QPF verification for 2007: COSMO-I7, COSMO-7, COSMO-EU, COSMO-ME, COSMO-I2, COSMO-IT.

Elena Oberto, Massimo Milelli

ARPA Piemonte, Via Pio VII 9, I-10135 Torino, Italy

1 Introduction

In this report we present the QPF verification of 4 model versions at 7 km resolution (COSMO-I7, COSMO-7, COSMO-EU, COSMO-ME) and 2 model versions at 2.8 km resolution (COSMO-I2, COSMO-IT). The observations come from the high resolution network of rain gauges of the COSMO dataset and Civil Protection Department, i.e. about 1300 stations. We show an update of the most recent results highlighting the failings and improvements of the model: the skills and scores are calculated considering 6h or 24h averaged cumulated observated/forecasted precipitation value over 90 meteo-hydrological basins that cover all the peninsula with the exception of two Southern regions: Sicilia and Puglia (see Fig. 1). In detail, we show:

- scores over a long period (seasonal trend);
- verification over last year (200706-200806).

Figure 1: Italian high resolution raingauge distribution.

2 Long period verification

We evaluated the seasonal skills (BIAS, ETS, FAR, POD) for 0.2, 2, 10, 20 mm/24h for the first and the second day starting from DJF 2004 to MAM 2008 for COSMO-I7 and COSMO-7, from MAM 2006 to MAM 2008 for COSMO-EU and from DJF 2006 to MAM 2008 for COSMO-ME. The reason for this discrepancy is the different availability of the data. In the following pictures only BIAS (see Fig. 2 left) and POD (see Fig. 2 right) for 20 mm/24h are shown and we can suggest the following remarks:

- there is a certain periodicity that shows in general that fall and summer have the best and worst skills respectively;
- the error periodicity is usually amplified with the increasing of the threshold;
- BIAS and FAR reduction and ETS increase have been achieved over a long period;
- POD obtains globally contradictory results whit general worsening for low thresholds and improvement for high thresholds, but only for COSMO-7 and COSMO-I7;
- COSMO-I7 and COSMO-7 have a similar error behavior, especially during last period, with a general worsening of the performance;
- COSMO-EU and COSMO-ME show similar trends and performances.

