EVALUATION OF YEAR 2007 OPERATIONAL WRF-NMM

PAPER P3B.6

Gozzini B., Bartolini G., Grifoni D., Messeri G., Pasi F., Piani F., Rossi M., Tei C.

CNR – IBIMET, Via Caproni, 8 50145 Firenze, Italy
LaMMa Consortium, Via Madonna del Piano, 10 50019 Sesto Fiorentino (FI), Italy

INTRODUCTION

Verification is a crucial point of meteorological research and operational forecasting activities. If the methodology is properly designed, verification results can effectively meet the needs of several working groups, including modellers, forecasters and users of forecast information (Casati et al., 2008).
The objectives of verification are: (i) to quantify the improvement of the forecast skill over the years; (ii) to compare the performance of different versions of a forecasting system in order to decide which is the best for operations; (iii) to understand where the problems are and what aspects of the system need refinements; (iv) to compare the relative value of two different systems for a specific category of users (Bougeault P., 2003).
A large number of regional weather services uses mesoscale limited-area models in their forecasting process. The quality of the forecasts is also related to the model capability in providing reliable precipitation, temperature, and wind forecasts at resolutions that tend to meet the needs of specific applications (e.g., agriculture, navigation, hydrology, flood forecasting) (Lagouvardos et al., 2003).
The main objective of this work is to analyse the skill of meteorological model chain currently running operationally at LaMMA (Laboratory for Meteorology and Environmental Modelling) Consortium, the regional weather service of Tuscany, Italy. A verification procedure has been developed in order to assess the ability of the model to provide accurate forecasts throughout the comparison against gridded analysis and available ground observations.

MATERIALS AND METHODS

LaMMa OPERATIONAL CHAIN AND SETTINGS

The WRF-NMM model is running operationally at the LaMMA Consortium for the regional weather forecasting service at a resolution of 0.07 deg (about 7,5 Km) over a domain covering central Europe (128x232 points), as shown in Figure 1. Initial and boundary conditions are given by the ECMWF model (T799) at 0.25 degree of resolution with data of 00UTC for 72-hours runs. Boundary conditions are updated with ECMWF forecasts every six hours.

Model configuration consists of 35 vertical levels unequally spaced from ground to 100 hPa (420 hPa is the limit of sigma to pressure) and with the first 10 levels being concentrated in the boundary layer (around 1.0 km above ground level).
Model is running with a 18 seconds time-step and other relevant physics options are: Ferrier microphysics, Geophysical Fluid Dynamics Laboratory (GFDL) Long and Shortwave radiation, NMM LSM Land surface, Janjic Similarity Surface layer, Mellor-Yamada-Janjic TKE Boundary-layer, Kain-Fritsch cumulus scheme. Land-use and soil category comes from the standard USGS categories (24 for land use and 16 for soil). Topography is derived from the global 30-second USGS topography data with a 4-point average.

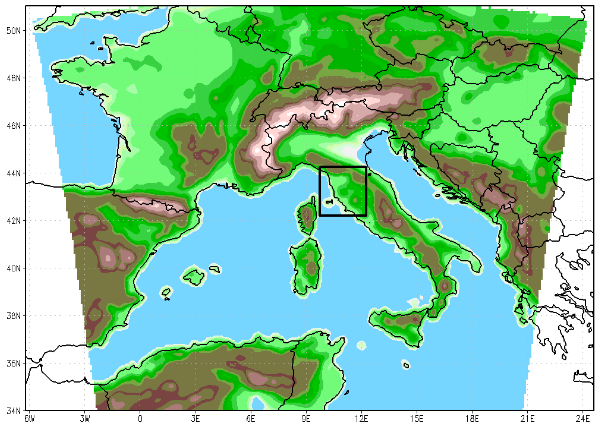


Fig. 1. Map of the operational WRF-NMM spatial domain compared to the study area

To evaluate the model performance the year 2007 was re-run with the operational configuration and 3.0 model version. The dedicated software MET (Model Evaluation Tools) version 2.0 by NCAR was used. The model was verified against upper-air gridded analysis fields provided by ECMWF (0.25 deg. of resolution) for what concerns the geopotential height, air temperature, relative air humidity and wind speed at 850 and 500 hPa levels. As for surface, fields model precipitation was verified against the regional network (Tuscany) of rain gauges, that counts more than 200 stations.

