Three models intercomparison for Quantitative Precipitation Forecast over Calabria(*)

S. Federico[1][2], E. Avolio[1][3], C. Bellecci[1][3], M. Colacino[2]
A. Lavagnini[2], C. Accadia[4], S. Mariani[4] and M. Casaioli[4]

[1] CRATI s.c.r.l. c/o Università della Calabria - 87036 Rende (CS), Italy
[2] CNR-ISAC - Via del Fosso del Cavaliere 100, 00133 Rome, Italy
[3] Università di Roma "Tor Vergata" Dept. STFE - Via del Politecnico 1, 00133 Rome, Italy
[4] Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici - Rome, Italy

Summary. — In the framework of the National Project "Sviluppo di distretti industriali per le Osservazioni della Terra" (Development of Industrial Districts for Earth Observations) funded by MIUR (Ministero dell'Università e della Ricerca Scientifica —Italian Ministry of the University and Scientific Research) two operational mesoscale models were set-up for Calabria, the southernmost tip of the Italian peninsula. Models are RAMS (Regional Atmospheric Modeling System) and MM5 (Mesoscale Modeling 5) that are run every day at Crati scrl to produce weather forecast over Calabria (`http://www.crati.it`). This paper reports model intercomparison for Quantitative Precipitation Forecast evaluated for a 20 month period from 1th October 2000 to 31th May 2002. In addition to RAMS and MM5 outputs, QBOLAM rainfall fields are available for the period selected and included in the comparison. This model runs operationally at "Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici". Forecasts are verified comparing models outputs with raingauge data recorded by the regional meteorological network, which has 75 raingauges. Large-scale forcing is the same for all models considered and differences are due to physical/numerical parameterizations and horizontal resolutions. QPFs show differences between models. Largest differences are for BIA compared to the other considered scores. Performances decrease with increasing forecast time for RAMS and MM5, whilst QBOLAM scores better for second day forecast.

PACS `92.60.Jq` – Water in the atmosphere (humidity, clouds, evaporation, precipitation).
PACS `93.00` – Geophysical observations, instrumentation, and techniques.
PACS `92.60.Wc` – Weather analysis and prediction.

(*) The authors of this paper have agreed to not receive the proofs for correction.
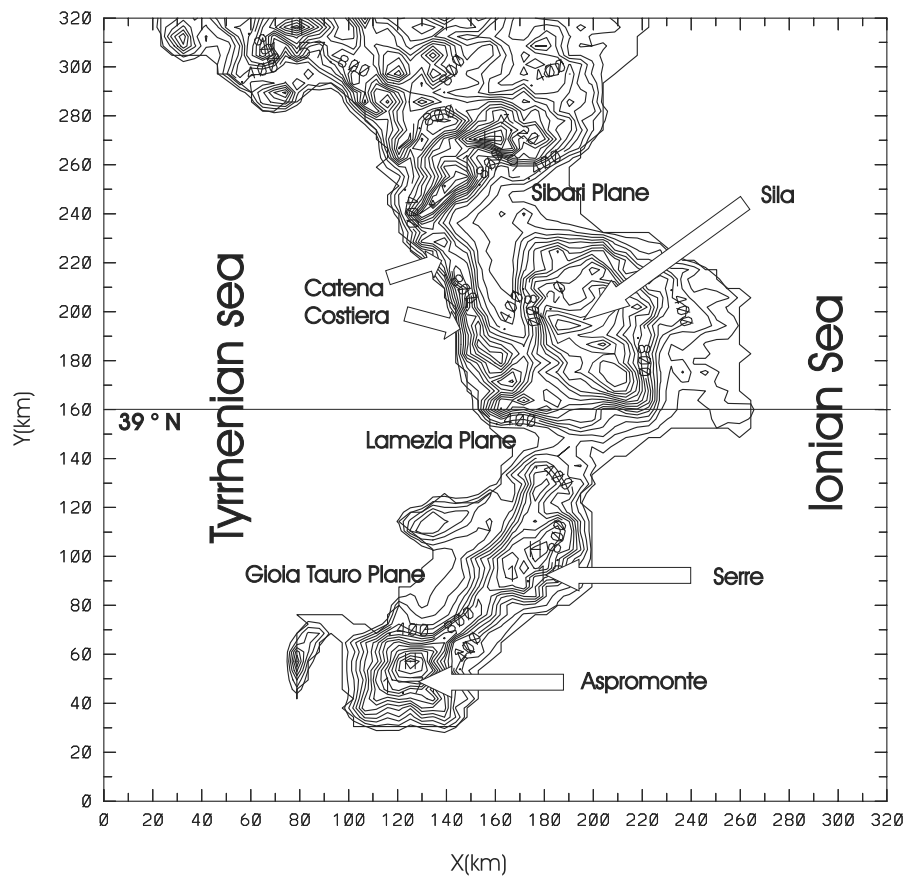
Fig. 1. – Topography of Calabria averaged over 10 km². Main features are also reported.

1. – Introduction

Improving Quantitative Precipitation Forecast (QPF) and flood forecasting is one of the most desired aspects in numerical weather prediction. Precipitation is a difficult field to predict quantitatively for several reasons. First of all the process depends on several factors like temperature, humidity, winds in highly nonlinear ways. Another complication is introduced by variable physiography, *i.e.* surface features such as topography, land water boundaries, vegetation and soil parameters. Numerical meteorological limited area models (LAMs) are operating in several forecasting centers over the world to give, among others, QPFs. At CRATI Scrl, two models are used, daily, to produce weather forecast over Calabria, the southernmost tip of the Italian peninsula: RAMS and MM5.

QBOLAM QPF fields are also available for comparison. The QBOLAM model runs operationally at Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT-Agency for Environmental Protection and Technical Service) in Rome, as a part of POSEIDON sea wave and tidal forecasting system [1], which includes also a wave model (WAM), an oceanographic model (POM) nested on a finite element model of the Venice Lagoon (VL-FEM) in cascade with QBOLAM output.

Calabria, fig. 1, ranges between $38°12'$ and $40°$ latitude North and between $15°30'$

and 17°15′ longitude East. The west coast of the region is bounded by the Tyrrhenian Sea and the east and south coasts by the Ionian sea. Apennines run North to South along the peninsula and are characterized by five main topographical features reaching 1.5–2.0 km elevations: Pollino, Catena Costiera, Sila, Serre, Aspromonte. The average width of the region is about 50 km in the West-East direction and 300 km in North-South direction. Three main valleys locate by the sea and most of agricultural and industrial sites are in those valleys.

In a previous paper [2], hereafter referred as FABC, the effects of enhancing horizontal grid resolution on QPF over Calabria were studied. Main conclusion was that physiographic features are important in determining rainfall patterns and amounts at all thresholds considered. The presence of mountains acts on rainfall by several mechanisms including low-level convergence associated with flow deflection around the topography [3-5] and terrain-triggered convection [6].

Another important aspect of QPF is that performances decrease while increasing forecast lead time. As pointed out by Lorenz [7,8] the atmosphere is a chaotic system and even an infinitesimally small perturbation introduced into the atmosphere at a given time will result in an increasingly large change in the evolution of the system so that small initial perturbations evolve in large differences with time. To cope with this problem, different techniques, like the ensemble forecasting [9,10], have been developed. However, the main point is that even for a perfect model, due to analysis errors, the forecast skill would degrade to zero with time. Analysis errors can be due to lack of complete coverage, measurement errors, errors in first guess and to the approximation of analysis technique. In addition, model errors, both physical and numerical, degrade performances.

The key issue of this paper, which represents an extension of FABC, is the verification of precipitation forecast. First of all models performances are assessed by non-parametric objective scores, then comparison between models is made and performances degradation is studied for each model as a function of integration time.

To cope with these tasks, daily forecasts for twenty months (from 1 October 2000 to 31 May 2002) are analyzed. For each model, starting from 1200 UTC ECMWF analysis/forecast cycle as initial and dynamic boundary conditions, a 60 h forecast is performed. The first 12 h are spin-up time whereas the remaining 48 h are the actual forecast. This time period is divided in two 24 h frames, 0–24 h and 24–48 h. These two frames are also referred to as 1D forecast and 2D forecast, respectively.

