

The 8 and 9 September 2002 flash flood event in France: a model intercomparison

S. Anquetin¹, E. Yates¹, V. Ducrocq², S. Samouillan¹, K. Chancibault¹, S. Davolio⁵, C. Accadia⁶, M. Casaioli⁶, S. Mariani⁶, G. Ficca⁷, B. Gozzini³, F. Pasi³, M. Pasqui³, A. Garcia⁴, M. Martorell⁴, R. Romero⁴, and P. Chessa⁷

¹Laboratoire d'étude des Transferts en Hydrologie et Environnement, Grenoble, France

²Centre National de la Recherche Météorologique, Météo-France, Toulouse, France

³Laboratorio Meteorologia e Modellistica Ambientale, Sesto Fiorentine, Italy

⁴Department of Physics Meteorology group, Palma de Mallorca, Spain

⁵ISAC, CNR, Bologna, Italy

⁶Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici, Roma, Italy

⁷SAR, Sardegna, Italy

Abstract. Within the framework of the European Interreg IIIB Medocc program, the HYDROPTIMET project aims at the optimization of the hydrometeorological forecasting tools in the context of intense precipitation within complex topography. Therefore, some meteorological forecast models and hydrological models were tested on four Mediterranean flash-flood events. One of them occurred in France where the South-eastern ridge of the French "Massif Central", the Gard region, experienced a devastating flood on 8 and 9 September 2002. 24 people were killed during this event and the economic damage was estimated at 1.2 billion euros.

To build the next generation of the hydrometeorological forecasting chain that will be able to capture such localized and fast events and the resulting discharges, the forecasted rain fields might be improved to be relevant for hydrological purposes.

In such context, this paper presents the results of the evaluation methodology proposed by Yates et al. (2005) that highlights the relevant hydrological scales of a simulated rain field. Simulated rain fields of 7 meteorological model runs concerning with the French event are therefore evaluated for different accumulation times. The dynamics of these models are either based on non-hydrostatic or hydrostatic equation systems. Moreover, these models were run under different configurations (resolution, initial conditions). The classical score analysis and the areal evaluation of the simulated rain fields are then performed in order to put forward the main simulation characteristics that improve the quantitative precipitation forecast.

The conclusions draw some recommendations on the value of the quantitative precipitation forecasts and way to use it for quantitative discharge forecasts within mountainous areas.

1 Introduction

The south-eastern ridge of the French "Massif Central" (see Fig. 1a for location) is frequently subject to flash floods. They mainly occur during the Autumn, and they affect a wide range of watershed scales (basins of typically 50 up to 500 km² embedded in the five main basins shown in Fig. 1b). The flash floods in this region are often due to intense and quasi-stationary mesoscale convective systems. The hilly topography of the region governs the distribution of the rain and the runoff production. Given the very short response time of flash floods, their forecast must rely on quantitative precipitation forecast.

One of the greatest challenges for numerical meteorological models is to improve significantly quantitative precipitation forecasting (QPF). As noted by Fritsch et al. (1998), QPF skill by the current operational models remains relatively low. Nevertheless, research advances open new perspectives and show that research non-hydrostatic models are now able to provide more realistic rainfall distributions at meso-scale for selected case studies (Ducrocq et al., 2002; Roebber and Eise, 2001; Stein et al., 2000) or during field experiments (e.g. the Mesoscale Alpine Program; Bougeault et al., 2001).

Within the framework of the European Interreg IIIB Medocc program, the HYDROPTIMET project dealt with the optimization of the hydrometeorological forecasting tools. Therefore, several meteorological forecast models and hydrological models were tested on four Mediterranean flash-flood events (14–16 November and 24–26 November 2002 in Italy; 8 and 9 September 2002 in France; 9 and 10 June 2000 in Spain). The choice of these events was motivated by the amount of the cumulated precipitation and the resulting damages in the region. More than 300 mm in 24 h were recorded for these four cases leading to a very fast an-

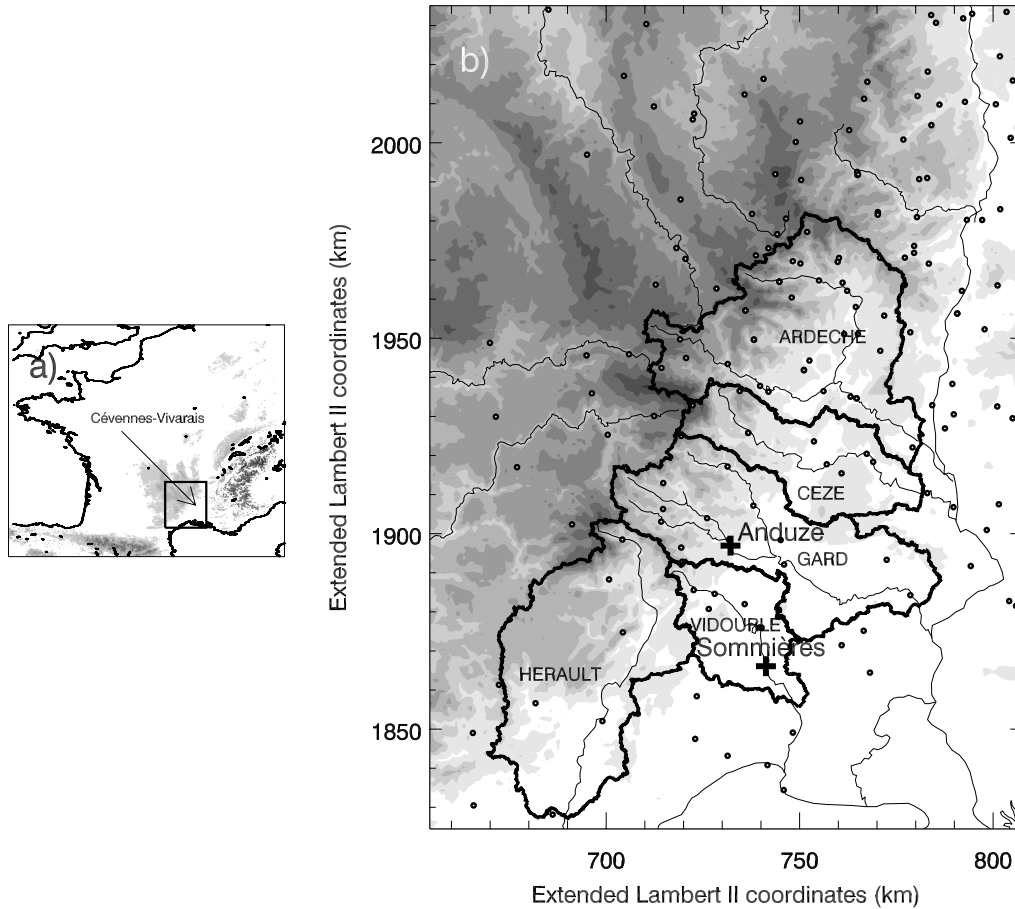


Fig. 1. (a) The black rectangle indicates the studied area located on the south-east edge of the Massif Central in France. (b) Relief of the studied area and main hydrographic network derived from a digital terrain model; the thick lines delineate the five main watersheds (Ardèche River (2500 km²); Cèze River (1054 km²); Gard River (1913 km²); Vidourle River (621 km²); Hérault River (2500 km²)); the thin lines indicate the main rivers; the 180 rain gauge stations used for the validation are indicated by the circles; Anduze and Sommières cities are shown (black crosses).

swer of the basins. Damages and more seriously deaths make these events important to test the hydrometeorological forecasting chain in order to improve the alerting system. A large variety of meteorological forecasting models was tested due to the origin of the different teams involved in this project. Both operational and research meteorological models were used in order to highlight the feasibility of the models to produce relevant simulated rain fields for hydrological purposes. The main idea was to stand out from the results of the simulations the important characteristics (dynamics, physical package, grid resolution, initial conditions) of the models that improve the QPF.

Within this general context, this paper presents the results of the simulations of the French test case that occurred in September 2002. Based on a recent work (Yates et al., 2005), the evaluation of 7 model runs performed under the HYDROPTIMET project aims at giving clues about which space scales can be considered as relevant for using simulated rainfall fields in hydrology. The next section is dedicated to the presentation of the meteorological situation and the resulting

observed rainfall. After the introduction to the 7 meteorological simulations evaluated in this paper (Sect. 3), the discussion is organized through 2 main evaluation approaches in Sect. 4. The classical point validation is based on the usual statistics and scores from contingency tables established for given thresholds, whereas the areal validation is based on the work of Yates et al. (2005) and proposes to scan space scales in order to identify the relevant space scale for hydrological purposes. The discussion (Sect. 5) proposes to highlight the different contributions of the different models to the QPF in order to draw some recommendations about the use of QPF from current and next generation Numerical Weather Prediction systems.