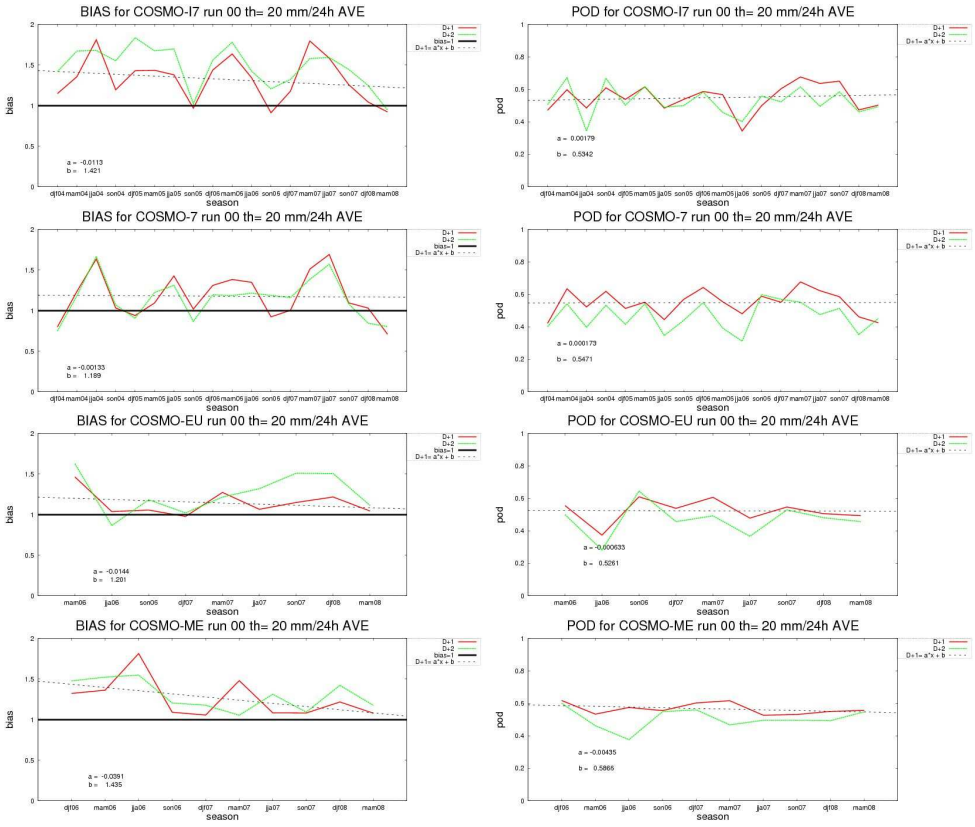

Figure 2: Seasonal trends (dashed line) for the various model versions, starting from DJF 2004 to MAM 2008 for 20 mm/24h threshold: D+1 (red) and D+2 (green) respectively. BIAS on the left side and POD on the right side.

3 Verification over last year

A focus on the verification over last year is reported here, in the second part of the work. Starting from June 2007 to June 2008 we analyze the feature and performance of the 6 model versions, 4 at 7 km resolution and 2 at 2 km resolution. A scheme to summarize the meaning of the main statistical indices used for categorical forecasts is shown in Fig. 3: we add a fifth index that takes into account the correct negative term in the contingency table and it represents the *specificity*, that is the percentage of correctly forecasted not-event.

		FORECASTED	
		NO	YES
OBSERVED	NO	A	B
	YES	C	D

MEANING	INDEX	RANGE	IDEAL VALUE
Overestimation/ underestimation	$BIAS = \frac{B+D}{C+D}$	[0, ∞]	1
% correctly forecasted events	$POD = \frac{D}{C+D}$	[0,1]	1
False alarme ratio	$FAR = \frac{B}{B+D}$	[0,1]	0
Eq. Threat Scores	ETS	[-1/3,1]	1
% correctly forecasted not-events (specificity)	$1\text{-}POFD = \frac{A}{B+A}$	[0,1]	1

Figure 3: Indices table.

A comparison between COSMO-7/COSMO-EU is reported in Fig. 4, with the five statistical indices and the number of cases, using a bootstrap technique developed by Hamill: in order to compare two model versions, it is necessary to have a confidence interval in order to assess the real differences between skills and scores. We consider the average precipitation values over each meteo-hydrological basin and the error bars indicate the 2.5th and 97.5th percentiles of a resampled distribution, applied to the *reference* model.

There are no significant differences between the two versions with the exception of a slightly improvement on average for COSMO-7. But, if we plot the season indices we can appreciate some strong difference (see Fig. 5):

- COSMO-EU seems to have a quite stable overestimation and the other indices do not show a strong periodicity;
- COSMO-7 shows a strong decreasing of BIAS and POD: the BIAS has a very big overestimation during last summer and a big underestimation during last spring. POD has a worsening during the latest two season.

