QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **134**: 1221–1242 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.291

Hydrometeorological ensemble simulations of flood events over a small basin of Majorca Island, Spain

A. Amengual,* R. Romero and S. Alonso

Grup de Meteorologia, Departament de Física, Universitat de les Illes Balears, Palma de Mallorca, Spain

ABSTRACT: A hydrometeorological modelling study is designed in order to assess the feasibility of high-resolution mesoscale-model-driven runoff simulations for a small basin of Majorca in the Balearic Islands. Four intense precipitation events, which caused flood events of different magnitudes over the Albufera basin (with a drainage area of 610 km^2), are analysed. The Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) runoff model is used to generate the hydrological simulations. The lack of flow measurements in the basin poses great difficulties in the evaluation of the rain-gauge-driven runoff simulations. Therefore the runoff model is run under the assumption that a best estimation of the hydrological model parameters, mainly related to the infiltration properties of the watershed, can be obtained from the high-resolution observational campaign developed by the Coordination of Information on the Environment (CORINE) Land Cover project. The non-hydrostatic fifth-generation Pennsylvania State University / NCAR Mesoscale Numerical Model (MM5) is used to provide quantitative precipitation forecasts for the events. The MM5-driven runoff simulations are compared against stream-flow simulations driven by the rainfall observations, thus employing the hydrological model as a validation tool. In addition to the control MM5 simulations, a multi-physics ensemble is carried out: various combinations of the physical parametrizations of the MM5 model (cloud microphysics, moist convection and boundary-layer schemes) are adopted, in order to better encompass the atmospheric processes leading to the high precipitation amounts. Results show that high-resolution numerical weather experiments in this area of complex orography accurately reproduce most of the extreme precipitation events under study, enabling potentially valuable discharge simulations, despite the small size of the basin. The value of the multi-physical model ensemble in conveying the uncertainty of precipitation, and therefore the discharge experiments, is also discussed. Copyright © 2008 Royal Meteorological Society

KEY WORDS small-sized basin; floods; hydrometeorological modelling; ensemble strategy

Received 26 March 2007; Revised 16 April 2008; Accepted 18 June 2008

1. Introduction

The Spanish Mediterranean area is affected virtually every year by flood events of diverse spatial and temporal scales, and most of them – mainly in autumn – can be classified as flash floods. The complex orography of the region (Figure 1) – formed by mountainous systems near the coast – acts as a natural barrier to the warm, moist Mediterranean air. This fact, together with the early-autumn intrusion of cold fronts, can lead to the generation of hazardous mesoscale convective systems (MCSs), resulting in heavy rainfall (see, for example, Romero *et al*. (2000) and references therein).

Two other important factors must be taken into account: large parts of the coastal areas are densely urbanized, and many of the small and medium-sized steep streams are ephemeral in this semi-arid environment and have short response times. The conjunction of all these circumstances can produce unexpected and extensive flood damage. Therefore it becomes important to study the feasibility of incorporating runoff forecasts before such events.

The short time-scales of flash-flood events, where the associated effects can develop within periods of very few hours after the rainfall, imply that the traditional warning systems based on rainfall observations do not provide the timely predictions required to implement the precautionary civil-protection measures that are expected (Siccardi, 1996). The possibility of increasing the lead times associated with runoff forecasting should be examined. One way in which this can be accomplished is by translating mesoscale-model rainfall forecasts into runoff forecasts (Anderson *et al*., 2002; Ferraris *et al*., 2002). Amengual *et al*. (2007) have applied this methodology to a flash-flood event that affected Catalonia in northeastern Spain (Figure 1), demonstrating the convenience of introducing runoff simulations driven by numerical weather prediction (NWP) into a medium-sized catchment (5040 km^2), characteristic of the Spanish Mediterranean environment. The relatively high predictability of the responsible MCS, together with the size of the basin, resulted in an enhanced predictability for that particular flash flood.

Many studies (e.g. Pessoa *et al*., 1993; Dolciné *et al*., 2001; Giannoni *et al*., 2003; Zhang and Smith, 2003)

* Correspondence to: A. Amengual, Dept. de Física, Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain.
E-mail: arnau.amengual@uib.es

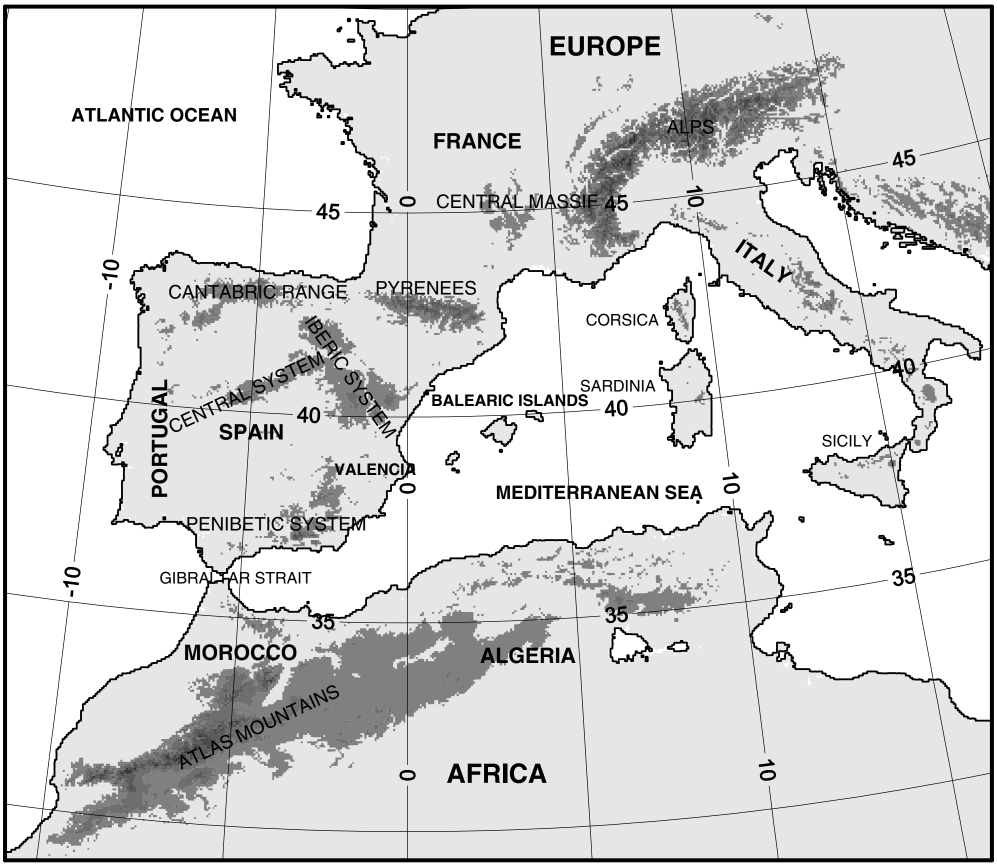

Figure 1. Main mountain systems of the western Mediterranean region. Major topographic features are shown starting at 1000 m.

have analysed the problem of hydrological modelling and forecasting of extreme floods using radar rainfall data for small basins (of the order of hundreds of square kilometres). But similar analyses using high-resolution rainfall simulations or quantitative precipitation forecasts (QPFs) provided by mesoscale numerical models are comparatively uncommon in the literature. This is a challenging task given the small hydrological scales involved, their associated uncertainties and the external-scale errors found in the numerical weather models.

In this context, we study four intense precipitation events that caused damaging flood episodes of different spatial and temporal scales in a catchment of Majorca, in the Balearic Islands (Figure 1). Some of these provide a good example of the hazardous effects of a rapid and sudden flow increase over a short period of time. For these events, we examine the feasibility of hydrometeorological-forecasting-model strategies, under a best guess of the large-scale atmospheric circulation. Specifically, the study is centred over the Albufera river basin, with a drainage area of 610 km^2. The first objective of this paper is to test the appropriateness of the aforementioned atmospherically-driven runoff simulations for that small basin, as an initial step towards extending the lead times associated with runoff forecasting of such extreme events. In particular, the HEC-HMS runoff model is forced with rainfall distributions derived from MM5 mesoscale simulations initialized with meteorological grid analyses from ECMWF and, for one of the cases, also from NCEP.

High-resolution precipitation simulation in extreme events is a remarkably difficult task, because many factors combine in determining such events and most of the precipitation is of a convective nature. NWP models have problems in triggering and organizing convection over the correct locations and times, due to the small-scale nature of many responsible atmospheric features (Kain and Fritsch, 1992; Stensrud and Fritsch, 1994a, 1994b). In the last decade, an important methodology has been implemented in order to improve the short-range QPFs: the use of an ensemble of model forecasts in order to further extend the space of possible outcomes. The aim of ensemble forecasting is to predict the probability of future weather events as completely as possible. This is motivated by the fact that forecasts are sensitive to both uncertainties in the initial and boundary conditions and model errors. Model contributions to these uncertainties are mainly due to the imperfect representation of the atmospheric physical processes in the model (Tribbia and Baumhefner, 1988). This issue is of major importance when the forecast is concerned with the precise locations and amounts of rain at small scales, which are directly affected by the uncertainties in the model parametrization schemes for the convection and moist microphysical processes (Kain and Fritsch, 1992; Wang and Seaman, 1997). It appears that the sensitivity of forecast accuracy to model parametrizations must be addressed for short-range forecasts involving convective events (Stensrud *et al.*, 2000). Examples of the use of a multi-physics ensemble strategy in order to take into account the model imperfections have been widely described (e.g. Stensrud *et al.*, 2000; Jones *et al.*, 2007). In this approach, different model physical parametrization schemes are combined to build varied versions of the mesoscale model and to produce an ensemble of simulations that start from the same initial conditions. An inherent assumption is that all the mesoscale-model configurations are equally skilful. If one of the model configurations is significantly less skilful than the others, then the ensemble members should be weighted unequally to obtain the best results. However, several studies (Wang and Seaman, 1997; Stensrud *et al.*, 2000; Bright and Mullen, 2002; Zhang and Zheng, 2004) have found no substantial differences in model performance using different well-tested physical schemes of the MM5 model in simulating diverse meteorological processes.

Following this methodology, the second objective of this paper is to assess the sensitivity of the small-scale features of the precipitation simulations to the uncertainties in the approximations of the physical parametrizations included in the MM5 mesoscale model. We have adopted an ensemble of MM5 experiments combining different parametrizations of cloud microphysics, moist convection and boundary-layer parametrizations, using large-scale analyses – rather than forecast data – as initial and boundary conditions in order to minimize the synoptic-scale errors. Then, it is expected that the influence on the hydrological response of the Albufera basin exerted by spatial and temporal errors in the simulated rainfall fields emerging from the physical parametrizations can be addressed. Furthermore, to study the impact of the large-scale uncertainties in the mesoscale-model performance, we have forced MM5 by using different initial and boundary conditions for one of the case studies. This test is a first attempt to assess the relative importance of these inaccuracies, when compared with the errors coming from the model formulation, for short-range modelling systems. We will also determine, from the ensemble of MM5-driven HEC-HMS simulations, the value of probabilistic hydrometeorological chains versus deterministic approaches when dealing with flood situations in the area. In addition to using a best guess for the MM5 initial and boundary conditions, we do not consider other sources of uncertainty acting in the problem, such as the hydrological model configuration; we state both assumptions. That is, the rainfall-runoff model will be treated under the perfect-model assumption and 'ideal' knowledge of the synoptic-scale dynamical forcing, in order to focus the study exclusively on the impacts of the NWP-model-formulation errors.

Validation of high-resolution precipitation fields is not straightforward, particularly for extreme events. If rain-gauge networks are not dense enough, they cannot resolve the small-scale features of the highly-variable precipitation fields driving floods. These limitations must be especially considered in this study, since we are dealing with a small basin – partially located in a mountainous area – with only a small number of automatic rain-gauge stations, and where meteorological radar is not available. Furthermore, a pointwise comparison between the observed and simulated rainfall fields is not always appropriate for hydrological purposes. In this study, we examine the performance of the spatial and temporal distributions of the simulated rainfall fields against the observed rainfall patterns at catchment scale by using a set of continuous and categorical verification indices over the sub-basins, which are employed as spatially-integrated surfaces. Furthermore, we compare the discharge experiments resulting from the one-way coupling between the meteorological and hydrological models with the rain-gauge-driven runoff simulations, thus employing the hydrological model as an advanced validation tool. This approach has been found especially suitable for the evaluation of high-resolution simulated precipitation fields (Benoit *et al.*, 2000; Jasper and Kaufmann, 2003; Chancibault *et al.*, 2006).

The paper is structured as follows. Section 2 contains a brief description of the study area; Section 3 describes the selected intense-rainfall episodes; Section 4 describes the meteorological model used to simulate the events and to design the ensemble of mesoscale simulations; Section 5 explains the hydrological tools used for the basin characterization; Section 6 presents and discusses the results; and finally, in Section 7 we assess the methodology used and directions for its further development.

