Observations needed for verification of additional forecast products

Clive Wilson and Marion Mittermaier – Met Office, FitzRoy Road, Exeter, EX2 5BB

Introduction – need for additional forecasts products

With the operational introduction of increasingly higher resolution models and improved realism of parameterization schemes has come demands for additional products to be routinely available. Whereas large scale fields and upper air parameters have previously been the core products there are now requirements for more weather and surface products.

Member and Co-operating States' user requirements for ECMWF products are discussed at the annual forecast users meetings and during the rounds of regular Member State visits conducted every two years. Additionally, following a discussion at the TAC in 2007, a request for user requirements was included in the questionnaire on the "Review of the verification measures applied to medium-range forecasting", sent to Member States and Co-operating States in August 2008. Input was received from 18 countries. The results of the questionnaire were discussed at the Expert Team meeting on Verification, 9-10 September 2008, and at the TAC in October 2008 (ECMWF/TAC/39(08)7). They have formed the basis of the user requirement for the "Review of the verification measures applied to medium-range forecasting" for which a subcommittee of the TAC has been formed.

The Expert Team concluded that consideration should be given to the development of additional products for the medium range. The proposed new products are:

- Visibility/fog
- Stability indices in addition to CAPE
- Freezing rain and/or freezing level
- Height of lowest significant cloud base
- Rainfall accumulations over long durations (several days or for specific events), or rainfall duration
- Classification/clustering/regime
- Calibrated probability products (percentiles) of model and observed climate for extreme events

Newly developed products would require a certain level of monitoring and routine verification to ensure consistent good quality. This talk considers what data is available for the verification of the new products and highlights difficulties and problems to be overcome based on the experience at the Met Office. Classification/clustering and calibrated probability products will not be addressed here.

Verification of Fog and Visibility

The most important product in terms of impact is fog, defined as visibility below 1 km. Conventional surface observations such as synop and ship report visibility. Many observing sites report a mixture of manual and automatic observations. For example in a study of 42 "long-term" UK stations, 20 were manned, but only 11 full time 24/7 and all but 4 had visiometers of the Belfont type. Observers are trained to give the least visibility from any direction and the reporting code allows for precision of 100m below 5km range but to 1km above 5km. This is not really practical and as can be seen in fig 1 with ~500m resolution at lower visibilities. This can lead to differences in verification scores at different thresholds. The forecasts are from the UK mesoscale and North Atlantic European model (12km resolution).

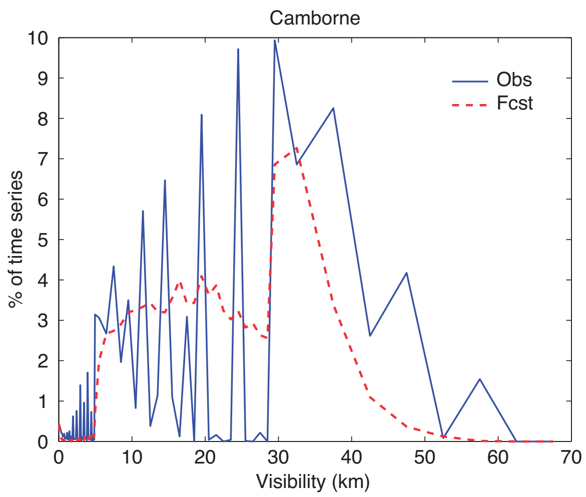

Fig. 1 Comparison of manual observations and model forecasts over 5 years

Automatic observations are closer to model forecasts in that they are continuous (figure 2). However in side-by-side comparison of the same instrument it was found that there could be a 25% offset. Also the model appears to overforecast good visibility, but this is not evident at the manned sites (eg Figure 1) and so is more likely pointing to the difficulty of visiometers measuring "clear" visibility. The differences in observations can lead to systematic differences in verification (Figure 3).