2 Presentation of the 8 and 9 September 2002 flash-flood event

On 8 and 9 September 2002, a heavy precipitation event was responsible of the most important flooding ever recorded in

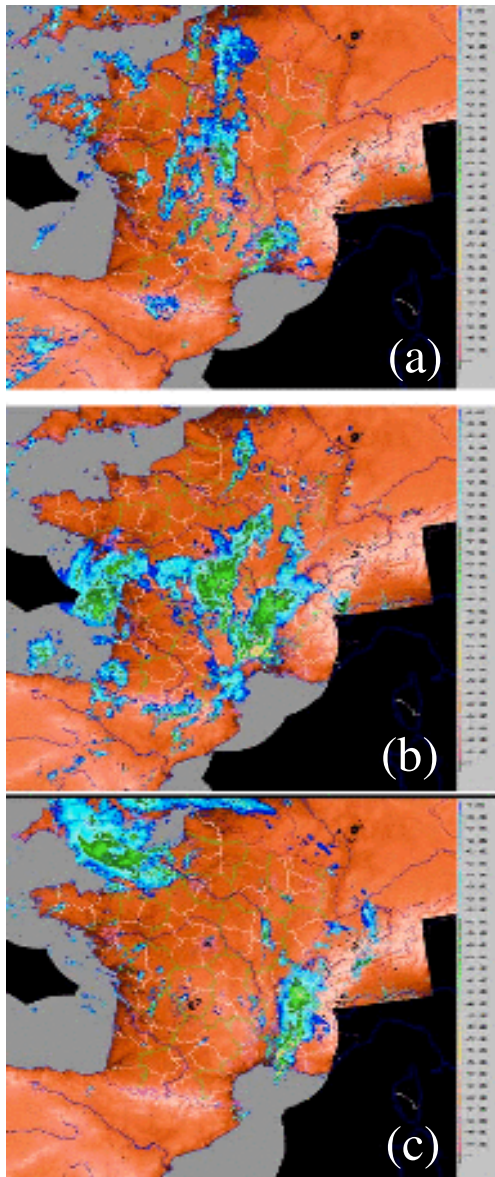


Fig. 2. Radar reflectivity composite from the network ARAMIS (METEO-FRANCE) at: (a) at 12:00 UTC, 8 September 2002, (b) 00:00 UTC, 9 September 2002, (c) 12:00 UTC, 9 September 2002.

the Cévennes-Vivarais region, in the southern part of France (Fig. 1). A major Mesoscale Convective System (MCS) affected the Gard area (Fig. 1): 24 people were killed during this event and economic damage was estimated at 1.2 billion euros (Huet et al., 2003). Delrieu et al. (2005) provided a detailed meteorological and hydrological description of this case. Here, a brief overview of the rainfall event is given below.

Figure 2 presents radar reflectivity composites during the event. First convective cells formed early in the morning of 8 September 2002 over the Mediterranean Sea (Fig. 2a). The convection progressed northward to form inland a MCS

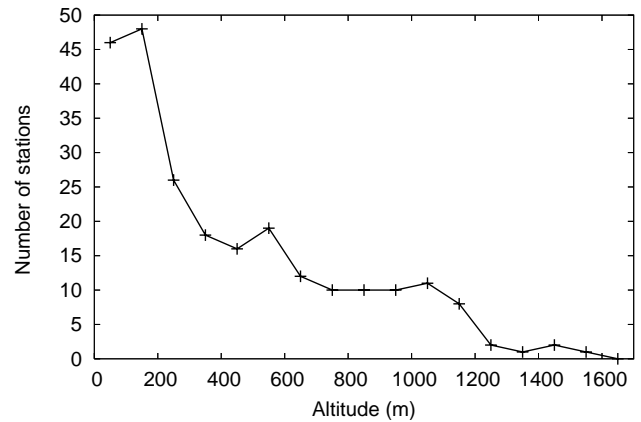


Fig. 3. Distribution of the rain gauges with altitude.

over the Gard region later (Fig. 2b). The quasi-stationary MCS stayed over the same region until approximately the next morning (Fig. 2c), and then progressed eastward with the surface cold front.

The rainfall observation network is made of 180 hourly rain gauges (circles in Fig. 1). The rain gauge density over the whole region is around 0.007 km^{-2} and as regards to the distribution with altitude presented in Fig. 3, this network gives a relatively reliable estimation of the rainfall distribution. Most part of the high total amounts of surface rainfall has been induced by the quasi-stationary MCS. The 24-hour accumulated precipitation reached almost 700 mm near Anduze (see location Fig. 1), with 500 mm recorded in less than 9 hours. The rain gauge observations have been spatially interpolated using the geostatistical method of kriging. This method has already been fitted and used for the Cévennes-Vivarais region (e.g. Lebel et al., 1987; Yates et al., 2005). The kriging method aims first at equalling to zero the mean estimator error and second at minimizing the mean estimation error (σ^2).

The kriged rain accumulation over the 24 h (12:00 UTC on 8 September 2002 to 12:00 UTC on 9 September 2002) is presented in Fig. 4. The confidence zones defined by the estimation variance threshold of 40% show the portions of the fields that are taken into account in the following analyses. As shown in Fig. 4, the main kriged rain field remains within the confidence zone and can therefore be used as the reference for the validation.

3 The meteorological simulations and analysis techniques

3.1 Presentation of the atmospheric simulations

Within the framework of the HYDROPTIMET project, two types of meteorological simulations were evaluated with respect to their relevance for hydrological applications. Two

Table 1. List of the meteorological simulations. DU00 stands for Ducrocq et al. (2000); KF90 makes reference to Kain and Fritsch (1990) whereas KF93 to Kain and Fritsch (1993); KU74 stands for Kuo (1974) and SS98 for Spencer and Stensrud (1998). -AN and -FC refer respectively to the analysis and the forecasts of the associated meteorological service (ARPEGE, ECMWF and NCEP).

Model Partner	Grid size	Convective scheme	Initial conditions	Boundary conditions	Father model	Convective scheme	Father initial cond.	Father bound. condi.
BOLAM _S Opera., Hydro. SAR	0.05°	KF90	0.18°BOLAM _S nesting	0.18°BOLAM _S every 1 h	0.18°BOLAM _S Hydro.	KF90	ECMWF-AN 07/09/02 at 12:00 UTC	ECMWF-FC every 6 h
BOLAM _A Opera., Hydro. APAT	0.1°	KU74	0.3°BOLAM _A nesting	0.3°BOLAM _A every 3 h	0.3°BOLAM _A Hydro.	KU74	ECMWF-AN 07/09/02 at 12:00 UTC	ECMWF-FC every 6 h
BOLAM _I Resea., Hydro. ISAC-CNR	0.06°	KF90 improved with SS98	0.2°BOLAM _I nesting	0.2°BOLAM _I every 1.5 h	0.2°BOLAM _I Hydro.	KF90 improved with SS98	ECMWF-AN 08/09/02 at 00:00 UTC	ECMWF-FC every 6 h
MOLOCH Resea., Non-Hydro ISAC-CNR	2.2 km	no	0.06°BOLAM _I nesting	0.06°BOLAM _I every 1 h	0.06°BOLAM _I Hydro.	KF90 improved with SS98	0.2°BOLAM _I 08/09/02 at 00:00 UTC	0.2°BOLAM _I every 1.5 h
RAMS Resea., Non-Hydro. LaMMA	2 km	no	8 km RAMS nesting	8 km RAMS every 1 h	8 km RAMS Non-Hydro.	KF93	ECMWF-AN 08/09/02 at 00:00 UTC	ECMWF-FC every 6 h
MM5 Resea., Non-Hydro. UIB	6 km	no	18 km MM5 two-way nesting	18 km MM5 two-way nesting	18–54 km MM5 Non-Hydro.	KF90	NCEP-AN 07/09/02 at 12:00 UTC	NCEP-AN every 12 h
MesoNH Resea., Non-Hydro. CNRM and CNRS	2.4 km	no	9.5 km MesoNH two-way nesting	9.5 km MesoNH two-way nesting	9.5 km MesoNH Non-Hydro.	KF90	ARPEGE-AN 08/09/02 at 12:00 UTC + DU00	ARPEGE-FC every 3 h

model runs (BOLAM_S and BOLAM_A) were performed using the BOLAM model in the operational configuration as defined by the operational services. The five other simulations (BOLAM_I, MOLOCH, RAMS, MM5 and MesoNH) used a model in its research mode and were, therefore, based on the best configuration of the model that it can have. The interest of such simulations is to identify the relevant components that might be incorporated in the next generation of operational models. They also provide more realistic simulated fields that can be used to analyze the main physical mechanisms responsible of such extreme events (this point will not be discussed in this paper).

Table 1 and the two next subsections give more details on the numerical configurations of the simulations. For more details on the description of the models (BOLAM_{I,A,S}; MOLOCH; MM5; RAMS), one can refer to Mariani et al. (2005). A detailed description of the MesoNH run configuration for precipitation events is given by Ducrocq et al. (2002). Most of the simulations started at 00:00 UTC on 8 September 2002 and lasted 36 h. BOLAM_A simulation started at 00:00 UTC on 8 September 2002 and lasted 48 h. MesoNH simulation started at 12:00 UTC on September 2002 and lasted 24 h. Only the 24 common simulated hours between 12:00 UTC, 8 September 2002 to 12:00 UTC, 9 September 2002 will be analysed. This time window corresponds to the main period of the rainy event.

