, a case study observed during the VERTIKATOR field campaign (Vertical Transport and Orography, Lugauer and co-authors (2003)) in the northern Alpine forelands on 9 July 2002 has been examined.

3 Results and discussion

Ahead of an eastward propagating cold front, pre-frontal convection developed in the northern Alpine region in the afternoon of 9 July 2002. The cloud signature of a convective cell that has been initiated two hours before in the northern Alps is clearly visible in the IR image across southern Bavaria at 16:00 UTC (Fig. 1). The elongated cloud band across eastern France marks the cold front.

Model-forecast synthetic IR images of each representative member of the 10 clusters, the ensemble mean, and its spread are displayed for a subdomain in Fig. 2. While most of the clusters capture the synoptic scale cloud pattern (outside the subdomain, not shown) there are large differences in the pre-frontal convection and the position of the cold front at 16:00 UTC. Visual intercomparison of the observed and synthetic IR images gives a first, subjective ranking of the realism of the different clusters. The mean ranking based on a subjective evaluation by 8 scientists is given in Table 1. The top scoring of clusters 2, 10 and 4, reproducing the convection and the corresponding cloud signature in this region, can be

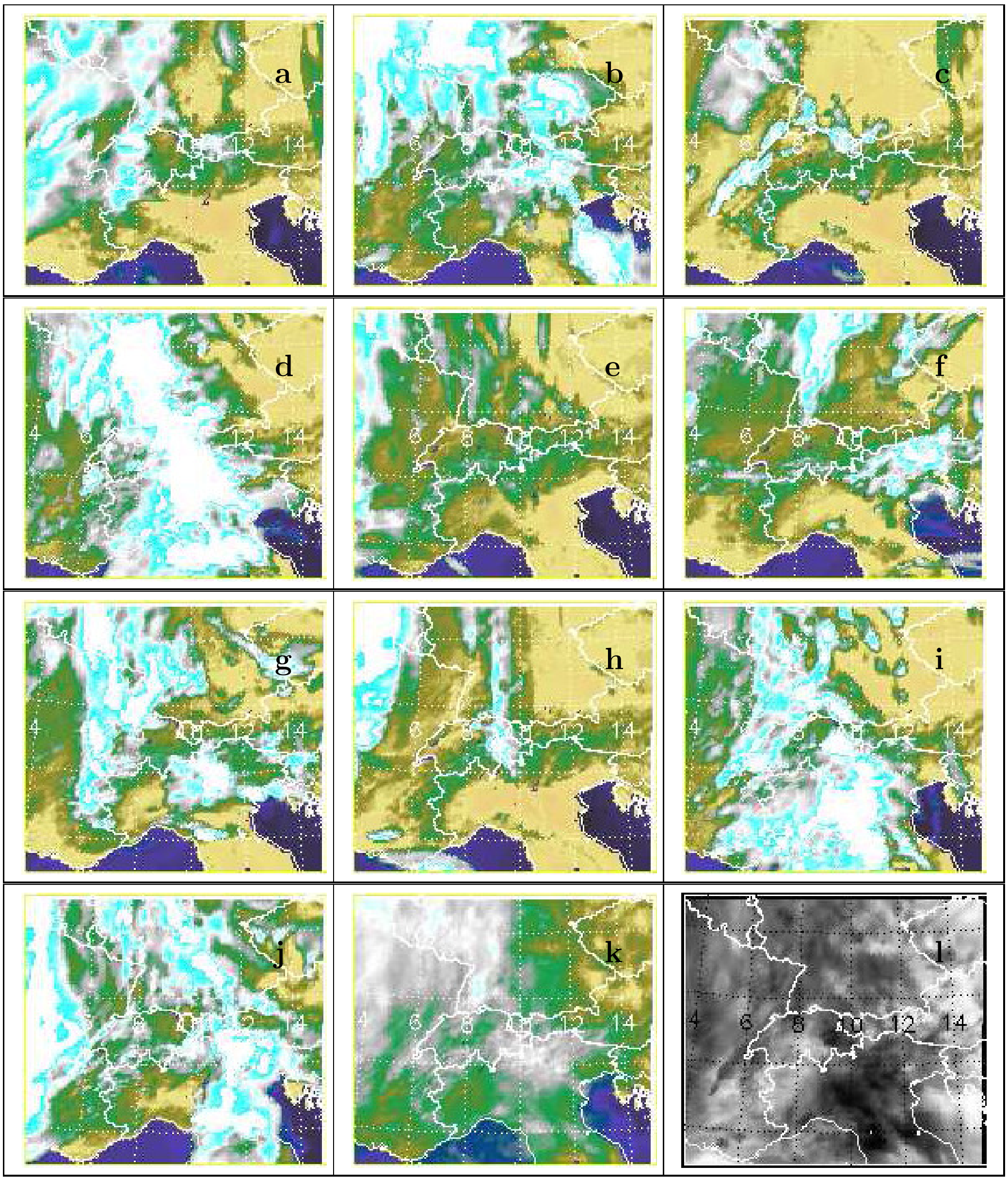

Figure 2: Forecast IR synthetic satellite imagery of the representative members of each cluster on 9 July 2002 16:00 UTC (LM +52h forecast range): (a-j) the individual ensemble members ranging from 1 to 10, (k) the ensemble mean and (l) ensemble spread (dark colored areas denote large spread).

confirmed in Fig. 2b,j,d. Likewise the low scoring of cluster 3 is evidently due to the severe underprediction of clouds in that area. The rank correlation between the scientists is quite high (0.85) confirming a good agreement among themselves and pointing towards a clear ranking of the clusters.

