**Natural Hazards
and Earth
System Sciences**

Predictive ability of severe rainfall events over Catalonia for the year 2008

A. Comellas[1], L. Molini[1], A. Parodi[1], A. Sairouni[2], M. C. Llasat[3], and F. Siccardi[1,4]

[1]CIMA Research Foundation, Savona, Italy
[2]SMC (Meteorological Service of Catalonia), Barcelona, Spain
[3]GAMA (Meteorological Hazards Analysis Team), Department of Astronomy & Meteorology, Faculty of Physics, University of Barcelona, Spain
[4]University of Genoa, Italy

Abstract. This paper analyses the predictive ability of quantitative precipitation forecasts (QPF) and the so-called "poor-man" rainfall probabilistic forecasts (RPF). With this aim, the full set of warnings issued by the Meteorological Service of Catalonia (SMC) for potentially-dangerous events due to severe precipitation has been analysed for the year 2008. For each of the 37 warnings, the QPFs obtained from the limited-area model MM5 have been verified against hourly precipitation data provided by the rain gauge network covering Catalonia (NE of Spain), managed by SMC. For a group of five selected case studies, a QPF comparison has been undertaken between the MM5 and COSMO-I7 limited-area models. Although MM5's predictive ability has been examined for these five cases by making use of satellite data, this paper only shows in detail the heavy precipitation event on the 9–10 May 2008. Finally, the "poor-man" rainfall probabilistic forecasts (RPF) issued by SMC at regional scale have also been tested against hourly precipitation observations. Verification results show that for long events (>24 h) MM5 tends to overestimate total precipitation, whereas for short events (≤ 24 h) the model tends instead to underestimate precipitation. The analysis of the five case studies concludes that most of MM5's QPF errors are mainly triggered by very poor representation of some of its cloud microphysical species, particularly the cloud liquid water and, to a lesser degree, the water vapor. The models' performance comparison demonstrates that MM5 and COSMO-I7 are on the same level of QPF skill, at least for the intense-rainfall events dealt with in the five case studies, whilst the warnings based on RPF issued by SMC have proven fairly correct when tested against hourly observed precipitation for 6-h intervals and at a small region scale.

Throughout this study, we have only dealt with (SMC-issued) warning episodes in order to analyse deterministic (MM5 and COSMO-I7) and probabilistic (SMC) rainfall forecasts; therefore we have not taken into account those episodes that might (or might not) have been missed by the official SMC warnings. Therefore, whenever we talk about "misses", it is always in relation to the deterministic LAMs' QPFs.

1 Introduction

The prediction of precipitation has been one of the main objectives of meteorology since its beginnings as a modern discipline some fifty years ago (Harper et al., 2007). At the same time, precipitation is one of the most difficult variables to be predicted due to its complex triggering factors (Amengual et al., 2007; Casati et al., 2008; Pappenberger et al., 2008; Rotach et al., 2009) and its own increasing chaotic behavior as spatial and temporal resolutions become finer and finer (Adlerman and Droegemeier, 2002; Bryan et al., 2003; Roberts and Lean, 2008; Rezacova et al., 2009; Llasat and Siccardi, 2010).

Quantitative precipitation forecasts (QPF) are traditionally undertaken by using either the deterministic or the probabilistic Numerical Weather Prediction (NWP) approaches (Golding, 2000). Both approaches are motivated and guided by the problem of uncertainty in the prediction of the spatio-temporal evolution of precipitation-related processes, which becomes crucial when the forecast is for a severe-rainfall event potentially leading to flash flooding (Doswell et al., 1996; Ferraris et al., 2002).

For this reason, and even more in the Mediterranean context of steep coastal orography and dense infrastructure development, it is very important to understand the

Correspondence to: A. Comellas
(albert.comellas@cimafoundation.org)

mechanisms through which uncertainty in NWP-based QPFs affects severe-rainfall warnings issued by the National Meteorological Services and Agencies (Levy and Hall, 2005; Norbiato et al., 2008; Molini et al., 2009). Such severe-rainfall warnings are usually based on the detailed inspection and consideration of the different large-scale and/or mesoscale NWP models' outputs available, complemented by poor-man approaches of different complexity aimed to take into account the forecaster(s) expertise and knowledge of local and fine-scale meteorological processes, inducing certain weather scenarios for a given area of interest (Ebert, 2001).

Then, it becomes crucial to gain a deeper understanding of the overarching topic of the hydro-meteorological predictive ability, referring to how close a particular model can predict the true state of the system and be affected by data quality and availability, parameterization of sub-grid processes, model approximations, and resolution (Kain et al., 2008; Roberts, 2008).

It is the purpose of this paper to analyze, on one hand, the uncertainty in MM5's (PSU/NCAR Mesoscale Model 5) QPFs for the set of rainfall intensity/depth warnings issued by the Meteorological Service of Catalonia (SMC) during 2008, as well as to perform, from the aforementioned predictive ability perspective, a simple comparison with COSMO-I7's (COnsortium for Small-scale MOdeling model) QPFs for some of the 2008 events taken as case studies. On the other hand, it aims to study the accuracy of SMC's "poor-man" rainfall probabilistic forecast (RPF) system, which is based on the combination of different limited-area models (LAM) operated by SMC (MM5, WRF, MASS) outputs received from outside (COSMO's from the German Weather Service, DWD) and in-house experience. The RPF issuance, in this case, is only human-based, whereas the approach described by Ebert (2001) is more complex and the outcome of an automatic model ensemble.

The paper is structured as follows: Sect. 2 describes the technical tools used here, as well as background information on the SMC's warning system and on the NWP models compared in this study; Sect. 3 explains the working procedures to get the results, with special emphasis on the observational dataset (rain gauge network and remote-sensing technology) available; Sect. 4 presents the results of the deterministic forecasts (MM5) analysis with special emphasis on the event of May 2008; Sect. 5 deals with MM5 and COSMO-I7 performances comparison for the five case studies; Sect. 6 exposes the findings of the SMC's RPFs analysis at regional level; and finally, Sect. 7 presents the conclusions of the overall results.

**Table 1.** Thresholds set by SMC corresponding to the different precipitation warnings and levels (data source: http://www.meteo.cat/).

Precipitation type	Level 1	Level 2
Rain Intensity	$> 20\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$	$> 40\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$
Accumulated Rainfall	$> 100\,\mathrm{mm}\,24\,\mathrm{h}^{-1}$	$> 200\,\mathrm{mm}\,24\,\mathrm{h}^{-1}$

2 Observational and predictive tools

2.1 SMC's warning system

In Catalonia, the first official step in order to activate the meteorological or hydro-meteorological warning chain is taken by the SMC. This happens when this institution has gathered enough predictive evidence that a potentially-dangerous situation can take place or a risk threshold can be surpassed. At that moment, they issue a *meteorological situation of risk* warning (SMR), which comes out via the SMC's website and some TV and radio stations. The warning, which depends on the weather type expected to be dangerous, is transmitted to the CECAT (the Control Center of the Catalan Civil Protection), which is responsible for undertaking the necessary procedures with the objective of preparing civil population and infrastructures for the hazardous weather expected. If the warning refers to precipitation, the Catalan Water Agency (ACA) is also warned. There are different types of warnings (for rain intensity, accumulated rainfall, 24 h accumulated snowfall, wind, sea waves, etc.), each of them being given the level 1 or 2, in function of the threshold to be overcome. Most of the SMR warnings imply only level 1, while level 2 is kept for very extreme scenarios. Each meteorological hazard has its own thresholds for level 1 and level 2 (set appropriately by SMC). Table 1 shows the thresholds corresponding to warnings for precipitation.

