# QPF verification of Italian COSMO-models using different parameters of the precipitation distribution

Maria Stefania Tesini

*Hydro meteorological Regional Service ARPASIMC, Bologna, Italy*
*email: mstesini@arpa.emr.it*

## 1 Introduction

The interpretation of the verification results of precipitation forecasts is often a difficult task because the choice of the verification methods, the reference thresholds or the statistics used to quantify the performance highlights different aspects of the numerical forecast.

In this work the ability of models to represent the shape of the precipitation field within a predefined geographical area, in terms of frequency of exceeding a certain threshold of rain and not in a spatial sense, has been investigated.

The verification methodology applied consists in the comparison of forecast and observation in terms of same parameters of their statistical distribution, evaluated after the precipitation values are aggregated over predefined geographical areas, representing the Italian warning areas for hydro-meteorological events.

Some Italian versions of the COSMO model (COSMO-I7, COSMO-I2, COSMO-ME) and the global model IFS-ECMWF were considered in this study.

The results may provide an interpretation key to a better use of models QPF, especially with respect to high rainfall events.

## 2 Dataset

Observed precipitation data consisted of more than 1500 rain-gauges made available by the Italian National Department of Civil Protection. The dataset cover almost all the Italian peninsula, (see figure 1(a) ) even if it is not homogeneous both in space and time.

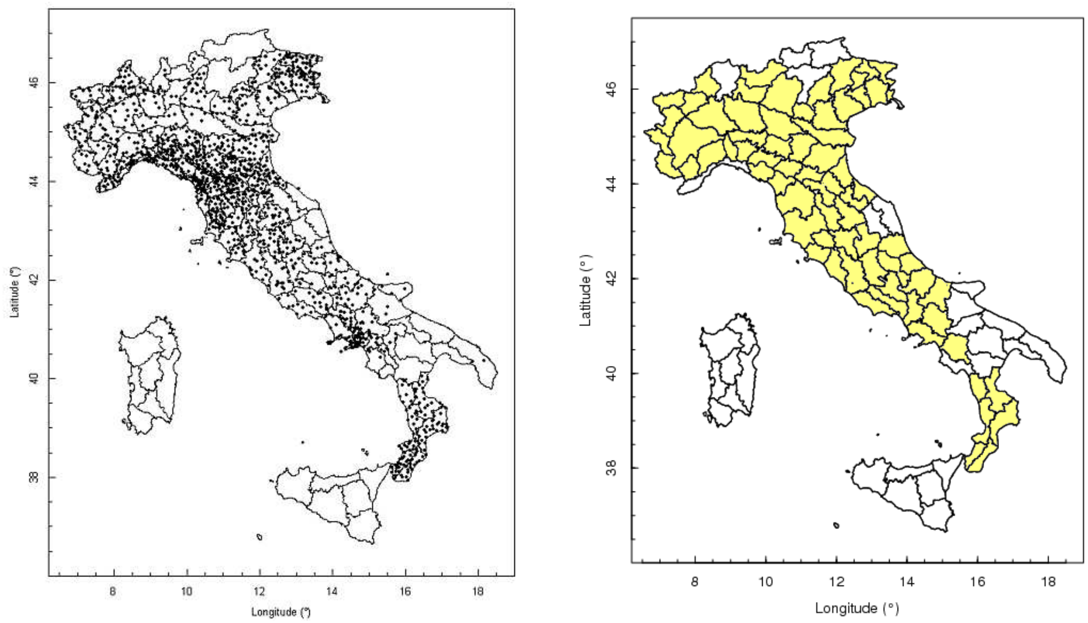
In this work we considered only precipitation values accumulated over 24h, starting at 00 UTC from March 2010 to April 2011.

The Italian implementations of the COSMO models involved in the verification are:

- COSMO-I7 ( 7 Km horizontal resolution performed at Arpa-SIMC)
- COSMO-ME (7 Km horizontal resolution operated at CNMCA)
- COSMO-I2 ( 2.8 Km horizontal resolution performed at Arpa-SIMC)

The global model IFS-ECMWF( about 16 Km of horizontal resolution), providing the boundary conditions for both the 7 Km COSMO-Models, is also taken into account in this work as a term of comparison to assess the added-value of the higher resolution models.

For all the models the the 24 hours accumulated precipitation at +24h, +48h and +72h for the 00 UTC run has been considered ( for COSMO-I2 only +24h and +48h ).




