

---

# Intercomparison of mesoscale meteorological models for precipitation forecasting

E. Richard<sup>1</sup>, S. Cosma<sup>1</sup>, R. Benoit<sup>2</sup>, P. Binder<sup>3</sup>, A. Buzzi<sup>4</sup> and P. Kaufmann<sup>3</sup>

<sup>1</sup>Laboratoire d'Aérodynamique, CNRS/UPS, 14 Avenue E. Belin, F-31400 Toulouse, France

<sup>2</sup>Recherche en Prévision Numérique, 2121 TransCanada Highway, Dorval, QC, Canada H9P 1J3

<sup>3</sup>MeteoSwiss, P.O. Box 514, CH-8044 Zürich, Switzerland

<sup>4</sup>Istituto di Scienze dell'Atmosfera e del Clima, CNR, Via Gobetti 101, I-40129 Bologna, Italy

Email for corresponding author: rice@aero.obs-mip.fr

---

## Abstract

In the framework of the RAPHAEL EU project, a series of past heavy precipitation events has been simulated with different meteorological models. Rainfall hindcasts and forecasts have been produced by four models in use at various meteorological services or research centres of Italy, Canada, France and Switzerland. The paper is focused on the comparison of the computed precipitation fields with the available surface measurements for three meteorological situations which lead to severe flash floods over the Toce-Ticino catchment in Italy (6599 km<sup>2</sup>) or the Ammer catchment (709 km<sup>2</sup>) in Germany. All four models reproduced the occurrence of these heavy precipitation events. The accuracy of the computed precipitation appears to be more case-dependent than model-dependent. The sensitivity of the computed rainfall to the boundary conditions (hindcast v. forecast) was rather weak, indicating that a flood forecasting system based on a numerical meteo-hydrological simulation was feasible in an operational context.

**Keywords:** meteorological models, precipitation forecast

## Introduction

Advanced predictions of heavy rainfall are an essential prerequisite for physically based hydrological forecasts and are fundamental for early warning of floods that can cause important material and human damage, especially in mountainous areas where the meteorological and hydrological risks might be very acute. Despite the increasing skill and spatio-temporal resolution of numerical weather prediction models, the quantitative forecast of precipitation is still a very challenging and difficult task.

An accurate rainfall forecast requires the understanding of the complex and highly nonlinear interaction between convection and the synoptic environment (Emanuel and Raymond, 1993): the large-scale environment determines and modulates the intensity of convection while the convective transport of moisture and momentum affects synoptic structures (as cyclones, fronts, cut-off lows,...). Among the different approaches for studying the interaction between convection and large scales, the use of high resolution models that can simulate cumulus convection

explicitly has produced promising results. The horizontal mesh size needed for such simulations must be smaller than a few kilometres (Molinari and Dudek, 1992), implying that the hydrostatic approximation currently used in most operational models must be discarded. Such models have recently emerged in the research community and are being implemented progressively at different meteorological centres.

In the framework of the RAPHAEL project, the results of four atmospheric models have been used to explore the potential benefit of a coupled meteo-hydrological modelling system for flood forecasting. These four models are fairly representative of the numerical tools currently available for precipitation forecasting. They range from classical, well proven, hydrostatic models to more sophisticated research models. In this paper, precipitation values obtained with these different numerical tools are compared and checked against the available surface observations. The comparison is carried out for three episodes of heavy rainfall that occurred either in the Southern Alps over the Toce-Ticino

catchment, or in the Northern Alps over the Ammer catchment. This study took place in the preparation phase of the Mesoscale Alpine Programme (MAP, Bougeault *et al.*, 2001).

## The selected episodes

For the RAPHAEL project, seven flood events were selected, three on the Ammer catchment (709 km<sup>2</sup>) in the Bavarian Alps and four on the Ticino-Toce area (6599 km<sup>2</sup>) upstream of Lago Maggiore in north-east Italy. All of them have been simulated with a coupled meteo-hydrological approach involving different atmospheric models and different hydrological models. All the hydrological models simulated the seven episodes. But in view of the large cost involved in operating high-resolution meteorological models, the simulations of the seven precipitation events have been shared among the different meteorological models. However, some important events have been run by all the atmospheric models to provide a common basis for an intercomparison of the meteorological results. These events are referred to as TT1, TT3 and AM3. The first two affected the Toce-Ticino area whereas the latter occurred over the Ammer region.

### TT1 : THE BRIG FLOOD

From 22 to 24 September, 1993, several meso-scale convective systems generated intense precipitation events from the eastern Pyrenees to the south-western Alps resulting in local flooding and severe damage, particularly over the city of Brig in Switzerland along the Rhone river. The average 24 h precipitation over the Italian Piedmont region was 90 to 120 mm from the 22nd to the 24th with maxima of 120 to 160 mm just west of Ticino province at Pallenzano for a three-day total of 3 400 mm.

### TT3 : THE PIEDMONT FLOOD

The disastrous flood that affected mainly the southern Piedmont area in November 1994 was associated with heavy rain that fell in a relatively small area in successive episodes spanning the period from 3 to 6 November 1994. On the 4th, precipitation peaks were present around and north of Genoa, with values consistently exceeding 100 mm accumulated in 24 hours. On the following day, the wettest area where values larger than 200 mm were reported, was displaced some 50 km to the west, corresponding to the highest part of river Tanaro and Bormida basins (Maritime Alps). Another distinct maximum is visible also over the Alps north-west of Turin, with some stations reporting more than 200 mm in 24 hours.

### AM3 : SUMMER CONVECTION

The episode selected for the Ammer region took place from 17 to 20 July, 1997. Precipitation resulted from summer convective storms that occurred on consecutive days. The intensities were not extremely high, but the final storm happened over saturated soil and led to the flooding of the Ammer catchment.

## The meteorological models

For the above episodes, precipitation computations have been performed with four different meteorological models: SM, BOLAM3, MC2, and MESO-NH.

The Swiss Model SM is the operational model used by Meteo-Swiss for short range weather prediction. It is a hydrostatic, limited-area model with horizontal resolution of 0.125 degrees (i.e. 14 km). This model has been run according to its operational setup (see Kaufmann *et al.*, 2002).

BOLAM3 is developed at the ISAC-CNR of Bologna for research purposes but is also used operationally by different Italian weather services. This model is formally based on a hydrostatic system of equations but includes non-hydrostatic corrective terms (Buzzi *et al.*, 1994; Buzzi and Malguzzi, 1997). The BOLAM3 simulations were carried out over two (and partially three) nested domains with increasing horizontal resolutions of 35 km, 10 km and 3.5 km.

MC2 is the Mesoscale Compressible Community model developed at RPN in Canada. It is a non-hydrostatic, fully compressible model, integrated with a semi-implicit, semi-Lagrangian numerical scheme which makes it very efficient in terms of computing cost (Benoit *et al.*, 1997). This model was used very successfully in real time and at very high resolution during the MAP field experiment (Benoit *et al.*, 2001). The MC2 simulations were carried out over two (and partially three) nested domains of horizontal resolutions of 50 km, 10 km and 3 km.

The Meso-NH model is a joint development of Météo-France and Laboratoire d'Aérodynamique. The model is non-hydrostatic, based upon an anelastic system of equations and is able to simulate all scales ranging from turbulent large eddies to the synoptic scale (Lafore *et al.*, 1998; Stein *et al.*, 2000). The nesting procedure was based upon two (and partially three) domains with horizontal mesh-sizes of 50 km, 10 km and 2 km.

## Initial and boundary conditions

Mesoscale meteorological models are limited area models (as opposed to global models). Therefore, they require not only initial conditions but also time-dependent boundary

conditions. The initial conditions are obtained by interpolating on the model grid a meteorological analysis which results from a data assimilation scheme combining model and observations. These analyses are performed on the global scale by the main meteorological services such as for instance the European Centre for Medium Range Weather Forecast (ECMWF) and are usually available every six hours. For RAPHAEL, mainly ECMWF analyses have been used except for SM which operationally uses the Europa Model (EM) analysis provided by the German weather service. Some of the MC2 simulations were also based upon the EM analyses.