The SMC issues a SMR warning when they predict the overcome of any of the thresholds associated to the given meteorological hazard within the next 36 h, with a temporal resolution of every 6 h. The SMC's level of confidence and (un)certainty in the prediction is expressed as a probability range (in %) of the given threshold being overcome or not, which is spatially resolved at a so-called *comarca* scale (Catalonia is composed of 41 of them; they are the minimum administrative region with an average extension of approximately 800 km^2). Such probabilistic forecasts, corresponding to the SMC's "poor-man" RPF system, can split (or not) the territory of Catalonia in any or all of the following 3 categories: very likely (probability superior to 70 %), likely (probability between 30 % and 70 %), and possible (probability inferior to 30 %).
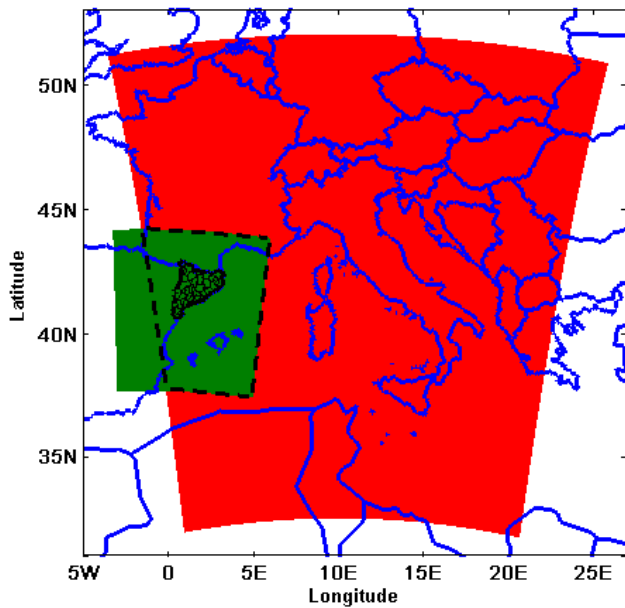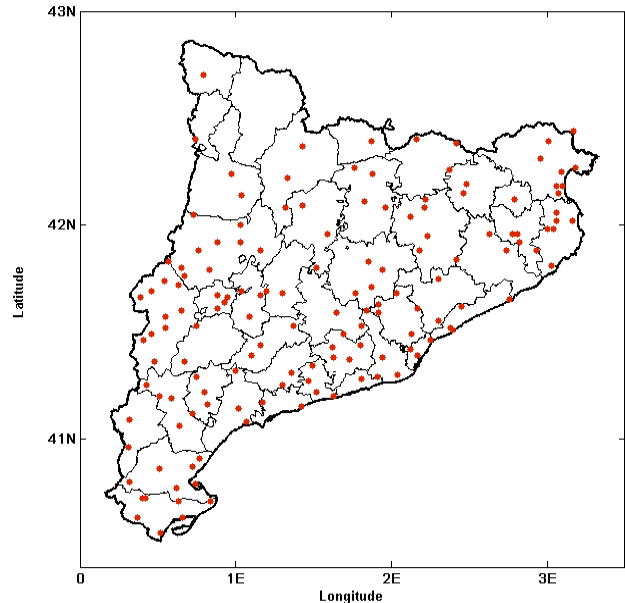
Fig. 1. MM5's domain (in green) superimposed to COSMO-I7's (in red). A dashed black line encloses the overlapping area. Catalonia and its comarca division is showed in black lines.

Fig. 2. Location of the 146 automatic rain gauges (red dots) from the XEMA network all over Catalonia in 2008, with the comarca division (thin black line). This represents an average density of ~1 rain gauge $200\,\mathrm{km}^{-2}$.

2.2 NWP models used

Two limited-area models, MM5 and COSMO-I7, are used in this study to compare their respective predictive ability under scenarios of severe rainfall events. Such a comparison, however, is far from being perfect, as both models use different parameterizations and have different spatial resolutions and integration domains. Thus, it is sought in this paper to perform a model intercomparison from an operational point of view, assuming that each model has been run under its particular best possible configuration. MM5 is one of the models run operatively at SMC, while COSMO-I7 is run at ARPA-SIMC, the regional hydro-meteorological service of Emilia-Romagna, and it is also used as reference LAM within the Italian Civil Protection hydro-meteorological forecasting chain (Italy Official Gazette, 2004).

The two models have overlapping domains over Catalonia (Fig. 1). They are nonhydrostatic and fully compressible, and were run twice a day at 00:00 and 12:00 UTC, with boundary conditions supplied by the ECMWF global model every 3 h, with a spatial resolution of 0.25° ($\sim$25 km). MM5 has a horizontal resolution of 12 km and 30 vertical levels, while COSMO-I7 has a 7-km horizontal resolution and 50 vertical levels. For sub-grid scale processes, MM5 and COSMO-I7 use, respectively, the following parameterizations: Grell (1993) and Tiedtke (1989) for convection, Schultz (1995) and a three-category ice scheme (Doms and Forstner, 2004) for microphysics, MRF (Hong and Pan, 1996) and Mellor and Yamada (1974) for Planetary Boundary Layer (PBL hereafter), Dudhia (1989) and Rit-

ter and Geleyn (1992) for radiation, and Noah Land-Surface Model (Chen and Dudhia, 2001) and TERRA (Jacobsen and Heise, 1982) for soil surface fluxes.

3 Methodology

This work has been guided by the list of all 37 early rainfall warnings issued by SMC during 2008. For each episode, the MM5 rainfall prediction outputs (corresponding to the area in green color, Fig. 1) have been kindly provided by SMC. Five cases have been selected in order to compare MM5's and COSMO-I7's (provided by CIMA Research Foundation) performances, which have been chosen following the considerations specified below. The observed rainfall fields (precipitation as measured by rain gauges interpolated over either model grid) used for verifying QPFs from both models have been derived from the hourly data of the 146 rain gauges present in the XEMA network, courtesy of SMC. The 2008 network is represented in Fig. 2.

With the objective of giving MM5 QPFs some flexibility in order to predict the observed maximum mean rainfall amounts as closely as possible, we have searched every observed episode (i.e., the maximum mean rainfall depth for the episode duration) within a moving time window beginning 12 h before the episode start and ending 12 h after the episode end (as in Molini et al., 2009). Hence, a model lag of a few hours has been considered as acceptable only if its rainfall forecast was appropriate when being verified

against observations for the episode duration and within the time window of ±12 h.

The 37 events have been separated into two groups according to their duration: short and long episodes. Short episodes are those whose duration is 24 h or less, and long episodes are those lasting more than 24 h. This criterion is motivated by the fact that long ones are expected to be largely driven by large-scale forcings and thus more stratiform in nature, while those short ones are mainly affected by local-scale effects and consequently more of convective behavior (Done et al., 2006; Molini et al., 2010).

In agreement with Casati et al. (2008), for every episode the bias or mean error (ME), the mean absolute error (MAE), and the root mean square error (RMSE) have been calculated and a scatterplot of MM5's QPF against observations has also been made. Thus, the main features of every episode (model overestimation, underestimation, anticorrelation, good agreement or insignificant values, or a combination of these) have been identified and classified from such scatterplots. Predicted and observed rainfall fields have also been plotted, and underwent visual inspection in order to visually (*subjectively*) verify the QPFs against observations, in terms of rainfall location and rainfall depth.

A detailed analysis of five case studies has been done. They have been chosen for their representativity of typical NWP QPF errors in the form of either large overestimation (2 long episodes) or large underestimation (basically 2 missed short episodes), except one case study that was special for its high average accumulations but which was acceptably predicted. In order to perform such analyses, we have used the satellite data described below, as well as the scatterplots of predicted against observed precipitation. An extra verification index has been calculated for these cases: the O/P ratio, corresponding to the total summed observed rainfall depth (O) over its predicted value (P). Thus, a value of O/P ~ 1 approximately translates into a *good* forecast, <1 into overestimation, and >1 into underestimation. However, this index is used only as a rough guide, as it deals with total values and thus it could still be 1 but as result of over and underestimations cancelling each other out. Table 2 shows the main features of these five events studied, and Fig. 3 displays the time evolution of their 3-hourly maximum observed precipitation.

Uncertainty has been analyzed for each case study following the methodology proposed by Molini et al. (2009), which requires knowledge of satellite-measured fields (Mirza et al., 2008) of columnar cloud liquid water and columnar water vapor contents (Advanced Microwave Scanning Radiometer, AMSR-E, dataset), as well as surface wind speed vectors (NASA's Quick Scatterometer, QuikScat, dataset).

The spatial mean values of columnar water vapor and columnar cloud liquid water, for the satellite pass times, have been measured over the sea area enclosed between longitudes $0°$ and $5°$ E and latitudes $39°$ N and $43°$ N for the five case studies. This region has been considered a representative
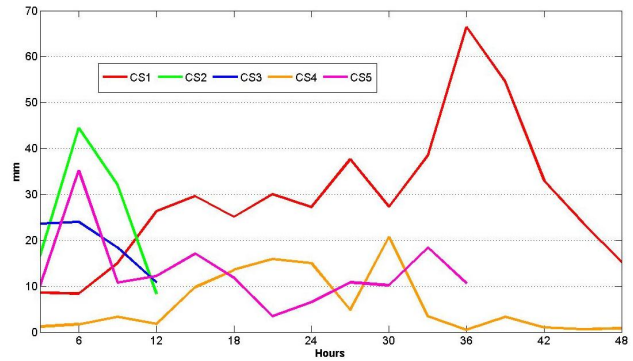
Fig. 3. Evolution of 3-hourly maximum observed precipitation over Catalonia for the five case studies.

enough atmospheric volume over the sea surrounding Catalonia as to regard these mean values as meaningful when advected over the territory under southerly to easterly air flow regimes, which are the most usual ones when heavy rainfall occurs in the region (Llasat, 1987; Llasat and Puigcerver, 1994). Some previous analyses have also shown the strong role of the sea in the convective precipitation produced in Catalonia and, particularly, in this selected sea area (Llasat and Puigcerver, 1997; Rigo and Llasat, 2007).

For the analysis of the probabilistic approach to intense-rainfall forecasting of the warning episodes during 2008, we verified RPFs against hourly observations (that is, if the given threshold was overcome or not in every rain gauge of the comarcas under warning). Therefore, in order to allow some flexibility to the forecasts, we checked also the comarcas neighboring those warned ones, and considered the probabilistic forecast to be correct when at least one rain gauge reached the threshold or overcame it. This procedure was repeated for every 6-h interval from every episode in 2008, and the results were classified as function of the probability range (3 categories; see Sect. 2.1) given to every time interval.