Performances are compared by scores computed from models outputs and raingauge measurements available from the ARPACAL (Calabria regional agency for environmental protection) network. Measurements are available at 75 locations and are shown in fig. 2. Data refer to daily cumulated rainfall. Statistical significance of differences between models' scores is assessed by the application of the statistical test proposed by Hamill [11, 12].

2. – Models set-up

2˙1. *RAMS model*. – For details on RAMS model the reader should refer to Pielke *et al.* and Cotton *et al.* [13,14]. Grids configuration is shown in fig. 3. Horizontal grids spacings are 30 km and 6 km for the first and second grid, respectively. Grid nesting uses a two-way interactive procedure [15]. Communication from the coarse to the nested grid is accomplished immediately following a timestep on the coarse grid. Twenty-five vertical levels, up to 12500 m in the terrain following coordinate system, are used in simulations. Levels are not equally spaced: within the PBL (Planetary Boundary Layer)
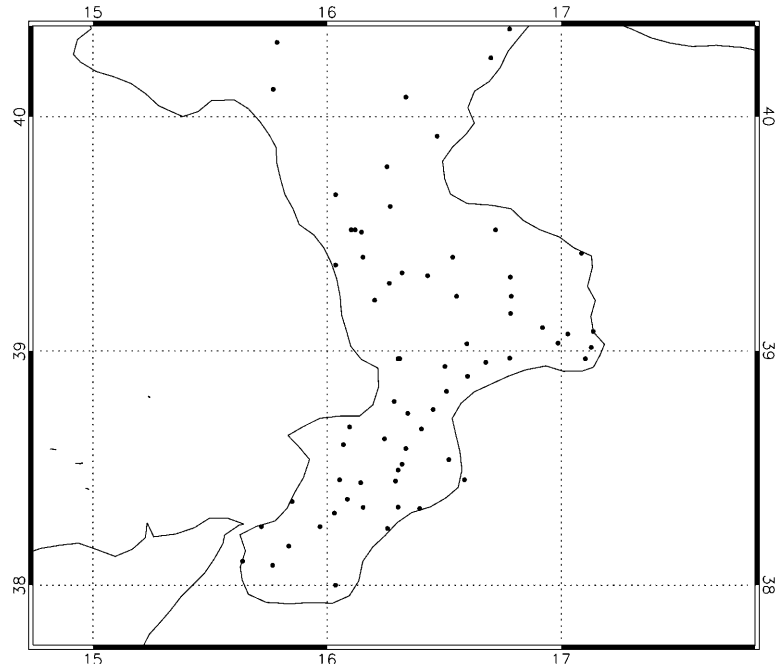
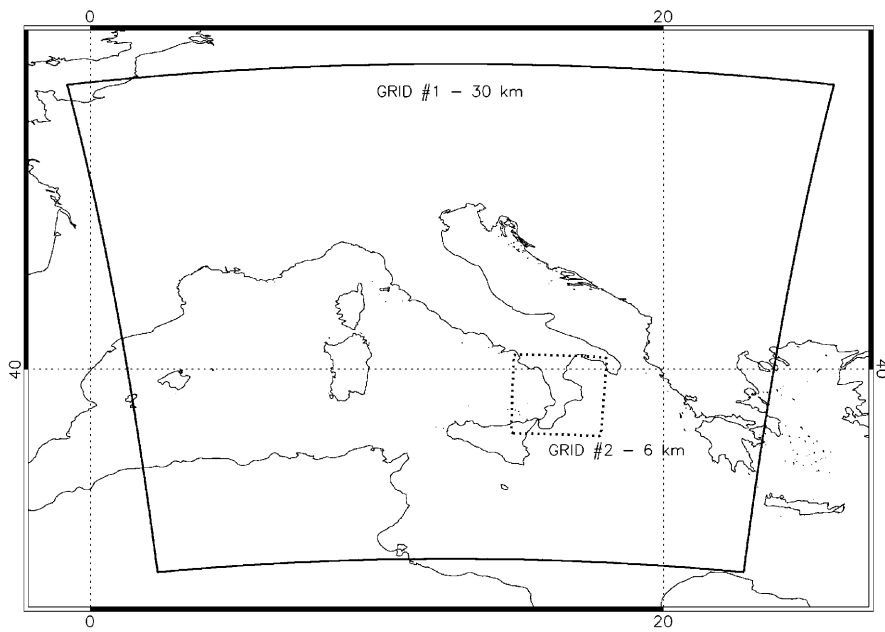Fig. 2. – ARPACAL network raingauges used to compute scores.

Fig. 3. – RAMS grids. Horizontal resolutions are 30 km and 6 km for first and second grid, respectively. Two-way interactive procedure is used for grids communication.
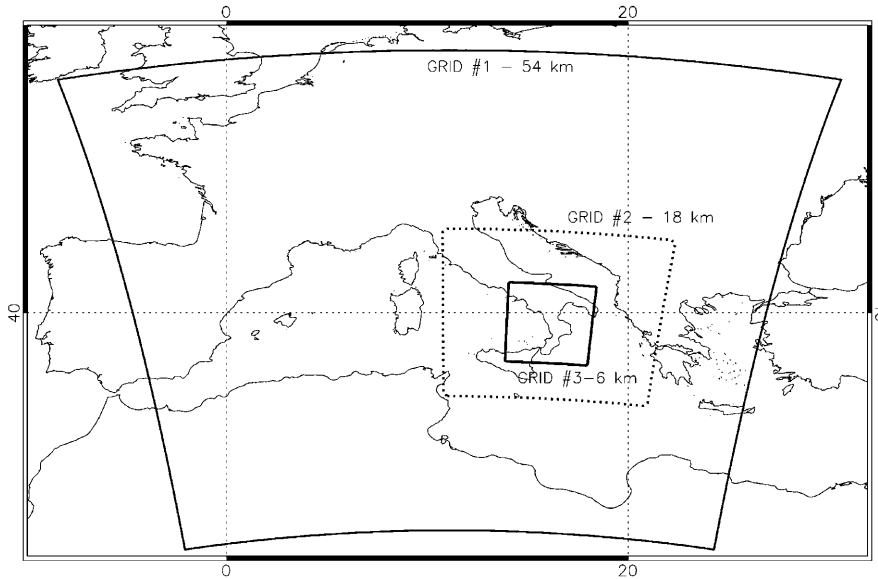
Fig. 4. – MM5 grids. Horizontal resolutions are 54 km, 18 km and 6 km for first, second and third grid, respectively. Two-way interactive procedure is used for grids communication.

layers run about 50–200 m thick, whereas in the middle and upper troposphere they are 1000 m thick.

The parameterization of the surface-atmosphere diabatic processes is described in Walko [16]. Nonconvective precipitation is computed from explicit prognostic equations for eight hydrometeors: total water, rain, pristine, cloud particles, ice, snow, hail and aggregates. Convective precipitation is parameterized following Molinari and Corsetti [17] who proposed a simplified form of the Kuo scheme that accounts for updrafts and downdrafts. Convection parameterization is applied to the first grid only.

2˙2. *MM5 model.* – For details on MM5 model the reader should refer to the relevant bibliography [18]. Grid configuration is shown in fig. 4. Horizontal grid-box spacing is 54 km, 18 km and 6 km for first second and third grids, respectively. All MM5 simulations use the explicit moisture scheme of Hsie *et al.* [19], with improvements to allow for ice-phase microphysics below $0°C$ [20]. Kain-Fritsch [21] cumulus parameterization is applied to 54 km and 18 km resolution grids. Planetary boundary layer is parameterized using the MRF scheme [22].

Twenty-four full sigma levels, unevenly spaced, are used in the vertical with maximum resolution in the PBL. Five-minute–averaged terrain data are analyzed for the 54 km and 18 km grids, while 30 s resolution data are analyzed for third MM5 grid, using a Cressman scheme with 1.5 grid point influence radius. A 30 s land use dataset from USGS is used to initialize 24 surface categories.

2˙3. *QBOLAM model.* – QBOLAM is a parallel version of the hydrostatic, primitive-equation, atmospheric limited area model BOLAM [23] implemented on a 128-processor QUADRICS APE-100 machine.

