A Comparison of Ensemble Strategies for Flash Flood Forecasting: The 12 October 2007 Case Study in Valencia, Spain

A. AMENGUAL AND D. S. CARRIÓ

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

G. RAVAZZANI

Gruppo di Idrologia Fisica, Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milan, Italy

V. HOMAR

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

(Manuscript received 1 December 2016, in final form 17 January 2017)

ABSTRACT

On 12 October 2007, several flash floods affected the Valencia region, eastern Spain, with devastating impacts in terms of human, social, and economic losses. An enhanced modeling and forecasting of these extremes, which can provide a tangible basis for flood early warning procedures and mitigation measures over the Mediterranean, is one of the fundamental motivations of the international Hydrological Cycle in the Mediterranean Experiment (HyMeX) program. The predictability bounds set by multiple sources of hydrological and meteorological uncertainty require their explicit representation in hydrometeorological forecasting systems. By including local convective precipitation systems, short-range ensemble prediction systems (SREPSs) provide a state-of-the-art framework to generate quantitative discharge forecasts and to cope with different sources of external-scale (i.e., external to the hydrological system) uncertainties. The performance of three distinct hydrological ensemble prediction systems (HEPSs) for the small-sized Serpis River basin is examined as a support tool for early warning and mitigation strategies. To this end, the Flash-Flood Event–Based Spatially Distributed Rainfall–Runoff Transformation–Water Balance (FEST-WB) model is driven by ground stations to examine the hydrological response of this semiarid and karstic catchment to heavy rains. The use of a multisite and novel calibration approach for the FEST-WB parameters is necessary to cope with the high nonlinearities emerging from the rainfall–runoff transformation and heterogeneities in the basin response. After calibration, FEST-WB reproduces with remarkable accuracy the hydrological response to intense precipitation and, in particular, the 12 October 2007 flash flood. Next, the flood predictability challenge is focused on quantitative precipitation forecasts (QPFs). In this regard, three SREPS generation strategies using the WRF Model are analyzed. On the one side, two SREPSs accounting for 1) uncertainties in the initial conditions (ICs) and lateral boundary conditions (LBCs) and 2) physical parameterizations are evaluated. An ensemble Kalman filter (EnKF) is also designed to test the ability of ensemble data assimilation methods to represent key mesoscale uncertainties from both IC and subscale processes. Results indicate that accounting for diversity in the physical parameterization schemes provides the best probabilistic high-resolution QPFs for this particular flash flood event. For low to moderate precipitation rates, EnKF and pure multiple physics approaches render undistinguishable accuracy for the test situation at larger scales. However, only the multiple physics QPFs properly drive the HEPS to render the most accurate flood warning signals. That is, extreme precipitation values produced by these convective-scale precipitation systems anchored by complex orography are better forecast when accounting just for uncertainties in the physical parameterizations. These findings contribute to the identification of ensemble strategies better targeted to the most relevant sources of uncertainty before flash flood situations over small catchments.

Corresponding author e-mail: Arnau Amengual, arnau.amengual@uib.es

DOI: 10.1175/JHM-D-16-0281.1

1. Introduction

Flash floods are among the most devastating natural hazards in terms of human, social, and economic losses. These extreme events can happen extraordinarily rapidly, and the response time for any preventive measure ought to be short. One major scientific challenge of the international Hydrological Cycle in the Mediterranean Experiment (HyMeX; www.hymex.org) program is to improve the understanding of hydrometeorological extremes in the Mediterranean (Drobinski et al. 2014). The Spanish Mediterranean is a flash flood–prone region during late summer and early autumn as high precipitation rates persist for several hours over individual basins. This persistence is often associated with prominent orography that anchors quasi-stationary mesoscale convective systems (MCSs; Doswell et al. 1996; Kolios and Feidas 2010). Furthermore, the particular geographical settings of this semiarid region, with many small- to medium-sized steep and densely urbanized coastal catchments, further reduce the hydrological response times and increase flood risks. In addition, most of these rivers are often dry during the warm season, exacerbating unexpected and extensive flood damage (Camarasa Belmonte and Segura Beltrán 2001; Amengual et al. 2007, 2015).

The 12 October 2007 flash flood in Valencia, eastern Spain, is a paradigmatic example of the hazardous consequences of rapid flow increases. That day, the central-eastern part of Valencia was impacted by long-lasting convective rainfall that affected most of its internal catchments, resulting in serious material and human damages. Specifically, we focus on the Serpis River basin, which is small in size and responds quickly to extreme rainfall events (Figs. 1, 2). Our first objective is to examine the hydrological response of this semiarid and karstic basin to intense precipitation and, in particular, to the 12 October 2007 flash flood. Persistent low antecedent soil water contents and high soil moisture capacities are characteristic of Mediterranean Spain at the end of the warm season. Additional hydrological uncertainties arise as heavy rainfalls and large precipitation amounts over karstic areas result in a highly nonlinear rainfall–runoff transformation (Borga et al. 2007).

Hydrological forecasting systems based only on rainfall observations do not provide forecasts with sufficient lead time to take effective precautionary civil protection measures. The use of quantitative precipitation forecasts (QPFs) by short-range and high-resolution numerical weather prediction (NWP) models is a primary tool to further extend the hydrometeorological forecasting lead times beyond the watershed response times. Nowadays, state-of-the-art convection-permitting NWP models realistically capture the initiation and intensification of convectively driven rainfalls with similar spatial and temporal scales to the flash flood–prone catchments. QPFs can be directly used to drive rainfall–runoff models without the need to implement additional downscaling procedures (Verbunt et al. 2007; Amengual et al. 2008; Vincendon et al. 2011; Addor et al. 2011).

However, the accurate numerical prediction of deep moist convective phenomena is challenging owing to the imperfect representation of several atmospheric processes leading to extreme precipitation rates: convection, planetary boundary layer (PBL), land physics, and moist microphysical processes (Stensrud et al. 2000; Jankov et al. 2005; Amengual et al. 2008; Tapiador et al. 2012). Not only is the parameterization of physical processes inexact, but any misrepresentation of the atmospheric state across the relevant scales strongly penalizes the quality of the forecasts in such nonlinear systems (Toth and Kalnay 1993; Mullen and Baumhefner 1988; Houtekamer and Derome 1995; Du et al. 1997). Indeed, errors of any origin can grow rapidly during the quantitative precipitation forecasting and steer toward misleading predictions, especially when fast-growing modes, such as those leading mesoscale convective developments, are dominant for the predicted field. Therefore, QPF is highly sensitive to errors in the initial conditions (ICs), lateral boundary conditions (LBCs), and model physical parameterizations.

Short-range ensemble prediction systems (SREPSs) aim at forecasting the probability of weather extremes with accuracy, reliability, and precision. Uncertainties in the representation of the atmospheric state are most often encompassed by running NWP models with perturbed ICs/LBCs (Buizza 2003; Grimit and Mass 2007). Similarly, the subspace of physical parameterization uncertainties is explored by running combinations of subgrid schemes, usually considered equally skillful. Short-range QPFs can also be improved by applying data assimilation (DA) techniques (Kistler et al. 2001; Uppala et al. 2005). DA algorithms aim to determine the atmospheric state and associated uncertainties by coupling model forecast information with multiple sources of observations and their respective errors. The ensemble Kalman filter (EnKF) is an ensemble DA approach to nonlinear Kalman filtering based on Monte Carlo techniques (Evensen 2003; Guillijns et al. 2006), which has been shown to improve the sampling of the IC error space in both meso- and storm-scale ensemble data assimilation systems (Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Tanamachi et al. 2013; Marquis et al. 2014; Sippel et al. 2013). The EnKF has an unquestionable potential to produce valuable hydrometeorological predictions in Mediterranean Spain, as it combines the skill for an appropriate mesoscale ensemble generation strategy and the transference of information from land regions toward

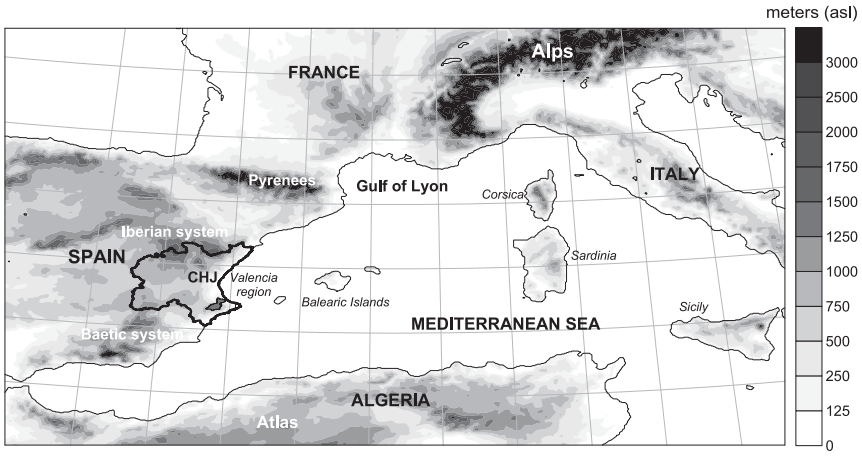

FIG. 1. Configuration of the computational domain used for the WRF numerical simulations. Main geographical features mentioned in the text are shown. The thick continuous black line shows the CHJ region where the Serpis River basin is located (highlighted in shaded gray).

relatively sparsely observed maritime areas (Carrió and Homar 2016). Both aspects pose serious challenges to forecasters in the region.

The distribution of plausible atmospheric states, represented by SREPSs, are used to build hydrological ensemble prediction systems (HEPSs) in order to convey these external-scale uncertainties down to the hydrological system. That is, the inclusion of independent information from a distribution of atmospheric scenarios aims at increasing the skill of HEPSs. However, the identification of the most suitable methods for generating HEPSs and the quantification of their added value are still under investigation (Cloke and Pappenberger 2009; see special issue of *Hydrological Processes*, 2013, Vol. 27, No. 1). The second objective of this study is to evaluate the predictive skill of three distinct ensemble generation strategies for the 12 October 2007 flash flood. To this end, we build ensembles for short-range flash flood forecasting purposes based on perturbed ICs/LBCs (PILB), multiple physical schemes (MPS), and EnKF ensemble techniques. The experiments not only shed light on the most relevant sources of uncertainty in hydrometeorological modeling, but also contribute to the discussion about the optimal design of HEPSs in small-sized Mediterranean river basins affected by heavy rainfall conductive to flash floods.

The rest of the paper is structured as follows. Section 2 consists of a brief description of the study area and the observational networks. Section 3 presents the hydrometeorological episode. The description of the PILB, MPS, and EnKF ensembles is provided in section 4. The hydrological tools and the basin characterization are described in section 5. Results are discussed in section 6, and the final section summarizes the main conclusions and provides further remarks.

2. The study area and databases

a. Overview of the Serpis River basin

The Serpis River basin is one of the Mediterranean catchments managed by the Confederación Hidrográfica del Júcar (CHJ) demarcation (Figs. 1, 2). CHJ administers an extension of 42 851 km^2 over east-central Spain and comprises most of the Spanish central rivers that flow into the Mediterranean Sea, with Júcar being the most important. The Serpis River basin has an extension of 802.6 km^2 and a length of 74.5 km at the basin outlet in the city of Gandia. The catchment extends from the northeasternmost part of the Baetic system, then flows northeasterly through a set of narrow gullies toward the coastal plain of Gandia to finally end in the Mediterranean. Maximum heights are roughly 1450 m in the headwaters and then pass through a height transition until 300–700 m in the middle basin (Fig. 2). The upper and middle catchment is principally formed by karstified limestone and marls. The lower basin is formed by carbonate strata and by alluvial quaternary deposits at the river mouth (Delgado et al. 2006). The number of inhabitants in the basin is nearly 230 000. Alcoi and Gandia are the main urban areas, with more than 60 000 and 75 000 inhabitants, respectively. Main land uses are forest (48.8%), agricultural (47.1%), and urban (3.2%).

The Valencia region has a semiarid Mediterranean climate with an average annual precipitation ranging from 300 to above 1000 mm. Seasonal rainfall distribution is typical of a Mediterranean region, with a period of summer drought and wet periods mainly in autumn and spring. Summer and autumn episodic heavy rainfalls can account for a very large fraction of

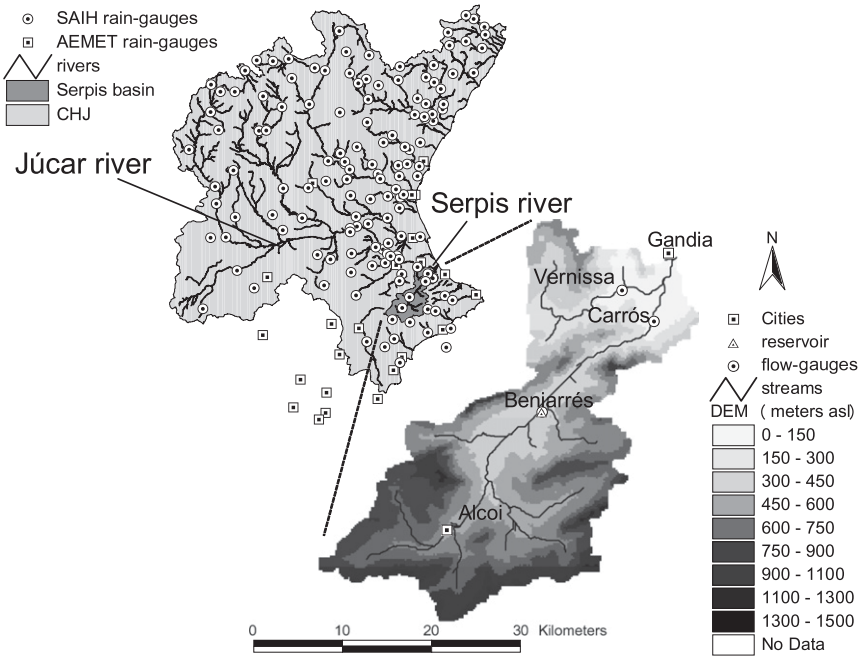

FIG. 2. At the upper left, a distribution of the rain gauges from the SAIH (131 stations) of the CHJ and AEMET (25 stations). It includes a total of 156 automatic rainfall stations distributed over an area of 45 000 km^2. The Serpis River basin is highlighted in shaded dark gray. At the lower right, a digital terrain model of the Serpis River basin with a cell size of 200 m. Main tributaries, stream gauges, and the reservoir are shown. Also included are locations mentioned in the text.