Concerning the boundary conditions, two types of runs have been performed, the so-called forecast and analysis modes. These terms are based on the source of the lateral boundary data used for the coarser run. If they come from a forecast provided by a large-scale or global scale model, then the mesoscale run is also a forecast, in the sense that it uses observed data only for its initial state. If the lateral boundary data come from an assimilation scheme (such as the ECMWF analyses), then the mesoscale run is no longer a forecast but a hindcast, in the sense that it uses observed data indirectly for both initial and boundary conditions. Hindcast allows the re-creation of a past event which is never possible in real-time forecasting. The quality of the lateral boundaries is expected to be higher in the analysis mode, while making some tests of the forecast mode is crucial to demonstrate the benefit of the numerical approach for predicting the floods. Because knowledge and experience with meteo-hydrological coupled models is rather limited, it was decided that for RAPHAEL it would be safer to proceed along the 'hindcast' route for the majority of the mesoscale model runs while also performing forecasting experiments for a subset of cases for intercomparison and to provide a number of examples in 'pseudo-operational' mode. In these cases, the possible degradation exhibited by the forecast compared with the analysis runs will be ascribed to the large scale or global scale model forecast errors.

## Numerical strategy

At the mesoscale, the typical range of a forecast is 24 h to 48 h and it is generally considered that, during the first few hours, the results are not very reliable because of the so-called model 'spin-up time' (Kasahara *et al.*, 1991). To cover the three- to four-day period of the selected precipitation events, it was necessary to perform a sequential set of nested runs. This numerical strategy, based upon a one-way nesting procedure, is illustrated in Fig.1 for the TT1 and TT3 episodes. For the TT1 (TT3) episode, a first coarse grid ( $\Delta x = 50$  km for MC2 and MESONH, and  $\Delta x = 35$  km for

BOLAM3) simulation starts on 21 September 1993, 12 UTC (3 November 1994, 00 UTC), i.e. some 12 hours before the onset of precipitation. Then a nested simulation with increased resolution ( $\Delta x = 10$  km) starts 6 hours later and is integrated for 30 hours. This sequence is repeated three times to cover the complete event. The starting times of each sequence are selected in such a way that the 10 km resolution simulations overlap by 6 hours. Each of the finest resolution runs provides hourly outputs of the meteorological surface parameters (including precipitation). Blanking the first 6 hours of each sequence and joining the different sequences led to a final file that covers the 72 hours of the precipitation event.

Except for the resolution of the coarse simulations, this strategy was identical for BOLAM3, MC2 and MESO-NH. The SM model, due to operational constraints, had less flexibility and was integrated on a single domain with a 14 km resolution. Each segment was integrated for 48 hours. For this model, the final meteorological file was obtained by blanking the first and last 12 h of each segment.

A similar strategy was used for the AM3 event. This episode is slightly shorter than TT1 and TT3. The 10 km grid-mesh simulations were performed with only two segments of 48 h. The starting time of the first coarse grid integration was then on 17 July 1997, 00 UTC.

In addition, as a sensitivity experiment, a short sequence of each event was simulated with a higher horizontal resolution, 2 km for MESO-NH, 3 km for MC2 and 3.5 km for BOLAM.

## Precipitation verification

One major challenge for numerical mesoscale models is to achieve the best possible forecast/ hindcast of the space and time precipitation distribution. Therefore, considerable attention is devoted to a thorough evaluation of computed precipitation versus observed precipitation. This task is not always straightforward, because measured precipitation is often not available at the desirable time and space resolution. Despite being a surface meteorological parameter, precipitation is still difficult to measure. High-resolution measurements provided by surface rain gauges are mostly available as daily precipitation sums only which do not give enough information on the time evolution. Furthermore, the density of the rain gauge network is highly variable in space and becomes coarser in mountainous areas.

For the purpose of the RAPHAEL project and of the Mesoscale Alpine Programme, a special effort was made to collect all the available precipitation observations. For the Toce-Ticino area, and for the time period considered in this study, about 120 stations provided precipitation

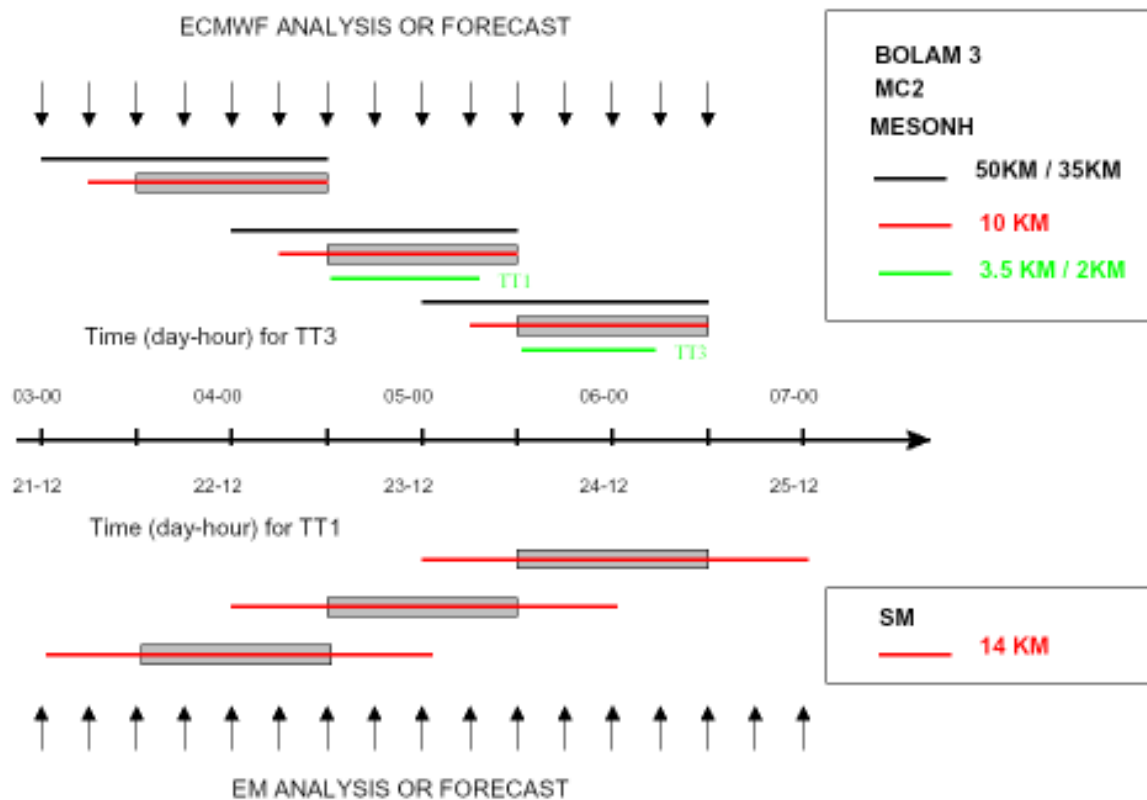


Fig. 1. Sketch of the numerical strategy used for the TT1 and TT3 episodes, with BOLAM3, MC2 and MESO-NH (top), and SM (bottom). The vertically pointing arrows show the frequency at which the initial and lateral boundary conditions are provided by the ECMWF or EM model. The horizontal black (red) lines indicate starting time and duration of each coarse (fine) grid simulation segment. The shaded rectangles show the results that were retained in the final output file. The green horizontal lines refer to the short segments that were simulated at very high resolution.

measurements. The density of stations is, on average, one per 110 km<sup>2</sup>, a relatively high value for such a wide mountain area, and is homogeneous over the domain of the analysis (Bacchi and Ranzi, 2000). These stations, operated by different services, have various standards of observation, with a recording frequency ranging from one to 24 hours. Among the 120 stations, only 35 were operated on an hourly basis. Therefore, the precipitation will be evaluated in two steps. First, computed daily precipitation will be compared with the observed daily precipitation, available from 06 UTC to 06 UTC at the 120 station locations. Then, emphasis will be put on the time evolution by comparing computed with observed hourly precipitation for the subset of the 35 stations which provided hourly measurements.

For the Ammer region, the density of measurements is much less. Only 10 stations are located in the study area but all of them provide hourly measurements. This single data set was used to verify both spatial distribution and time evolution, but the small number of verification points may weaken the conclusions.