2. The study area

2.1. Overview of the Albufera basin

The Albufera basin is the most important of the hydrographic catchments in the Balearic Islands in terms of size, river length, mean flow and socioeconomic activities. It is located in Majorca, the biggest of the islands, and is composed of the Almedrà and Sant Miquel ephemeral river basins. The Albufera basin extends from the Tramuntana range, with heights close to 1500 m, to the central plain. This central plain constitutes the main agricultural area of Majorca. The last sections of both ephemeral rivers flow into the Albufera's natural park, a wetland located in the northeastern part of the basin, extending over an area of 17 km^2. The main economic activities in the catchment area are tourism (along the basin coastline – this is the leading activity) and agriculture (in wide areas of the central plain) (MEDIS, 2006). Although the Albufera river basin extends over 610 km^2, we have modelled the catchment upstream from the junction of the Almedrà and Sant Miquel rivers in the natural wetland, with a drainage area of 607.4 km^2 and a maximum length of 42.1 km (Figure 2). The Sant Miquel river basin (141.7 km^2) can be classified as mountainous, characterized by steep streams, short concentration

DOI: 10.1002/qj

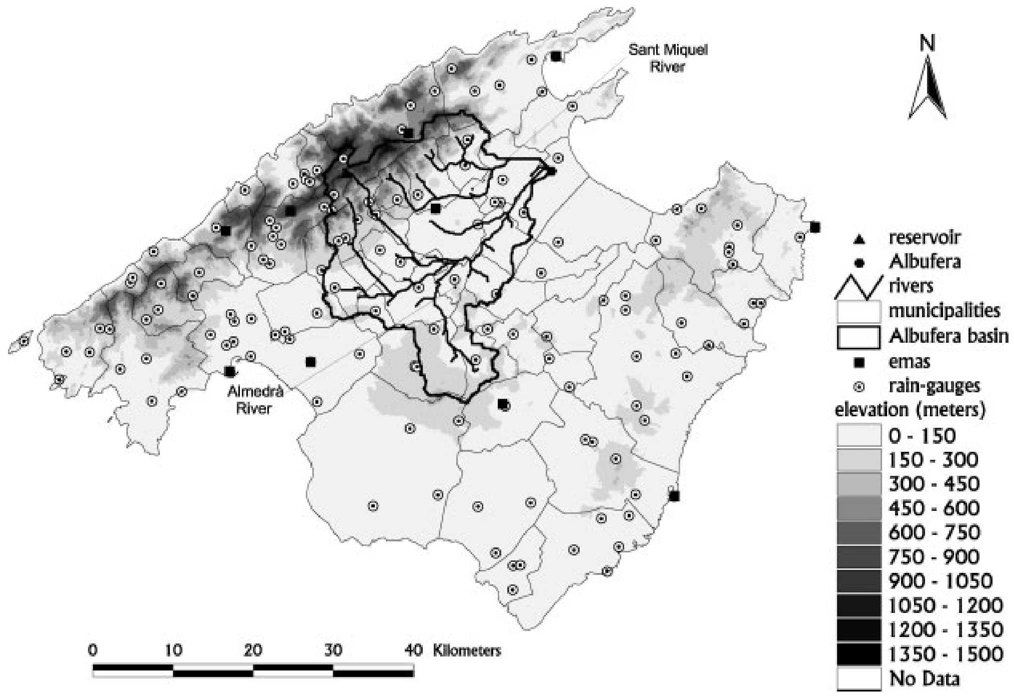

Figure 2. Spatial distribution of the rain-gauge network of the Spanish Institute of Meteorology (INM) in Majorca Island. It includes 100 stations, which provide 24 h accumulated values (circled dots) and 10 automatic stations ('emas') (black squares). The Almedrà and Sant Miquel rivers, which compose the Albufera basin, are highlighted, as well as the basin outlet (black circle) and the reservoir mentioned in the text (black triangle). The digital terrain model of the watershed has a cell size of 50 m.

times and high flow velocities. By contrast, the Almedrà river basin (465.7 km^2), although in part composed of elevated terrain, mainly flows through the central plain of Majorca, an area with moderate slopes and consequently with longer concentration times and lower flow velocities.

The hydrographic catchment is divided into several climatic areas, because of the diversity of the pluviometric records due to the varying altitude. Annual rainfall in the Albufera basin can range from quantities exceeding 1000 mm in the Tramuntana range (over 900 m) and 700 mm over pre-mountainous areas (with elevations of 500–900 m) to about 600 mm in the plain area. The rainfall regime is typical of the Mediterranean region, with most of the heavy-rainfall episodes occurring between September and December, but with occasional events in spring and winter. These extreme daily rainfall episodes can represent a large fraction of the annual amounts.

2.2. The rain-gauge network

The raw precipitation data consist of 24 h (0700–0700 UTC) accumulated values at 140 climatic stations deployed by the Spanish Institute of Meteorology (INM) in the Balearic Islands (see Figure 2 for the locations of the Majorca stations). Of these 140 stations, about 40 lie inside or near the watershed boundaries. Precipitation is also recorded every 10 min in 12 additional automatic rain gauges ('emas' in Figure 2). The automatic stations located inside or very close to the Albufera basin have been used, first, to accumulate the 10 min series into 1 h series, and secondly, to build hourly series for the rest of the INM network. Thus, the temporal frequency of the precipitation data has been increased to permit hydrological applications. To downscale the daily accumulations into hourly values using the automatic stations' 1 h series, we apply the following inverse-distance-weighted equation on each day of the selected flood episodes:

$$p_{\mathrm{st}\,j}(t) = \frac{\sum_i \dfrac{p_{\mathrm{emas}\,i}(t)}{d_{ij}} \dfrac{P_{\mathrm{st}\,j}}{P_{\mathrm{emas}\,i}}}{\sum_i \dfrac{1}{d_{ij}}},$$

where $p_{\mathrm{st}\,j}(t)$ is the derived hourly value at time step t for daily station j, $p_{\mathrm{emas}\,i}(t)$ is the 1 h value at time step t for automatic station i, $P_{\mathrm{st}\,j}$ is the 24 h accumulated value at station j, $P_{\mathrm{emas}\,i}$ is the daily accumulation at automatic station i, and d_{ij} is the distance between daily station j and automatic station i. It is considered a reasonable approximation that the temporal distributions of rainfall at daily stations and neighbouring automatic stations should not differ significantly, because of the typical size of the meteorological disturbances driving large-rainfall events (see Section 3). It is worth noting that the spatial distribution of automatic and daily rain gauges covers the various climatic areas of the basin reasonably well.

3. Description of the intense-precipitation episodes

Romero *et al.* (1999) presented a classification of the atmospheric-circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. That study pointed out the synoptic-scale disturbances bearing important rainfall accumulations over the Balearics. These are associated with general northerly or northeasterly surface winds produced by low-pressure centres to the east of the islands and troughs or cut-off cyclones aloft over the western Mediterranean.

We have selected four intense-precipitation episodes from among the flood events of highest magnitude since the late 1980s, when the automatic stations were installed. All the events are characterized by cold cut-off lows at mid-to-upper-tropospheric levels. The circulation at lower levels over northern Majorca had a general southeast–northwest pressure gradient, which imposed the above-mentioned surface wind regime (Figure 3). The Albufera river basin – especially its upper part – was directly affected by substantial rainfall accumulations (Figure 4).

The four cases (described below) represent a sample of heavy-rainfall episodes resulting in floods of diverse spatial and temporal scales. The first two cases, because

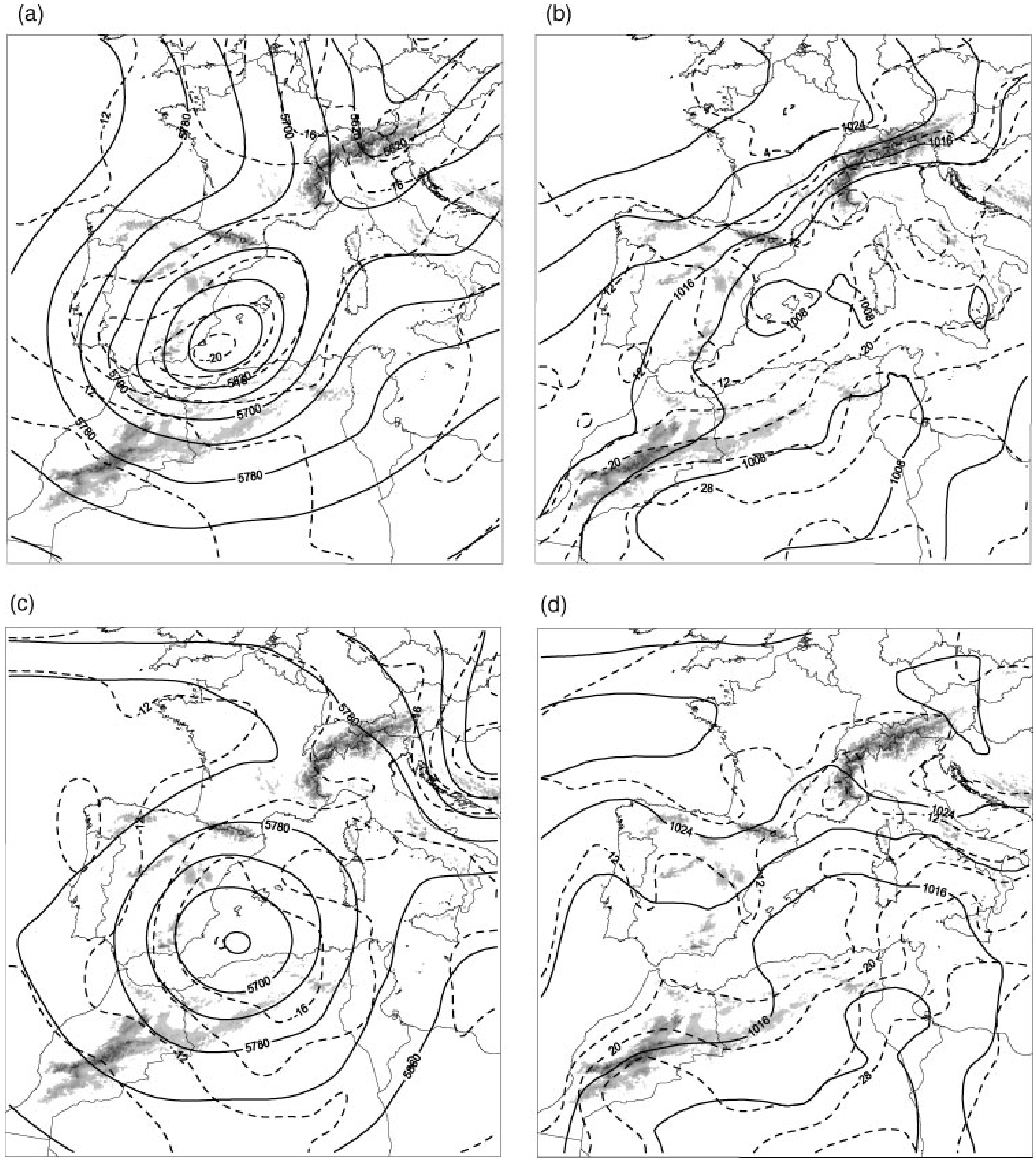

Figure 3. ECMWF analyses for the four cases considered: (a, c, e, g) 500 hPa geopotential height (solid line, in gpm) and 500 hPa temperature (dashed line, in °C); and (b, d, f, h) sea-level pressure (solid line, in hPa) and 925 hPa temperature (dashed line, in °C); for: (a, b) 8 October 1990 at 1200 UTC; (c, d) 9 October 1990 at 1200 UTC; (e, f) 11 November 2001 at 0000 UTC; (g, h) 4 April 2002 at 0000 UTC. The main orographic systems are highlighted.

DOI: 10.1002/qj

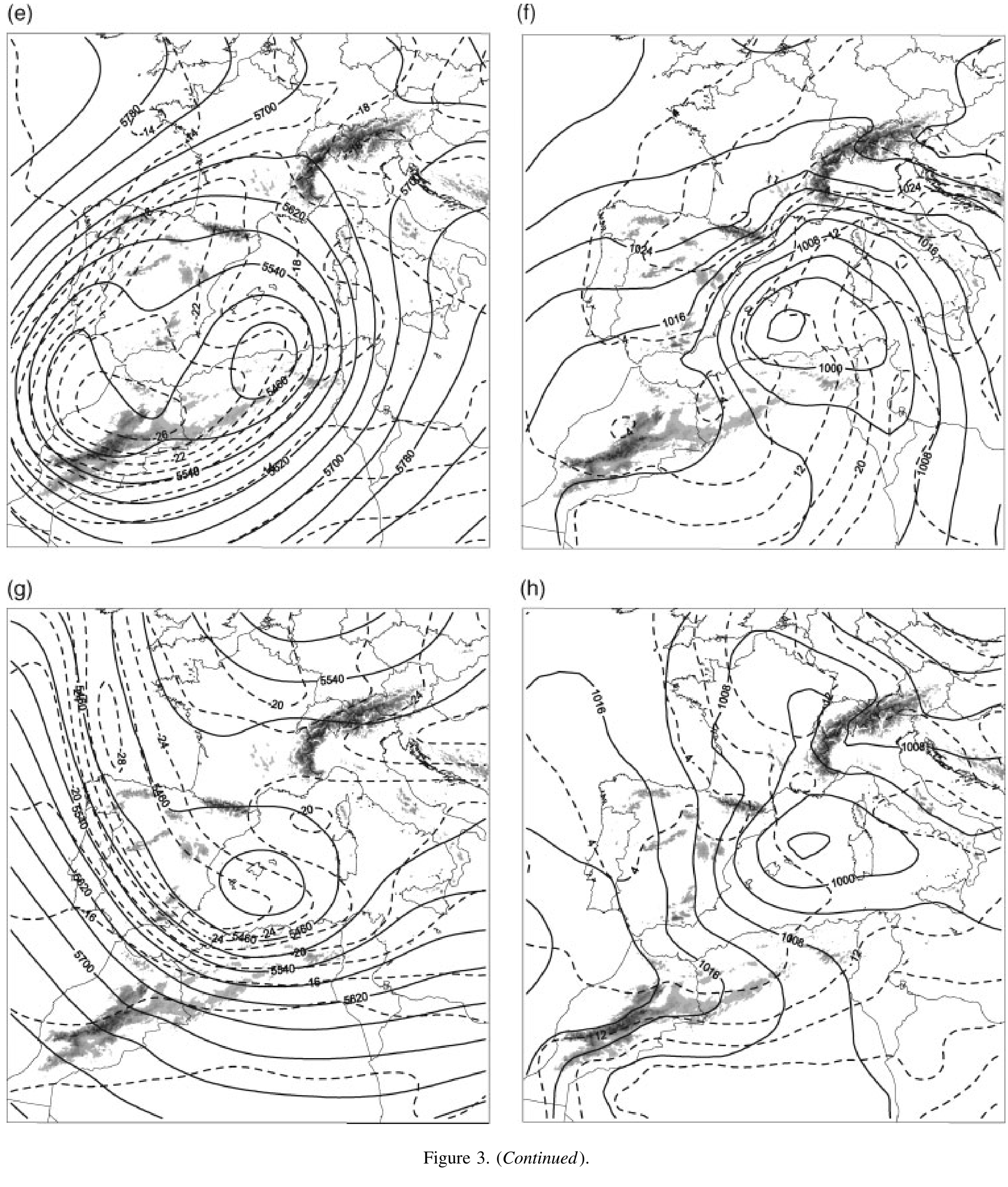

Figure 3. (*Continued*).

of their convective nature, produced exceptional and sudden rising flows; the last two cases produced more sustained, stratiform-like precipitation rates over longer periods, but also led to notable discharges at the Albufera basin outlet.