the annual amounts (Romero et al. 1998; Pastor et al. 2010). The Serpis River basin is located in one of the rainiest regions in Valencia, with average annual precipitation ranging from 700 to 900 mm, owing to its particular geographical setting. The catchment is situated between the Baetic mountain ranges, which are oriented from southwest to northeast, and the southeast to northwest coastline (Fig. 1). The east to northeast facing of the basin produces an efficient rainfall response to the entrance of easterly and northeasterly moist flows, mainly associated with subsynoptic-scale rain-bearing Mediterranean cyclones. These episodes bring very copious and convectively driven rainfall. The flow regime of the Serpis catchment is typical of the semiarid Mediterranean Spain, passing from large periods of very low flows to sporadic flash floods. Being aware of this highly irregular river regime, the CHJ hydraulic division built a reservoir with a capacity of 27.0 cubic hectometers (hm^3) for water supply and flood control purposes at Beniarrés town, located in the middle of the watershed (Fig. 2).

b. *Meteorological and hydrological data*

We analyze data from 156 automatic rain gauges, belonging either to the Automatic Hydrologic Information System (SAIH) network of the CHJ or to the Spanish Agency of Meteorology [Agencia Estatal de Meteorología (AEMET)]. These stations provide 5-min accumulations and cover the entire CHJ (Fig. 2), and 40 of them lie within the Serpis River basin or near its close vicinity. Series of 2-m temperature from 12 additional automatic ground stations of the AEMET network are used as well. Raw runoff data at 5 min intervals are also available for three flow gauges located along the catchment. These stream gauges are integrated in the SAIH network and are deployed 1) at Rótova city, in the Vernissa affluent (labeled as Vernissa by the CHJ); 2) just upstream of the Beniarrés reservoir (Beniarrés); and 3) at Assut d'en Carrós close to Villalonga town (Carrós) in the Serpis River. Their respective drainage areas are 113.0, 505.3, and 594.2 km^2 (Fig. 2).

3. Description of the 11–12 October 2007 hydrometeorological episode

During 9–10 October 2007, an Atlantic upper-level closed low was displaced from northwestern France and south of the British Islands toward the western Mediterranean, moving over the eastern Iberian Peninsula on

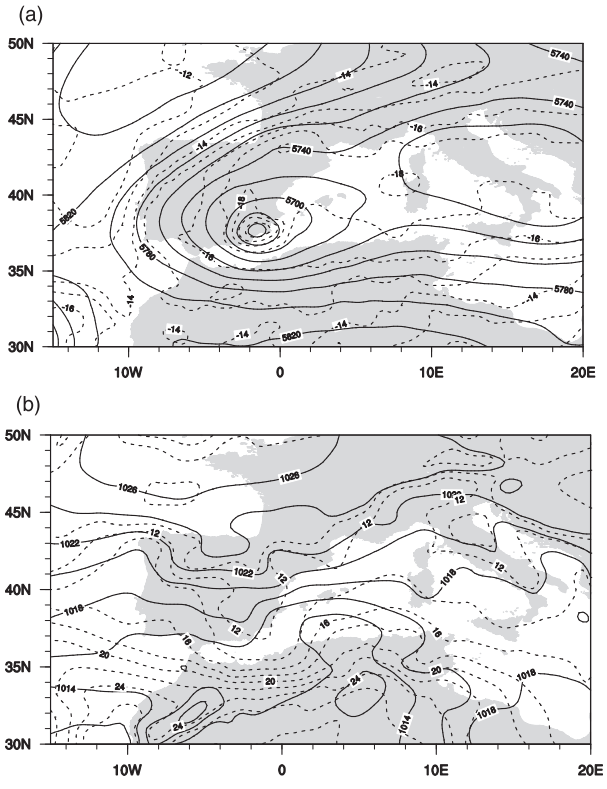

FIG. 3. ECMWF analysis at 0000 UTC 12 Oct 2007: (a) geopotential height (gpm; solid) and temperature (°C; dashed) at 500 hPa and (b) mean sea level pressure (hPa; solid) and temperature (°C; dashed) at 925 hPa.

11 October. At the surface, a strong anticyclone remained stationary across western and northern Europe on 11 October, resulting in an easterly flow advection from central-eastern Europe to the western Mediterranean for the next 2 days (Fig. 3). The sensible and latent heat fluxes from the relatively warm sea surface of the Mediterranean Sea increased the convective available potential energy (CAPE) of this overlying air mass. Together with the intrusion of high lapse rates in the lower to midtroposphere, the intense air–sea moisture exchanges, the low-level flow channeling, and uplift from the complex orography and the land–sea contrasts promoted the conditional instability of the air and its elevation above the lifting condensation and free convection levels, resulting in the triggering of deep moist convective activity. Observed 42-h accumulated precipitation (from 0000 UTC 11 October to 1800 UTC 12 October), exceeded 400 mm in two rain gauges, 300 mm in three additional stations, and 200 mm in seven additional ones. The most intense rates were recorded during the early morning of 12 October, with hourly accumulations surpassing 90 mm (Fig. 4a). More details about this episode can be found in Pastor et al. (2010), including surface and satellite measurements and NWP model simulations highlighting the importance of orography as the triggering mechanism for the development of the quasi-stationary convective systems.

Subsequent flooding from the torrential precipitation resulted in one fatality, 40 people rescued, and 243 evacuated, in addition to 1200 damaged dwellings and the collapse of several bridges. Economic losses were estimated at over EUR 100 million. Regarding the Serpis River basin, 42-h rainfall accumulations reached 330 mm over the northernmost part and the areal-averaged accumulation was close to 200 mm (Table 1; Fig. 4a). At Vernissa, two almost consecutive maximum discharges of 206.9 and 314.4 $m^3 s^{-1}$ were recorded at 0700 and 1000 UTC 12 October, with times to peak of 2 and 5 h, respectively (Fig. 5a). The very steep slopes of the rising limbs denote the extraordinary increase of the discharge rates. At Beniarrés and Carrós, peak flows at 1400 UTC 12 October were 255.9 and 201.4 $m^3 s^{-1}$, respectively (Figs. 5b,c). Thus, the peak discharge at Beniarrés was significantly abated by the reservoir prior to its downstream propagation. Note the remarkable magnitude of the flash flood at Vernissa, where the peak discharge for a 25-yr return period $Q_{p_{25}}$ is 167 $m^3 s^{-1}$ (MAGRAMA 2011).

4. Meteorological tools

Atmospheric numerical simulations for this event are produced with the Weather Research and Forecasting (WRF) Model, version 3.4 (Skamarock et al. 2008). We define a single computational domain of 767 × 575 grid points centered in the western Mediterranean and spanning the entire Mediterranean Spanish coast (Fig. 1). A horizontal resolution of 2.5 km, 50 vertical levels, and an integration time step of 12 s is used in all WRF runs, which allow for deep moist convective systems with a relevant entity to be explicitly resolved (Weisman et al. 1997; Bryan et al. 2003; Roberts and Lean 2008; Zheng et al. 2016). WRF forecasts span over 42 h, from 0000 UTC 11 October to 1800 UTC 12 October 2007. This time period encompasses the initiation phase and the mature evolution of the convective systems for this episode. Finally, hourly QPF outputs from the experimental atmospheric ensembles force the Flash-Flood Event–Based Spatially Distributed Rainfall–Runoff Transformation–Water Balance (FEST-WB) model, resulting in the corresponding experimental HEPSs examined here.

a. PILB experiment

The operational European Centre for Medium-Range Weather Forecasts ensemble prediction system (ECMWF-EPS) aims at sampling the distribution

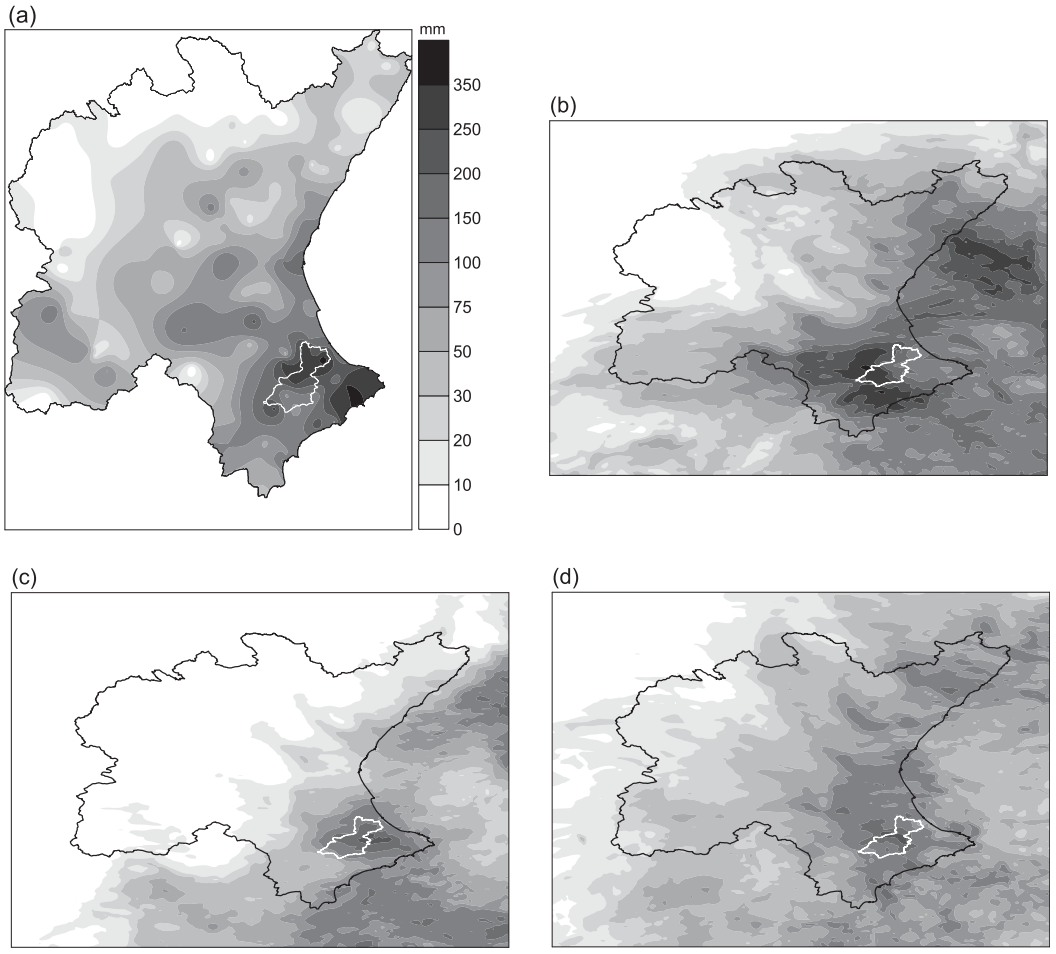

FIG. 4. Accumulated precipitation in the CHJ demarcation according to (a) rain gauges, (b) MPS percentile 90, (c) PILB percentile 90, and (d) EnKF percentile 90. The 42-h accumulated precipitation is valid at 1800 UTC 12 Oct 2007. Note that the spatially distributed accumulated precipitation in (a) has been obtained after applying kriging with a linear model for the semivariogram fit.

of plausible atmospheric states, given the bulk of observational and modeled information available (Buizza and Palmer 1995; Molteni et al. 1996). In particular, the global T639L62 ECMWF-EPS (horizontal spatial resolution of ~50 km) consists of 50 members generated by perturbing a deterministic analysis with the singular vector technique plus an unperturbed (i.e., reference) forecast. Out of these 51 members, and in order to encompass the maximum number of plausible synoptic scenarios affecting the region, we dynamically downscale

TABLE 1. Main hydrometeorological features of the 12 Oct 2007 flash flood for the different hydrometric section areas of the Serpis River basin. The 42-h (from 0000 UTC 11 Oct to 1800 UTC 12 Oct) rainfall amounts are expressed as area-averaged values. In Gandia, total runoff and peak discharge have been estimated from the observation-driven runoff simulation. Note that we have accounted for runoff volume stored in the Beniarrés reservoir according to the observation-driven runoff simulation, but it has not been possible to consider its abating effect in the observed and observation-driven peak discharges at Carrós and the basin outlet.

