Verification and properties of COSMO-I7 QPF

in comparison with ECMWF model

mstesini@arpa.emr.it
ccacciamani@arpa.emr.it

Maria Stefania Tesini[1,2], Carlo Cacciamani[1]
[1] ARPA-SIM Regional Hydro-Meteorological Service of Emilia-Romagna, Bologna, Italy
[2] CIMA Research Foundation, International Center of Environmental Monitoring, Savona, Italy

INTRODUCTION

Limited Area Models are able to produce Quantitative Precipitation Forecast (QPF) with high spatial and temporal resolution, showing a large variability also in the amount of rain falling in a restricted area.
However frequent errors in time and space positioning make difficult a grid-point based use of models QPF.

A more correct use of QPF should be approached in a different way, in order to appreciate the properties and the additional information provided by LAMs respect to coarser resolution models.

As a starting point we devised a strategy based on the aggregation of observations and forecasts that fall into a predefined geographical area, several times wider than the model grid-box .
Here we present how COSMO-I7 (7 km horizontal resolution) works on the Italian territory in comparison with ECMWF model (50 km and 25 km horizontal resolution) focusing on relevant rain events useful for the Italian Civil Protection monitoring alert system

AIMS

1. Assessment of observed and forecast precipitation **climatology** over well-defined areas, study of its distribution function (*pdf*) and evaluation of summarizing quantities such as areal mean, maximum values and quantiles, in order to analyse the capability of the models in reproducing precipitation statistical properties

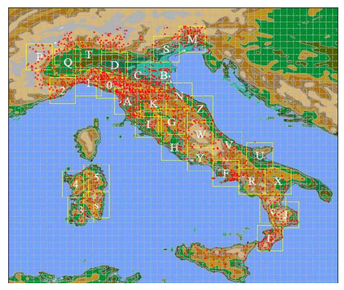
2. Verification of the **"day-by-day"** quality of the QPFs, making use both of descriptive methods and of quantitative measures (such as POD,FAR,TS and BIAS) in each defined area, throughout the selected period and thresholds deduced by the climatological results obtained in the previous point

METHODOLOGY

We considered **24 hours cumulated precipitation (00-24 UTC)** throughout seasons of several years (Summer 2005,2006,2007; Spring 2005,2006,2007; Autumn 2005,2007) even if we present here only the results concerning **Autumn 2005 and 2007**

QPF of the following modelling chains have been taken into account:

- **RUN 00 UTC COSMO-I7(+24h)**
  - Italian version of Lokal Model (LM)
  - ~7 km horizontal resoltion
  - *for detail concerning COSMO Consortium and technical characteristics of LM and COSMO-I7 implementation see the COSMO web site: http://www.cosmo-model.org/*
- **RUN 12 UTC ECMWF-IFS (+12/+36)**
  - ~50 km horizontal resoltion (SON 2005)
- **RUN 00 UTC ECMWF-IFS (+24)**
  - ~25 km horizontal resoltion (SON 2007)

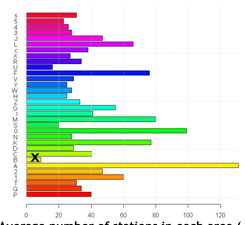


The observation dataset was composed by about 1000 rain gauges of the National Department of Civil Protection (DPCN) network (red points in the map)

We devised a verification methodology by aggregating observed and predicted precipitation in boxes of 1°x 1° (labelled boxes in the map)

The choice of the size and position of the areas has been performed according to different rules:

- the boxes have to be enough large in order to contain a high number of observation points (ranging from 20 to over 100, depending on location and period of time considered)
- the boxes have to be homogeneous as much as possible in terms of geographic-territorial characteristics



Average number of stations in each area (SON 2007)