#### SPATIAL DISTRIBUTION OF THE DAILY PRECIPITATION

##### TT1

Figure 2 shows the spatial distribution of the precipitation daily sums for a two-day sequence of the TT1 event, computed with the four numerical models run in the analysis mode. The available rain gauge observations have been reported on each panel for comparison. The geographical domain covers an area of roughly 130 km × 130 km and is centred on the Ticino-Toce catchment. It should be kept in mind that this domain represents less than 10% of the full numerical domain, and corresponds to a detail of a much wider image. Such a zooming effect will emphasise the differences between the numerical results.

On the first day, the differences are quite substantial. The precipitation pattern is quite spotty for MESO-NH and SM and more organised for BOLAM and MC2. On the second day, all models agree in displaying a wide band of precipitation, SW/NE-oriented and much more intense than on the previous day. However, the maximum value varies

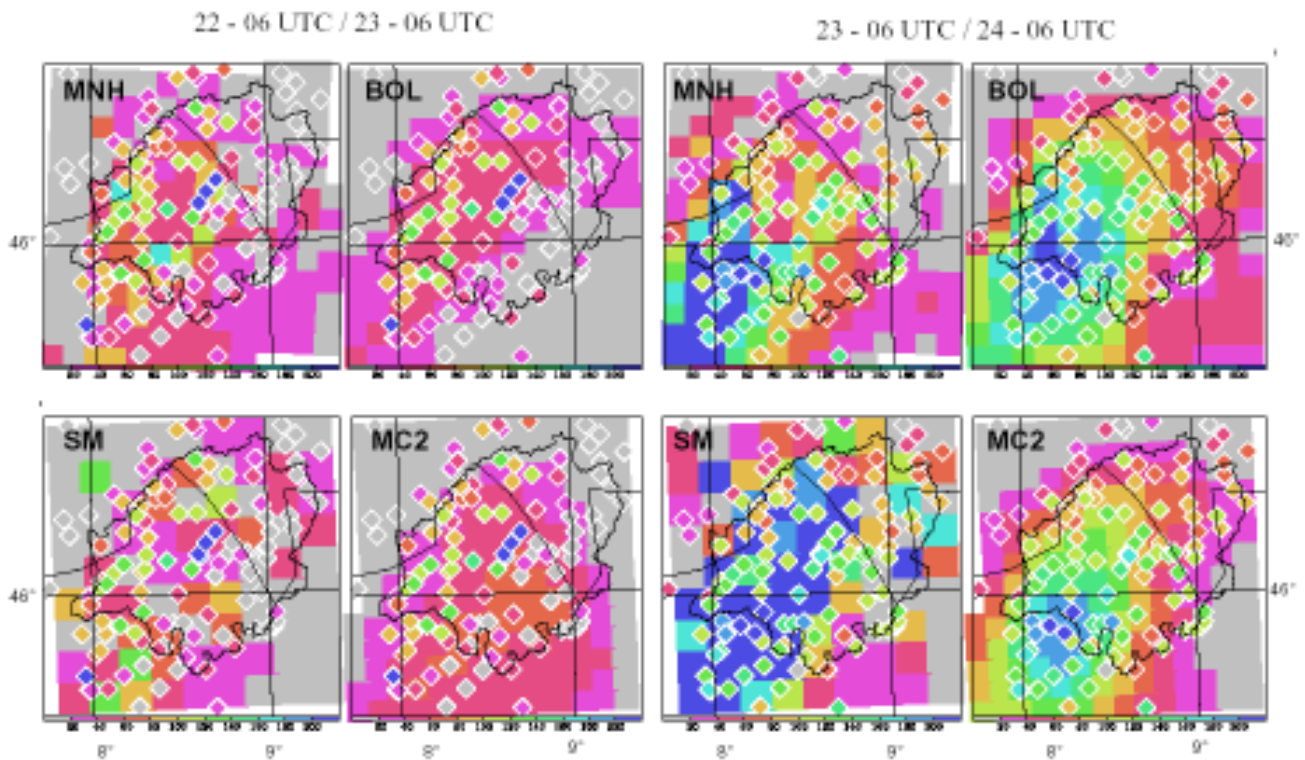


Fig. 2. 24 h accumulated precipitation (in mm) computed in the analysis mode with the four numerical models MESO-NH (MNH), BOLAM3 (BOL), SM (SM) and MC2 (MC2) for TT1 event: from 93/09/21 06 UTC to 93/09/22 06 UTC on the left and from 93/09/22 06 UTC to 93/09/23 06 UTC on the right. On each panel the corresponding rain gauge measurements (white contoured diamonds) are superimposed. The border between Italy and Switzerland, and the limit of the Ticino-Toce catchment are drawn in black.

by twice as much, from 471 mm (MESONH) to 191 mm (MC2). The mean value varies with the same amplitude, from 131 mm (SM) to 67 mm (MC2). On these plots, the comparison with the local observations provided by the rain gauges remains subjective. Nevertheless, some major discrepancies can be identified. On the first day, all the models underestimate the intensity of the precipitation core. Three stations consistently report values in excess of 200 mm whereas model results hardly reach 100 mm in this area. It is also clear that on the second day, MESO-NH does not catch the extension of the precipitation band towards the north-east correctly.

### TT3

Figure 3 is similar to Fig. 2 but presents the results of the TT3 event. The precipitation pattern is fairly similar to that obtained for TT1 and exhibits the same banded structure localised on the first Alpine foothills. The overall consistency between the different model fields seems to be stronger than for the TT1 although some noticeable differences remain. On the first day, the precipitation band computed by SM is much narrower than that obtained with the other models. Differences in intensity are specially

marked on the second day, with much higher values for BOLAM3 than for the other models, with both maximum and mean values being at least 50% larger. When compared with the measurements, all the model results locate the precipitation core correctly.

All the TT3 event computations were carried out in both analysis and forecast mode. In the forecast mode, the boundary conditions of the coarse grid run were obtained from the ECMWF forecast for MESO-NH and BOLAM, whereas for SM and MC2 they were derived from the EM forecast. It should be noted that for MC2, unlike the other models, the switch to the forecast mode implies not only a change in the boundary conditions but also in the initial state. Figure 4 presents the results obtained in forecast mode and can be compared directly with Fig. 3. The precipitation pattern is only weakly modified by the change in the boundary conditions. But all the computations carried out in the forecast mode yield more intense precipitation. On average, precipitation has been increased by 30%. This increase is entirely due to the inaccuracy of the ECMWF forecast for this event. The impact on the mesoscale simulations is significant but remains limited.