Hydrometric section	Total precipitation (mm)	Total runoff (mm)	Peak discharge ($m^3 s^{-1}$)	Unit peak discharge ($m^3 s^{-1} km^{-2}$)	Runoff ratio
Vernissa	229.3	80.7	314.4	2.8	0.35
Beniarrés	174.2	21.9	255.9	0.5	0.13
Carrós	185.3	35.8	201.4	0.3	0.19
Gandia	196.2	47.2	537.8	0.7	0.24

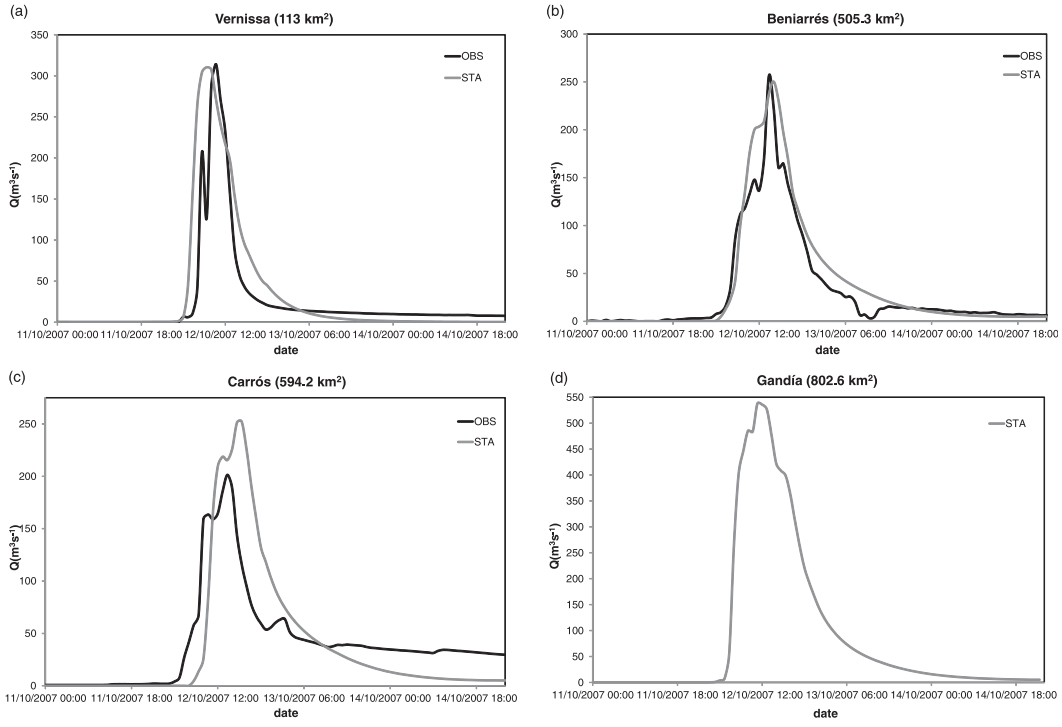

FIG. 5. Observed (OBS) and station-driven (STA) runoff simulation for the 12 Oct 2007 episode at (a) Vernissa, (b) Beniarrés, and (c) Carrós flow gauges. (d) Ground-station-driven runoff discharge is also shown at Gandia at the Serpis basin outlet.

the 20 ECMWF-EPS members exhibiting maximum IC/LBC perturbations over the numerical domain. This is an attempt to compensate for the mismatch between the synoptic-scale error growth optimization time for the singular vectors and the subsynoptic error growth more relevant for short-range basin-scale predictions (Stensrud et al. 2000; Tapiador et al. 2012).

To this end, we have implemented a method based on principal component analysis and *k*-means clustering. Using the 500-hPa geopotential height and 850-hPa temperature, all 50 ECMWF-EPS members are classified in 20 clusters, and the 20 closest members to the centroids are used as initial and boundary fields for the PILB mesoscale ensemble. Thus, we rely on the sampling of the IC/LBC uncertainty subspace provided by the global system (Marsigli 2009) and target it over the area of interest. Boundary fields are updated every 3 h in the WRF Model. Physical parameterizations are identical across PILB members and include the WRF single-moment 6-class microphysics scheme (WSM6) including graupel (Hong and Lim 2006), the 1.5-order Mellor–Yamada–Janjić (MYJ) boundary layer scheme (Janjić 1994), the Dudhia shortwave scheme (Dudhia 1989), the RRTM longwave scheme (Mlawer et al. 1997), the unified Noah land surface model (Tewari et al. 2004), and the Eta similarity surface-layer model (Janjić 1994). Note that the computational domain, the vertical levels, and this physical setting match the operational configuration routinely used by the research Meteorology Group at the University of the Balearic Islands (http://meteo.uib.es/wrf).

b. *MPS experiment*

Sensitivity analyses of mixed physics ensembles reveal that no single model configuration systematically outperforms any other one, because meteorological variables are sensitive to various processes that are simulated differently by competitive parameterization schemes (e.g., Jankov et al. 2005; Evans et al. 2012). In such convectively driven episodes, cumulus parameterizations would be a logical candidate for direct uncertainty sampling. However, as convection is explicitly resolved in our experiments, PBL and microphysical subgrid processes are the next determinant factors for deep moist convective activity. Therefore, the 20 members of the MPS experiment are constructed by using the unperturbed ECMWF-EPS member as initial and lateral boundary conditions, and by combining the following five microphysics and four PBL schemes:

- Microphysics schemes: WSM6, New Thompson, and NSSL two-moment with three cloud condensation

nuclei (CCN) prediction values of 0.5×10^9, 0.75×10^9, and $1.0 \times 10^9\,\mathrm{cm}^{-3}$.
- Planetary boundary schemes: Yonsei University (YSU), MYJ, Mellor–Yamada–Nakanishi–Niino level 2.5 (MYNN), and Total Energy–Mass Flux (TEMF).

All these schemes involve the simulation of explicitly resolved water, cloud, and precipitation processes, including mixed-phase transformations (i.e., the interaction of ice and water). However, each microphysical parameterization treats differently the interaction among six moisture species (i.e., water vapor, cloud water, rain, cloud ice, snow, and graupel); the physical processes of rain droplet production, fall, and evaporation; the cloud water accretion and autoconversion; condensation; and saturation adjustment and ice sedimentation (Skamarock et al. 2008). WSM6 solves the differences between the liquid and solid phase dependent on temperature. Water and rain are treated separately from ice and snow (Hong and Lim 2006). The New Thompson scheme adds rain number concentration and uses the Gamma distribution and dependent intercept parameters for the raindrop size distribution (Thompson et al. 2008).

On the other side, the two-moment NSSL scheme predicts average graupel particle density, allowing this to span the range from frozen drops to low-density graupel (Mansell et al. 2010). Given the specific geographical setting of the western Mediterranean Basin, we also sample the CCN uncertainty as the variability of aerosol concentration—depending on the origin of the air mass affecting the area—is clearly influential and not considered by standard tabulated aerosol parameters. That is, Saharan, continental central European, Atlantic, or purely Mediterranean air masses have radically different aerosol characteristics, affecting the moist physical processes over the Mediterranean region (Clarke et al. 1997). We have considered the above-mentioned three different levels of CCN to account for this variability and to cope with the inaccuracy of using tabulated CCN data (Hudson 1993).

PBL schemes are used to parameterize the subgrid turbulent vertical fluxes of heat, momentum, and moisture within the boundary layer and throughout the atmosphere (Pielke and Mahrer 1975). The PBL representation is a determinant factor in accurately simulating mesoscale weather phenomena owing to the critical role that these fluxes exert in the unfolding of severe phenomena. The choice of a PBL scheme can substantially affect temperature and moisture profiles in the lower troposphere and the effects of turbulence in daytime convective conditions (Hu et al. 2010; Coniglio et al. 2013).

The YSU scheme is a first-order, nonlocal scheme with a countergradient term in the eddy-diffusion equation, which enhances mixing in the stable boundary layer (Hong et al. 2006). On the other side, the MYJ scheme employs a 1.5-order turbulence closure model to represent turbulence above the surface layer. This scheme determines eddy-diffusion coefficients from prognostic turbulent kinetic energy (Janjić 1994). MYNN treats condensation physics in the boundary layer by considering liquid-water potential temperature and total water content. In addition, MYNN allows for partial condensation in a model grid to assure proper interaction with microphysics and radiation (Nakanishi and Niino 2006). Additionally, the TEMF scheme uses eddy diffusivity and mass flux concepts to determine vertical mixing, and it includes a subgrid total energy prognostic variable, giving more realistic profiles for shallow convection (Angevine et al. 2010).

c. *EnKF experiment*

The Data Assimilation Research Testbed (DART; Lanai version) package, developed by the National Center for Atmospheric Research (NCAR; Anderson et al. 2009), was used to implement an ensemble data assimilation system. EnKF is a sequential filter method that minimizes the variance of the resulting atmospheric analysis, given the errors of the prior fields and observational data. EnKF is typically implemented as a cyclic data assimilation system that consists of two elemental phases (Houtekamer and Mitchell 1998): an assimilation step, in which available observations are ingested accounting for the ensemble and observation covariances, and a forecast step, consisting in the time advancing of the newly generated ensemble of states to the next assimilation step. One of the most attractive properties of an EnKF is the background error covariance calculation, derived dynamically from the ensemble covariance, which dramatically reduces its calculation costs with respect to climatologically based approaches.

First, we have selected the 20 ECMWF-EPS members exhibiting maximum IC/LBC perturbations over the numerical domain on 1200 UTC 10 October 2007. Next, our EnKF experimental ensemble design consists of hourly assimilation windows over the next 12 h (i.e., until 0000 UTC 11 October, performing a total of 13 assimilation steps). Finally, the resulting analyses at 0000 UTC are integrated forward for 42 h. In addition to encompassing errors in the initial and lateral boundary conditions, the EnKF ensemble accounts for uncertainties in the representation of subgrid physical processes as well. For the PBL, the same schemes as for the MPS experiment are used. However, the differing number of microphysical species in the parameterizations used in the MPS ensemble impedes its use in the assimilation phases of the EnKF experiment. Thus, the diversity in the microphysical processes is accounted for

using just the NSSL scheme, but with five different CCN concentrations: 0.1×10^9, 0.25×10^9, 0.5×10^9, 0.75×10^9, and $1.0 \times 10^9\,\mathrm{cm}^{-3}$.

The observational databases were provided by the Meteorological Assimilation Data Ingest System (MADIS) of the National Oceanic and Atmospheric Administration's National Weather Service (NOAA/NWS). In particular, we use radiosonde, METAR, marine, and ACARS data. Further details on the design of the EnKF ensemble are provided by Carrió and Homar (2016). Also note that all three ensemble strategies are designed to mimic an operational forecasting framework. Operational EnKF systems separate clearly the assimilation process from the forecast process, running almost independently. Here, all three experimental ensembles could have been run operationally shortly after 0000 UTC 11 October, rendering a comparable set of ensemble forecasts.

5. Hydrological tools

a. Hydrological model and basin characterization

We simulate the hydrological response of the Serpis River basin with the FEST-WB model (Rabuffetti et al. 2008). FEST-WB is physically based and accounts for evapotranspiration, infiltration, surface runoff, subsurface flow, and flow routing. The computational domain is discretized with a regular-squared mesh. The hydrological model computes soil moisture fluxes by solving the water balance equation at each grid point. In particular, the evolution of the soil moisture θ_{ij} for the generic point at (i, j) is

$$\frac{\partial \theta_{ij}}{\partial t} = \frac{1}{Z_{ij}}(P_{ij} - R_{ij} - D_{ij} - \mathrm{ET}_{ij}), \tag{1}$$

where P is the precipitation rate, R is the runoff flux, D is the drainage flux, ET is the evapotranspiration rate, t is time, and Z is the soil depth. Runoff is calculated according to a modified Soil Conservation Service curve number (SCS-CN; USDA 1986) method extended for continuous simulation (Ravazzani et al. 2007, 2016). Thus, the maximum potential retention S is updated at the beginning of a storm as a linear function of the degree of saturation ε. That is,

$$S = S_1(1 - \varepsilon) + S_3\varepsilon, \tag{2}$$

where S_1 and S_3 are the values of S when the soil is dry and wet (i.e., antecedent moisture condition I and III, respectively). The actual evapotranspiration is calculated as a fraction of the potential rate tuned by the beta function that, in turn, depends on soil moisture content (Montaldo et al. 2003). Potential evapotranspiration is computed according to a modified version of the Hargreaves–Samani equation (Ravazzani et al. 2012). The surface and subsurface flow routing is based on the Muskingum–Cunge method in its nonlinear form with the time-variable celerity (Montaldo et al. 2007). The minimum amount of input atmospheric data required to run FEST-WB is precipitation and temperature, which are interpolated from the ground stations to the model grid points by means of a kriging algorithm applied using a linear model for the semivariogram fit.

The physiographic basin characteristics required for the implementation of FEST-WB include a digital elevation model (DEM) and land-use and lithology maps. DEM and soil properties layers are provided by the CHJ hydraulic division and the Spanish Geological Survey (Instituto Geológico y Minero de España) databases. The Coordinated Information on the Environment (CORINE) land-cover dataset provides the land-use information (Bossard et al. 2000), and the SCS-CN map is obtained from the Spanish Ministry of Agriculture, Food and Environment. From these layers, the following basin parameters are derived: flow direction, slope, aspect, residual and saturated soil moisture, pore size distribution index, saturated hydraulic conductivity, wilting point, field capacity, and soil depth.

Because of the Beniarrés reservoir, the Serpis River basin cannot be modeled under a natural regime (Fig. 2), as it results in important hydrograph diffusion effects on the flood waves. Flow routing through the detention pond is modeled by means of the fourth-order Runge–Kutta method for the level pool scheme (Ravazzani et al. 2014). Initial elevation, storage capacities, maximum outflows, and water elevations have been provided by the CHJ hydraulic division. Finally, continuous observation–driven FEST-WB simulations are carried out at hourly time steps and with a horizontal grid resolution of 200 m.

b. Model calibration strategy

Model calibration focused on peak discharge and timing as well as runoff volume, which are strongly dependent on infiltration and routing processes. In the semiarid Mediterranean Spain, sparse vegetation together with torrential convective precipitation, which easily exceeds the high initial soil infiltration capacity, favor fast Hortonian flows and rapid flow velocities in the river streams during the warm season (Camarasa Belmonte and Segura Beltrán 2001; Amengual et al. 2007, 2009, 2015). As pointed out by Borga et al. (2007), heterogeneities in the hydraulics of the basin response to flash floods arise as a consequence of the systematic decrease of catchment reaction with increasing rainfall amounts. Two major factors regulating these heterogeneities are the expansion of the

stream network to unchanneled topographic elements during flash flooding and the increase of flow velocity with discharge. Consequently, we have first calibrated the Strickler coefficients associated with the hillslope flow routing $k_{s_{hs}}$ and channel flow routing k_{s_c}, thus modulating the overland flow velocity V_{hs} and channel flow velocity V_c.

Initially dry soils and high infiltration capacities enhance the nonlinear response of runoff to intense precipitation and large rainfall amounts over the Serpis River basin, resulting in a noticeable mitigation of the magnitude of the runoff discharges and volumes (Table 1). These high initial infiltration losses and large soil moisture storage capacities are associated with the predominant presence of limestone, marl, and carbonate strati in the basin, mainly favoring the recharge of deep aquifers (Camarasa Belmonte and Segura Beltrán 2001). As mentioned in section 5a, the runoff generation processes are described according to a modified SCS-CN method extended for continuous simulation. Following the procedure by Borga et al. (2007), we have selected the following infiltration parameters for calibration to better encompass the strong nonlinearities: curve number (CN), infiltration storativity S_0, initial abstraction ratio λ, and saturated soil hydraulic conductivity K_s.