We will now briefly describe the synoptic situation and derived rainfall distributions for each case.

3.1. Case 1: 7–10 October 1990, first phase

This was a long episode that lasted about 72 h, which for practical reasons we have split into two different phases. The first phase consisted of a mid-to-upper-level cold cut-off cyclone located to the southwest of the Balearics. The associated low-level cyclone provided a northeasterly surface current towards Majorca (Figure 3(a,b)). The high rainfall rates, with accumulated values above 100 mm from 1600 to 1800 UTC on 8 October and total accumulations during the first 48 h over 240 mm (Figure 4(a)), together with the fact that the heavy rainfall fell over the northern part of the basin, resulted in a flash flood over the Sant Miquel catchment and the last sections of both rivers, with hazardous effects for the coastal urbanizations.

DOI: 10.1002/qj

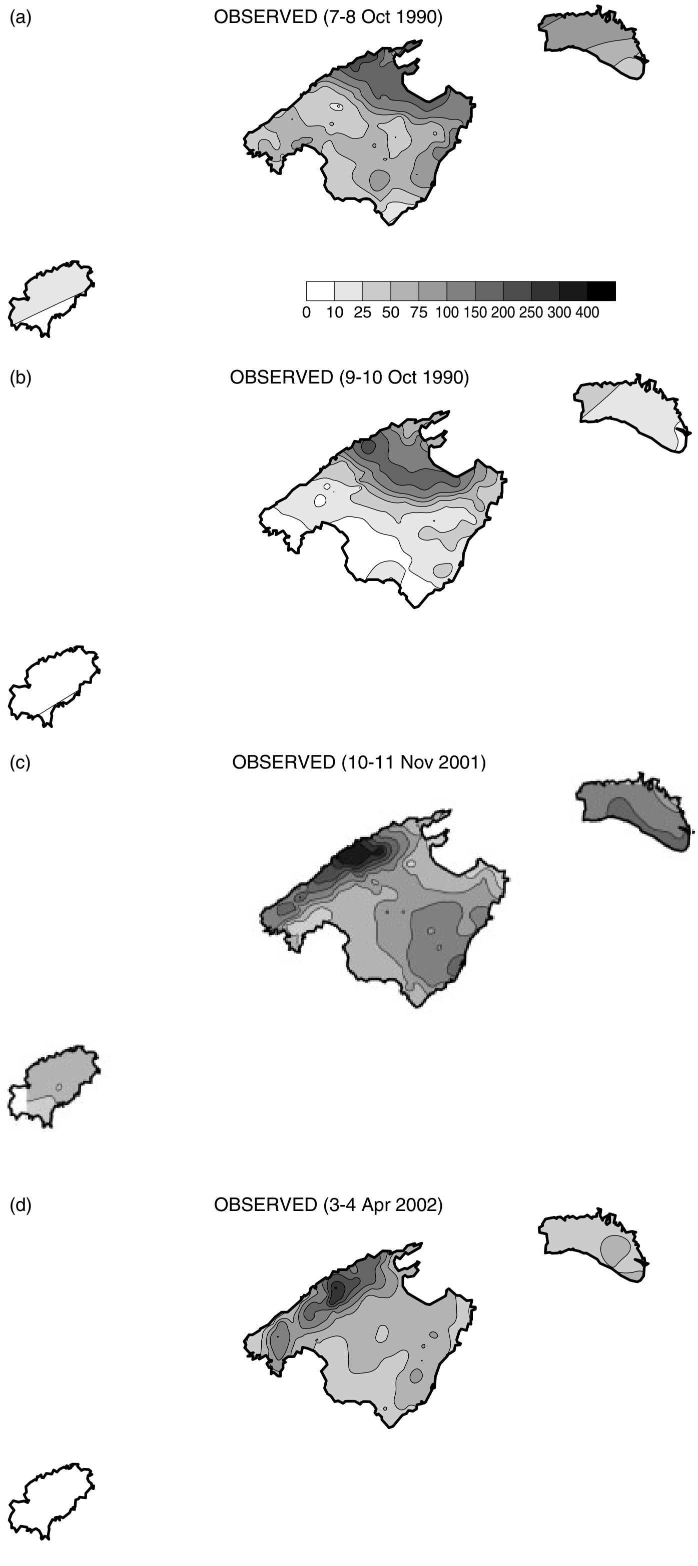

Figure 4. Observed accumulated precipitation (in mm), for: (a) 7–8 October 1990; (b) 9–10 October 1990; (c) 10–11 November 2001; (d) 3–4 April 2002.

3.2. Case 2: 7–10 October 1990, second phase

During the second phase of this episode, the depression at the 500 hPa level remained stationary but weaker (Figure 3(c)). At low levels, the southeast–northwest pressure gradient maintained the northeasterly surface current (Figure 3(d)). This second phase was of short duration (a few hours), but produced extraordinary rainfall rates, with an hourly maximum close to 115 mm at 2200 UTC on 9 October. The spatial distribution of the rainfall was quite similar to that of the first phase, but with a slight southwestward shift. Thus,

DOI: 10.1002/qj

the entire northeastern part of the basin was affected by the sudden event, with cumulative rainfall amounts over 235 mm. The heavy precipitation was caused by an intense convective system that remained quasi-stationary over the same area. The subsequent flash flood affected several locations along the rivers; but again, the areas of coastal dwellings were the most damaged (Figure 4(b)).

3.3. Case 3: 10–11 November 2001

This case produced accumulated precipitation values close to 240 mm in 24 h, mainly between 2000 UTC on 10 November and 0500 UTC on 11 November, and total amounts up to 400 mm. During the event, the whole basin collected substantial quantities of precipitation, although the mountainous range was the most affected (Figure 4(c)). At 500 hPa, two embedded depressions in the large-scale trough were located to the south of the Iberian Peninsula (Figure 3(e)). The low-level cyclone was placed to the southeast of the Balearics over a zone of marked baroclinicity (Figure 3(f)). The associated strong winds produced a severe sea storm, substantial material losses, and four fatalities.

3.4. Case 4: 3–4 April 2002

This episode produced accumulated precipitations over 230 mm in 24 h, and total amounts in the period near 300 mm. The maximum amounts were recorded on the Tramuntana range between 1730 UTC on 3 April and 0930 UTC on 4 April. The observed rainfall pattern is quite similar to that of case 3, but with less rainfall collected over the Albufera basin (Figure 4(d)). The mid-to-upper-tropospheric low was situated over the western Mediterranean, very near the Balearics (Figure 3(g)). The associated surface cyclone was situated to the east of the Balearic Islands, providing northerly winds over the Albufera river basin (Figure 3(h)).

4. Meteorological tools

The non-hydrostatic MM5 numerical model is used to perform the meteorological simulations. This is a high-resolution short-range weather-forecast model jointly developed at Pennsylvania State University and NCAR (Dudhia, 1993; Grell *et al.*, 1995). The model domains are configured as in the real-time operational version used at the University of the Balearic Islands (UIB) (http://mm5forecasts.uib.es). Simulations are designed using 24 vertical σ-levels and three spatial domains with 121×121 grid points centred at the Balearic Islands (Figure 5). Their horizontal resolutions are 22.5 km, 7.5 km and 2.5 km. The finest domain spans the entire region of the Balearic Islands and the surrounding sea, and is used to supply the high-resolution rainfall fields to drive the hydrological simulations. The interaction between the domains follows a two-way nesting strategy (Zhang and Fritsch, 1986).

For initialization and provision of boundary conditions, large-scale analyses are interpolated to the MM5 coarse domain: ECMWF analyses are used for all the cases, and an additional experiment for the 10–11 November 2001 episode using NCEP analyses is also included. The latter serves as a test of the sensitivity of high-resolution rainfall simulations to the initial conditions. Thus, the MM5-ECMWF simulations use the ECMWF analyses, with a spatial resolution of 0.3° and an update period of 6 h, while the MM5-NCEP simulation uses the NCEP analyses, with a spatial resolution of 2.5° and an update period of 12 h. In both cases, the first-guess analysis fields interpolated onto the MM5 model grid are improved using surface SYNOP and upper-air RAOB observations with a successive-correction objective-analysis technique (Benjamin and Seaman, 1985). The tendencies along the model coarse-domain boundaries, specified by the differences of the fields between the analyses 6 h and 12 h apart, are applied using a Newtonian relaxation approach (Grell *et al.*, 1995).

The control simulation of the four episodes follows the same physics options as the UIB operational runs. To represent the moist-convection effects, the Kain–Fristch

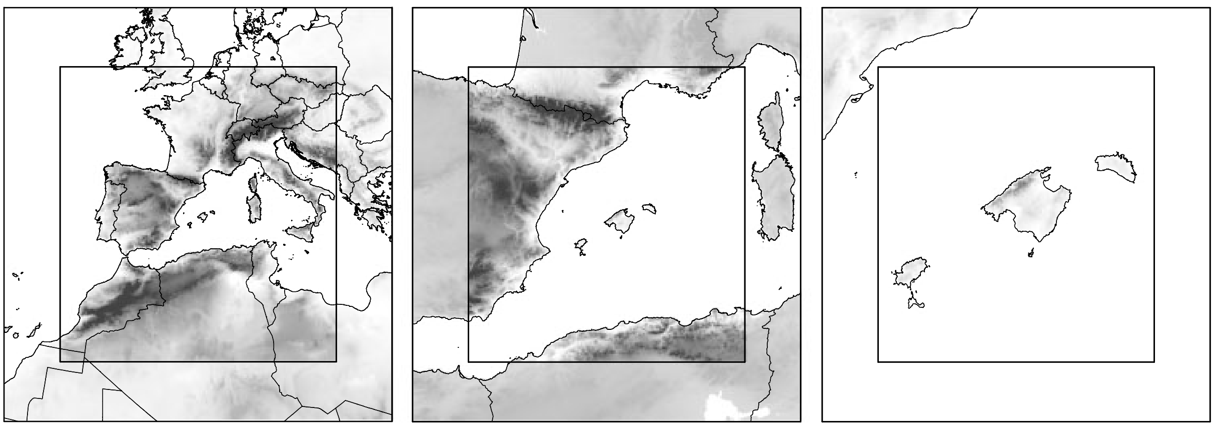

Figure 5. Configuration of the three computational domains used for the MM5 numerical simulations (inner squares), with horizontal resolutions of (left to right) 22.5 km, 7.5 km and 2.5 km.

DOI: 10.1002/qj

parametrization scheme (Kain, 2004) is used in the large domain, while convection is explicitly resolved in the second and third domains, thanks to the high horizontal resolutions. Explicit microphysics is represented in all domains, with prediction equations for cloud and rainwater fields, cloud ice and snow, allowing for slow melting of snow, supercooled water, graupel and ice number concentration (Reisner *et al.*, 1998). The planetary-boundary-layer physics is formulated using a modified version of the scheme of Hong and Pan (1996). Surface temperature over land is calculated using a force-restore slab model (Blackadar, 1979; Zhang and Anthes, 1982); over sea it remains constant during the simulations. Finally, longwave and shortwave radiative processes are formulated, taking into account the cloud cover (Benjamin, 1983). These control simulations are not intended to represent the most commonly-used configuration for the MM5 model, nor to have the best model skill for the study cases, but simply to match the configuration used in the UIB operational runs.

In addition to the control MM5 simulations, the multi-physics ensemble is run with different combinations of three models' physical parametrizations (explicit microphysics, moist convection and boundary-layer schemes), in order to better encompass the atmospheric processes leading to the high precipitation amounts. Following previous research, the Hong–Pan parametrization scheme (option 5 in the MM5 model) has been selected for the boundary-layer turbulence, and kept fixed. Then the multi-physics ensemble is defined as all possible combinations of five well-tested explicit moisture schemes (options 4, 5, 6, 7 and 8 in the model) and the inclusion or absence of the Kain–Fritsch convection scheme (option 8 in the MM5 model) in the second domain. Experiments with and without parametrized convection in the second domain are intended to account for the uncertainty about whether a 7.5 km resolution can resolve convection appropriately without a convection scheme.

The ten resulting experiments are labelled as follows:

- microphysics schemes (4, 5, 6, 7, 8) and Hong–Pan scheme (5) with Kain–Fritsch convection scheme in the first domain – simple ice (MM5-4-5), mixed-phase (MM5-5-5), graupel (MM5-6-5), Reisner-graupel (MM5-7-5, control), and Schultz (MM5-8-5);
- microphysics schemes (4, 5, 6, 7, 8) and Hong–Pan scheme (5) with Kain–Fritsch convection scheme (8) also in the second domain – MM5-4-5-8, MM5-5-5-8, MM5-6-5-8, MM5-7-5-8, and MM5-8-5-8, respectively.

All these MM5 simulations comprise a 48 h forecast period covering each of the flood episodes under study: three sets of experiments are performed to better encompass the episode of 7–10 October 1990 (cases 1 and 2) starting at 0000 UTC on 7, 8 and 9 October 1990; the simulations for case 3 start at 0000 UTC on 10 November 2001; and the simulations for case 4 start at 0000 UTC on 3 April 2002.