4 Analysis of MM5 forecasts

### 4.1 Verification of forecasted rainfall fields

The year 2007 and the beginning of 2008 are remembered, in Catalonia, as a period of severe drought (Altava-Ortiz et al., 2008). As mentioned in the climatic bulletins from SMC (which the reader may find at http://www.meteo.cat/), approximately from summer 2007 until spring 2008 all Catalonia suffered the effects of precipitation being well below average. This situation then changed, as the spring 2008 was rainier than the mean and the rest of the year was irregular in terms of precipitation (in function of the region). On the whole, however, 2008 is considered as normal (average rainfall) or rainy (above the average rainfall) in most of the country. In total during that year, SMC issued 37 warnings (all of

Table 2. Main rainfall features of the five events studied measured with the rain gauge network (XEMA). The rainfall depths are from the rain gauge with the maximum value for the given period of time.

CS	DATE (of 2008)	DUR. (h)	12 h Max. Rainfall Depth (mm)				Total max. Rainfall Depth (mm)
1	9–10 May	48	39	77	104	71	181
2	25–26 May	12	75				75
3	26–27 June	12	36				36
4	12–13 Oct	48	8	55	27	3	90
5	28–29 Oct	36	40	27	42		76

Table 3. Verification indices (Bias, MAE and O/P) corresponding to MM5's and COSMO-I7's QPF performances for the five case studies analysed in this study. The ratio O/P, corresponding to the total summed observed rainfall depth over its predicted value for all XEMA rain gauges, is also shown.

| CS		MM5			COSMO-I7		
	Duration (h)	Bias (mm)	MAE (mm)	O/P	Bias (mm)	MAE (mm)	O/P
1	48	6	38	0.92	35	45	0.66
2	12	−21	21	86	−17	18	5.6
3	12	−10	10	8.8	−11	11	35
4	48	9	11	0.19	1	4	0.75
5	36	1	10	0.96	0	8	1.01
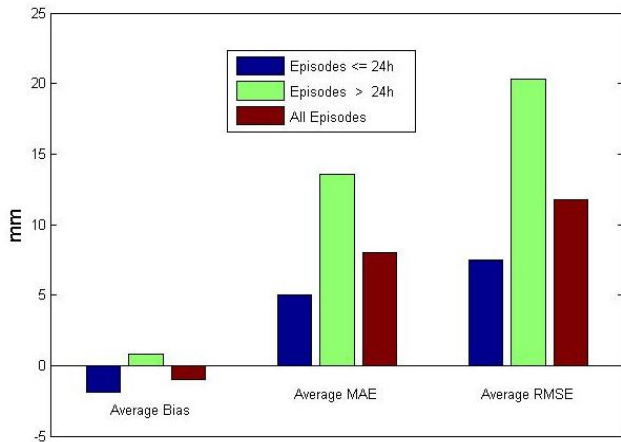
Fig. 4. Verification indices (bias, MAE and RMSE) for MM5's QPFs against rainfall observations for all warning episodes during 2008.

level 1) for rain intensity and/or rainfall accumulation. Of these, 25 warnings corresponded to short episodes (duration ≤24 h), and 12 warnings to long episodes (duration >24 h).

Figure 4 shows the average verification indices for MM5's performances for short, long, and all warned episodes during 2008. The average bias (which means model underestimation if negative and model overestimation if positive) has been found to be small but positive (0.8 mm) for long episodes, and larger but negative (−1.9 mm) for short episodes. The average bias for the whole set of episodes is thus negative (since the number of short episodes approximately doubles that of long ones, short episodes weight more when calculating any average over the whole set of episodes). Generally, then, it can be said that short episodes (≤24 h) tend to be underestimated by the model, while long episodes (>24 h) tend to be overestimated. The average MAE and RMSE have been found to be much bigger for long episodes than for short ones, showing that the former are more easily affected by errors caused by geographic misplacement of precipitation and/or by errors in the rainfall depth forecasted. Short episodes may carry big errors of precipitation misplacing too, but due to the smaller magnitude of the predicted rainfall depth associated to this kind of episodes, the overall error is usually much smaller. Moreover, it's quite common that mesoscale NWP models provide forecasts with realistic small-scale precipitation patterns, but with amplitude and gradients which may be somewhat misplaced (due to the convective parameterization) when they are verified with common verification methods (Sairouni et al., 2007).
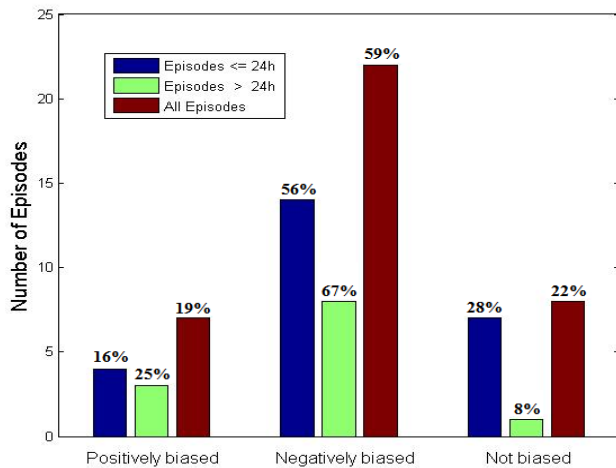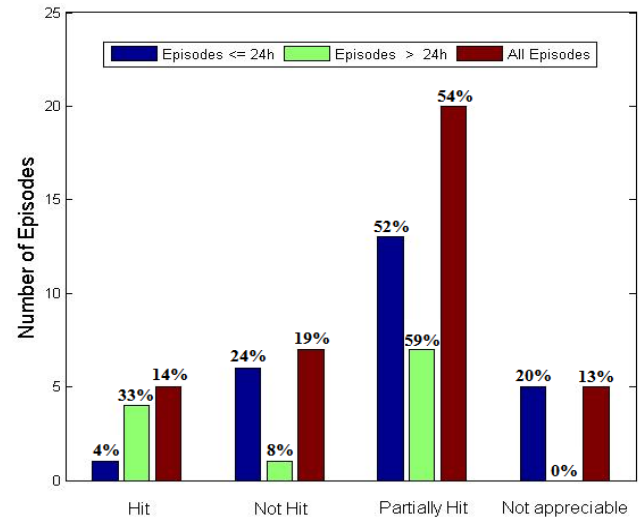
Fig. 5. Bias tendency for MM5's QPFs against rainfall observations for all warning episodes during 2008.

Fig. 6. Visual verification considering the maps of MM5's QPFs and rainfall observations for all warning episodes during 2008.

Figure 5 displays all episodes grouped corresponding to their biases: positively, negatively or not biased at all. Of all short episodes, 16 % were positively biased, 56 % were negatively biased, and 28 % were not biased. We see, therefore, that the majority of them was negatively biased, which matches the fact that the average bias for such episodes was also negative. Of all long episodes, 25 % were positively biased, 67 % negatively biased, and 8 % had no bias. However, the average bias for long episodes was positive, which means that this 25 % of positively biased episodes had rather large biases (i.e., were heavily overestimating) and made up for and overcame the weight of the 67 % of episodes negatively biased. From all episodes, 8 had zero bias, of which 7 were short episodes and just 1 was a long one. This is because short episodes are more likely to present cancellation as a result of the usually small values implied in this kind of event. Long episodes are less likely to show zero bias, as they imply bigger amounts of precipitation and so less probability of cancellation.

Considering now the patterns shown in the scatterplots of predicted against observed precipitation, the episodes have been classified as overestimated, underestimated, anticorrelated (those ones displaying both strong overestimation and underestimation simultaneously but in different regions), "well-agreeing" or as small-valued (not shown). However, these features are not exclusive; a given episode can exhibit one or more of them. In this way, of all short episodes, 28 % showed overestimation, 48 % underestimation, 24 % anticorrelation, 12 % showed good agreement, and 32 % of them were small-valued. Underestimation is thus the main feature of MM5 performances also under this criterion. Such a shortage of acceptable rainfall forecasts for short episodes may be due to the fact that, on one hand, they can be affected by the spin-up problem in NWP models, while on the other hand,
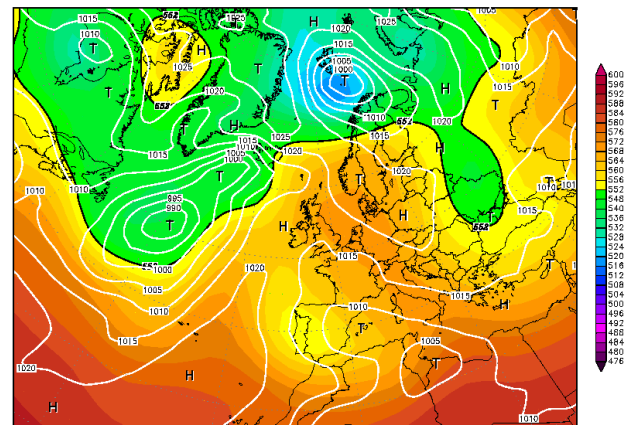
Fig. 7. Geopotential height (shaded, in dam) at 500 hPa and MSLP (contours, in hPa) for 00:00 UTC 10 May 2008 (source: www.wetterzentrale.de).

this type of episodes is implicitly of convective nature, and hence is difficult to be forecasted correctly by the NWP models. It is known that typical convective scales are basically smaller than the models' resolution, and their parameterization usually proves not good enough: consequently, models usually underestimate this kind of events (Zhang et al., 2003; Kain et al., 2008).