(a) Available rain-gauges

(b) Verification areas

Figure 1: Location of available rain-gauges in Italy and geographical areas division

## 3 Methodology

The verification domain has been divided into about one hundred geographical areas, representing the warning areas of the Italian Civil Protection (see figure 1(b)).

For each day of the verification period and for each area, the maximum value and the median of all the model grid-points that fall in the area, have been computed. Corresponding values have been evaluated also from the observations.

Verification was then performed using a categorical approach: the "yes-no" event was defined according to the condition that the maxima and/or the medians of forecast and observed precipitation distribution exceed a preselected threshold (e.g. maximum greater than 25 mm and median greater than 5 mm in the 24 hours period). Usual quality measures such as Probabilty of Detection, False Alarm Ratio,Threat Score (also known as Critical Success Index) and Bias Score have been derived from the entries of the 2x2 contingency table.

The representativeness of observational data set must be taken into consideration when interpreting results. In fact the choice of the median implies that at least on half of the area the QPF had a value above the reference, regardless of the resolution of the model(even though the points may not be contiguous), while for the observed value this was not always true because the stations are not uniformly distributed over the area. Even when considering the maximum, it is important to remind that the density of stations on the territory was not homogeneous and higher precipitation values could be missed.

## 4 Results

The results are summarized using a particular type of graphic, the Performance Diagram (Robber,2009), in which it is exploited the geometric relationship between four measures of dichotomous forecast performance: probability of detection (POD), the success ratio (SR, defined as 1-FAR), bias score and threat score (TS, also known as the Critical Success Index). For good forecasts, POD, SR, bias and TS approach unity, such that a perfect forecast lies in the upper right of the diagram. Deviations in a particular direction will indicate relative differences in POD and SR, and consequently bias and TS. An immediate visualization of differences in performance are thus obtained. The influence of sampling variability is estimated using a form of resampling with replacement bootstrapping from the verification data. The 95th percentile range for SR and POD are plotted as "cross-hairs" about the verification point and the variation in bias and CSI is simultaneously displayed. One thousand new samples of the same size as the original are created using the sampling frequencies of observed and forecast "yes" and "no" entries (i.e. the marginal frequencies), and the 25th and 975th accuracy measures are computed from these "climatological" samples to generate the 95th percentile range (description retrieved from the website of WWRP/WGNE Joint Working Group on Forecast Verification Research *http://www.cawcr.gov.au/projects/verification/Roebber/PerformanceDiagram.html*).