Recall that in the original SCS-CN formulation, S is a site storage index related to the CN as (Ponce and Hawkins 1996):

$$S = S_0\left(\frac{100}{\mathrm{CN}} - 1\right). \tag{3}$$

The use of S_0 as a calibration parameter allows employing the spatial distribution of CN values to correctly simulate the observed flood water balance (Borga et al. 2007). Note that the default value of S_0 is 254 mm. The initial abstraction is specified as a percentage of S:

$$I_a = \lambda S. \tag{4}$$

In the original formulation, $\lambda = 0.2$ is considered as the standard value (Ponce and Hawkins 1996). However, given the aforementioned physical specificities of the Serpis River basin, we have considered λ as a further calibration parameter as well.

FEST-WB has been subjected to calibration by comparing observation-driven simulated runoff against recorded runoff for an independent set of flash floods. To this aim, hourly observed discharge series have been first examined for the entire 2002–12 period. Three episodes have been selected for calibration tasks at Vernissa. At Beniarrés, two additional independent events have been chosen for calibration purposes. These episodes have been selected based on having similar hydrometeorological features to our case study (Table 2). Thus, the 12 October 2007 flash flood has been employed for verification purposes. Continuous observation–driven FEST-WB simulations have been carried out for the same period to extract the selected flash floods. To this end, we have employed the observational network described in section 2b, although not all the ground stations were continuously operational for the entire period. In addition, impacts on the natural river flow from real-time reservoir operations (i.e., water redistribution and diversion and actual water levels) have not been encompassed, as these data series are not available. Thus, we have modeled the Beniarrés dam by using the aforementioned technical characteristics provided by the CHJ hydraulic division.

TABLE 2. Main hydrometeorological features of the episodes for different stream gauges used for the calibration of the FEST-WB model. Total rainfall amounts are expressed as area-averaged values.

Stream gauge	Date	Total precipitation (mm)	Peak discharge ($m^3 s^{-1}$)
Vernissa	7 May 2002	232.6	190.2
	16 Apr 2003	81.5	51.3
	28 Sep 2009	156.4	90.6
Beniarrés	9 Oct 2008	116.3	94.2
	23 Nov 2011	105.0	165.2
Carrós	9 Oct 2008	122.0	48.9
	23 Nov 2011	110.0	59.7

The hydrological model has been manually calibrated by adopting the trial and error approach and by using a set of objective functions, with the aim of minimizing discrepancies in simulated peak discharge, runoff volume, and time to peak. That is, the calibration effectiveness has been evaluated by computing the relative errors in volume (EV) and peak (EP) discharges at the flow gauges. For a general assessment of a model's performance, other indices that are well known in the literature have been computed as well: the Nash–Sutcliffe efficiency criterion (NSE; Nash and Sutcliffe 1970) and the root-mean-square error (RMSE). Finally, the values of the calibrated parameters found at Beniarrés were extended to the basin outlet, as discharge observations at Carrós could not be used for calibration because they were affected by upstream dam regulation.

After calibration, an improvement of goodness-of-fit indices characterizes all hydrometric sections. That is, reasonably satisfactory results have been obtained, even though some errors still remain (Table 3). In particular, FEST-WB exhibits moderate underestimations in the observed peak discharges and flood volumes at Vernissa and Beniarrés. Table 3 displays greater errors in NSE at Carrós as a consequence of lacking precise information

TABLE 3. Mean values of the statistical indices of the peak discharge and flood volume errors before and after calibration for the observation-driven runoff simulations at the indicated stream gauges. Std devs are shown in parentheses. Negative EV and EP scores denote model underestimation.

Stream gauge	Statistical index	Before calibration	After calibration
Vernissa	NSE	0.32 (0.35)	0.41 (0.28)
	RMSE ($m^3 s^{-1}$)	14.59 (6.11)	14.62 (8.63)
	EV	−0.32 (0.40)	−0.19 (0.25)
	EP	−0.65 (0.11)	−0.25 (0.18)
Beniarrés	NSE	−0.03 (1.13)	0.51 (0.03)
	RMSE ($m^3 s^{-1}$)	21.79 (2.78)	20.45 (11.26)
	EV	1.23 (2.04)	−0.01 (0.88)
	EP	−0.26 (0.38)	−0.37 (0.39)
Carrós	NSE	−4.27 (6.05)	−1.01 (0.90)
	RMSE ($m^3 s^{-1}$)	21.25 (15.56)	15.20 (0.90)
	EV	1.06 (1.35)	−0.25 (0.47)
	EP	0.90 (0.68)	0.33 (0.55)

of real-time reservoir operations. This fact is primarily reflected in advances/delays in times to peak and in under/overestimations in runoff volumes. In general, FEST-WB exhibits an overall agreement among the observed and simulated floods. These results confirm the ability of the hydrological model to adequately simulate the hydrological response of the Serpis River basin to torrential precipitation events. The FEST-WB parameterization for the different hydrometric sections is shown in Table 4.

6. Results

a. *Observation-driven runoff simulation*

The observation-driven FEST-WB simulation (STA) for the 12 October 2007 flash flood spanned from 0000 UTC 1 August to 0000 UTC 15 October 2007. The initial warm-up period allows a good initialization of the soil moisture content. This observation-driven runoff simulation provides a basic assessment of the ability of the hydrological model to reproduce the observed streamflows for this extreme flood (Table 5, Fig. 5). An accurate peak discharge is obtained at Vernissa, albeit a noticeable overestimation of the runoff volume is produced as well. In addition, the STA simulates a peak discharge 2 h ahead of time. Despite these shortcomings and the extreme observed peak discharge (above Q_{p25}, Table 2), the model competently reproduces the flashy hydrological response to the extreme precipitation registered in this subcatchment (Fig. 5a). The STA also accurately captures the observed maximum flow and time to peak at Beniarrés, even if a remarkable overestimation in the runoff volume still remains. As expected owing to the abating effects of the Beniarrés dam, more significant inaccuracies are found at Carrós: an important overestimation in peak discharge as well as a delay of 3 h in the time to peak (Figs. 5b,c). According to this simulation, maximum peak discharge at the (ungauged) basin outlet exceeded 535 $m^3 s^{-1}$ (Fig. 5d), very close to the 10-yr return period value Q_{p10} (580 $m^3 s^{-1}$; MAGRAMA 2011).

After calibration, the FEST-WB parameterization succeeds when simulating the highly nonlinear runoff production, the Hortonian infiltration excess mechanism, and the fast times to peaks and flood wave celerities in the river channels. That is, a multisite calibration of the model parameters is necessary to account for both the diversity of physiographic and hydrologic features and heterogeneities in the hydraulics of the Serpis River basin response. The hydrological model correctly simulates small subsurface contributions as well (Fig. 5). Indeed, saturated subsurface layers are rare for flash flooding in the semiarid Mediterranean Spain, as discharge into the river channels mainly occurs through overland flow, resulting in small subsurface storm flows (Puigdefabregas et al. 1998). These results point out an effective adjustment of the observation-driven peak discharges and runoff volumes as well as a suitable dynamical formulation of the hydrological model. Therefore, the subsequent WRF-driven runoff experiments can be safely carried out, as the hydrological response of the Serpis basin to the 12 October 2007 flash flood is well simulated by FEST-WB. That is, we have confirmed the suitability of the hydrological model as a test tool for precipitation forecasts and their associated streamflow responses. This allows examining confidently the performance of the distinct HEPSs as a nonstructural tool for early warning systems.

b. *Ensemble prediction systems*

Prior to assessing the value of the hourly QPFs to drive the distributed hydrological model, we analyze the ability of each ensemble system to reproduce the general aspects of the rainfall episode. To this end, we use an Akima method (Akima 1978, 1996) to interpolate 42-h simulated precipitation amounts—from 0000 UTC 11 October to 1800 UTC 12 October—to all available rain gauges over the CHJ area. First, the skill of each ensemble member is quantified by means of Taylor diagrams (Taylor 2001). Then, the skill of the probabilistic forecasts is assessed by means of relative operating characteristic (ROC) curves (Stanski et al. 1989; Schwartz et al. 2010).

Owing to the frequently poor skill of short-range convective forecasts, the ensemble spread becomes an additional important issue to quantify predictability. Therefore, we also focus our attention on the generation

TABLE 4. Calibrated FEST-WB parameters at the different hydrometric sections. Variables CN, S_0, λ, and K_s are shown as the area-averaged values. Std devs are shown in parentheses. Note that CN(II) denotes curve numbers for normal antecedent conditions. Also note that $V_{c_{max}}$ is case dependent and tabulated values correspond to the 12 Oct 2007 flash flood.

	Hydrometric sections			
	Vernissa	Beniarrés	Carrós	Gandia
$k_{s_{hs}}$ ($m^{1/3} s^{-1}$)	15	10	10	15
k_{s_c} ($m^{1/3} s^{-1}$)	22–30	22–30	22–30	22–30
V_{hs} ($m\,s^{-1}$)	0.19	0.19	0.19	0.19
$V_{c_{max}}$ ($m\,s^{-1}$)	4.83	3.36	3.27	3.47
CN(II)	71.4 (11.2)	63.8 (10.4)	61.4 (11.7)	63.1 (10.9)
S_0 (mm)	254.0 (0.0)	280.0 (0.0)	280.0 (0.0)	272.2 (11.3)
λ	0.20 (0.00)	0.40 (0.00)	0.40 (0.00)	0.34 (0.10)
K_s ($\times 10^{-5}\,m\,s^{-1}$)	0.8 (1.4)	0.6 (1.0)	0.6 (1.0)	0.7 (1.1)

of spread by each experimental ensemble. We use the root-mean-square difference (rmsd) of the hourly simulated precipitation SP_{rmsd} (Scherrer et al. 2004), which provides a quadratic measure of the spread and highlights the presence of extremes in the ensemble. It is computed as the average of the sum of the root-mean-square difference between each ensemble member and the remaining ones as follows:

$$T_{ij} = \mathrm{rmsd}(W_i, W_j),$$
$$\Delta = \sum_i \sum_j T_{ij} \quad \text{for} \quad i < j, \quad \text{and}$$
$$\mathrm{SP}_{\mathrm{rmsd}} = \frac{2\Delta}{N(N-1)}, \tag{5}$$

where W_i and W_j are the values of the field under analysis (i.e., 42-h accumulations) for ensemble members i and j. Note that T_{ij} stands for the root-mean-square difference between ensemble members, and N is the size of the EPS. As T_{ij} is symmetric, Δ is just the sum over all $i < j$.

In general, all three experiments produce, to some extent, high 42-h accumulated rainfall amounts over the CHJ area, although the PILB and EnKF fall shorter than MPS at reaching the highest observed amounts (Fig. 4). Indeed, the high end values rendered by all three ensembles—but especially for MPS—would indicate the potential for torrential accumulations. This provides a baseline quality check for the ensemble designs proposed in this study, as all three can raise warning flags for hazardous scenarios, but the reliability of the finer details must be investigated. A key feature in the mesoscale organization of the moisture flows is the maritime impinging of the southerly low-level jet (LLJ) over the coastal slopes in the southern part of the CHJ domain (Pastor et al. 2010). While most members of the MPS experiment produce this LLJ, most members of PILB and EnKF experiments miss this important triggering mechanism. This becomes an important aspect for the accurate forecast of the maritime and inland precipitation structures of this episode.

A quantitative verification of the 42-h cumulative rainfall amounts over the whole CHJ area reveals that MPS and PILB members produce more precise spatial distributions than EnKF components, with the best members of both experiments reaching correlations as high as 0.75 over this relatively small area (Fig. 6a). Regarding the predicted amounts, MPS better captures the range of observed variability, as it correctly generates higher amounts than PILB and EnKF over a larger portion of the central and coastal parts of the CHJ. It is noteworthy that a few EnKF members also exhibit a realistic variability in the precipitation field, although with much less frequency than MPS (Fig. 4).

MPS features the largest ensemble spread, whereas PILB exhibits the lowest value due to rendering the most important underestimations in the rainfall amounts (Fig. 6b). Focusing on EnKF, it shows higher rainfall accumulations and spread than PILB, yielding a more realistic estimate of the rainfall spatial variability. Although EnKF members render moderate deterministic scores, uncertainty indications derived from the ensemble variability show promising skill.

A more detailed analysis of the ensemble spread reveals that the highest SP_{rmsd} are produced by the MPS strategy on the first surge of precipitation during the last

TABLE 5. Statistical indices of the peak discharge and flood volume errors for the 11–12 Oct 2007 flash flood and the rain-gauge-driven runoff simulation at the indicated stream gauges. Negative EV and EP scores denote model underestimation.