From all long episodes, 25 % showed overestimation, 33 % underestimation, 33 % anticorrelation, 42 % good agreement, and none was small-valued. Not surprisingly, this 25 % of overestimated episodes coincides with the 25 % of long episodes positively biased, whereas such coincidence does not happen with those showing underestimation. This

is because the average bias of those long episodes positively biased is quite high (19 mm), but the average bias of those long episodes negatively biased is not much lower than zero (−6 mm), so some of them can be considered as well-forecasted episodes instead of as underestimated.

Another method, although more subjective and partially inspired by Ebert (2008), has been used to verify rainfall forecasts against observations: to visually compare the maps of MM5's QPFs against observed precipitation for every episode. If the areas affected by precipitation and the magnitudes of both prediction and observation coincide approximately, the episode was regarded as "hit" (i.e., correctly forecasted); if this only happened in some parts of the territory, then the episode was regarded as "partially hit"; if the areas of predicted and observed precipitation did not coincide at all, the episode was regarded as "not hit" (i.e., wrongly forecasted; either as "false alarm" or as a "missed" episode). When values were simply too small as to judge reliably, the episode was classified as "not appreciable". The results of this visual verification are shown in Fig. 6.

Under this method, it follows that among long episodes, 33 % were regarded as "hit", 59 % as "partially hit", and 8 % as "not hit"; among those short ones, only 4 % were regarded as "hit", 52 % as "partially hit", 24 % as "not hit", and 20 % as "not appreciable". Likewise, we can state that 80 % (20 %) of those "hit" episodes corresponded to long (short) ones, and that of those regarded as "not hit", 14 % (86 %) corresponded to long (short) episodes. From "partially hit" episodes, 35 % (65 %) were due to long (short) episodes.

Concluding this section, it can be said that a vast majority of long episodes were either correctly forecasted by MM5 or were correct at least to some extent, and that only a small minority were missed or false alarms. Contrarily, short episodes had a very low "hitting" rate, although most of the rainfall forecasts for these episodes were partially correct. Short episodes' forecasts also presented a rather high index of missed/false alarm cases.

4.2 Case study analysis: the example of 9–10 May 2008

This case study (CS, 00:00 UTC 9 May to 00:00 UTC 11 May), which was the most severe in 2008 in terms of average rainfall depth all over Catalonia, involved clear model overestimation on the whole, although there were two distinct areas of over- and underestimation. The synoptic situation was dominated by a mid-level trough at 500 hPa penetrating to the Iberian peninsula from the northwest and advecting positive vorticity over Catalonia at mid-levels, together with a humid eastern flux at surface over Catalonia caused by a weak low-pressure centre coming from North Africa (Fig. 7). Such elements constitute a rather typical example of a situation potentially leading to large amounts of precipitation over the northeastern Iberian peninsula (Llasat, 1987; Doswell et al., 1996). All in all, the maximum total precipitation measured

in Catalonia during the episode (48 h) was nearly 200 mm, with a maximum rainfall intensity of 33.4 mm h^{-1}.

MM5's QPF for this long episode was quite good in the central and western areas of Catalonia and it pointed out well the locations of the precipitation maxima, but there was large rainfall overestimation in some coastal areas and underestimation in the southwestern corner and northern areas of the country (rainfall maps in Fig. 8, scatterplot in Fig. 14-left, and bias map in Fig. 15-left). The mean absolute error was 36 mm and the bias +6 mm, which logically indicates that the overestimation was cancelled out, to a large extent, by underestimation, as the bias is indeed small and positive. The scatterplot in Fig. 14-left shows a group of some grid points overestimating largely (those well above the bisector line), while a group of many more points stands below the bisector underestimating, thus compensating the big overestimation of the rest of the points.

When analyzing the columnar water vapor and columnar cloud liquid water fields both predicted by MM5 and measured by satellite-borne AMSR-E sensor, as well as the predicted and observed surface-wind fields, it seems that the factor responsible for the model inaccuracies is most probably the cloud liquid water columnar content (Derbyshire et al., 2004). In Fig. 9 we can see the evolution during this episode of the atmospheric water vapor content in the Western Mediterranean as measured by AMSR-E sensor (25 km resolution), and in Fig. 10 as predicted by MM5. It can be seen that both evolutions of columnar water vapour are of the same order of magnitude and are well placed in space and time. Figures 11 and 12 show, respectively, also the corresponding observed and predicted evolutions for the cloud liquid water columnar content. MM5 always produced more scattered, more localized and less extensive columnar cloud liquid water fields than what satellite measurements testify.

If we now look at the spatially-averaged values of columnar water vapor and columnar cloud liquid water contents over a significant sea area surrounding Catalonia (Fig. 13), we see that MM5 did a good job in forecasting the former, while it did not do so with the latter (MM5 heavily underestimated this variable by a factor of up to 5). This is probably due to inaccurate representation of mixed-phase clouds in cloud parameterizations (Curry et al., 2000).

No outstanding difference has been found when checking the predicted and observed superficial wind fields, thus kinematically, the model performed well and the advection of variables was represented correctly, at least at the surface (no figures shown).

Concluding this case study, we have seen how the columnar content of cloud liquid water is the variable most likely to have caused MM5's QPF failure. However, the actual state of this (and also of water vapor) columnar quantity over land remains unknown for this episode, knowing which would greatly help in the process of verifying the model and understanding why its QPF had flaws.
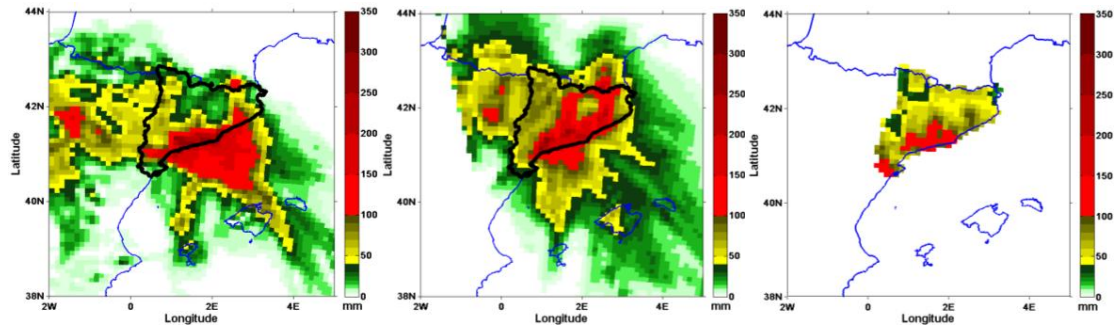
Fig. 8. MM5's QPF (left), COSMO-I7's (centre) and observed rainfall (right) for the whole 9–10 May 2008 event (case study 1). For clarity, the observations domain limit has been highlighted (black line) in the models' prediction maps.
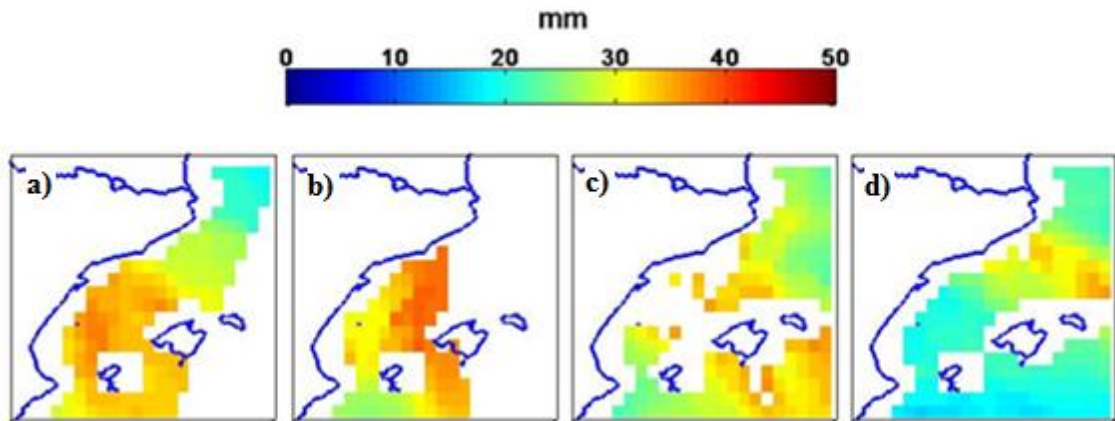
Fig. 9. Columnar water vapor measured by AMSR-E during the episode, each image corresponding to passes at: **(a)** 02:24 UTC 9 May, **(b)** 13:36 UTC 9 May, **(c)** 01:30 UTC 10 May, and **(d)** 12:42UTC 10 May (source: http://www.remss.com/).