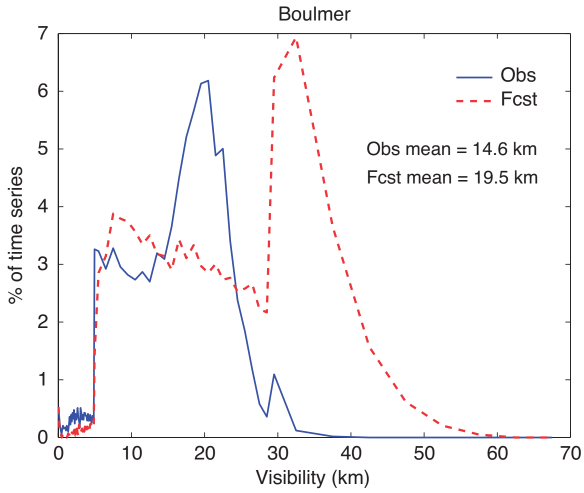

Fig. 2 Comparison of automatic observations and model forecasts over 5 years

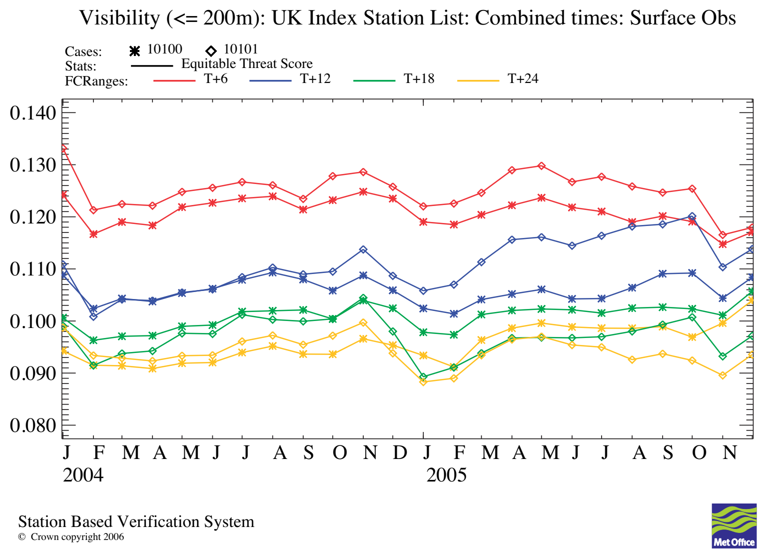

Fig. 3 Comparison of monthly Equitable threat scores for visibility below 200m with all observations (x) and with manual only (◊)

Satellite observations can detect fog and/or low cloud at night from the differences in the emissivities of the fog/cloud tops and the underlying surface at 3.9 and 10.8 microns wavelength (figure 4). An attempt is also made to separate out freezing fog (top colder than -1 deg C) from water fog (top warmer than +1 deg C) The satellite data alone cannot determine the difference between fog and stratus, since there is no way of telling if the cloud base reaches the surface. However, sometimes the coverage of the fog (e.g. in river valleys) shown in the image can give confidence to the assignment of fog. Only night-time images can be used, as during the daytime reflected solar radiation also influences the 3.7 micron channel radiances. This can also give problems around dawn/dusk. Other problems are spurious fog identification if the threshold difference is set too low and contamination by overlying ice cloud which has significant absorption in the 3.9 micron band.

