Comparison of precipitation forecasts by two operational NWP models:

The global model ARPEGE and the smaller mesh LAM

ALADIN

Marielle Amodei and Joël Stein

DPrevi/COMPAS Météo-France

Objectives:

1. The 24 hours accumulated precipitation forecasted by ARPEGE and ALADIN is verified and compared over France according to the methodology recommended by the WGNE (Working Group on Numerical Experiment) in order to:
   - Define a set of relevant scores to verify the precipitation forecasted by the French operational high-resolution LAM, ALADIN today, AROME in the future.
   - Evaluate differences between 2 models and determine if they are statistically significant.

2. The daily raingauges will be replaced by the ANTILOPE hourly analysis mixing raingauges with radar data.

Future:

- The AROME prototype of the future high-resolution model will be verified according to this methodology.

Observations and models used:

- The temporal period contains 9 trimesters from December 2004 until February 2006.
- The climatological state network (Figure 1) contains almost 4000 observations per day. These observations, spaced by about 10 km, are averaged on the 0.2° verification grid to provide the reference data.
- The persistence of the precipitation of the previous day provides the trivial forecast used to compute skill scores.
- The spatial resolution of the uniform verification grid is 0.2° x 0.2°.
- The ARPEGE forecasts, available over all 372 points and the ARPEGE area are interpolated on a 0.2° grid.
- The global ARPEGE model provides lateral boundary conditions to ALADIN. The physical packages are similar for the 2 models. Since the 29th of July 2005, the ALADIN model has its own 3DVAR assimilation and assimilates the high-resolution SEVIRI radiances from Meteosat 8.

Fuzzy approach:

In order to avoid the “double penalty” which affects the high resolution models, a statistical post-processing is performed to allow a probabilistic treatment of the forecasts (Roebert 2004, Thao et al. 2005). A probability to exceed a given threshold, is computed in a square of non grid points for the forecasts and the persistence. The verification is performed by computing the Brier Skill Score (BSS) and Rank Probability Skill Score (RPSS). Both are estimated according to the persistence. 2 BSS are computed: BSS-SS single observation – neighbourhood forecast gives us a local verification

BSS-N0 neighbourhood observations– neighbourhood forecast gives us a regional verification.

The neighbourhood characteristic size (NCS) vary from 60 km to 242 km. Both BSS are plotted for two seasons (JUL 2006 and JUL 2004) (Figure 5). By comparing both figure, we conclude that winter precipitation is better forecasted than summer one because of the more convective nature of the summer rainfall.

The comparison between the two models is shown for the BSS-N0 along the whole period for extreme thresholds values (0.2 mm/d, 20 mm/d) and 2 NCS (60 km, 242 km) (Figure 6). The largest neighbourhood always achieves the largest forecast improvement whatever the threshold. The BSS-N0-ALADIN is 10% larger than the BSS-N0-ARPEGE.

The fuzzy approach shows an improvement of the ALADIN model for the rain detection (0.2 mm/d) during the last two trimesters after the ALADIN assimilation introduction. This is coherent with the classical HSS scores. For the 20 mm/d threshold, the ALADIN improvement after JJA 2005 is greater for BSS than for HSS.

This is related to a reduction of the double penalty importance allowed by the fuzzy method. Nevertheless, the impact is not as important for very high resolution models because the resolution ratio between ARPEGE and ALADIN is only 2.3 and they share the same physical package.

A more synthetic information is given by the Rank Probability Skill Score (RPSS) (Figure 7). PRSS-SS and RPSS-NO are compared in the same way as the BSS. Reference is provided by the persistence. RPSS of the two models and of the persistence have been multiplied by a factor 15 and plotted together with the RPSS. These curves confirm that the RPSS variations are mainly due to the PRSS variations and the ALADIN improvement is clear after the introduction of its proper assimilation. We also notice the improvement in term of skill scores when a larger verification box is used.

Deterministic approach:

Verification is performed in a deterministic framework. Scores are computed over every trimester from the contingency tables for the different thresholds 0.2, 1, 2, 5, 10, 20 mm/day over France. A bootstrap resampling technique is used to compute a box plot for the different scores but also to test the significance of the differences between both models according to Hamm (1989).

The temporal evolution of the HSS (Figure 4) shows the significant superiority of ALADIN for the rain detection (threshold 0.2 mm/day). For stronger precipitation (20 mm/day), this is the contrary. The improvement occurring after JJA 2005 for the ALADIN model is permanent for both thresholds and is due to 3DVAR assimilation introduced in ALADIN on the 25/07/2005. Nevertheless, it only leads to a reduction of the gap between both models for strong precipitation.

The comparison between the two models is shown for the BSS-N0 along the whole period for extreme thresholds values (0.2 mm, 20 mm) and 2 NCS (60 km, 242 km) (Figure 6). The largest neighbourhood always achieves the largest forecast improvement whatever the threshold. The BSS-N0-ALADIN is 10% larger than the BSS-N0-ARPEGE.

The fuzzy approach shows an improvement of the ALADIN model for the rain detection (0.2 mm/d) during the last two trimesters after the ALADIN assimilation introduction. This is coherent with the classical HSS scores. For the 20 mm/d threshold, the ALADIN improvement after JJA 2005 is greater for BSS than for HSS.

This is related to a reduction of the double penalty importance allowed by the fuzzy method. Nevertheless, the impact is not as important for very high resolution models because the resolution ratio between ARPEGE and ALADIN is only 2.3 and they share the same physical package.

A more synthetic information is given by the Rank Probability Skill Score (RPSS) (Figure 7). PRSS-SS and RPSS-NO are compared in the same way as the BSS. Reference is provided by the persistence. RPSS of the two models and of the persistence have been multiplied by a factor 15 and plotted together with the RPSS. These curves confirm that the RPSS variations are mainly due to the PRSS variations and the ALADIN improvement is clear after the introduction of its proper assimilation. We also notice the improvement in term of skill scores when a larger verification box is used.

This is related to a reduction of the double penalty importance allowed by the fuzzy method. Nevertheless, the impact is not as important for very high resolution models because the resolution ratio between ARPEGE and ALADIN is only 2.3 and they share the same physical package.

A more synthetic information is given by the Rank Probability Skill Score (RPSS) (Figure 7). PRSS-SS and RPSS-NO are compared in the same way as the BSS. Reference is provided by the persistence. RPSS of the two models and of the persistence have been multiplied by a factor 15 and plotted together with the RPSS. These curves confirm that the RPSS variations are mainly due to the PRSS variations and the ALADIN improvement is clear after the introduction of its proper assimilation. We also notice the improvement in term of skill scores when a larger verification box is used.

Conclusion:

- The probabilistic approach has been used to evaluate the differences between the global ARPEGE model and the LAM ALADIN. BSS-N0 against persistency is a good candidate to measure the impact of higher resolution and assimilates the high resolution SEVIRI radiances from Meteosat 8.

- This is related to a reduction of the double penalty importance allowed by the fuzzy method. Nevertheless, the impact is not as important for very high resolution models because the resolution ratio between ARPEGE and ALADIN is only 2.3 and they share the same physical package.

A more synthetic information is given by the Rank Probability Skill Score (RPSS) (Figure 7). PRSS-SS and RPSS-NO are compared in the same way as the BSS. Reference is provided by the persistence. RPSS of the two models and of the persistence have been multiplied by a factor 15 and plotted together with the RPSS. These curves confirm that the RPSS variations are mainly due to the PRSS variations and the ALADIN improvement is clear after the introduction of its proper assimilation. We also notice the improvement in term of skill scores when a larger verification box is used.