5. Hydrological tools

5.1. Rainfall-runoff model and input data

This study is carried out using the physically-based HEC-HMS rainfall-runoff model developed by the US Army Corps of Engineers (USACE-HEC, 1998). The model has been implemented in a semi-distributed and event-based configuration. HEC-HMS utilizes a graphical interface to build a semi-distributed watershed model and to set up rainfall and control variables for the simulations. Figure 2 shows the digital terrain model for the Albufera watershed together with the main watercourses forming the Almedrà and Sant Miquel river basins. The watershed has been segmented into 35 sub-watersheds, with an average size of 17.4 km^2 and a total extension of 607.4 km^2 at the junction of the Almedrà and Sant Miquel rivers (Figure 2).

HEC-HMS is forced using a single hyetograph for each sub-basin. This hyetograph is built in two steps: first, a spatial distribution of rainfall is generated from 1 h accumulated values at INM rain gauges (see Section 2.1 and Figure 2) using the kriging interpolation method with a horizontal grid resolution of 250 m; then, the temporal rainfall series is calculated for each sub-basin as the areal average of the gridded rainfall within the sub-catchment. The same methodology is used to assimilate simulated rainfall fields in HEC-HMS (see Section 6), except that atmospheric-model grid-point values are used instead of the INM network observations.

5.2. Theoretical background and model set-up

The hydrological model calculates runoff volume by subtracting from the rainfall the water volume that is lost through interception, infiltration, storage, evaporation and transpiration. The loss rate is calculated using the Soil Conservation Service (SCS) 'curve number' (SCS-CN) (e.g. USDA, 1986). This method assumes that the storm runoff volumes are proportional to the rainfall volumes exceeding an initial abstraction threshold, in terms of the ratio of the accumulated infiltration to a storage capacity. With this assumption and the continuity principle, the cumulative volume of storm-flow becomes nonlinearly related to the excess rainfall volume, which is a function of cumulative rainfall, soil cover, land use and antecedent moisture (Chow *et al.*, 1988; Bacchi *et al.*, 2002). The SCS-CN model has the advantage that with a single parameter (the storage capacity) it reproduces two phenomena that are systematically observed during floods: an initial loss of rain and an increase in the efficiency of the basin in producing runoff as a response to the rainfall input (Ranzi *et al.*, 2003). A synthetic unit hydrograph (UH) provided by SCS is used to convert rainfall excess into direct runoff on a watershed. This SCS-UH relates the peak discharge to the time to the UH peak in terms of the sub-basin area and a conversion constant. The flood hydrograph is routed using the Muskingum method (Chow *et al.*, 1988; USACE-HEC, 2000). The Albufera basin contains one reservoir located

DOI: 10.1002/qj

in the area upstream of the Almedrà river (Figure 2) but with no contribution downstream, and consequently it has not been modelled.

The lack of flow measurements in the basin poses great difficulties in the evaluation of the rain-gauge-driven runoff simulations. Therefore it has not been possible to carry out a calibration and verification task for the model. The runoff model has been run under the assumption that a best estimation of the initial model parameters can be obtained from the high-resolution observational campaign developed by the CORINE Land Cover project (Bossard *et al.*, 2000). The curve numbers, and thus the initial abstractions, have been assigned for all the sub-basins from that experimental database. All the rainfall-runoff model simulations have been run for a 72 h period with a 2 min time step. These periods cover the four flood events and the subsequent hydrograph tails, beginning at: 0000 UTC on 7, 8 and 9 October 1990 for cases 1 and 2; 0000 UTC on 10 November 2001 for case 3; and 0000 UTC on 3 April 2002 for case 4.

6. Results and discussion

6.1. Rain-gauge and MM5-control-driven runoff simulations

We first drive the stream-flow simulations using precipitation observations, in order to assess the performance of the model for the selected episodes. The lack of stream gauges in the river basin poses great difficulties in evaluating the hydrological simulations. The only existing information corresponds to field estimations of the floods of 9–10 October 1990 at various locations close to the rivers. The estimated peak discharge for the Sant Miquel river was 260 $m^3\ s^{-1}$ when it was overflowing, and for the Almedrà river was 366 $m^3\ s^{-1}$ the estimated peak flows were estimated by Grimalt; and the application of these equations was carried out by Grimalt as well (Grimalt, 1992). We apply the empirical Riggs equation (Riggs, 1976) for times when the flood waves are contained within the river channels, and the empirical Williams equation (Williams, 1978) at river sites where active flood planes were found.

Rain-gauge-driven simulation yields maximum discharges of 356 $m^3\ s^{-1}$ and 346 $m^3\ s^{-1}$ for the Sant Miquel and Almedrà rivers, respectively, at the basin outlet. Thus, it seems that the hydrological-model set-up satisfactorily captures the initial basin conditions, at least in terms of its effects on the attained peak discharges linked to this episode. A small simulation peak-discharge error, of only −5.5%, is obtained for the Almedrà river basin. With respect to the Sant Miquel river, the inaccuracy for the maximum flow at the basin outlet is larger, close to 37%, but the river did overflow and the hydrological model cannot take this effect into account. It is worth noting that the simulated peak discharges of the two river basins coincide in time at the basin outlet, yielding a total peak outflow over 700 $m^3\ s^{-1}$ (Figure 6(b)). This is a remarkable flow considering the small size of the whole watershed.

Note the peak discharge obtained for the 7–8 October simulation, where again both river flows coincide,

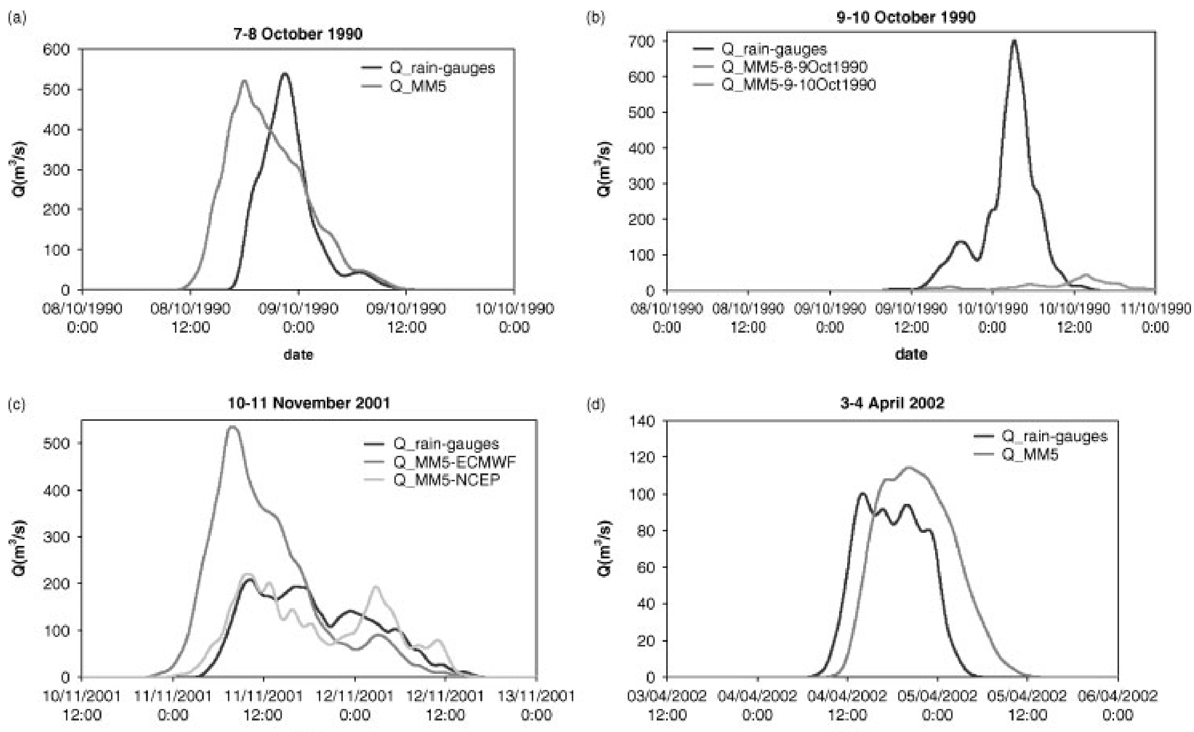

Figure 6. Rain-gauge-driven and MM5-control-driven runoff simulations, for: (a) 7–8 October 1990; (b) 9–10 October 1990; (c) 10–11 November 2001; (d) 3–4 April 2002. Panel (b) shows the mesoscale-model-driven runoff experiments for 8–9 and 9–10 October 1990. Panel (c) shows the discharge runs for the mesoscale-model simulations driven by ECMWF and NCEP analyses for 10–11 November 2001.

DOI: 10.1002/qj

reaching a maximum value of about 540 $m^3 s^{-1}$ on 8 October at 2230 UTC (Figure 6(a)). The rain-gauge-driven simulation of the 10–11 November 2001 episode shows a signal characterized by several peak discharges, the maximum of which is above 208 $m^3 s^{-1}$ (on 11 November 2001 at 1010 UTC), corresponding mainly to the contribution of the Sant Miquel river. The second-highest peak occurs four hours later, with a stream-flow close to 200 $m^3 s^{-1}$, due principally to the Almedrà-river discharge (Figure 6(c)). Note also the quicker response of the shorter and steeper Sant Miquel river basin in comparison with the Almedrà watershed when the heavy rainfall does not exclusively affect the last sections of both rivers (contrast, for example, cases 2 and 3, Figure 4(b,c)). Finally, the simulation of 3–4 April 2002 produces the lowest peak among the set of simulations, with a value up to 100 $m^3 s^{-1}$, on 4 April at 1400 UTC, due again to the contribution of the Sant Miquel river. The Almedrà basin's peak contribution is obtained later, at 2000 UTC, with a maximum stream-flow above 90 $m^3 s^{-1}$ (Figure 6(d)).

The MM5 mesoscale model provides the high-resolution precipitation fields for the episodes (MM5 control – see Figure 7). Runoff simulations driven by these rainfall patterns are then compared against the rain-gauge-driven runoff simulations. The skill of the resulting runoff experiments is expressed in terms of the Nash–Sutcliffe efficiency (NSE) criterion (Nash and Sutcliffe, 1970), a 'goodness-of-fit' measure widely used in hydrological model validation (Dolciné *et al.*, 2001; Jasper and Kaufmann, 2003). The NSE takes values in the range $(-\infty, 1]$, with higher values indicating a better agreement of the model results with the observations. The performance of the runoff simulations is also checked by means of the relative error of total volume at basin outlet, expressed as a percentage (EV). Positive and negative values of EV indicate overestimation and underestimation, respectively, of the volume by the model. In addition, to assess the skill of the MM5 mesoscale runs, we compare the spatial and temporal distributions of the simulated rainfall volumes against the rain-gauge-derived volume patterns. The spatial comparison is done using the 35 sub-basins as accumulation units for each episode, and the temporal comparisons use hourly accumulations for the whole basin. The degree of agreement between simulated and observed rainfall distributions is measured using the NSE efficiency criterion and the root mean square error (RMSE).

Tables I and II present the skill of the spatial and temporal rainfall volume distributions for the set of MM5-7-5 (control) simulations, and Figure 7 shows the accumulated-rainfall patterns of these simulations (compare with Figure 4). With regard to the spatial distributions, the experiments for 7–8 October 1990, 10–11 November 2001 using NCEP analysis, and 3–4 April 2002 show the best performances, while moderate errors are found for the 10–11 November 2001 MM5-ECMWF simulation. With regard to the timing, only the 10–11 November 2001 MM5-NCEP and the 3–4 April 2002 simulations present a reasonable agreement with the observed rainfall series. The runs for 8–9 and 9–10 October 1990 are not able to precisely match either the spatial or the temporal rainfall distributions of this convective episode, and the 10–11 November 2001 MM5-ECMWF experiment presents an over-forecasting of the precipitation amounts over the basin (compare Figures 4 and 7). It seems that, as a general feature for this set of episodes, the mesoscale model determines the spatial rainfall distributions more precisely than the temporal rainfall distributions. In addition, a noticeable impact of the November 2001 multi-analysis experiment is obtained.

Table III and Figure 6 summarize the MM5-control-driven runoff simulations in this complex orographic basin. Some of the experiments reproduce reasonably well the rain-gauge-driven floods, in spite of the small size of the basin, thus allowing the production of potentially valuable discharge predictions. Specifically, for the episode of 7–8 October 1990 (Figure 6(a)), the MM5 and rain-gauge-driven runoff simulations are quite similar in terms of peak discharge – with a slight difference close to 16 $m^3 s^{-1}$ – but with an important advance on the time to peak (more than 4 h). This fact, together with the wide overestimation of the runoff volume, causes a penalty in the NSE index. Better statistical scores are found for the event of 3–4 April 2002 (Figure 6(f)). The MM5-driven runoff simulation shows a moderate error in forecasting the time to peak, with a delay to the first maximum of 3 h, but better agreement is found in terms of the maximum peak discharge (with a relative error of 14.2%) and the runoff volume. The most suitable results are obtained for the 10–11 November 2001 MM5-NCEP experiment (Figure 6(e)). The simulation accurately matches the rain-gauge-driven hydrograph, with a NSE score of 0.84, an error in volume of only 1.7%, and a small overestimation of the peak discharge (below 15 $m^3 s^{-1}$), together with a slight advance in time (about 30 min). However, the mesoscale-model-driven runoff runs are very deficient in the MM5-ECMWF experiments for 8–9 and 9–10 October 1990 and 10–11 November 2001. The first two simulations completely miss the flash-flood event, resulting in a severe underestimation of the flow (Table III, Figure 6(b)), whereas the last run largely overestimates the flood at the basin outlet (Table III, Figure 6(c)). Therefore, it remains an important question whether our multi-physics probabilistic strategy could potentially provide better short-range prediction guidance when dealing with these unsuccessful flood simulations.