Equations are discretized on a horizontal Arakawa-C grid, rotated in order to minimize
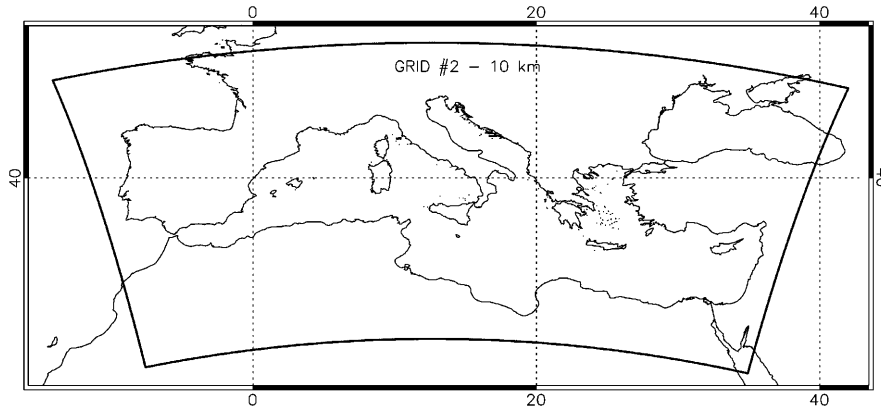
Fig. 5. – QBOLAM second grid. Horizontal resolution is about 10 km. Grid 2 is nested in a parent grid through a one-way interactive communication.

grid anisotropy [2], with 40 levels in the vertical (sigma-level vertical Lorenz [24] scheme). The model uses a split-explicit integration scheme [25]: a forward-backward time integration scheme [26] is used for the term describing gravity waves; a forward-backward Advection Scheme [27] for the horizontal and vertical advection terms. A fourth-order diffusion and a second-order divergence damping are applied. Moreover, some parameterization schemes are simpler than in other BOLAM versions due to massive parallelization issues. Boundary-layer fluxes are represented by analytic formulae [28] based on Monin-Obukhov similarity theory. A simplified radiation scheme [29, 30] is used. Convection parameterization is based on the Kuo [31] scheme. A three-layer soil model provides lower boundary conditions.

In the operational configuration, model is run on two one-way nested grids whose horizontal grid-box spacing is about 0.3° and 0.1°, respectively. The horizontal inner grid domain covers the entire Mediterranean Sea and is reported in fig. 5.

Initial and dynamic boundary conditions, for all models, are from 1200 UTC ECMWF analysis/forecast cycle. Large-scale dataset resolution is half degree.

3. – Verification methodology and data

3˙1. *Scores*. – In this subsection we shortly introduce scores used in this paper. A thorough review can be found in Hamill and Accadia *et al.* [11, 12].

Scores are useful for evaluating deterministic gridded precipitation forecasts as these

TABLE I. – *Contingency table of possible events.*

		Observed	
		Yes	No
	Yes	a	b
Forecast	No	c	d

reported in this paper. They are generated from contingency tables, reported in table I, as follows. Precipitation space is divided in four, mutually exclusive and exhaustive sets: *hits* (a) represent locations number in which both rain forecasts and measurements are greater than or equal to a threshold; *false alarms* (b) represent locations number where the model is above a threshold whereas measurement is under the same threshold; *misses* (c) represent locations number where the measurement is above a threshold and forecast value is under the same threshold; *correct no forecasts* (d) represent locations number where the model and measurements are both under the threshold.

Starting from the contingency table we define BIA:

$$\text{BIA} = \frac{a+b}{a+c}.$$

BIA measures if the model overforecasts or underforecasts precipitation frequency over an area for a selected threshold. If BIA > 1 the model overestimates the precipitation area, if BIA < 1 the model underestimates this area. For a perfect forecast BIA $= 1$.

The most widely used score is Equitable Threat Score (ETS) defined as

$$\text{ETS} = \frac{a - a_\text{r}}{a + b + c - a_\text{r}},$$

where a_r is the expected number of correct forecasts above the threshold in a random forecast where forecast occurrence/non-occurrence is independent of observation/non-observation. It is defined as

$$a_\text{r} = \frac{(a+b)(a+c)}{a+b+c+d}.$$

For a perfect forecast ETS is equal to 1, while it is less than or equal to zero for a useless forecast.

Another score discussed in this paper is the Hanssen-Kuipers score that is the difference between the probability of detection (POD) and false alarm rate (F):

$$\text{HK} = \text{POD} - \text{F} = \frac{a}{(a+c)} - \frac{b}{(b+d)}.$$

POD is the number of yes forecasts when the event occur, whilst F is the frequency of yes forecasts when the events do not occur. HK ranges for -1 to 1. It is independent of the event and non-event marginal distributions. An HK equal to 1 is associated with a perfect forecast, a value of -1 means that *hits* (a) and *correct no forecast* (d) are zero and a value equal to zero is associated with a constant forecast.

3‘2. *Forecast and verification data*. – To verify forecasts, we use data provided by the former Servizio Idrografico e Mareografico Nazionale (SIMN; Italian National Hydrographic and Marigraphic Service) for a total of 75 raingauges spread over Calabria region and primarily located in the major Calabria hydrographic basins. After performing a series of quality-control procedures, observations were accumulated on a daily basis, starting from 0000 UTC.

Not all data were available during the period considered. Average stations number for each month considered in our analysis is reported in fig. 6. For these stations, every
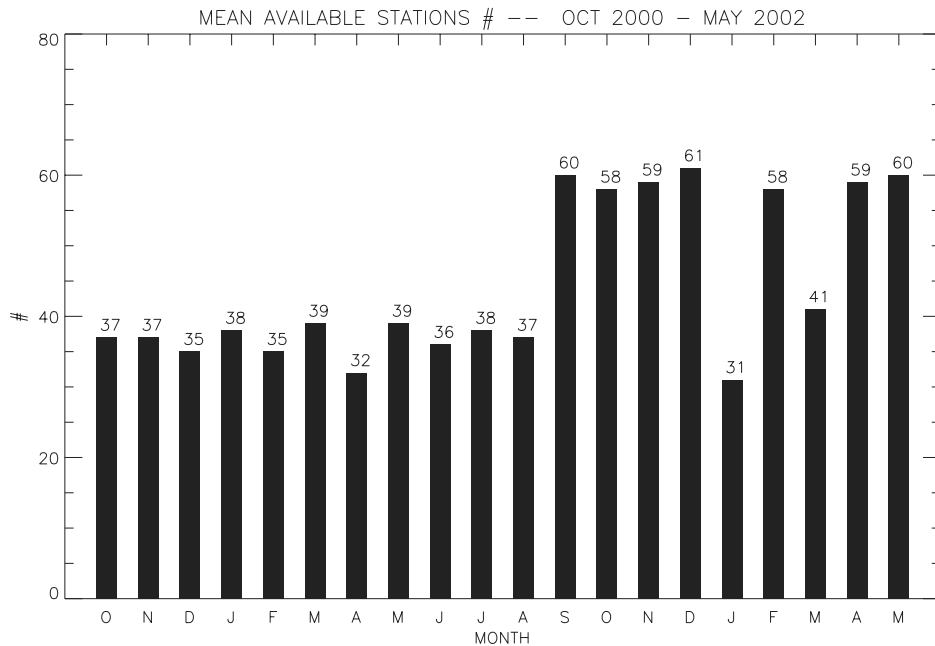
Fig. 6. – Available station number averaged for each month considered in this paper. Mean values are reported over histogram bars.

day of the verification period, a contingency table is computed. Precipitation forecasts are cumulated over 24 h and observations are assigned to the nearest grid point. If more than one station is assigned to a grid point, measurements are averaged and assigned to the grid point.

To compare forecasts coming from different competitor models or from different integration days of the same model, we use the hypothesis test developed by Hamill [11] that is based on a resampling technique. With this technique the idea is to build an artificial data set distribution, consistent with the null hypothesis, by resampling collected data using a computer-based method. The technique is the same used in FABC and Accadia *et al.* [2, 12] and is not further discussed. In this paper we consider a 5% rejection level for all performed hypothesis tests.

This kind of study could be considered incomplete without analyzing possible time-space correlation of forecast errors.