In the other case studies analyzed, it was also found that the cloud liquid water columnar content was the most poorly represented variable by MM5, and to a lower degree, water vapor columnar content and surface-wind fields.

5 MM5/COSMO-I7 model comparison

The results of the comparison between both LAM models' QPFs are presented here for the five case studies analysed. Table 3 shows how MM5 and COSMO-I7 models performed similarly, on the whole, when forecasting the rainfall fields.

From the verification indices point of view, MM5 did reasonably better than COSMO-I7 in CS1, while in CS4 it happened the other way around. For CS5 both models did rather well, and for CS2 and CS3 they performed equally poorly, basically missing these episodes (Table 3). However, when comparing both models' performances by looking at their scatterplots individually, we achieve the findings that follow.

It is realized that MM5 did not perform as well as the bias and MAE indices suggest for case study 1, as the point spread

appears bigger than that of COSMO-I7, although this model widely overestimated precipitation (points above bisector), while MM5 did a better job in this sense as corroborated by the bulk of points near – both above and below – the bisector line (Fig. 14). CS2's scatterplots show that COSMO-I7 was slightly smoother than MM5 in providing a good forecast and got several grid points' rain depths fairly well, although rainfall was still missed by the model in the majority of points; MM5 simply missed this event in all grid points. In CS3 the two models did equally poorly, basically missing the precipitation everywhere or just producing negligible values. In CS4 both overestimate rainfall, although MM5 does it with much higher values than COSMO-I7 (thereby the worse verification indices). CS5 scatterplots (Fig. 17) demonstrate the very good skill of COSMO-I7 in predicting this episode's rainfall field, while MM5 did not perform badly either.

Figure 8-centre gives a further hint on the way COSMO-I7 performed in CS1, and indeed not largely differently from MM5 (Fig. 8-left). Figure 15 displays both models' grid point biases for the whole CS1. Both of them underestimate severely in the southwest corner of Catalonia, while
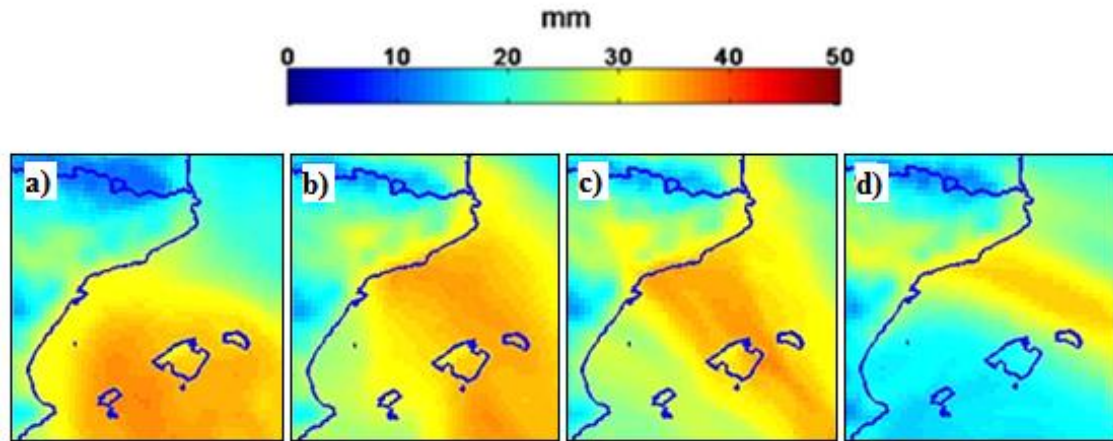
Fig. 10. Columnar water vapor predicted by MM5 during the episode, where each image corresponds to the nearest hourly model output to the satellite passes: **(a)** 02:00 UTC 9 May, **(b)** 14:00 UTC 9 May, **(c)** 01:00 UTC 10 May, and **(d)** 13:00 UTC 10 May.
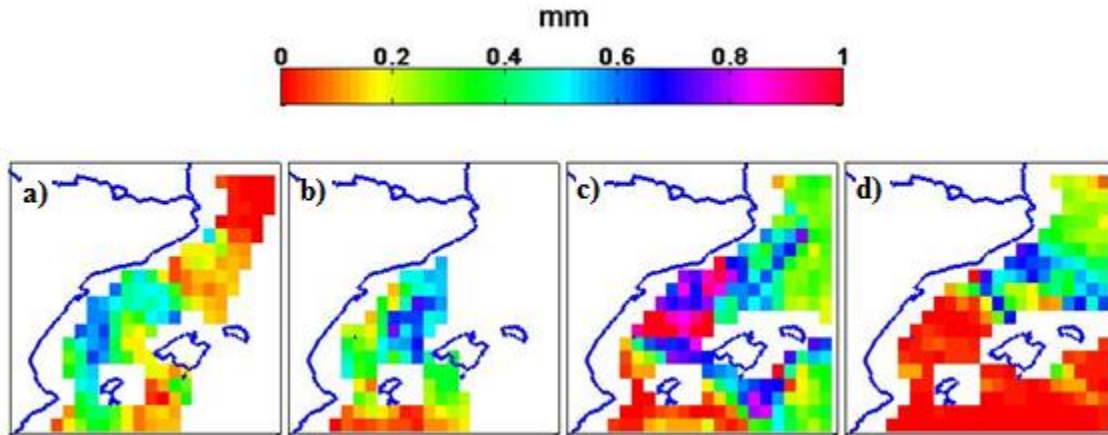
Fig. 11. Columnar cloud liquid water measured by AMSR-E during the episode, each image corresponding to passes at: **(a)** 02:24 UTC 9 May, **(b)** 13:36 UTC 9 May, **(c)** 01:30 UTC 10 May, and **(d)** 12:42 UTC 10 May (source: http://www.remss.com/).

overestimating heavily at the central and southern coasts. Nevertheless, while COSMO-I7 basically overestimates everywhere in the West and North of the territory, MM5 manages it rather well in such places, although it underestimates in central and some northern areas of Catalonia.

Case study 5 (06:00 UTC 28 October to 18:00 UTC 29 October) was characterized by the skillful performances of both models. Figure 16 shows both models' QPF maps against the actual observation field; it is visible how COSMO-I7 gets a better general picture of the precipitation observed during this episode than MM5 does. Figure 17 displays the scatterplots of both models' rainfall predictions against observations. Further from the good verification indices, the grid points in COSMO-I7's scatterplot follow the bisector line more closely than those in MM5's. Figure 18 helps to confirm the better skill of COSMO-I7 in this case study; MM5's

rainfall bias field shows several areas of over- and underestimation around the country, whereas COSMO-I7's is more homogeneous with areas of near-zero biases and it exhibits fewer areas of over- or underestimation.

6 Results of the probabilistic regional-scale analysis

For all 37 episodes, each with its given probability range (>70 %, between 30 % and 70 %, or < 30 %), an analysis was carried out taking into account the multiple 6-h intervals into which all episodes are split in the SMC warning bulletins. As described thoroughly in the methodology section, the procedure consisted in looking for rain gauges reaching or overcoming the warning threshold for hourly accumulations within the warned comarcas and their neighbors, and during the events' time windows. If that happened
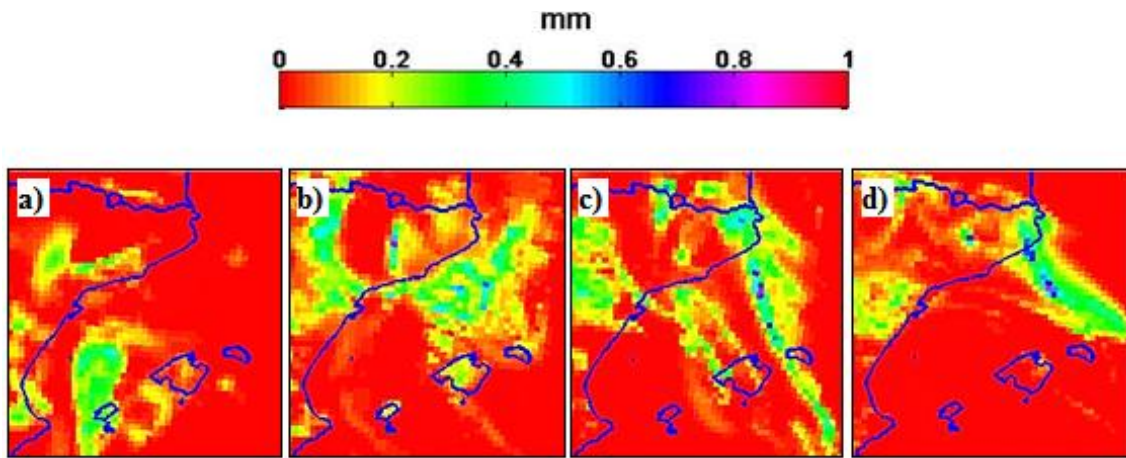
Fig. 12. Columnar cloud liquid water predicted by MM5 during the episode, where each image corresponds to the nearest hourly model output to the satellite passes: **(a)** 02:00 UTC 9 May, **(b)** 14:00 UTC 9 May, **(c)** 01:00 UTC 10 May, and **(d)** 13:00 UTC 10 May.
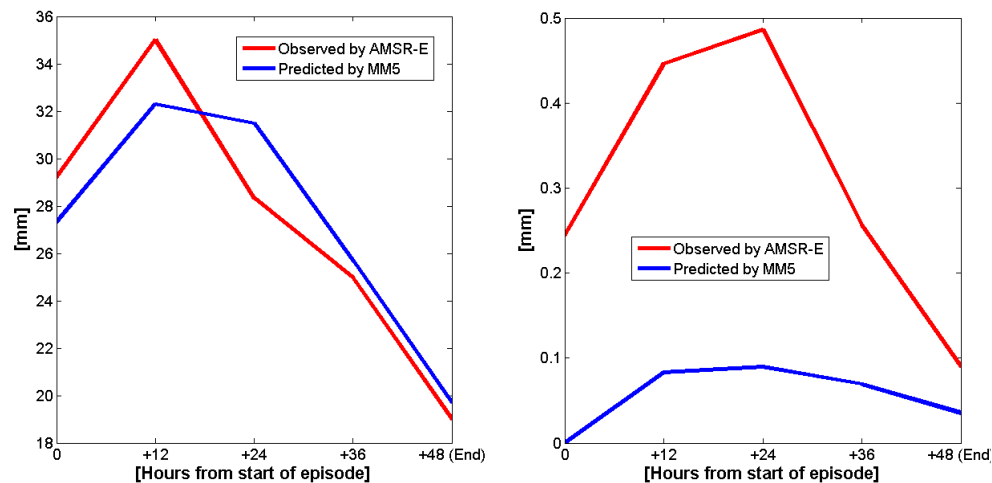
Fig. 13. Evolution of observed and predicted average values over the sea for columnar water vapor (left) and columnar cloud liquid water (right) during the episode.