6.2. Multi-physics ensemble of MM5-driven runoff simulations

Following the motivation and methodology explained in Section 4, nine additional experiments are performed for each episode in order to produce the MM5 multi-physics ensemble. To evaluate the derived runoff simulations, the aforementioned statistical indices are used. These

DOI: 10.1002/qj

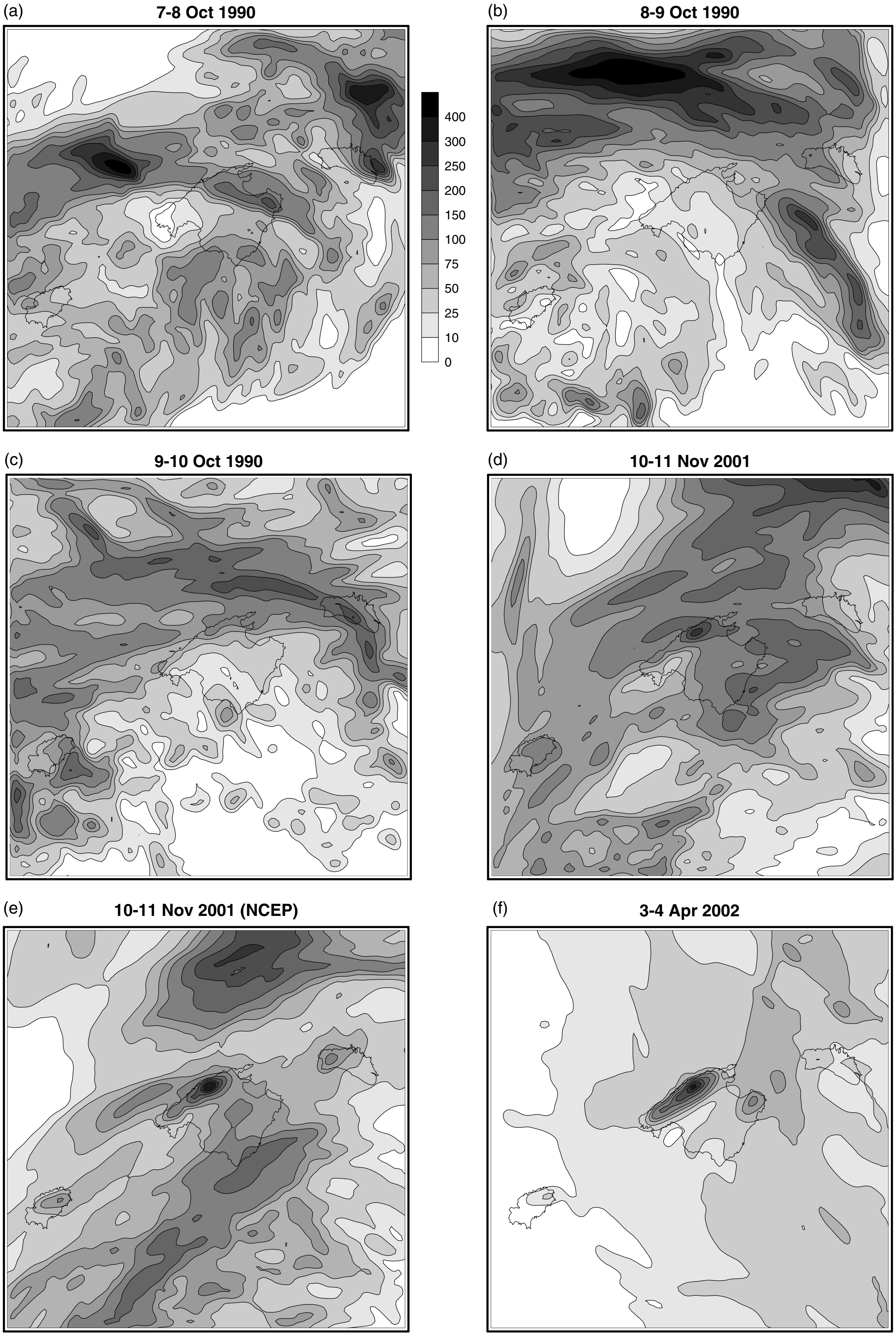

Figure 7. Spatial distribution of accumulated precipitation over the Balearics, for MM5-control 48 h simulations: (a) 7–8 October 1990; (b) 8–9 October 1990; (c) 9–10 October 1990; (d) 10–11 November 2001; (e) 10–11 November 2001 (NCEP); (f) 3–4 April 2002.

Table I. Error indices applied to the spatial rainfall volume distributions produced by the ensemble of MM5 simulations of the episodes under study. RMSE values are in hm^3. The best simulations according to these indices are shown in bold.

	7–8 Oct 1990		8–9 Oct 1990		9–10 Oct 1990	
	NSE	RMSE	NSE	RMSE	NSE	RMSE
MM5-4-5	**0.72**	**0.58**	−0.10	2.50	−0.12	2.53
MM5-5-5	−0.41	1.30	−0.10	2.50	−0.17	2.59
MM5-6-5	0.51	0.76	−0.01	2.40	−0.12	2.53
MM5-7-5 (control)	**0.80**	**0.49**	−0.73	3.14	−0.45	2.88
MM5-8-5	0.30	0.91	−0.42	2.84	−0.11	2.52
MM5-4-5-8	0.45	0.81	**0.13**	**2.23**	−0.18	2.59
MM5-5-5-8	**0.68**	**0.62**	0.06	2.32	−0.12	2.53
MM5-6-5-8	**0.87**	**0.40**	−0.09	2.50	0.35	1.92
MM5-7-5-8	0.33	0.90	0.01	2.38	0.10	2.26
MM5-8-5-8	**0.75**	**0.55**	0.11	2.26	**0.51**	**1.67**
Mean	0.79	0.50	−0.07	2.47	0.02	2.37
	10–11 Nov 2001		10–11 Nov 2001 (NCEP)		3–4 Apr 2002	
	NSE	RMSE	NSE	RMSE	NSE	RMSE
MM5-4-5	**0.63**	**0.68**	0.50	0.80	0.38	0.70
MM5-5-5	−1.05	1.61	−0.38	1.32	0.55	0.60
MM5-6-5	0.54	0.76	−0.29	1.28	0.41	0.69
MM5-7-5 (control)	0.33	0.92	**0.86**	**0.42**	**0.91**	**0.26**
MM5-8-5	0.59	0.72	**0.76**	**0.55**	**0.76**	**0.44**
MM5-4-5-8	0.52	0.78	0.46	0.82	0.44	0.67
MM5-5-5-8	0.51	0.79	0.41	0.86	**0.71**	**0.48**
MM5-6-5-8	**0.70**	**0.62**	0.40	0.87	0.41	0.69
MM5-7-5-8	0.46	0.83	−0.96	1.58	**0.92**	**0.26**
MM5-8-5-8	−0.63	1.44	−0.26	1.26	**0.81**	**0.39**
Mean	0.56	0.75	0.57	0.73	0.85	0.35

are complemented with an additional set of skill scores in order to test the quality of hydrometeorological-chain simulations. Specifically, the frequency-bias score (BIAS), the probability of detection (POD), the false-alarm rate (F) and the relative-operating-characteristics (ROC) score are calculated for various precipitation and runoff volume thresholds (Jolliffe and Stephenson, 2003; Wilks, 2006). These skill indices are calculated using the six hydrometeorological experiments in order to increase the statistical significance of the results as follows:

- the rainfall volumes of the MM5 control simulations are compared against the observed rainfall volumes;
- the rainfall volumes of the ensemble means are employed;
- the volumes by the members of the ensembles are used for comparison with the observations.

All these rainfall volumes are accumulated at hourly time steps, using the sub-basins as accumulation units. With regard to the discharge volumes, the same methodology is followed except that we instead compare the hourly runoff volumes produced at each sub-basin by the MM5-driven runoff simulations against the rain-gauge-driven runoff volume accumulations. In this case, the scores are calculated using only the hourly data that are non-zero in at least one of the two compared series, in order to prevent an artificial improvement of the ROC values.

It is worth remarking that, statistically, the ensemble mean provides a better forecast than any individual ensemble member, because errors in the individual forecasts tend to cancel when averaged (Epstein, 1969; Leith, 1974). Moreover, previous research studies (Du *et al.*, 1997; Stensrud *et al.*, 2000; Jones *et al.*, 2007) have pointed out that, despite the computational limitations and burdens of generating large ensembles, a clear improvement due to ensemble averaging can be obtained with small ensembles (typically with 8–19 members). Therefore, ROC scores for the ensemble means of rainfall and runoff volumes have also been computed in order to highlight the benefits of a simple ensemble average in comparison with the control experiments.

The probabilistic results provided by the ensemble strategy have been represented as cumulative distribution functions for the maximum peak discharges plotted on a Gumbel chart (Ferraris *et al.*, 2002). Each peak flow value is equally likely, by our assumption of equally-skilful model configurations. Although no hydrometeorological forecasting chain is currently implemented for

DOI: 10.1002/qj

Table II. Error indices applied to the temporal rainfall volume distributions produced by the ensemble of MM5 simulations of the episodes under study. RMSE values are in hm^3. The best simulations according to these indices are shown in bold.

	7–8 Oct 1990		8–9 Oct 1990		9–10 Oct 1990	
	NSE	RMSE	NSE	RMSE	NSE	RMSE
MM5-4-5	**0.03**	**2.87**	−0.19	4.94	−0.09	4.73
MM5-5-5	−1.83	4.90	**−0.10**	**4.75**	−0.22	5.00
MM5-6-5	−0.57	3.66	−0.21	4.99	−0.22	5.01
MM5-7-5 (control)	−0.41	3.47	−0.19	4.94	−0.13	4.81
MM5-8-5	−0.25	3.27	−0.24	5.05	**−0.06**	**4.66**
MM5-4-5-8	−0.28	3.30	−0.23	5.02	−0.19	4.94
MM5-5-5-8	−0.32	3.35	−0.70	5.91	−0.20	4.95
MM5-6-5-8	−0.67	3.76	−0.20	4.96	−0.25	5.06
MM5-7-5-8	−0.83	3.95	−0.22	5.01	−0.20	5.06
MM5-8-5-8	−1.22	4.35	−0.37	5.30	−0.38	5.32
Mean	−0.28	3.30	−0.13	4.82	−0.13	4.82
	10–11 Nov 2001		10–11 Nov 2001 (NCEP)		3–4 Apr 2002	
	NSE	RMSE	NSE	RMSE	NSE	RMSE
MM5-4-5	−0.67	2.38	**0.45**	**1.36**	0.10	1.20
MM5-5-5	−2.03	3.21	0.11	1.74	−0.07	1.31
MM5-6-5	−0.03	1.87	−0.65	2.37	−0.40	1.50
MM5-7-5 (control)	−0.02	1.86	**0.31**	**1.54**	**0.37**	**1.01**
MM5-8-5	−1.21	2.74	0.30	1.54	0.07	1.22
MM5-4-5-8	−0.25	2.06	0.13	1.72	0.12	1.18
MM5-5-5-8	−1.68	3.02	0.20	1.65	0.08	1.21
MM5-6-5-8	**0.11**	**1.74**	0.22	1.63	−0.45	1.52
MM5-7-5-8	0.08	1.77	−0.73	2.43	**0.35**	**1.01**
MM5-8-5-8	−3.15	3.76	−2.77	3.58	0.03	1.25
Mean	−0.10	1.93	0.53	1.26	0.13	1.18

civil-protection purposes in the Albufera basin, we have considered suitable for the present study the introduction of a hypothetical warning discharge threshold. This threshold, $Q_{\mathrm{th}} = 100$ m^3 s^{-1}, is the lowest peak discharge found among the four rain-gauge-driven runoff simulations.

We will now briefly discuss the results, case by case.

6.2.1. *Case 1: 7–10 October 1990, first phase*

For this case study, some members of the ensemble outperform the control simulation in terms of the spatial or temporal rainfall distributions (Tables I and II, Figure 8(a)), the runoff volume, and the time to peak (Table III, Figure 9(a)). However, the control simulation still shows the best reproduction of the maximum discharge (Figure 6(a)).

From a probabilistic point of view, for the runoff ensemble, Figure 10(a) shows that the probability of peak discharge exceedance for the rain-gauge-driven runoff simulation is close to 0.3, and the probability of exceeding Q_{th} would be 1. This demonstrates the clear benefit of an ensemble that estimates the range of the atmospheric probability density function through the inclusion of the mesoscale-model physics uncertainties.

6.2.2. *Case 2: 7–10 October 1990, second phase*

With regard to the 9–10 October 1990 flash-flood episode, the two sets of MM5 ensemble simulations (for 8–9 and 9–10 October) are rather similar: the maximum rainfall amounts are located in the northwestern part of the domain, quite far from the Albufera basin (Tables I and II, Figure 8(b,c)). Only one member is sufficiently accurate to reproduce the rain-gauge-driven discharge (Figure 9(b,c)). This member pertains to the 9–10 October 1990 experiment, and depicts a peak disagreement of only 30 m^3 s^{-1}, but with a remarkable overestimation of the runoff volume (Table III). It is also worth noting that one member from the 8–9 October experiment approaches the rain-gauge-driven simulation. However, it is less accurate in terms of peak discharge, with a difference up to 200 m^3 s^{-1}, and an advance of the time to peak of 3 h.