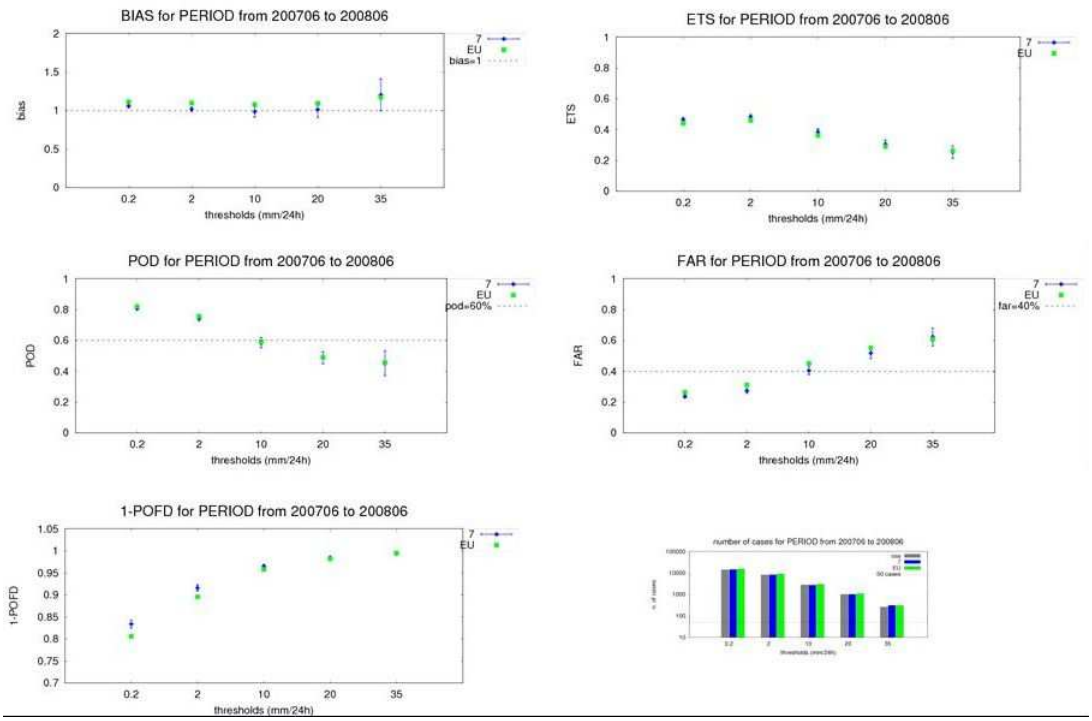

Figure 4: COSMO-EU/COSMO-7: BIAS, ETS, POD, FAR and number of cases during 200706-200806.

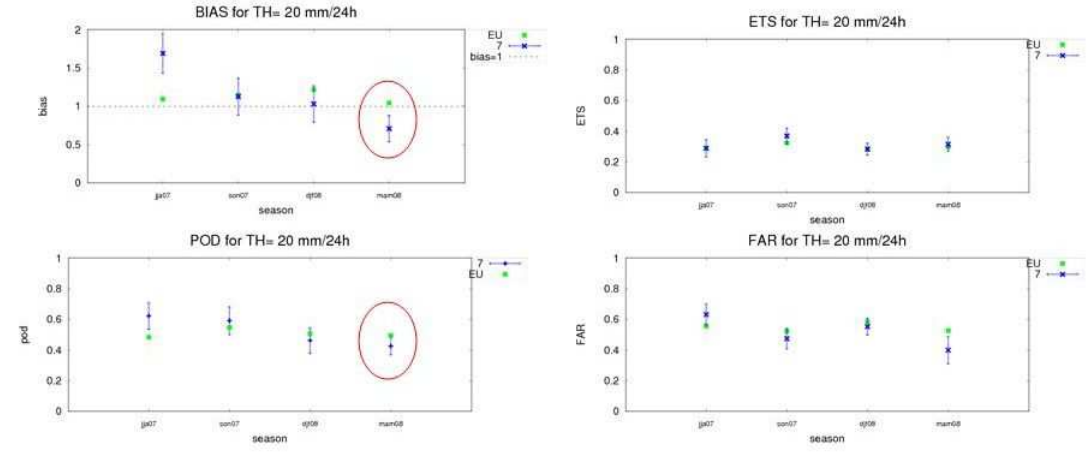

Figure 5: COSMO-EU/COSMO-7: BIAS, ETS, POD, FAR seasonal trend during last year.

Similarly, a comparison between COSMO-I7/COSMO-ME over the whole period is not so significant but, some peculiarity is noticeable if we look at the seasonal indices (see Fig. 6):

- COSMO-ME seems to have a behavior similar to COSMO-EU, with a quite stable overestimation without periodicity;
- in addition, COSMO-I7 seems to have a behavior similar to COSMO-7, with strong BIAS and POD decrease and with a general worsening during the latest seasons.

