Progresses on LAMI, LM-DWD, aLMo verification over Northern Italy

M. Turco, E. Oberto, P. Bertolotto

ARPA Piemonte, Italy

1 Comparative high resolution verification over Piedmont

1.1 The method

In this section we compare the performances over Piedmont of the three versions of LM (LAMI, LM-DWD, aLMo), using cumulated precipitation maps for each season from June 2003 to November 2004 and objective methods of verification using statistical indices. In this kind of verification we consider only Piedmont region because of the high density, spatial homogeneity, and high operativeness percentage of the observational network.

This non-GTS network is very dense and it is composed by more than 350 rain gauges. In order to obtain the cumulated QPF maps, we use ordinary Kriging techniques to interpolate the observed values over Piedmont, taking into account only the rain gauges with a 100% percentage of valid data in a season ($\sim$ 250 stations). In Fig. 1 we show the percentage of valid data over the whole considered period.

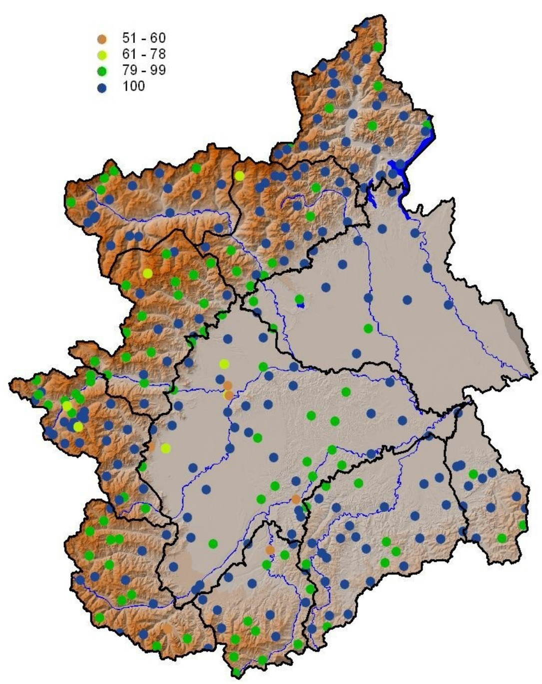

Figure 1: Percentage of rain gauges valid data used for the verification over the whole period and Piedmont warning areas.

The observed precipitation in each season from June 2003 to November 2004 is compared with the forecasted cumulated precipitation in order to obtain a visual comparison of the three versions of LM (LAMI, LM-DWD, aLMo) in terms of overestimation and underestimation. We consider the 24h cumulated precipitation every day for the first 24h of forecast time.

At the same time we perform an objective verification with statistical indices (BIAS, ETS) among the three model versions and the observations. We consider eleven basins, with mean size of 3000 km^2 that represent groups of neighboring hydrological catchments used as warning areas (see Fig. 1).

We average 24h forecasted and observed rain amounts over these basins in order to obtain contingency tables and statistical indices from June 2003 to November 2004. We compare the results fixing the 24h threshold and calculating the contingency tables season by season (using the first and the second 24 hours of forecast time).

In order to compare two model versions, a confidence interval is necessary to assess the real differences between skill scores. This is computed using a bootstrap technique, first developed by Hamill; for an exhaustive description of the method see Hamill, 1999. This test requires that the time series considered have negligible autocorrelations. Spatial correlation of data is taken into account by calculating the scores without considering any geographical partition of the data; time correlation is considered by choosing an accumulation period (24h) larger than the time correlation scale (as shown in Accadia et.al., 2003).

1.2 Main results

In Fig. 2 the seasonal cumulated precipitation (observed and forecasted) are exhibited.

From the maps of Fig. 2 we can derive different results:

- globally the best performances are in fall 2003, fall 2004 and over the plains;
- better performances of LM-DWD in summer (jja 2003 especially) with respect to the others, presumably because of the soil moisture external analysis (it has to be pointed out that in summer 2003 all the versions used the I.C. and B.C. from GME-DWD);
- there is a constant overestimation of LAMI and aLMo over the North-Western Alpine region of Piedmont;
- LM-DWD overestimated over the North-Western Alpine region too, but not in summer 2003 and not after the introduction of the prognostic rain scheme (summer 2004 and fall 2004);
- in summer 2004 aLMo dramatically overestimated.