STATISTICAL VERIFICATION OF UPPER AIR FIELDS

Root-Mean-Square Error (RMSE) and Mean Error (BIAS) were daily, monthly and seasonally calculated over the whole domain. Since the considered domain includes Alps and other mountain chains exceeding or intersecting the 850 hPa level (height of about 1500m), model grid points above 1000 m were masked out.

STATISTICAL VERIFICATION OF PRECIPITATION

Recently in literature, there is a growing interest in accurate precipitation forecasts because of the large impact of rain on agriculture, outdoor activities, traffic, hydroelectric power generation, or flooding preparedness. Reliable precipitation forecasts are particularly important in mountainous regions because of the increased likelihood of heavy precipitation.
So, special attention is paid to the verification of quantity precipitation forecasts (QPF). The precipitation forecast is considered to be one of the most important model output fields provided by numerical weather prediction models because precipitation has direct (and often disastrous) impacts on human activities.
A large variety of verification scores are used operationally to verify QPFs (Wilson, 2001; Nurmi, 2004; Jolliffe and Stephenson, 2003) and for this study we choose to compute the highly recommended different skill scores in literature. The basis for the calculation of these scores is a two-way contingency table (Yes/No). A contingency table usually shows frequencies for particular combinations of values of two discrete random variables X and Y (tab. 1).

Table 1 – Contingency table

	Observed YES	Observed NO
Forecasted YES	**a**	**b**
Forecasted No	**c**	**d**

where:
a = number of stations where observed and forecasted precipitation are above a threshold (hits)
b = number of stations where forecasted precipitations are above a threshold (hits) while the observed are below (false alarm)
c= number of stations where observed precipitations are above a threshold (hits) while the forecasted are below (misses)
d= number of stations where forecasted and observed precipitations are below threshold (correct rejections)

we calculated:

<u>AREAL BIAS</u>
This is the ratio between the total number of stations in which the model forecasted an event (precipitation above a certain threshold) and the total number of stations where that event was observed. Areal Bias is equal to 1 for a perfect forecast, above or below 1 for model overestimation and underestimation, respectively.

$$\frac{(a+b)}{(a+c)}$$

<u>CRITICAL SUCCESS INDEX (THREAT SCORE)</u>
It is the ratio between the number of stations where the event was correctly forecasted (hits) and the number of those where the event occurred or was forecasted. Its value is 1 for a correct forecast.

$$\frac{a}{(a+b+c)}$$

<u>HIT RATE (PROPORTION CORRECT)</u>
It is the fraction of correct forecast (including correct rejection) on the total number of stations. Its value is 1 for a correct forecast.

$$\frac{a+d}{(a+b+c+d)}$$

<u>POD</u>
It is the fraction of events that were correctly forecasted to occur. Its value is 1 for a correct forecast.

$$\frac{a}{(a+c)}$$

<u>FAR</u>
It is the fraction of forecasted events for which the event did not occur. Its value is 0 for a correct forecast

$$\frac{b}{(a+b)}$$

<u>QUANTITY BIAS</u>

$$\frac{1}{N}\sum_{i=1}^{n}(f_i - o_i)$$

Where f_i is a single precipitation forecast and o_i is the corresponding precipitation observation.

<u>MAE</u>

$$\frac{\sum|f_i - o_i|}{n}$$

Where f_i is a single precipitation forecast, o_i is the corresponding precipitation observation and n is the number of observing stations.
The first five skill indices were calculated for five distinct thresholds values of precipitation (0.1, 0.2, 5, 10 and 20 mm).
The last two indices were calculated for five ranges of the observed precipitation amounts (0.1-2; 2-5; 5-10; 10-20; >20 mm).
All the statistical verification were performed using MET 2.0 point-stat module with distance-weighted mean. The forecast value at *P (station)* is a weighted sum of the values in the W x W square (where W is assumed 2, that is the 4 closest model grid points). The weight given to each forecast point is the reciprocal of the square of the distance (in grid coordinates) from *P*. The weighted sum of forecast values was normalized by dividing by the sum of the weights.