3.1.1 The operational model runs

The Bologna Limited Area Model (BOLAM) from Servizio Agrometeorologico Regionale (SAR) (later referred to BOLAM_S) was run operationally by the Sardinia region. BOLAM (Davolio et al., 2004) is a primitive equation, sigma coordinate, hydrostatic model. The operational procedure is based on two runs of the model. The first run is obtained through a direct nesting of the 0.18° BOLAM_S into the ECMWF global model. The second run is obtained nesting the 0.05° BOLAM_S into the 0.18° BOLAM_S using the aforementioned output as initial and boundary conditions. The 0.18° BOLAM_S initial condition is supplied from the ECMWF analysis at 12:00 UTC on 7 September 2002. The initial condition for the 0.05° BOLAM_S is the 12 h BOLAM_S father forecast.

The QUADRICS Bologna Limited Area Model (QBOLAM, later referred to BOLAM_A) was run by the Agenzia per la Protezione dell’Ambiente e per i Servizi Tecnici (APAT). QBOLAM is a parallel version of BOLAM. It runs operationally at APAT in Rome (Italy) as part of POSEIDON sea wave and tidal forecasting system (Speranza et al., 2004). The operational simulation is based on the run of two one-way nested grids. The outer grid has a horizontal spacing of 0.3° and is initialized by the ECMWF analysis. The boundary conditions are given by the ECMWF forecasts initialized with the same analysis. The inner domain has a horizontal

spacing of 0.1° . The operational, 60-h run starts daily with the 12:00 UTC analysis; however, the inner domain run starts 12 h later (spinup time) so that the higher resolution run has a 48-h forecast range, starting from 00:00 UTC of the day after the analysis.

3.1.2 The research model runs

The BOlogna Limited Area Model (later referred to BOLAM_I) and MOdello LOcale (MOLOCH) are used for scientific purposes at the Istituto di Scienze dell'Atmosfera e del Clima-Consiglio Nazionale delle Ricerche (ISAC-CNR) at Bologna (Italy). The BOLAM_I simulation is based on the run of two one-way nested grids. The outer grid has a horizontal spacing of 0.2° while the inner grid resolution is 0.06° . The initial condition is supplied from the ECMWF analysis at 00:00 UTC, 8 September 2002. Boundary conditions are supplied from ECMWF forecasts from 12:00 UTC, 7 September 2002. The boundary conditions are imposed using a relaxation scheme (Lehman, 1993). MOLOCH is a non-hydrostatic high-resolution model that integrates the fully compressible set of equations. MOLOCH is nested into the higher resolution BOLAM_I simulation (i.e. 0.06°), with lateral boundary values updated every hour.

The Regional Atmospheric Modeling System (RAMS) is used at the Laboratorio di Meteorologia e Modellistica Ambientale (LaMMA) at Florence (Italy). RAMS is a non-hydrostatic fully compressible model. The simulation configuration is based on the run of two one-way nested grids at 8 and 2 km spacing, respectively. The initialization is based on the ECMWF analysis whereas the six-hourly ECMWF forecasts provide boundary conditions on the coarser grid, using both a lateral nudging, on a strong 6h time scale, and a inner-domain light nudging on a 12 h time scale. The outer grid forecast provides one-way forcing for the 2-km grid spacing inner domain. Boundary conditions are provided hourly with a lateral nudging.

The Fifth Generation Mesoscale Model (MM5) is run for research purposes at the Departament de Fisica, Universitat de les Illes Balears (UIB). MM5 is a non-hydrostatic model. The simulation configuration is based on three two-way nested grids at 54, 18 and 6 km, respectively. NCEP analysis has been used every 12 h (initial and boundary conditions). The 54-km MM5 outputs are every 6 h, whilst for the other two domains outputs are every 3 h.

The MesoNH Atmospheric Simulated System (Lafore et al., 1998) is a joint effort of Centre National de Recherches Météorologiques at Météo-France and Laboratoire d'Aérodynamique à Toulouse (France). MesoNH is designed as a research tool for small and meso-scale atmospheric processes. It is an anelastic non-hydrostatic model. For this study, the model was run over two two-way nested grids with horizontal spacing of 9.5 and 2.4 km, respectively. The initial state is provided by the mesoscale initialization of Ducrocq et al. (2000). This fine scale initialization is composed of two steps; the first consists of mesoscale surface observation analysis. An optimal interpolation scheme analyses mesonet

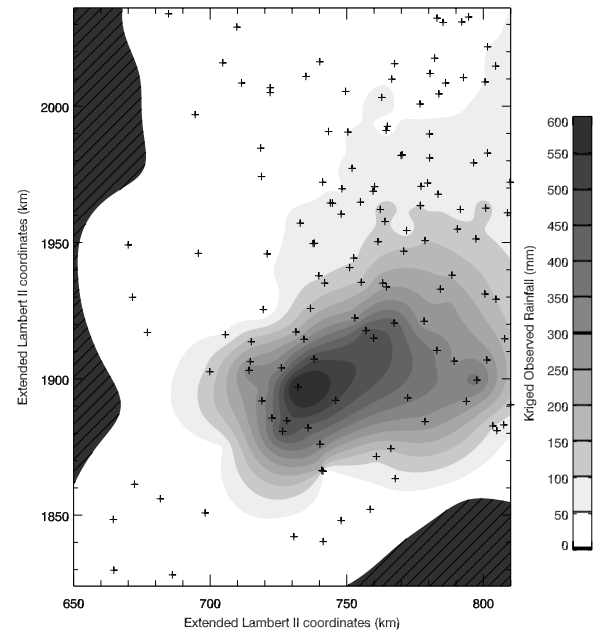


Fig. 4. 24-hour rain accumulated from 12:00 UTC, 8 September 2002 to 12:00 UTC, 9 September 2002: kriging of the 180 rain gauges values (+ symbols). The shaded surfaces correspond to points where less than 40% of variance is explained by the interpolation.

surface observations which are in average spaced by about 30 km. We analyse the mesonet surface observations valid for 8 September 2002 at 12:00 UTC. The background is provided by the 12:00 UTC ARPEGE analysis for 8 September 2002. The second step consists of an adjustment of humidity and hydrometeors based on a pre-analysis of the cloudy and rainy areas associated with the developing storms of 8 September 2002 at 12:00 UTC. This pre-analysis is derived from the radar reflectivities and from the IR METEOSAT brightness temperature. The benefit of this mesoscale initialization for intense rainfall event forecast has been previously demonstrated (Ducrocq et al., 2002).

3.2 Evaluation strategy

The analysed time window starts at 12:00 UTC, 8 September 2002 and ends at 12:00 UTC, 9 September 2002. The evaluation strategy is based on the analysis of the hourly simulated rain fields (i.e. 24 terms) obtained by the models and their comparison with the observed rain fields. The comparisons have been also performed on three accumulation time steps: 6-hour (i.e. 4 terms), 12-hour (i.e. 2 terms) and 24-hour (i.e. 1 term). Due to the difference of grid size between the models, all the simulated rain fields have been interpolated on a $1 \times 1 \text{ km}^2$. Therefore, for each term, the number of available precipitation values is approximately equal to $n_x \times n_y = 213$. The validation is made in 2 steps: (i)

Table 2. Schematic 2×2 contingency table for the definition of scores.

	Forecast < threshold	Forecast ≥ threshold
Reference < threshold	<i>a</i>	<i>b</i>
Reference ≥ threshold	<i>c</i>	<i>d</i>

the *point validation* that compares the kriged observed values at each cell location to the simulated values at the same location and (ii) the *areal validation* based on the comparison between integrated measured values on a surface to the integrated simulated values on the same surface. This latter validation methods appears more relevant for hydrological purposes (Yates et al., 2005). Statistics are calculated on $n_{term} \times n_x \times n_y$ values where n_{term} is the number of terms corresponding to each time step. The validation that uses the comparison between the rain values measured at the locations of the raingauges to the simulated values at the nearest grid element gives the same tendency that the proposed point validation method, and will not be presented here.

3.2.1 Point validation methodology

For point validation, the commonly used comparison criteria rely on both continuous and threshold statistics.

By continuous statistics, we mean criteria that use distribution functions (Figs. 5a and 6a) of the reference (observed) and the simulated values. This comparison is insensitive to the spatial distribution of the rain structure of the compared fields. The additional use of scatter-diagrams (Figs. 5b and 6b) allows to take into account the structure. The comparison can be reduced to bulk scores, such as the bias (BIAS) or the square of the correlation coefficient (R^2) defined as follow :

$$\text{BIAS} = \frac{1}{N} \sum_{i=1}^N (P_i - R_i) = m_P - m_R$$

$$R^2 = \left(\frac{\sum_{i=1}^N (R_i - m_R)(P_i - m_P)}{N \cdot \sigma_R \cdot \sigma_P} \right)^2$$

where R_i and P_i are respectively the i^{th} reference and predicted values among N , m_R and m_P the corresponding means and σ_R and σ_P the standard deviations of the reference and predicted rainfall values. R^2 is sensitive to the structure of the rain field, not to its amplitude.

