Working Group 2 of COST-717: Using radar observations in parameterization and validation of atmospheric models

Minutes of Meeting held on 15 September, 2001 at Civil Engineering Department,

University College Dublin

Clive Wilson, Met Office, UK

The agenda was accepted as follows:

Agenda of the WG2 Meeting

----

1. Welcome & introduction of new members

2. Progress Reports WA3: Validation of models and parameterisation schemes
   - Carl Fortelius

   Verification of precipitation predicted by HIRLAM using BALTRAD data.
   - Clive Wilson

   Further verification of precipitation forecasts
   - Daniela Rezacova

   Application of LM to several cases of severe convection - present state of the work

3. Progress Reports WA2: Approaches to verification
   - Christian Keil

   Use of satellite data for quality control of LM simulations
   - Clive Wilson

   Final report on verification methods

4. Progress Reports WA1: Radar data
   - Dieter Fruehwald

   Poster of Elmar Weigl: Real-time adjustment scheme of radar precipitation data at DWD
   - Joan Bech

   Estimation of propagation conditions
   - Sergio Barbosa

   Use of Kalman filter in radar raingauge adjustment
   - Uta Gjertsen

   Radar precipitation estimates - towards a gauge adjustment method
   - Zbynek Sokol

   Estimation of daily area precipitation using radar and gauge data
   - Marco Gabella

   Real-time estimation of daily precipitation with radar during the Oct 2000 Piedmont flood.
   - Stanislaw Moskowicz

   Bayesian approach for merging radar and gauge rainfall data and its application for model rainfall verification

5. Discussion of work planned over the next 6 months

6. Matters to be raised at MCM & STSMs

7. Any other business:

- Manpower estimates
- Material for Technical Committee Met.
- Website

- Attending: Sergio Barbosa , Joan Bech , Carl Fortelius, Dieter Fruehwald, Marco Gabella, Uta Gjertsen, Elisabeth Gerard, Christian Keil ,Vassiliki Kotroni, Peter Meischner, Silas Michaelides, Daniel Michelson, Stanislaw Moskowicz, Daniela Rezacova, Zbynek Sokol, Clive Wilson, Jaques van Gorp

The Chairman, Dieter Fruehwald, welcomed members. He and the group paid their respects in memory of Hugo Ottoy whose untimely death had deprived the group of his valuable contributions.

Progress Reports

Daniel Michelson proposed taking the progress reports in reverse Working Area order in view of the truncated time (the session planned for the previous day had been cancelled because of the Day of Mourning following the terrorist attacks in the USA on Tuesday 11 September). This would enable WA3 & 2 to be given more attention as WA1 was usually covered first.

2 WA 3 Validation of Models and Parametrisation schemes

Carl Fortelius showed results from extending his study comparing Hirlam model forecasts with the gauge adjusted BALTRAD network.

He has now taken the reference Hirlam forecasts run in delayed mode at ECMWF and compared to BALTRAD data. The model is rather coarse (0.5 degrees) in order to cover the large domain (Canada to Africa, south N. Atlantic to Siberia) but includes full data assimilation. To minimise problems with the radar data only the Finnish network was used, for which earlier studies showed to have smaller errors because of relatively flat terrain and homogeneous radars. 24-h daily precipitation forecasts from 00Z for T+18 to +42 h were compared to radar estimates of total precipitation from June 2000 to present. The model forecasts were obtained from 90 model grid boxes; the radar estimates were areally averaged to the model grid and a 31-day running mean applied to both model and radar estimates. The total bias for the area showed underforecast totals in summer 2000 and overforecasting (by up to 100%) in winter. Frequency histograms revealed the radar having more events with >3mm/day than the model in summer 2000. In winter there were far too many model forecasts of weak amounts ~1mm/day. Although the winter totals may be affected by overshooting radar beams and missing snow Carl thought the model still produced excessive total precipitation with a frequency bias of 1.5. A change to the formulation had been made this April and the coming winter would be monitored for impacts.

Further work planned included comparison of higher resolution forecasts from the Finnish operational Hirlam model and investigation of marine/terrestrial differences.

