Application and verification of ECMWF products 2013

Royal Meteorological Institute of Belgium – Pascal Mailier

3. Verification of products

3.1 Objective verification

3.1.2 ECMWF model output compared to other NWP models

The accuracy (root mean squared error) of ECMWF 2-m temperature forecasts of the year 2012 is compared briefly with the accuracy of the corresponding 2-m temperature forecasts produced by the RMI's own high-resolution (7 km) limited-area model Alaro. A Kalman filter has been applied to all forecasts in order to remove systematic biases. Results are presented for Uccle (06447).

The plot in Fig. 1 indicates that, after some processing, the performance of ECMWF 2-m temperature forecasts is as good as, or even slightly better than the performance of Alaro. This result holds at a majority of Belgian stations.

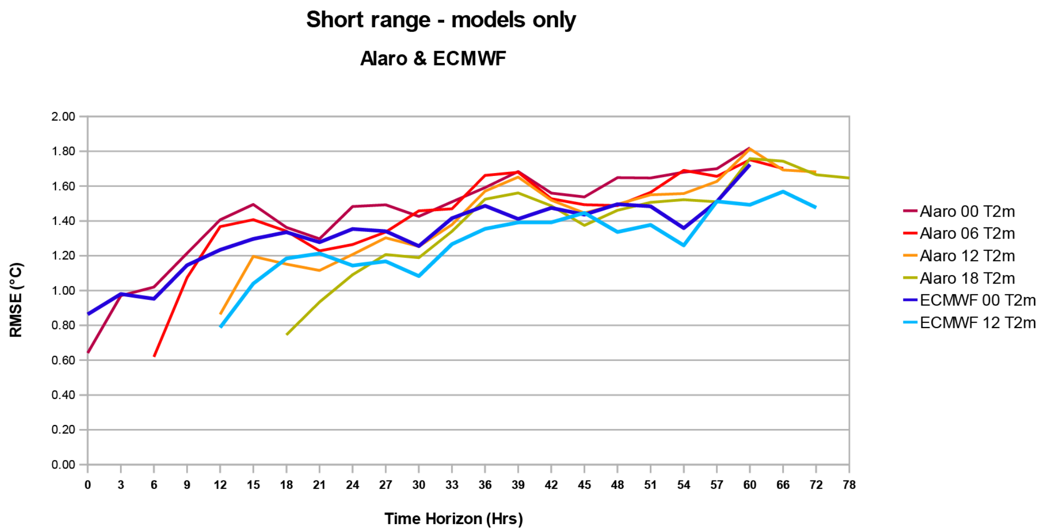

Figure 1 - Root mean squared error (RMSE) of 2-m temperature short-range forecasts for Uccle (06447) produced by the ECMWF high-resolution model (blue) and Alaro (other colours) during the year 2012.