Panels (b) and (c) of Figure 10 demonstrate a remarkable improvement in the rainfall simulation when an ensemble strategy is used: the poor detection of the control simulations are partially alleviated. Even though none of ensembles' members is able to reproduce the rain-gauge-driven peak discharge – emphasizing the low predictability of this hydrometeorological event – an

DOI: 10.1002/qj

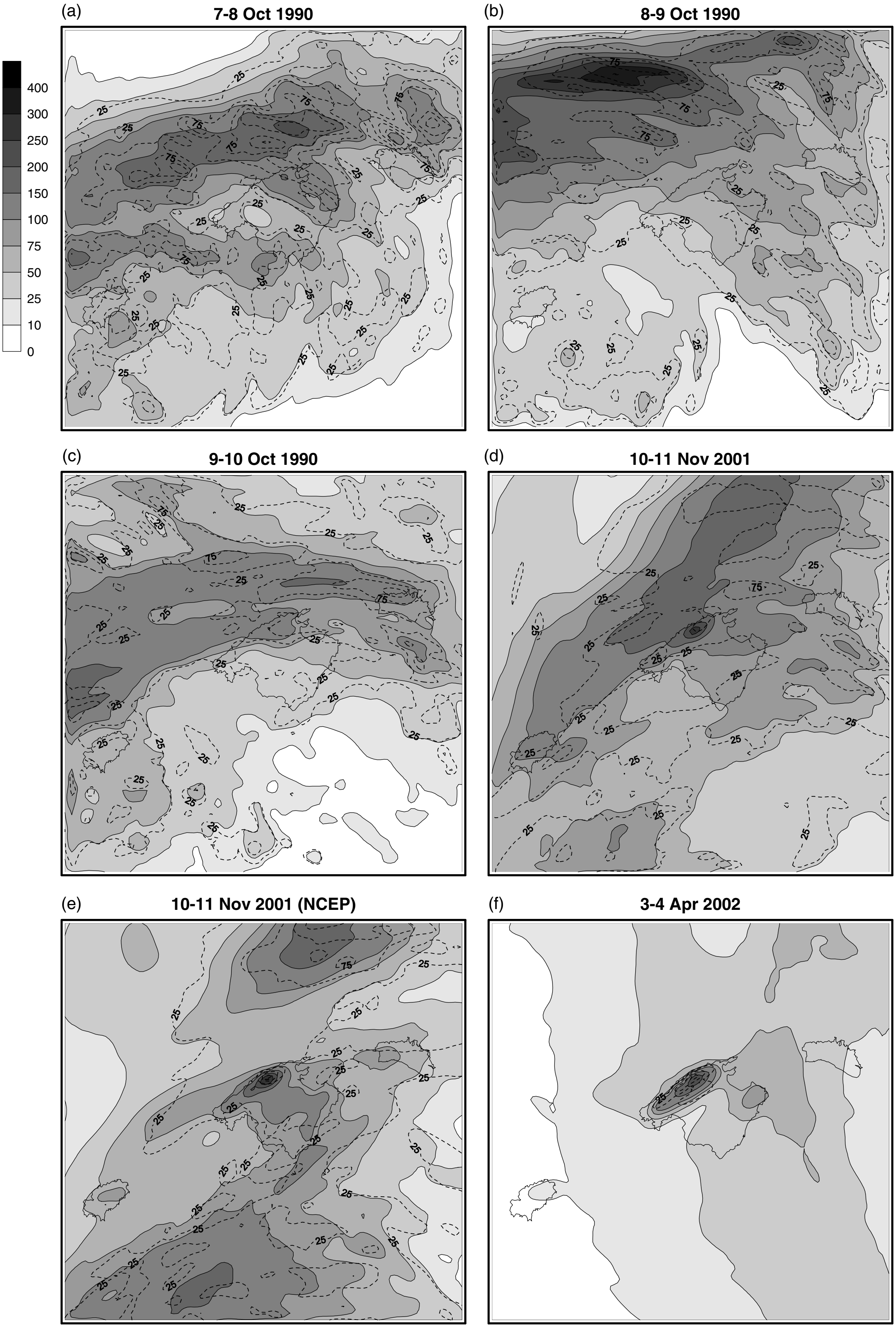

Figure 8. Ensemble mean (shaded contours, in mm) and ensemble standard deviation (dashed line, in mm at 25 mm intervals), for the accumulated precipitation over the Balearics, for: (a) 7–8 October 1990; (b) 8–9 October 1990; (c) 9–10 October 1990; (d) 10–11 November 2001; (e) 10–11 November 2001 (NCEP); (f) 3–4 April 2002.

DOI: 10.1002/qj

Table III. NSE and EV (%) at the Albufera basin outlet for the ensemble of MM5-driven runoff simulations and the selected episodes. The best simulations according to these indices are shown in bold.

	7–8 Oct 1990		8–9 Oct 1990		9–10 Oct 1990	
	NSE	EV	NSE	EV	NSE	EV
MM5-4-5	**0.75**	−26.9	0.22	−64.0	0.05	−70.2
MM5-5-5	−0.73	118.7	0.06	−70.6	−0.08	−75.3
MM5-6-5	0.31	−51.4	−0.02	−56.0	−0.14	−65.1
MM5-7-5 (control)	0.28	55.4	−0.2	−98.7	−0.15	−91.6
MM5-8-5	0.31	36.5	−0.17	−88.9	−0.13	−60.0
MM5-4-5-8	0.36	−56.0	0.28	−35.8	0.17	−60.5
MM5-5-5-8	**0.63**	**−19.2**	**0.61**	**−4.7**	0.28	−62.7
MM5-6-5-8	0.38	−26.9	**0.53**	−43.3	**0.56**	**−5.2**
MM5-7-5-8	**0.58**	**−19.8**	0.26	−47.8	0.34	−36.5
MM5-8-5-8	−0.21	46.9	0.1	−47.9	**0.69**	50.4
Mean	0.68	5.7	0.26	−55.8	0.3	−47.7
	10–11 Nov 2001		10–11 Nov 2001 (NCEP)		3–4 Apr 2002	
	NSE	EV	NSE	EV	NSE	EV
MM5-4-5	0.49	41.0	**0.71**	58.8	−2.75	193.1
MM5-5-5	−5.29	190.5	0.51	147.5	−0.89	129.3
MM5-6-5	0.03	59.0	0.52	135.1	−2.82	187.4
MM5-7-5 (control)	−1.53	73.1	**0.84**	**1.7**	**0.60**	27.8
MM5-8-5	−0.46	43.4	**0.74**	**−15.8**	0.50	−42.4
MM5-4-5-8	**0.57**	**18.5**	−1.44	−56.1	−2.38	183.0
MM5-5-5-8	−0.55	48.2	−1.28	−56.3	−0.02	78.8
MM5-6-5-8	**0.71**	**−15.4**	**0.63**	71.8	−2.88	188.8
MM5-7-5-8	**0.62**	−51.8	0.48	169.0	**0.57**	**13.5**
MM5-8-5-8	−9.05	172.6	**0.52**	142.0	**0.55**	−32.5
Mean	−0.23	57.9	0.78	59.8	−0.14	92.7

important improvement is found, and the probability of exceeding Q_{th} would be close to or above 0.8 for both ensembles. In agreement with Stensrud *et al.* (2000), we find that model physics largely control the evolution of this convectively-driven weather event.

6.2.3. *Case 3: 10–11 November 2001*

As explained in Section 4, the set of simulations for the 10–11 November 2001 episode consists of multi-physics ensembles initialized with ECMWF and NCEP analyses. It appears that the two ensembles (MM5-ECMWF and MM5-NCEP) present a similar performance with regard to the spatial distribution of rainfall, and a great homogeneity among their members (Table I, Figure 8(d,e)), as well as for the total precipitated volume over the basin: the observed volume was 60.9 hm^3, whereas the mean volumes obtained by the ECMWF and NCEP experiments are 73.2 hm^3 and 74.0 hm^3 respectively. Nevertheless, the members of the MM5-NCEP ensemble show a better reproduction of the temporal rainfall distribution, together with a greater uniformity of results (Table II), and therefore the ensemble of MM5-NCEP-driven runoff simulations presents more members with a best 'goodness of fit' (Table III, Figure 9(d,e)). Some members of the MM5-ECMWF-driven runoff ensemble display reasonable agreement in terms of the peak discharge and the stream-flow volume when compared against rain-gauge-driven runoff, thus outperforming the control experiment.

In addition, the MM5-ECMWF-driven runoff ensemble presents a considerable increase in forecasting skill with respect to the deterministically-driven runoff simulation, and the probabilities of exceedance of the rain-gauge-driven runoff peak and the threshold peak flow are above 0.9 and up to 1 respectively (Figure 10(d)). Rain-gauge-driven peak flow and Q_{th} show a probability of being exceeded greater than 0.8 for the MM5-NCEP-driven runoff ensemble (Figure 10(e)). In contrast with the MM5-NCEP deterministic experiment, which is very accurate, there is a clear overestimation of the runoff by the MM5-ECMWF deterministic run. This problem, however, is notably alleviated by most of the MM5-ECMWF ensemble members. It is worth noting that the effects of the external-scale uncertainties related to the initial and boundary conditions (as measured by the difference between MM5-ECMWF and MM5-NCEP control runs) are smaller than the effects due to the model-physics uncertainties (as measured by the spread found in both ensembles). This is consistent with the results of Stensrud *et al.* (2000), where the variance of

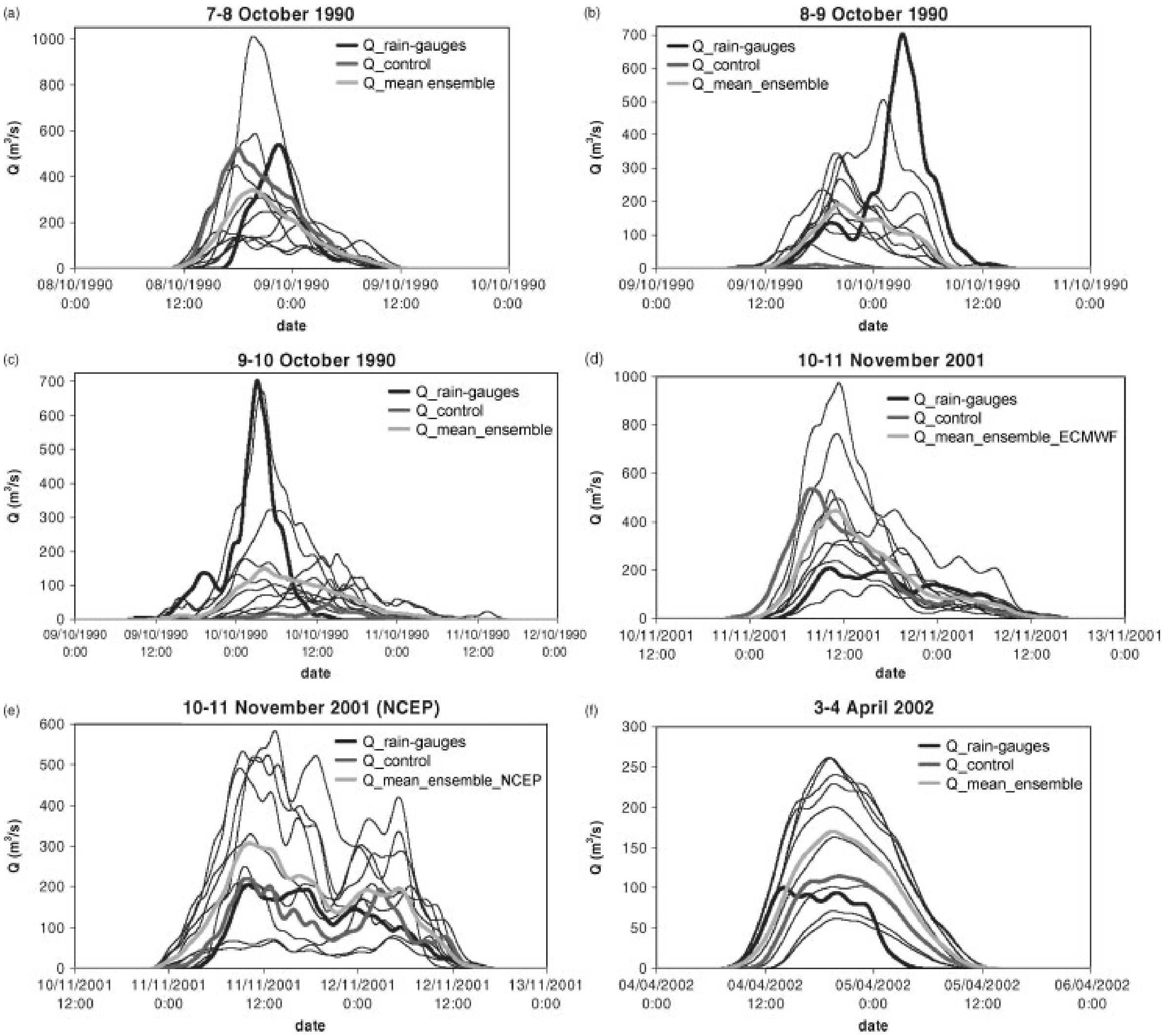

Figure 9. Ensemble of the multi-physics MM5-driven runoff simulations, for: (a) 7–8 October 1990; (b) 8–9 October 1990; (c) 9–10 October 1990; (d) 10–11 November 2001; (e) 10–11 November 2001 (NCEP); (f) 3–4 April 2002. The thin lines correspond to the nine additional ensemble members.

a multi-physics experiment exceeded that produced by an initial-condition experiment in a short-range-ensemble-forecast (SREF) modelling system.

6.2.4. Case 4: 3–4 April 2002

The atmospheric ensemble for the 3–4 April 2002 case presents the greatest similarities in the simulated rainfall patterns among its members, and some of them accurately match the observed pattern (Table I, Figure 8(f)), but this feature is lost when the temporal distributions are evaluated (Table II). Attending to the resulting runoffs, three simulations are clearly the most suitable (Table III, Figure 9(f)), with only small differences in their skill scores. The low temporal skill of the mesoscale model results in a remarkable delay in the timing of the peak discharges. Moreover, the important over-forecasting of the precipitation amounts produces excessive flow volumes. These facts are reflected in the poor performance of the ensemble of simulated hydrographs.