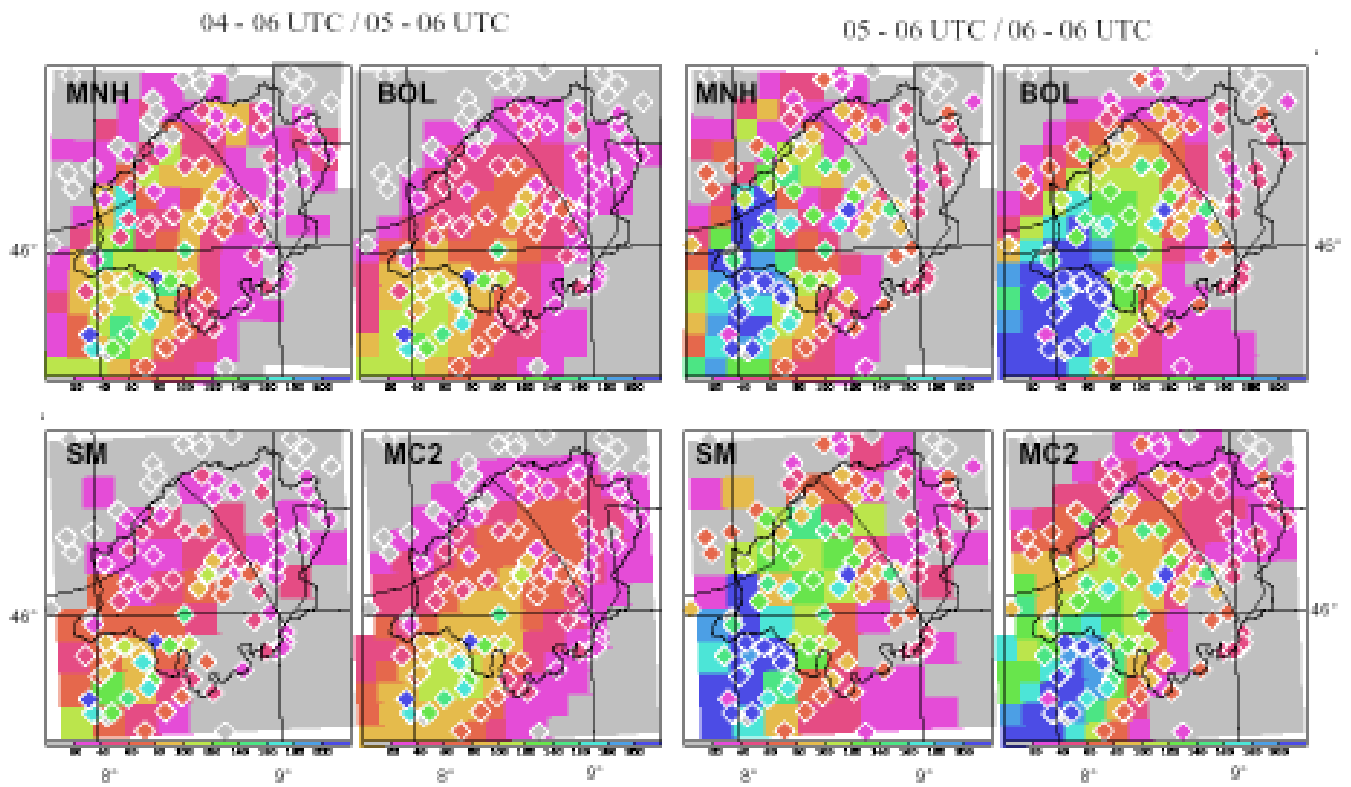


Fig. 3. Same as Fig.2 but for the TT3 event: from 94/11/04 06 UTC to 94/11/05 06 UTC on the left and from 94/11/05 06 UTC to 94/11/06 06 UTC on the right.

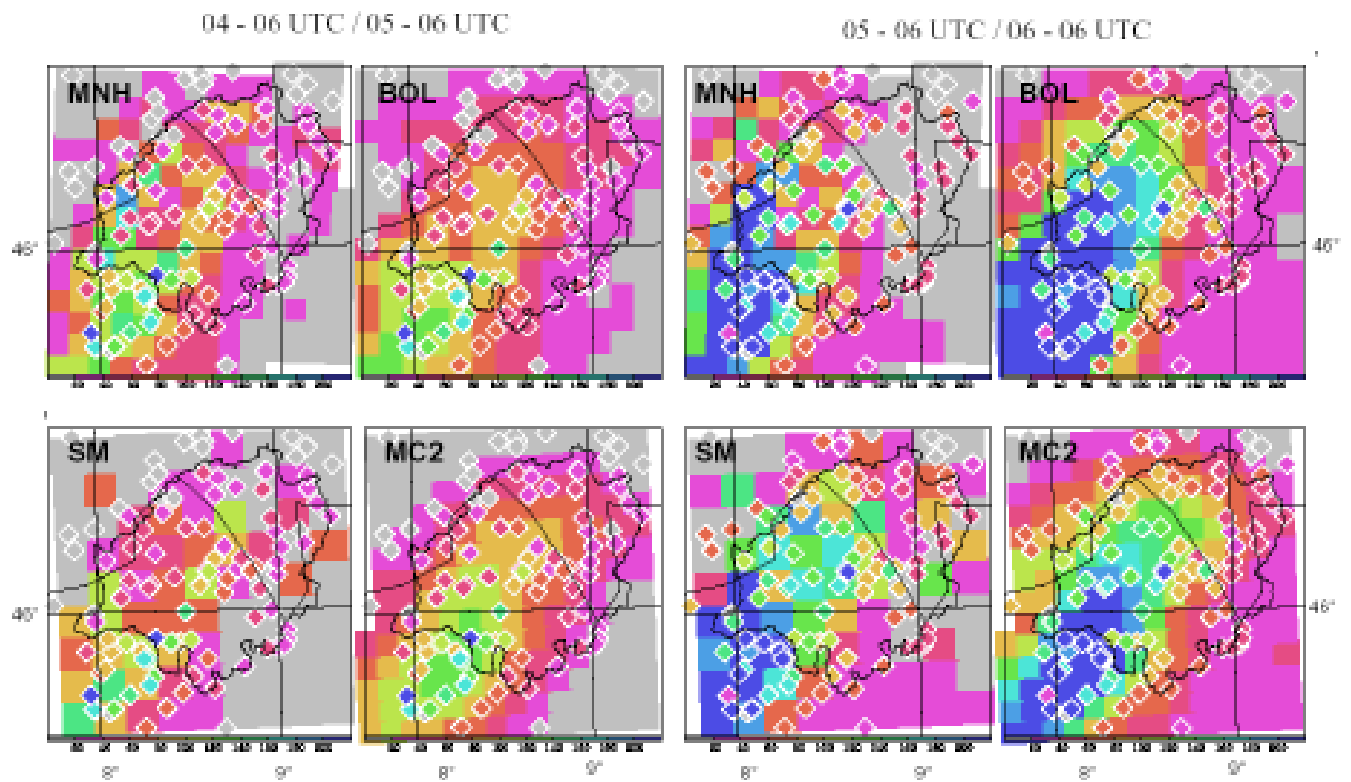


Fig. 4. Same as Fig.3 but for numerical results obtained in the forecast mode.

*AM3*

The results obtained for the AM3 event are shown in Fig. 5. The area displayed still covers a region of 130 km × 130 km but is much wider than the Ammer catchment. In terms of precipitation, this event is much less intense than TT1 or TT3 (note that the scale now ranges from 0 to 100 mm instead of 0 to 200 mm). For this episode, there is a lot of scatter between the different model results. During the two days, most of the models show a banded pattern over the mountainous area (south of the domain), aligned roughly with the German-Austrian border. But the local maxima vary a lot in intensity and location. It can also be noticed that the MC2 results depart significantly from other model results, especially on the second day.

The AM3 results are also available in the forecast mode, shown in Fig. 6. Even for this case, where the precipitation is essentially convective, the model results are only slightly modified by the change in boundary conditions. Only the MC2 fields present a substantial modification between the two modes but it is likely that this has to be ascribed more to the change in the initial state than to the boundary conditions.

STATISTICAL ANALYSIS

A more quantitative and objective assessment of the model results can be obtained by computing some statistical parameters. The model results have been interpolated at the rain gauge measurement points and are compared with the observations. This method of comparison is basic and quite crude, but any alternative would involve a spatial interpolation of the unevenly spaced rain gauge data which might introduce further bias.

Different parameters such as mean value, bias, and correlation coefficient have been computed and are reported in Tables 1 and 2. In this analysis, about 120 verification points are considered for TT1 and TT3 versus only 10 for AM3. Results obtained for this latter episode are given for completeness although they might not be very representative.

In the analysis mode, if the anomalous value of 165% (MC2, second day of AM3) is excluded, the relative bias spans from -52% (SM, first day of TT1) to +46% (SM, second day of TT1). There is no systematic evidence for any model towards a positive or negative bias and it is also interesting to note that on the first day of TT1, all the models