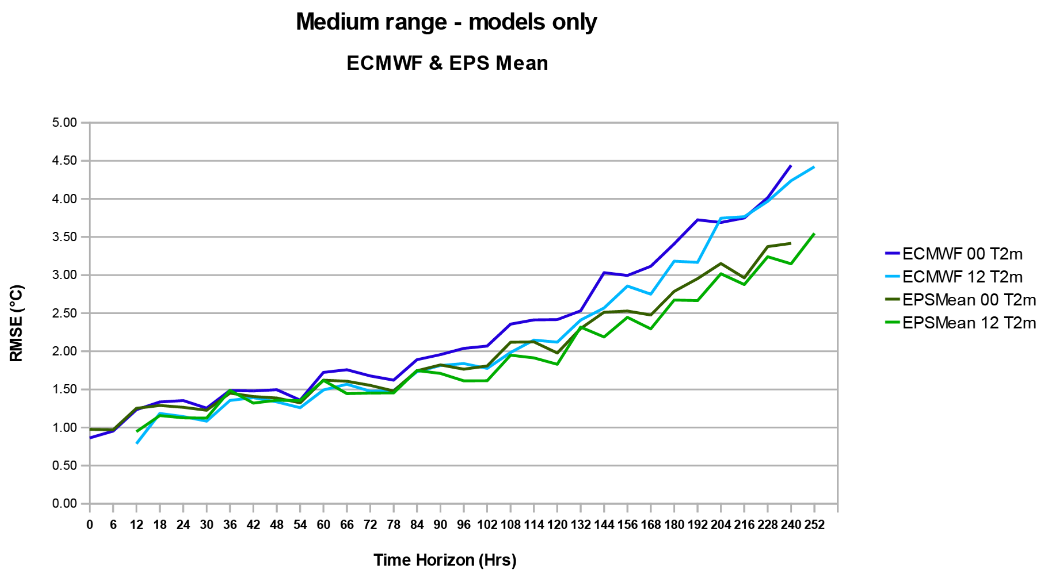

Figure 2 – Root mean squared error (RMSE) of 2-m temperature medium-range forecasts for Uccle (06447) produced by the ECMWF high-resolution model (blue) and the EPS mean (green) during the year 2012.

In Fig. 2, a comparison is made between the ECMWF high-resolution forecasts and the EPS mean (all forecasts have been Kalman-filtered). As expected, the EPS mean performs better in the medium range. Despite the lower EPS resolution, the EPS mean accuracy appears strikingly close to that of the ECMWF high-resolution forecasts in the short range.

Figures 3 and 4 compare the performances of minimum and maximum 2-m temperature forecasts from Alaro, the ECMWF high-resolution model and the EPS. In both cases, the forecasts produced by ECMWF emerge as more accurate.

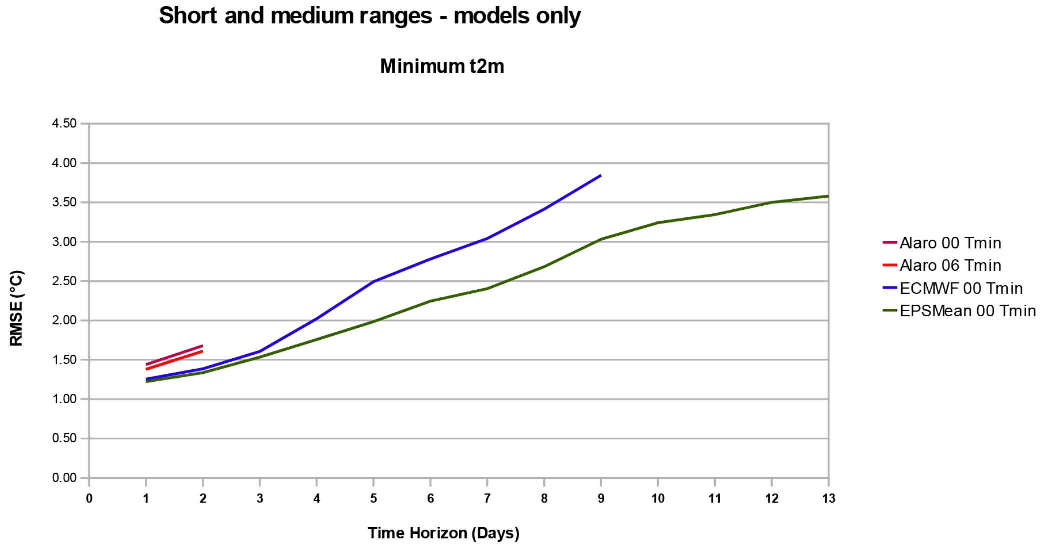

Figure 3 - Root mean squared error (RMSE) of minimum 2-m temperature forecasts for Uccle (06447) produced by Alaro (red), the ECMWF high-resolution model (blue) and the EPS mean (green) during the year 2012.

