Verification of the operational high-resolution model AROME

J. Stein, M. Amodei, I. Sanchez and F. Pouponneau
Météo-France/DPREVI/COMPAS/COM

After the preliminary tests of scores of quantitative prediction forecasts using neighbourhood (Roberts and Lean 2008) for the hierarchy of operational models (Amodei and Stein 2009), the method has been applied to other parameters: brightness temperature and wind gusts. The long temporal series of daily forecasts made by the high resolution model AROME from December 2008, where it became operational and February 2011, were collected to monitor the AROME quality (horizontal resolution 2.8 km) and compared with the ALADIN model results (horizontal resolution 10 then 7.5 km). The reference is always given by observations: from 600 to 1800 surface stations covering France and Meteosat 9 data over the simulation domain.

The increase of the Brier skill score against the persistence forecast with increasing neighbourhood is presented in Figure 1 for the 6-hourly accumulated rain for the whole year 2010 for 4 different thresholds and 4 different lead times.

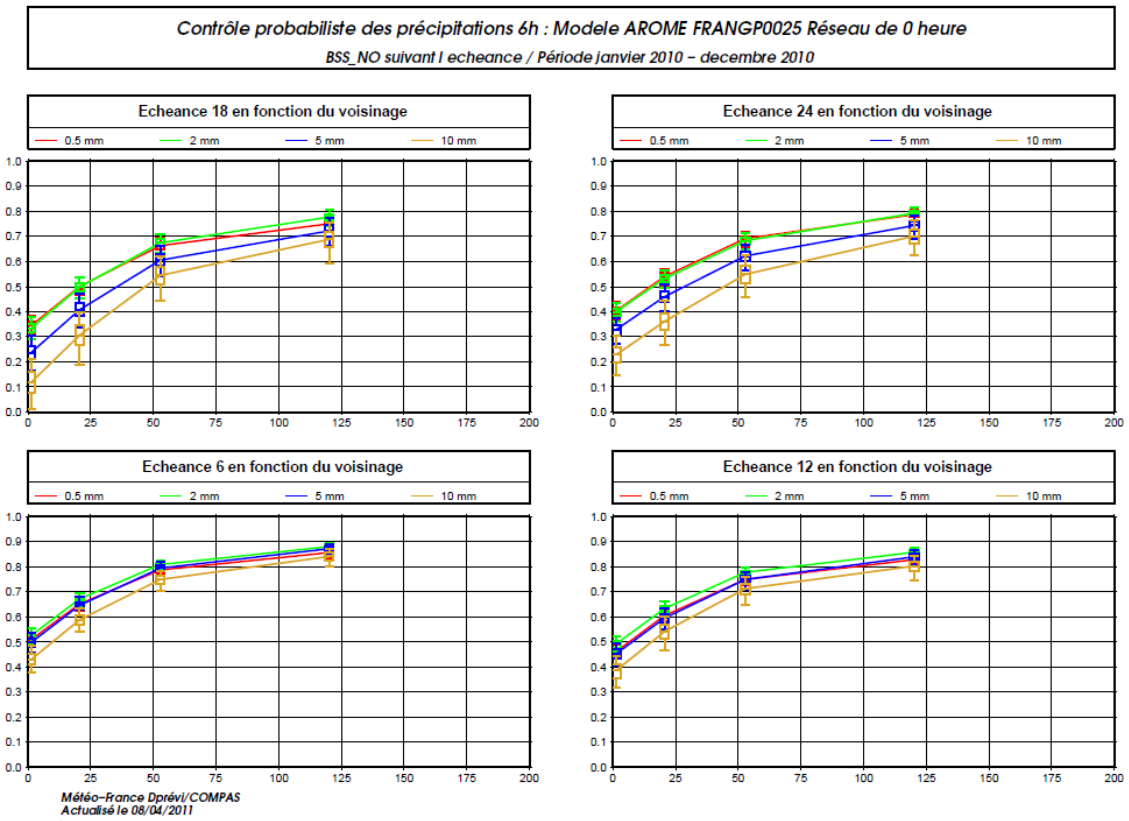

Figure 1: Brier skill scores against the persistence forecast in function of the neighbourhood size for 4 different thresholds: 0.5 mm /6h (red), 2 mm/6h (green), 5 mm/6h (bleu) and 10 mm/6h (beige). The box-plots collects the quantiles 5, 25, 75 and 95% obtained through a bootstrap technique. The scores are computed for the run starting at 0 UTC at 4 lead times: 6 h (bottom left), 12 h (bottom right), 18 h (top left) and 24 h (top right).