in at least one rain gauge, the "poor-man" probabilistic forecast was considered "hit" or validated for that specific 6-h interval.

Of all warnings for rain intensity (threshold of $20\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$, i.e., $\sim 25\,\mathrm{mm}\,\mathrm{h}^{-1}$ as found by Llasat, 2001), it was found that 67 % of the events proved correctly forecasted when the given probability was up to >70 %; 23 % proved correctly forecasted when the given probability was 30 %–70 %; and only 8 % proved correctly forecasted when the given probability was <30 %.

As far as warnings for rainfall accumulation (threshold of $100\,\mathrm{mm}\,\mathrm{day}^{-1}$) are concerned, the dataset of only one year seems too limited to draw meaningful conclusions from this simple statistical analysis: none of the events was given the maximum level of probability; 33 % (1 out of 3) of them proved correctly forecasted when the given probability was

between 30 % and 70 %; and none (0 out of 1) proved correctly forecasted when the given probability was <30 %.

Therefore, it can be said that when the warning events were due to rain intensity, the probability given by SMC of the threshold being overcome or not matched reasonably well with the percentage of observed events overcoming the threshold over the total number of events with the same given level of probability; thereby it was proven that SMC had been rather successful when issuing rain intensity warnings during 2008. And although just very few warnings were due to rainfall accumulation, the given probability was also shown to agree with the percentage of observed events "hit" over the total set of this kind.
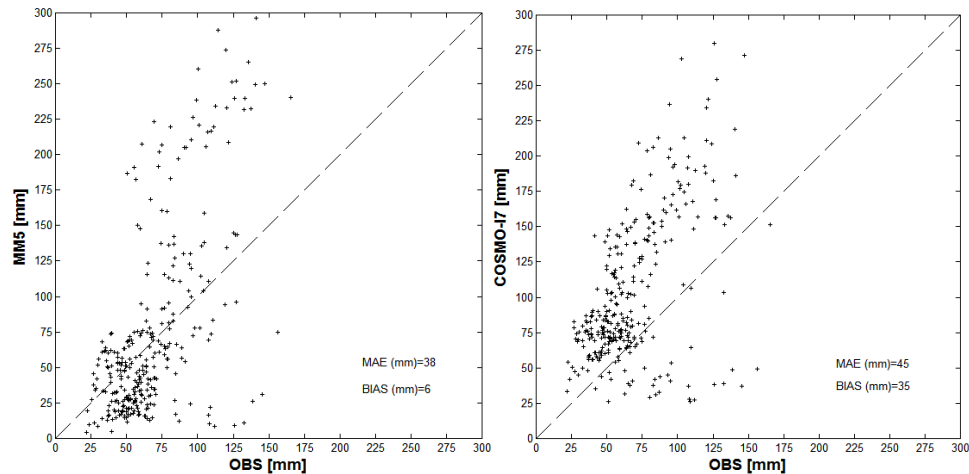
Fig. 14. Case study 1 scatterplots of MM5's (left) and COSMO-I7's (right) QPFs against observed rainfall, showing both models' overestimations and underestimations.
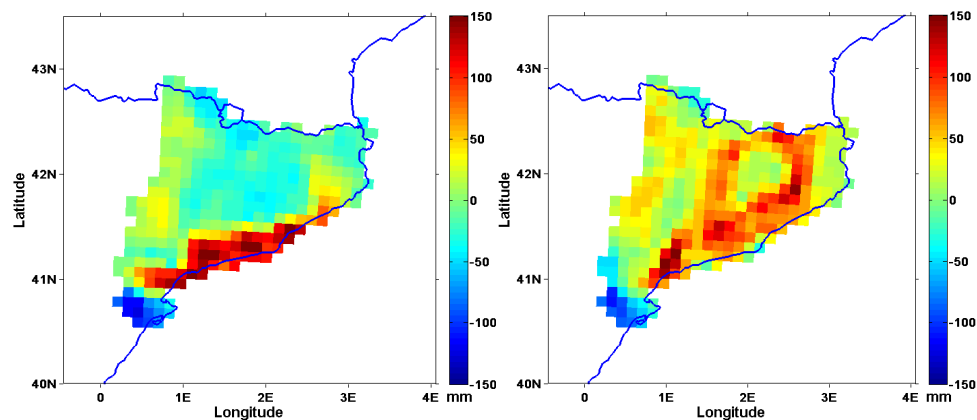

Fig. 15. Rainfall prediction bias fields over Catalonia for MM5's (left) and COSMO-I7's (right) QPFs for case study 1.

If we now group the warning events into the respective episodes in which they were embedded, we find that a total of 67 % of long episodes contained at least one "hit" event, while only 20 % of short episodes did so. For our results, this fact means that episodes longer than 24 h carried a higher probability of 'seeing' intense-rain events at some point during their duration in comparison to this probability for short episodes (those lasting 24 h or less), which was clearly much lower.

This is somewhat surprising, since long episodes tend to cause large rainfall accumulations, but are not usually associated with intense-rain events; contrarily, short episodes, being typically those of convective nature, are characterized specifically by presenting high rain intensities during their short duration (Done et al., 2006; Molini et al., 2010). However, and judging by the RPFs performance, this was not the

case for the 2008 dataset, since long episodes had a high hitting rate for rain intensity forecasts while short episodes had a modest hitting rate for this type of forecasts. Nevertheless, we should not be alarmed by this fact, as it is well known that convective phenomena are intrinsically difficult to forecast accurately, and we already saw in the last section how the model MM5, on average, underestimated observed precipitation fields for short-episode warnings (48 % of short episodes were underestimated, exactly). Taking these facts into account, then, it becomes clear that SMC did rather well (probably even better than MM5's forecast by itself) in issuing the probabilistic forecast warnings for intense-rain events.