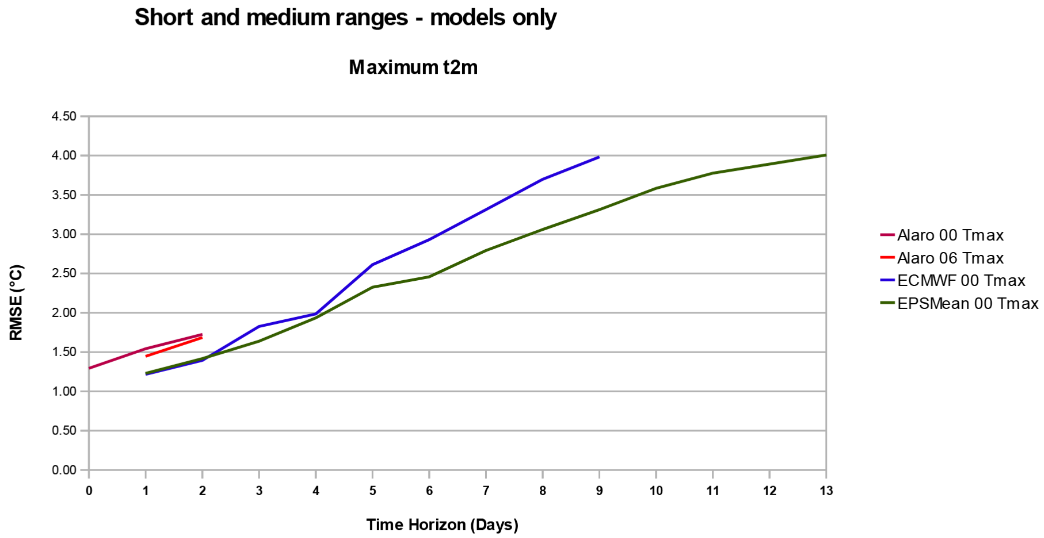

Figure 4 - Root mean squared error (RMSE) of minimum 2-m temperature forecasts for Uccle (06447) produced by Alaro (red), the ECMWF high-resolution model (blue) and the EPS mean (green) during the year 2012.

The performance of 10-m wind short-range forecasts is examined in Fig. 5. Here too, the forecasts have been de-biased using a Kalman filter. Again, the ECMWF high-resolution model and the EPS mean show better or equivalent accuracy most of the time when compared with Alaro.

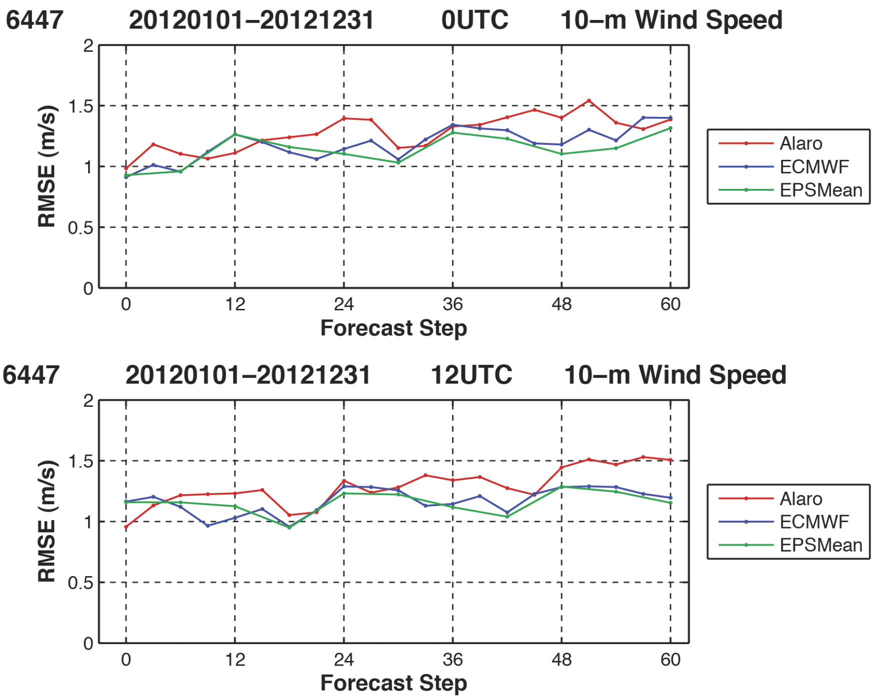

Figure 5 - Root mean squared error (RMSE) of 10-m wind short-range forecasts for Uccle (06447) produced by the ECMWF high-resolution model (blue), the EPS mean (green) and Alaro (red) during the year 2012.