3˙3. *Spatial correlation*. – Because of spatial correlation of forecast errors, individual grid box elements cannot be considered independent. Indeed, considering also the limited extent of the verification area, it is likely that an event missed in some area of the country is missed in a different area too. To verify this hypothesis, we divide Calabria in two parts. The first data set contains raingauges north of 39° N, the second one is formed by remaining stations. This subdivision leaves, roughly, the same amount of stations in northern and southern Calabria. The Spearman rank correlation coefficient between these two areas is computed for all models, and a two-side significance test of its deviation from zero is performed. This value is called p. A rank correlation associated with a p value close to 1 means that correlation is not meaningful while a value of p close to zero

TABLE II. – *Spearman rank correlation of ETS between northern and southern Calabria and p values computed from a two-tailed significance test.*

Threshold (mm)	ETS RAMS	p RAMS	ETS MM5	p MM5	ETS QBOLAM	p QBOLAM
1.0	0.36	< 0.01	0.36	< 0.01	0.21	< 0.01
5.0	0.48	< 0.01	0.27	< 0.01	0.25	< 0.01
10.0	0.61	< 0.01	0.37	< 0.01	0.24	< 0.01
20.0	0.38	< 0.01	0.31	< 0.01	0.17	0.05
30.0	0.18	0.18	−0.20	0.11	0.04	0.69
40.0	0.12	0.52	0.13	0.47	0.10	0.46
50.0	0.56	0.04	0.11	0.64	0.08	0.63

represents a strong correlation.

Table II shows ETS and p values for RAMS, MM5 and QBOLAM models. Non-negligible correlations exist between the two areas and all grid points on a given day must be treated as a single sample.

3˙4. *Temporal correlation*. – Temporal correlations must be assessed because the bootstrap technique, adopted in this paper to compare RAMS, MM5 and QBOLAM models, require the assumption of temporal statistical independence of the errors. ETS, BIA and HK scores were computed for each day of November 2001, the wettest month in our dataset, and a lag-one Spearman rank correlation is performed to verify if there were statistical correlation. Both positive and negative values are found for the models and the two-tailed test p values are shown in table III for BIA and ETS. Results show little forecast error time correlation. Similar results were obtained for HK.

TABLE III. – *Two-side significance test p values for lag-one Spearman rank correlation computed for BIA and ETS for 24 h accumulation time.*

Threshold (mm)	BIA RAMS	BIA MM5	BIA QBOLAM	ETS RAMS	ETS MM5	ETS QBOLAM
1.0	0.70	0.45	0.20	0.82	0.10	0.82
5.0	0.41	0.72	0.35	0.55	0.54	0.23
10.0	0.55	0.74	0.79	0.83	0.48	0.85
20.0	0.38	0.55	0.23	0.44	0.93	0.39
30.0	0.58	0.87	0.99	0.54	0.39	0.74
40.0	0.34	0.79	0.25	0.89	0.47	0.59
50.0	0.87	0.97	0.32	0.48	0.67	0.76

4. – Results

In this paper we consider the following rainfall thresholds in a day: 1 mm, 5 mm, 10 mm, 20 mm, 30 mm, 40 mm, 50 mm. They are a compromise between precipitation falling in a day over Calabria and the field variability. First of all it must be noticed that the comparison reported in this paper refers to the models set-up and to the area considered. LAM performances are strictly related to these features and results obtained in this work cannot be extended to other areas.

Figure 7a shows first integration day BIA for MM5 and RAMS. Models overestimate precipitation for all thresholds. BIA is larger for MM5 and highlights its greater "wet-ness" compared to RAMS. A contingency table analysis reveals that *hits* and *false alarms* are greater for MM5 for all thresholds. Scores are different at 5% level for 1 mm, 40 mm and 50 mm thresholds.

ETS values for first integration day, fig. 7b, are well above the useless value (ETS $\leqslant 0$) for all thresholds but they decrease with increasing rainfall. This is a typical behaviour of LAM QPF forecast. Scores are similar and differences are not significant at 5% test level.

HK is shown in fig. 7c for RAMS and MM5 first integration day. Both models are well above the useless threshold (HK $\leqslant 0$) and there is an improvement of MM5 performances compared to RAMS. Differences are never statistically significant at 5% level and MM5 improvement is a consequence of larger MM5 BIA. To better understand this point, we remember that HK is the difference between POD and F. POD is a function of *hits* (a) and *misses* (c), while F is a function of *false alarms* (b) and *correct no forecasts* (d). Let ($\Delta a, \Delta b, \Delta c, \Delta d$) be the differences between MM5 and RAMS for (a, b, c, d) at a fixed threshold, it can be easily shown that $\Delta a, \Delta b, \Delta c$ and Δd are related as follows:

$$(1) \qquad\qquad\qquad \Delta a = -\Delta c \,; \qquad\quad \Delta b = -\Delta d \,.$$

For our models set-up, larger MM5 BIA values produce an increase of *hits* and *false alarms* and an equal decrease of *misses* and *correct no forecasts*. While *hits* change in contingency tables increases POD, *false alarms* variation produces a minor change in F because $d \gg b$. Indeed, the number of locations with no forecast/observed precipitation, *i.e.* contingency table d term, are an order of magnitude greater than a, b and c. So, even if there is an increase in *false alarms,* F is dominated by *correct no forecast*. Hence, HK is larger for MM5 compared to RAMS because changes in contingency tables increase POD and leaves F nearly unchanged. This point must be considered in decision processes.

RAMS and QBOLAM BIA for first forecast day is shown in fig. 8a. QBOLAM has larger BIA than RAMS for all thresholds and differences are statistically significant. In-spection of contingency tables shows that BIA increase is due to larger *hits* and *false alarms* forecast by QBOLAM. Despite to larger *hits*, RAMS ETS are larger than QBO-LAM for first integration day, as reported in fig. 8b. This is related to the increase of QBOLAM *false alarms* that lowers ETS values.