Furthermore, another factor should be included when verifying probabilistic forecasts against observations: the spatial distribution of the rain gauges over the territory. Figure 2 (Sect. 3) displays the location of the 146 rain gauges over
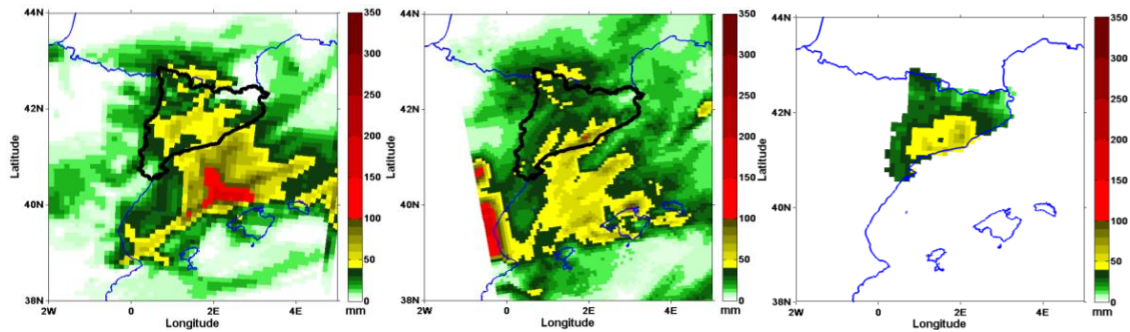
Fig. 16. MM5's QPF (left), COSMO-I7's (centre) and observed rainfall (right) for the 28–29 October 2008 event (case study 5). For clarity, the observations domain limit has been highlighted (black line) in the models' prediction maps.
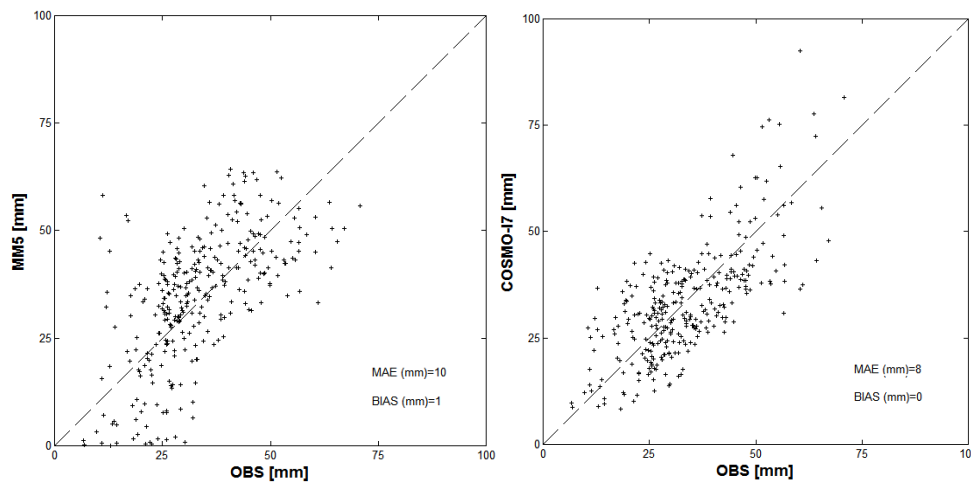
Fig. 17. Case study 5 scatterplots of MM5's (left) and COSMO-I7's (right) QPFs against observed rainfall, both exhibiting good skill.

Catalonia, which clarifies that their density is not homogeneous throughout the country; some areas have a high rain gauge concentration while in others their presence is much more sparse. This fact represents, of course, an extra drawback when it comes to detecting precipitation in the areas with less coverage, especially because they are mountainous areas and hence more prone to intense convective activity. Furthermore, it can also be problematic to properly retrieve precipitation in such areas using radar data due to substantial rainfall underestimation at long range, as pointed out by Trapero et al., 2009. As a result, if the average rain gauge density is roughly one rain gauge per area of $200 \, \text{km}^2$ –which is a scale somewhat bigger than that of typical convective activity in our latitudes–, and moreover we add those poorly-covered mountainous regions to the overall picture of the observation network over Catalonia, we can objectively conclude that some events will be completely missed by our observations, while many others will just be detected partially and not at their highest intensity.

Accepting this fact, we can better understand why it is meaningful that *only* the 20 % of short-episode warnings could be successfully verified from the point of view of the intense-rain threshold. And we should value appropriately all the results drawn before in this section considering this actual constraint that affects the episodes' detection and their measurement.

7 Conclusions

The present study has intended to put some light on the topic of predictive ability, both for the deterministic and the (simplified, "poor-man") probabilistic approaches, for severe rainfall events over Catalonia. In order to do that, MM5's rainfall predictions have been verified against hourly precipitation data provided by the rain gauge network (XEMA) of Catalonia for the episodes corresponding to intense-rain and rainfall accumulation warnings issued by SMC in 2008. Five case studies have examined the model's uncertainty by making use of satellite data, and a QPF comparison has been
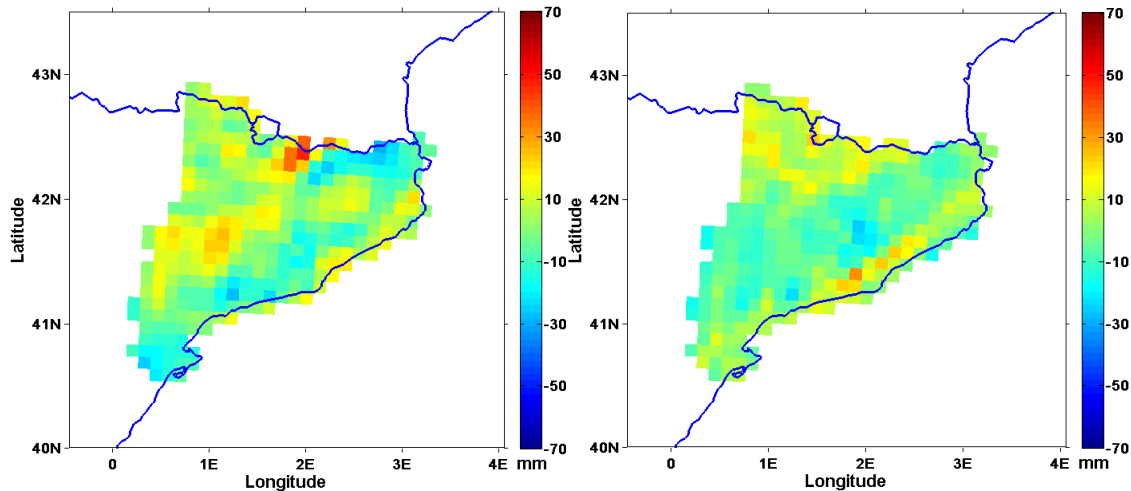
Fig. 18. Rainfall prediction bias fields over Catalonia for MM5's (left) and COSMO-I7's (right) QPFs for case study 5.

undertaken between the MM5 and COSMO-I7 models for these cases. SMC's simplified probabilistic forecasts at regional level have also been tested against hourly rainfall observations.

The results from the verification analysis of MM5's QPFs against observations show that, for long events (>24 h), MM5 tends to overestimate the total precipitation observed, whereas for short events (≤ 24 h) the model tends to underestimate it. For the total 12 episodes classified as long, the average MAE gives $13 \div 14$ mm and the average bias $+0.8$ mm, which is, as a percentage of MAE, about 6 %, which is quite a good result. However, since there are more episodes negatively biased than positively biased, it is concluded that episodes with a positive bias were indeed much superior in magnitude to those with a negative bias. For the total 25 episodes classified as short, the average MAE is in the order of 5 mm and the average bias about -2 mm, or, as percentage of MAE, ~ 40 %, which is a rather bad result. For this type of episodes, those with a negative bias are a majority, while a minority is either positively biased or not biased at all. Furthermore, it can be said, on one hand, that a large number of long episodes have been either correctly forecasted by MM5 or at least correct to some extent, and that just very few of them were false alarms or missed episodes. On the other hand, short episodes have been associated to a very low "hitting" rate, although most of the rainfall predictions were partially correct; this type of episodes has shown a rather high index of false alarms and missed events. Altogether, these facts suggest that further research is needed to continue improving NWP models so they can achieve a more skillful representation of small-scale convective processes.

The case studies analysed in this work conclude, on the whole, that most of MM5's QPF errors are most likely triggered by a poor representation of some of its microphysical species (cloud liquid water and, to a lesser degree, water va-

por) along with errors in low-level wind field prediction. As has been shown, these factors were of crucial importance in some cases because they caused severe model overestimation and underestimation of the observed rainfall fields. However, water vapor is usually much better predicted than cloud liquid water, presumably because the model often failed to generate realistic cloud fields as a result of imbalance between the cloud microphysics scheme and other physical processes, like radiation (Hong et al., 2004). Actually, some of these variables' errors may be associated simply with the model's data assimilation inaccuracies, which would add some further uncertainty to the whole prediction process.

A comparison of the models' performance has been done for the five case studies between MM5 –12 km horizontal resolution– and COSMO-I7 –7 km– models. It was demonstrated that both of them are able to predict intense rainfall to a similar extent of skill for the cases analysed. This may be explained by the fact that while Catalonia is in the centre of MM5's domain but only on the western fringe of COSMO-I7's, the latter model has a higher horizontal resolution than the former. As a matter of fact, thus, a similar degree of QPF skill is demonstrated between the two LAM models.

The rainfall intensity/depth warnings based on "poorman" probabilistic forecasts and issued by SMC have proven fairly successful when tested against hourly observed precipitation at comarca scale. From the viewpoint of the rain gauges' observations, most of the warnings for long episodes have hit at least one intense-rain event, while only a minority of warnings for short episodes has been able to do so. This may be due to a variety of factors, such as the intrinsic difficulty in the forecast of small-scale convection (as we have seen, not even MM5's QPFs give good enough results for short episodes), and the observational "loss" of some events resulting from the sparsity of the rain gauges network, particularly in the Pyrenees.

References

Adlerman, E. J. and Droegemeier, K. K.: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters, Mon. Weather Rev., 130, 2671–2691, 2002.