3.1.3 Post-processed products

The EVMOS (Error-in-Variables Model Output Statistics) post-processing of the EPS (Van Schaeybroeck and Vannitsem, 2012) now runs operationally at the RMI. The post-processed EPS forecasts were benchmarked against the corresponding Kalman-filtered EPS forecasts for a couple of surface weather variables at 11 stations of reference for the period 20110620 to 20120614. The Kalman filter (KF) provides an excellent benchmark as it is computationally much lighter than EVMOS. Furthermore, the KF requires little additional data whereas the EVMOS relies on the availability of large volumes of reforecast data. Results for the predicted 2-m temperature and 10-m wind speed (forecasts starting from 00UTC, 6-hourly timesteps) at a typical Belgian station (Uccle, 06447) are shown in the figures below. On its public web site, the RMI summarises the EPS forecasts with diagrams showing the ensemble mean and 70% central prediction intervals. Accordingly, a set of deterministic and probabilistic scores was used to test the predicted ensemble mean and intervals.

Figure 6 shows the bias and root mean squared error of the 2-m temperature ensemble-mean forecasts (top) together with the Pearson product-moment correlation coefficient between ensemble-mean forecasts and observations (bottom). Both the EVMOS post-processing and the KF lead to improved accuracy through bias correction, more particularly during the late and early hours (peaks around midnight). There is little or no improvement around noon when the bias is negligible. The linear association between ensemble-mean forecasts and observations is as good as unaffected by the post-processing. The benefit of the EVMOS post-processing is above all visible within a time horizon of one week. The KF is nearly as good as the EVMOS, which indicates that much of the errors at those time ranges are dominated by systematic biases. At time horizons beyond one week, where the stochastic component of the error dominates, only the EVMOS still brings some marginal improvement.

Figure 7 shows the hit rate and interval score (Gneiting and Raftery, 2007) on 70% prediction intervals. The EVMOS and the KF improve the quality of the prediction intervals slightly out to one week. The KF appears more efficient than the EVMOS to produce better hit rates in the short range whereas only the EVMOS improves the prediction intervals marginally after one week. The most striking feature is the low hit rate (systematically below 70%) despite the corrections, more particularly in the first week of the forecast.

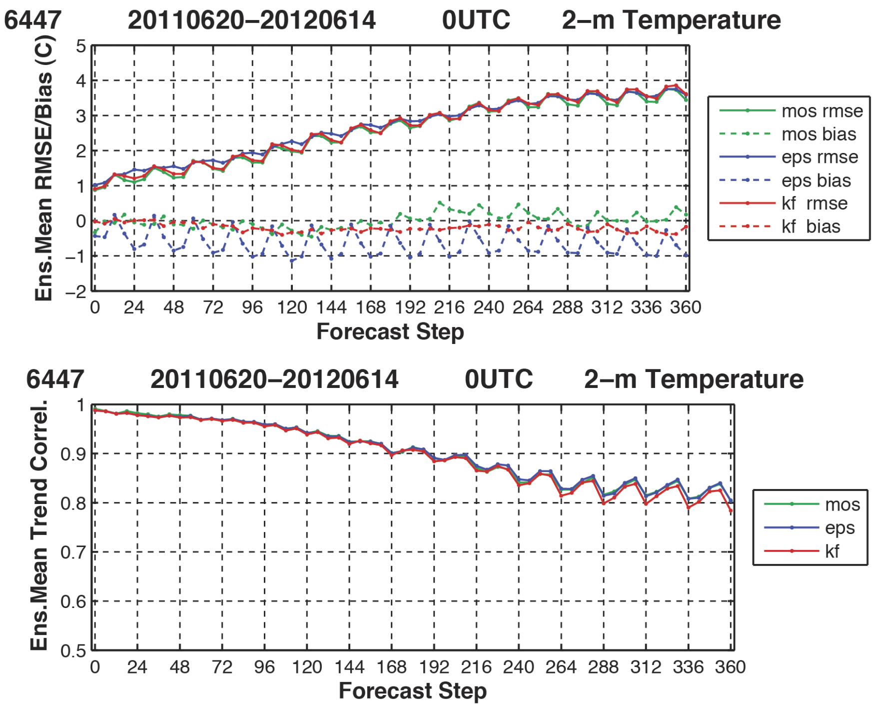

Figure 6 - Top: Bias (mean error, dotted lines) and accuracy (root mean squared error, solid lines) of the uncorrected (blue), EVMOS post-processed (green) and Kalman-filtered (red) 2-m temperature ensemble-mean forecasts. Bottom: Linear association (correlation coefficient) between ensemble-mean forecasts and observations.