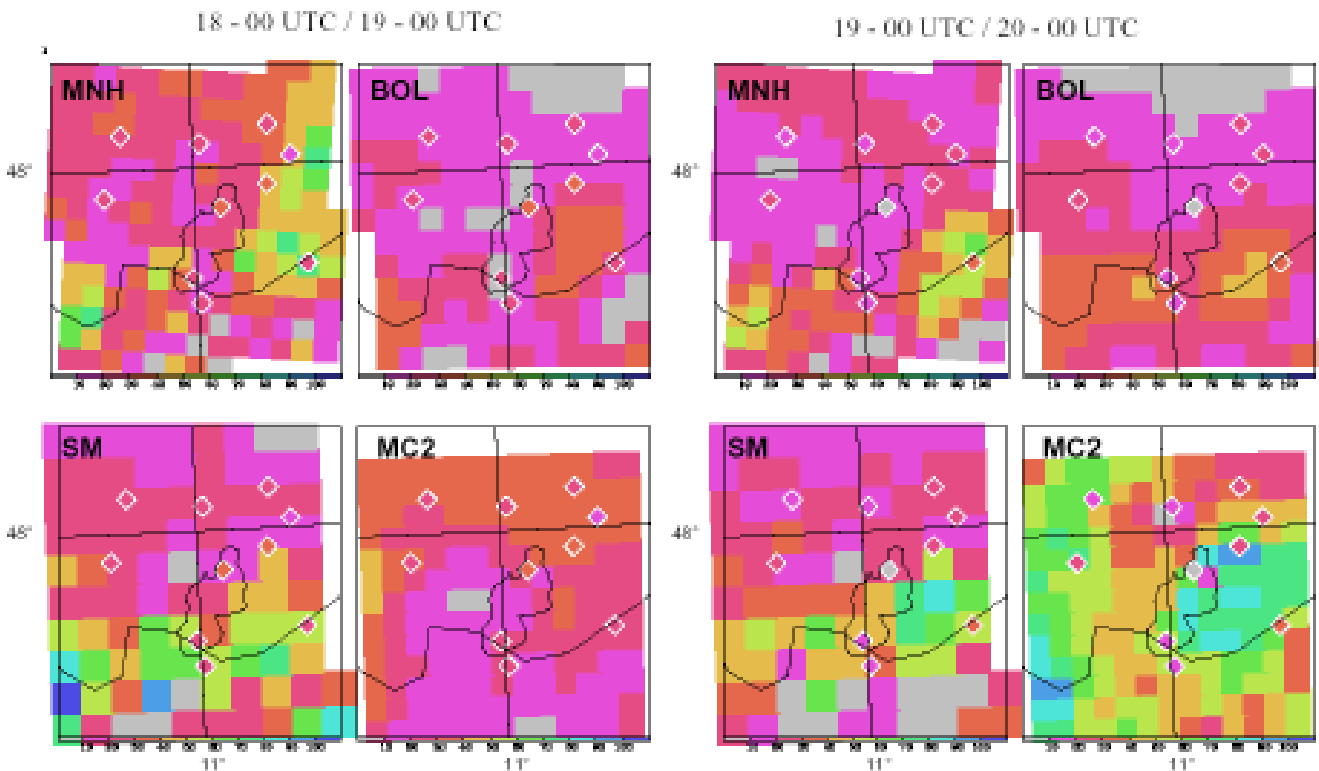


Fig. 5. Same as Fig.2 but for the AM3 event: from 97/07/18 00 UTC to 97/07/19 00 UTC on the left and from 97/07/19 00 UTC to 97/07/20 00 UTC on the right. The boarder between Germany and Austria and the limit of the Ammer watershed are drawn in black.

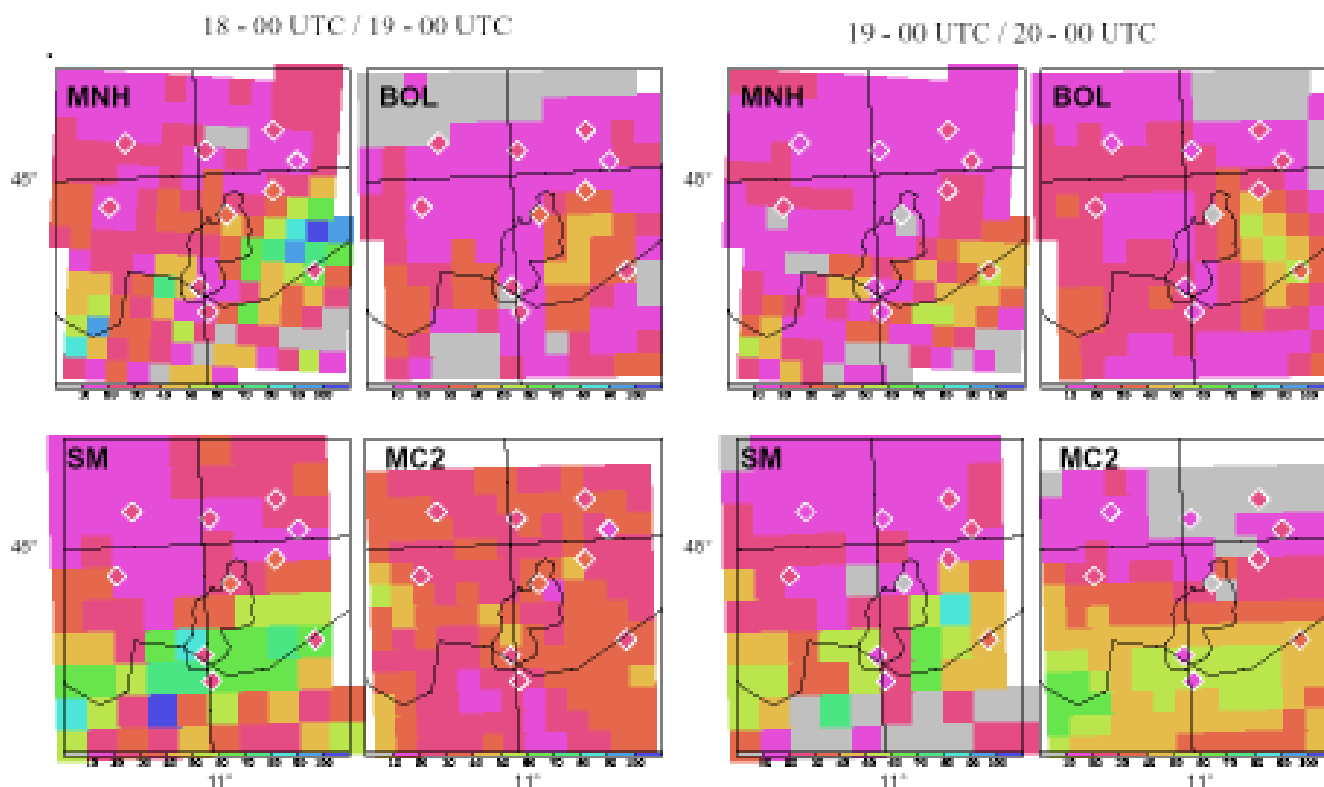


Fig. 6. Same as Fig.5 but for numerical results obtained in the forecast mode.

Table 1. Area average and relative bias of the 24h precipitation computed with the four models (MESONH, BOLAM3, SM and MC2) run in the analysis (Ana) or forecast (For) mode, for the TT1, TT3 and AM3 episodes. The 24h time period goes from 06 UTC to 06 UTC of the next day for TT1 and TT3, and from 00 UTC to 00 UTC of the next day for AM3. Model results have been interpolated at the observation point locations (120 points for TT1 and TT3, 10 points for AM3).

	TT1 22/23-09-93		TT1 23/24-09-93		TT3 04/05-11-94		TT3 05/06-11-94		AM3 19-07-97		AM3 20-07-97	
	Mean (mm)	Bias (%)	Mean (mm)	Bias (%)	Mean (mm)	Bias (%)	Mean (mm)	Bias (%)	Mean (mm)	Bias (%)	Mean (mm)	Bias (%)
Observations	54	-	103	-	56	-	97	-	24	-	20	-
MESO-NH (Ana)	42	-22	91	-12	59	+5	67	-31	30	+25	22	+10
BOLAM3 (Ana)	29	-46	104	+1	49	-13	102	+5	15	-38	2	+10
SM (Ana)	32	-52	150	+46	36	-35	84	-13	31	+29	27	+35
MC2 (Ana)	36	-41	84	-18	56	0	76	-21	28	+17	5	+16
MESO-NH (For)	-	-	-	-	62	+10	87	-10	25	+4	18	-10
BOLAM3 (For)	-	-	-	-	63	+12	138	+42	17	-29	25	+25
SM (For)	-	-	-	-	46	-18	101	+4	33	+57	25	+25
MC2 (For)	-	-	-	-	64	+14	113	+16	31	+29	27	+35

have a strong negative bias. For the TT3 episode, in the forecast mode, as shown previously in Fig. 4, the mean precipitation has increased for all the models by 20% to 50%. The biases, essentially negative in the analysis mode, are now positive.