Figure 8c reports HK score for RAMS and QBOLAM first day forecast. HK are similar for both models and differences are never statistically significant, if we exclude 1 mm and 50 mm thresholds. This result is attained, however, by large differences in POD and F (not shown). Indeed QBOLAM has larger POD and larger F than RAMS but differences between POD and F, *i.e.* HK score, are similar. More in detail, QBOLAM POD is larger than RAMS for all thresholds and differences are significant up to 20 mm
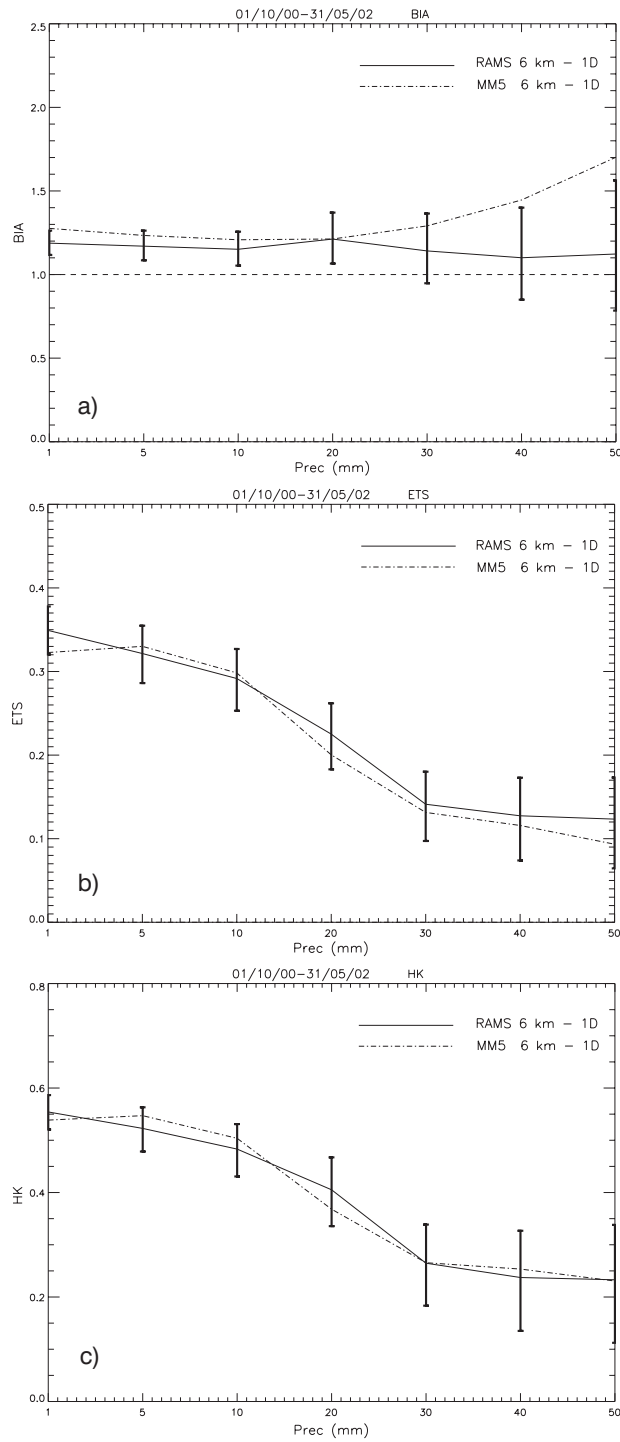
Fig. 7. – a) RAMS and MM5 BIA for 1D forecast; b) as in a) for ETS; c) as in a) for HK.
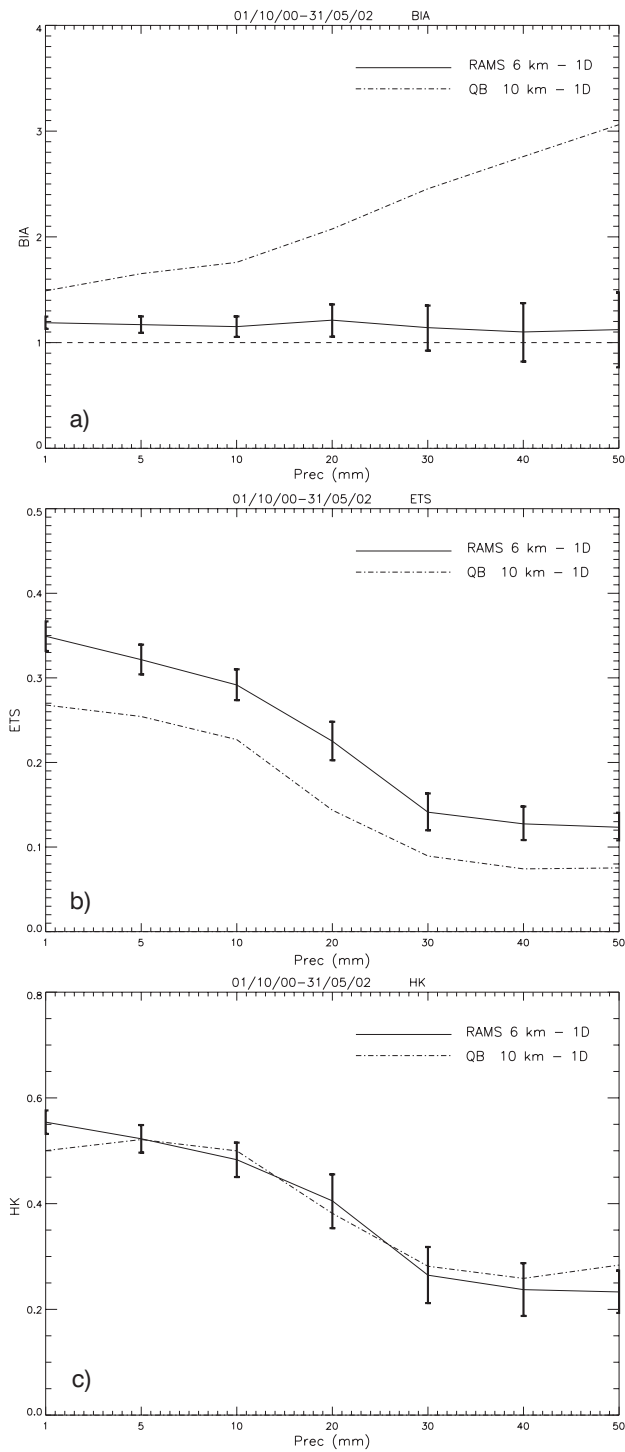
Fig. 8. – a) As in fig. 7a) for RAMS and QBOLAM; b) as in fig. 7b) for RAMS and QBOLAM; c) as in fig. 7c) for RAMS and QBOLAM.
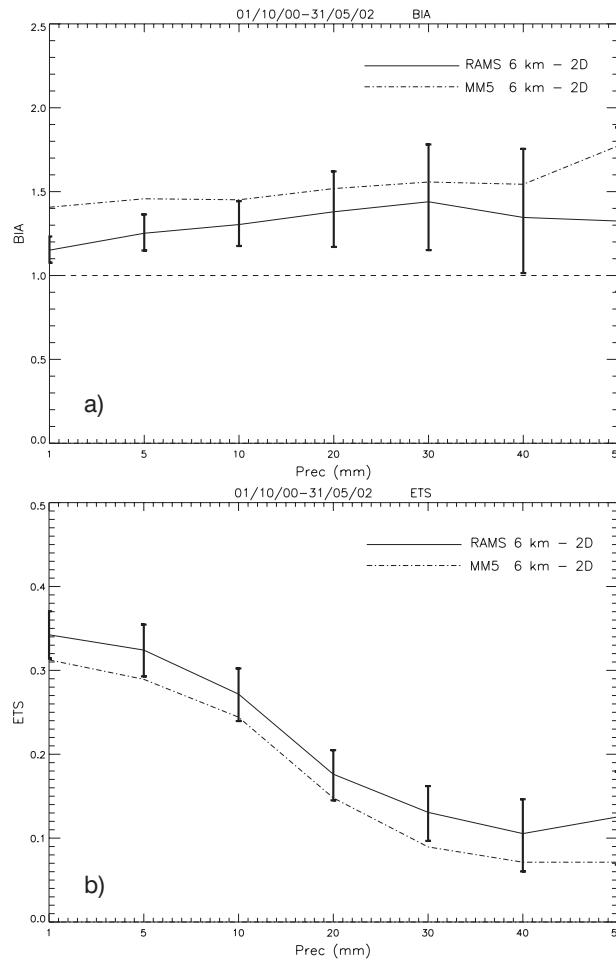
Fig. 9. – a) RAMS and MM5 BIA for 2D forecast; b) as in a) for ETS.

thresholds and for 50 mm daily rainfall. QBOLAM gives also more *false alarms* at all thresholds and differences are statistically significant for all thresholds.

RAMS and MM5 BIA for the second integration day, fig. 9a, are greater than one for both models. Larger values are obtained for MM5 and differences are statistically significant for 1 mm, 5 mm and 10 mm daily rainfall. From figs. 7a and 9a it follows that MM5 has larger BIA than RAMS for models set-up used in this paper. In addition there is a drift to "wetness" with increasing forecast time.

Figure 9b shows RAMS and MM5 ETS for 2D forecast. Comparison with fig. 7b evidences a decreased performance for the second integration day, mainly for MM5. RAMS outperforms MM5 for all thresholds and differences are statistically significant for 1 mm, 5 mm and 30 mm daily rainfall. HK, not reported, shows similar values for thresholds less than 20 mm whilst for higher precipitation RAMS has better performances. Differences are never statistically significant at 5% test level.