In addition we can look at the seasonal model performances using statistical indices. In Fig. and the three versions are compared season by season in terms of BIAS score and Equitable Threat Score for a fixed threshold of $10mm/24h$. We consider the average precipitation values over the hydrogeological basins shown in Fig. 1. The error bars in Fig. and indicate 2.5th and 97.5th percentiles of resampled distribution, applied to the "reference" model. Note that the bootstrap technique is symmetric: the "reference" model could be the "competitor" model and viceversa. The score difference is statistically significant if it lies outside a given confidence interval (95% in this case) on the PDF produced by means of a bootstrap resampling technique.

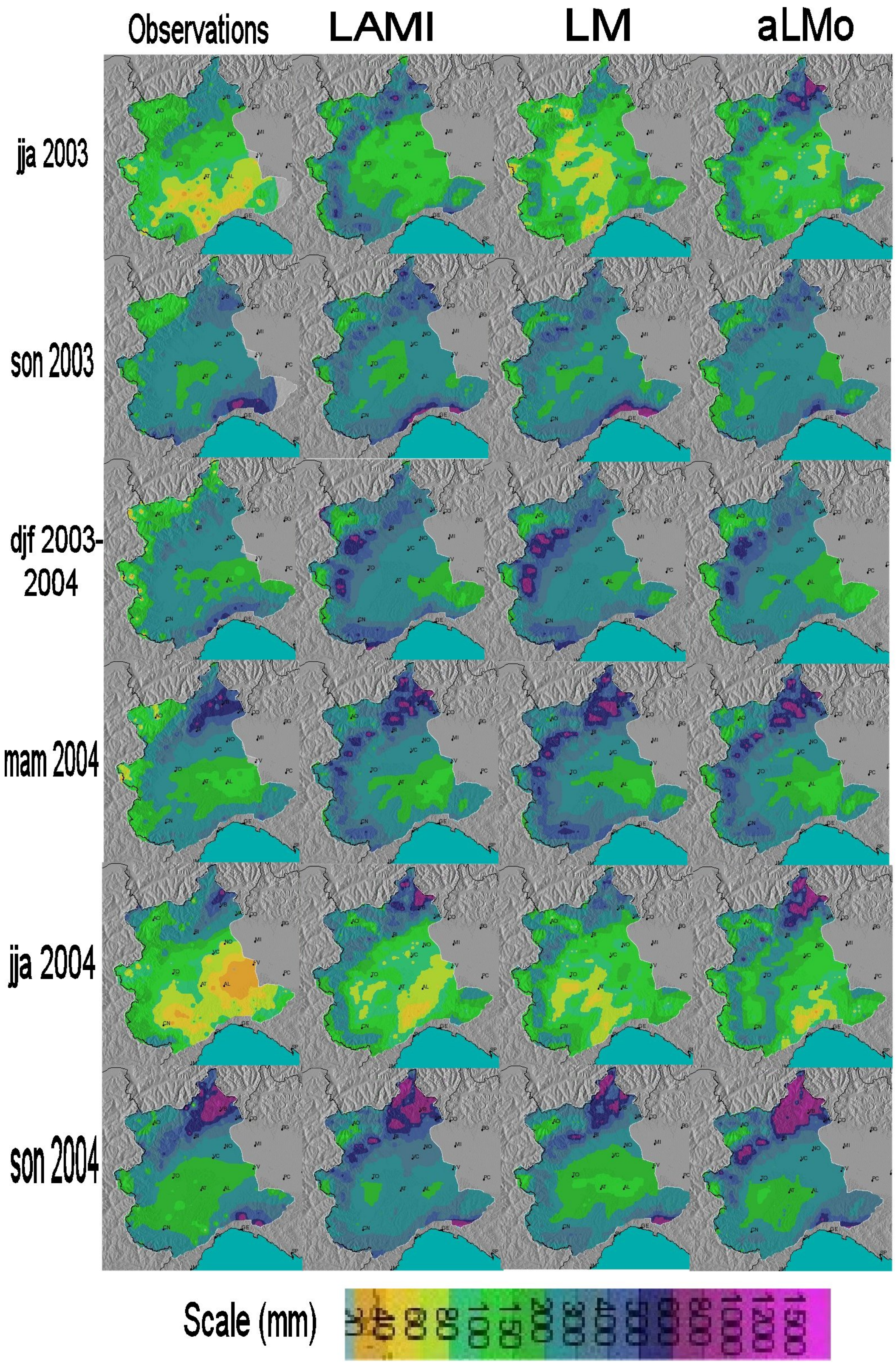

Figure 2: Comparison between observed and forecasted precipitation for each season in the period 200306-200411

