WGNE assessment of Quantitative Precipitation Forecasts from Operational Numerical Weather Prediction Models over the U.K.

M. Goeber,* C.A. Wilson
Met Office, London Road, Bracknell, RG12 2SZ, U.K.

Up to 3-day forecasts of daily precipitation accumulation from the 12 UTC run of 7 global, operational numerical weather prediction models as well as the mean and median forecast of those models were verified over the U.K. for more than two years (different samples between models because of transmission problems etc.). The model data were up/down-scaled by box-averaging to a common resolution of $96 * 96km^2$. The forecasts were compared against upscaled daily accumulations derived from quality controlled and corrected radar observations (Harrison *et al* (2000)) comprising the British Isles and adjacent waters.

Marginal, joint and conditional (denoted by a bar '|') probabilities p of events in certain categories were computed on the basis of monthly and total contingency tables, respectively. Plots are presented of the frequency bias FB, odds ratio θ and likelihood ratios L according to the following definitions (Stephenson (2000), Göber *et al* (2003)):

$$FB = \frac{p(f)}{p(o)}; \quad L(rain) = \frac{p(o|f)}{p(\bar{o}|f)}; \quad L(no-rain) = \frac{p(\bar{o}|\bar{f})}{p(o)|\bar{f})}; \quad \theta = L(rain) * L(no-rain)$$

where f (o) denotes the forecast (observed) event and $\bar{f}$ $(\bar{o})$ denotes the not forecast (not observed) event.

Fig. 1 shows that most models are reasonably well calibrated, with the MEAN and MEDIAN (MM) forecasts showing the behavior expected from a smoothed forecast. The skill of the forecasts ($\log\theta$) increases slightly with accumulation. The main cause of this is the stronger drop in false alarm rate $p(\bar{o}|f)$ than the drop in hit rate $p(o|f)$ with threshold (not shown). Note, that the odds ratio corrects for the 'base-rate' effect, which is a strong influence in other scores like Equitable Threat Score, i.e. it accounts for the fact that for *rare* events one can not get lots of hits and there are lots of potential cases to issue a false alarm (Göber *et al* (2003)). The MM forecasts are generally better than the best single model. A split of the odds ratio into forecasts of the event and non-event reveals that the MM forecasts are good in both categories whereas single models are good in either forecasting the event or the non-event. Figs. 2 of the time evolution of the monthly scores show a substantial variability of the monthly performance of the models themselves and between the models. Again, the MM forecasts perform relatively well in most months.

References

Göber, M, C. A. Wilson, S. F. Milton, and D. B. Stephenson, 2003: Fairplay in the verification of operational quantitative precipitation forecasts. *J. Hydrol.*, *submitted for publication*.

Harrison, D. L., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.*, **6**, 135–144.

Stephenson, D. B., 2000: Use of the 'odds ratio' for diagnosing forecast skill. *Wea. Forecasting*, **15**, 221–232.