Compared to the first integration day, QBOLAM BIA show a sensible improvement
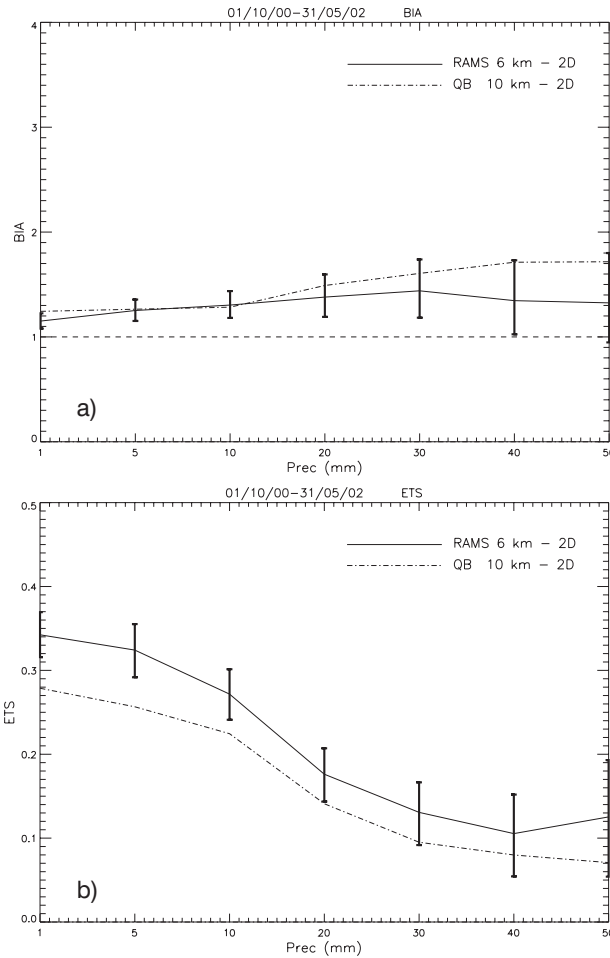
Fig. 10. – a) RAMS and QBOLAM BIA for 2D forecast; b) as in a) for ETS.

for the second integration day, as shown in fig. 10a that reports BIA score for RAMS and QBOLAM 2D. Performances are similar for both models and differences are not statistically significant, excluding 1 mm threshold. Considering ETS, fig. 10b, RAMS has better performances for all thresholds and they are statistically significant up to 20 mm threshold. ETS differences are reduced compared to 1D forecast, fig. 8b.

In the previous discussion MM5 and QBOLAM models were compared to RAMS that was assumed as "reference" model. It is interesting to discuss, briefly, MM5 and QBO-LAM comparison. For first day, MM5 and QBOLAM BIA differences are statistically significant for all thresholds and MM5 scores better. This is due to the overestimation of *hits* and *false alarms* that characterize QBOLAM forecast for this day, as noticed above. First integration day ETS shows better performances for MM5. Differences are statistically significant up to 40 mm thresholds. QBOLAM has larger *hits* and *false alarms* than MM5 that result in larger POD and F. There follows a greater HK for QBOLAM but differences are not statistically significant, excluding 1 mm threshold.

For the second integration day, BIA is better for QBOLAM. Results are statistically

significant up to 20 mm threshold. Reduced differences, compared to 1D forecast, are found for ETS and they are statistically significant for 1 mm and 5 mm daily rainfall. QBOLAM HK is lower than MM5 up to 20 mm threshold and differences are statistically significant for 1 mm, 5 mm and 10 mm.

A discussion of QBOLAM performances must take into account the fact that model grid-size is about 10 km compared to 6 km used for RAMS and MM5. During the past two decades, a number of studies have examined the effects of horizontal resolution on forecast accuracy of QPF and few studies have demonstrated that high-resolution forecast produces better-behaved and more realistic structures, evaluated subjectively [32]. High resolution appears to be most useful for strongly forced convection associated with fronts, topography, etc. [33, 34]. However few studies have demonstrated that forecast accuracy, measured objectively, can decrease if grid spacing is less that 15–10 km. Even if it is not possible, by the simulations considered in this paper, to separate effects due to different models physical/numerical parameterizations from those introduced by enhanced horizontal resolution, comparison between RAMS, MM5 and QBOLAM suggests that this is not the case for Calabria and a resolution < 10 km is an important factor for objective QPF scores improvement.

A second important issue is that QBOLAM is hydrostatic. From Calabria main features, fig. 1, there are valleys and mountain slopes where nonhydrostatic effects could be important.

These two features help to explain differences between RAMS/MM5 and QBOLAM scores. The former are nonhydrostatic and have a greater horizontal resolution. This result confirms, at least partially, the main conclusion of FABC.

Considering models performances as a whole, differences are found for ETS and HK scores but they are usually inside the confidence interval or close to its bounds. Largest differences are for BIA.

To gain more insight into models differences, we applied the BIA adjustment technique to compare forecasts. Indeed, as noticed in previous works [11,12,35] comparison of ETS and HK from competing forecasts may be misleading if their biases are dissimilar. "Wet" BIA may result in a comparatively larger skill score compared to "dry" BIA [35]. Hence we investigate to what extent differences found in ETS and HK are an effect of biases differences.

BIA adjustment technique is applied using RAMS as "reference" model. Basically we achieve similar biases by adjusting the forecast precipitation thresholds of MM5 and QBOLAM in order to obtain similar values to RAMS.

We do not report results for RAMS and MM5 1D forecast because ETS and HK are similar and BIA adjustment reduces even more these differences. Figures 11a and 11b report BIA and ETS for RAMS and MM5 2D forecast when BIA adjustment is applied. Comparing ETS values, fig. 11b, better performances are achieved by RAMS and differences are statistically significant, if we exclude 1 mm and 10 mm thresholds. Even if RAMS has better performances than MM5 differences are inside the error bars or close to their limits. Comparing fig. 11b and fig. 9b it follows that BIA adjustment produces minor changes for ETS. For HK, not shown, changes are larger and show better RAMS performances. Differences are statistically significant, at 5% level, for 5 mm, 30 mm and 50 mm thresholds.

Figure 12a and 12b show BIA and ETS for RAMS and QBOLAM (first day forecast) when BIA adjustment is applied. Comparing these figures with 8a and 8b it follows that, despite large changes in BIA, ETS values are almost unchanged. In this case, however, *false alarms* and *hits* are both reduced for 1D QBOLAM adjustment compared to the
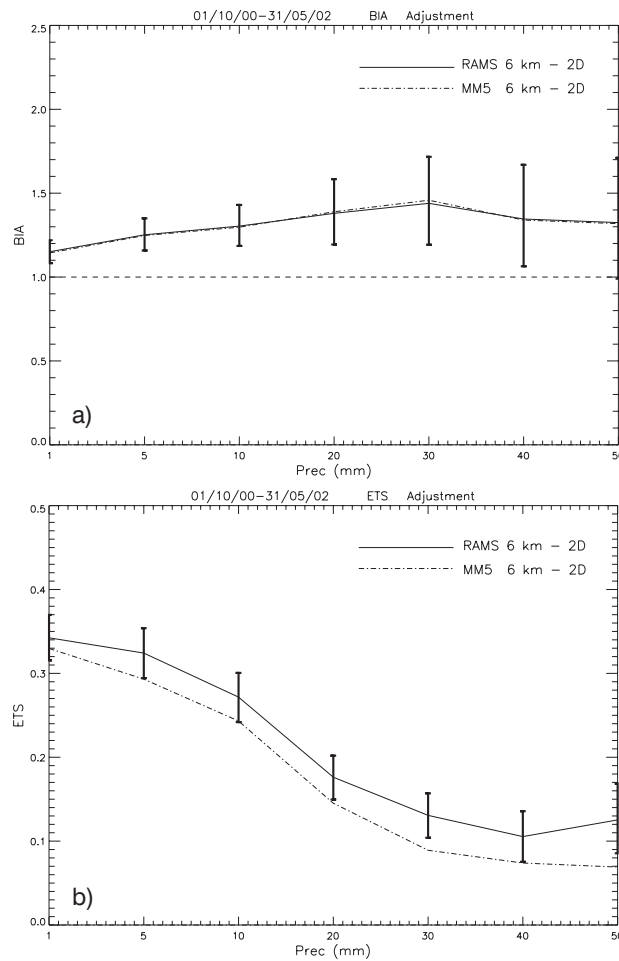
Fig. 11. – a) RAMS and MM5 BIA after adjustment for second day forecast; b) as in a) for ETS.

nonadjusted case. For 2D forecast, as can be inferred comparing with results shown in figs. 10a and 10b, adjustment produces minor changes for ETS due to lower BIA differences compared to 1D forecast.

In summary BIA adjustment for MM5 and QBOLAM changes elements in contingency tables but results for ETS and HK are similar to nonadjusted case.

It is well known that meteorological models performances lower with forecast time. This problem was also considered in FABC with respect to RAMS QPF over Calabria. Considering this issue, this paper represents an extension of FABC for two reasons. First of all, the period considered is extended to 20 months, second scores are reported for MM5 and QBOLAM too.