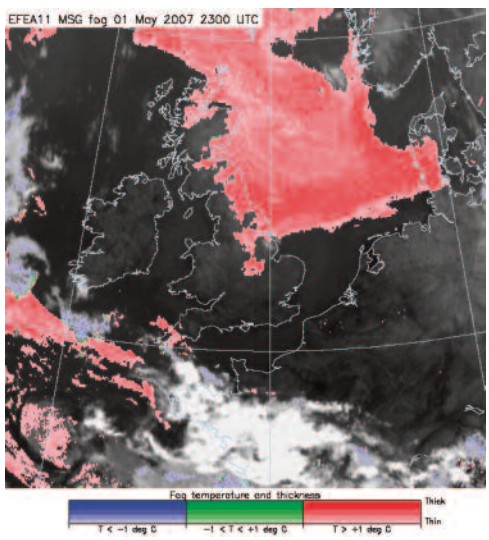

Fig. 4

Freezing level

Freezing level can be verified against radiosonde temperature reports. Typical rms errors are ~180m at T+12. Mittermaier and Illingworth, 2003 compared freezing levels, derived from steps in the returns from the vertically pointing 94 GHz radar, to model forecasts from the UM mesoscale model and ECMWF operational model. The errors were found to be rmse of 147m and 15m bias for the UM 0-5h and rmse of 318m and 58m bias for ECMWF 12-36h forecasts.

Lifting Index

CAPE and other indexes can be derived from radiosonde observations. A lifting index is defined as the difference between the environmental air temperature at 500 hPa and the temperature of an air parcel lifted from the surface to 500 hPa. The air parcel lifted from the surface is defined for a 100 hPa-thick layer immediately above the model surface, although the model's skin temperature is contributes to the definition of the mean temperature of this layer. An MSG satellite derived index can also be found. The background model forecast (from the previous 6-hourly run) is used as the first guess profile, and a 1D-Var calculation is carried out on each pixel to adjust the profile such that the radiance simulations carried out from this adjusted profile provide the best match to the observed MSG radiances. Note that the lifted index is only calculated for cloud free areas (Figure 5 left). The comparison with the first guess model only lifting index (Figure 5 right) shows large regions of agreement (green) though some areas have been identified such as Northern Germany where the satellite indicate greater instability than the model. Widespread convection was evident in the regions of large index later that day. As with other remotely sensed observations the satellite has the advantage of a much larger geographical coverage than radiosoundings.