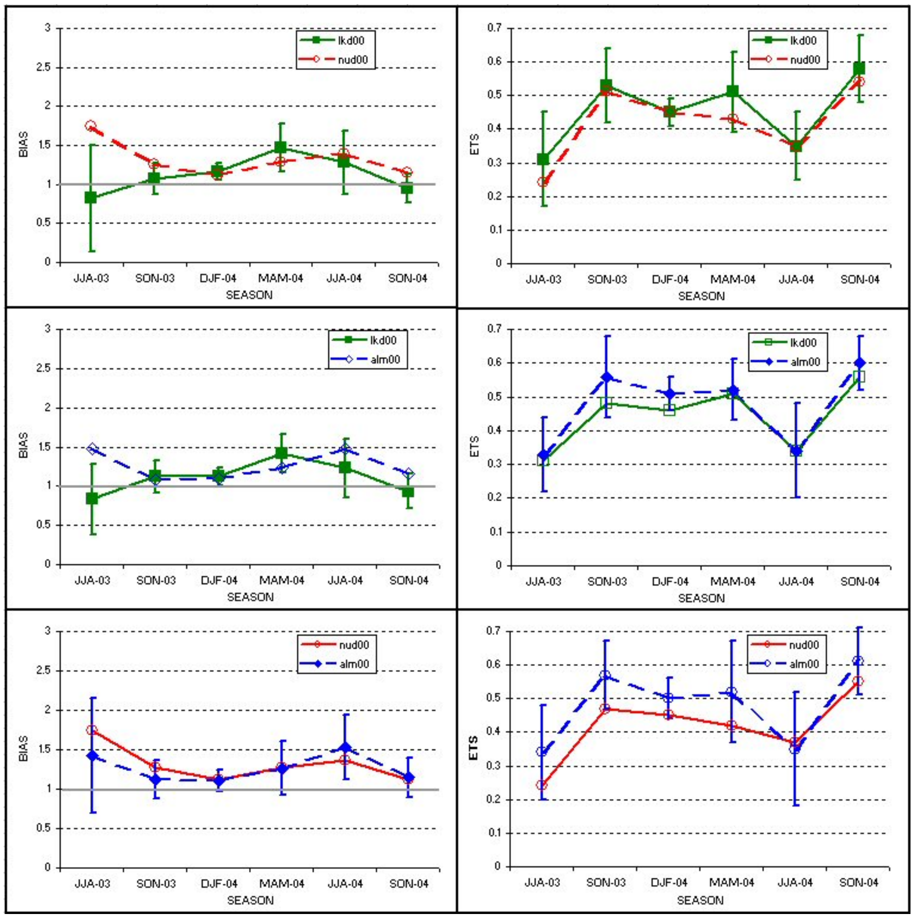

Figure 3: Seasonal trend of BIAS and ETS in the period 200306-200411 – Threshold=$10mm/24h$ – Forecast time: $00/+24$. The error bars indicate the 2.5th and the 97.5th percentiles of resampled distribution, applied to the "reference" model.

In the first 24 hours (day 1) of forecast time we note that:

- generally the model tends to overestimate the precipitation;
- globally the worst performances are in summer;
- also in this case we confirm the behavior of LM-DWD in summer 2003, presumably because of the soil moisture external analysis;
- LM-DWD recorded the best BIAS score also in fall 2004, this can be due to the presence of the prognostic rain scheme (aLMo introduced it only at the end of November 2004);
- there are no significant differences in ETS plots, with the exception of aLMo superior ETS values in some seasons (son-2003, djf-2004).