MET point-stat module was applied to 19 case studies with relevant amounts of 24-h accumulated precipitation over Tuscany Region, occurred during 2007 (table 2). The verification of accumulated precipitation has been performed considering different model forecasts (day0 = 0h+24h forecast, day1 = 24h+48h forecast, day2 = 48h+72h forecast) against the available rain gauge data of the network of Hydrological Service of Tuscany Region (Centro Funzionale) (Fig. 2).

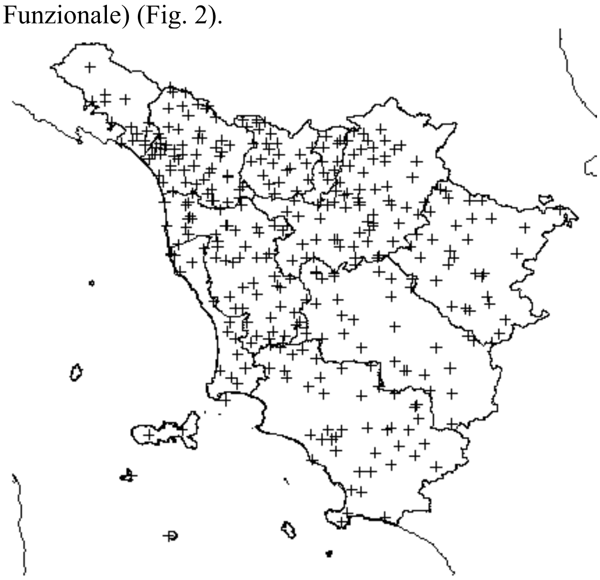


Figure 2 – Distribution in Tuscany of hydrological service weather station network

Table 2. List of the 19 case studies

Case	24h precipitation exceding
8 Jan	60/80 mm over North Eastern part of Tuscany
23 Jan	80 mm at 6 stations
7 Feb	60 mm at 10 stations
12 Feb	80 mm at 6 stations
25 Feb	80 mm over the Northern part of Tuscany
19 Mar	80 mm over North Eastern part of Tuscany
4 May	100 mm at 14 stations
5 May	100 mm at 8 stations
28 May	60 mm over many areas in Tuscany
8 Aug	100 mm over the Northern area of Tuscany
20 Aug	50 mm at 7 stations
23 Aug	40 mm over many areas of Tuscany
4 Sep	50 mm over the Central and North area of Tuscany
17 Sep	80 mm over the North part of Tuscany
27 Sep	80 mm over the North area of Tuscany
26 Oct	100 mm at 7 stations
30 Oct	80 mm at 6 stations
24 Nov	100 mm over the North Eastern part of Tuscany
8 Dec	60 mm over Florence in Tuscany

RESULTS

UPPER AIR FIELDS

Figure 3 shows the RMSE and BIAS of air temperature at 850 hPa levels for t+12, t+24, t+36, t+48 and t+60 forecast hours as a monthly average values for the period January-December 2007 over the entire available domain.

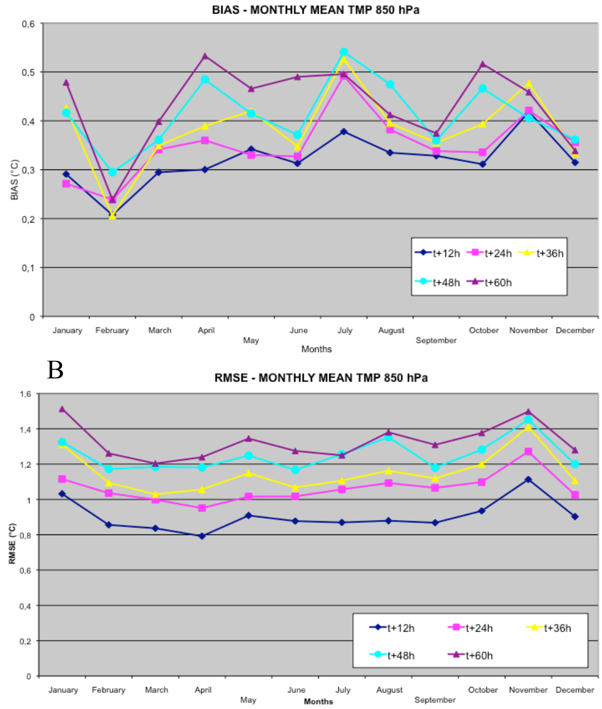


Figure 3 – Monthly average of BIAS (A) and RMSE (B) of air temperature at 850 hPa levels for +12, +24, +36, +48 and +60 forecast hours for January-December 2007 over the entire available domain.