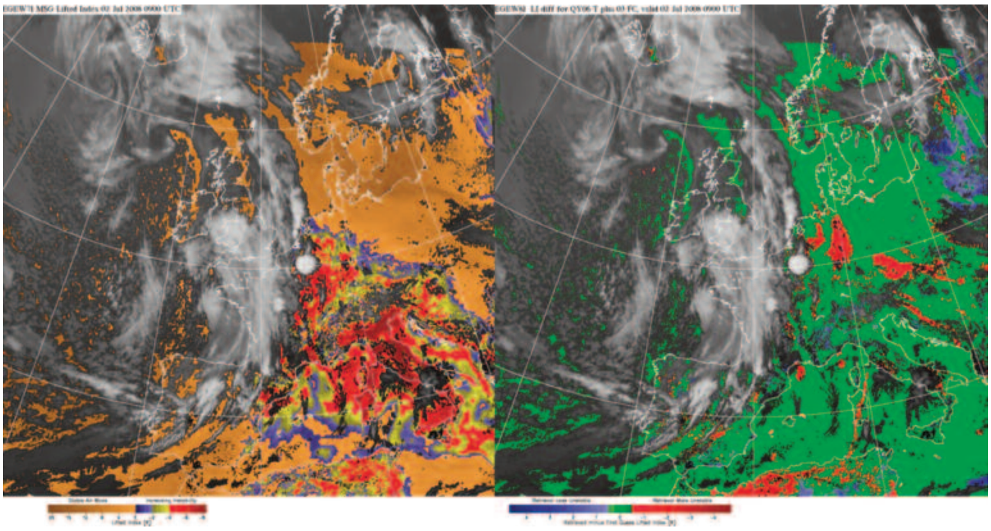

Fig. 5

Clouds

For verification of cloud amount and cloud base surface observations are most often used. However there are appreciable differences between observers and cloud detector instruments. The following problems have been identified for automated cloud observations: observations of medium and high cloud limited; too little cloud reported when it rained with under-estimation worse when it snowed; well scattered cloud poorly represented; CBH too high. The observing practice at either close to totally clear or overcasts is to bias away from these if there is a single patch of cloud or clear sky evident and report either 1/8 or 7/8. This is clearly evident in Figure 6. Automatic observations do not shy away from the extremes, but have relatively more clear sky, possible as a result of the limited field of view compared to the hemispheric view of the observer and to inability to detect small amounts of high cloud. Similarly there is an artificial cloud ceiling in the cloud base height (Figure 7) at ~2000m. (NB the zero cloud has been excluded from the CDF diagram). Because of the different characteristics of manual and automatic observations the model forecasts may either be assessed as predicting too little or too much cloud (Figure 8) and achieve a higher/lower ETS (Figure 9).

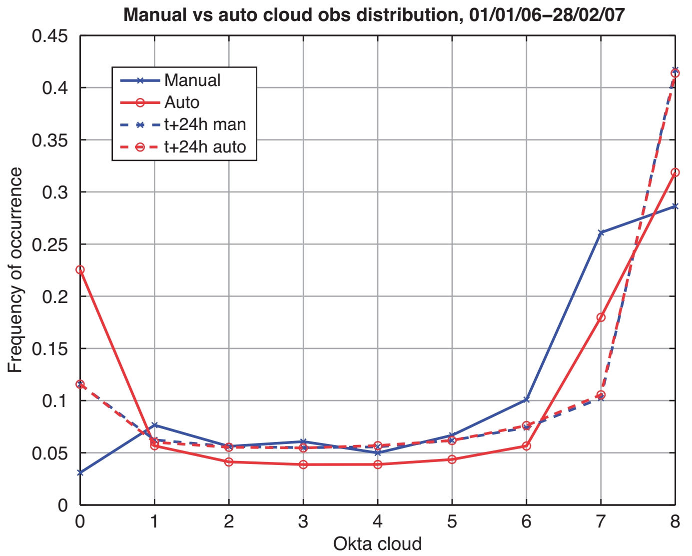

Fig. 6 Comparison of distribution of manual and automated cloud observations by okta, and NAE model forecasts

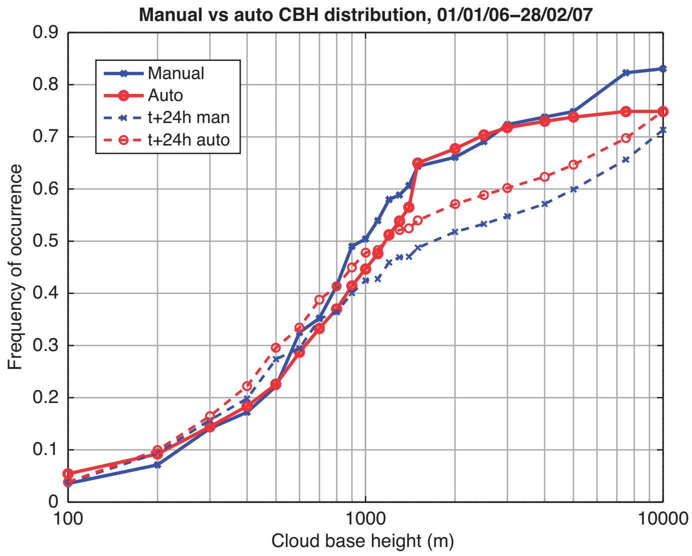

Fig. 7 Cumulative distributions of cloud base heights for manual and automatic observations and NAE model forecasts