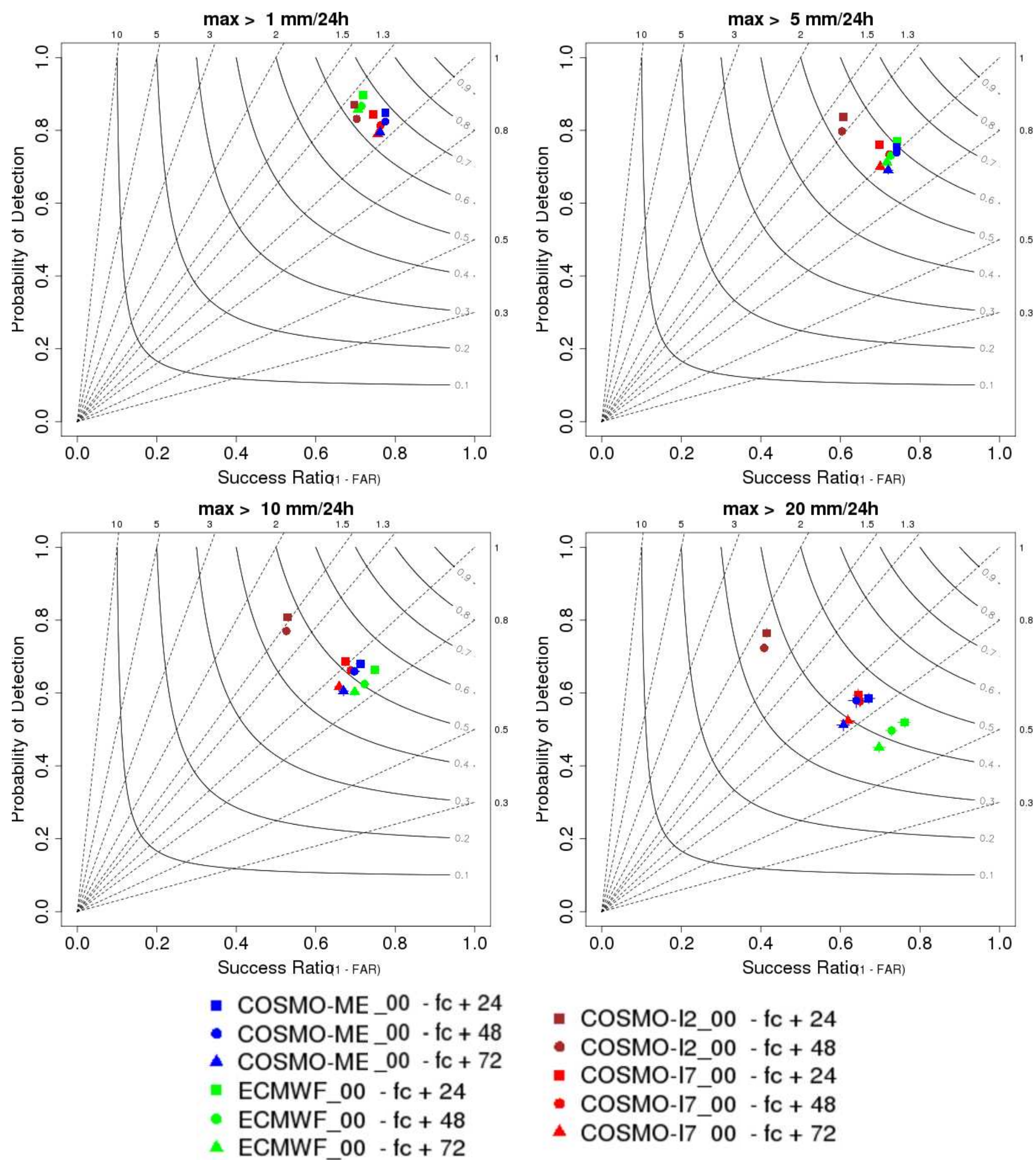



Figure 2: Performance diagrams for the event "maximum value greater than reference thresholds of 1,5,10,20 mm/24h". In the performance diagram are summarized the Success Ratio (1-FAR)in the x-axis, POD in the y-axis. Dashed lines represent Bias Scores with labels on the outward extension of the line, while labelled solid contours are TS. Sampling uncertainty is given by the "cross-hairs", not very visible in this case because the sampling variability is small.

In figure 2 are shown the performance results of models for the event "maximum value exceeding the reference threshold" for the threshold 1,5,10 and 20 mm in the 24 hours period. At the lower threshold, that means that at least in one point of the area the rain is greater than 1 mm/24h, the threat score of all models is between 0.6 and 0.7, which is definitely a good score, but they show a tendency to overforecast the event. The tendency is more pronounced for IFS-ECMWF and COSMO-I2, which present also a higher number of false alarm, despite a higher POD. Increasing the reference threshold the performances change: ECMWF and the 7 km COSMO models reduce the BS, reaching an unbiased condition, while COSMO-I2 increases the overestimation of the events. The POD of COSMO-I2 is the higher, but the SR decrease. Scores of other models are slightly reduced, although still good, in particular for COSMO-ME. A further increase in the reference threshold to 10 mm/24h and 20 mm/24h shows a general worsening in the POD and TS, but while COSMO-I2 further increases the BS and the number of false alarm, the other models tends to underforecast the event, especially the global model. The difference between the 7 km COSMO models and IFS-ECMWF becomes more pronounced with the increase of the threshold, indicating a greater difficulty for the global model in reproducing relatively high rainfall events.