RMSE, which represents the magnitude of the mean error, shows that t+12h forecast has the best results with a mean value of about 0.9 °C over the entire year, except for November and January with values of more than one (still acceptable). As expected the accuracy of the model decreases with forecast time, so t+24h and t+36h are about 1.1 degrees and t+48h and t+60h around 1.3 degrees.

Also BIAS, which represents the fluctuation of the errors, shows the best performance for the t+12h forecasts while the t+60h forecast seems to be, in general, the worst prediction with the exception of February, July and November. Differences between t+24h, t+36h and t+48h forecast are quite small. The positive values of BIAS highlight a general overestimation of the model.

To understand the distribution of the two indices over the whole domain, average seasonal BIAS and RMSE values for t+12h forecast hours were visualised as maps obtained using GrADS software (Fig. 4).

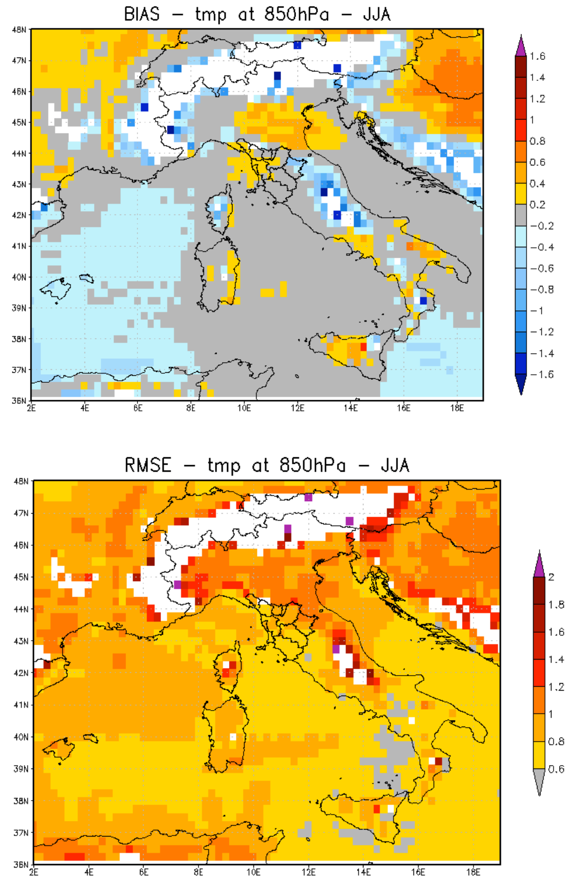
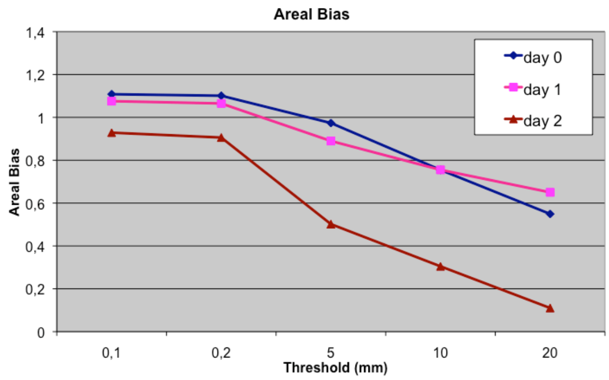


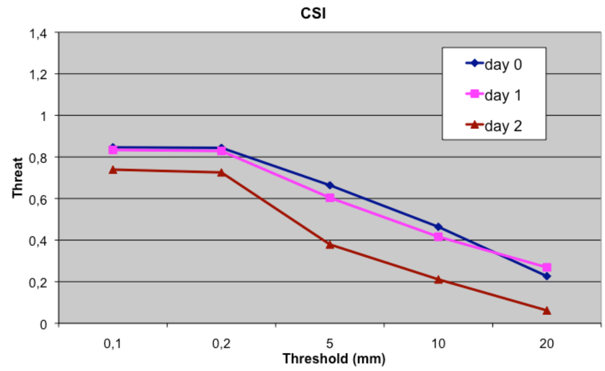
Figure 4 – Summer average values of BIAS and RMSE of temperature at 850 hPa levels for +12 forecast hours for January-December 2007.