In order to study this key weather forecasting issue, it is necessary to assess whether scores differences between competing models are statistically significant. In this case, however, competitors are the 1D and 2D forecasts for each model so the bootstrap hypothesis test, at 5% level, was applied to each model, for the first and second integration
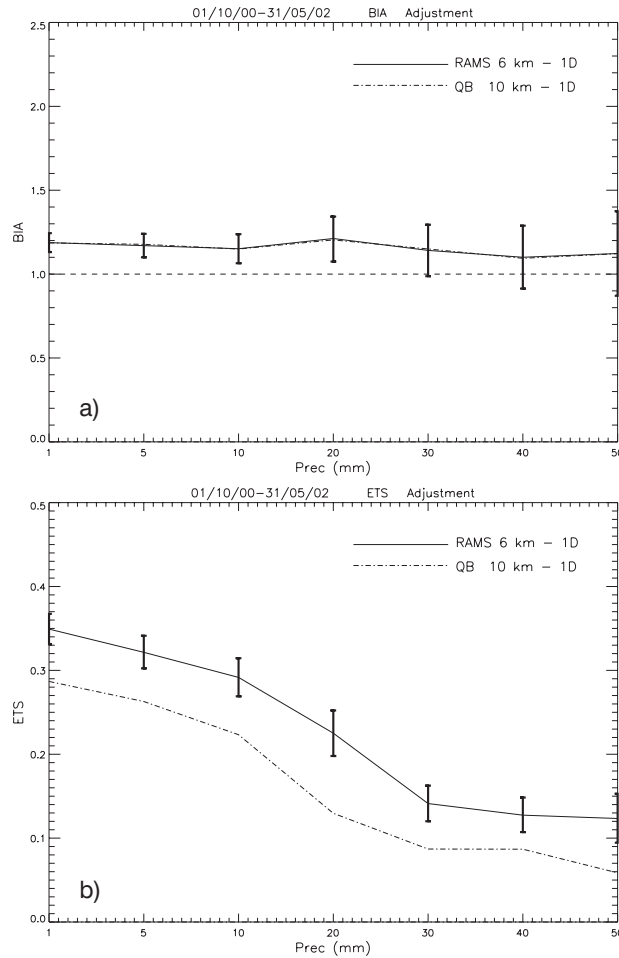
Fig. 12. – a) RAMS and QBOLAM BIA after adjustment for first day forecast; b) as in a) for ETS.

day outputs. Results are shown for ETS only, whereas BIA is briefly discussed.

RAMS ETS scores are reported in fig. 13a. There is a performance decrease for this model. In fact, ETS is usually lower for the second forecast day. Differences are statistically significant for 10 mm, 20 mm and 40 mm thresholds. However performance deterioration is not large, ETS values are close to error bars or inside their range and forecast is useful (ETS $\geqslant$ 0) for all thresholds for the second integration day. So, even if there are signs of performance lowering, we can conclude that results between 1D and 2D forecast are similar for RAMS model. BIA score (see figs. 7a, 9a) shows a drift to "wetness" and differences are significant for all thresholds but 1 mm and 50 mm. Greater biases are due, however, to a *false alarms* increase while *hits* are nearly stationary.

A different behaviour is shown by MM5 model, fig. 13b. In this case ETS values are lower for the second integration day for all thresholds. Differences are statistically significant for all thresholds but 1 mm and 50 mm. The analysis of contingency tables reveals that MM5 performance lowering is related mainly to a *false alarms* increase and
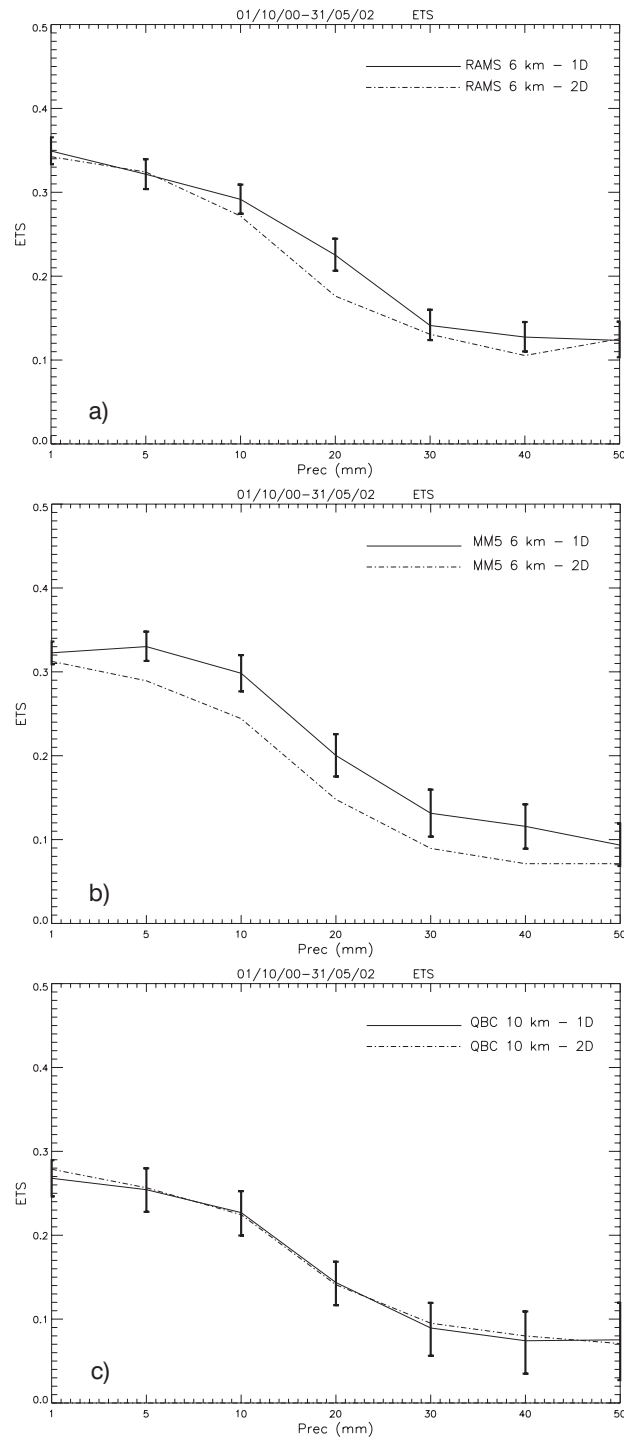
Fig. 13. – a) RAMS ETS for 1D and 2D forecasts; b) as in a) for MM5; c) as in a) for QBOLAM.

to a decrease of *hits* between the first and second integration day. This determines also a model drift to "wetness" mainly for thresholds less that 40 mm, as can be also inferred comparing figs. 7a and 9a. BIA differences between first and second day forecasts are significant up to 40 mm thresholds and MM5 second day BIA is larger for all thresholds.

A completely different behaviour is shown by QBOLAM, fig. 13c. ETS values are not statistically different between the first and second integration day and are very similar. In addition, as reported in figs. 8a and 10a, there is a large improvement of second day BIA compared to the first integration day. This is related to a decrease in both *hits* and *false alarms* between the first and second integration days that improves BIA leaving ETS constant.

5. – General comments

In the previous discussion we have highlighted differences between models rather than their similarities. However, despite differences that characterize models physical/numerical parameterizations and horizontal resolutions (we believe that vertical resolutions are enough for all models to represent basic processes involved in QPFs over Calabria), results are often similar and scores are close to error bars or inside their range. This is, at least partially, due to the synoptic conditions that are shared by all models and suggests the need of mesoscale/regional asynoptic measurements to better exploit limited area models parameterizations, mainly for complex topography countries.

We have presented results for several objective scores because the goals and measures of verification are inevitably dependent on the needs of the users. A satisfactory mesoscale forecast for one might be a complete failure for another. For example a farmer needing to make a decision about irrigation might not care when precipitation fell, as long as it occurs the day following planting. On the other hand, there are applications where exact timing and positioning is the main issue, as in cases of ceremonies or important sport events. Thus different scores can be useful to give value of high-resolution numerical forecast according to its applications.