However, the probability of exceedance of the rain-gauge-driven peak discharge is above 0.8 (Figure 10(f)). These results would have been found suitable in a hypothetical real-time hydrometeorological forecasting framework, thanks to the high probability of surpassing Q_{th}. In fact, Anderson *et al.* (2002) have pointed out that runoff predictions for use in emergency-management directives may not need to exactly match the peak discharges or the timing, but must reach suitable thresholds to cause the appropriate directives to be enacted. Ferraris *et al.* (2002) have argued for similar requirements within an operational civil framework for the Tanaro river basin in northwestern Italy.

6.2.5. POD, F, BIAS and ROC skill indices

Table IV and Figures 11 and 12 present the results for the rest of the applied scoring techniques. As mentioned above, the skill indices are applied to all the experimental ensembles in order to increase their statistical significance. With regard to the rainfall accumulations, the ensemble mean proves to be the best for the POD, although the increased skill for the POD induces a rise in the F score, and a moderate over-forecasting of rain

DOI: 10.1002/qj

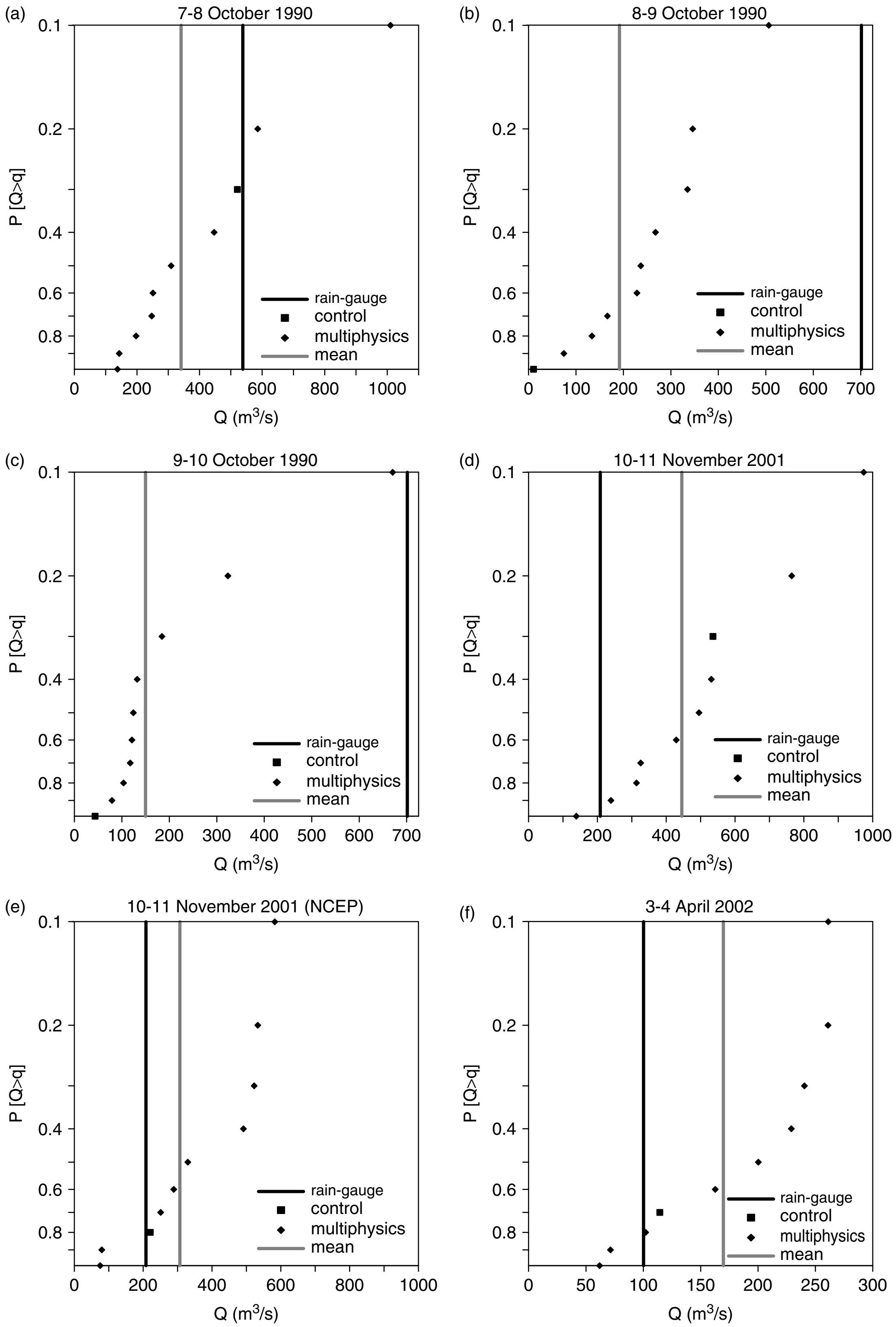

Figure 10. Peak discharge exceedance probability, plotted on a Gumbel chart, for the hydrometeorological experiments: (a) 7–8 October 1990; (b) 8–9 October 1990; (c) 9–10 October 1990; (d) 10–11 November 2001; (e) 10–11 November 2001 (NCEP); (f) 3–4 April 2002. The cumulative distribution functions of peak discharge are shown at the Albufera river-basin outlet. The vertical black line indicates the rain-gauge-driven maximum peak flow at the Albufera river-basin outlet, for each of the hydrometeorological events. The vertical grey line indicates the maximum peak discharge from the ensemble mean.

amounts at low and medium thresholds can be observed considering the BIAS index. The ensemble mean presents the highest ROC score, as a consequence of the aforementioned smoothing effect from averaging the rainfall fields (Table IV). ROC scores for the control and ensemble simulations are rather similar, and higher PODs for the ensemble experiments can be appreciated only at low precipitation volumes. Moreover, the set of ensembles gives higher F for all thresholds, when compared with the control simulations. These control simulations give a systematic under-prediction of the precipitation amounts at all thresholds, which is in part alleviated by the ensemble experiments (Figure 11(c)).

With respect to the driven runoff forecasts, Table IV shows again the highest performance in terms of ROC score for the ensemble mean, thanks to the highest POD scores at small and medium values, together with an appreciable tendency of decreasing F at increasing volumes. Note also the over-prediction at low and medium thresholds, and under-prediction at high thresholds, for the ensemble mean (Figure 12). Furthermore, we find an improvement of the ROC scores when using the ensembles strategy instead of the deterministic control simulations. It appears that the ensembles yield slightly higher POD and smaller F indices at low thresholds, and a smaller underestimation of runoff volumes at medium to high thresholds.

Finally, it is worth noting that the introduction of a convection scheme in the second domain is only beneficial for some of the high-resolution rainfall simulations (Tables I and II). For example, an improvement in the spatial distribution for the 9–10 October 1990 event is found for one of the experiments, but the reproduction of the heavy-rainfall timing is rather deficient. This fact leads to a slight improvement in the simulation of the flash-flood event in terms of runoff (Table III). The enhanced representation of the physical processes resulting from the parametrized convection is of reasonable benefit to the 10–11 November 2001 ensemble based on ECMWF analyses, since the wide areas with large rainfall amounts (Figure 7(d)) become better constrained to the Albufera basin (not shown). However, no benefit is obtained for the 10–11 November 2001 ensemble based on NCEP analyses, and only a slight improvement can be noticed for some experiments of the 7–8 October 1990 episode. These results reinforce the idea of considering the members of the physics ensemble with and without parametrized convection in the second domain as equally skilful.

Table IV. ROC scores for the control simulations, ensemble mean and full ensemble of all the hydrometeorological experiments, for hourly rainfall and runoff volumes.

	Control	Mean	Ensemble
rainfall volume	0.67	0.74	0.68
runoff volume	0.57	0.77	0.72

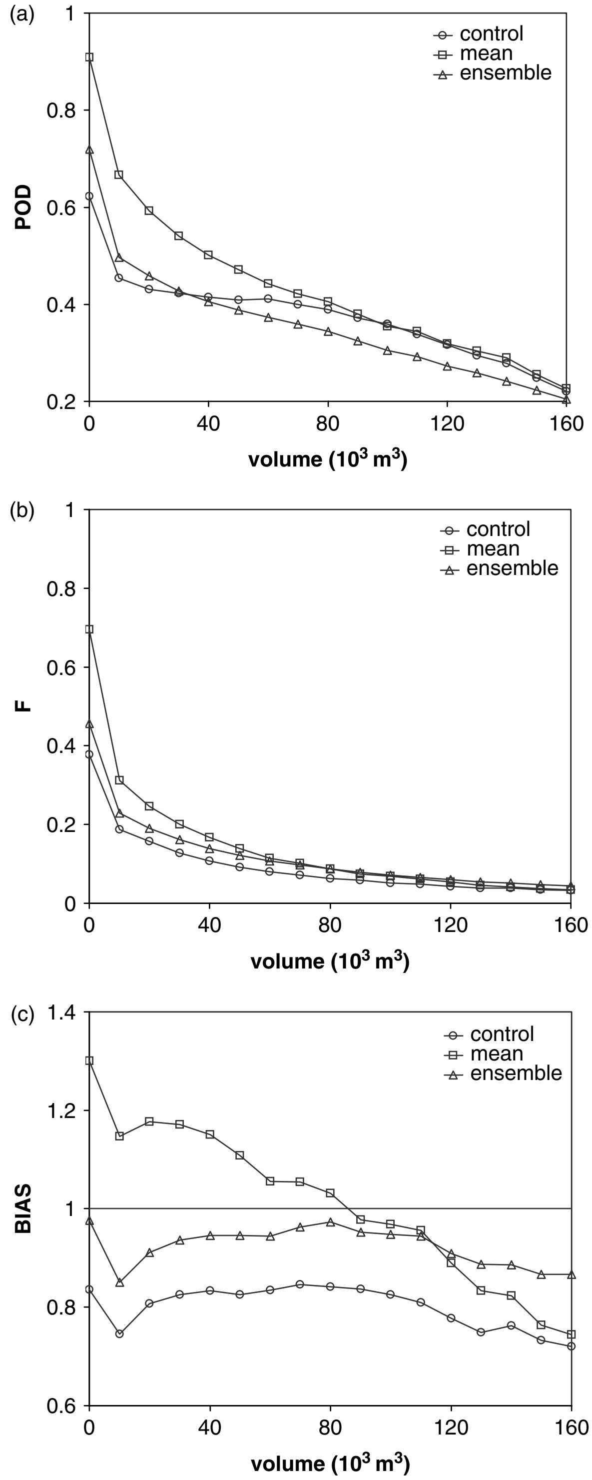

Figure 11. (a) POD, (b) F and (c) BIAS skill scores for different precipitation-volume thresholds obtained by the ensemble of MM5-driven runoff discharge simulations for all the hydrometeorological experiments.

7. Conclusions and further remarks

We have analysed the feasibility of runoff simulations driven by a high-resolution non-hydrostatic mesoscale atmospheric model over the small Albufera river basin of Majorca. By using analyses (rather than forecasts) to drive the model, we assume a best scenario for the synoptic-scale environment. We have considered

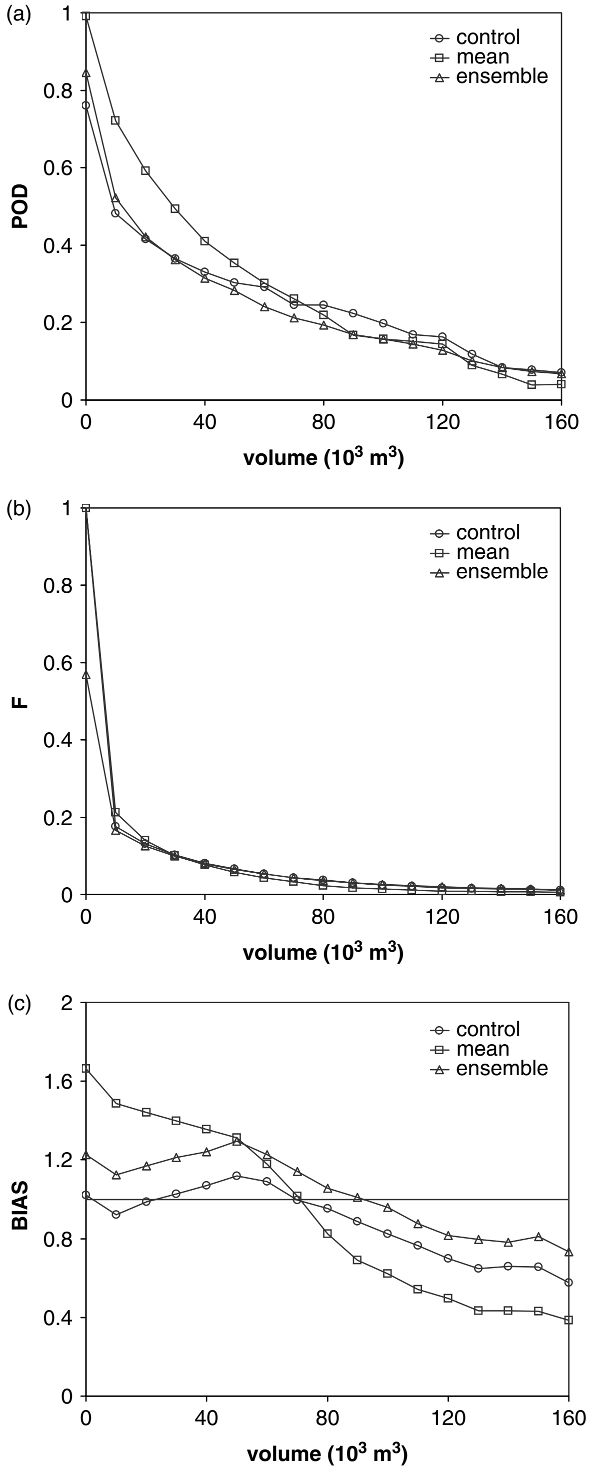

Figure 12. (a) POD, (b) F and (c) BIAS skill scores for different runoff-volume thresholds obtained by the ensemble of MM5-driven runoff discharge simulations for all the hydrometeorological experiments.

four intense-rainfall events, which resulted in floods of varying spatial and temporal scales. These kinds of intense precipitation events – often highly localized and convectively driven – present short recurrence periods in Mediterranean Spain as a whole, and so the conclusions drawn could be widely applicable to other territories of the region as well.