Clive Wilson showed further verification of precipitation forecasts using the UK composite radar precipitation analyses from the Nimrod system, and contrasted the results with those presented at the last meeting in Graz. The overall total accumulation from the UK 12 km mesoscale model agreed well with the accumulated radar estimates for the period January to July 2001, as it had for the Autumn 2000 period. However the marine/terrestrial contrast was somewhat different with now underforecasting over marine areas of Southwest England and Northeast Scotland for rain/norain threshold (0.4 mm /6 h), whilst for moderate/heavy (>8mm/6h) there was still a positive frequency bias. Sea areas for other regions showed a consistent positive bias at both thresholds. As in autumn 2000 there was generally an underforecasting of amounts over western land areas. Many of these regions contain orography and a comparison for a few cases between daily analyses from a high-density gauge network and the radar estimates show the radar to have higher values. This work will be extended to a much larger sample.

Skill and contingency scores for the autumn winter 2000/01 period were contrasted with those from February- August. The former period was unusually wet with nearly 50% rain occurrence whilst the latter was more normal with around 30% occurrence. Scores for verification against radar analyses, synoptic observations and radar analysed at the station locations only were compared. For the wetter period all the scores were closer than for the drier period when the scores using synoptic stations were generally lower. The interpretation of these differences is that the station observations are more representative during the wet period than during the "summer" period where there is likely to be more smaller scale variation. The radar estimates are more similar to the model forecasts in representing an areal average.

The routine verification of operational forecasts with Nimrod analyses has now commenced (since August 2001). Also the DWD Lokal model forecasts over the UK have been verified using the offline verification for the year 2001 to date. The Met Office and DWD will discuss the results.

Daniela Rezacova presented comparisons between Czech radar estimated precipitation and forecasts from two nested versions of the DWD Lokal model, L-LM, at 14 km 20 levels and S-LM, 2.8 km 35 levels, for 14 severe convection cases. L-LM was run from objective analyses based on successive correction method combined with optimum interpolation technique and received its boundary data from analyses at 12-h intervals. The 2.8 km, S-SM, was initialised and received its lateral boundary data from L-LM. The most extreme case was for 22-23/07/98 where flash floods had resulted from line convection with about 200 mm in 10 h. The present results were limited to cloud and precipitation forecasts. Unusually the higher resolution model predicted smaller amounts than the coarser resolution model. Early results from tests without parameterisation indicated less precipitation. The group encouraged her to examine other dynamical fields such as vertical velocity, vertical stability to determine how the model was behaving. Future work in the next 6 months would look at simulating radar reflectivity, calculating contingency table scores, verification by using pattern recognition, and simulations with higher resolution of 1 km.

3 WA2: Approaches to verification

Dieter Fruehwald showed an example of subjective verification of the German LM model against radar precipitation rate estimates (qualitative information based on 6 categories).

Christian Keil reported further results of the study of comparing Lokal model simulations for the Christmas 1999 storm (Lothar) with Meteosat satellite and radar data. The approach is from model to observations. Radiative fluxes are derived from the model temperature and humidities (cf. Morcrette, 1991). A "pyramidal" image matching between the model-forecast images and Meteosat images is then made. First both synthetic and observed images are Gaussion filtered; then the images are matched progressively from large to small scales.

Displacement vectors are found, which when applied to the synthetic images give a better agreement with the observed. The directional and magnitude variances of the vectors can then provide a skill measure. He illustrated the technique using the operational and best reruns of the Lokal model for the Christmas 1999 storms.

Clive Wilson reported briefly that the review and survey of current methods of precipitation forecast verification was finished. It did not claim to be comprehensive, but all the common scores were included and key references given. Some examples of drawbacks and strengths of different scores were illustrated using Met Office results; also a few examples of use of radar composites were included. The review would be put on the COST 717 web site and as newer techniques were developed the review would be updated.

The chairman reported on the WMO workshop on QPF held at CHMI, Prague 14-16 May 2001. The draft recommendations of the working groups are included here in the Appendix.

4 WA1: Radar data

A poster on real-time adjustment scheme for DWD by Elmar Weigl was displayed.

Joan Bech presented results of a study using 653 radiosonde ascents at Barcelona to calculate the vertical gradient of refractivity ($\partial N/\partial z$) in order to identify regimes when sub-, standard and super- refraction and ducting occurred. Both conventional TEMP messages and 10s resolution data were examined. Cumulative frequencies of $\partial N/\partial z$ in the first km showed anomalous propagation (AP) by super-refraction occurred on about 10% of occasions. Ducting was most intense over the Mediterranean in spring and autumn. The mesoscale model forecasts (MASS) could give some indication of intense AP events. Future work would aim to account for AP in beam shielding procedures and look at the joint occurrence of AP and rainfall events.

Sergio Barbosa described a Kalman filter approach to adjust radar to gauges and produce point and spatial estimates of precipitation. The eventual aim was an automatic online adjustment.