We have presented results for traditional objective scores based on verification at fixed observing locations or grid points. This traditional approach is greatly influenced by timing and spatial errors as well as deficiencies of the observing network used for comparison. Several studies that deal with precipitation forecast verification show that increased horizontal resolution generally produces better defined mesoscale structures and gradients which are more realistic. On the other hand, errors in timing or position of these mesoscale features usually amplify errors evaluated objectively using "classical" verification scores. Thus, even if structures are more realistic in high-resolution simulations, there is a point where the existence of timing and position errors may result in poorer objective scores for high-resolution domain. Even if we cannot definitively assess its value, results of this paper, together with those of FABC, indicate that the point of diminishing returns could be less than 6 km for Calabria and suggest that a "one size fits all" approach, in which all areas of a country are run with the same resolution, may not be an efficient use of computer resources.

As horizontal resolution approaches the point where objective scores start to lower, increasing emphasis should be given to evaluating short-term forecasts coming from different models, *i.e.* short-term ensemble forecast. There are several positive results about this approach [36, 37], however the story is far from the end and the value of short-term ensemble forecast still has to be assessed. At CRATI Scrl work is in progress to assess the performances of a multimodel ensemble short-term forecast, based on RAMS and MM5, compared to high-resolution forecast.

6. – Conclusions

In the framework of the project "Sviluppo di Distretti Industriali per le Osservazioni della Terra" we investigated the performances of RAMS, MM5 and QBOLAM limited area models for quantitative precipitation forecast over Calabria. Three main issues are discussed: performances of LAMs for QPF over Calabria, comparison between models and performances decrease with increasing forecast time. Large-scale forcing is shared by all models and differences are due to physical/numerical parameterizations and resolutions. Results can be summarized as follows:

– RAMS, MM5 and QBOLAM are useful for all QPF thresholds considered in this paper and for both integration days. BIA shows a tendency to "wetness" for all models. This behaviour is more evident for QBOLAM first integration day and MM5 second day forecast.

– About model intercomparison, differences are found mainly for BIA while they are reduced for other scores. Excluding BIA, RAMS and MM5 have similar scores for the first integration day but their difference increases with time.

– Differences between RAMS/MM5 and QBOLAM are larger. Even if it is difficult to separate different contributions of models physical/numerical parameterizations and horizontal grid resolutions to scores, differences are likely related to two main issues: the first one is horizontal resolution, the second QBOLAM hydrostatic approximation.

– Different behaviours characterize models performance with time. RAMS has a larger BIA for the second integration day, however it exhibits minor changes in ETS and HK. MM5 shows forecast deterioration with increasing time. Indeed as time progresses there are more *false alarms* and less *hits*. QBOLAM has better performances for the second integration day due to a reduction of *hits* and *false alarms*. This gives better BIA and stationary ETS.

* * *

REFERENCES

[1] Speranza A., Accadia C., Casaioli M., Mariani S., Monacelli G., Inghilesi R., Tartaglione N., Ruti P. M., Carillo A., Bargagli A., Pisacane G., Valentinotti F. and Lavagnini A., *Nuovo Cimento*, **27** (2004) 329.
[2] Federico S., Avolio E., Bellecci C. and Colacino M., *Nuovo Cimento C*, **26** (2003) 663.
[3] Buzzi A., Tartaglione N. and Malguzzi P., *Mon. Weather Rev.*, **126** (1998) 2369.
[4] Rotunno R. and Ferretti R., *J. Atmos. Sci.*, **58** (2001) 1732.
[5] Smith R. B., *The influence of mountains on the atmosphere*, in *Advances in Geophysics*, Vol. **21** (Academic Press) 1979, pp. 87-230.

[6] FEDERICO S., BELLECCI C. and COLACINO M., *Nuovo Cimento C*, **26** (2003) 7.
[7] LORENZ E. N., *J. Atmos. Sci.*, **20** (1963) 130.
[8] LORENZ E. N., *Tellus*, **17** (1965) 321.
[9] MOLTENI F. and PALMER T. N., *Q. J. R. Meteorol. Soc.*, **119** (1992) 269.
[10] TOTH Z. and KALNAY E., *Bull. Am. Meteorol. Soc.*, **74** (1993) 2317.
[11] HAMILL T. M., *Weather Forecasting*, **14** (1999) 155.
[12] 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. and CREMONINI R., *Nuovo Cimento C*, **26** (2003) 61.
[13] PIELKE R. A., COTTON W. R., WALKO R. L., TREMBACK C. J., LYONS W. A., GRASSO L. D., NICHOLLS M. E., MURRAN M. D., WESLEY D. A., LEE T. H. and COPELAND J. H., *Meteorol. Atmos. Phys.*, **49** (1992) 69.
[14] COTTON W. R., PIELKE R. A. SR., WALKO R. L., LISTON G. E., TREMBACK C. J., JIANG H., MCANELLY R. L., HARRINGTON J. Y., NICHOLLS M. E., CARRIO G. G. and MCFADDEN J. P., *Meteorol. Atmos. Phys.*, **82** (2003) 5.
[15] WALKO R. L., TREMBACK C. J., PIELKE R. A. and COTTON W. R., *J. Appl. Meteorol.*, **34** (1994) 994.
[16] WALKO R. L., BAND L. E., BARON J., KITTEL T. G., LAMMERS R., LEE T. J., OJIMA D., PIELKE R. A. SR., TAYLOR C., TAGUE C., TREMBACK C. J. and VIDALE P. L., *J. Appl. Meteorol.*, **39** (2000) 931.
[17] MOLINARI J. and CORSETTI T., *Mon. Weather Rev.*, **113** (1985) 485.
[18] GRELL G. A., DUHIA J. and STAUFFER D. R., A description of the fifth-generation Penn State / NCAR mesoscale model (MM5), NCAR Technical note - NCAR/TN-398+STR, (1994).
[19] HSIE E. Y., ANTHES A. and KEYSER D., *J. Atmos. Sci.*, **41** (1984) 2581.
[20] DUHIA J., *J. Atmos. Sci.*, **46** (1989) 3077.
[21] KAIN J. S. and FRITSCH J. M., *J. Atmos. Sci.*, **47** (1990) 2784.
[22] HONG S. Y. and PAN H. L., *Mon. Weather Rev.*, **124** (1996) 2322.
[23] BUZZI A., FANTINI M., MALGUZZI P. and NEROZZI F., *Meteorol. Atmos. Phys.*, **53** (1994) 53.
[24] LORENZ E. N., *Tellus*, **12** (1960) 364.
[25] GADD A. J., *Q. J. R. Meteorol. Soc.*, **104** (1978) 569.
[26] MESINGER F., *Atmos. Phys.*, **50** (1977) 200.
[27] MALGUZZI P. and TARTAGLIONE N., *Q. J. R. Meteorol. Soc.*, **125** (1999) 2291.
[28] LOUIS J. F., TIEDTKE M. and GELEYN J. F., *A short history of the PBL parameterization at ECMWF*, in *Proceedings of ECMWF Workshop on PBL Parameterization*, Readings 25-27 Nov. 81 (1982), pp. 59-80.
[29] PAGE J. K., *Prediction of Solar Radiation on Inclined Surfaces* (D. Reidel Publishing Company) 1986.
[30] RUTI P. M., CASSARDO C., CACCIAMANI C., PACCAGNELLA T., LONGHETTO A. and BARGAGLI A., *Beitr. Phys. Atmos.*, **70** (1997) 201.
[31] KUO H. L., *J. Atmos. Sci.*, **31** (1974) 1232.
[32] KAIN J. S., BALDWIN M. E., JANISH P. R., WEISS S. J., KAY M. P. and CARBIN G. W., *Weather Forecasting*, **18** (2003) 847.
[33] COLLE B. A., MASS C. F. and WESTRICK K. J., *Weather Forecasting*, **6** (2000) 730.
[34] MASS C. F., OVENS D., WESTRICK K. and COLLE B. A., *Bull. Am. Meteorol. Soc.*, **83** (2002) 407.
[35] MASON J., *Aust. Meteorol. Mag.*, **37** (1989) 75.
[36] STENSRUD D. J., BROOKS H. E., DU J., TRACTON M. S. and ROGERS E., *Mon. Weather Rev.*, **127** (1999) 433.
[37] GRIMIT E. P. and MASS C. F., *Weather Forecasting*, **17** (2002) 192.