RMSE is always lower than 2 degrees and the highest errors occurs close to the areas with complex orography. Over Tuscany, RMSE is lower than 1 degrees over its plains and hills, while exceed 1 degree (but lower than 1.5 degrees) near Apennines.
BIAS shows the same behaviour: maximum values are close to the Alps (northern border of the domain) and Apennines (in particular over Gran Sasso, around 2500 meters, in Central Italy) and close to the Massif Central (at the western boundary of model domain). Where the topography is far from reliefs, model reflects observation fields very well. This fact is particularly evident in the open sea.
The same statistics for the two pressure levels (500 and 850 hPa) has been computed for the other parameters (geopotential height, relative air humidity and wind speed).

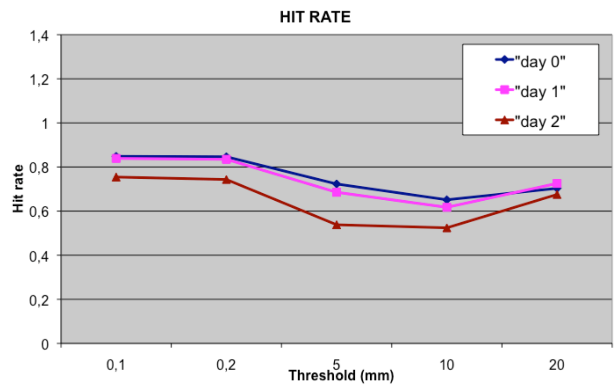
PRECIPITATION



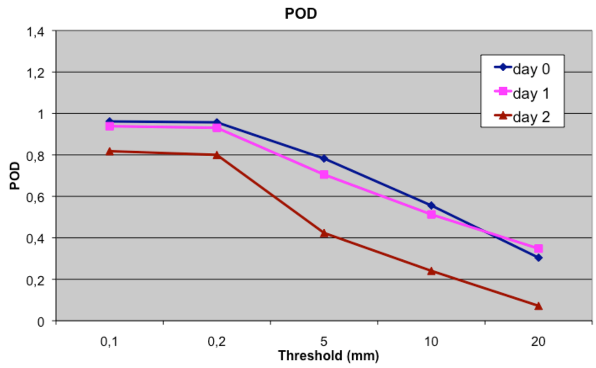
AREAL BIAS is very close to 1 in the threshold from 0.1 to 5 mm. For higher amounts of rain the areal bias decrease indicating an under prediction of the areal extent of precipitation by the model. The areal bias values confirms the reduction of performance of the model forecasted precipitation at day2 while the skills at day0 and day1 are very similar.



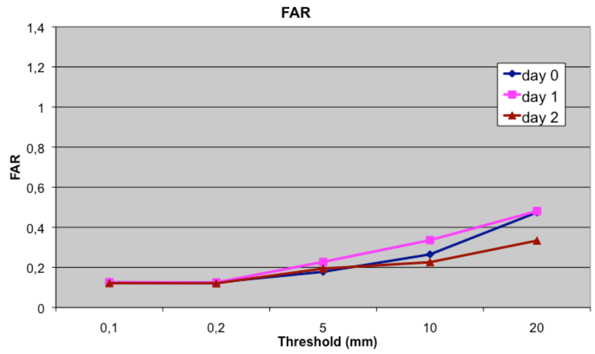
CRITICAL SUCCESS INDEX shows a good forecast skill up to the 5mm threshold. Even for large precipitation amounts the model has skill, with threat scores ~ 0.3. Also in this case the threat score values confirm the reduction of performance of the model forecasted precipitation at day2 while the skills at day0 and day1 are very similar.