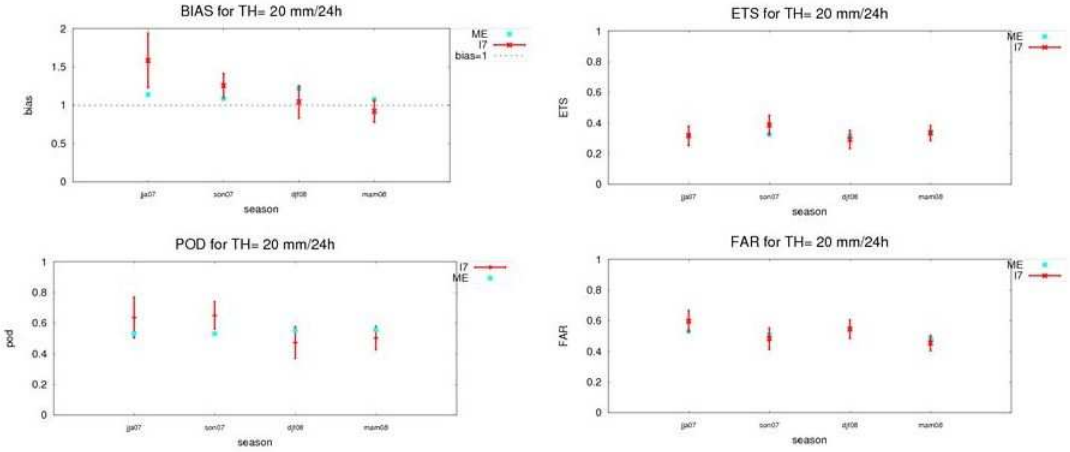

Figure 6: COSMO-I7/COSMO-ME: BIAS, ETS, POD, FAR seasonal trend during last year.

Now we present the results for 2.8 km model versions over a complete period (1 year), with a comparison between COSMO-I7/COSMO-I2, the Italian model versions (7 km and 2.8 km) run in Bologna, and COSMO-ME/COSMO-I2, the Italian model version (7 km and 2.8 km) run in Rome. The COSMO-I7 and COSMO-I2 seasonal error is shown in Fig. 7, where COSMO-I2 seems to follow COSMO-I7 trend, with a strong positive bias last summer (probably because the deep convection parameter was on at that time) and a negative bias this spring.

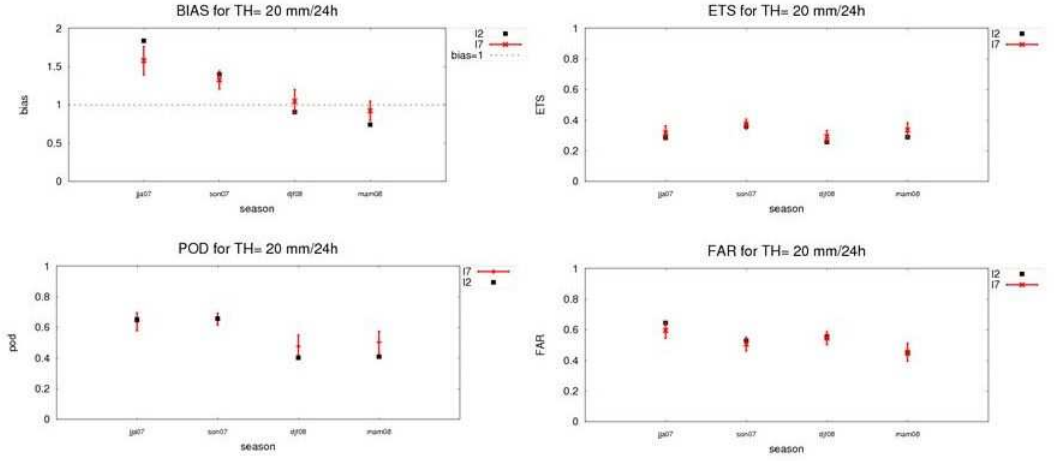

Figure 7: BIAS, ETS, POD, FAR seasonal trend during last year for COSMO-I7/COSMO-I2.

Concerning to the comparison between COSMO-ME/COSMO-IT, as we note in Fig. 8, the two models seem to have a similar trend with a BIAS greater than one.