Uta Gjertsen gave a brief verbal report on application of the BALTRAD method of gauge adjustment to the Oslo radar. There was a substantial improvement in bias but still larger scatter than for BALTRAD system, due to more substantial topography. The climate gauge reports would provide more information on spatial variability and enable better methods for dealing with sectors blocked to the radar beam.

Zbynek Sokol presented results of estimating daily areal precipitation over Czech river basins using 3 years of data (April-Sept 96-98) and 2 km radar estimates. In pixels with gauges, a gauge /radar factor was calculated and interpolated to the radar domain pixels, and then the radar adjusted to give an areal estimate. Various regression methods were also used to derive areal precipitation. The study concluded that raw radar estimates could be in error by 500%; the pixel adjustment method was much improved and competitive with the regression methods.

Marco Gabella described his weighted multiple regression method of gauge adjustment. Each radar estimate is adjusted depending on 3 variables: distance from radar, minimum height of target above ground, height of ground at each radar pixel. He had applied the technique for

the October 13-16 2000 flood in Piedmont, adjusting the radar with coefficients calculated from the previous day.
A Baysesian approach to merge radar and gauge data was taken by Stanislaw Moskowicz for the Polish radar. The bias between radar and gauge was significantly reduced, but the rms error was little affected. Additional radar parameters - path integrated rainfall and height of reflectivity measurement - could easily be included. Verification of 2 limited-area models (ALADIN, UKMO_POL) was performed for the resulting 6-h precipitation totals for the summer period April-mid September 2000. Both models generally predicted more rain than observed. For all thresholds and light rain the Aladin model had a slight advantage but at longer lead-time the UKMO_POL provided significantly better forecasts.

5 Work Plan for next 6 months

Since time was short it was agreed that this would be communicated by email to the chairman following the meeting.

6 STSMs

Carl Fortelius would report to the MC on the visit of Guenther Haase to FMI to implement the radar simulation model within Hirlam.

7 AOB

The Chairman for the progress report collected manpower resource estimates of the contributions from each institution to COST 717.

Material for Andrea Rossa to present to the Technical Committee Met. were sought.

Working Papers

C. Wilson - survey of current verification techniques to be posted on the web site following the Dublin meeting.

Appendix: WMO Workshop on the Verification of Quantitative Precipitation Forecasts, Prague 14-15 May 2001 Recommendations

Working Group 1 (verification methods & data)

<u>Separation of errors and error structure functions</u>

As is common in data assimilation, errors should be separated into forecast error and observation error in order to make a more objective evaluation of the effective departure of the forecast from the observations. A good database is necessary in order to specify the error structures of precipitation data. However, this crucial condition is not reasonably satisfied for the following reasons: i) the rain gauge networks are too coarse, particularly for the verification of meso-scale models, and are often not centralized to enable processing and exchange of data; ii) the coverage of radar measurements is limited primarily to the most populated areas and is often inhomogeneous due to different technology deployed by different countries. Despite these severe limitations some countries produce objective analyses of precipitation, and there have been initial crude attempts to estimate the error structures associated with the precipitation data. Nevertheless the observation network needs to be enhanced and the methods for preprocessing of radar data (noise reduction, attenuation correction, etc.) improved and possibly homogenized so that the instantaneous data error structure can be better determined.

<u>Separation of scales (issue of representative forecast and observations)</u>

The verification should be done on scales that are appropriate for the forecast. For example, the observed and/or forecast rain patterns may be decomposed using Fourier series, wavelets or other techniques. A similar sort of filtering needs to be applied when comparing the radar and rain gauge precipitation data.

Available data

There exist good quality data sets composed of "climatological" networks of rain gauges measuring 24 h accumulated precipitation. However, these data are often not exchanged between the national meteorological services and research laboratories. These data could include radar observations, which effectively enhance the temporal resolution of the observed precipitation patterns. The best possible use of radar data must be emphasized. It is absolutely necessary to standardize the methods of producing good-quality radar rainfall estimates so that different user groups can work with the same kind of data.

Verification method and samples to be used

The choice of the method depends on the user of the verification output. Given the "client" and the information needed, an appropriate verification method and split(s) of scores might be chosen. It is necessary to determine the important questions for the verification users and to tailor the verification process to answer these questions. A diagnostic verification approach provides the tools to answer the verification questions.