A fixed radius of 50 km has been chosen for the neighbourhood. This corresponds to the typical size of French administrative area and of part of the delivered final forecasts. The probabilistic scores of accumulated rain and wind gusts have been averaged with equal weights for 4 lead times from 6 to 24 hours to form 2 of 3 ingredients for a synthetic score for AROME. The temperature measured at 2m AGL is the third element of the synthetic score: it is evaluated by an rms error normalized by

the rms of the persistence forecast. The evolutions of the synthetic score and its 3 components are presented on Figure 2. It shows some improvements of the model related to 2 main changes in the operational suite in April and November 2010. The first change corresponds to the start of the assimilation of radar reflectivities, which leads to an improvement of the scores until 12 hours. The second change (enlargement of the domain) does not really improve the AROME forecasts as measured by theses 3 scores. Of course the synthetic score is a summary of a long list of scores and analysis of scores mixing classical and probabilistic scores.

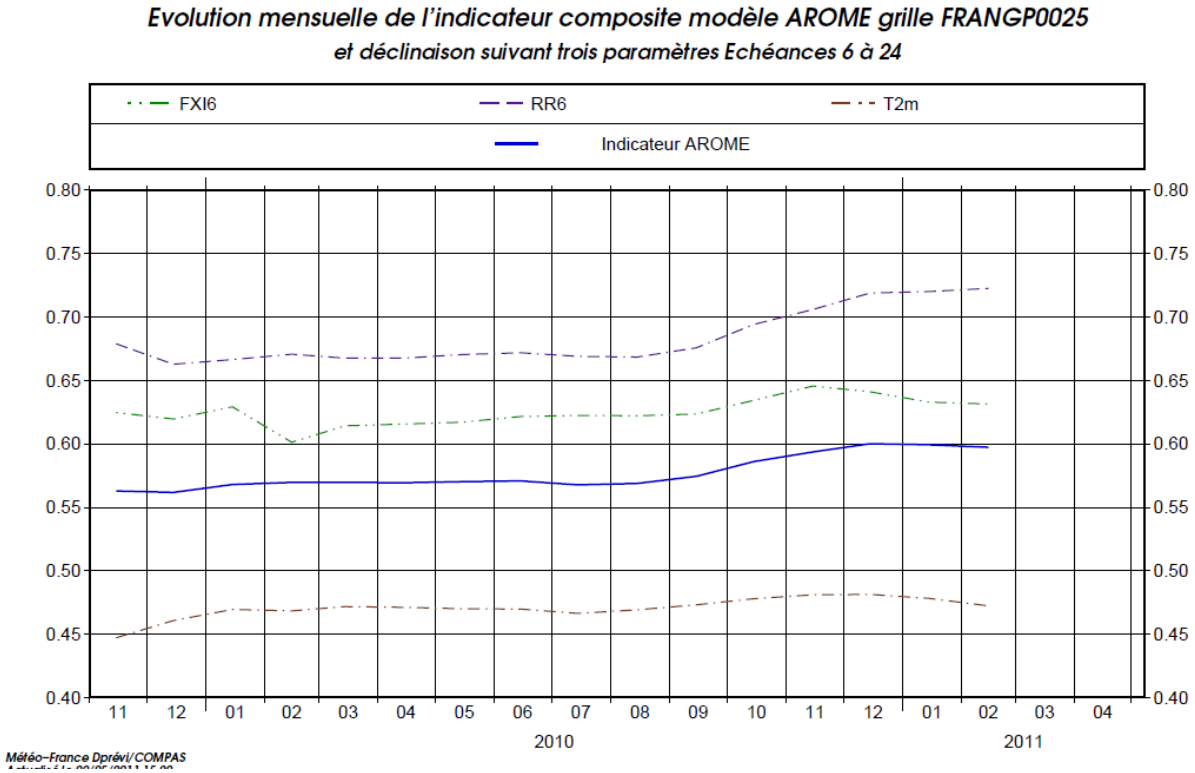

Figure 2: temporal evolution of the synthetic score (bleu curve) with the contributions of the different components all normalized the persistence forecast : maximum wind gust during the last 6 hours (dashed green curve), 6 hours accumulated rain (dashed purple curve) and 2m temperature (dashed brown curve). Every score is the mean of the 4 scores corresponding to the lead times 6, 12, 18 and 24 hours.

The evaluation of the quality of the nebulosity is also made through the computation of the brightness temperatures derived by the RTTOV-cloud radiative transfer scheme. Probabilistic scores were used to verify the forecasted fields against Meteosat 9 data for the 10.8 µm IR channel. This fourth component is also followed during the same period and shows the same evolution as the rain parameter. The temporal series is too short to decide whether specific evolution should be isolated from this fourth parameter.

The use of a neighbourhood was fundamental to obtain better scores for the high resolution model AROME (from 3 to 5 %) in comparison to the ALADIN model (not shown) for the rain and the wind gusts.