By threshold statistics, we mean binary criteria usually calculated on contingency tables established for a given threshold (Table 2). In this study, three thresholds have been chosen based on the 70, 80 and 90 quantiles (not shown) of the distribution function of the observed rain for the 4 accumulation time steps. These thresholds are presented in Table 3.

We will discuss on the evolution of three classical scores computed from these tables and based on the three different

thresholds (Table 3). Namely, the forecast accuracy (ACC), the frequency bias (FBIAS) and the false alarm ratio (FAR) (Mason, 1989) can be computed according to the following expressions:

$$\text{FBIAS} = (b + d)/(c + d)$$

$$\text{FAR} = b/(b + d)$$

$$\text{ACC} = (a + d)/(a + b + c + d)$$

These scores can be interpreted as follow:

- FBIAS compares the occurrence of the model and reference precipitation above the given threshold. It does not take into account localization errors. A perfect forecast has FBIAS=1.
- FAR gives the rate of ill predicted locations among the locations where the forecast is over the threshold; a perfect forecast has FAR=0.
- ACC gives the rate of all well-predicted locations with respect to the threshold; a perfect forecast has ACC=1. The FAR and ACC scores need generally to be examined together.

3.2.2 Areal validation methodology

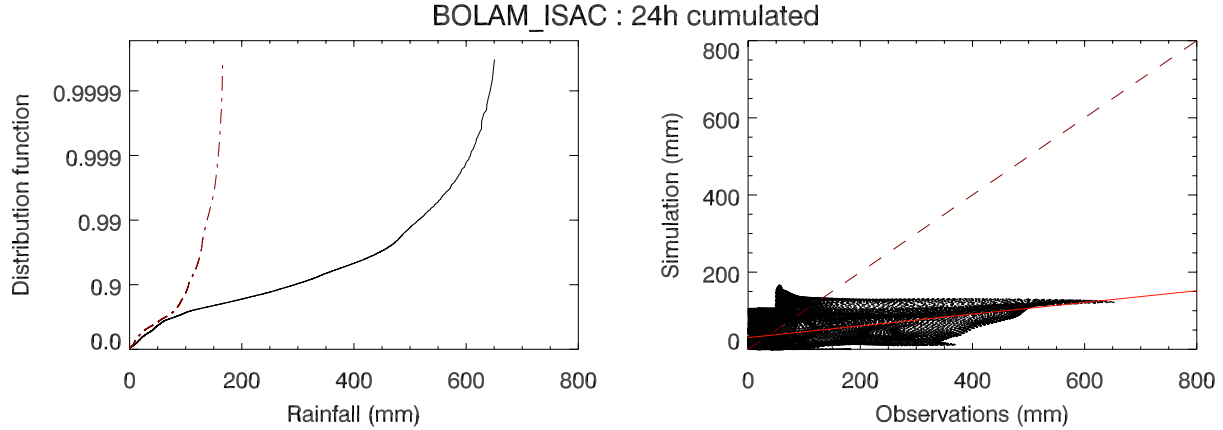
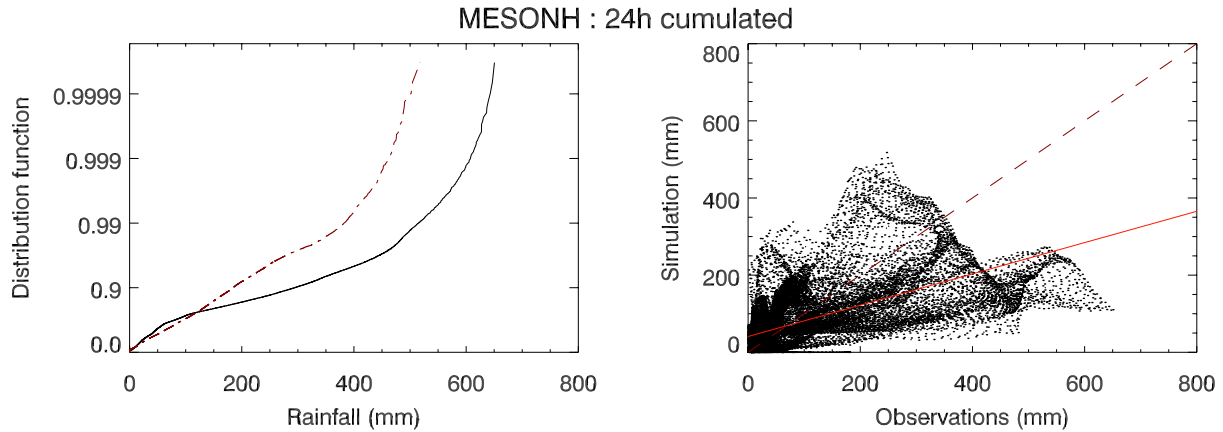
The multi-scale approach, as proposed by Yates et al. (2005), scans space scales with the correlation coefficient R^2 (calculated as defined previously). The domain of study, D_T , is divided according to the natural drainage network. The delineation of areas constituted by two hill slopes draining water towards a river reach yields a set of surfaces that are sound for hydrologic applications. The example map presented in Fig. 7 was obtained from a digital terrain model with a resolution of 75 m over the $D_T=32\,000\text{ km}^2$ of the Cévennes-Vivaraïs region. The flow pathway algorithm based on the TOPMODEL concepts (Beven and Kirkby, 1979) was used to automatically draw the stream channels and the associated hill slopes. As this hydrologic subdivision is submitted to the natural constraints of topography, the different cells at a given scale do not have exactly the same surface. Therefore, they are distributed into classes to allow statistical processing. The repartition into classes has to be consistent with the statistical analysis; each class must contain enough surfaces; the range of surface size within each class must be as narrow as possible. These two criteria can be respected by making classes with regular width with respect to the logarithm of the surface: each cell of the class C_j has its surface between $10^{\frac{j-1}{10}}$ and $10^{\frac{j}{10}}$.

4 A first comparison between the observed and the simulated rain fields

Figures 8 and 9 show the comparison between the 24 h accumulated rain fields ending at 12:00 UTC on 9 September

Table 3. Rain thresholds (mm) used to compute the scores for the 4 accumulation time steps.

Accumulation time step	1-hour			6-hour			12-hour			24-hour		
Thresholds (mm)	4	7	15	24	38	73	43	73	154	88	161	294

**Fig. 5.** Example of results obtained with an operational model run (i.e. *BOLAM_S*). Left: Distribution functions of the 24-hour accumulated observed precipitation (solid line) and of the simulated precipitations (dashed-dotted line). Right: Simulated versus reference 24-hour accumulated precipitation. The solid line represents the best fit on the simulation. The dashed line is the 1×1 ideal curve.**Fig. 6.** Example of results obtained with a research model run (i.e. *MesoNH*). Left: Distribution functions of the 24-hour accumulated observed precipitation (solid line) and of the simulated precipitations (dashed-dotted line). Right: Simulated versus reference 24-hour accumulated precipitation. The solid line represents the best fit on the simulation. The dashed line is the 1×1 ideal curve.

2002 from observations and simulated by operational and research model runs, respectively.

All the simulations performed with different configurations of BOLAM present a large underestimation of the precipitation amount (Fig. 8 and Fig. 9), as may be expected from a hydrostatic model in case of simulation of an intense convective event. The BOLAM_I simulation (the research configuration of the hydrostatic model) seems to give

the highest amount of the 24 h accumulated field. Nevertheless, for the 3 BOLAM runs, the rainfall is mainly produced over the Massif Central relief and not over the plain where the MCS was observed.

Concerning the non-hydrostatic models, Fig. 9, MOLOCH, MM5 and MesoNH simulations produce intense rainfall patterns displaced to the north along the foothills of the Massif Central, while RAMS produce a

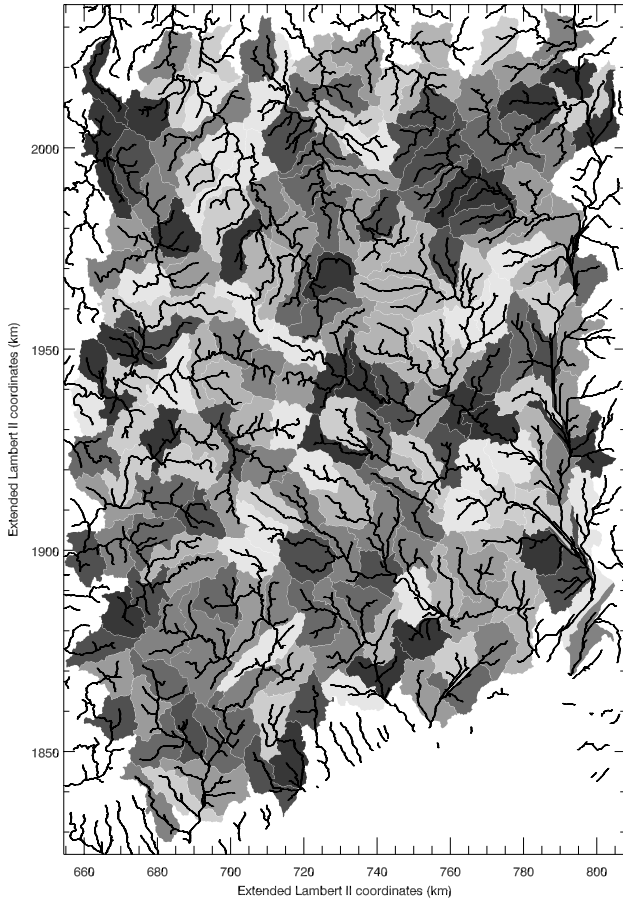


Fig. 7. Illustration of the validation surfaces used in this study. As an example, the elementary surfaces are here 100 times smaller than the total domain. The colors are only for reading purposes.

weak maximum of rainfall close to the coast. However the rainfall intensities obtained with non-hydrostatic simulations are closer to the observations than those obtained with the hydrostatic simulations.