Regarding the second 24 hours (day 2) of forecast time, the results are:

- globally, as seen for the first 24 hours of forecast time, the worst performances are in summer;

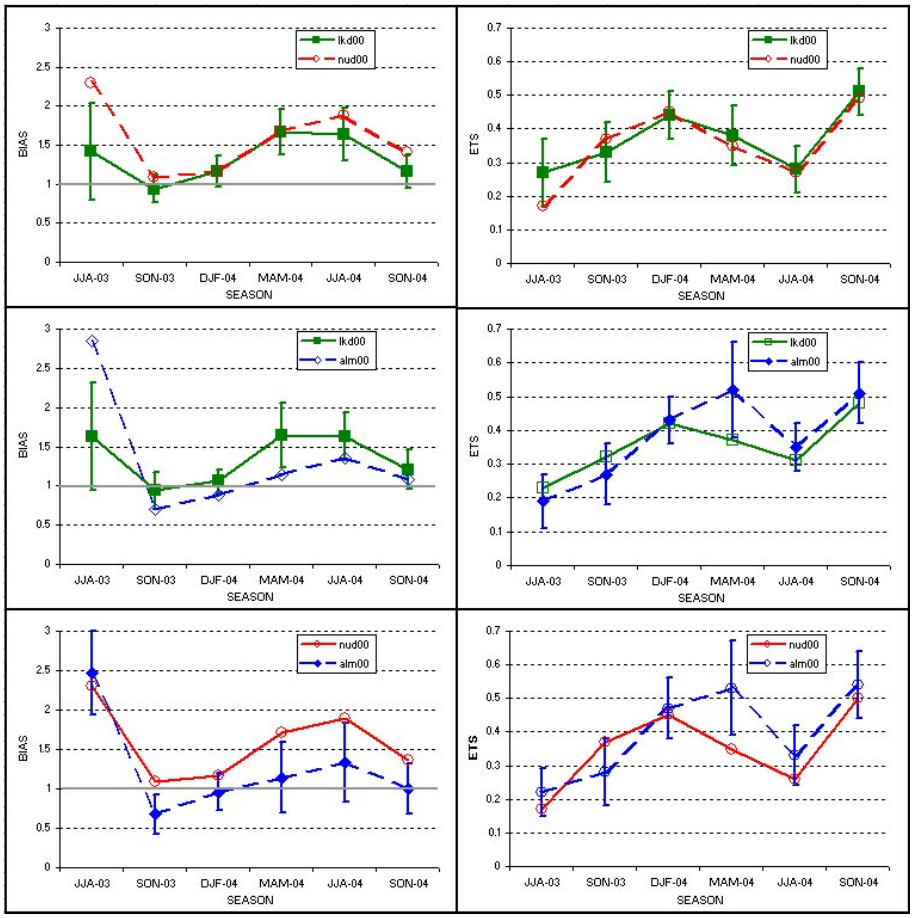

Figure 4: Seasonal trend of BIAS and ETS in the period 200306-200411 – Threshold=$10mm/24h$ – Forecast time: $+24/+48$. The error bars indicate the 2.5th and the 97.5th percentiles of resampled distribution, applied to the "reference" model.

- aLMo gives results better than the others in terms of BIAS score in all the seasons with the exception of summer 2003, this may be due to the IFS initial and boundary conditions;
- in summer 2003 LM-DWD showed the best scores;
- no significant differences in the ETS comparison, except for the superiority of aLMo in spring 2004.

2 Comparative high resolution verification over Northern Italy

2.1 The method

The aim of this section is to compare the performance of the three model versions (aLMo, LAMI, LM-DWD) over a common domain composed by 47 meteo-hydrological basins (see Fig. 5) of Northern Italy, in which fall 1023 station points (see Fig. 6). Further, we summarize the main results obtained during the last two years, presenting only the most representative and significant outcomes. We take into account a common period in which