The correlation coefficient gives additional information on the relationship between observed and computed precipitation. The different values are listed in Table 2. The correlation between observed and computed precipitation is not very strong, with a coefficient reaching at best 0.75.



Table 2. Correlation coefficients between the observed and computed 24h precipitation. Computations were carried out with the four models (MESO-NH, BOLAM3, SM, and MC2) run in the analysis (Ana) or forecast (For) mode, for the TT1, TT3 and AM3 episodes.

	<i>TT1</i> 22/23-09-93	<i>TT1</i> 23/24-09-93	<i>TT3</i> 04/05-11-94	<i>TT3</i> 05/06-11-94	<i>AM3</i> 19-07-97	<i>AM3</i> 20-07-97
MESO-NH (Ana)	.31	.50	.58	.58	0.07	0.64
BOLAM3 (Ana)	.54	.65	.75	.69	0.23	0.26
SM (Ana)	.04	.50	.70	.67	-0.30	0.19
MC2 (Ana)	.16	.68	.74	.70	0.14	0.33
MESO-NH (For)	-	-	.56	.60	0.35	0.62
BOLAM3 (For)	-	-	.73	.69	0.40	0.48
SM (For)	-	-	.66	.67	-0.24	0.09
MC2 (For)	-	-	.72	.73	0.54	0.13

The relative comparison is interesting. It confirms the particularly bad performance of the models for the first day of TT1 and for the whole of the AM3 event. Results are, in general, much better for TT3. On average, BOLAM3 seems to produce better results than the other three models, being more consistent in its performance. For TT3, when the computations are done in the forecast mode, results are modified weakly. There is a slight but systematic degradation for the first day, and for some models a slight improvement for the second day. In the case of AM3, results are too random to be discussed in detail.

When assessing precipitation computations, meteorologists often use skill scores instead of correlation. They are supposed to be more informative in the case of semi-bounded variables (like precipitation) for which correlation is not fully meaningful. Skill scores are computed from contingency tables. A 2-class contingency table considers the respective proportions of observed and simulated precipitation below or above some specified threshold. This evaluation used the Heidke skill score, defined as  $HS = (GF - RF) / (N - RF)$  where GF is the number of correct simulated values (i.e. the number of times where the simulation and verifying values fell into the same class), N is the total number of observations, and RF the number of simulated values expected to be correct just by chance (computed from the marginal totals of the contingency table). With this score, a perfect simulation would give  $HS = 1$ , whereas a simulation leading to a negative HS should be considered as worse than random.

Score computations are presented in Fig.7, only for TT1 and TT3. For the two days of each episode, and for the four models, Heidke skill scores have been computed from two-class contingency tables in which the class-limit or threshold has been varied from 20 to 200 mm with a 20 mm increment. The scores are plotted as a function of the class-limit. In general, scores are much weaker for TT1 than for TT3. For

TT1, the best scores hardly reach 0.5 whereas they exceed 0.7 for TT3. As already noticed, the results obtained for the first day of TT1 are fairly bad, specially for MC2 and SMI, for which the scores become negative as soon as the class limit reaches 40 mm. Results are slightly better on the second day and the spread between the different models is reduced. Significantly better results are obtained for TT3. On the first day, the scores drop rapidly when the threshold exceeds 100 mm but only very few stations report values higher than 100 mm compared with the second day on which half of the stations measured more than 100 mm. On this last day, the scores stay fairly high, even for a threshold of 200 mm, meaning the models are able to depict quite well the heavy precipitation core. The score analysis also confirms the weak impact of the computation mode. Results are quite similar between analysis and forecast mode. Most of the time, the difference between two different models is larger than the difference between a forecast and a hindcast of the same model.

#### HOURLY PRECIPITATION

Among the 120 stations included in the Toce Ticino analysis domain, 35 provide hourly measurements and allow a more detailed evaluation of the model results. The time evolution of the average of the gauge hourly values is presented in Fig.8 (left panel of each frame). Seventy-two hours are displayed for each episode. The observations are compared with the four model results obtained in the analysis mode for TT1, and in both analysis and forecast modes for TT3. For completeness again, the results are also presented for AM3.

In the case of TT1, the observations show two main precipitation peaks. The first one is narrow and only lasts for 12 h. It is missed totally by all the models. The second peak is wider and spans 24 h. All the models are able to

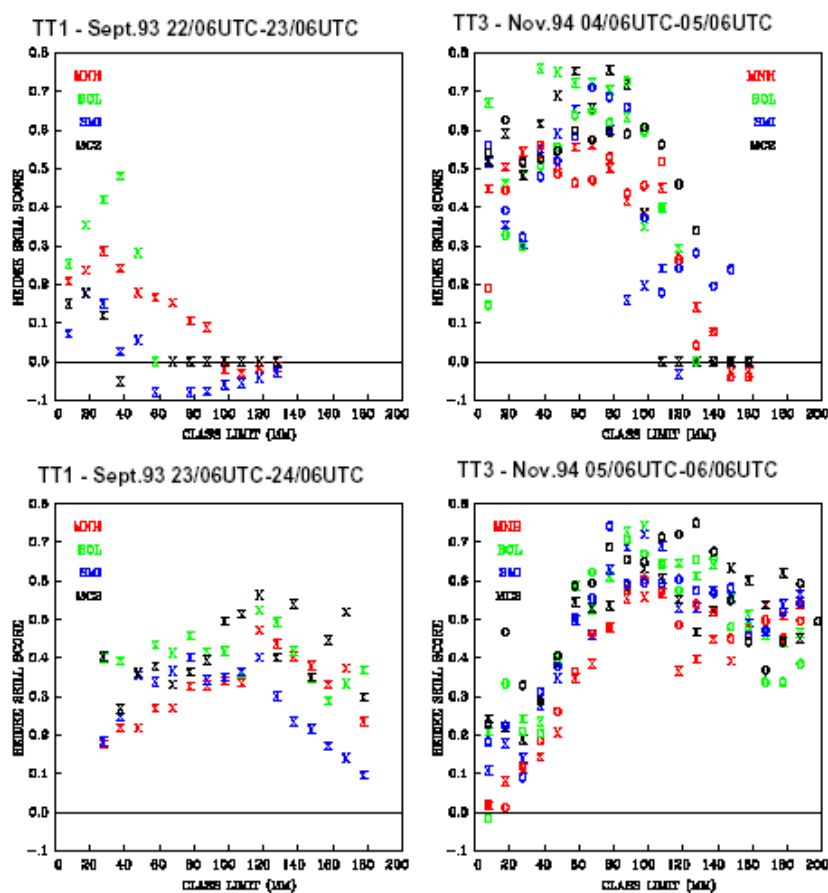


Fig. 7. Heidke skill scores as a function of the class limit, for the daily precipitation, for TT1 and TT3, and for the four numerical models. Results obtained in the analysis mode and forecast mode are distinguished by crosses and circles respectively.

catch it, but make it start a few hours earlier than observed, and tend to overestimate it by 10 to 20%. The time evolution of the correlation coefficients between observations and the different computations is shown in the right-hand panels of the same figure. The correlations, quite weak during the first 36 hours, increase during the second half of the episode. On average, BOLAM3 results are significantly better than the other results.

In the case of TT3, the observations show a continuous increase in the hourly precipitation for the first 48 h followed by a well marked peak. In the analysis mode, the model results reproduce the general trend, but tend to build somewhat artificial peaks during the first 48 hours. Globally, the correlation coefficients are stronger for this episode than for TT1. In the forecast mode, the time evolution of the precipitation is very similar to that obtained in the analysis mode but most model results now exhibit a clear indication of overestimation. Only MESO-NH performs better in the forecast mode than in the analysis mode and, except for this model, the switch to forecast mode has little impact on the correlation.

For the AM3 event, precipitation is featured by a succession of moderate (1 to 2 mm h<sup>-1</sup>) and narrow (6 h) peaks, quite characteristic of convective events. In general, the models reproduce this behaviour well, but do not catch the right timing of the individual peaks. Except for MC2, the computed intensities are not far from those observed. The correlation plot is quite illustrative. Correlation coefficients may reach acceptable values during a few hours but then drop to negative values during the next few hours: there is no time consistency. These results show clearly that the models are not able to reproduce the detailed timing of the event. For this event, the results obtained in the forecast mode are slightly better, especially for the MC2 model. But again, the agreement or disagreement with observations appears to be random.