	NSE	RMSE ($m^3\,s^{-1}$)	EV	EP
Vernissa	0.59	38.21	0.29	−0.01
Beniarrés	0.89	17.13	0.18	−0.02
Carrós	0.30	37.78	−0.05	0.25

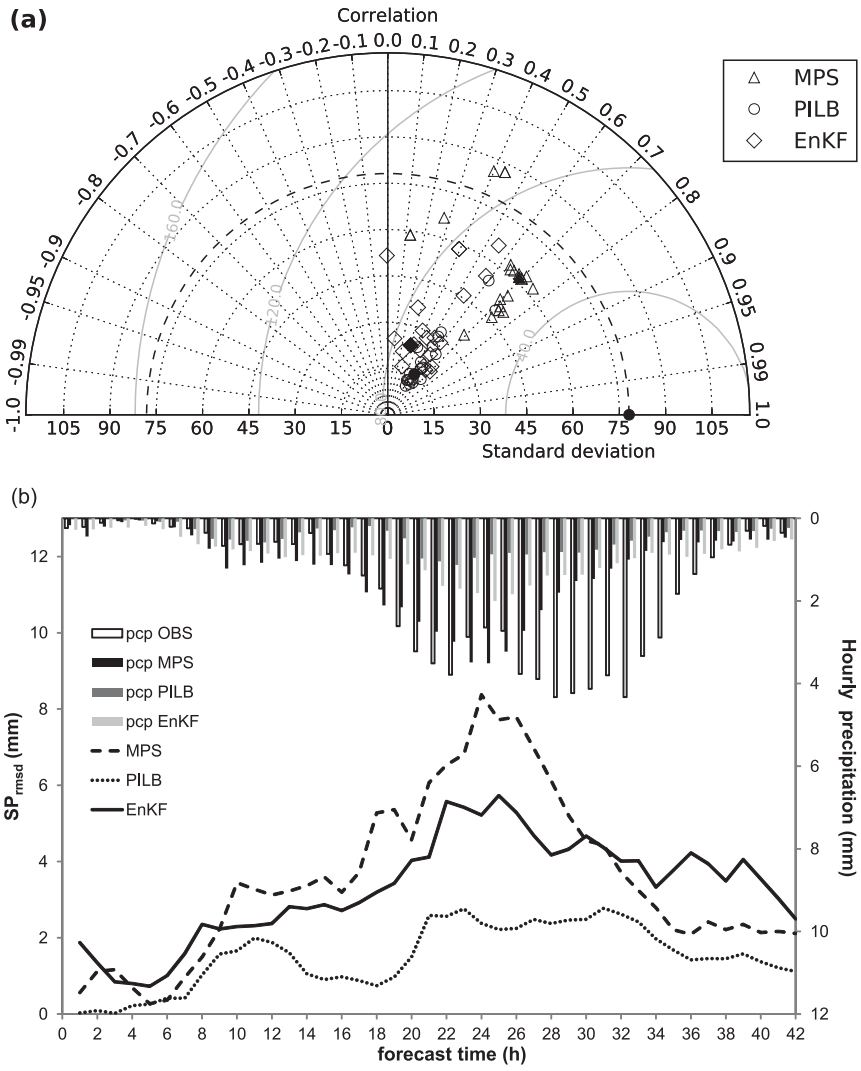

FIG. 6. (a) Taylor diagrams of the MPS, PILB, and EnKF ensembles for the 42-h accumulated precipitation over the CHJ. Filled black symbols denote ensemble medians. (b) Ensemble spread evolution of the different strategies for the hourly simulated rainfalls from 0000 UTC 11 Oct to 1800 UTC 12 Oct over the CHJ. Dashed, dotted, and continuous black lines denote the MPS, PILB, and EnKF ensembles, respectively. Hourly observed and ensemble mean precipitation is shown as vertical bars and std devs (mm) and centered RMS differences (mm) are given.

hours on 11 October (Fig. 6b). This suggests that sampling only on the uncertainty of the microphysical and boundary layer processes in organizing convection outperforms the benefits of accounting for an enhanced diversity in the ICs/LBCs or the application of a DA technique that samples both IC/LBC and physics diversity. Except for the first 8 and last 10 h of prediction, in which EnKF issues higher levels of spread under small precipitation rates, the MPS produces significantly larger spread throughout the episode. Most likely, the influence of complex geographical settings on the subsynoptic flows narrows the dispersion in the PILB experiment, whereas diversity introduced by MPS in the boundary layer and moist processes better encompass the broad range of plausible scenarios at the regional scale.

Besides the prediction of forecast uncertainty of deterministic products through the spread–skill assumption, ensembles provide a discrete statistical sample of the forecast probability density function. This precious information allows us to derive probabilistic forecasts of any dichotomous, discrete, or continuous predictands (Wilks 2006). We analyze the skill of each experimental ensemble in predicting the odds for three relevant accumulations for this episode: 50, 100, and 150 mm in 42 h.

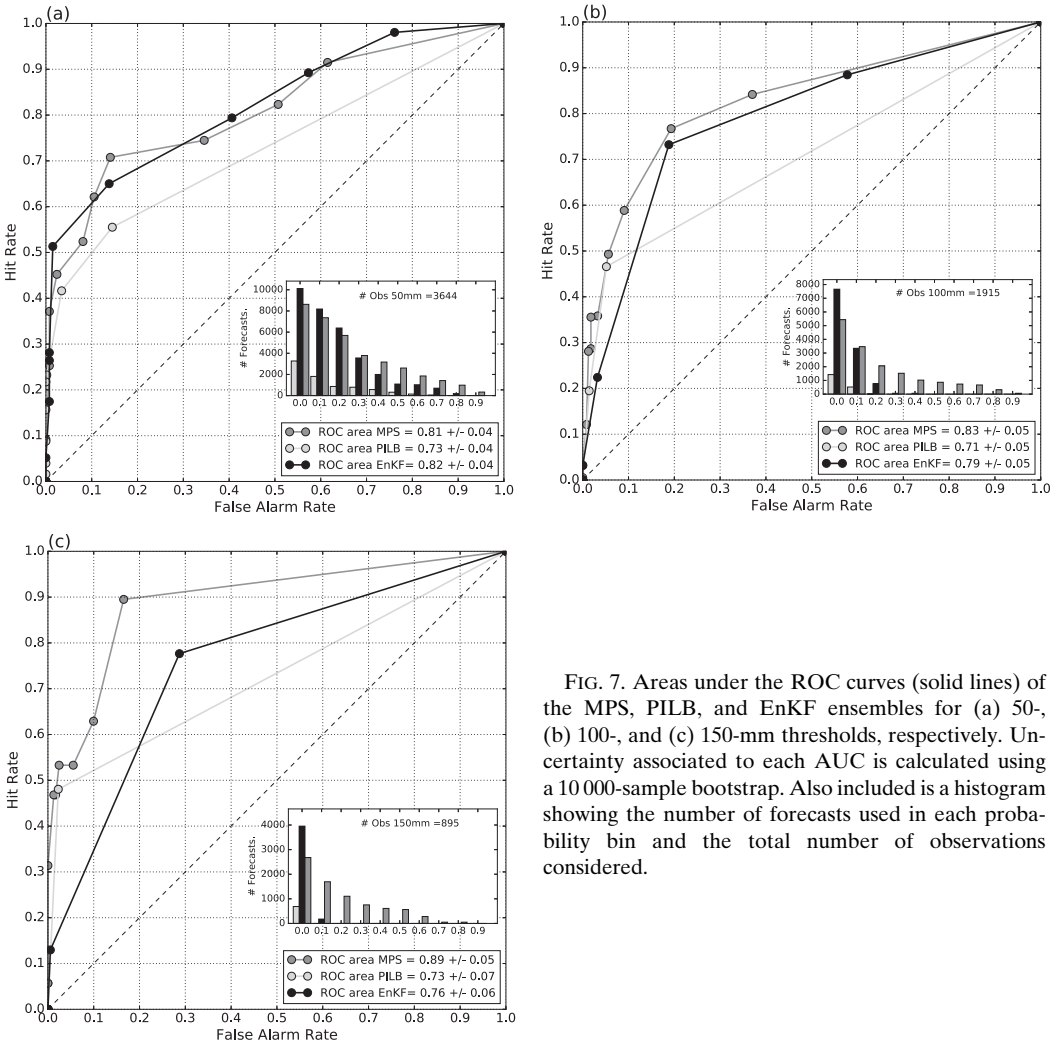

FIG. 7. Areas under the ROC curves (solid lines) of the MPS, PILB, and EnKF ensembles for (a) 50-, (b) 100-, and (c) 150-mm thresholds, respectively. Uncertainty associated to each AUC is calculated using a 10 000-sample bootstrap. Also included is a histogram showing the number of forecasts used in each probability bin and the total number of observations considered.

At each rain gauge, the probability of exceedance of the threshold is computed and verified against the dichotomous observation. ROC curves indicate the true hit rate of a probabilistic forecast at varying false alarm rates (Schwartz et al. 2010), and the area under the curve (AUC) measures the ability of the system to discriminate between the occurrence or nonoccurrence of the event.

Results show that for moderate accumulations (50 mm; Fig. 7a), EnKF and MPS are statistically indistinguishable, with AUCs exceeding 0.8. This reveals a surprisingly high quality of the probabilistic forecasts of 42-h moderate accumulations for the two experiments that account for physical processes diversity. Missing the sampling of this source of errors in the ensemble design renders significantly poorer probabilistic forecasts, as shown by the PILB AUCs. At higher precipitation thresholds (Figs. 7b,c), the number of observation–forecast pairs decreases, although verification scores are kept statistically significant for all probability levels. At the 100-mm threshold, MPS and EnKF scores are not distinguishable at the 95% confidence level, but the expected value of AUC for the MPS already shows slightly better quality than for the EnKF.

This is accentuated at the 150-mm accumulation level, for which MPS produces the most reliable probabilistic forecasts with an impressive 0.89 discrimination skill. The results rendered by EnKF and PILB for these extreme values are not statistically distinguishable at the 95% confidence level, indicating the relevance of physical parameterization diversity, especially on the microphysical processes for the correct representation of uncertainties on high-end precipitation rate scenarios. MPS not only produces competitive scores at moderate precipitation amounts but also shows remarkably high levels of accuracy for extreme accumulations (exceeding 100 and 150 mm in 42 h).

Regarding the spatial distribution of the predicted rainfall fields, the analysis at the regional scale is

informative on how each prediction system handles the episode. Evidently, an accurate hydrological forecast over a small-sized river basin, such as the Serpis, has specific requirements on the location, amount, and timing of the simulated, convectively driven precipitation systems. Focusing on catchment scale, the MPS and PILB experiments initiate precipitation over the basin 8–10 h earlier than observed, with EnKF initiating the latest, around 1400 UTC. The MPS members render realistic cumulative precipitation profiles even though they underestimate the maximum observed precipitation rates between 27- and 37-h forecast times (Fig. 8a). However, the resulting predictive guidance is excellent on basin-averaged precipitation input for the hydrological model.

On the other hand, the PILB and EnKF strategies render clear underestimations, missing the highest central precipitation rates (Figs. 8b,c). Again, this confirms the benefits of sampling only the subspace of parameterized physics uncertainty when the synoptic and large mesoscale dynamical and thermodynamical environment is sufficiently accurate in the global unperturbed member. How general this aspect is in the western Mediterranean remains for further research with other case studies in the basin. On the contrary, the PILB does not show sufficient spread over the basin, and maximum hourly simulated precipitation is persistently negatively biased.

This cumulative perspective can be misrepresentative, as errors may balance over time and provide an optimistic estimation of the predicted precipitation amounts and spreads. A much more ambitious predictability challenge is the accurate representation of hourly accumulations. When the spatial distribution and variability of hourly rainfall rates predicted by the ensemble members are verified, most MPS ensemble members stand out over the PILB and EnKF members (Fig. 9a). Again, MPS elements are generally able to produce better spatial distribution variability at the highest-frequency precipitation rates, while PILB and EnKF individual runs strongly underestimate both amounts and standard deviations. Regarding ensemble-mean values, the hourly hyetograph shows how MPS simulates the highest intensities about 2 h ahead of time and with lower precipitation rates. PILB and EnKF produce much poorer forecasts of areal-averaged hourly rainfall. In terms of ensemble spreads, MPS features the largest SP_{rmsd} values, while the PILB and, to a lesser extent, EnKF strategies present lower ensemble spreads (Fig. 9b).

c. *Hydrological ensemble prediction systems*

Hourly time series of precipitation and temperature from each experiment have been embedded in the simulation period of the FEST-WB model. The MPS-, PILB-, and EnKF-driven runoff ensembles—labeled as MPS-HEPS, PILB-HEPS, and EnKF-HEPS, respectively—have been run for a 96-h period, starting at 0000 UTC 11 October 2007. Thus, the hydrological model has run the last 54 h without forcing precipitation in order to safely encompass the 12 October flash flood and the subsequent hydrograph tails. The resulting quantitative discharge forecasts (QDFs) are compared against the observation-driven runoff simulation for the same period rather than against the observed runoffs. By doing so, we exclusively focus the analysis on uncertainties coming from the SREPS inputs, narrowing down the errors originated in the hydrological model. That is, the observation-driven runoffs are considered pseudorunoff observations, which allow focusing on the external-scale (i.e., SREPS originated) errors. In any case, observation-driven simulated runoffs were satisfactory across the basin, especially when simulating the observed peak discharges (Fig. 5).

The MPS-HEPS simulates significant flood peaks at all stream gauges, with some individual members exhibiting reasonably satisfactory accuracy (Fig. 10). Poorer results are obtained for the PILB-HEPS members and, in particular, for the EnKF-HEPS experiments: the moderate to strong hourly QPF underpredictions at the basin scale strongly impact the hydrometeorological system (Figs. 11, 12). Even if the PILB ensemble features less spread than EnKF, it better localizes the maximum hourly rainfall amounts over the basin. Clearly, the highly nonlinear hydrological response—related to the aforementioned threshold effects on runoff generation—results in an apparent degradation of QDFs for both systems. Hence, ensemble medians barely reproduce the significant observation-driven discharges at all the hydrometric sections. Next, the hourly simulated runoffs for all the HEPSs are verified against the 1-h observation-driven flow discharge at the basin outlet by means of NSE, EV, and EP. The spread generated by the different HEPSs strategies is also quantified by means of the SP_{rmsd}.