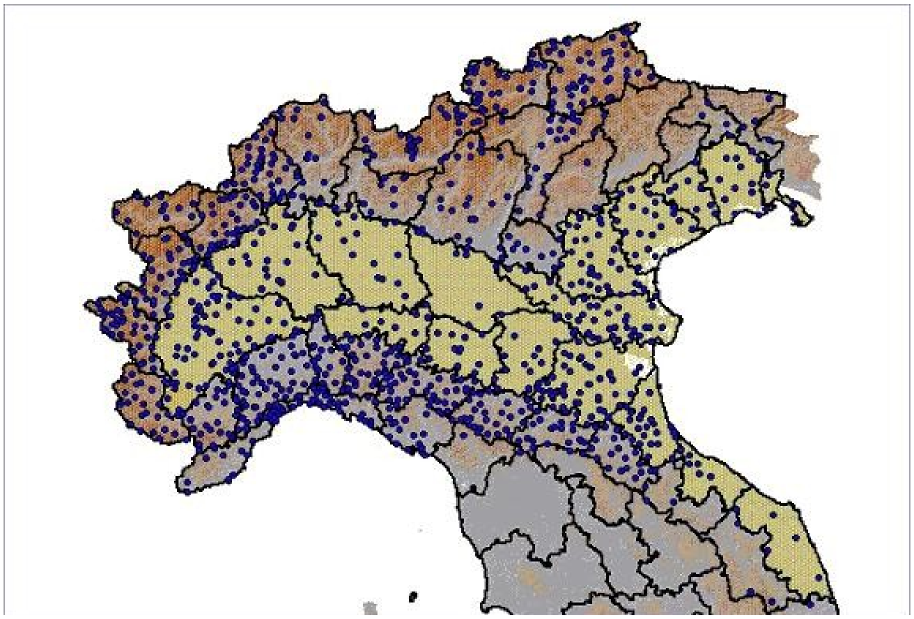

Figure 5: Map of Northern Italy: 47 meteo-hydrological basins, 30 brown areas representing mountainous basins, 17 yellow areas representing plain basins.

REGION	AVAILABILITY PERIOD	N. OF STAZIONS	AREA (km2)	RESOLUTION (km)	ORIGIN	USE
Piemonte + Valle D'Aosta	Jan '01- Dec '04	352	28700	9	Cosmo Project	used
Liguria	Jan '01- Dec '04	76	5400	8	Cosmo Project	used
Prov. Auton.	Jan '01- Dec '04	88	7400	9	Cosmo Project	used
Prov. Auton.	Jan '01- Dec '04	20	6200	18	Cosmo Project	used
Marche	Jan '01- Dec '04	12	9700	2	Cosmo Project	used
Veneto	Jan '02- Dec '04	89	18000	14	Cosmo Project	used
Friuli	Jan '02- Dec '04	27	7800	17	Cosmo Project	used
Emilia Romagna	Dec '00- May '03	42	22000	23	Cosmo Project	used
	Jun '03- Dec '04	271		9	Cosmo Project	used
Lombardia	Jan'02- Nov '04	88	23900	16	Civil Protection	used
Sardegna	Jan '02- Dec '04	49	24000	22	Cosmo Project	not used/out of common domain

Figure 6: Available Italian dataset table.

all the model versions are available, from June 2003 to December 2004, where we have got a complete dataset for 00UTC runs. Regarding to 12UTC runs, we hold a comparable period of data only for LM-DWD, so we will focus our efforts on 00UTC runs. We calculate skills and scores considering 24h cumulated precipitation averaged over basins for several precipitation thresholds, for +24h and +48h forecast time. In addition, in order to test the model sensitivity over complex terrain and the capability to distinguish a complex orography, we subdivide the 47 meteo-hydrological basins into two big subset, mountain (30 meteo-hydrological basins, Fig. 5 brown areas) and plain (17 meteo-hydrological basins, Fig. 5 yellow areas) areas. Therefore we estimate skill and scores considering 24h cumulated precipitation averaged over these different kind of basins for several precipitation thresholds, for +24h and +48h forecast time. Moreover, in order to evaluate the model changes and improvements, we show seasonal scores with a fixed chosen threshold (5 mm) considering 6h cumulated precipitation averaged over basins.