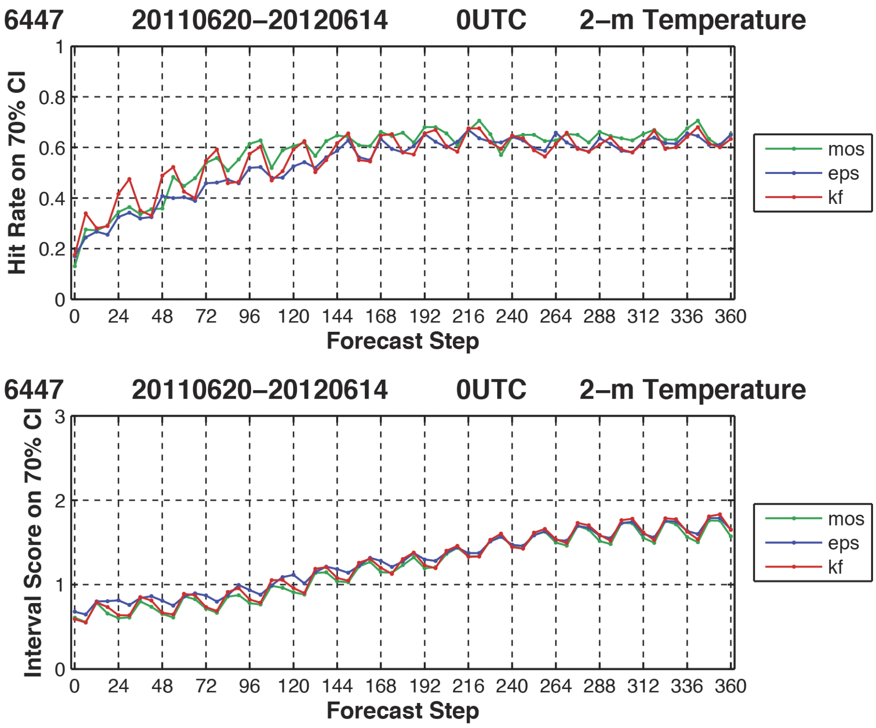

Figure 7 – Hit rate (top) and interval score (bottom) on 70 prediction intervals of 2-m temperature. Colours as in Fig. 6.

The bias, root mean squared error and correlation coefficient of the ensemble-mean forecasts of the 10-m wind speed are presented in Fig. 8. The EVMOS improves the overall accuracy of the forecasts more than the KF. Systematic biases still account for a significant part of forecast errors, but not so much as in the case of 2-m temperature forecasts. The KF does not improve or even deteriorates the association between ensemble-mean forecasts and observations.