Using ECMWF analyses to initialize the hydrometeorological chain, it is possible to obtain reasonable runoff simulations at the basin outlet for some of the episodes. In addition, an ensemble of MM5 experiments with varying microphysical, moist-convection and boundary-layer parametrizations has been adopted, in order to mitigate the low forecasting skill of the deterministic runoff simulations in some events (such as 9–10 October 1990 and 10–11 November 2001 using ECMWF analyses). The use of an ensemble strategy has thus enabled us to further extend the short-range prediction guidance when dealing with flood-simulation situations for the Albufera river basin. The ensemble of simulated rainfall fields displays moderate spatial and temporal variabilities, as well as significant changes in the precipitation amounts. Some members of the ensemble outperform the control simulation, and reduce the biases at the Albufera outlet, where the control experiments would not have produced enough accurate runoff simulations. A multi-analysis experiment has also been introduced for the 10–11 November 2001 event, in order to examine the sensitivity of the hydrometeorological chain to the initial and boundary meteorological conditions. For this particular case, it is found to be better to initialize the MM5 mesoscale model with the NCEP analyses, but this result cannot be generalized. Presumably, the better performance of the NCEP-based simulations in terms of peak discharges and their timing is simply a particularity of this meteorological situation, and not an inherent aspect of this analysis dataset.

The performance of the mesoscale model has been assessed from a comparison of simulated and observed rainfall distributions in space and time over the subbasins, as well as in terms of the MM5-driven runoff discharges. Thus, the one-way coupling between the meteorological and hydrological models is regarded as a validation tool for the simulated rainfall distributions. The value of a multi-physical model ensemble in conveying the uncertainty of the small-scale features in precipitation, and thus of the discharge simulations, has also been proven. This provides a good example of the potential benefits of more general SREF modelling systems aimed at accounting for the forecast variance associated with the physical parametrizations or the uncertainties in the initial conditions.

The lack of flow-gauge measurements and any estimated runoff peak (except for the 9–10 October 1990 episode), together with the scarcity of automatic pluviometric stations over the Albufera catchment, entails great difficulties in constructing a set of reference rain-gauge-derived runoff simulations. Nevertheless, the perfect-hydrological-model assumption allows us to consider these unverified rain-gauge-driven discharges as 'observed flows' when evaluating sets of mesoscale-model-driven runoff simulations. Obviously, in the framework of a hydrometeorological forecasting chain, the reliability and skill of the rainfall-runoff model must be improved, and it would be highly desirable to get more information on the flood events affecting the Albufera river basin. The expected future increase in the number

DOI: 10.1002/qj

of automatic stream and rain gauges in the catchment will help in addressing the uncertainties related both to the spatial and temporal variabilities found in the model's initial conditions (the infiltration mechanism) and to the dynamical formulation (the channel routing). In spite of the current limitations, however, the benefits from hydrometeorological analyses such as the present one are of greater significance than their possible weaknesses, given the hazardous consequences and relatively short recurrence periods of these kinds of extreme hydrometeorological events.

Finally, it is important to note that the hydrometeorological chain has been designed without the intervention of any precipitation-assimilation technique connecting the meteorological and hydrological models, but using a very high-resolution mesoscale model. Although the one-way-coupled runoff simulations show reasonable skill for most of the evaluated episodes, the 9–10 October 1990 event provides a good example of the difficulties involved in precise detection of a convectively-driven episode affecting a small basin, even in the framework of an ensemble strategy. Precise reproduction of the timing and location of these episodes is a challenging problem, which could be addressed by implementing assimilation methodologies. Among these techniques are various applications of statistical downscaling (e.g. Hewitson and Crane, 1992; von Storch and Zwiers, 1999; Wilks, 1999; Antolik, 2000; Clark and Hay, 2004) and disaggregation techniques (Deidda *et al*., 1999; Deidda, 2000; Ferraris *et al*., 2002). These lines of research appear to be of great interest in relation to future attempts to develop suitable hydrometeorological-chain forecasting systems for the Balearic Islands.

Acknowledgements

Dr Peter Clark, associate editor of the *Quarterly Journal of the Royal Meteorological Society*, and two anonymous reviewers are deeply acknowledged for their valuable comments, which helped to improve the quality of this manuscript. The Water Agency of the Environmental Department of the Government of the Balearic Islands is acknowledged for providing the CORINE data. The Spanish Institute of Meteorology (INM) is also acknowledged for providing the precipitation data. Figure 1 was produced by Alberto Martín from the Meteorology Group of the University of the Balearic Islands. This work has been sponsored by CGL 2005-03918/CLI (PRECIOSO) Spanish and INTERREG IIIB-MEDOCC 2003-03-4.3-I-079 (AMPHORE) European projects.

References

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

Anderson ML, Chen ZQ, Kavvas ML, Feldman A. 2002. Coupling HEC-HMS with atmospheric models for prediction of watershed runoff. *J. Hydrol. Eng.* **7**: 312–318.

Antolik MS. 2000. An overview of the National Weather Service's centralized statistical quantitative precipitation forecast. *J. Hydrol.* **239**: 306–337.

Bacchi B, Buzzi A, Grossi G, Ranzi R. 2002. 'Flood forecasting in a midsize catchment in the southern Alps: Recent experiences on the use of coupled meteorological and hydrological models'. Pp. 201–208 in *Proceedings of the Third EGS Plinius Conference*, Baja Sardinia. Consiglio Nazionale delle Ricerche: Rome.

Benjamin SG. 1983. *Some Effects of Heating and Topography on the Regional Severe Storm Environment*. Ph.D. thesis, Pennsylvania State University, 265 pp. (Available from University Microfilm, 300 N. Zeeb. Rd, PO Box 1346, Ann Arbor, MI 46801-1346, USA.)

Benjamin SG, Seaman NL. 1985. A simple scheme for improved objective analysis in curved flow. *Mon. Weather Rev.* **113**: 1184–1198.

Benoit R, Pellerin P, Kouwen N, Ritchie H, Donaldson N, Joe P, Soulis ED. 2000. Toward the use of coupled atmospheric and hydrologic models at regional scale. *Mon. Weather Rev.* **128**: 1681–1706.

Blackadar AK. 1979. High resolution models of the planetary boundary layer. *Adv. Environ. Sci. Eng.* **1**: 50–85.

Bossard M, Feranec J, Otahel J. 2000. 'CORINE Land Cover Technical Guide – Addendum 2000'. Technical Report 40, European Environment Agency, Copenhagen, Denmark. 105 pp.

Bright DR, Mullen SL. 2002. The sensitivity of the numerical simulation of the southwest monsoon boundary layer to the choice of PBL turbulence parameterization in MM5. *Weather and Forecasting* **17**: 99–114.

Chancibault K, Anquetin S, Ducrocq V, Saulnier G-M. 2006. Hydrological evaluation of high-resolution precipitation forecasts of the Gard flash-flood event (8–9 September 2002). *Q. J. R. Meteorol. Soc.* **132**: 1091–1117.

Chow VT, Maidment DR, Mays LW. 1988. *Applied Hydrology*. McGraw-Hill. 572 pp.

Clark MP, Hay LE. 2004. Use of medium-range numerical weather prediction model output to produce forecasts of stream-flow. *J. Hydrometeorol.* **5**: 243–262.

Deidda R. 2000. Rainfall downscaling in a space-time multifractal framework. *Water Resour. Res.* **36**: 1779–1794.

Deidda R, Benzi R, Siccardi F. 1999. Multifractal modelling of anomalous scaling laws in rainfall. *Water Resour. Res.* **35**: 1853–1867.

Dolciné L, Andrieu H, Sempere-Torres D, Creutin D. 2001. Flash flood forecasting with coupled precipitation model in mountainous mediterranean basin. *J. Hydrol. Eng.* **6**: 1–9.

Du J, Mullen SL, Sanders F. 1997. Short-range ensemble forecasting of quantitative precipitation. *Mon. Weather Rev.* **125**: 2427–2459.

Dudhia J. 1993. A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Weather Rev.* **121**: 1493–1513.

Epstein ES. 1969. Stochastic dynamic prediction. *Tellus* **21**: 739–759.

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

Giannoni F, Smith JA, Zhang Y, Roth G. 2003. Hydrologic modeling of extreme floods using radar rainfall estimates. *Adv. Water Resour.* **26**: 195–203.

Grell GA, Dudhia J, Stauffer DR. 1995. 'A description of the fifth-generation of the Penn State/NCAR mesoscale model (MM5)'. NCAR Tech. Note NCAR/TN-398+STR, Boulder, USA.

Grimalt M. 1992. *Geography of the Risk in Majorca: The Floods*. Ph.D. thesis, Institut d'Estudis Baleàrics. Conselleria de Cultura, Educació i Esports. Govern de les Illes Balears. 359 pp. (In Catalan.)

Hewitson BC, Crane RG. 1992. Large-scale atmospheric controls on local precipitation in tropical Mexico. *Geophys. Res. Lett.* **19**: 1835–1838.

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

Jasper K, Kaufmann P. 2003. Coupled runoff simulations as validation tools for atmospheric models at the regional scale. *Q. J. R. Meteorol. Soc.* **129**: 673–692.

Jolliffe IT, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley: Chichester, England. 240 pp.

Jones MS, Colle BA, Tongue JS. 2007. Evaluation of a mesoscale short-range ensemble forecast system over the northeast United States. *Weather and Forecasting* **22**: 36–55.

Kain JS. 2004. The Kain–Fritsch convective parameterization: an update. *J. Appl. Meteorol.* **43**: 170–181.

DOI: 10.1002/qj

Kain JS, Fritsch KM. 1992. The role of the convective 'trigger function' in numerical forecasts of mesoscale convective systems. *Meteorol. Atmos. Phys.* **49**: 93–106.

Leith CE. 1974. Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.* **102**: 409–418.

MEDIS. 2006. 'Towards sustainable water use on Mediterranean islands: addressing conflicting demands and varing hydrological, social and economic conditions'. Münster, Germany. Final Report, European Project Contract No. EVK1-CT-2001-00092. 285 pp. (Available at http://www.uni-muenster.de/Umweltforschung/medis/index.html.)

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models. Part I: A discussion of principles. *J. Hydrol.* **10**(3): 282–290.

Pessoa LM, Brass RL, Williams ER. 1993. Use of weather radar for flooding forecasting in the Sieve river basin: A sensitivity analysis. *J. Appl. Meteorol.* **32**: 462–475.

Ranzi R, Bacchi B, Grossi G. 2003. Runoff measurements and hydrological modelling for the estimation of rainfall volumes in an Alpine Basin. *Q. J. R. Meteorol. Soc.* **129**: 653–672.

Reisner J, Rasmussen RJ, Bruintjes RT. 1998. Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Q. J. R. Meteorol. Soc.* **124B**: 1071–1107.

Riggs HC. 1976. A simplified slope-area method for estimating flood discharges in channels. *J. Res. US Geol. Surv.* **4**: 285–291.

Romero R, Sumner G, Ramis C, Genovés A. 1999. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *Int. J. Climatol.* **19**: 765–785.

Romero R, Doswell CA, Ramis C. 2000. Mesoscale numerical study of two cases of long-lived quasistationary convective systems over eastern Spain. *Mon. Weather Rev.* **128**: 3731–3751.

Siccardi F. 1996. Rainstorm hazards and related disasters in the western Mediterranean region. *Rem. Sens. Rev.* **14**: 5–21.

Stensrud DJ, Fritsch JM. 1994a. Mesoscale convective systems in weakly forced large-scale environments. Part II: Generation of a mesoscale initial condition. *Mon. Weather Rev.* **122**: 2068–2083.

Stensrud DJ, Fritsch JM. 1994b. Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Weather Rev.* **122**: 2084–2104.

Stensrud DJ, Bao J-W, Warner TT. 2000. Using initial and model physics perturbations in short-range ensemble simulations of mesoscale convective events. *Mon. Weather Rev.* **128**: 2077–2107.

Tribbia JJ, Baumhefner DP. 1988. The reliability of improvements in deterministic short-range forecasts in presence of initial state and modeling deficiencies. *Mon. Weather Rev.* **116**: 2276–2288.

USACE-HEC. 1998. 'HEC-HMS Hydrologic Modeling System user's manual'. US Army Corps of Engineers Hydrologic Engineering Center, Davis, California, USA. 188 pp.

USACE-HEC. 2000. 'Hydrologic Modeling System HEC-HMS: Technical reference manual'. US Army Corps of Engineers Hydrologic Engineering Center, Davis, California, USA. 157 pp.

USDA. 1986. 'Urban hydrology for small watersheds'. Technical release 55 of the Nature Resources Conservation Service. US Department of Agriculture. Washington DC, USA. 164 pp.

Von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press. 494 pp.

Wang W, Seaman NL. 1997. A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Weather Rev.* **125**: 252–278.

Wilks D. 1999. Multisite downscaling of daily rainfall with a stochastic weather generator. *Climate Res.* **11**: 125–136.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, second edition. Elsiever. 627 pp.

Williams GP. 1978. Bankfull discharge of rivers. *Water Resour. Res.* **14**(6): 1141–1154.

Zhang DL, Anthes RA. 1982. A high resolution model of the planetary boundary layer: Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteorol.* **21**: 1594–1609.

Zhang DL, Fritsch JM. 1986. Numerical simulation of the meso-β scale structure and evolution of the 1977 Johnstown flood. Part I: Model description and verification. *J. Atmos. Sci.* **43**: 1913–1943.

Zhang Y, Smith JA. 2003. Space-time variability of rainfall and extreme flood response in the Menomonee river basin, Wisconsin. *J. Hydrometeorol.* **4**: 506–517.

Zhang DL, Zheng WZ. 2004. Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *J. Appl. Meteorol.* **43**: 157–169.

DOI: 10.1002/qj