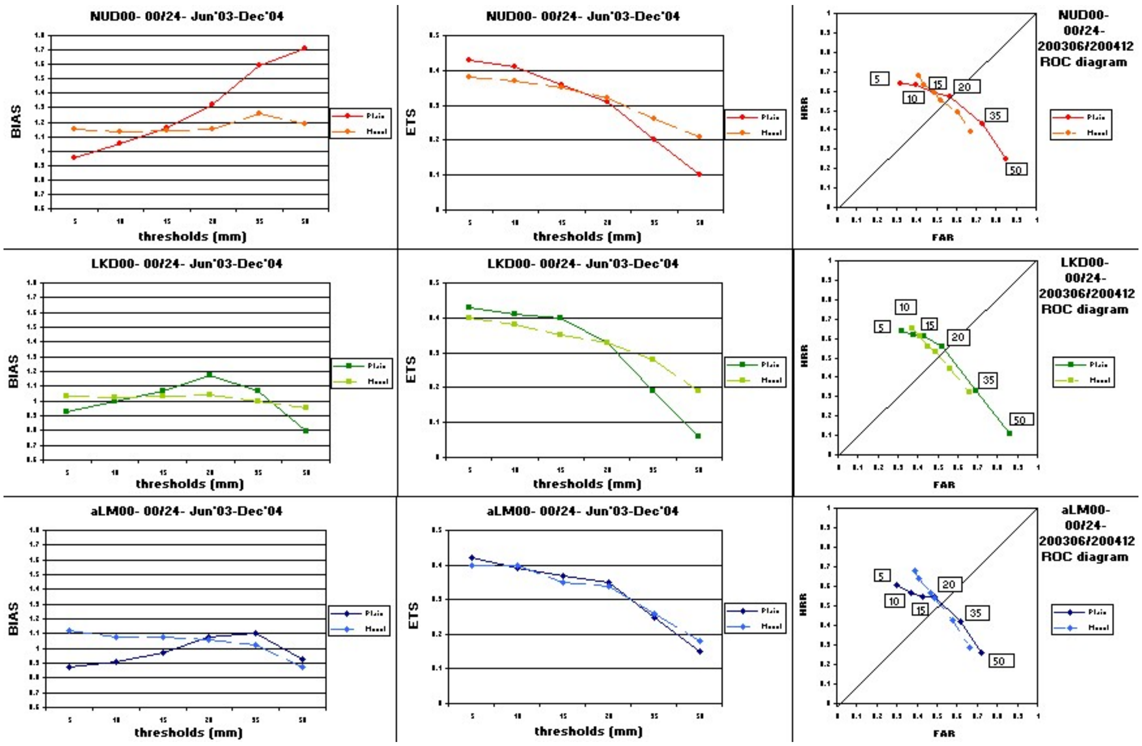

Figure 7: Skill and Scores for 24h average cumulated precipitation over mountain and plain basins. Runs 00UTC +24h forecast time. Jun'03-Dec'04.

2.2 Main results

First of all we want to analyze the three model version performance over mountainous and flat basins, with respect to 24h cumulated average precipitation, for +24h (see Fig. 7) and +48h (see Fig. 8) forecast time. We can deduce some main features concerning to the three versions behavior, using statistical indices as BIAS, ETS and ROC diagram:

- the BIAS for small precipitation amount on plain areas is greater than on mountainous areas;
- this trend seem to be inverted as thresholds increase: in fact we generally obtain a greater BIAS on plain than on mountain for high thresholds;
- in particular, with regard to the first 24h, we find a BIAS on mountain more or less constant with respect to threshold and around 1;
- on the other hand, we derive better scores (see ETS and ROC diagram) on plain areas than mountain areas up to 10/15mm, above these thresholds the trend is inverted.

To summarize, we note a different behavior in the model performance over mountain and plain zones: in the first case there is probably a balance of the up-wind/down-wind orographic effect that seems to compensate the underestimation/overestimation systematic error; in the second case the results are worse in term of BIAS especially for medium/high thresholds, but better in term of ETS for low/medium thresholds. In addition, according to the comparative study among the three model versions, we observe a fairly similar scores in term of ETS and ROC diagram, whereas we obtain a different results for BIAS skills, in fact LAMI model tends to overestimate more than the others particularly for high thresholds.