As previously suggested by member-wise verification (Figs. 6, 9), MPS-HEPS outperforms clearly with respect to PILB-HEPS and EnKF-HEPS in terms of both peak discharges and runoff volumes (Fig. 13a). However, NSE is below 0.1 for the ensemble median, owing to the noticeable underestimations in peak discharges and runoff volumes at the basin outlet. PILB-HEPS and, especially, EnKF-HEPS exhibit worse ensemble skills as a result of the aforementioned remarkable rainfall underestimations over the Serpis basin. Their ensemble dispersions are clearly lower than the MPS-HEPS variance as well (Fig. 13b). Obviously, lower hourly rainfall amounts result in less runoff production as a consequence of the highly nonlinear transformation. It is worth noting that the larger MPS and MPS-HEPS spreads are attributable to the temporal dispersion of the hourly simulated

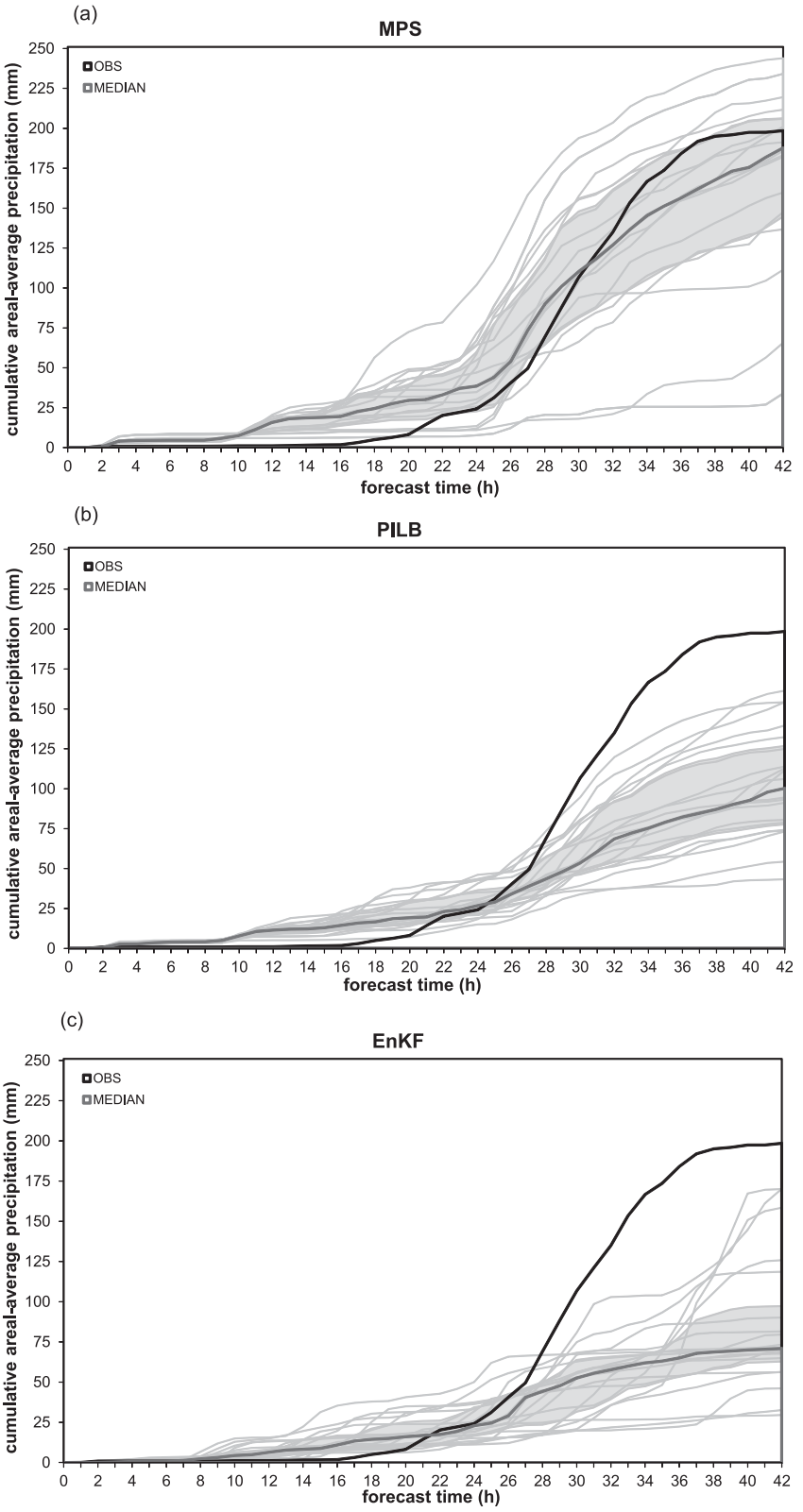

FIG. 8. Cumulative area-averaged precipitation over the Serpis River basin (802.6 km^2) for the (a) MPS, (b) PILB, and (c) EnKF ensembles. The ensemble members are shown as thin gray lines. The ensemble median is denoted as a thick gray line. The gray shading represents the p_{25}–p_{75} interquartile range.

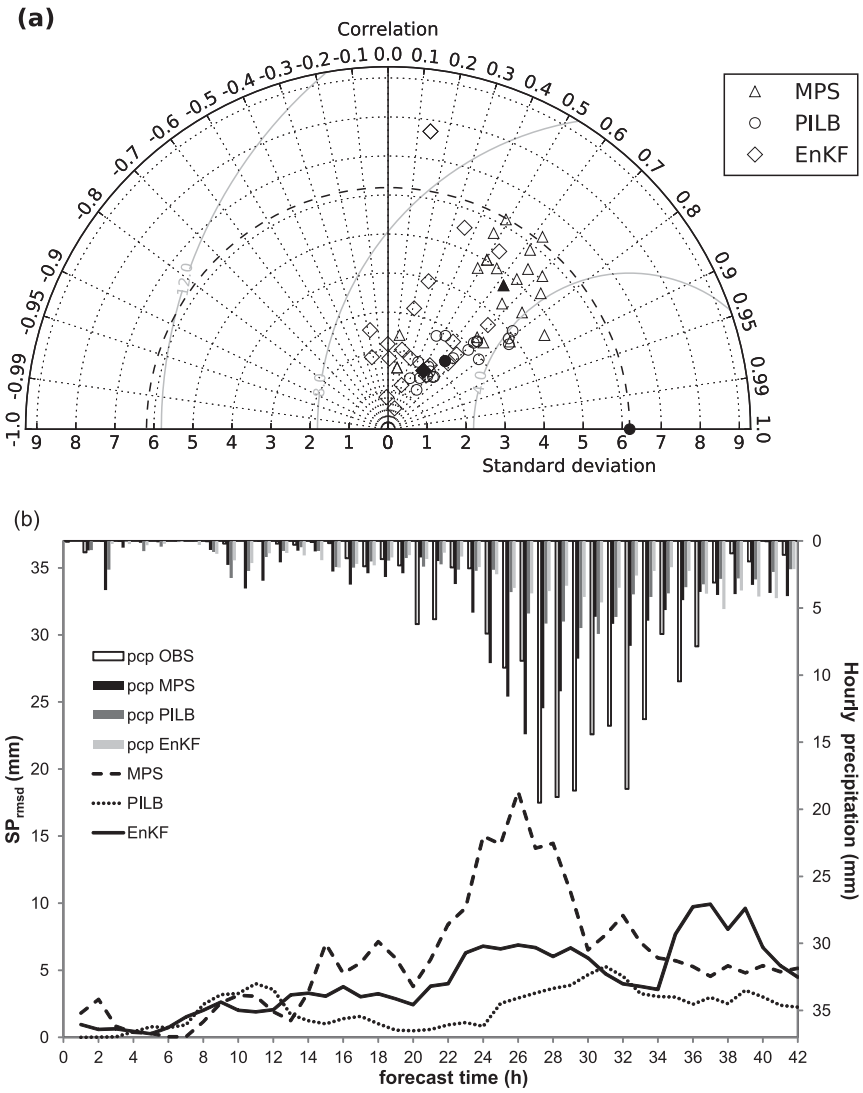

FIG. 9. (a) Taylor diagrams of the MPS, PILB, and EnKF ensembles for the hourly area-averaged precipitation over the Serpis River basin and from 0000 UTC 11 Oct to 1800 UTC 12 Oct. Filled black symbols denote ensemble medians. (b) Ensemble spread evolution of the different strategies for the hourly simulated rainfalls from 0000 UTC 11 Oct to 1800 UTC 12 Oct over the Serpis River basin. Dashed, dotted, and continuous black lines denote the MPS, PILB, and EnKF ensembles, respectively. Hourly observed and ensemble mean precipitation is shown as vertical bars and std devs (mm) and centered RMS differences (mm) are given.

convectively driven rainfalls over different forecast time steps, producing uneven initializations of the storm-driven flows (Figs. 9b, 13b).

Finally, we provide additional arguments to illustrate the usefulness of accounting for advanced probabilistic ensemble prediction systems when dealing with flash floods over Mediterranean Spain. For flood mitigation purposes, CHJ established official warning discharge thresholds at $Q_{\mathrm{CHJ}} = 20$ and $30\,\mathrm{m^3\,s^{-1}}$ for Vernissa and Carrós, respectively. These streamflow safety levels are used to preventively monitor the subsequent evolution of the river flows. When further increases in river discharges become important, the protocol to alert civil protection authorities is triggered. In addition to these administrative warning levels, we consider two additional predetermined thresholds. These alerts correspond to discharge return periods of 5 and 10 years at the different hydrometric sections (Q_{p_5} and $Q_{p_{10}}$; Table 6), as derived from prior flood risk assessments analysis completed by the CHJ (MAGRAMA 2011). Peak discharge exceedance probabilities $[P(Q > q)]$ for these supplementary warning thresholds quantify the risk of facing

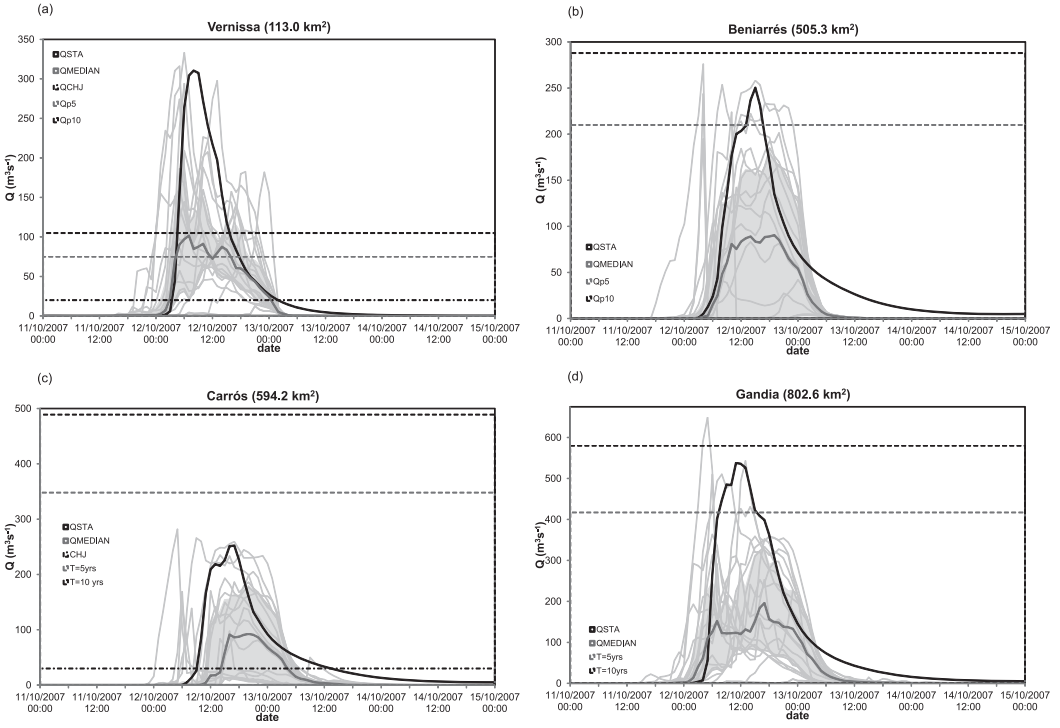

FIG. 10. MPS-HEPS at (a) Vernissa, (b) Beniarrés, and (c) Carrós stream gauges and (d) Serpis outlet. The thick black and gray solid lines correspond to the observation-driven and ensemble median discharges, respectively. The hydrological ensemble members are shown as thin gray lines. The gray shading represents the p_{25}–p_{75} interquartile range. Gray and black horizontal dashed lines show Q_{p_5} and $Q_{p_{10}}$, respectively. Note that dotted and dashed black lines at Vernissa and Carrós correspond to the pre-alert monitoring flows set by the CHJ hydraulic division.

infrequent hazardous situations like the 12 October 2007 flash flood (Figs. 10–12). For the sake of a reference when discussing these values, AEMET issues a warning when the probability of occurrence of extreme weather exceeds 0.20.

In regard to the CHJ warning levels, $[P(Q > Q_{\mathrm{CHJ}})]$ are 0.90, 0.65, and 0.30 for the MPS-, PILB-, and EnKF-HEPS strategies, respectively, at Vernissa (Table 6). Similarly, peak discharge exceedance probabilities are 0.85, 0.55, and 0.30 at Carrós for the MPS-, PILB-, and EnKF-HEPSs, respectively. Furthermore, MPS-HEPS issues unequivocal probabilities of 0.85 and 0.75 at Vernissa for $P(Q > Q_{p_5})$ and $P(Q > Q_{p_{10}})$, but not exceeding 0.30 for $P(Q > Q_{p_5})$ at Beniarrés and Gandia. In line with previous results, lower peak discharge exceedance probabilities are found at the remaining locations for all HEPS experiments, and none of the ensemble strategies render high $P(Q > q)$ for Q_{p_5} and $Q_{p_{10}}$ at Carrós (Table 6). Indeed, observed peak discharges also fall far short of both thresholds.

Even with the aforementioned inaccuracies, the three distinct hydrological ensemble strategies have proven useful for conveying proper information to civil protection and emergency decision-makers up to 42 h ahead before this particular natural hazard. However, only the multiple physics strategy points clearly at its extreme character.

7. Conclusions and further remarks

The Spanish Mediterranean lands are persistently affected by hazardous flash floods. Many small- to medium-sized basins are highly populated and often dry or with very low flows during the warm season. As return periods for damaging events over individual catchments go beyond the decade, the risks associated with these ephemeral streams are often forgotten. Sudden increases in flow rates, rapid flow velocities, and high peak discharges can cause sudden and substantial damage to human life and property.

Furthermore, strong nonlinearities emerge in the hydrological responses of these semiarid catchments to intense rainfalls, being an arduous task to properly characterize them to yield accurate hydrological forecasts. After the warm season, predominantly low initial soil moisture contents together with high soil moisture storage capacities owing to karstified substrates result in low runoff ratios and differential spatial responses.