In figure 3 are shown the results obtained by requiring that two conditions are simultaneously verified. More precisely, the event was defined as "maximum above 25 mm/24h when the median is greater than a predetermined reference threshold (e.g. 1,5,10,20 mm/24h). The required condition implies that the rain exceeds 25 mm/24h at least in one grid-point/station but also that the precipitation exceeds the specified threshold in half of the grid-points/stations of the area. In this way, taking into account the problems of representativeness of the observational dataset, as previously mentioned, the ability of the models to reproduce some feature of the precipitation distribution over a region (in terms of quantity and not from a spatial point of view) has been investigated. The most salient aspect is that gradually increasing the threshold of the condition on the median, the BS tends to move closer to 1. COSMO-I2 reduces the overestimation of the events while the other models reduce the underestimation of the events, especially IFS-ECMWF. The threat score of the COSMO-I7, COSMO-ME and IFS-ECMWF does not undergo specific changes up to the 10 mm/24h threshold, maintaining approximately the same value with the increase of the reference threshold for the condition on the median, even if improvements in the POD are compensated by deterioration in SR. In the same cases COSMO-I2 reaches TS comparable to that of the other models improving the SR while reducing the POD. It therefore seems that models generally improve the performance in the identification of maximum precipitation when precipitation is spread over much of the area considered. Analyzing the condition of "maximum and median greater than 25 mm/24h mm/24h greater than 20" (bottom right graph in figure 3) we are in a different situation: COSMO-I2 tends to underforecast events, the scores of all the COSMO model have slightly worsened while IFS-ECMWF behaves better than the other models. So, in the case relatively high and widespread precipitation, higher-resolution models do not seem to add particular value in respect to the global model.

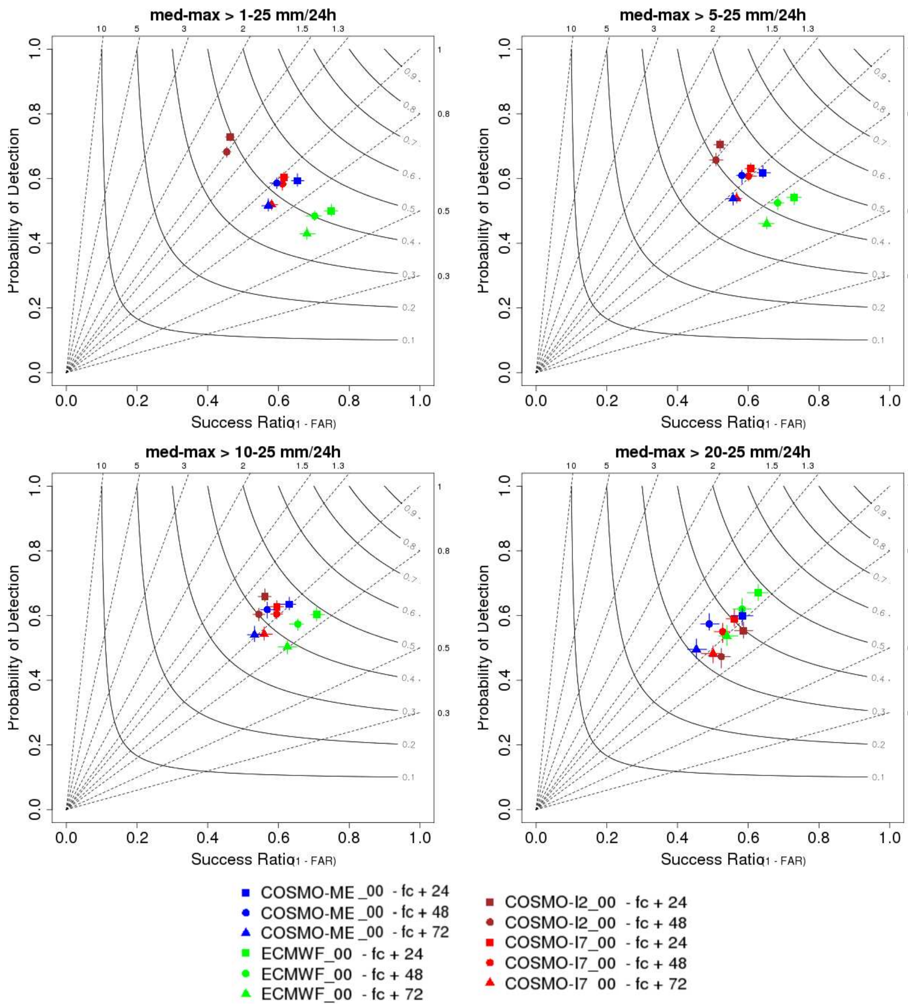



Figure 3: Performance diagrams as in figure 2, but for the event "maximum value greater than 25 mm/24h when the median exceed the reference threshold of 1,5,10 or 20 mm/24h".