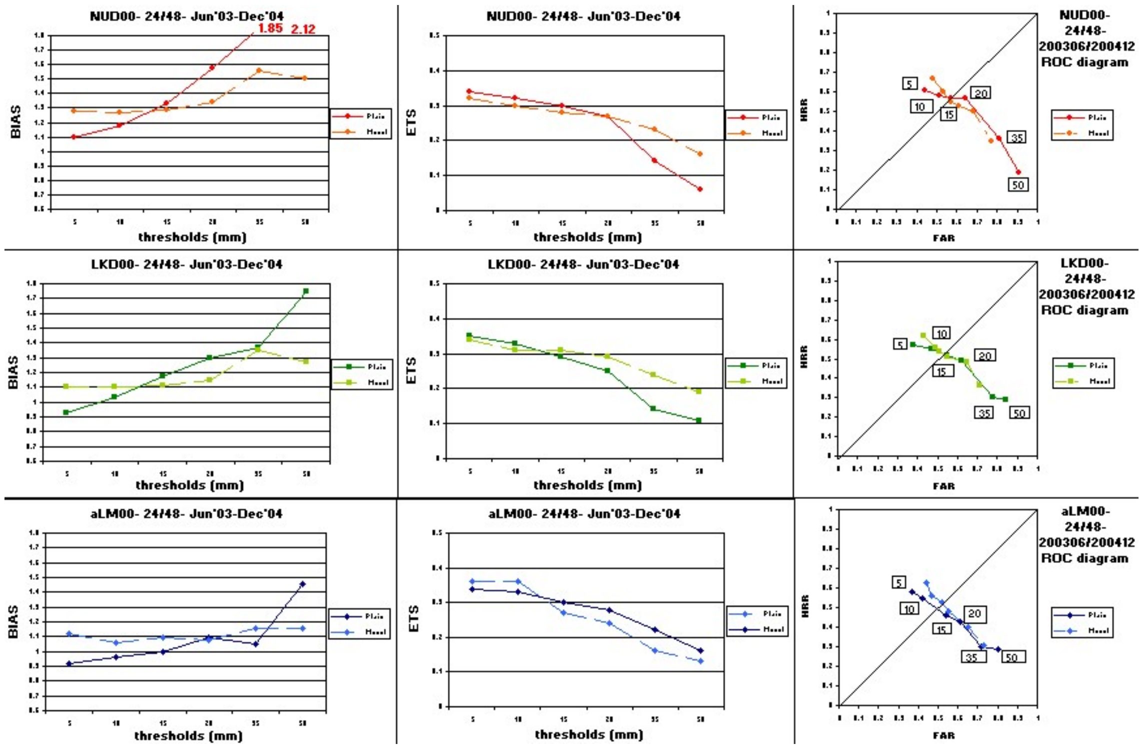

Figure 8: Skill and Scores for 24h average cumulated precipitation over mountain and plain basins. Runs 00UTC +48h forecast time. Jun'03-Dec'04.

Concerning the daily cycle and seasonal trend, we present BIAS and ETS graphs considering 6h cumulated precipitation averaged over basins for a fixed precipitation thresholds (5mm) to investigate the three versions behavior. Some differences exist (see Fig. 9). In particular, regarding to summer 2003, that was an exceptional dry season, we note some remarkable features:

- a similar trend for all the models, with a pronounced diurnal cycle, with a big overestimation in the morning and underestimation around nocturnal hours (this characteristics can be found also in JJA2004);
- during the strong convection peaks, LM-DWD seems to perform better than the others, probably because of the soil moisture analysis, that plays a basic role during summer seasons;
- a very bad ETS, maybe due to a low models capability to localize and predict accurately the precipitation pattern when it is larger the convective contribution.

Further, we note some differences among the model versions initialized by GME and aLMo, which has boundary condition from ECMWF (IFS-frames) since September 2003. In fact aLMo gives the lowest BIAS index for fall 2003, winter and spring 2004, and the highest ETS (at least from winter 2004). It has to be pointed out that LM-DWD BIAS behavior tends to be similar to aLMo behavior during the last two considered season (that is after the introduction of the prognostic rain scheme).