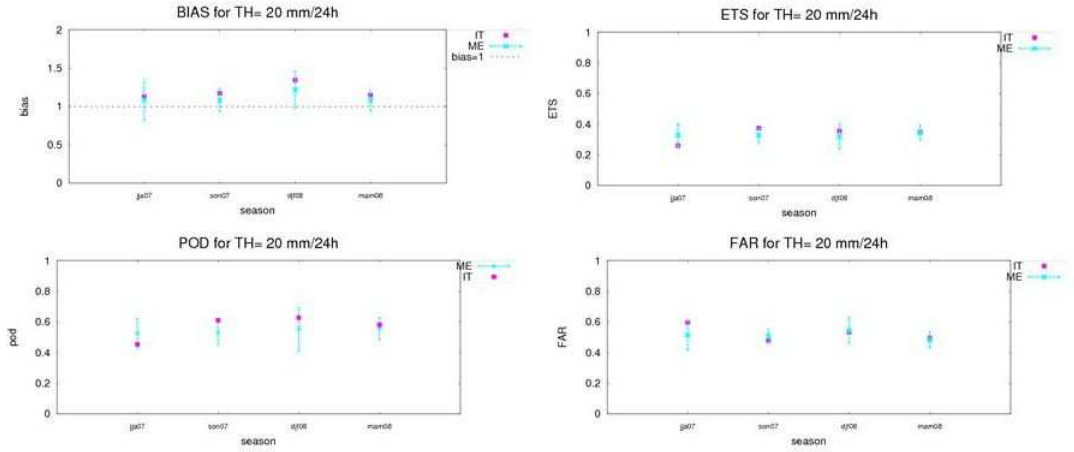

Figure 8: BIAS, ETS, POD, FAR seasonal trend during last year for COSMO-ME/COSMO-IT.

Finally, we show in Fig. 9 some results concerning the diurnal cycle: we considered 6h averaged cumulated observated/forecasted precipitation from June 2007 to June 2008 for all the six model versions with different thresholds, starting from 0.2 mm up to 10 mm. For low thresholds all the models present a bias overestimation peak at midday, while the best value peak for POD, ETS and FAR occurs around the afternoon. Moreover, it is noticeable a spin-up problem for all the models especially for COSMO-I7 and COSMO-I2 that seems to disappear for higher thresholds.

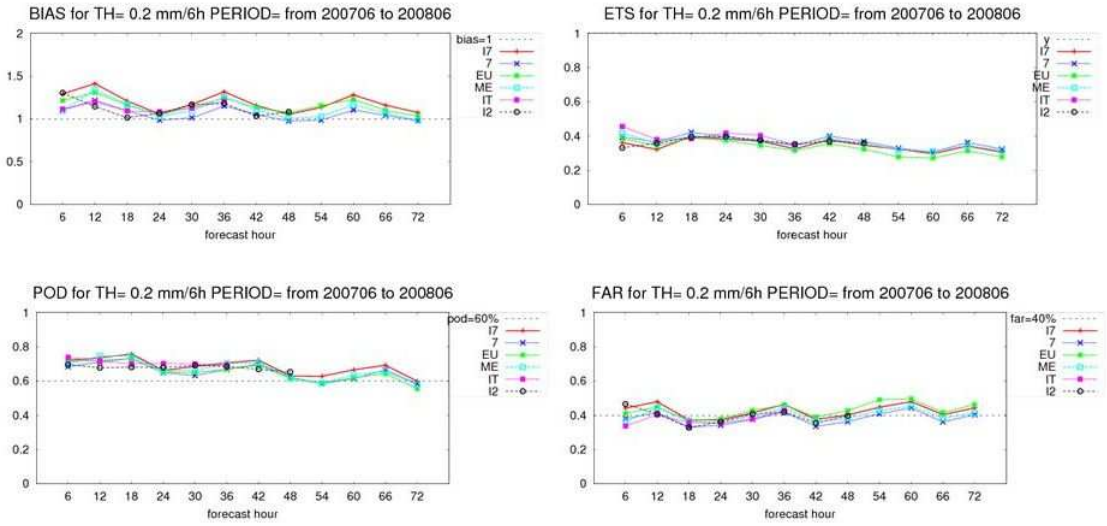

Figure 9: BIAS, ETS, POD, FAR for 6h cumulated precipitation of the six model versions.