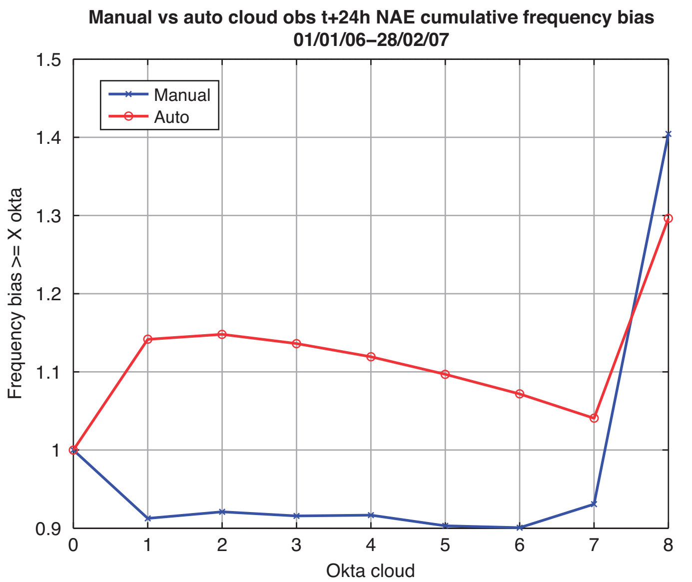

Fig. 8 Cumulative frequency biases against manual and automatic cloud observations

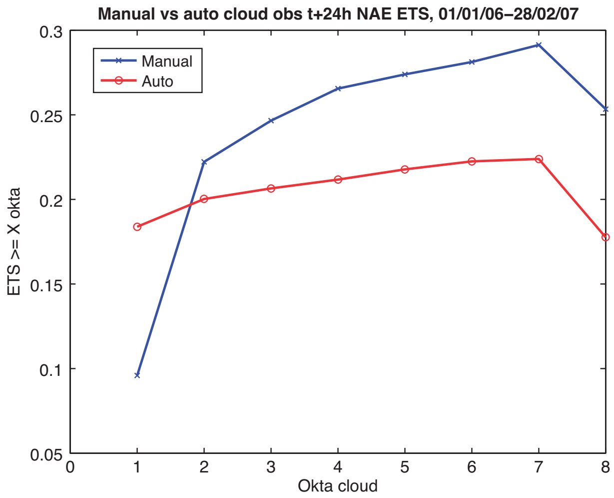

Fig. 9 ETS for model forecasts against manual and automatic observations for cloud cover

Long duration and Specific events

Key specific events in this category are pluvial flooding and snow events. Long duration events such as drought periods are also of interest. For all these events good estimates of accumulated precipitation are required. Both gauge and radar based estimates are useful. The OPERA radar hub will be an important resource in future (Figure 10), although problems of consistency and biases have to be addressed.

Radar total estimates of accumulations for the Morpeth Flood (6th September 2008, Figure 11) show the UK 4km and 1.5km models to have overpredicted the intensity although the maxima locations are generally good. The radar peaks agreed well with the gauge totals of 70-80 mm. The Met Office produce monthly analyses using all the climatological observing gauges (Figure 12 left) which with a monthly accumulation estimate from radar composites (Figure 12 right) are useful to evaluate the mean totals from the model forecasts and help to identify systematic errors.