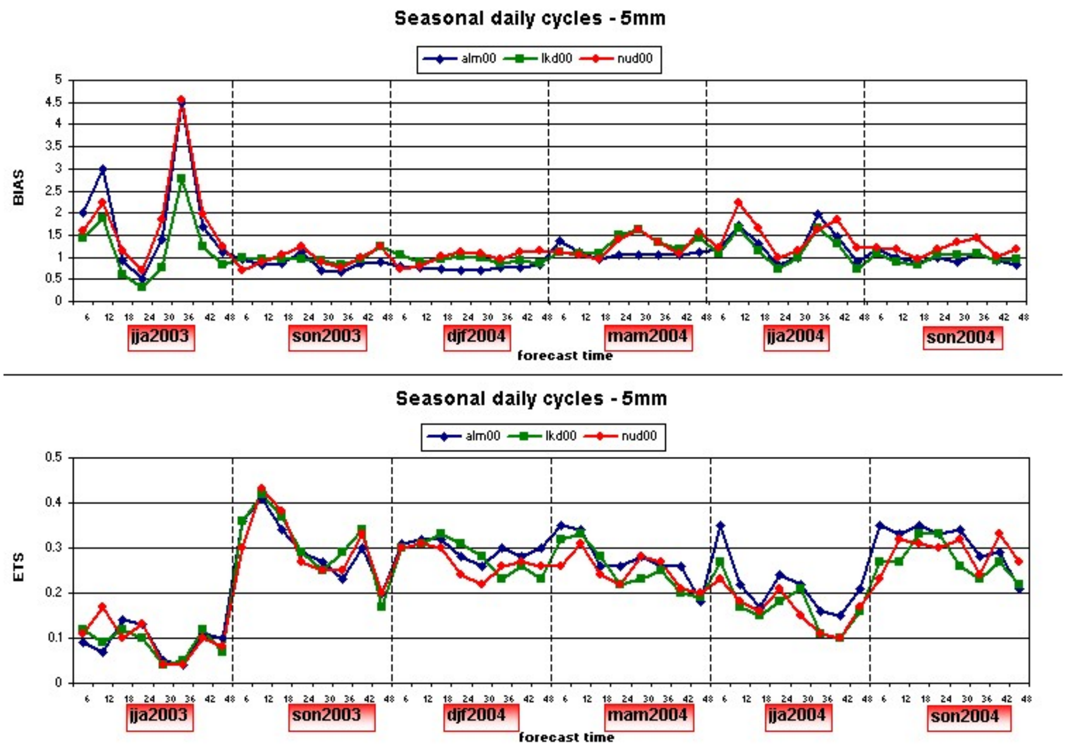

Figure 9: Seasonal daily cycles for 6h average cumulated precipitation. Runs 00UTC. Jun'03-Nov'04. Threshold=5mm

3 General conclusions

The main goal of this report is to compare the performance of the three model versions (aLMo, LAMI, LM-DWD). We consider two common domain (Northern Italy and Piedmont) and a period in which all the model versions are available (from June 2003 to December 2004). The main outcomes are:

- a strong skill dependence on the model changes. In particular aLMo has a trend different from the others, may be due to the IFS initial and boundary conditions. Considering the second day forecast for Piedmont, aLMo gives the best score for all the seasons except for summer 2003;
- LM-DWD gives the best scores in summer 2003 presumably because of the soil moisture external analysis. Good results for the last two seasons considered probably due to the prognostic rain scheme;
- regarding Northern Italy there is a systematic QPF difference among plain and mountain areas; about the eye-ball verification, we note an overestimation of LAMI and aLMo over the North-Western Alpine region of Piedmont, while over the whole mountainous area the results seem to be better, but this may be due to a balance between upwind and downwind effect.

References

Accadia C., Casaioli M., Mariani S., Lavagnini A., Speranza A., De Venere A., Inghilesi R., Ferretti R., Paolucci T., Cesari D., Patruno P., Boni G., Bovo S., Cremonini R., 2003.

Application of a statistical methodology for limited area model intercomparison using a bootstrap technique. *Il Nuovo Cimento*, Vol. 26, Issue 01, p.61, 2003.

Hamill, T.M., 1999. Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, 14, 155-167, 1999.