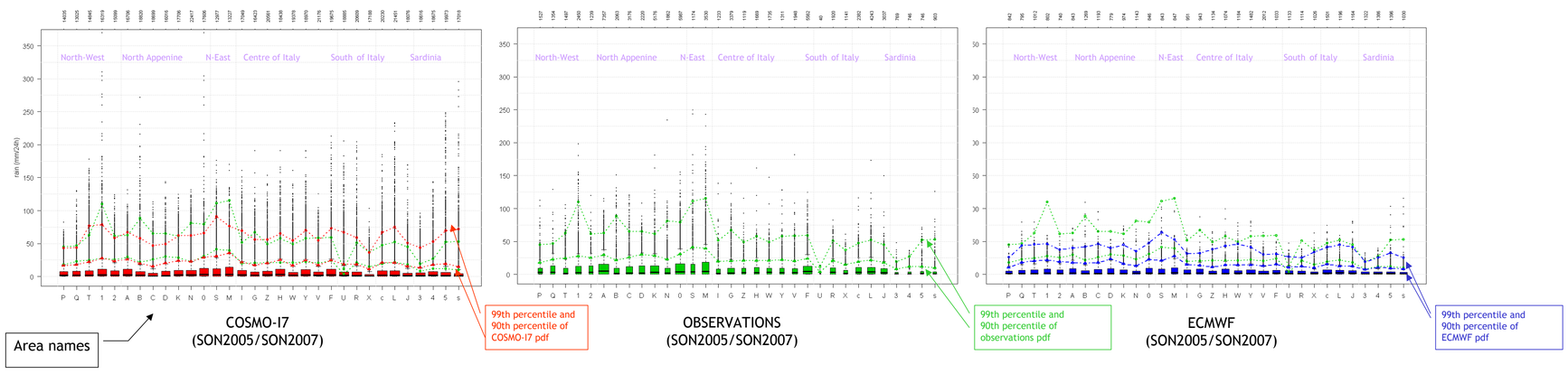
AIM 1

"Climatological" behaviour

We estimated the distribution function of the observed/predicted precipitation field taking in account all the points that fall in each area throughout the examined periods (SON 2005 and SON 2007 in this case).

Each area contains a lot of observation points as well as model grid-points: we assume that the occurrence of precipitation in each of these points is equiprobable.

Open problem: which should be the dimension of the area in order to consider well-founded this assumption?
We are working on this topics, but for this work we followed a pragmatic approach for the choice of the dimension on the box: we should have enough forecast and observation points to perform significative statistics in each area



In the above boxplots the distribution of COSMO-I7 QPF (red), observations (green) and ECMWF QPF (blue) for each area of the study are displayed.
Green dotted lines in each plot represent the 99th and 90th percentile of the observed rain distribution, while the red and blue lines refer respectively to the model QPF distributions

The distribution of observed rain shows that the 90% of rain events throughout SON2005 and SON2007 is generally lower than about 20 mm/24hours, with a large variability due to geographical position of the areas.
Events greater than 50 mm/24hours represent only 1% of the total, but it is noteworthy that these are just the events in which the Civil Protection is interested in.

COSMO-I7 distribution seems to be fairly realistic, at least since the 90th percentile, even if in many areas also the 99th percentile is well described. Greater differences are due to outliers, with large overestimation of the maxima in many areas. However, it should be noted that it could often depend on a problem of spatial representation, such as in costal areas that present sea points not covered by observation.
The features of the observed pdf , and in particular the spread of the tail of the distribution, seem to be reasonably well reproduced.