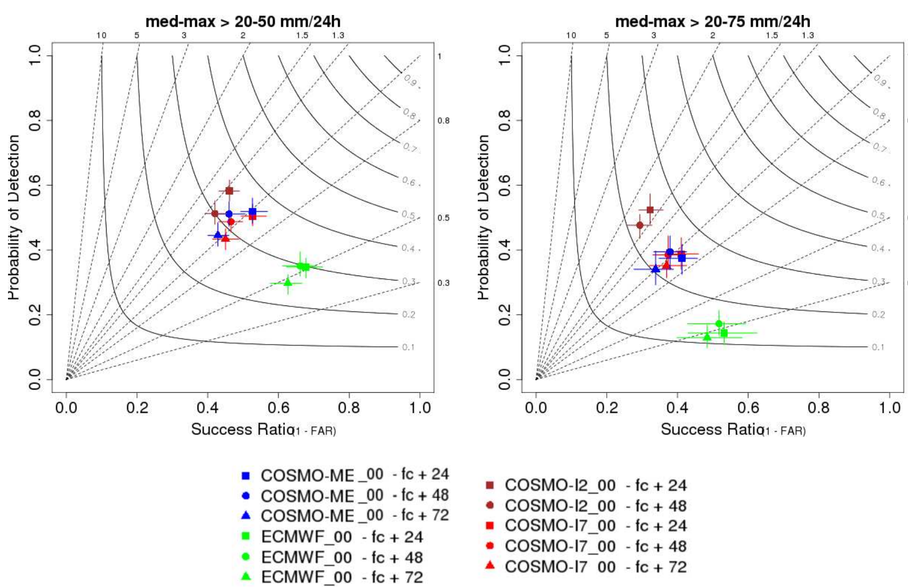



Figure 4: Performance diagrams as in figure 2 , but for the event "maximum value greater than 50 mm/24h and 75 mm/24h when the median exceed the reference threshold 20 mm/24h".

But if the condition on the median is kept fixed and the reference threshold for the condition on the maximum is increased, as shown in the graphs of figure 4, we note that the COSMO models present POD value better than IFS-ECMWF, even if the number of false alarm is large, in particular for COSMO-I2 which also overforecast the events. It should be stressed that the representativeness of the observational data set can be important in this case: the increase in the number of false alarms could depend on the fact that higher precipitation values may have not been actually measured, thereby supporting models that generally underestimates them. We can say that errors of underestimation are surely true while errors of overestimation should be false and need better investigations.

### 5 Conclusion

In this study we investigated whether the models are able to represent the distribution of precipitation within an area from a quantitative point of view . Some parameters of the forecast and observed precipitation distributions were then evaluated and compared using a categorical approach.

The QPF was first verified by imposing a single condition on the maximum of precipitation then adding a second condition (occurred simultaneously) on the median of the precipitation distributions, in order to select cases of a more widespread rain. In general, the Italian versions of the COSMO model seem to aptly describe the characteristics of the precipitation distribution that have been considered.

The 7 km models (COSMO-I7 and COSMO-ME) show a Bias Score very close to 1 under all the conditions, giving the impression to reproduce the distribution of precipitation in a fairly realistic way. The behavior of the two models is very similar, except the first 24h of forecast where the COSMO-ME is slightly better.

COSMO-I2 is able to capture values of precipitation punctually high, even if it has a large number of false alarms. The overestimations are reduced under the condition that gradually increasing precipitation occur in at least half of the area.

IFS-ECMWF behavior is very sensitive to the reference threshold, especially in detecting the maximum value of the precipitation distribution. When low thresholds are considered the number of events is overestimated and the POD and TS are very good, while for high value of precipitation a strong underestimation of the events and a general worsening in the scores are observed. The global model perform better than the COSMO models in case of widespread and uniform precipitation, but it seems to have some difficulty in the description of the precipitation distribution in case of high maxima (e.g. greater than 50 mm/24h).

It is important to remember that the results may be influenced by the spatial inhomogeneity of the observational datasets, especially in the assessment of the overestimation of the number of events and false alarms.

## References

[2] Marsigli,C., Montani,A., Paccangnella,T., 2008: A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorol. Appls.*, **15**, 125-143.

[2] Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601-608.

[3] Tesini,MS., Cacciamani,C., Paccagnella,T., 2010: Statistical properties and validation of Quantitative Precipitation Forecast. *COSMO Newsletter*, No. **10**, 45-54.