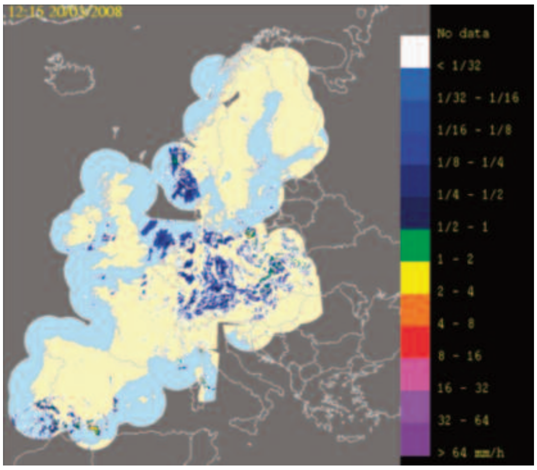

Fig. 10 OPERA composite

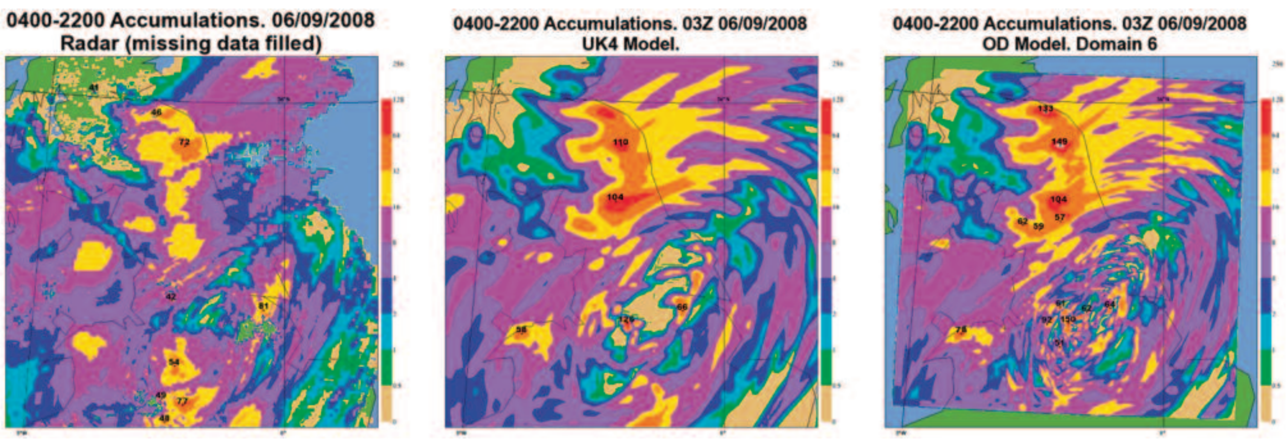

Fig. 11 Morpeth Flooding total accumulations

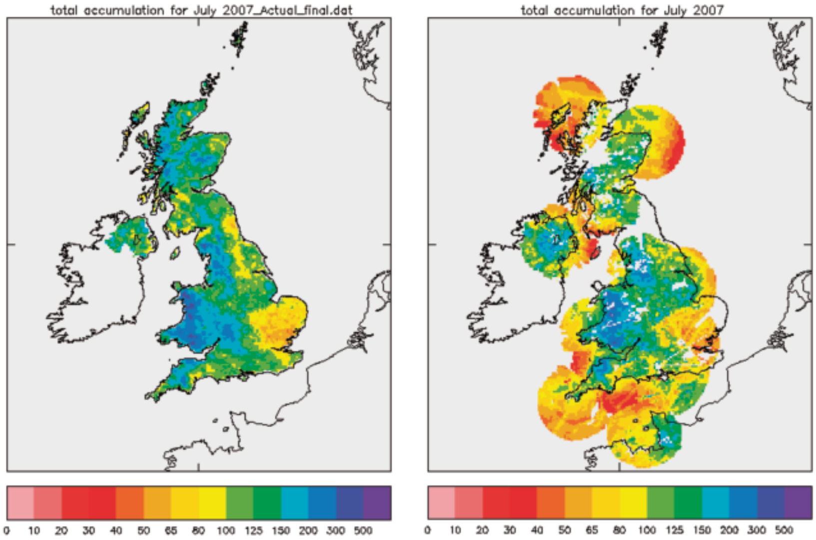

Fig. 12 Total accumulations for July 2007, gauges (left) and radar (right)

Conclusions

There are a range of observations suitable for verification of the new products. Difficulties and uncertainties arising from their characteristics have been highlighted, and especially the impact on verification scores. Increasing use will be made of remotely sensed radar and satellite observations as these provide greater spatial and temporal coverage to evaluate higher resolution model forecasts.

References

Mittermaier M.P. and **A.J. Illingworth**, 2003: Comparison of model-derived and radar-observed freezing level heights: Implications for vertical reflectivity profile correction schemes. *QJRMS*, **129**, January A No. 587, 83-96.