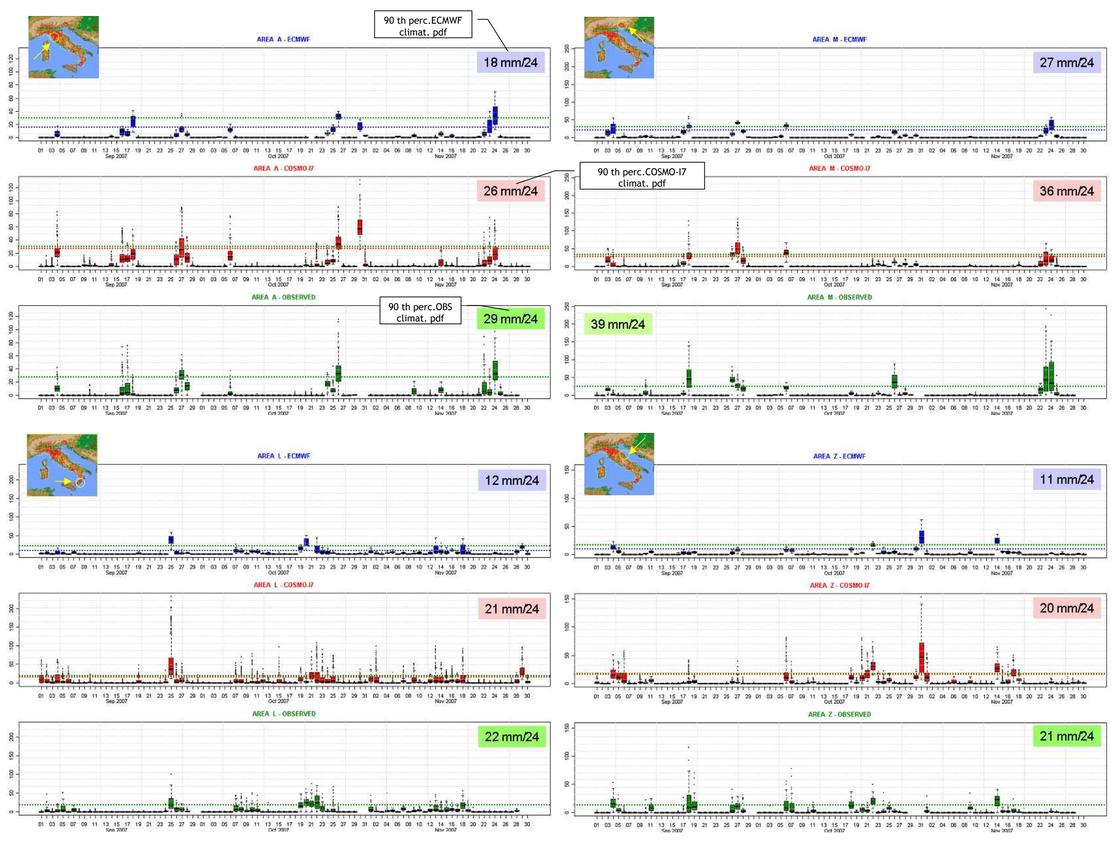
On the other hand, the spread of ECMWF *pdf* do not cover all the range of the observed values, with large underestimation both of 90th and 99th quantiles.

AIM 2

"Day-by-day" behaviour: qualitative description

We assessed the quality of the "day-by-day" QPF representing the daily distribution of the observed rain and models QPF in a single area. We considered all the model grid-points and all the observed rains that fall daily (00-24UTC) within each box.
Using boxplot we can appreciate the capability of the model to reproduce the variability of the rain field over the selected area: it is possible to obtain information about the spread of the distribution and a roughly description of key measures that define it, such as the median, the quartiles, maximum and minimum values, and a range of values identified as "outliers".

In the graphs below are shown the time-series of the daily forecasted/observed rain distribution in four different areas for ECMWF (blue), COSMO-I7 (red) and observations (green). In order to define a reference threshold we added in the graph some dotted coloured lines (colours as before), indicating the 90th percentile of the climatological distribution estimated for each dataset , for the selected area, in the previous point.



COSMO-I7 is more realistic than ECMWF in reproducing the intra-box variability. However, COSMO-I7 presents both a large number of false alarms and high "spikes". On the other hand, ECMWF presents a greater number of missed alarms, especially if we choose high reference thresholds.
According to most standard verification measures, COSMO-I7 forecast would have poor quality, but it might be very valuable to the forecaster since it provides information on the distribution and variability of the rain field over the considered region.

"Day-by-day" behaviour: quantitative description

In order to quantify the quality of the models QPF we use dichotomous statistics (i.e. yes-or-no forecasts). We chose different criteria to define the event, according to the properties that we want to investigate, than we generate a contingency table that shows the frequency of "yes" and "no" forecasts and occurrences.

For example we defined the event in 2 different way:

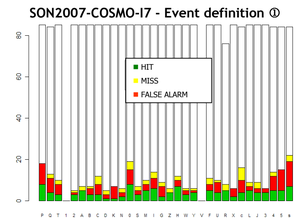
① at least 10 percent of the points (**observed/forecasted**) belonging to the area of interest should present rain values greater the **90th percentile of the climatological pdf of the observations**. The threshold depends on considered region (see green dotted lines in the daily boxplot)

② at least 10 percent of the points (**observed/forecasted**) belonging to the area of interest should present rain values greater then **a relative threshold: 90th percentile of climatological pdf of observation for the observations ,** and **90th percentile of the climatological pdf of model QPF for model.** The threshold depends on the selected area. (see red (COSMO-I7) and blue (ECMWF) dotted lines in the daily boxplot)
Note: *This is a "renormalisation" procedure, to take into account the different climatology of observed and forecasted precipitation.*