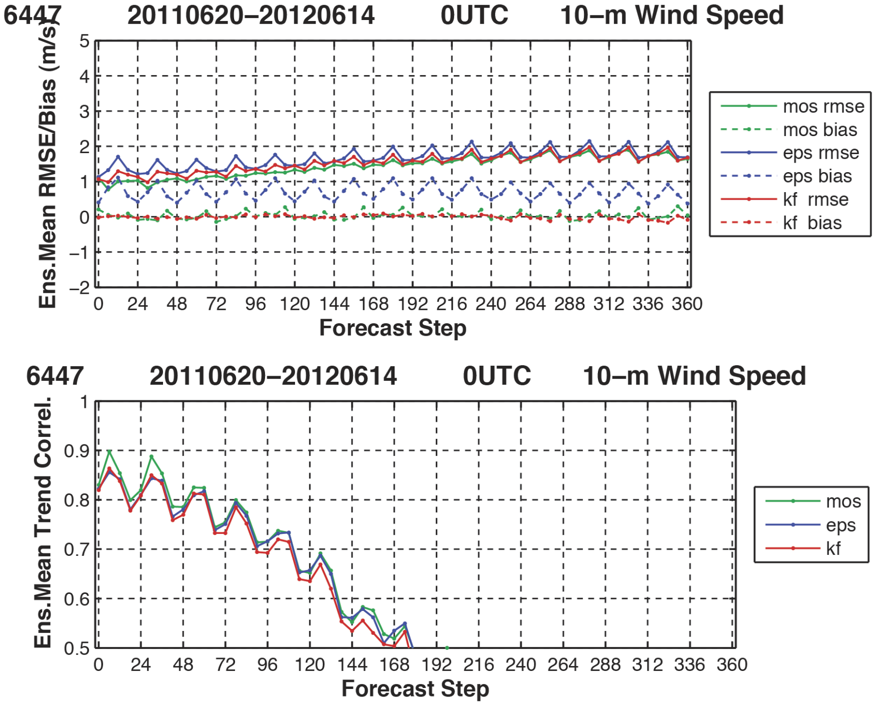

Figure 8 – Same as in Fig.6 for 10-m ensemble-mean forecasts.

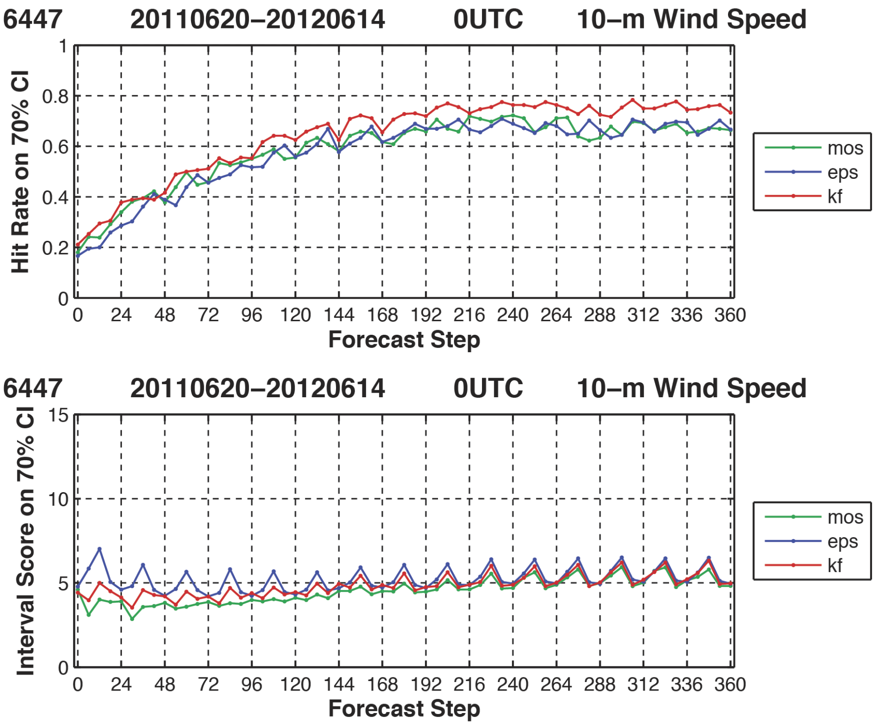

Figure 9 – Same as in Fig. 7 for 10-m ensemble-mean forecasts.

Like in the case of 2-m temperatures, the hit rate is far too low in the first few days of the forecast as can be seen from Fig. 9. The EVMOS consistently produces better interval scores than the KF. But none of these schemes manages to improve the hit rate convincingly though.

4. References to relevant publications

Van Schaeybroeck, B. and **S. Vannitsem**, 2012: Toward post-processing ensemble forecasts based on hindcasts. *Publication scientifique et technique IRM N° 061*, 22pp.

Gneiting, T. and **E. Raftery**, 2007: Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Americ. Stat. Assoc.*, **102**(477), 359-378.