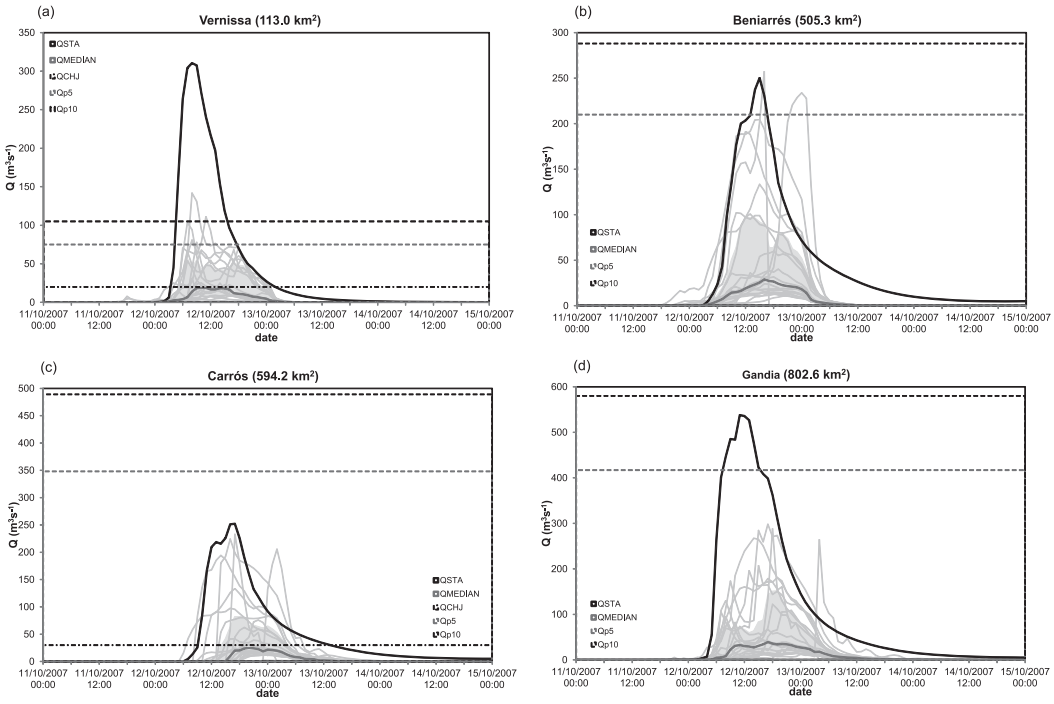

FIG. 11. As in Fig. 10, but for the PILB-HEPS.

Additional uncertainties arise from the highly complex rainfall–runoff transformation before heavy precipitation and large rainfall accumulations. To cope with these uncertainties, we have adopted the calibration procedure first introduced by Borga et al. (2007).

On the other hand, the use of convection-permitting NWP models allows us to simulate the triggering and subsequent development of convectively driven precipitation effectively and to further extend quantitative discharge forecasting beyond the concentration time of flash flood–prone basins. However, forecasts of quantitative convectively driven precipitation are remarkably challenging. The degrees of freedom to initialize the system increase and nonlinear processes dominate crucial aspects such as the location and intensity of the QPFs. Indeed, the imperfect representation of the responsible atmospheric processes and chaotic forecast sensitivity to misrepresentations of the preceding atmospheric states strongly penalize the forecast products. Within this context, HEPSs have arisen as valuable tools for encompassing these external-scale uncertainties, producing probabilistic forecasts of flash flood occurrence.

In line with the major scientific goal of the HyMeX program of improving flood early warning procedures and mitigation measures, we have first examined in depth the hydrological response of the Serpis River basin to heavy rainfalls. Next, we have evaluated the predictive skill of three distinct HEPSs strategies for the 12 October 2007 flash flood. Regarding the SREPS design, the PILB ensemble accounts for uncertainties in the atmospheric conditions causing the flash flood. PILB was generated by dynamically downscaling the 20 ECMWF-EPS members with maximum IC/LBC dispersions over the area of interest. The MPS ensemble encompasses only inaccuracies in model physical parameterizations. MPS was built from several combinations of equally skillful moist microphysical and PBL schemes. Finally, an ensemble DA technique uses the same design as a state-of-the-art EnKF ensemble, coping with both sources of external-scale uncertainties.

The main conclusions of this work are as follows:

- A multisite and novel calibration approach for the hydrologic model parameters was necessary to properly encompass the response of the semiarid and karstic Serpis River basin to heavy precipitation. After this, FEST-WB successfully reproduced the basin response to the 12 October 2007 flash flood. As the physio- and hydrographic features of this catchment are similar to many other river basins over the Mediterranean, this calibration method could be extended to these for further testing and application.
- Regarding the verification of the QPFs, MPS and EnKF strategies show indistinguishable skill at producing moderate accumulations for the larger scales over 42 h. However, just the pure physics diversity

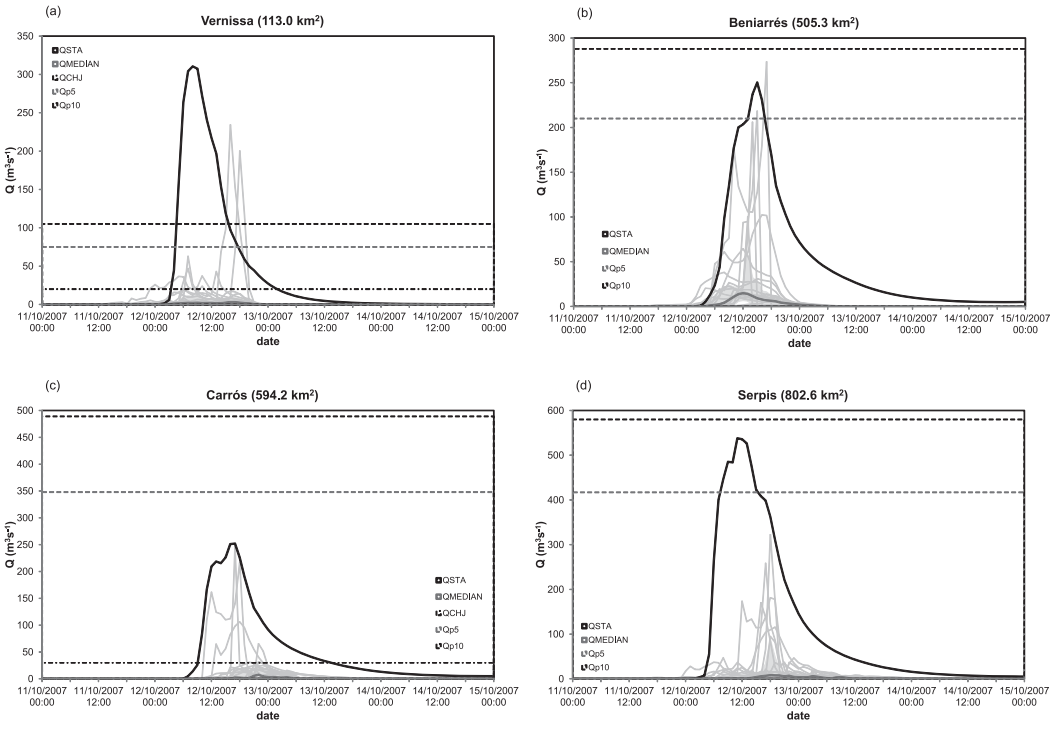

FIG. 12. As in Fig. 10, but for the EnKF-HEPS.

ensemble is proven to produce overall better forecasts in terms of the highest precipitation rates and accumulated amounts over the entire area and, in particular, over the Serpis River basin.

- Subsequent MPS-driven QDFs have shown remarkably high peak discharges and runoff volumes, exhibiting relatively large peak discharge exceedance probabilities. In summary, MPS-HEPS results depict more precisely the extreme nature of this episode and have provided the most robust prediction tool for early warning procedures. Certainly, when initial analyses provide a sufficiently accurate synoptic-scale environment, only sampling the uncertainties in the most relevant physical processes for deep moist convection renders the best QPF guidance to HEPS.
- Although EnKF-HEPS and, to a lesser extent, PILB-HEPS have fallen short of producing reliable probabilities of relevant runoff levels, both have provided an indication of plausible hazardous scenarios. Indeed, the three experimental HEPSs are proven valuable

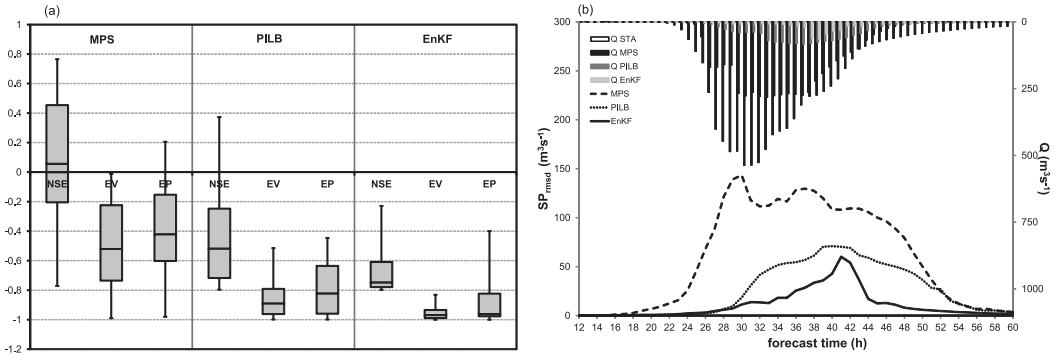

FIG. 13. (a) Statistical scores for the MPS-HEPS, PILB-HEPS, and EnKF-HEPS experiments. The boxes denote the p_{25} and p_{75} interquartile ranges, the middle horizontal lines show the ensemble median, and the whiskers display the best and the worst ensemble members. (b) Ensemble spread evolution of the different strategies for the hourly simulated discharges. Dashed, dotted, and continuous black lines denote the MPS-HEPS, PILB-HEPS, and EnKF-HEPS, respectively. Hourly observation- and ensemble mean–driven discharges are shown as vertical bars.

TABLE 6. Peak discharge exceedance probabilities [$P(Q > q)$] for the MPS-HEPS, PILB-HEPS, and EnKF-HEPS experiments at the indicated stream gauges. Variables Q_{p_5} and $Q_{p_{10}}$ denote peak discharge exceedance probabilities of the 5- and 10-yr return periods, respectively, and Q_{CHJ} indicates the pre-alert monitoring flows set by the CHJ hydraulic division.

Hydrometric section	Peak discharge	$P(Q > q)$ MPS-HEPS	$P(Q > q)$ PILB-HEPS	$P(Q > q)$ EnKF-HEPS
Vernissa	$Q_{CHJ} = 20\ m^3\ s^{-1}$	0.90	0.65	0.30
	$Q_{p_5} = 75\ m^3\ s^{-1}$	0.85	0.25	0.10
	$Q_{p_{10}} = 105\ m^3\ s^{-1}$	0.75	0.10	0.10
Beniarrés	$Q_{p_5} = 210\ m^3\ s^{-1}$	0.30	0.10	0.10
	$Q_{p_{10}} = 288\ m^3\ s^{-1}$	<0.05	<0.05	<0.05
Carrós	$Q_{CHJ} = 30\ m^3\ s^{-1}$	0.85	0.55	0.30
	$Q_{p_5} = 348\ m^3\ s^{-1}$	<0.05	<0.05	<0.05
	$Q_{p_{10}} = 489\ m^3\ s^{-1}$	<0.05	<0.05	<0.05
Gandia	$Q_{p_5} = 417\ m^3\ s^{-1}$	0.25	<0.05	<0.05
	$Q_{p_{10}} = 580\ m^3\ s^{-1}$	0.05	<0.05	<0.05

within a realistic civil protection management framework: all exceeded the official CHJ warning thresholds at Vernissa and Carrós. This simple application enlightens the benefits of using advanced ensemble strategies to surpass predetermined warning levels before extreme floods.

Obviously, this single case study does not allow us to reach general conclusions about the predictability of this type of event or about the optimal hydrometeorological forecasting strategy in an operational framework, but it clearly points out important aspects to take into account in future statistical studies. The 12 October 2007 flash flood is a prototype of long-lasting and orographically driven convective systems that are responsible for the most hazardous flash floods over the western Mediterranean. The predictability analysis of this paradigmatic event allowed intercomparing the performance of three competitive and popular ensemble strategies and discerning the detailed differences among them. This particular event may allegedly exemplify the paradigm transition toward a more convectively centered approach for the search of relevant sources of uncertainty in hazardous flash flood events. Future directions point toward the confirmation of the spatial—over different basins—and temporal—over a number of flash flood events—generality of the presented findings before these can be safely and confidently transferred to operations and breakthrough improvements in risk management strategy development can be achieved.

Acknowledgments. Dr. Faisal Hossain, scientific editor of the *Journal of Hydrometeorology*, and three anonymous reviewers are deeply acknowledged for their valuable comments that helped to improve the quality of this manuscript. The Confederación Hydrográfica del Júcar (CHJ), and especially Onofre Gabaldó, head of the hydraulic division and Carmen González, from the SAIH operational center, are acknowledged for providing data. The Spanish Agency of Meteorology (AEMET) is also acknowledged for providing data from the automatic weather stations. This work has been sponsored by CGL2011-24458 [Improving Short-Range Mediterranean Severe-Weather Forecasts by Means of Adaptive Observations and Advanced Ensemble Methods in Mediterranean Experiment (MEDEX) Phase II and HyMeX (PREDIMED)] and CGL2014- 52199-R [Future Regional Impacts of Climate Change Associated to Extreme Weather Phenomena (EXTREMO)] Spanish projects, which are partially supported by Fondo Europeo de Desarrollo Regional (FEDER) and Secretaría de Estado de Investigación, Desarrollo e Innovación. Author D. Carrió received funding from the Conselleria d'Innovació, Recerca i Turisme del Govern de les Illes Balears and FEDER (Decision 18607). High-resolution numerical forecast experiments were carried out at BSC MareNostrum III supercomputer under Red Española de Supercomputación (RES) virtual infrastructure.