Altava-Ortiz, V., Barrera-Escoda, A., and Llasat, M. C.: Monitoratge, classificació i reconstrucció de registres de sequera a Catalunya, in: Primeres Jornades de Meteorologia i Climatologia de la Mediterrània Occidental, Barcelona, Spain, 28 November 2008.

Amengual, A., Romero, R., Gómez, M., Martín, A., and Alonso, S.: A hydrometeorological modeling Study of a flash-flood event over Catalonia, Spain, J. Hydrometeorol., 8, 282–303, 2007.

Bryan, G. H., Wyngaard, J. C., and Fritsch, J. M.: Resolution requirements for the simulation of deep moist convection, Mon. Weather Rev., 131, 2394–2416, 2003.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., and Mason, S.: Forecast verification: current status and future directions, Meteorol. Appl., 15–1, 3–18, 2008.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR, MM5 modeling system. Part I: Model implementation and sensitivity, Mon. Weather Rev., 129, 569–585, 2001.

Curry, J. A., Hobbs, P. V., King, M. D., Randall, D. A., Minnis, P., Isaac, G. A., Pinto, J. O., Uttal, T., Bucholtz, A., Cripe, D. G., Gerber, H., Fairall, C. W., Garrett, T. J., Hudson, J., Intrieri, J. M., Jakob, C., Jensen, T., Lawson, P., Marcotte, D., Nguyen, L., Pilewskie, P., Rangno, A., Rogers, D. C., Strawbridge, K. B., Valero, F. P. J., Williams, A. G., Wylie, D.: FIRE Arctic clouds experiment, B. Am. Meteorol. Soc., 81, 5–30, 2000.

del Carmen Llasat, M. and Siccardi, F.: A reflection about the social and technological aspects in flood risk management – the case of the Italian Civil Protection, Nat. Hazards Earth Syst. Sci., 10, 109–119, doi:10.5194/nhess-10-109-2010, 2010.

Derbyshire, S. H., Beau, I., Bechtold, P., Grandpeix, J. Y., Piriou, J. M., Redelsperger, J. L., and Soares, P. M. M.: Sensitivity of moist convection to environmental humidity, Q. J. Roy. Meteor. Soc., 130, 3055–3079, 2004.

Doms, G. and Forstner, J.: Development of a kilometer-scale NWP-system: LMK., COSMO Newsletter, 4, 159–167, 2004.

Done, J. M., Craig G. C., Gray, S. L., Clark, P. A., and Gray, M. E. B.: Mesoscale simulations of organized convection: Importance of convective equilibrium, Q. J. Roy. Meteor. Soc., 132, 737–756, 2006.

Doswell, C. A. III., Brooks, H. E., and Maddox, R. A.: Flash flood forecasting: An ingredients-based methodology, Weather Forecast., 11, 560–581, 1996.

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, J. Atmos. Sci., 46, 3077–3107, 1989.

Ebert, E. E.: Ability of a poor man's ensemble to predict the probability and distribution of precipitation, Mon. Weather Rev., 129, 2461–2480, 2001.

Ebert, E. E.: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework, Meteorol. Appl., 15, 51–64, 2008.

Ferraris, L., Rudari, R., and Siccardi, F.: The uncertainty in the prediction of flash floods in the northern Mediterranean environment, J. Hydrometeorol., 3, 714–727, 2002.

Golding, B. W.: Quantitative precipitation forecasting in the UK, J. Hydrol., 239, 286–305, 2000.

Grell, G. A.: Prognostic evaluation of assumptions used by cumulus parameterizations, Mon. Weather Rev., 121, 764–787, 1993.

Harper, K., Uccellini, L. W., Morone, L., Kalnay, E., and Carey, K.: 50th anniversary of operational numerical weather prediction, B. Am. Meteorol. Soc., 88, 639–650, 2007.

Hong, S., Dudhia, J., and Chen, S.: A revised approach to ice microphysical processes for the bulk parametrization of clouds and precipitation, Mon. Weather Rev., 132, 103–120, 2004.

Hong, S. and Pan, H.: Nonlocal boundary layer vertical diffusion in a medium range forecast model, Mon. Weather Rev., 124, 2322–2339, 1996.

Jacobsen, I. and Heise, E.: A new economic method for the computation of the surface temperature in numerical models, Contr. Atmos. Phys., 55, 128–141, 1982.

Kain, J. S., Weiss, S. J., Bright, D. R., Baldwin, M. E., Levit, J. J., Carbin, G. W., Schwartz, C. S., Weisman, M. L., Droegemeier, K. K., Weber, D. B., and Thomas, K. W.: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP, Weather Forecast., 23, 931–952, 2008.

Levy, J. K. and Hall, J.: Advances in ?ood risk management under uncertainty, Stoch. Env. Res. Risk A., 19, 375–377, 2005.

Llasat, M. C.: Heavy Rain Events in Catalonia: Genesis, Evolution and Mechanism (in Spanish), Ph.D. thesis, Department of Astronomy and Meteorology, University of Barcelona, Spain, 1987.

Llasat, M. C.: An objective classi?cation of rainfall events on the basis of their convective features: application to rainfall intensity in the northeast of Spain, Int. J. Climatol., 21, 1385–1400, 2001.

Llasat, M. C. and Puigcerver, M.: Meteorological factors associated with floods in the north-eastern part of the Iberian Peninsula, Nat. Hazards, 9, 81–93, 1994.

Llasat, M. C. and Puigcerver, M.: Total rainfall and convective rainfall in Catalonia, Spain, Int. J. Climatol., 17, 1683–1695, 1997.

Mellor, G. L, and Yamada, T.: A hierarchy of turbulence closure models for planetary boundary layers, J. Atmos. Sci., 31, 1791–1806, 1974.

Mirza, C. R., Koike, T., Yang, K., and Graf. T.: Retrieval of atmospheric integrated water vapor and cloud liquid water content

over the ocean from satellite data using the 1-D-Var ice cloud microphysics data assimilation system (IMDAS), IEEE T. Geosci. Remote, 46 (1), 119–129, 2008.

Molini, L., Parodi, A., Rebora, N., and Craig, G. C.: Classifying severe rainfall events over Italy by hydrometeorological and dynamical criteria, Q. J. Roy. Meteor. Soc., 137, 148–154, 2010.

Molini, L., Parodi, A., and Siccardi, F.: Dealing with uncertainty: an analysis of the severe weather events over Italy in 2006, Nat. Hazards Earth Syst. Sci., 9, 1775–1786, doi:10.5194/nhess-9-1775-2009, 2009.

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., and Anquetin, S.: Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, J. Hydrol., 362, 274–290, 2008.

Pappenberger, F., Scipal, K., and Buizza, R.: Hydrological aspects of meteorological verification, Atmos. Sci. Lett., 9, 43–52, 2008.

President of the Council of Ministers Directive, 2/27/2004, Italy Official Gazette, Rome, 2004.

Rezacova, D., Zacharov, P., and Sokol, Z.: Uncertainty in the area-related QPF for heavy convective precipitation, Atmos. Res., 93, 238–246, 2009.

Rigo, T. and Llasat, M. C.: Analysis of mesoscale convective systems in Catalonia (NE of Spain) using radar for the period 1996–2000, Atmos. Res., 83, 458–472, 2007.

Ritter, B. and Geleyn, J. F.: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, Mon. Weather Rev., 120, 303–325, 1992.

Roberts, N. M.: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model, Meteorol. Appl., 15–1, 163–169, 2008.

Roberts, N. M. and Lean H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Mon. Weather Rev., 136, 78–97, 2008.

Rotach, M. W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H. S., Behrendt, A., Bouttier, F., Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M. A., Marsigli, C., McTaggart-Cowan, R., Montani, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos-Muñoz, D., Schär, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V., and Zappa, M.: MAP D-PHASE: Real-time demonstration of weather forecast quality in the Alpine region, B. Am. Meteorol. Soc., 90, 1321–1336, 2009.

Sairouni, A., Moré, J., Toda, J., Miró, J. R., Aran, M., and Cunillera, J.: Verificació dels models de mesoscala operatius al Servei Meteorològic de Catalunya (Verification of the operational NWP mesoscale models at the Meteorological Service of Catalonia), Nota d'Estudi de l'SMC num. 71, Dept. de Medi Ambient i Habitatge de la Generalitat de Catalunya, Barcelona, Spain, 78 pp., 2007.

Schultz, P.: An explicit cloud physics parameterization for operational numerical weather prediction, Mon. Weather Rev., 123, 3331–3343, 1995.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., Mitrescu, C., and the Cloudsat Science Team: The CloudSat mission and the A-train: A new dimension of space-based observations of clouds and precipitation, B. Am. Meteorol. Soc., 83, 1771–1790, 2002.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Weather Rev., 117, 1779–1800, 1989.

Trapero, L., Bech, J., Rigo, T., Pineda, N., and Forcadell, D.: Uncertainty of precipitation estimates in convective events by the Meteorological Service of Catalonia radar network, Atmos. Res., 93, 408–418, 2009.

Zhang, F., Snyder, C., and Rotunno, R.: Effects of moist convection on mesoscale predictability, J. Atmos. Sci., 60, 1173–1185, 2003.