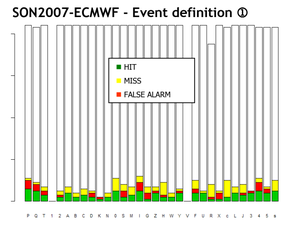
FCST \ OBS	YES	NO
YES	hits	false alarm
NO	misses	correct rejection

- *hit* - event forecast to occur, and did occur
- *miss* - event forecast not to occur, but did occur
- *false alarm* - event forecast to occur, but did not occur
- ☐ *correct rejection* - event forecast not to occur, and did not occur

SON2007-COSMO-I7 - Event definition ①

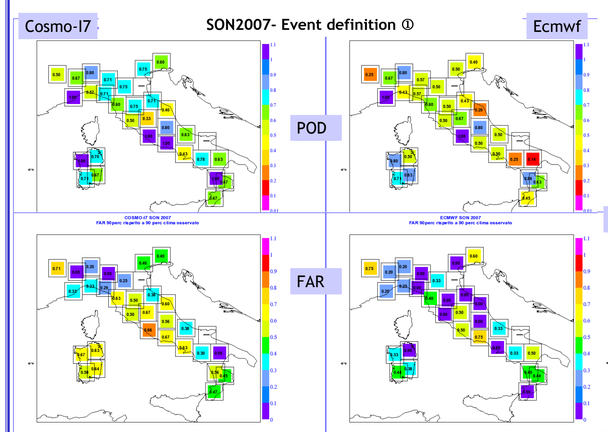


SON2007-ECMWF - Event definition ①



The contingency table is a useful way to see what types of errors are being made

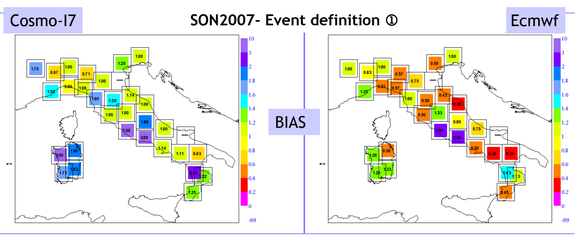
SON2007- Event definition ①



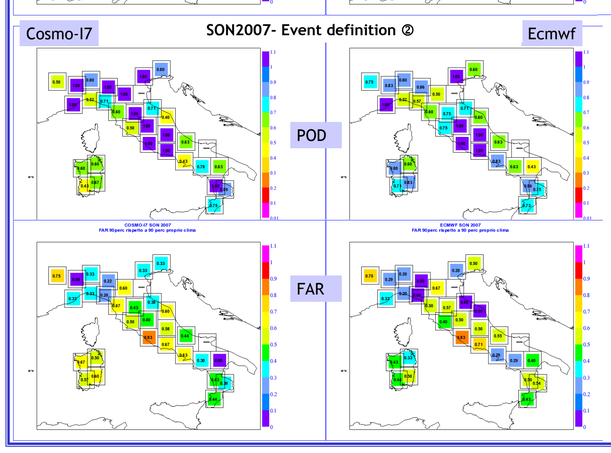
From the elements of the contingency table we calculate some categorical statistics, such as POD,FAR and BIAS as to describe particular aspects of forecast performance.

POD=hits/(hits+misses)
FAR=false alarms/(hits+false alarms)
BIAS=(hits+false alarms)/(hits+misses)

SON2007- Event definition ①



SON2007- Event definition ②



It's interesting to note the strong dependence of the models performance on the geographical position of the boxes.

We observe different results according to the period of the year and in connection with particular weather type occurrence.
In some areas models exhibit various behaviour depending on season. Troughout the same season they perform differently according to different weather regimes. This aspect is even more evident for COSMO-I7 because of the stronger interaction of the synoptic flow with orography

POD improves when we considered separately the observations threshold and the model threshold. This kind of recalibration process affects mainly ECMWF, because the difference between model and observed pdf is more significant, especially in the tail distribution. However the number of false alarm increase, in particular for ECMWF.

4th International Verification Methods Workshop
June 4 – 10, 2009
Finnish Meteorological Institute, Helsinki, Finland
Tutorial Session: June 4-6 Scientific Workshop: June 8-10