It may also be useful to stratify the data samples (both in space and time) in order to extract specific error signals in the forecasts. For example, orography might be one possible criterion for stratification. Other approaches for stratifying the forecasts might include considering each station or each day separately, keeping in mind the need for adequate sample size.

Concerning the evaluation of meso-scale or other forecasts of extreme precipitation events it was agreed that radar data should be used more effectively, as stated above. In cases where the extreme event is simulated with approximately the correct amplitude but mislocated in space, decomposition of the total error into components related to amplitude and location may be useful to get a more realistic picture of forecast performance. In addition, objective verification could be complemented by subjective evaluation done by operational forecasters. In order to profit from such subjective evaluation, however, it would be necessary to work out the "subjective" evaluation procedure.

In order to assess improvements in the forecasts, especially from very high-resolution models, the higher moments of the forecast precipitation field should be looked at and verified. The structures of the forecast field should be studied qualitatively as well.

We should investigate the possibility of interpreting observations as probability and investigate appropriate verification methodologies for using this information. Confidence intervals should be included with reported verification statistics. A standard of reference, such as the performance of the persistence forecast, should also be reported.

WGII: User's needs and use of verification results

In order to provide efficient communication of verification results to end users, the working group expresses the need to set a common language. From the presentations prepared for this workshop, a common set of definitions and notations will be proposed that should serve as common framework to communicate the verification results. The common framework should start from basic definitions and gradually introduce further complexity. Relations between the different quantities should be underlined, with multiple, equivalent definitions introduced when possible. Extra information brought by a given score should be detailed, as well as limitations (e.g. how the score can be "cheated" or biased). The number of scores should be minimised in order to maximise their usability, (simple – more complex, etc…). Information on the magnitude of errors as well as the frequency characteristics should be equally represented. Examples of when a given score should *not* be used (because they would provide misleading results) should also be given.

Partitions of the different measures and scores were felt particularly useful to help the interpretation of the verification data by the end user. Several examples could be found in the lectures given at this workshop, for example:

- The partition of errors between displacement, amplitude and pattern (B. Ebert); it was suggested that information on the temporal mismatch should also be added;
- The partition of the Brier Score between resolution, reliability and uncertainty (L. Wilson); this is by many aspects similar to the partition of RMSE;
- Although quantification of the partition of errors between model, observation and representativity errors are far from simple, research aiming at this type of partition (e.g. using methodologies developed for data assimilation) should be strongly supported.

Another important feature of the verification set is that confidence tests in the verification values themselves should be given to the users. This is very similar to the crucial importance of bringing information on the confidence (probability) together with the forecast as stressed by A. Murphy (1993).

The value/ decision making approach was recognised as a very important contribution to bringing the verification results into a form that users can easily understand. The limitations of the approach for a large set of practical decision making situations were however stressed. These limitations are particularly severe for extreme events situations, when decisions have to be made without the support of a large sample of similar cases in the past on which a careful cost-loss analysis can be calibrated. These prescriptive models should be accompanied by more descriptive models of decision making.

It was agreed that the most important users of the verification results should be the forecasters. The delivery and feedback of verification results to the forecasters in a *timely*, *comprehensive* and *clear* manner on a consistent basis should be promoted. Forecasters are usually very interested in identifying and learning about key biases in the model forecasts that they can use in the process of adding value to the purely numerical forecast. Verification results should also allow them to identify key biases in their own (final) forecasts, therefore additionally contributing to the improvement of the quality of the end products. This however usually means a stratification of the verification in the form of conditional probabilities that cannot always provide unambiguous results due to the limited sample size and/or the weakness of the relation under investigation. It is therefore crucial that together with the conditional bias statistics, information on the statistical significance of the results be brought to the forecaster's knowledge. More research on the forecast process and the effect of interventions such as providing verification information should be encouraged.

Although the forecasters are a very important group of users, it was recognised that their view of the forecast quality might also have its biases with respect to an end user point of view. Their synoptic view of the forecast might very well fit with a verification using pattern matching techniques for example. A user might have a purely local point of view that will penalise phase errors much more heavily.

Training was identified as another important issue. It has been suggested that an interactive Web application be produced with WMO support. This should allow a characterisation of some typical sets of data and relate them to the different scores in a graphical as well as mathematical way. The training will explicitly include the implications of verification for different value systems. General users, administrators and forecasters should be targeted, maybe using different sets of interactive examples. A CD-ROM version of the software should also be made available

It would also be very desirable in order to help the public make better use the forecast of quantitative precipitation if standard practices to report very simple verification measures (such as percentage of detection and of false alarms) could be promoted.