Most of the models place the maximum of precipitation over the Massif Central foothills and crests. Ducrocq et al. (2003) have shown that the location of the maximum of precipitation over the Gard plains is related to the facing of the cold pool generated by the rain evaporation with the moist southeasterly low level jet that prevailed during the event. Therefore, the difficulty to forecast the good location of the heavy precipitation lies probably in erroneous representations of the cold pool and/or moist low-level jets. Further investigations are currently carried out in order to understand this deficiency of the models: physical parameterizations (microphysical schemes, turbulence, air-sea surface fluxes ...), initial atmospheric and surface conditions (sea surface temperature, soil moisture contents ...) and orography representation (Ducrocq et al., 2004, 2005; Davolio et al., 2005).

This type of comparison could be extended to different

accumulation time steps (not presented here) and would demonstrate the capacity of the models to reproduce the dynamics of the rain fields. This is the objective which will be pursued in the following sections.

5 Validation and evaluation of the meteorological models

5.1 Point validation

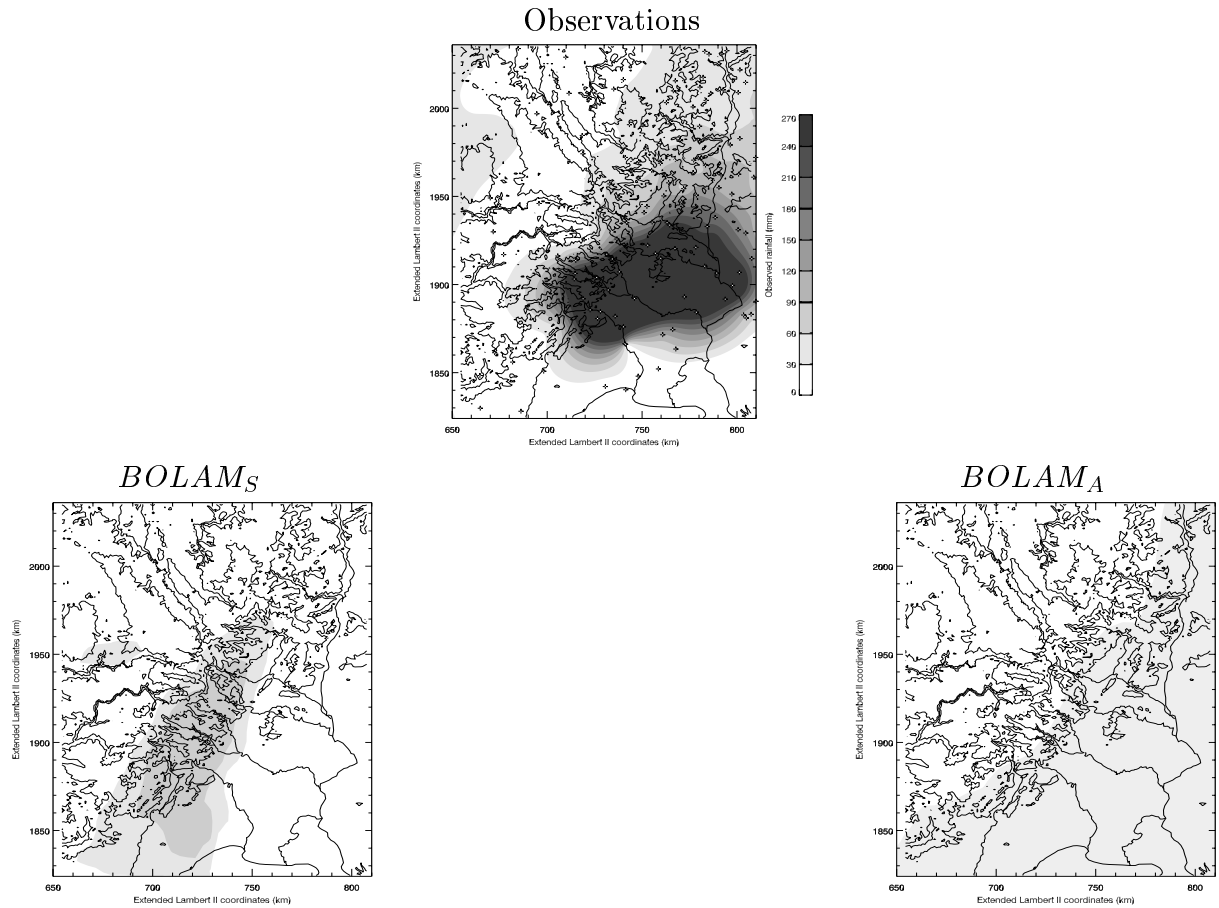
Figures 5 and 6 present an example of a distribution function and of a scatter-diagram for an operational simulation (BOLAM_S) and a research simulation (MesoNH), respectively. These two examples are representative of the results obtained with hydrostatic and non-hydrostatic simulations performed in this study. Non-hydrostatic models provide higher amount of rainfall leading to a lower bias on the distribution function plot. Moreover, the simulation of the location of the rain pattern is more accurate leading to a regressive line closer to the bisecting line in the scatter-diagram. The results of the analysis of the dispersion and the scatter diagrams drawn for the seven simulations as presented in Figs. 5 and 6, are summarized in Table 4. The BIAS and R^2 bulk scores are given for the 4 accumulation time steps and for the seven model runs, except for MM5 for which only the 3 h accumulated simulated rain fields were provided, therefore only the 6 h, 12 h, and 24 h accumulation time steps are available for this model. Table 4 suggests the following comments.

- All the models underestimate the rainfall at any time steps. Usually long time accumulated rainfall gives better correlation coefficient. Nevertheless, we observe very low correlation coefficients R^2 for the MM5 model associated to a large delocalization of the rain pattern.
- The BIAS obtained with non-hydrostatic models are usually lower than the one obtained with hydrostatic models, especially for the larger rainfalls. The conclusion on the R^2 coefficient is not straightforward. The results obtained with MesoNH are usually the best in terms of R^2 coefficient among the four non-hydrostatic models. The location of the simulated rain fields is improved by the mesoscale initial conditions used for the MESO-NH simulation as already mentioned by Chancibault et al. (2005)¹.
- The comparison between the two BOLAM simulations that have approximately the same numerical configuration (BOLAM_S and BOLAM_A) shows the impact of the convection scheme on the structure of the rain pattern. The Kain and Fritsch (1990) convection scheme seems to better reproduce the convection for such precipitation event within complex topography. The comparison between BOLAM_S and BOLAM_I runs tends to

¹Chancibault, K., Anquetin, S., Ducrocq, V., and Saulnier, G.-M.: Hydrological evaluation of high resolution precipitation forecasts of the Gard flash-flood (8–9 September 2002), Q. J. Roy. Meteor. Soc., submitted, 2005.

Table 4. Synthetic results of the analysis of the dispersion and scatter diagrams for the 4 accumulation time steps.