### Sensitivity to the horizontal resolution

As a sensitivity experiment, some additional runs were performed at very high-resolution with MESO-NH and BOLAM<sup>1</sup> with respective grid-mesh of 2 and 3.5 km. Owing

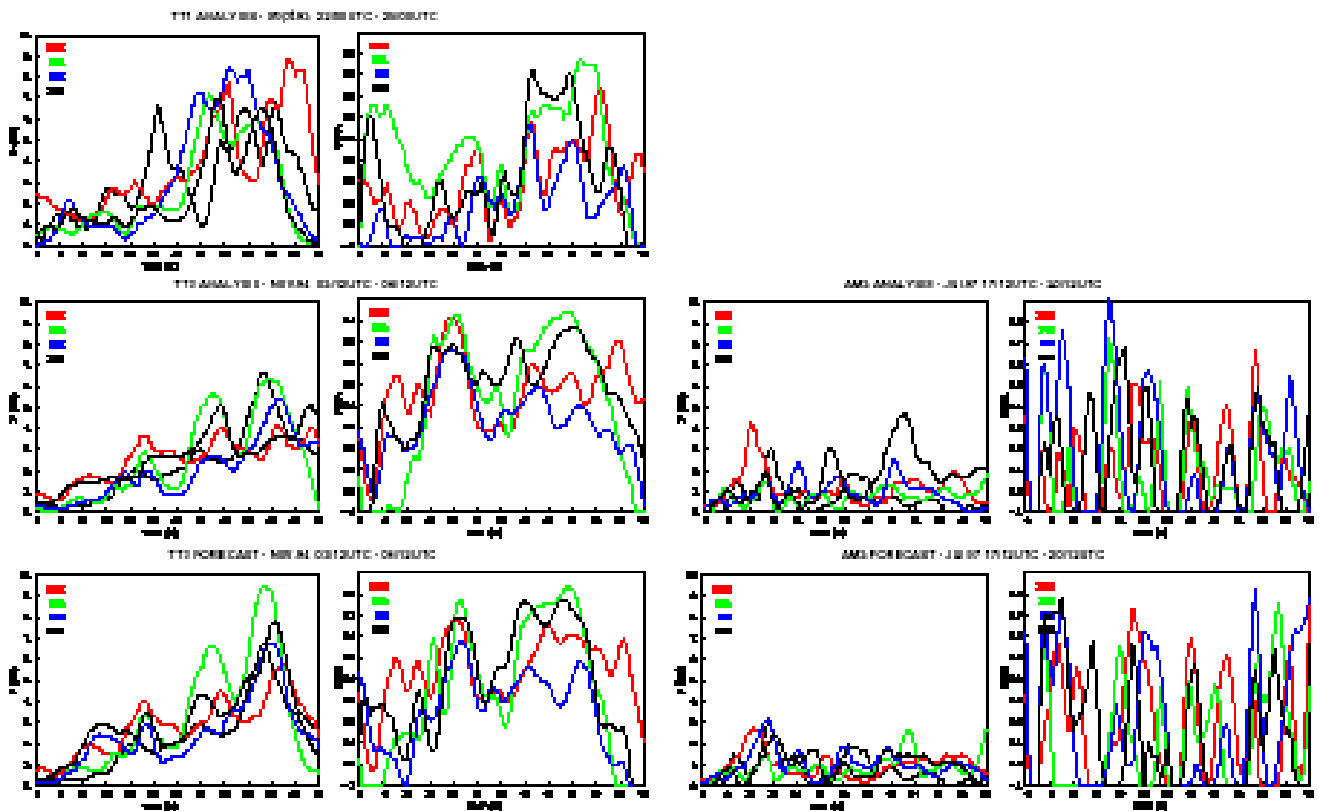


Fig. 8. Time evolution of the hourly area average precipitation ( $P$ ) and of the correlation coefficient ( $CORR$ ) for the TT1, TT3, and AM3 episodes and for the four numerical models. The thick black curve on each left panel corresponds to the observations.

to their large computing cost, these simulations were carried out only for a short time segment of 18 h for each episode. This segment was centred on the main precipitation peak, from 93/09/23 00 UTC to 93/09/23 18 UTC for TT1, and from 94/11/05 12 UTC to 94/11/06 06 UTC for TT3. The spatial distribution of the 18 h accumulated precipitation for each episode is shown in Fig. 9 which compares the results for the four models obtained with a 10 km grid mesh with those obtained at very high-resolution. When the resolution is increased, for both models and both episodes, the maximum values are increased by roughly 25% and the area average by 10%. As the resolution gets higher, the orographic forcing is represented better, slopes are steeper, and some extra lifting and associated condensation processes occur. With the increase in resolution, the precipitation patterns are not modified dramatically. Finer scale structures have developed, giving a more realistic appearance to the fields. They now present a level of heterogeneity comparable

to radar-derived precipitation fields. Some radar data were available for these two episodes but, unfortunately, they were of poor quality and could not be used for quantitative comparison.

The time evolution of the average (at gauge locations) of the hourly precipitation obtained with high resolution simulations for these 18 h segments is shown in Fig. 10. It can be compared with the observations and with the results obtained with the 10 km resolution simulations that are recalled on the figure. For both models and both episodes, precipitation has increased but the results, when compared to the observations, are not significantly better. For TT1, the precipitation peak, which was not reproduced in the 10 km simulations, is still absent. For TT3, the three to five hours delay at which the peak appears has not been reduced when the resolution was increased. The time evolution of the correlation coefficients between numerical results and rain gauge measurements is shown on the same figure. For BOLAM3, the high-resolution results are, in general, not as good as the 10 km resolution results. For MESO-NH, there is some improvement with the high resolution, but not in a systematic way. For the cases presented here, the benefit of high resolution simulations is not obvious, but

<sup>1</sup>MC2 also provided some very high-resolution simulations. However, they are not shown in this study because they were obtained with a different numerical protocol and cannot be compared with the 10 km resolution simulations presented in the previous section.

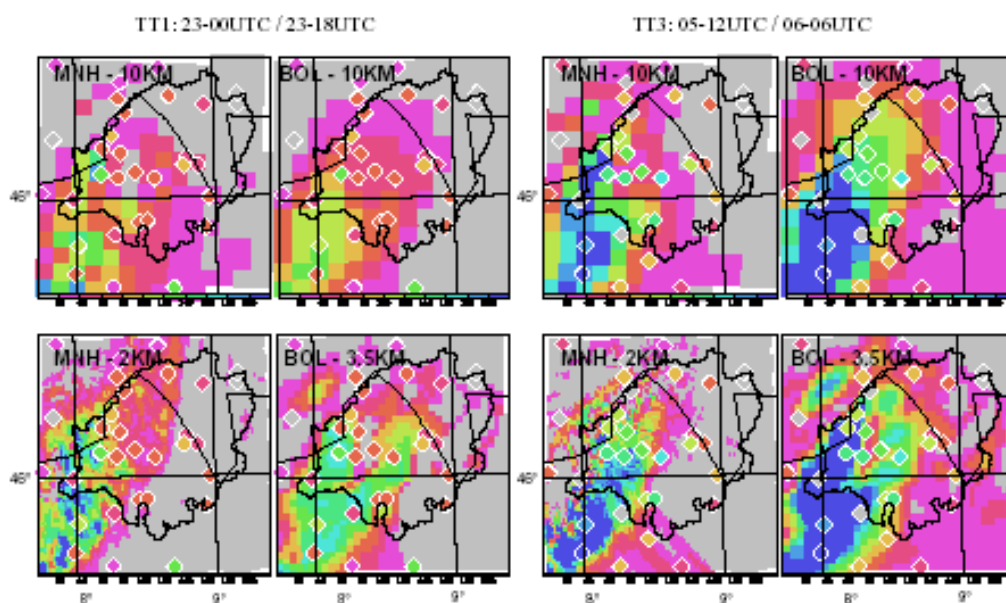


Fig. 9. 18h accumulated precipitation (in mm) computed with BOLAM3 run with a 10 and 3.5 km resolution and MESO-NH run with a 10 and 2 km resolution for TT1 event from 93/09/23 00 UTC to 93/09/23 18 UTC on the left, and for TT3 event from 94/11/05 12 UTC to 94/11/06 06 UTC on the right. On each panel are superimposed the corresponding rain gauge measurements (white contoured diamonds).