REFERENCES

Addor, N., S. Jaun, F. Fundel, and M. Zappa, 2011: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.*, **15**, 2327–2347, doi:10.5194/hess-15-2327-2011.

Akima, H., 1978: A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Trans. Math. Software*, **4**, 148–164, doi:10.1145/355780.355786.

——, 1996: Algorithm 761: Scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans. Math. Software*, **22**, 362–371, doi:10.1145/232826.232856.

Amengual, A., R. Romero, M. Gómez, A. Martín, and S. Alonso, 2007: A hydrometeorological modeling study of a flash flood event over Catalonia, Spain. *J. Hydrometeor.*, **8**, 282–303, doi:10.1175/JHM577.1.

——, ——, and S. Alonso, 2008: Hydrometeorological ensemble simulations of flood events over a small basin of Majorca Island, Spain. *Quart. J. Roy. Meteor. Soc.*, **134**, 1221–1242, doi:10.1002/qj.291.

——, ——, M. Vich, and S. Alonso, 2009: Inclusion of potential vorticity uncertainties into a hydrometeorological forecasting chain: Application to a medium size basin of Mediterranean Spain. *Hydrol. Earth Syst. Sci.*, **13**, 793–811, doi:10.5194/hess-13-793-2009.

——, V. Homar, and O. Jaume, 2015: Potential of a probabilistic hydrometeorological forecasting approach for the 28 September 2012 extreme flash flood in Murcia, Spain. *Atmos. Res.*, **166**, 10–23, doi:10.1016/j.atmosres.2015.06.012.

Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.

Angevine, W. M., H. Jiang, and T. Mauritsen, 2010: Performance of an eddy diffusivity–mass flux scheme for shallow cumulus boundary layers. *Mon. Wea. Rev.*, **138**, 2895–2912, doi:10.1175/2010MWR3142.1.

Borga, M., P. Boscolo, F. Zanon, and M. Sangati, 2007: Hydrometeorological analysis of the 29 August 2003 flash flood in the eastern Italian Alps. *J. Hydrometeor.*, **8**, 1049–1067, doi:10.1175/JHM593.1.

Bossard, M., J. Feranec, and J. Otahel, 2000: CORINE land cover technical guide—Addendum 2000. Tech. Rep. 40, European Environment Agency, 105 pp. [Available online at http://image2000.jrc.ec.europa.eu/reports/corine_tech_guide_add.pdf.]

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416, doi:10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2.

Buizza, R., 2003: Weather prediction: Ensemble prediction. *Encyclopedia of Atmospheric Sciences*, Academic Press, 2546–2557, doi:10.1016/B0-12-227090-8/00461-9.

——, and T. N. Palmer, 1995: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456, doi:10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

Camarasa Belmonte, A. M., and F. Segura Beltrán, 2001: Flood events in Mediterranean ephemeral streams (ramblas) in Valencia region, Spain. *Catena*, **45**, 229–249, doi:10.1016/S0341-8162(01)00146-1.

Carrió, D. S., and V. Homar, 2016: Potential of sequential EnKF for the short-range prediction of a maritime severe weather event. *Atmos. Res.*, **178–179**, 426–444, doi:10.1016/j.atmosres.2016.04.011.

Clarke, A. D., T. Uehara, and J. N. Porter, 1997: Atmospheric nuclei and related aerosol fields over the Atlantic: Clean subsiding air and continental pollution during ASTEX. *J. Geophys. Res.*, **102**, 25 281–25 292, doi:10.1029/97JD01555.

Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, doi:10.1016/j.jhydrol.2009.06.005.

Coniglio, M. C., J. Correia Jr., P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF Model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, doi:10.1175/WAF-D-12-00103.1.

Delgado, J., J. A. Peláez, R. Tomás, A. Estévez, C. López Casado, C. Doménech, and A. Cuenca, 2006: Evaluación de la susceptibilidad de las laderas a sufrir inestabilidades inducidas por terremotos: Aplicación a la cuenca de drenaje del río Serpis (provincia de Alicante). *Rev. Soc. Geol. Esp.*, **19**, 197–218.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, doi:10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005, doi:10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.

Drobinski, P., and Coauthors, 2014: HyMeX: A 10-year multidisciplinary program on the Mediterranean water cycle. *Bull. Amer. Meteor. Soc.*, **95**, 1063–1082, doi:10.1175/BAMS-D-12-00242.1.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459, doi:10.1175/1520-0493(1997)125<2427:SREFOQ>2.0.CO;2.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Evans, J. P., M. Ekström, and F. Ji, 2012: Evaluating the performance of a WRF physics ensemble over south-east Australia. *Climate Dyn.*, **39**, 1241–1258, doi:10.1007/s00382-011-1244-5.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

Grimit, E. P., and C. F. Mass, 2007: Measuring the ensemble spread–error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Wea. Rev.*, **135**, 203–221, doi:10.1175/MWR3262.1.

Guillijns, S., O. B. Mendoza, J. Chandrasekar, B. L. R. De Moor, D. S. Bernstein, and A. Ridley, 2006: What is the ensemble Kalman filter and how well does it work? *2006 American Control Conf.*, Minneapolis, MN, IEEE, 6 pp., doi:10.1109/ACC.2006.1657419.

Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42** (2), 129–151.

——, N. Yign, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.

Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181–2196, doi:10.1175/1520-0493(1995)123<2181:MFEP>2.0.CO;2.

——, and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1843, doi:10.1175/2010JAMC2432.1.

Hudson, J. G., 1993: Cloud condensation nuclei. *J. Appl. Meteor.*, **32**, 596–607, doi:10.1175/1520-0450(1993)032<0596:CCN>2.0.CO;2.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

Jankov, I., W. A. Gallus Jr., M. Segal, B. Shaw, and S. E. Koch, 2005: The impact of different WRF Model physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting*, **20**, 1048–1060, doi:10.1175/WAF888.1.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267, doi:10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kolios, S., and H. Feidas, 2010: Warm season climatology of mesoscale convective systems in the Mediterranean Basin using satellite data. *Theor. Appl. Climatol.*, **102**, 29–42, doi:10.1007/s00704-009-0241-7.

MAGRAMA , 2011: Mapa de caudales máximos. Tech. Memo., Ministerio de Agricultura, Alimentacion y Medio Ambiente, Madrid, Spain, 67 pp. [Available online at http://www.mapama.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/memoria_tecnica_v2_junio2011_tcm7-162773.pdf.]

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, doi:10.1175/2009JAS2965.1.

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine, 2014: An investigation of the Goshen County, Wyoming, tornadic supercell of 5 June 2009 using EnKF assimilation of mobile mesonet and radar observations collected during VORTEX2. Part I: Experiment design and verification of the EnKF analyses. *Mon. Wea. Rev.*, **142**, 530–554, doi:10.1175/MWR-D-13-00007.1.

Marsigli, C., 2009: COSMO Priority Project ''Short Range Ensemble Prediction System'' (SREPS): Final report. COSMO Tech. Rep. 13, Consortium for Small Scale Modelling, 33 pp. [Available online at http://www2.cosmo-model.org/content/model/documentation/techReports/docs/techReport13.pdf.]

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:10.1029/97JD00237.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, doi:10.1002/qj.49712252905.

Montaldo, N., V. Toninelli, J. D. Albertson, M. Mancini, and P. A. Troch, 2003: The effect of background hydrometeorological conditions on the sensitivity of evapotranspiration to model parameters: Analysis with measurements from an Italian alpine catchment. *Hydrol. Earth Syst. Sci.*, **7**, 848–861, doi:10.5194/hess-7-848-2003.

——, G. Ravazzani, and M. Mancini, 2007: On the prediction of the Toce alpine basin floods with distributed hydrologic models. *Hydrol. Processes*, **21**, 608–621, doi:10.1002/hyp.6260.

Mullen, S. L., and D. P. Baumhefner, 1988: Sensitivity to numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations. *Mon. Wea. Rev.*, **116**, 2289–2329, doi:10.1175/1520-0493(1988)116<2289:SONSOE>2.0.CO;2.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level 3 model: Its numerical stability and application to a regional prediction of advecting fog. *Bound.-Layer Meteor.*, **119**, 397–407, doi:10.1007/s10546-005-9030-8.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models. Part I: A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.

Pastor, F., I. Gómez, and M. J. Estrela, 2010: Numerical study of the October 2007 flash flood in the Valencia region (eastern Spain): The role of orography. *Nat. Hazards Earth Syst. Sci.*, **10**, 1331–1345, doi:10.5194/nhess-10-1331-2010.

Pielke, R., and Y. Mahrer, 1975: Representation of the heated planetary boundary layer in mesoscale models with coarse vertical resolution. *J. Atmos. Sci.*, **32**, 2288–2308, doi:10.1175/1520-0469(1975)032<2288:ROTHPB>2.0.CO;2.

Ponce, V. M., and R. H. Hawkins, 1996: Runoff curve number: Has it reached maturity? *J. Hydrol. Eng.*, **1**, 11–19, doi:10.1061/(ASCE)1084-0699(1996)1:1(11).

Puigdefabregas, J., G. del Barrio, M. M. Boer, L. Gutiérrez, and A. Solé, 1998: Differential responses of hillslope and channel elements to rainfall events in a semi-arid area. *Geomorphology*, **23**, 337–351, doi:10.1016/S0169-555X(98)00014-2.

Rabuffetti, D., G. Ravazzani, C. Corbari, and M. Mancini, 2008: Verification of operational quantitative discharge forecast (QDF) for a regional warning system—The AMPHORE case studies in the upper Po River. *Nat. Hazards Earth Syst. Sci.*, **8**, 161–173, doi:10.5194/nhess-8-161-2008.

Ravazzani, G., M. Mancini, I. Giudici, and P. Amadio, 2007: Effects of soil moisture parameterization on a real-time flood forecasting system based on rainfall thresholds. *Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management*, E. Boegh et al., Eds., IAHS Publ. 313, 407–416.

——, C. Corbari, S. Morella, P. Gianoli, and M. Mancini, 2012: Modified Hargreaves–Samani equation for the assessment of reference evapotranspiration in Alpine river basins. *J. Irrig. Drain. Eng.*, **138**, 592–599, doi:10.1061/(ASCE)IR.1943-4774.0000453.

——, P. Gianoli, S. Meucci, and M. Mancini, 2014: Assessing downstream impacts of detention basins in urbanized river basins using a distributed hydrological model. *Water Resour. Manage.*, **28**, 1033–1044, doi:10.1007/s11269-014-0532-3.

——, A. Amengual, A. Ceppi, V. Homar, R. Romero, G. Lombardia, and M. Mancini, 2016: Potentialities of ensemble strategies for flood forecasting over the Milano urban area. *J. Hydrol.*, **539**, 237–253, doi:10.1016/j.jhydrol.2016.05.023.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/2007MWR2123.1.

Romero, R., J. A. Guijarro, C. Ramis, and S. Alonso, 1998: A 30-year (1964–1993) daily rainfall data base for the Spanish Mediterranean regions: First exploratory study. *Int. J. Climatol.*, **18**, 541–560, doi:10.1002/(SICI)1097-0088(199804)18:5<541::AID-JOC270>3.0.CO;2-N.

Scherrer, S. C., C. Appenzeller, P. Eckert, and D. Cattani, 2004: Analysis of the spread–skill relations using the ECMWF ensemble prediction system over Europe. *Wea. Forecasting*, **19**, 552–565, doi:10.1175/1520-0434(2004)019<0552:AOTSRU>2.0.CO;2.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:10.1175/2009WAF2222267.1.

Sippel, J. A., S. A. Braun, F. Zhang, and Y. Weng, 2013: Ensemble Kalman filter assimilation of simulated HIWRAP Doppler velocity data in a hurricane. *Mon. Wea. Rev.*, **141**, 2683–2704, doi:10.1175/MWR-D-12-00157.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, doi:10.1175//2555.1.

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. 2nd ed., Research Rep. MSRB 89-5, WWW Tech. Rep. 8, WMO/TD 358, World Meteorological Organization. [Available online at http://www.cawcr.gov.au/projects/verification/Stanski_et_al/Stanski_et_al.html.]

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial and model physics perturbations in short-range ensemble simulations of mesoscale convective events. *Mon. Wea. Rev.*, **128**, 2077–2107, doi:10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

Tanamachi, R. L., L. J. Wicker, D. C. Dowell, H. B. Bluestein, D. T. Dawson, and M. Xue, 2013: EnKF assimilation of high-resolution, mobile Doppler radar data of the 4 May 2007 Greensburg, Kansas, supercell into a numerical cloud model. *Mon. Wea. Rev.*, **141**, 625–648, doi:10.1175/MWR-D-12-00099.1.

Tapiador, F. J., W. K. Tao, J. J. Shi, C. F. Angelis, M. A. Martinez, C. Marcos, A. Rodríguez, and A. Hou, 2012: A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain. *J. Appl. Meteor. Climatol.*, **51**, 489–504, doi:10.1175/JAMC-D-11-041.1.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, doi:10.1029/2000JD900719.

Tewari, M., and Coauthors, 2004: Implementation and verification of the unified NOAH land surface model in the WRF Model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 14.2a. [Available online at https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm.]

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

USDA, 1986: Urban hydrology for small watersheds. USDA TR-55, 164 pp. [Available online at https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044171.pdf.]

Verbunt, M., A. Walser, J. Gurtz, A. Montani, and C. Schär, 2007: Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies. *J. Hydrometeor.*, **8**, 897–909, doi:10.1175/JHM594.1.

Vincendon, B., V. Ducrocq, O. Nuissier, and B. Vié, 2011: Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting. *Nat. Hazards Earth Syst. Sci.*, **11**, 1529–1544, doi:10.5194/nhess-11-1529-2011.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548, doi:10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 648 pp.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zheng, Y., K. Alapaty, J. A. Herwehe, A. D. Del Genio, and D. Niyogi, 2016: Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) Model with an updated Kain–Fritsch scheme. *Mon. Wea. Rev.*, **144**, 833–860, doi:10.1175/MWR-D-15-0005.1.