	1-hour mean observed: 4.7 mm		6-hour mean observed: 24 mm		12-hour mean observed: 49 mm		24-hour mean observed: 93 mm	
	Bias (mm)	R^2	Bias (mm)	R^2	Bias (mm)	R^2	Bias (mm)	R^2
BOLAM _S	−3.8	.095	−19	.04	−37	.04	−71	.06
BOLAM _A	−3.6	.001	−20.	.006	−38.	.008	−69.	.16
BOLAM _I	−2.7	.053	−13	.11	−26	.18	−49	.22
MOLOCH	−2.5	.015	−13	.06	−37	.01	−71	.02
RAMS	−2.5	.017	−11	.03	−11	.03	−42	.08
MM5			−3.79	1.10^{-6}	−8.42	1.10^{-4}	−14.78	6.10^{-6}
MesoNH	−1.6	.01	−8.7	.02	−17	.01	−31	.10

**Fig. 8.** 24 h accumulated rainfields ending at 12:00 UTC, 9 September 2002. A comparison between the observed rainfall and simulated rainfields obtained with the operational model runs.

support this idea even if, in this case, the initialization is slightly modified. The improvement of the convective scheme proposed by Spencer and Stensrud (1998) slightly increases the quality of the simulation for all

the accumulation time steps, both in terms of amount and structure of the simulated rain fields. Therefore, the BOLAM_I simulation seems to be the best among the BOLAM simulations. Moreover, BOLAM_I simulation

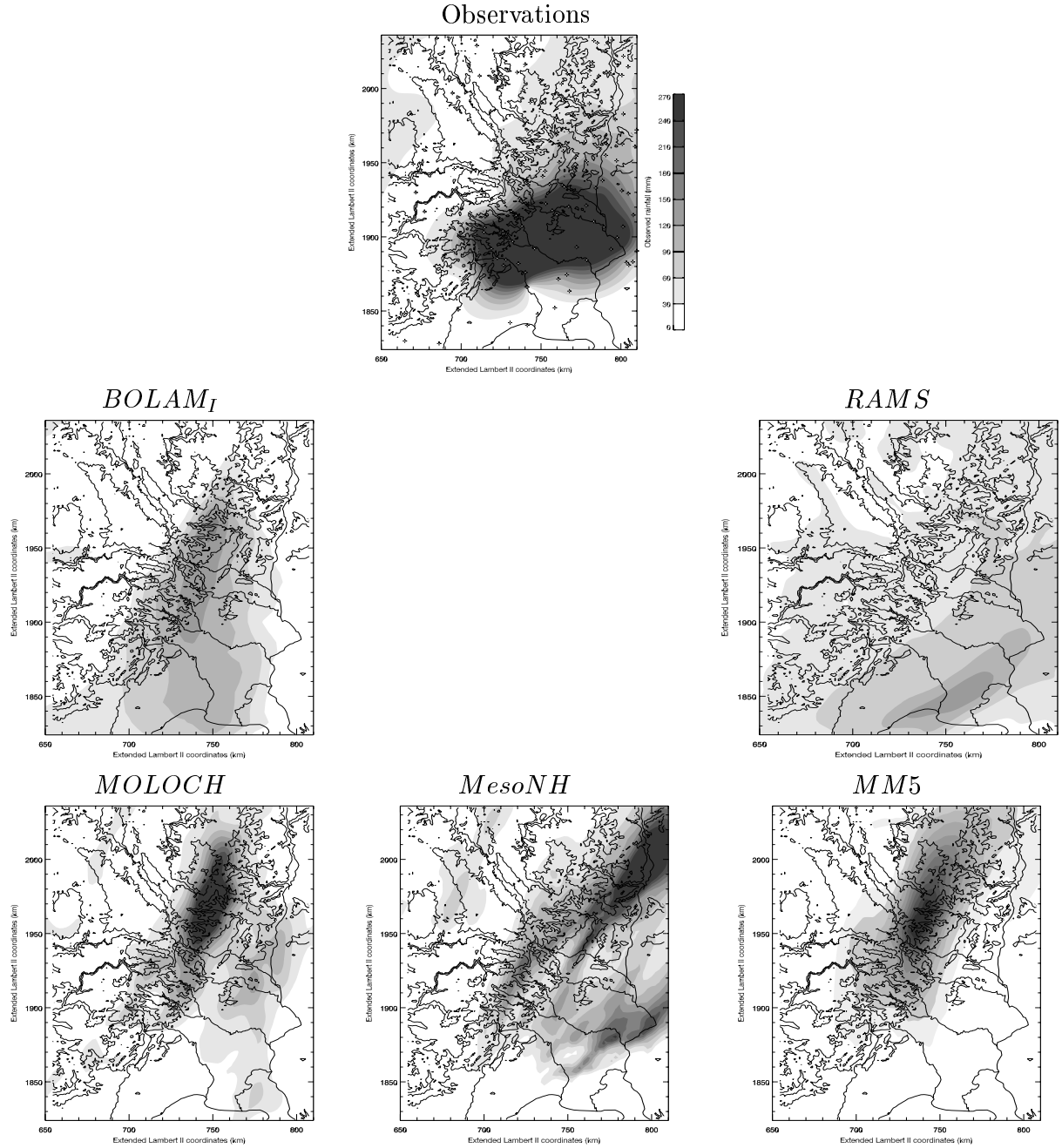


Fig. 9. 24 h accumulated rainfields ending at 12:00 UTC, 9 September 2002. A comparison between the observed rainfall and simulated rainfields obtained with the research model runs.

provides the best R^2 correlation coefficient. It shows that the shape of the simulated rain structure is closer to the observed rain pattern.

Tables 5, 6 and 7 give the scores obtained for the seven simulations and for the four accumulation time steps. To simplify and to make readable the intercomparison between the operational and research models, for each accumulation time step only the results obtained with the best hydrostatic model are compared to the results obtained with the best non-

hydrostatic model. This comparison is drawn in Figs. 10, 11 and 12. Based on Tables 5, 6 and 7 and Figs. 10, 11 and 12, the following comments can be made:

- (a) Due to the large underestimation of the simulated rain fields for the 7 simulations (i.e. Bias Table 4), the ACC values are bounded by the value of the quantile. Therefore, the values are roughly identical for all the simulations and equal to the quantiles (i.e. 0.7, 0.8 and 0.9).

Table 5. ACC score values for the seven model runs evaluated for the four accumulation time steps.

Accumulation time step Thresholds (mm)	1-hour			6-hour			12-hour			24-hour		
	4	7	15	24	38	73	43	73	154	88	161	294
BOLAM _S	0.68	.8	0.9	0.7	0.8	0.9	0.71	0.8	0.9	0.71	0.8	0.9
BOLAM _A	0.66	.77	0.88	0.7	0.8	0.9	0.71	0.8	0.9	0.72	0.8	0.9
BOLAM _I	0.65	.79	0.9	0.7	0.79	0.9	0.73	0.79	0.9	0.74	0.8	0.9
RAMS	0.68	.78	0.89	0.67	0.78	0.89	0.7	0.78	0.89	0.67	0.89	0.89
MOLOCH	0.65	.75	0.87	0.68	0.76	0.87	0.69	0.77	0.89	0.68	0.75	0.89
MM5				0.62	0.67	0.82	0.59	0.69	0.84	0.58	0.69	0.85
MesoNH	0.7	0.76	0.86	0.75	0.79	0.86	0.8	0.81	0.88	0.79	0.85	0.89

Table 6. FAR score values for the seven model runs evaluated for the four accumulation time steps.

Accumulation time step Thresholds (mm)	1-hour			6-hour			12-hour			24-hour		
	4	7	15	24	38	73	43	73	154	88	161	294
BOLAM _S	0.62	.62	–	0.41	–	–	0.32	–	–	0	–	–
BOLAM _A	0.86	.87	0.87	0.44	–	–	–	–	–	–	–	–
BOLAM _I	0.59	.94	1	0.5	0.83	–	0.44	0.80	–	0.4	1	–
RAMS	0.55	.72	0.97	0.58	0.70	1	0.5	0.89	–	0.68	–	–
MOLOCH	0.66	.78	0.88	0.55	0.72	0.89	0.54	0.70	1	0.59	1	1
MM5				0.81	0.84	0.80	0.82	0.79	0.60	0.91	0.77	0.61
MesoNH	0.49	0.64	0.83	0.35	0.54	0.81	0.34	0.48	0.74	0.37	0.30	0.67

Table 7. FBIAS score values for the seven model runs evaluated for the four accumulation time steps.

Accumulation time step Thresholds (mm)	1-hour			6-hour			12-hour			24-hour		
	4	7	15	24	38	73	43	73	154	88	161	294
BOLAM _S	0.23	0	0	0.11	0	0	0.13	0	0	0.03	0	0
BOLAM _A	0.27	.3	0.42	0.22	0	0	0	0	0	0	0	0
BOLAM _I	0.78	.07	0	0.65	0.07	0	0.8	0.09	0	0.69	0	0
RAMS	0.51	.27	0.06	0.44	0.22	0.08	0.34	0.16	0	0.24	0	0
MOLOCH	0.47	.47	0.41	0.48	0.44	0.32	0.58	0.42	0.12	0.45	0.23	0.13
MM5				0.82	0.84	0.80	0.83	0.8	0.61	0.92	0.78	0.61
MesoNH	0.59	0.67	0.75	0.64	0.67	0.68	1.13	1	0.35	1.14	0.69	0.28

- (b) For most of the thresholds and for the 2 other scores (FAR and FBIAS), the computed scores based on the research model simulated rain fields are better than the one based on the operational simulations.
- (c) Globally, all the models fail with the FAR score (i.e. a perfect simulation would have $FAR=0$). For the larger thresholds, the scores are even worst due to the large underestimation of the rainfall. Sometime, it is even not possible to compute the three scores due to the choice of the thresholds. The higher the threshold, the lower is the number of points entering the statistics, which explains why the scores may be misleading.
- (d) The FBIAS score is improved at large accumulation time step. Nevertheless, all the models fail for the largest thresholds.