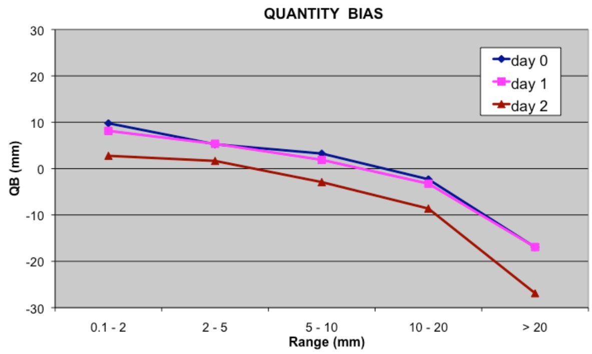
HIT RATE shows a good forecast skill. Obviously, for large precipitation amounts the HIT RATE values increase considering as correct forecast also the correct rejection.

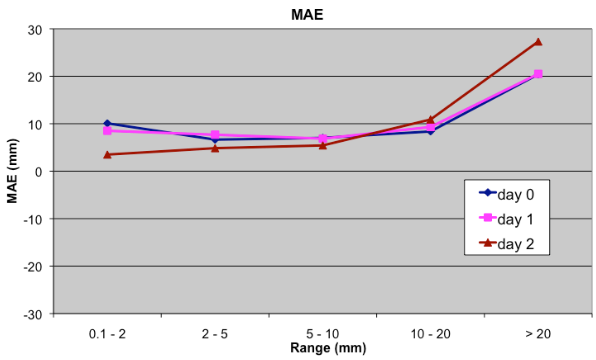


POD shows the best performance for the thresholds up to 5 mm, while decreases for the greater threshold of rain events. Also in this case the POD score confirms the reduction of performance of the model forecasted precipitation at day2 while the skills at day0 and day1 are very similar.



FAR is very close to 0 (perfect score) in the thresholds from 0.1 to 10 mm while for higher thresholds it, obviously increases.





The calculation of the QUANTITY BIAS together with MAE reveals that the forecasts overpredict the rain amounts for low and medium precipitation ranges. For these ranges this overprediction is less important for 24-h accumulated forecasted precipitation at day2. At the highest ranges the model underpredicts the rain amounts.

CONCLUSIONS

The aim of this study was the verification of WRF-NMM running operationally at LaMMa Consortium (Tuscany/Italy weather service) for year 2007. Several both upper air and surface fields forecast by WRF-NMM has been compared against ECMWF analysis and ground station observations by using the most common skill scores (BIAS, RMSE, POD, FAR, ..).
Upper air fields verification shows satisfactory results for a regional forecast (e.g. RMSE lower than 2° C for temperature at 850 hPa).
QPF verification, made on 19 case studies with cumulative precipitation above 60 mm in 24h, shows similar results to those found in other evidences.

REFERENCES

Lagouvardos K., Kotroni V., Koussis A:, Feidas H., Buzzi A. and Malguzzi P., 2003. *The Meteorological Model BOLAM at the National Observatory of Athens: Assessment of two-year Operational Use.* Journal of Applied Meteorology, vol. 42, pagg. 1667-1677.

Casati B., Wilson L.J., Stephenson D.B., Nurmi P., Ghelli A., Pocernich M., Damrath U., Ebert E.E., Brown B.G. and Mason S., 2008. *Review Forecast Verification: current status and future directions.* Meteorological Applications, 15, pagg. 3-18

Bougeault P., 2003. *The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events.*. Summary on the state of the art.
http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/Bougeault/Bougeault_Verification-methods.htm

Jolliffe, I.T., and Stephenson, D.B., 2003. *Forecast Verification. A Practitioner's Guide in Atmospheric Sciences*. Wiley and Sonss Ltd, 240 pp.

Nurmi, P., 2003. *Recommendations on the verification of local weather forecasts.* ECMWF Tech Memo. 430, 18pp.
http//:www.ecmwf.int/publications/library/ecpublications/:pdf/tm430.pdf

Wilson, C., 2001. *Review of current methods and tools for verification of numerical forecasts of precipitation.* COST717 Working Group Report on Approaches to verification.
http://pub.smhi.se/cost717/doc/WDF/_02_200109_1.pdf