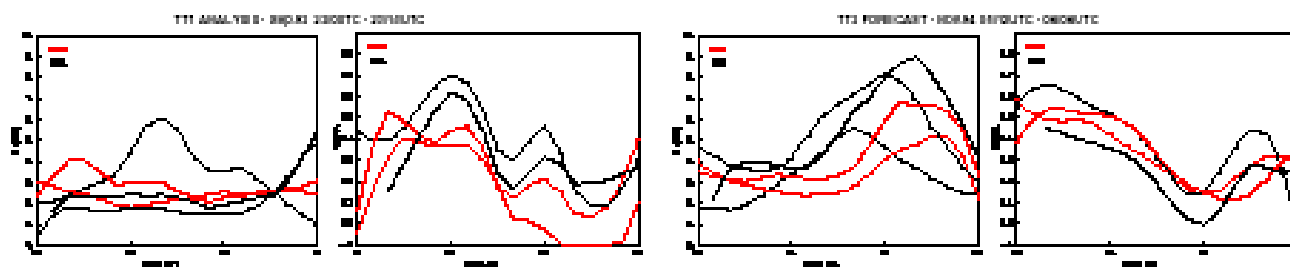


Fig. 10. Time evolution of the area average hourly precipitation ( $P$ ), and of the correlation coefficient ( $CORR$ ) for TT1 and TT3, during the time corresponding to the high-resolution simulations. The solid curves refer to the 10 km resolution simulations and the dashed curves to the high resolution simulations. The thick black curve on each left panel corresponds to the observations.

no definite conclusion can be obtained from such limited results. More recent experiments, performed on the heavy precipitation events documented during the MAP field experiment, provided very convincing results in favour of the high resolution modelling (e.g. Benoit *et al.*, 2002; Pradier *et al.*, 2002; Asencio *et al.*, 2004; Richard *et al.*, 2004).

## Conclusion

In the framework of the RAPHAEL project, a series of past heavy precipitation events has been simulated with four different meteorological mesoscale models, run at a 10 to 14 km resolution. Three of these events, which have been simulated by all the four models with a common protocol,

were selected to intercompare the different numerical results and to assess their quality. The first two events (TT1, Brig flood in 1993, and TT3, Piedmont flood in 1994) affected the Toce-Ticino region strongly whereas the last episode (AM3, 1997) was associated with a flood in the Ammer catchment. For each episode, computed precipitation was compared with all the available observations within an area of  $130 \times 130 \text{ km}^2$  centred either on the Toce-Ticino region or on the Ammer region. The density of the observations is fairly high in the Toce Ticino region but much weaker in the Ammer region.

Different statistical parameters have been computed, such as mean, bias, correlation coefficients and skill scores. In general, there is a fairly good consistency between the different models. For the 24 h precipitation, the relative mean

bias stays in the range  $-50\%/+50\%$ , indicating that all the models were able to reproduce the occurrence of the event, and to reproduce it with reasonable intensity. There is no systematic trend for a model to over- or under-estimate precipitation. The quality of the numerical results appears to be much more case-dependent than model-dependent. For instance, all models have difficulty in reproducing the observed precipitation in the first stage of the TT1 episode, and all models perform quite well for the TT3 episode. This finding is not surprising. At the mesoscale, numerical results depend strongly on the initial state. Errors in the analysis used to initialise the model are known to have much more impact on the results than the details of the model itself.

In the context of an operational flood forecast system based upon computed precipitation, it was crucial to show that the numerical results were not too degraded if the boundary conditions of the mesoscale model were derived from a forecast and not from an analysis. In that respect, the results are quite encouraging. The sensitivity to the change in the boundary conditions was found to be rather weak.

In a more prospective way, some attempt was made to perform very high resolution simulations, with a 2 to 3.5 km resolution. Owing to their high computing cost, they were carried out for short periods of time only. From the limited results, it is not clear whether or not they could lead to a major improvement. However, recent studies for the MAP episodes show much more promising results.

## Acknowledgements

This study was supported in part by the European Commission Programme *Environment and Climate* under contract ENV4-CT97-0552.

## References

- Asencio, N., Stein, J. and Chong, M., 2004. Analysis and simulation of local and regional conditions for the heavy rainfall over the Lago Maggiore. *Q. J. Roy. Meteorol. Soc.* in press.
- Bacchi, B. and Ranzi, R., 2000. RAPHAEL final report, 154pp. Available from the University of Brescia, Italy.
- Benoit, R., Desgagné, M., Pellerin, P., Desjardins, S. and Chartier, Y., 1997. The Canadian MC2: a semi-Lagrangian, semi-implicit wide-and atmospheric model suited for fine-scale process studies and simulations. *Mon. Wea. Rev.*, **125**, 2382–2415.
- Benoit, R., Schär, C., Binder, P., Chamberland, S., Davies, H.C., Desgagné, M., Girard, C., Keil, C., Kouwen, N., Luthi, D., Maric, D., Muller, E., Pellerin, P., Schmidli, J., Schubiger, F., Schwierz, C., Spenger, M., Walser, A., Willemsse, S., Yu, W. and Zala, E., 2002. The realtime ultrafinescale forecast support during the special observing period of the MAP. *Bull. Amer. Meteor. Soc.*, **83**, 85–109.
- Bougeault, P., Binder, P., Buzzi, A., Dirks, R., Houze, R., Kuetner, J. Smith, R.B., Steinacker, R. and Volkert, H., 2001. The MAP Special Observing Period. *Bull. Am. Meteor. Soc.* **82**, 433–462.
- Buzzi, A. and Malguzzi, P., 1997. The BOLAM III model : recent improvements and results. *MAP Newsletter*, **7**, 98–99.
- Buzzi, A., Fantini, M., Malguzzi, P. and Nerozzi, P., 1994. Validation of a Limited Area Model in cases of Mediterranean cyclogenesis: Surface fields and precipitation scores. *Meteor. Atmos. Phys.*, **53**, 137–153.
- Buzzi, A., Tartaglione, N. and Malguzzi, P., 1998. Numerical simulations of the 1994 Piedmont flood: Role of the orography and moist processes. *Amer. Meteor. Soc.*, **126**, 2369–2383.
- Emanuel, K.A. and Raymond, D.J., 1993. *The representation of Cumulus Convection in Numerical Models*, Meteor. Monogr. No. 46, American Meteorological Society. 146pp.
- Kasahara, A., Mizzi, A.P. and Donner, L.J., 1991. Impact of cumulus initialization on the spinup of precipitation forecasts in the tropics. *Mon. Weather Rev.*, **120**, 1360–1380.
- Kaufmann, P., Schubiger, F. and Binder, P., 2002. Precipitation forecasting by a mesoscale NWP model: Eight years of experience. *Hydrol. Earth Syst. Sci.*, **7**, 812–832.
- Lafore, J.P., Stein, J., Bougeault, P., Ducrocq, V., Duron, J., Fischer, C., Hérelil, P., Mascart, P., Masson, V., Pinty, J.P., Redelsperger, J.L., Richard, E. and Vila-guerau de Arellano, J., 1998. The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations. *Annales Geophysicae* **16**, 90–109.
- Molinari, J.M. and Dudek, M., 1992. Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Weather Rev.*, **120**, 326–344.
- Pradier, S., Chong, M. and Roux, F., 2002. Radar observations and numerical modeling of a precipitating line during MAP IOP 5. *Mon. Weather Rev.*, **130**, 2533–2553.
- Richard, E., Cosma, S., Tabary, P., Pinty, J.P. and Hagen, M., 2004. High-resolution numerical simulation of the convective system observed in the Lago Maggiore area on 17 September 1999 (MAP IOP2a). *Q. J. Roy. Meteorol. Soc.* in press.
- Stein, J., Richard, E., Lafore, J.P., Pinty, J.P., Asencio, N. and Cosma, S., 2000. Meso-NH simulations with grid-nesting and ice phase parameterization. *Meteor. Atmos. Phys.*, **72**, 203–221.