- (e) For most of the accumulation time steps the MesoNH FAR score is the best. This is mainly due to the detailed initialization procedure as proposed by Ducrocq et al. (2000) that improves significantly the location of the rain pattern as already showed (Chancibault et al., 2005¹) for hydrological purposes.
- (f) Due to the larger underestimation of the simulated rain fields obtained with the three hydrostatic models and the choice of the thresholds, it is not always possible to compute the FAR scores.

5.2 Areal validation

Based on the recent work of Yates et al. (2005) and due to the above conclusions on the point validation step, only the MOLOCH and MesoNH simulated rainfalls are evaluated

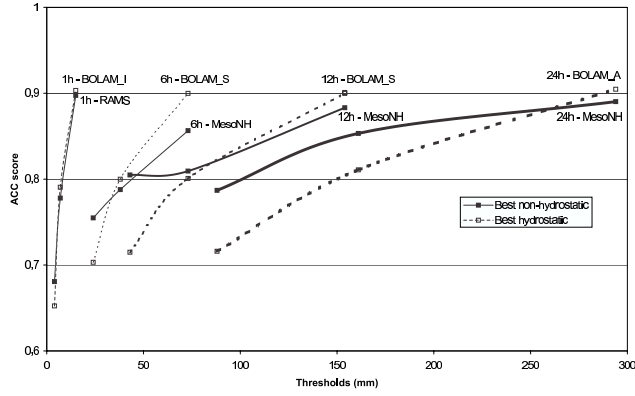


Fig. 10. Best evolution of the ACC scores for the hydrostatic (dotted lines) and non-hydrostatic (solid lines) models. The thinner lines are associated to the scores based on the 1 h accumulation time. Then, the bolder are the lines, the larger are the accumulation time steps. A perfect forecast has ACC=1.

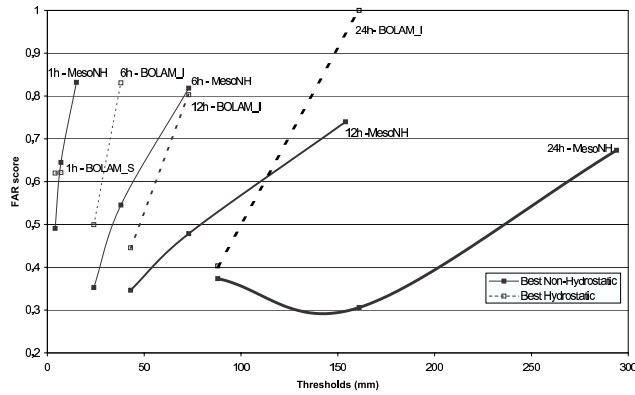


Fig. 11. Best evolution of the FAR scores for the hydrostatic (dotted lines) and non-hydrostatic (solid lines) models. The thinner lines are associated to the scores based on the 1 h accumulation time. Then, the bolder are the lines, the larger are the accumulation time steps. A perfect forecast has FAR=0.

in this section. These two simulated precipitation fields are the more accurate as regard to the continuous and threshold statistics.

Figure 13 shows the evolution of the correlation coefficient R^2 with the surfaces, has defined in Sect. 3.2.2, for the hourly and 24 h accumulated simulated rain fields. Due to the fast hydrological answer of the basins of the region, the hourly accumulation time is chosen to evaluate the capacity of the models to provide simulated rain fields that can be used as input of a hydrological model, whereas the results based on the 24 h accumulated rain fields will give information on the overall performance of the model to forecast the whole event.

For both models and for the two accumulation time steps,

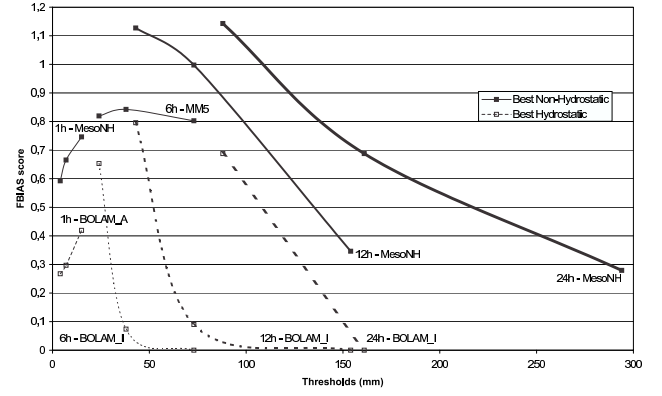


Fig. 12. Best evolution of the F1BIAS scores for the hydrostatic (dotted lines) and non-hydrostatic (solid lines) models. The thinner lines are associated to the scores based on the 1 h accumulation time. Then, the bolder are the lines, the larger are the accumulation time steps. A perfect forecast has F1BIAS=1.

the correlation coefficients are very low. This highlights the bad prediction of the location of the simulated rain pattern. For both models, the correlation coefficient for the 24 h accumulation time is higher than the one obtained for the hourly rain fields. Therefore, the use of the hourly simulated rain fields as input for hydrological models is not straightforward and the resulting discharges will be largely underestimated compared to the observations (Chancibault et al., 2005¹).

The MesoNH simulated rain fields provide a larger R^2 coefficient than MOLOCH for both accumulation time steps. As previously mentioned, this is due to the detailed initialization procedure (Ducrocq et al., 2000) that improves, at least for the first 12 h, the simulated rain dynamics as shown by Chancibault et al. (2005)³.

For both accumulation time steps and for the MOLOCH simulation, the R^2 evolution is very flat with the surface, whereas the curve of the R^2 evolution based on the MesoNH simulation presents an inflexion point at $S=150 \text{ km}^2$ for both time steps. Yates et al. (2005) show that this inflexion point highlights the critical surface for which the simulated rain fields can be stated relevant for hydrological purposes. Therefore, if the location of the simulated rain pattern is improved, the correlation coefficient R^2 value will be increased and it will be possible to rely on the hourly rain fields and to use them as input for discharge forecast for basins larger than 150 km^2 .

6 Conclusions

Within the framework of the European INTERREG-IIIb Medocc program, the HYDROPTIMET project aimed at the optimization of the hydrometeorological forecasting tools in the context of intense precipitation within marked topography.

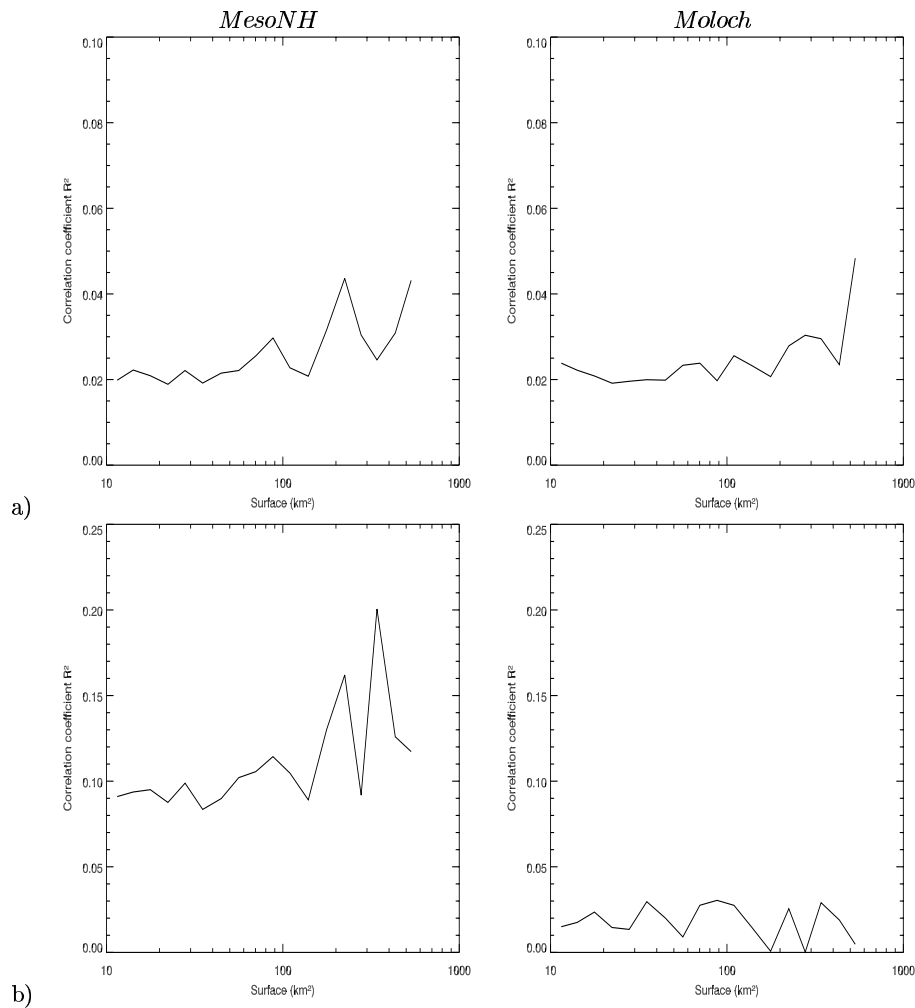


Fig. 13. Correlation coefficient R^2 evolution with the surfaces based (a) on the hourly and (b) on the 24 h accumulated simulated rainfall for the MOLOCH and MesoNH simulations.