Next, the Pyramidal Image Matching technique (Mannstein, personnel communication) is applied to weight the different clusters according to their correspondence to the observed satellite imagery. In essence, differently coarse-grained pixel elements are compared. Starting at the largest scale (one pixel element containing 8×8 LM grid cells), a displacement vector field that minimizes the total squared error in brightness temperature within the range of ± 2 pixel elements, that is within a range of about 250 km, is computed. Subsequently, this image processing is done at successively finer scales (pyramidal). Finally a displacement

Table 1: Rankings of the 10 clusters according to (i) subjective eyeball evaluation of 8 scientists, (ii) the cluster population (number of members per cluster) and (iii) the objectively calculated forecast quality index.

Rank	1	2	3	4	5	6	7	8	9	10	rank corr.
subjective	2	10	4	7	9	1	5	6	8	3	0.85
population	3	2	4	1	5	7	6	8	9	10	-0.29
FQI	7	2	9	4	10	1	8	6	3	5	0.77

vector for every pixel is obtained from the sum over all scales.

However, the mean vector length of the displacement vector field contains no one-to-one information of the forecast quality. For instance, imaging a forecast showing no (or very few) cloud features at all (subjectively a forecast failure, e.g. cluster 3) would result in a mean displacement vector equal (or close to) zero. On the other hand, a perfect forecast would result to zero as well. Thus, a quality measure is constructed containing different measures of quality: the objectively computed mean displacement *displ*, the ratio of forecast and observed cloud occurrence CC_{LM}/CC_{Sat} (below a threshold brightness temperature), and the spatial correlation of observed and forecast-matched cloud structures *corr*. This measure is the normalized forecast quality index FQI, attaining zero for a *perfect* forecast:

$$FQI = 0.33 * [displ + (1 - CC_{LM}/CC_{Sat})_{+} + (1 - corr)].$$

Application of the Pyramidal Image Matcher on the cloud pattern at 16:00 UTC allows an objective ranking of each cluster that is shown in Table 1, too. Comparison of the subjective ranking and the one obtained from the object-oriented algorithm shows that the image matching provides a reasonable error measure for phase errors: the subjectively top-scored clusters are within the top five of the objective technique, while the lowest scored cluster agree reasonably well, too. The rank correlation between the average subjective and the objective ranking attains 0.77, confirming the consistent results of both rankings. In contrast, the ranking based on cluster population (number of members per cluster) shows no correlation with the other rankings (-0.29).

At 16:00 UTC, i.e. after +52 h forecast time, the variance of clouds is not only confined to the pre-frontal and frontal regions. Instead the ensemble spread shown in Fig. 2l shows a considerable variance of brightness temperatures in large areas of the domain. Due to the long forecast range, there is considerable noise in the forecast. The persistence of skill of individual members can be assessed by computing the rank correlation with different lead times. For this episode the persistence of skill is about 12 hours owing to a change of weather regime in the region. This new synoptic scale weather system moving into the region from 00:00 UTC onwards developed to the violent *Berlin* storm on 10 July 2002 (Gatzen, 2004).

4 Summary

The Regional Ensemble System currently developed at DLR consists of the following main components: (i) the COSMO-LEPS system, (ii) forward operators to generate synthetic satellite imagery based on model fields, and (iii) a pattern recognition algorithm to measure the quality objectively.

Validating the ensemble output of different episodes of pre-frontal summertime convection in Bavaria using the novel forecast quality measure FQI leads to the following conclusions:

- Pyramidal image matching provides a plausible measure of forecast error, which is consistent with subjective rankings.
- COSMO-LEPS cluster populations are a poor indicator of local skill.
- Persistence of skill is about 12 hours owing to change of weather regime in region.

In future, additional case studies from the Schwarzwald (moderate orography) and the southern UK will be simulated, to explore the performance of the system in different predictability regimes. Next to satellite observations, radar data will be utilized to validate the ensemble output using the radar forward operator developed by Pfeifer et al. (2004). An opportunity has arisen to implement a stochastic convective parameterisation in the LM (Craig et al., 2005), and if initial tests are successful, this will be used to compare its contribution to ensemble spread to that of the EPS boundary conditions.

Acknowledgements

Meteosat data are copyrighted by EUMETSAT. The authors thank Andrea Montani (ARPA-SIM) for his guidance to run COSMO-LEPS.

References

Craig, G.C., B.G. Cohen and R.S. Plant, 2005: Statistical mechanics and stochastic convective parameterisation. In: Workshop on Representation of sub-grid processes using stochastic dynamic models, ECMWF, Reading, England.

Doms, G. and U. Schättler, 2002: A Description of the Nonhydrostatic Regional Model LM. Deutscher Wetterdienst, Offenbach am Main.

Gatzen, C., 2004: A Derecho in Europe: Berlin, 10 July 2002, *Weather and Forecasting*, 19, pp. 639–645.

Keil, C. A. Tafferner, and T. Reinhardt, 2005: Synthetic satellite imagery in the Lokal-Modell, *Atmos. Res.*, accepted.

Lugauer, M. and co-authors, 2003: An overview of the VERTIKATOR project and results of Alpine Pumping, In: *Proceed. Int. Conf. Alpine Meteorol.*, Brig, Switzerland, pp. 129–132.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlinear Processes in Geophysics*, 12, pp. 527-536.

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

Molteni, F. R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. Part I: definition of representative members and global-model experiments, *Quart. J. Roy. Meteor. Soc.*, 127, pp. 2069–2094.

Pfeifer, M., G. Craig, M. Hagen and C. Keil, 2004: A polarimetric radar forward operator, In: *Proceed. Third European Conference on Radar in Meteorology and Hydrology (ERAD)*, Visby, Sweden, pp. 494–498.

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, pp. 75-96.