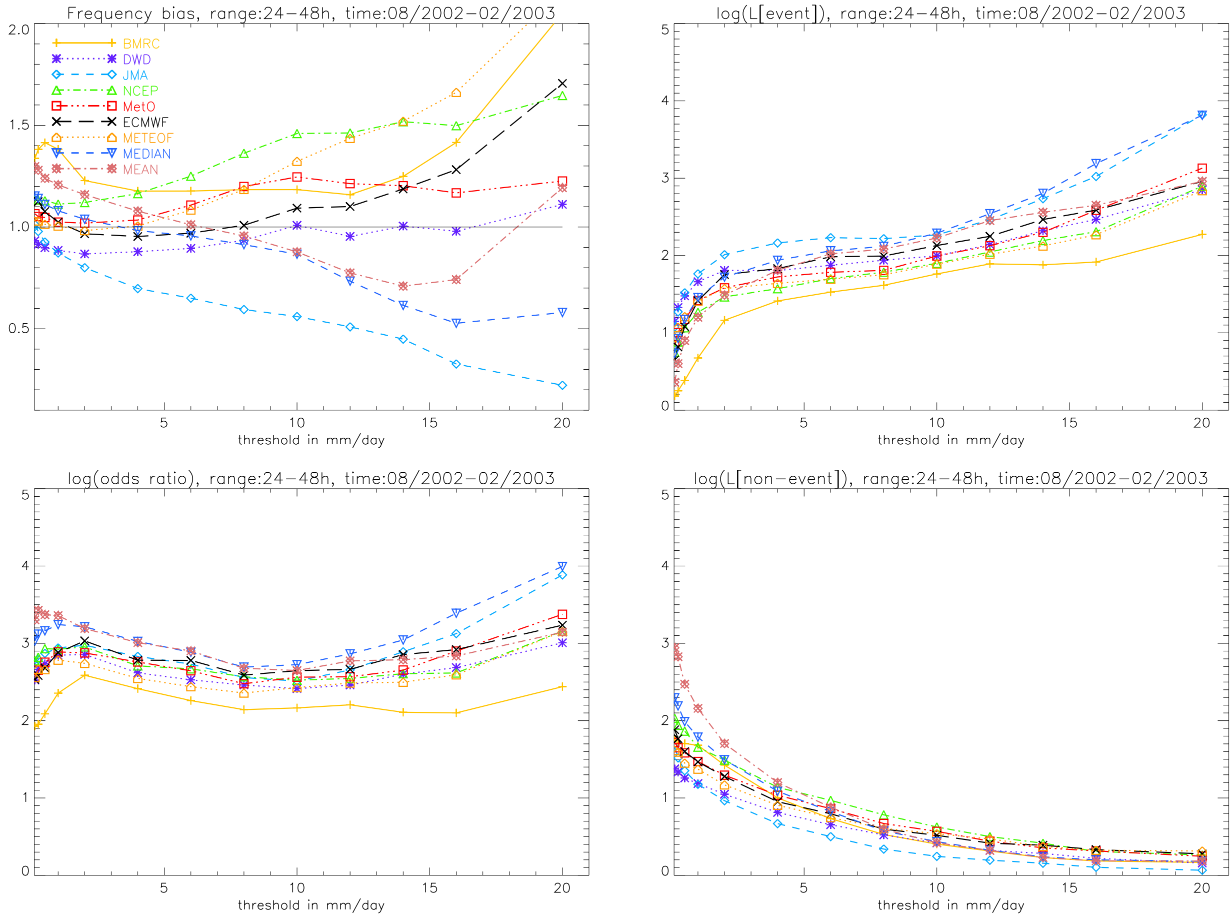

Figure 1: Frequency bias FB (upper left), odds ratio $ln\theta$ (lower left) and its split into the likelihood ratios $L(rain)$, $L(no-rain)$ (upper and lower right column) as a function of precipitation threshold for daily accumulations 2 days into the 12-UTC forecasts based on a sample from 08/02-02/03. Mean and Median are taken from all models available at a particular grid box and time.

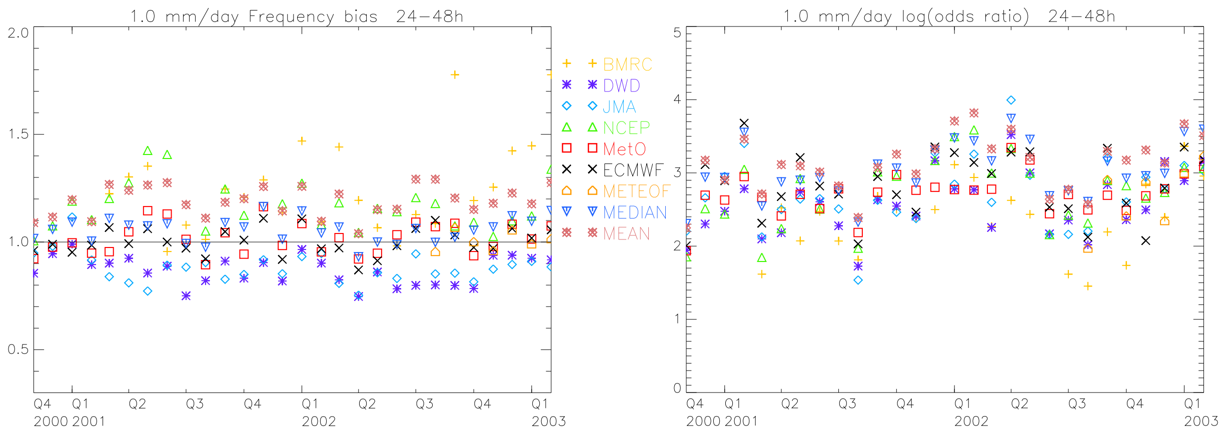

Figure 2: Monthly time series of frequency bias (left) and log(odds ratio) (right) for daily accumulations of more than 1.mm/day of day two of the 12-UTC forecasts. Note missing data and the longer time interval than Fig. 1.