Several meteorological models were tested on 4 meteorological situations chosen for their amount of the cumulated precipitation and the resulting damages. One of them occurred in the South-eastern part of France in the Cévennes-Vivarais region during September, 2002.

This paper presents the comparison between seven meteorological simulations that were run for this intense precipitation event. Two types of models are evaluated: (1) the operational models, based on the hydrostatic model BOLAM, that use different types of grid resolution and convective schemes; (2) the research models mainly non-hydrostatic which are based on higher grid resolution allowing an explicit resolution of convection. The main objective is to propose some recommendations to build the next generation of forecasted tools.

The evaluation procedure is twofold: the first approach is based on the classical point validation methodology using both continuous and threshold statistics; the second one is based on a multi-scale approach that highlights the relevant

atmospheric simulated scale for hydrological purposes.

The results show that all the simulated rain fields are underestimated. This bias is reduced for most of the research non-hydrostatic models compared to the one obtained with the operational models. Therefore, one of the conclusions is to put forward the need to increase the grid resolution of the operational models in order to explicitly resolve the convection that will lead to predict higher amount of rainfalls. Several National Weather Services already plan to run operationally by the end of this decade such non-hydrostatic models with horizontal resolution of the order of 1–3 km.

All the models fail on the rain pattern location prediction. All the models give a simulated rain fields too far in the north. This point is less critical when the initialization uses a mesoscale data assimilation, allowing to improve the first 12 hours of the event. Nevertheless, these results show the difficulty of using directly QPF for hydrological applications without taking into account uncertainties in location and rainfall amount of the forecast precipitation fields. The

INTERREG IIIb AMPHORE project aims at taking into account these uncertainties through testing ensemble meteorological forecasts for hydrological purposes.

Acknowledgements. The current study was funded by the INTERREG III-b European program within the framework of HYDROPTIMET project coordinated by the ARPA-Piedmont (Italy).

Edited by: A. Buzzi

Reviewed by: one referee

References

- Beven, K. and Kirkby, M.: A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 43–69, 1979.
- Bougeault, P., Binder, P., Buzzi, A., Dirks, R., Houze, R., Kuettner, J., Smith, R., Steinacker, R., and Volkert, H.: The MAP special observing period, *Bull. Amer. Meteorol. Soc.*, 82, 433–462, 2001.
- Davolio, S. and Buzzi, A.: A nudging scheme for the assimilation of precipitation data into a mesoscale model, *Weather and Forecasting*, 19, 855–871, 2004.
- Davolio, S., Buzzi, A., and Malguzzi, P.: Orographic influence of deep convection : case study and sensitivity experiments, *Proceeding of the 28th International conference on Alpine Meteorology (ICAM) and the annual Scientific meeting of the Mesoscale Alpine Programme (MAP)*, Croatie, 159–162, 2005.
- Delrieu, G., Ducrocq, V., Gaume, E., Nicol, J., Payrastré, O., Yates, E., Kirstetter, P.-E., Andrieu, H., Ayral, P.-A., Bouvier, C., Creutin, J.-D., Livet, M., Anquetin, S., Lang, M., Neppel, L., Obled, C., du Chatelet, J. P., Saulnier, G.-M., Walpersdorf, A., and Westbrook, W.: The catastrophic flash-flood event of 8–9 September 2002 in the Gard regio, France: a first case study for the Cévennes – Vivarais Mediterranean Hydrometeorological Observatory, *J. Hydrometeorol.*, 6, 34–52, 2005.
- Ducrocq, V., Lafore, J.-P., Redelsperger, J.-L., and Orain, F.: Initialization of a fine-scale model for convective-system prediction: a case study, *Q. J. Roy. Meteor. Soc.*, 126, 3041–3065, 2000.
- Ducrocq, V., Ricard, D., Lafore, J., and Orain, F.: Storm-scale numerical rainfall prediction for five precipitating events over France: On the importance of the initial humidity field, *Wea. Forecasting*, 17, 1236–1256, 2002.
- Ducrocq, V., Chancibault, K., Habets, F., and Anquetin, S.: Mesoscale modelling of a cooding storm. Application to the extreme flood of Gard, 5th EGS Plinius Conference, Ajaccio, Corsica, France, 2003.
- Ducrocq, V., Lebeaupin, C., Thouvenin, T., Giordani, H., Chancibault, K., Anquetin, S., and Saulnier, G.: L'évènement des 8–9 septembre 2002: situation météorologique et simulation à mésoéchelle (in French), *La Houille Blanche*, 6, 86–92, 2004.
- Ducrocq, V., Lebeaupin, C., Ricard, D., and Giordani, H.: Torrential rain events over the Cévennes-Vivarais region, *Proceeding of the 28th International conference on Alpine Meteorology (ICAM) and the annual Scientific meeting of the Mesoscale Alpine Programme (MAP)*, Croatie, 254–258, 2005.
- Fritsch, J., Houze, R. J., Adler, R., Bluestein, H., Bosart, L., Brown, J., Carr, F., Davis, C., Johnson, R., Junker, N., Kuo, Y., Rutledge, S., Smith, J., Toth, Z., Wilson, J., Zipser, E., and Zrnic, D.: Quantitative precipitation forecasting: report of the 8th prospectus development team, US Weather Research Program, *Bull. Amer. Meteorol. Soc.*, 79, 285–299, 1998.
- Huet, P., Martin, X., Prime, J.-L., Foin, P., Laurain, C., and Canard, P.: Retour d'expérience des crues de septembre 2002 dans les départements du Gard, de l'Hérault, du Vaucluse, des Bouches-du-Rhône, de l'Ardèche et de la Drôme. (In French), Tech. rep., Ministère de l'écologie et du développement durable, République Française, 2003.
- Kain, J. and Fritsch, J.: A one-dimensional entraining/detraining plume model and its application in convective parameterization., *J. Atmos. Sci.*, 47, 2784–2802, 1990.
- Kain, J. and Fritsch, J.: Convective parameterization for mesoscale models: The Kain-Fritsch scheme, *Meteorological Monography*, 46, 165–170, 1993.
- Kuo, H.: Further studies of the parameterization of the influence of cumulus convection on large scale flow, *J. Atmos. Sci.*, 31, 1232–1240, 1974.
- Lafore, J., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., Fischer, C., Hérel, P., Mascart, P., Masson, V., Pinty, J., Redelsperger, J., Richard, E., and Vilà-Guerau de Arellano, J.: The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations, *Ann. Geophys.*, 16, 90–109, 1998.
- SRef-ID: 1432-0576/ag/1998-16-90.**
- Lebel, T., Bastin, G., Obled, C., and Creutin, J.-D.: On the accuracy of areal rainfall estimation: a case study, *Wat. Resour. Res.*, 23, 2123–2138, 1987.
- Leheman, R.: On the choice of relaxation coefficients for Davies lateral boundaries scheme for regional weather prediction models, *Meteorol. Atmos. Phys.*, 52, 10–14, 1993.
- Mariani, S., Casaioli, M., Accadia, C., Llasat, M., Pasi, F., Davolio, S., Elementi, M., Ficca, G., and Romero, R.: A limited area model intercomparison on the Montserrat-2000 flash-flood event using statistical and deterministic methods, *Nat. Hazards Earth Syst. Sci.*, 5, 565–581, 2005.
- SRef-ID: 1684-9981/nhess/2005-5-565.**
- Mason, I.: Dependence of the critical success index on sample climate and threshold probability, *Aust. Meteor. Mag.*, 37, 75–81, 1989.
- Roebber, P. and Eise, J.: The 21 June 1997 flood: storm-scale simulations and implications for operational forecasting, *Wea. Forecasting*, 16, 197–218, 2001.
- Spencer, P. and Stensrud, D.: Simulating flash flood events: importance of the subgrid representation of the convection, *Mon. Wea. Rev.*, 126, 2284–2912, 1998.
- Speranza, A., Accadia, C., Casaioli, M., Maraini, S., Inghilesi, R., Tartaglione, P., Ruti, M., Carillo, A. and Bargagli, A., Valentinotti, F., and Lavagnini, A.: PSEIDON: An integrated system for analysis and forecast of hydrological, meteorological and surface marine fields in the Mediterranean area, *Nuovo Cimento*, 27C, 329–345, 2004.
- Stein, J., Richard, E., Lafore, J., Pinty, J., Asencio, N., Cosma, S., and Volkert, H.: High-resolution non-hydrostatic simulations of flash-flood episodes with grid-nesting and ice-phase parameterization, *Heavy Precipitation in the Alpine Region (HERA)*, *Meteor. Atmos. Phys.*, 72, 203–221, 2000.
- Yates, E., Anquetin, S., Ducrocq, V., Creutin, J.-D., Ricard, D., and Chancibault, K.: Point and areal validation of forecasted precipitation, *Meteorol. Appl.*, in press, 2005.